

# Immunological Data Standards & Repositories

R. BURKE SQUIRES



This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

#### You are free to:

**Share** — copy and redistribute the material in any medium or format  
**Adapt** — remix, transform, and build upon the material  
 for any purpose, even commercially.



The licensor cannot revoke these freedoms as long as you follow the  
 license terms.

#### Under the following terms:

- Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

#### Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity](#), [privacy](#), or [moral rights](#) may limit how you use the material.

Please Note: parts of this work have been borrowed from other individuals and may contain a different license.

# Objectives

Enable you to name one or more immunologically relevant data standards

Enable you to search for immunologically relevant data repositories

Enable you to take a(nother) step in becoming a reproducible scientist

# Outline

## Immunological Data

- Standards
- Repositories

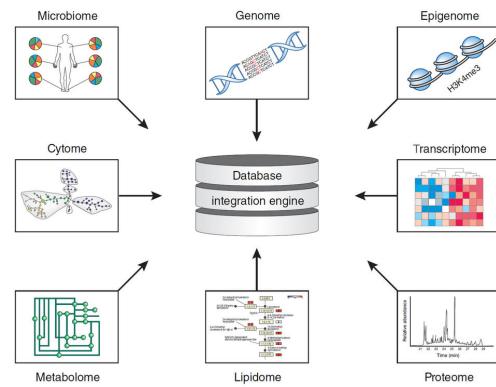


Image: Schultze, J. L. Teaching 'big data' analysis to young immunologists. *Nature Immunology* **16**, 902–905 (2015).

## Immunological Data

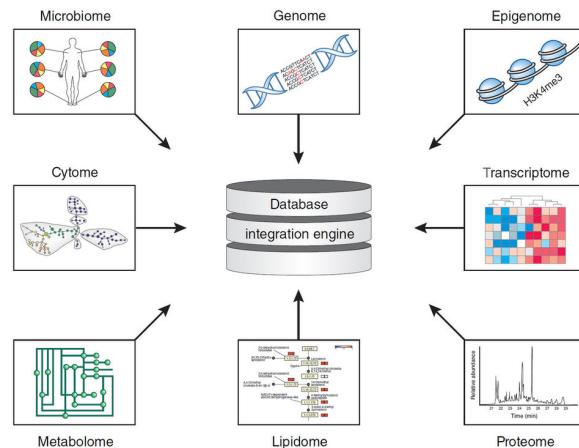
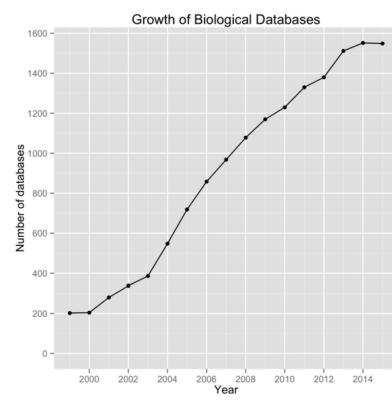
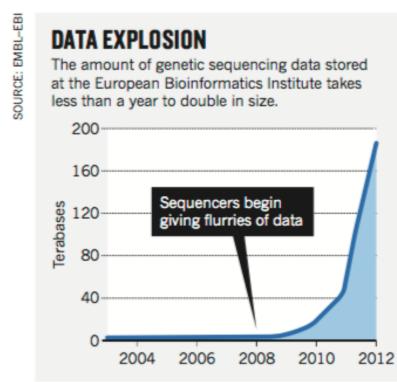


Image: Schultze, J. L. Teaching 'big data' analysis to young immunologists. *Nature Immunology* 16, 902–905 (2015).

## Growth in Data and Databases



Number of databases in the NAR database issue, up to 2015 (from @AlexBateman1)

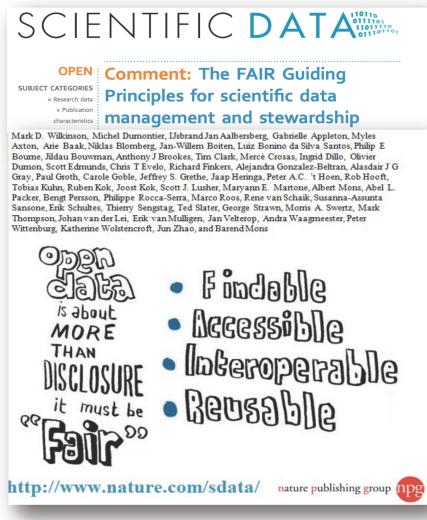
## Better Data = Better Science (Five Reasons To Protect Your Data)



<http://blogs.lse.ac.uk/impactofsocialsciences/2014/01/21/five-reasons-to-protect-your-scientific-data/>

## How Do We “Get” Better Data

# The FAIR Principles



 [biorxiv.org](http://www.biorxiv.org)  
Information Resources

Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018 (2016).

## But in all fairness, not all data is FAIR!

*Nature Genetics* 41, 149 - 155 (2009)  
Published online: 28 January 2008 | doi:10.1038/ng.295 

### Repeatability of published microarray gene expression analyses

See associated Correspondence: Baggerly, *Nature* 467, 401 (September 2010)

John P A Ioannidis<sup>1,2,3</sup>, David B Allison<sup>4</sup>, Catherine A Ball<sup>5</sup>, Issa Coulibaly<sup>6</sup>, Xianggui Cui<sup>2</sup>, Aedin C Culhane<sup>6,7</sup>, Mario Falchi<sup>8,9,10</sup>, Cesare Furiniello<sup>11,9</sup>, Laurence Game<sup>11</sup>, Giuseppe Jurman<sup>10</sup>, Jon Mangion<sup>11</sup>, Tapan Mehta<sup>9</sup>, Michael Nitzberg<sup>9</sup>, Grier P Page<sup>6,12</sup>, Enrico Petretto<sup>11,13</sup> & Vera van Noort<sup>13</sup>

Given the complexity of microarray-based gene expression studies, guidelines encourage transparent design and public data availability. Several journals require public data deposition and several public databases exist. However, not all data are publicly available, and even when available, it is unknown whether the published results are reproducible by independent scientists. Here we evaluated the replication of data analyses in 18 articles on

The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis. Repeatability of published microarray studies is apparently limited. More strict publication rules enforcing public data availability and explicit description of data processing and analysis should be considered.

 [biorxiv.org](http://www.biorxiv.org)  
Information Resources

# We Are Talking About Reproducible Science Here!

---

## The Need For Data Standards

---

What do we mean by standards

Why have them?

Where can you find data standards?

What are some immunologically relevant data standards?

## Defining Data Standards

"Agreed-upon conventions for doing 'something', established by community consensus or an authority, e.g. managing a process or delivering a service"

- Susanna-Assunta Sansone, PhD

## Data Standards: Interoperability Standards

Nuts and Bolts

Standardization initiated in 1864 by William Sellers

Only widely adopted a century later!



Susanna-Assunta Sansone, PhD

# Data Standards: Interoperability Standards

- Agreed-upon specifications, guidelines or criteria designed to ensure data and any other digital object (such as code, algorithms, workflows, models, software, or journal articles) are FAIR

## FAIR

- Findable
- Accessible
- Interoperable
- Reusable



Susanna-Assunta Sansone, PhD

Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016).

# Data Standards: Interoperability Standards – Fundamentals

Essential for **discovery** of data and any other digital object, but also **citation** and **credit**

Enable the operational processes, underlying their (re)use

- such as exchange, aggregation, integration, comparison

**Optimal interoperability** is reached when all processes are automated for both human and machine, this **requires**

- metadata:** or descriptors for the digital objects to understand what it is, where to find it, how to access it etc.
- identifiers:** unique, resolvable and versionable identifiers are essential elements of the digital world

Susanna-Assunta Sansone, PhD

## Data Standards: Interoperability Standards – Realizing Potential

**Identifiers and metadata** must be implemented by technical experts in tools, registries, catalogues, databases, services

- to find, store, manage (e.g., mint, track provenance, version) and aggregate (e.g., interlink and map etc.) these digital objects.

**Implementations** are essential to make standards ‘**invisible**’ to users, such as researchers, who often have little or no familiarity with them

Susanna-Assunta Sansone, PhD

## Data Standards: Metadata Standards – Fundamentals

**Metadata - data about the data!**

**Descriptors** for a digital object that help to understand what it is, where to find it, how to access it etc.

The type of metadata depends also on the digital object

The **depth and breadth** of metadata varies according to their purpose (discovery, citation, credit)

- e.g. reproducibility requires richer metadata

Susanna-Assunta Sansone, PhD

## What is Wrong With This Data?

(Notice I did not ask IF anything is wrong with this data...)

	A	B	C	D	E
1		Group1	Group2		
2		Day 0			
3	Sodium	139	142		
4	Potassium	3.3	4.8		
5	Chloride	100	108		
6	BUN	18	18		
7	Creatine	1.2	1.2		
8	Uric acid	5.5*	6.2*		
9		Day 7			
10	Sodium	140	146		
11	Potassium	3.4	5.1		
12	Chloride	97	108		

Credit to: Iain Hrynaszkiewicz

.... which this isn't...

	S1Sh.cuo			
	A	B	C	D
1		Group1	Group2	
2		Day 0		
3	Sodium	139	142	
4	Potassium	3.3	4.8	
5	Chloride	100	108	
6	BUN	18	18	
7	Creatine	1.2	1.2	
8	Uric acid	5.5*	6.2*	
9		Day 7		
10	Sodium	140	146	
11	Potassium	3.4	5.1	
12	Chloride	97	108	

Credit to: Iain Hrynaszkiewicz

# Data Standards: Metadata Standards – Fundamentals

....this is much clearer!

Table_S1_Shanghai_blood.xls						
	A	B	C	D	E	F
1	Parameter	Day	Control	Treated	Units	P
2	Sodium	0	139	142	mEq/l	0.82
3	Sodium	7	140	146	mEq/l	0.70
4	Sodium	14	140	158	mEq/l	0.03
5	Sodium	21	143	160	mEq/l	0.02
6	Potassium	0	3.3	4.8	mEq/l	0.06
7	Potassium	7	3.4	5.1	mEq/l	0.07
8	Potassium	14	3.7	4.7	mEq/l	0.10
9	Potassium	21	3.1	3.6	mEq/l	0.52
10	Chloride	0	100	108	mEq/l	0.56
11	Chloride	7	97	108	mEq/l	0.68
12	Chloride	14	101	106	mEq/l	0.79

Susanna-Assunta Sansone, PhD

# Data Standards: Searching for Standards

Where do we find them?

Biosharing.org is now FAIRsharing.org

FAIRsharing.org

- Standards
- Databases
- Policies
- Collections
- Add...

## Exercise

### STEPS

1. Go to: <https://FAIRsharing.org>

### QUESTIONS

What type of information is contained in FAIRsharing?

When was the last update?

## What Are Some Immunologically Relevant Data Standards?

Flow Cytometry Data File Format Standards (FCS 3.1)

ISAC Classification Results File Format (CLR)

ISAC standard for representing gating descriptions in flow cytometry (Gating-ML)

Minimum Information for Biological and Biomedical Investigations (MIBBI)

- Minimum Information About a Bioinformatics investigation (MIABI)
- Minimum Information about a Flow Cytometry Experiment (MIFlowCyt)

LINCS Extended Metadata Standard

- Antibody reagents
- Cell lines
- Differentiated cells
- Embryonic stem cells
- iPSCs
- Nucleic acid reagents
- Other reagents
- Primary cells
- Proteins
- Small molecules

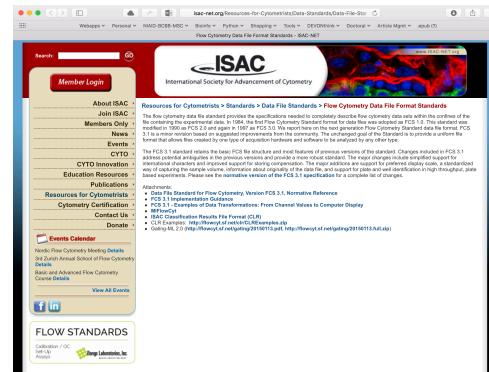
# Flow Cytometry Data File Format Standards (FCS)

The flow cytometry data file standard provides the specifications needed to completely describe flow cytometry data sets within the confines of the file containing the experimental data.

Goal of the Standard is to provide a uniform file format that allows files created by one type of acquisition hardware and software to be analyzed by any other type.

FCS Is implemented in

- FlowRepository
- The Immunology Database and Analysis Portal - OpenImmport



# ISAC Classification Results File Format (CLR)

CLR has been developed to exchange the results of manual gating and algorithmic classification approaches in a standard way in order to be able to report and process the classification.

Although it was originally designed for the field of flow cytometry, it is applicable in any domain that needs to capture either soft or unambiguous classifications of virtually any kinds of objects.

A screenshot of a PDF document titled "Classification Results File Format (CLR)". The document is from the International Society for Advancement of Cytometry (ISAC). It includes sections such as "Document Status", "Copyright and License", "Disclaimer of Liability", and "Implementation". The document specifies that it is a Candidate Recommendation and notes that it has not changed since CLR 1.0 version 110216, only the format documentation and examples have been improved. It also states that the document can be used under the terms of the Creative Commons Attribution-ShareAlike 3.0 Unported license. The document is dated September 3, 2014, and is version 1.0. A footer at the bottom right indicates it is supported by NIH/NIBMS supplement award 1R01NS080400.

<http://isac-net.org/PDFS/65/659585c7-9ef4-4152-829c-859aad25aaaf.pdf>

# NIH Library of Integrated Network-Based Cellular Signatures (LINCS)

The Library of Integrated Network-Based Cellular Signatures (LINCS) Program aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents.

## Data portal

- <http://lincsportal.ccs.miami.edu/dcic-portal/>

## Cells

- <http://lincsportal.ccs.miami.edu/cells/>

# Minimum Information About a Flow Cytometry Experiment (MIFlowCyt)

## Minimum Information about a Flow Cytometry Experiment (MIFlowCyt)

- Outlines the minimum information required to report about flow cytometry experiments
- Represents the community consensus
- 33 coauthors from 19 institutions
- ISAC Recommendation
- Required/recommended by Cytometry A and Nature

ORIGINAL ARTICLE

Cytometry



### MIFlowCyt: The Minimum Information About a Flow Cytometry Experiment

Jamie A. Lee,<sup>1,1</sup> Josef Spidlen,<sup>2</sup> Keith Boyce,<sup>3</sup> Jennifer Cai,<sup>1</sup> Nicholas Croshie,<sup>4</sup> Mark Dahlén,<sup>5</sup> Jeff Furlong,<sup>6</sup> Maura Gasparrini,<sup>7</sup> Michael Goldberg,<sup>8</sup> Elizabeth M. Goralsky,<sup>9</sup> Bill Hyun,<sup>10</sup> Kirstin Jansen,<sup>11</sup> Tobias Kollmann,<sup>10</sup> Megan Kong,<sup>12</sup> Robert Leis,<sup>13</sup> Brian McEvilly,<sup>14</sup> Thomas D. Morrissey,<sup>15</sup> Michael M. Murray,<sup>16</sup> Gary N. Palmer,<sup>17</sup> John P. Paton,<sup>18</sup> Jason McWenney,<sup>19</sup> David Pash,<sup>20</sup> Barclay Purcell,<sup>21</sup> Yu Qian,<sup>22</sup> Binnan Sun,<sup>23</sup> Christopher Smith,<sup>2</sup> Olga Tchurukina,<sup>24</sup> Anne Wertheimer,<sup>25</sup> Peter Wilkinson,<sup>26</sup> Christopher Wilson,<sup>8</sup> James Wood,<sup>22</sup> Robert Zegers,<sup>27</sup> The International Society for Advancement of Cytometry Data Standards Task Force, Richard H. Schuermann,<sup>28</sup> Ryan R. Brinkman<sup>24</sup>

<sup>1</sup> Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA  
<sup>2</sup> Terry Fox Laboratory, BC Cancer Agency, Vancouver, British Columbia, Canada  
<sup>3</sup> Cellomics, Inc., San Francisco, CA, USA  
<sup>4</sup> Immature Tolerance Network, Pittsburgh, Pennsylvania 15238  
<sup>5</sup> Vysis Technologies, Monroe 2124A, Monroe, NJ, USA  
<sup>6</sup> Mergen Inc., Thousand Oaks, California 91320  
<sup>7</sup> Department of Immunology, University of Washington, Seattle, Washington 98195  
<sup>8</sup> BD Biosciences, San Jose, California 95131

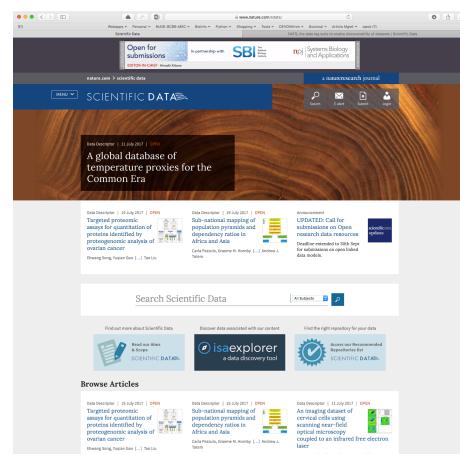
<sup>9</sup> Abstract  
An fundamental tenet of scientific research is that published results are open to independent validation and refutation. Minimum data standards aid data providers, users, and publishers by providing a specification of what is required to unambiguously interpret flow cytometry data. The Minimum Information About a Flow Cytometry Experiment (MIFlowCyt) standard, stating the minimum information required to report a flow cytometry experiment, was developed by a cross-disciplinary international collaborative group of bioinformaticians, computational statisticians, software developers, instrument manufacturers, and clinical scientists. Adoption of MIFlowCyt by the scientific and publishing communities will facilitate greater understanding and reuse of FCM data. © 2008 International Society for Advancement of Cytometry

# Data as a Primary Resource

## Data as a Primary Resource

It gets its own journal – from Nature Publishing Group!

### *Scientific Data Journal*



# Not Everyone is Happy With Data Reuse: “Research Parasites”

On January 21, 2016, the Editors of the New England Journal of Medicine published an editorial that characterized [some] scientists who re-analyzed published data sets as “research parasites”



Longo, D. L. & Drazen, J. M. Data Sharing. *N Engl J Med* 374, 276–277 (2016).

# Immunological Repositories

# Repositories

---

**DATA (META)REPOSITORIES**

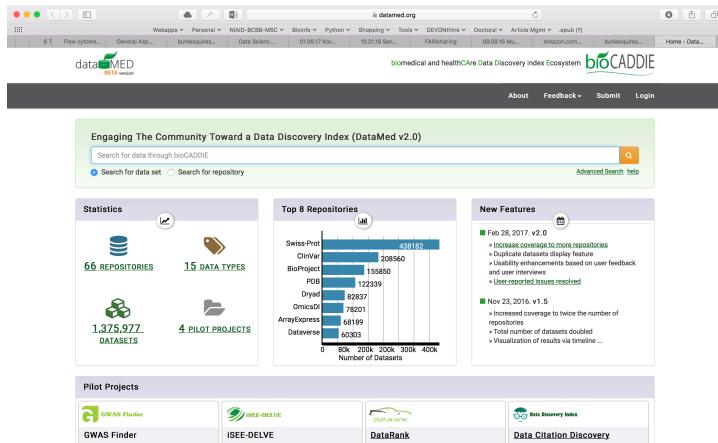
- DataMed**
  - Data repository search engine
  
- Nucleic Acid Research (NAR)**
  - Database Issues
  - Website Issues

**TOOL (META)REPOSITORIES**

- OMICSTools**
  - Immunological tools
  
- LabWorm**
  - Immunological tools

## DataMed

---



The screenshot shows the DataMed v2.0 interface. At the top, there's a search bar with the placeholder "Engaging The Community Toward a Data Discovery Index (DataMed v2.0)". Below the search bar are two radio buttons: "Search for data set" (selected) and "Search for repository". To the right of the search bar are links for "About", "Feedback", "Submit", and "Login".

Below the search bar, there are three main sections:

- Statistics:** Shows "66 REPOSITORIES" and "1,375,977 DATASETS".
- Top 8 Repositories:** A bar chart showing the number of datasets for each repository. The data is as follows:
 

Repository	Number of Datasets
Swiss-Prot	438182
ClinVar	208560
BioProject	155850
DBPedia	122339
Dryad	42837
Omics2D	3601
ArrayExpress	6119
Dataverse	60303
- New Features:** A section listing changes from February 28, 2017 (v2.0) and November 23, 2016 (v1.5). The changes include increased coverage, duplicate datasets display, user feedback integration, and visualization of results via timeline.

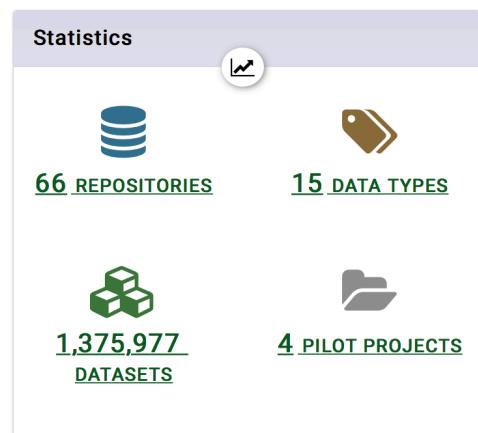
At the bottom, there's a "Pilot Projects" section featuring four projects: GWAS Finder, ISEE-DELVE, DataRank, and Data.Citation\_Discovery.

At the very bottom of the page, a reference is provided: Ohno-Machado, L. et al. Finding useful data across multiple biomedical data repositories using DataMed. *Nat Genet* **49**, 816–819 (2017).

# Repositories: DataMed

“Immunology”

- Repositories
  - Dryad (98)
  - BioProject (71)
  - PDB (67)
  - ArrayExpress (55)
  - OmicsDI (53)
  - ImmPort (45)
  - RGD (16)
  - dbGaP (12)
  - GEMMA (4)
  - ProteomeXchange (1)
  - VectorBase (1)
  - Zenodo (1)



Ohno-Machado, L. et al. Finding useful data across multiple biomedical data repositories using DataMed. *Nat Genet* **49**, 816–819 (2017).

## Exercise

### STEPS

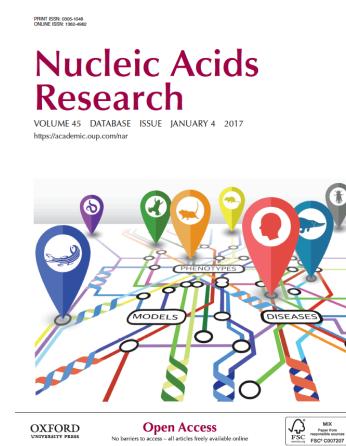
1. Go to: <https://datamed.org>
2. Search for “immunology” repositories
3. Search for “immunology” data sets

### QUESTIONS

- What type of information is contained in Datamed?
- When was the last update?
- How many repositories did you find?
- How many data sets did you find?
- What types are data sets there?
- Name something that you would like to look at further?

# Nucleic Acid Research (NAR) Database Issue

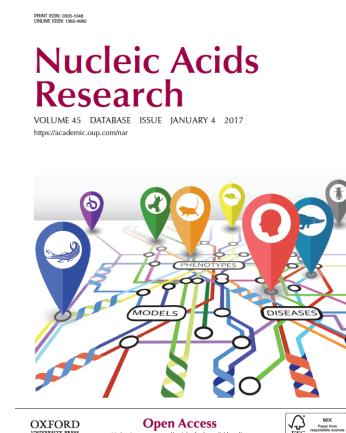
Nucleotide Sequence Databases	Microarray Data and other Gene Expression Databases
RNA sequence databases	Proteomics Resources
Protein sequence databases	Other Molecular Biology Databases
Structure Databases	Organelle databases
Genomics Databases (non-vertebrate)	Plant databases
Metabolic and Signaling Pathways	Immunological databases
Human and other Vertebrate Genomes	<ul style="list-style-type: none"> <li>◦ <a href="http://www.oxfordjournals.org/nar/database/c">http://www.oxfordjournals.org/nar/database/c</a></li> </ul>
Human Genes and Diseases	<ul style="list-style-type: none"> <li>Cell biology</li> <li>◦ <a href="http://www.oxfordjournals.org/nar/database/c">http://www.oxfordjournals.org/nar/database/c</a></li> </ul>



<http://www.oxfordjournals.org/nar/database/>

## Nucleic Acid Research (NAR) Database Issues: Immunological Databases

AAgAtlas	IEDB-3D	IPD-KIR - Killer-cell Immunoglobulin-like Receptors
ALPSbase	IL2Rbase	IPD-MHC
AntigenDB	IMGT	MHCBN
AntiJen	IMGT/GENE-DB	MHCPEP
BCIpep	IMGT/HLA	MPID-T2
bNAber	IMGT/LIGM-DB	MUGEN Mouse Database
dbMHC	IMGT/mAb-DB	DIGIT
DIGIT	ImmuneNet	Protegen
Epitome	InnateDB	SABDab
GPX-Macrophage Expression Atlas	Interferon Stimulated Gene Database	SuperHaptens
HaptenDB	IPD - Immuno Polymorphism Database	VBASE2
HPTAA	IPD-ESTDAB	
IEDB	IPD-HPA - Human Platelet Antigens	



<http://www.oxfordjournals.org/nar/database/cat/14>

## Exercise - Later

---

### STEPS

1. Go to:  
<http://www.oxfordjournals.org/nar/database/>

### QUESTIONS

- What type of information is contained in NAR?  
When was the last update?

## Data Repositories

---

## Reasons to Look at These...

Make you familiar with the repositories to:

- Download publically available data that you may find useful
  - Enhance grant application or data sharing plan
  - Upload your lab data
    - by choice
    - if required by grant or contract

# Immunology Database and Analysis Portal (ImmPort)

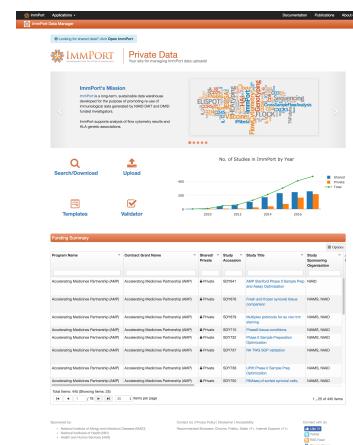
## A long-term, sustainable data warehouse

Developed for the purpose of promoting re-use of immunological data

Data generated by NIAID DAIT and DMID funded investigators

Private data and pre-release data are stored in private workspaces of investigators at the ImmPort site located at NIAID

- <http://import.niaid.nih.gov>



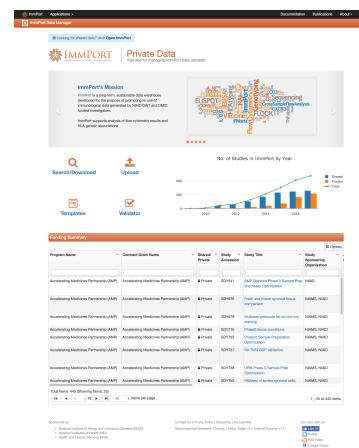
# Immunology Database and Analysis Portal (ImmPort)

“The goals of the ImmPort project are to:

- Provide an open access platform for research data sharing
  - Create an integrated environment that broadens the usefulness of scientific data and advances hypothesis-driven and hypothesis-generating research
  - Accelerate scientific discovery while extending the value of scientific data in all areas of immunological research
  - Promote rapid availability of important findings, making new discoveries available to the research community for further analysis and interpretation
  - Provide analysis tools to advance research in basic and clinical immunology”

Private data and pre-release data are stored in private workspaces of investigators at the ImmPort site located at NIAID

- <http://immpport.niaid.nih.gov>



# Exercise

STEPS

1. Go to: <http://www.immport.org>
  2. Click on private data

## QUESTIONS

What type of information is contained in ImmPort?

When was the last update?

What four activities are possible with private data?

How many shared studies are there with  
“influenza” in the title?

# Exercise

## STEPS

1. Go to: <https://www.immporgalaxy.org>

## QUESTIONS

What type of information is contained in ImmPort?

When was the last update?

What four activities are possible with private data?

How many shared studies are there with "influenza" in the title?

# FlowRepository

A database of flow cytometry experiments where you can query and download data collected and annotated according to the MIFlowCyt standard.

# Exercise

## STEPS

1. Go to: <https://flowrepository.org>
2. Search for “flowcap” data sets

## QUESTIONS

What type of information is contained in ImmPort?

When was the last update?

How many public data sets are there?

How many “flowcap” data sets are there?

# Immune Epitope Database (IEDB)

Free resource that offers easy searching of experimental data characterizing antibody and T cell epitopes studied in humans, non-human primates, and other animal species.

Epitopes involved in infectious disease, allergy, autoimmunity, and transplant are included.

Hosts tools to assist in the prediction and analysis of B cell and T cell epitopes.

A free resource, funded by a contract from the NIAID

**Summary Metrics**

Peptide Epitopes	305,461
Neutralizing Epitopes	1,233
T Cell Epitopes	321,919
B Cell Assays	354,663
T Cell Assays	68,635
Epitope Source Organisms	2,612
Restricting MHC Alleles	743
References	16,622

**Epitope Analysis Tools**

- T Cell Epitope Prediction**: Predict MHC-I or MHC-II binding, T-cell receptor usage, and T-cell receptor diversity.
- B Cell Epitope Prediction**: Predict IgM or IgG epitope usage, predict decoy epitopes, and predict B-cell epitope conservation.
- Epitope Coverage**: Analyze across sets of epitopes, including Population Coverage, Conservation Across Antigens, and Clusters with Similar Sequences.

<http://www.iedb.org/>

## Exercise

### STEPS

1. Go to: <http://www.iedb.org>
2. Search for all epitopes by the organism "influenza virus all"

### QUESTIONS

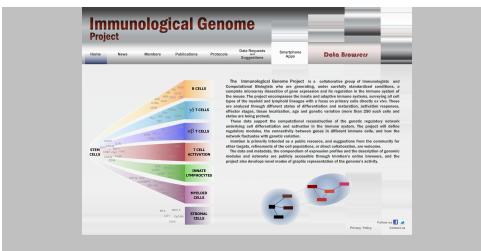
- What type of information is contained in IEDB?
- When was the last update?
- How many epitopes for influenza are there total?

## Immunological Genome Project (ImmGen)

Goal of the project is to computationally reconstruct the gene regulatory network in immune cells

A gene-expression database for all characterized immune cells in the mouse

Compendium of microarray data currently include over 250 immunologically relevant cell types, from all lymphoid organs and other tissues which are monitored by immune cells



## Exercise

### STEPS

1. Go to: <https://www.immgen.org>

### QUESTIONS

What type of information is contained in IEDB?  
When was the last update?

## VDJ Server

A free, scalable resource for performing immune repertoire analysis and sharing data

VDJServer currently provides the following analysis components

- Read pre-processing, such as quality filtering and demultiplexing, using the VDJPipe and/or pRESTO applications
- V(D)J assignment using the IgBlast application
- Repertoire analysis using the RepCalc application

The screenshot shows the VDJ Server homepage. At the top right is a circular logo with scientific icons and the text "VDJ SERVER". To its right is a "WELCOME!" box containing a brief description of the service and links for "LOGIN", "Forgot password?", "Create Account", "Send us Feedback", and "Documentation". Below this is a "NEWS & ANNOUNCEMENTS" section featuring a thumbnail of a video titled "April 26, 2017 - Announcing VDJServer Release 1.0.3". At the bottom are four large buttons labeled "UPLOAD", "ANALYZE", "PUBLISH", and "SHARE", each with a small icon and a brief description.

<https://vdjserver.org/>

## Exercise

---

### STEPS

1. Go to: <https://vdjserver.org>

### QUESTIONS

What type of information is contained in VDJ Server?

When was the last update?

## Tool (Software) Repositories

---

# OmicsTools

---

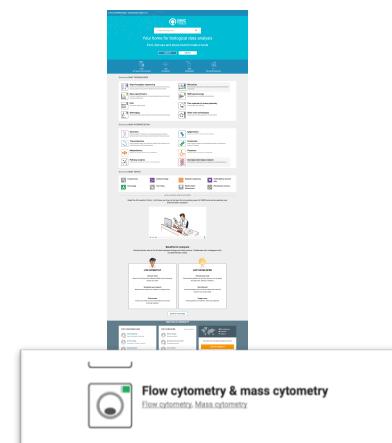
Find, discuss and share bioinformatics tools

19,051 Software And Databases

1,255 Categories

Browse By

- Omic Technologies
- Omic Interpretation
- Omic Topics



# OmicsTools

---

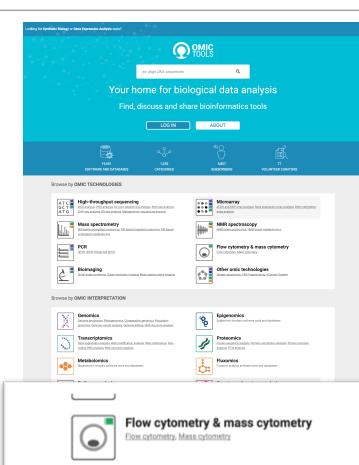
Omic Technologies

- Flow cytometry (81 tools)
- Mass cytometry (15 tools)

Omic Interpretation

Omic Topics

- Immunology



# Exercise

## STEPS

1. Go to: <https://omicstools.com>
2. Search for “rnaseq”
3. Search for “flow”

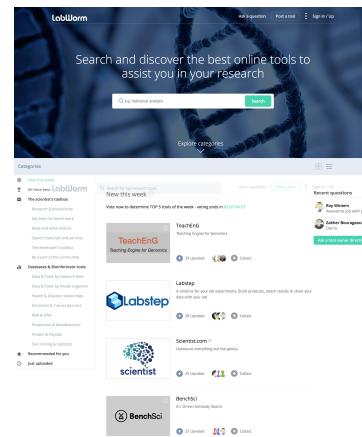
## QUESTIONS

- What type of information is contained in OmicsTools?
- When was the last update?
- How many “rnaseq” resources are found?
- How many “flow” resources are found?

# Labworm

Search and discover the best online tools to assist you in your research

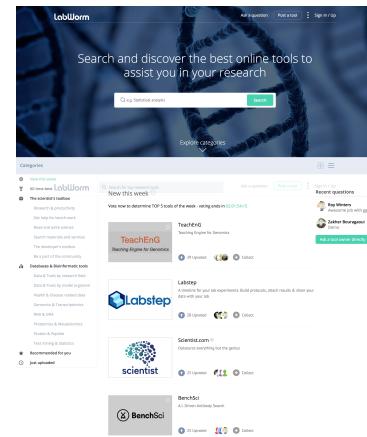
- The scientist's toolbox
  - Research & productivity
  - Get help for bench work
  - Read and write science
  - Search materials and services
  - The developer's toolbox
  - Be a part of the community
- Databases & Bioinformatic tools
  - Data & Tools by research field
  - Data & Tools by model organism
  - Health & Disease related data
  - Genomics & Transcriptomics
  - RNA & DNA
  - Proteomics & Metabolomics
  - Protein & Peptide
  - Text mining & Statistics



# Labworm - Immunology

## Collections

- Immunological Tools
  - Margaret Jordan
- Tools
  - IMGT
  - IMGT/HLA
  - Immgene
  - Immune Epitope Database
  - Kabat Database
  - Immuno Polymorphism Database
  - InnateDB
  - Bcepred
  - INTERFEROME
  - ABCpred
  - Antigenic



<https://labworm.com/collection/Margaret.Jordan/immunological-tools>

# Exercise

## STEPS

1. Go to: <https://labworm.com>
2. Search for “immune” resources

## QUESTIONS

- What type of information is contained in LabWorm?
- When was the last update?
- How many “immune” resources are there?

# GitHub

GitHub is a development platform inspired by the way you work. From [open source](#) to [business](#), you can...

host and review code

manage projects

build software alongside millions of other developers.

The screenshot shows a GitHub repository page for 'immunology-informatics'. The repository name is at the top. Below it, there are tabs for 'Pull requests', 'Issues', 'Marketplace', and 'Git'. A search bar and a 'Create new file' button are also present. The main area displays a list of files and their details:

- `01_intro.Rmd`: Add intro and READMEs, 18 days ago
- `02_intro.R`: intro forming, 8 days ago
- `03_intro.html`: Update package that need to be installed, 18 days ago
- `04_intro.ipynb`: intro, 18 days ago
- `05_downstream_analysis`: Add intro and READMEs, 18 days ago
- `06_impute.R`: initial commit of file, 7 days ago
- `07_pca.Rmd`: Add intro and READMEs, 18 days ago
- `08_gwas.R`: initial pipeline, 18 days ago
- `09_loess.R`: initial commit, 22 days ago
- `10_SESAME.Rmd`: Initial commit for subresource, 22 days ago

Below the files, there is a section for 'README.md' which contains the following text:

```
Immunology Informatics
Big Data Analysis in Immunology
Tutorials
These tutorials are part of the [American Association of Immunologists (AAI) Course in Big Data Analysis in Immunology](AAI Course in Big Data Analysis in Immunology).
Please feel free to issue pull requests and improve these materials. Please note that much of this material has been authored by many talented individuals. Please see the license in each tutorial for acknowledgement.
```

# Exercise

## STEPS

1. Go to: <https://github.com>
2. Search for “flow cytometry” resources
3. Select “R” resources only

## QUESTIONS

What type of information is contained in LabWorm?

When was the last update?

How many “flow cytometry” resources are there?

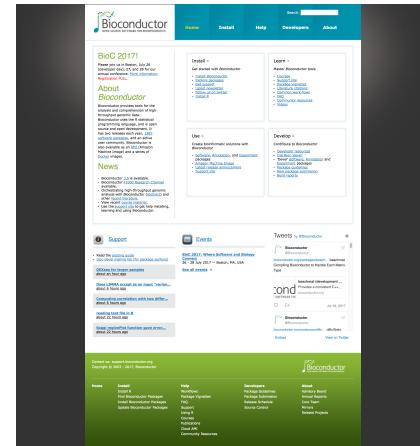
How many are “R” resources?

# Bioconductor

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data.

Bioconductor uses the R statistical programming language, and is open source and open development.

It has two releases each year, [1383 software packages](#), and an active user community.



# Exercise

## STEPS

1. Go to: <https://www.bioconductor.org>
2. Search for “flow cytometry” resources

## QUESTIONS

- What type of information is contained in LabWorm?
- When was the last update?
- How many “flow cytometry” resources are there?
- How many are “R” resources?

# Stackoverflow

Programming question and answer website  
Highly curated!

The screenshot shows the Stackoverflow homepage with a search bar at the top. Below it is a list of questions from various users. One prominent question is about "Adding Java's casting usage with Reference to MySQL". Other visible questions include "Matrix - what Matrix must be positive definite", "How do I send an image file via email using Java?", and "Uncaught promise happens only on Travis CI". On the right side, there's a sidebar for "Looking for a job?" featuring job listings for "Front End Web Developer" and "Software Engineer". Below the sidebar, there's a section for "Related Tags" and "Hot Network Questions".

# Exercise

STEPS	QUESTIONS
<ol style="list-style-type: none"> <li><b>1.</b> Search (Google) for           <ol style="list-style-type: none"> <li><b>1.</b> “how do I loop through files in r”</li> </ol> </li> </ol>	<p>What websites rise to the top? Stack overflow</p> <ul style="list-style-type: none"> <li>◦ How many stars does the first answer have?</li> </ul>

# Reproducible Research

---

## Becoming a Reproducible Scientist...

---

Remember – take it one step at a time

- Good – Document where you get your data, the steps you use to produce your analysis
- Better – Script your analysis
- Best – Scripts your analysis, upload your scripts to GitHub, contribute well document data to a data repository

# Additional Resources

---

## Resources

---

Big Data 2 Knowledge (BD2K) Lecture series

- <https://bigdatau.ini.usc.edu/data-science-seminars>

Wikipedia (via wikiwand)

- Computational immunology
  - [http://www.wikiwand.com/en/Computational\\_immunology](http://www.wikiwand.com/en/Computational_immunology)
- Flow cytometry bioinformatics
  - [http://www.wikiwand.com/en/Flow\\_cytometry\\_bioinformatics](http://www.wikiwand.com/en/Flow_cytometry_bioinformatics)

“Awesome” Resource Lists of all types (many bioinformatics or computational)

- <https://github.com/sindresorhus/awesome>

Medical clipart – share alike

- <http://smart.servier.com>

## Resources Used

BD2K Guide to the Fundamentals of Data Science Series Talks

- <https://www.youtube.com/channel/UCKIDQOa0JcUd3K9C1TS7FLQ>
- [Metadata Standards](#), Dr. Susanna-Assunta Sansone from University of Oxford
- [Introduction to Big Data and the Data Lifecycle](#), Dr. Mark Musen from Stanford University

## Papers of Interest

Genser, B., Cooper, P. J., Yazdanbakhsh, M., Barreto, M. L. & Rodrigues, L. C. A guide to modern statistical analysis of immunological data. *BMC Immunology* 2007 8:1 8, 27 (2007).

- “This paper will help the immunologist to choose the correct statistical approach for a particular research question.”

Zhang, G. L., Sun, J., Chitkushev, L. & Brusic, V. Big Data Analytics in Immunology: A Knowledge-Based Approach. *BioMed Research International* 2014, 1–9 (2014).

## Objectives

---

Enable you to name one or more immunologically relevant data standards

Enable you to search for immunologically relevant data repositories

Enable you to take a step in becoming a reproducible scientist

## Q & A

---