

---

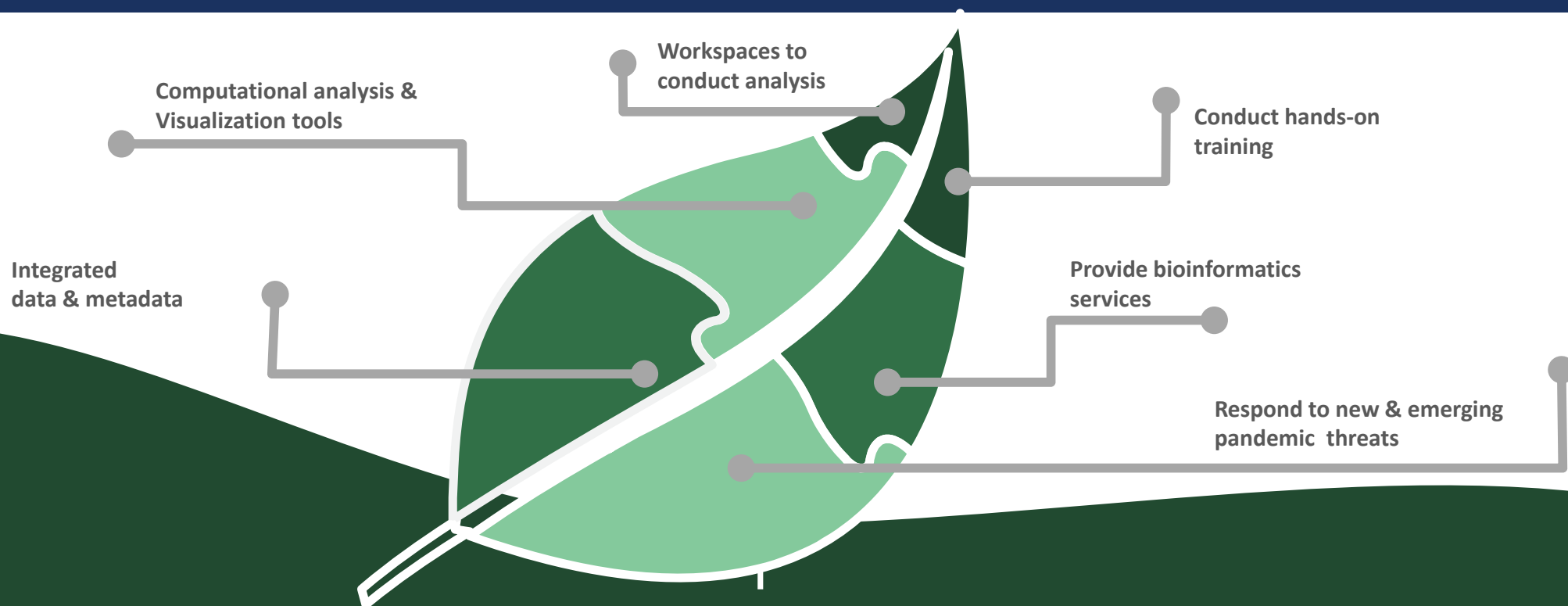
# BIOINFORMATICS RESOURCE CENTERS

A knowledgebase of Integrated Data, Metadata, & Analysis  
Tools to Support Data-enabled Infectious Disease Research

Ishwar Chandramouliswaran  
Office of Data Science & Emerging Technologies  
NIAID, NIH, DHHS



# KEY CAPABILITIES



The BRCs provide integrated bioinformatics resources in support of basic & applied infectious disease research

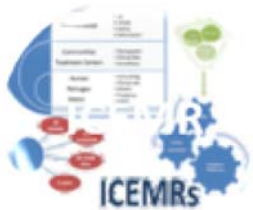
# THE CENTERS ORGANIZED BY PATHOGEN GROUPS

Institute name	PI name	BRC	Pathogen
University of Notre Dame	Mary Ann McDowell	VectorBase	Invertebrate vectors of human pathogens
University of Pennsylvania	David Roos	EuPathDB/FungiDB	Eukaryotic/fungal pathogens
Northrop Grumman	Richard Scheuermann	ViPR/IRD	Viruses/Influenza
University of Chicago	Rick Stevens	PATRIC	Bacteria

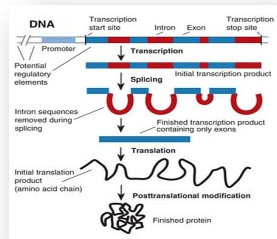


# COLLABORATE TO INTEGRATE DATA ACROSS NIAID PROGRAMS

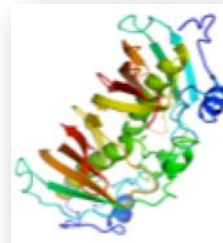
## Sequencing



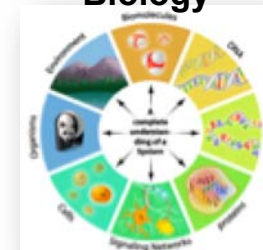
## Functional Genomics



## Structural



## Systems Biology



## Bioinformatics

### Genomic Research Resources

Genomic/Omics Data Sets, Databases, Bioinformatics Tools, Biomarkers, 3D Structures, Protein Clones, Predictive Models

To address key questions in microbiology and infectious disease

To identify new targets and develop new strategies for vaccines, diagnostics and therapeutics

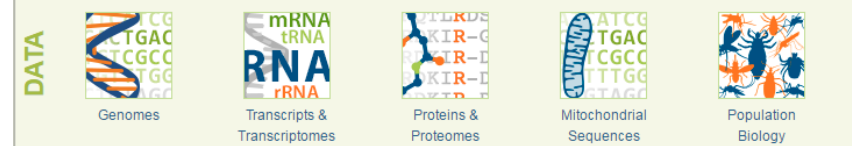


# VECTORBASE BRC FOR INVERTEBRATE VECTORS OF HUMAN PATHOGENS

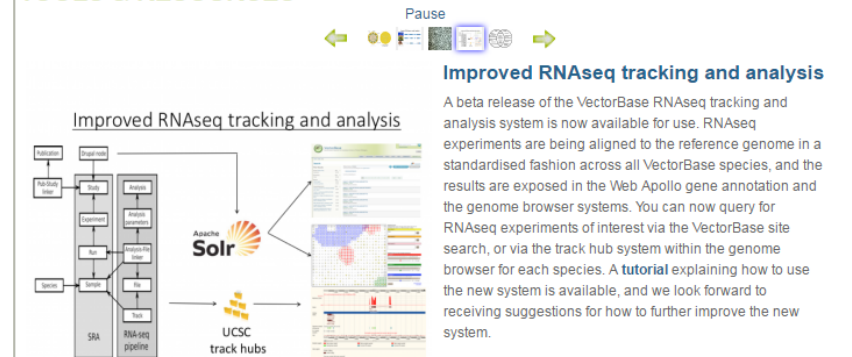
- **An unique and integrated bioinformatics resource for vector community**
  - Genome of **35 organisms** with annotated gene features, genome variation, transcript and protein expression
  - **Population** genetics data integrated with variation and **insecticide-resistance** phenotypes, field-associated samples from **surveillance** studies, and pathogen **transmission** data
  - Transcriptomes (RNA-seq) and gene expression data
  - Pathways
- **Data access, analysis and visualization tools**
  - Data-centric search system
  - **Computational tools** made accessible through a Galaxy instance implemented in VB
  - Expression map
  - Ontology browser

## Welcome to VectorBase!

VectorBase is a NIAID Bioinformatics Resource Center providing genomic, phenotypic and population-centric data to the scientific community for invertebrate vectors of human pathogens.



## TOOLS & RESOURCES



## SUBMIT YOUR DATA TO VECTORBASE

VectorBase is not just for accessing data and using tools - we

Tweets by @VectorBase



# EUPATHDB

## BRC FOR INVERTEBRATE VECTORS OF HUMAN PATHOGENS

### An unique resource for protozoan & fungal pathogens



- A comprehensive collection genomic and functional **genomic data, isolate data and phylogenomics**
  - genome sequences and annotation, strand-specific RNA-seq data, splice junction predictions, phosphoproteomic data, high-throughput phenotyping data, single nucleotide polymorphism, and expression quantitative trait loci data.
- Sophisticated **search strategies** using metadata
- Computational **analysis** on Galaxy

**EuPathDB** provides a unified entry point for the Eukaryotic Pathogen Bioinformatics Resource Center where you can ask questions that leverage data across organisms using orthology:  
<http://EuPathDB.org>

**OrthoMCL** allows you to identify phylogenetically conserved or unique proteins among genomes from all branches of life:

<http://OrthoMCL.org>

Organism specific websites allow you search diverse data types from multiple species:

**AmoebaDB** Amoebozoa specific resource including *Entamoeba* - *Acanthamoeba* coming soon:

<http://AmoebaDB.org>

**CryptoDB** *Cryptosporidium* specific resource:  
<http://CryptoDB.org>

**GiardiaDB** *Giardia* specific resource includes assemblages A, B and E:

<http://GiardiaDB.org>

**MicrosporidiaDB** Includes multiple microsporidia species:

<http://MicrosporidiaDB.org>

**PiroplasmaDB** *Babesia* and *Theileria* specific resource:

<http://PiroplasmaDB.org>

**PlasmoDB** Resource for *Plasmodium* species:  
<http://PlasmoDB.org>

**ToxoDB** *Toxoplasma*, *Neospora* and *Eimeria* resource:

<http://ToxoDB.org>

**TrichDB** Resource for *Trichomonas vaginalis*:

<http://TrichDB.org>

**TriTrypDB** Resource for kinetoplastida including *Trypanosoma* spp. and *Leishmania* spp.:

<http://TriTrypDB.org>



# VIRUS PATHOGEN RESOURCE (ViPR) INFLUENZA RESEARCH DATABASE (IRD)

- **A comprehensive collection of influenza and multiple virus family related data.**
  - sequences and annotations, immune epitopes, 3D protein structures, HTP omics data, clinical phenotype, surveillance, serology, host factor data, and curated antiviral drug data.
- **A suite of analytical and visualization tools**
  - phylogenetic tree, sequence variation determination, metadata-driven Comparative Analysis Tool for Sequences (meta-CATS), short peptide identification, PCR primer design, Sequence Feature and Phenotypic Variant Type (PVT) annotation, HA clade classification, HA subtype numbering conversion, surveillance data visualization, protein structure visualization, and host factor data enrichment analysis
- **Personal workbench spaces for data storage, sharing and analysis**
- **An infrastructure for antiviral drug data management and analysis.**
  - Drug interaction site curation
  - Integration of host factor data with drug target
  - Anti-viral drug resistance risk assessment tool

The image shows two screenshots of web interfaces. The top screenshot is the Influenza Research Database (IRD) homepage, featuring a world map with red and green markers indicating surveillance data. The bottom screenshot is the Virus Pathogen Resource (ViPR) homepage, which displays a search bar, a list of data types (Genomes, Genes & proteins, Immune epitopes, 3D protein structures, Host Factor Data, Antiviral Drugs), and a section for analyzing data online (Sequence Alignment, Phylogenetic Tree, Sequence Variation (SNP), Metadata-driven Comparative Analysis, BLAST). It also includes a 'Save to Workbench' section with options to store and share data, integrate data with ViPR, and create a custom search alert.

**IRD Influenza Research Database**

**Surveillance**  
Human, avian and non-human mammalian surveillance data can be searched based upon location and various host characteristics for download and display on a map with bird flyway overlay. Create a report showing counts of surveillance records grouped by user-driven criteria.

**Key Highlights:**

- Human, avian and non-human mammalian surveillance data
- User-driven serotype comparison
- Display data on Google Map with flu prevalence color coded and bird flyways overlaid
- View results by metadata matrix report

**ViPR Virus Pathogen Resource**

**Search**  
Search our comprehensive database for:  
• Genomes  
• Genes & proteins  
• Immune epitopes  
• 3D protein structures  
• Host Factor Data  
• Antiviral Drugs  
Browse All Search Types

**Analyze**  
Analyze data online:  
• Sequence Alignment  
• Phylogenetic Tree  
• Sequence Variation (SNP)  
• Metadata-driven Comparative Analysis  
• BLAST  
Browse All Tools

**Save to Workbench**  
Sign up for a workbench to:  
• Store and share data  
• Combine working sets  
• Integrate your data with ViPR data  
• Store and share analyses  
• Custom search alert  
Sign In



# PATHOGEN RESOURCE INTEGRATION CENTER (PATRIC) BRC FOR BACTERIAL SPECIES



## Integrated Data Collections

Genomes, Transcriptomes (uArray and RNAseq), Proteomics, Metabolomics, Curated gene collections, Protein-Protein Interactions, Tn-Seq

AMR panel data, Disease & isolation phenotype metadata,, Virulence, Drug Targets, Biochemical Pathways

DATA WORKSPACES SERVICES HELP

**Data Types**  
Antibiotic  
Resistance  
Genomes  
Genomic  
Features  
Pathways  
Protein Families  
Specialty Genes  
Transcriptomics  
  
**Download Data**  
FTP Server

**Specialty Data Collections**  
PATRIC Collaborations  
PATRIC DBPs  
NIAID Clinical  
Proteomics  
NIAID Genome  
Sequencing  
NIAID Structural  
Genomics  
NIAID Systems Biology  
NIAID Functional  
Genomics

**Personal Workspace** for analysis of User data and “Virtual Integration” of user data in the context of all the public datasets

PATRIC / home / Special Collections / NIAID Systems Biology Centers / Omics4TB

Name	Size	Owner	Members	Created
Parent folder				
MTB TF fitness		PATRIC	Public	5/6/17, 12:02 PM
MTB TF overexpression		PATRIC	Public	5/6/17, 12:12 PM
MTB TRIP relative fitness profiling		PATRIC	Public	3/25/18, 11:29 AM
MTB hedges		PATRIC	Public	5/6/17, 12:09 PM
		PATRIC	Public	5/6/17, 12:11 PM
		PATRIC	Public	3/25/18, 9:29 AM

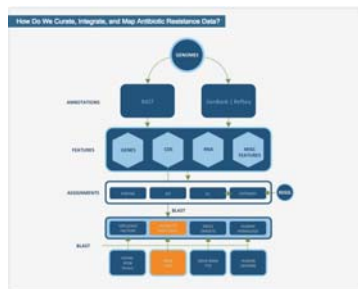
Antibiotic Resistance /  
Susceptibility Phenotype  
Predictions

## Computational Analysis Services

SERVICES HELP All Data Types Find a gene genome

**Genomes**  
Assembly  
Annotation  
BLAST  
Similar Genome Finder  
Variation Analysis  
Tn-Seq Analysis  
Phylogenetic Tree  
Metagenome Binning

**Protein Tools**  
Protein Family Sorter  
Protein Comparison  
  
**Metabonomics**  
Comparative Pathway  
Model Reconstruction  
  
**Data**  
ID Mapper



What do we mean by...

**Antibiotics:** Antibiotics are a type of antimicrobial drug used in the treatment and prevention of bacterial infections. PATRIC provides basic information about commonly used antibiotics, including their chemical and physical properties, pharmacology, and mechanism of action. In addition, each antibiotic is linked to other relevant data available in PATRIC, such as AMR phenotypes for genome, AMR genes, and AMR regions. Below are some examples.

**AMR Phenotypes:** AMR phenotypes refer to the resistance or susceptibility of a given organism to one or more antibiotics. PATRIC collects AMR phenotype data generated using antimicrobial susceptibility testing methods (AST) from published studies and collaborators. In addition, we also provide predicted AMR phenotypes using machine learning classifiers. See AMR phenotype data for selected genes:

- *ampicillin*
- *streptomycin*
- *chloramphenicol*

**AMR Genes:** AMR genes refer to the genes implicated in or associated with the resistance to one or more antibiotics. The resistance may result from the presence or absence of a gene or specific mutations acquired spontaneously or through evolution over time. We integrate and map known antibiotic resistance genes from the following sources:

- *AMR*
- *AMR*
- *AMR*

**AMR Regions:** AMR regions refer to the small genomic regions implicated in or associated with the resistance to one or more antibiotics. The AMR regions are computationally predicted using machine learning classifiers used to predict AMR phenotypes. They may map to existing genes or intergenic regions and may help identify new AMR genes or understand AMR mechanisms.

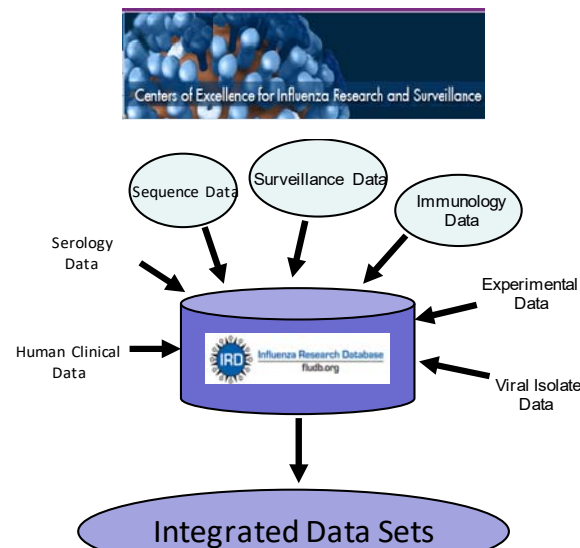
[View all AMR genes](#)

Summary of Omics4TB Datasets

Experiment	PMID	Data Source
The DNA-binding network of Mycobacterium tuberculosis	25081035	• BioProject PRJNA255884 Raw sequence data for ChIP-seq experiments (BAM format) • MTH Network Portal ChIP-seq (Sorted and Indexed ChIP-seq files)
Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network	25389855	• GEO Accession: GSE59008 Transcription factor overexpression data from ring microscopy experiments • PATRIC: Ring-microscopy data • MTH Network Portal TPOE Sample information for each experiment
A high-resolution network model for global gene regulation in Mycobacterium tuberculosis	25230956	• MTH Network Portal TPOE Data sets and additional information
A comprehensive map of genome-wide gene regulation in Mycobacterium tuberculosis	25077515	• PATRIC: ChIP-seq ChIP-seq experiment information page • PATRIC: TPOE Expression TPOE Experiment information page • GEO Accession: GSE59008 Transcription factor overexpression data from ring microscopy experiments • MTH Network Portal TPOE Sample information for each experiment • PATRIC: TPOE Expression TPOE Experiment information page
Network analysis identifies Rv0226 and Rv0227 as regulators of isoniazid tolerance in Mycobacterium tuberculosis	25173154	• GEO Accession: GSE57349 Expression profiling data
AMR-18-regulated molecular network orchestrates cell fate in the innate and adaptive immune response to Mycobacterium tuberculosis	26048045	• PATRIC: Proteomics Proteomics
Abscissa Proteome Composition and Dynamics during dormancy and resuscitation of Mycobacterium tuberculosis		• PATRIC: Proteomics Proteomics
MTH Transcriptional Response Induction Phenotype (TRIP) Screen		• PATRIC: TRIP Screen



# RESPOND TO NEEDS OF ID RESEARCHERS



- Clinical & Epidemiological Data
- PopBio Data & Tools

**Zika Virus**

About Us Community Announcements Links Resources Support

SEARCH DATA ANALYZE & VISUALIZE WORKBENCH VIRUS FAMILIES HELP Wordbench Sign In

**Zika Virus**

Taxonomy: Group IV ((+ssRNA); Flaviviridae; Flavivirus; Zika virus most similar to St. Louis encephalitis virus

Virion: enveloped, spherical, ~50 nm in diameter with an electron dense core of ~30 nm

Genome: ~10.8 kb positive-sense, single-stranded RNA

Proteome: single polyprotein, co- & post-translationally cleaved into 11 mature proteins

Infection: initiated by binding of the viral E protein to a host receptor

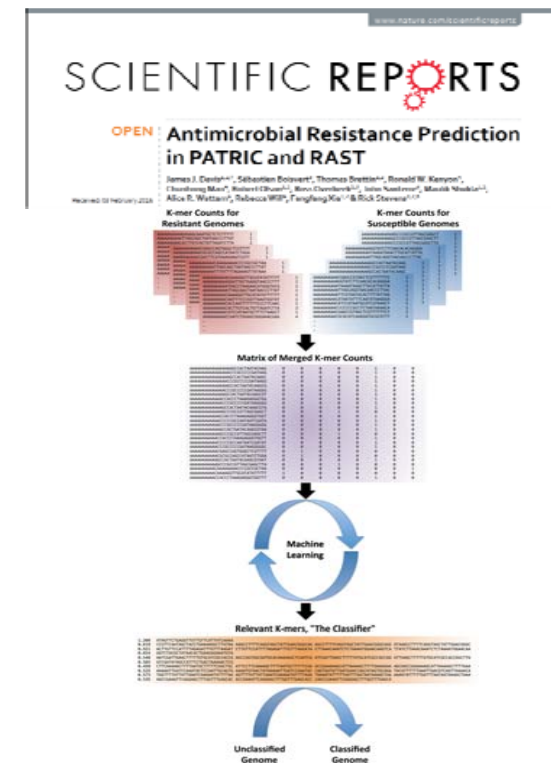
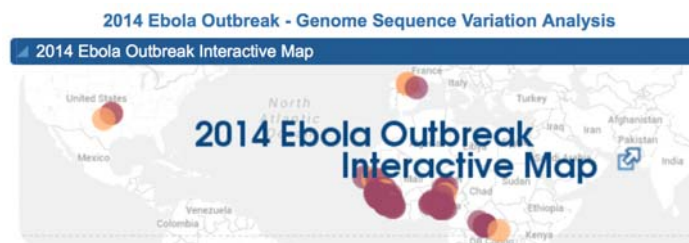
RNA Transcript: 5' methyl cap, no 3' poly-A tail

Transmission: primarily through the bite of an infected Aedes mosquito; rarely from mother to child; infrequently through blood transfusion and sexual contact

Epidemiology: sporadic outbreaks in Africa, Southeast Asia, and the Pacific Islands before 2015; major outbreaks have been occurring in the Americas since May 2015. Currently, 30 countries have active transmission of the Zika virus while numerous others have reported travel-associated cases.

Clinical: the most common symptoms of Zika virus infection are fever, rash, joint pain, and conjunctivitis; a link to neurological syndromes, including Guillain-Barre syndrome, and birth defects is strongly suspected.

References: David M. Kriple, Peter M. Howley, et al. Fields' Virology, 5th ed. 2007 Lippincott Williams & Wilkins, USA; <http://www.cdc.gov/zika>

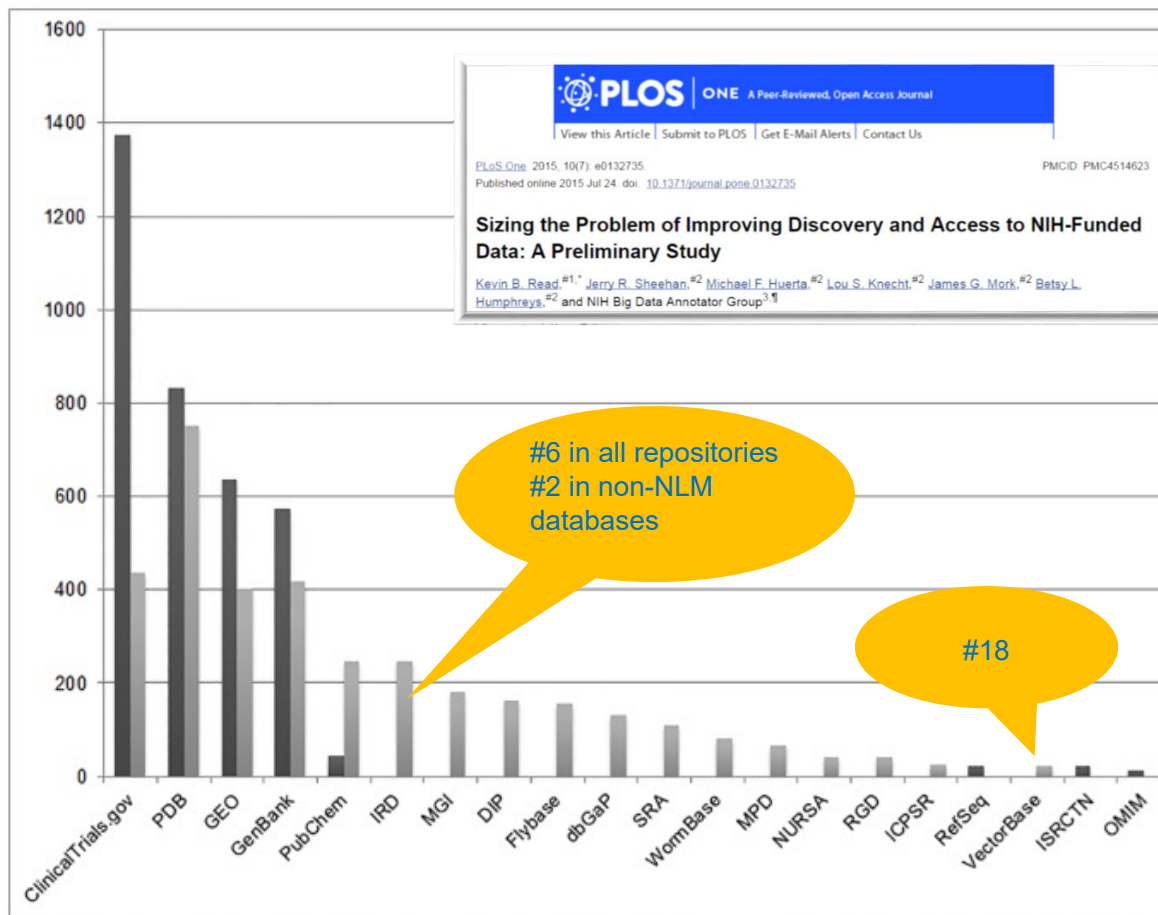


- AMR data collection
- AMR prediction tools

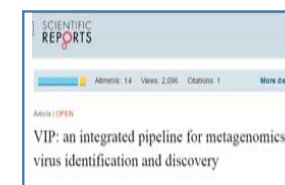
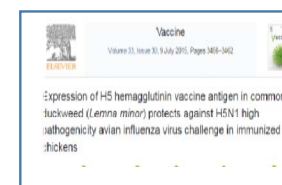
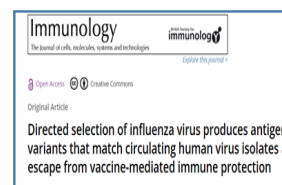
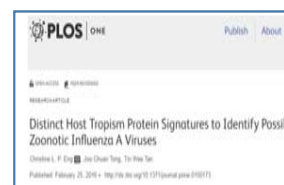
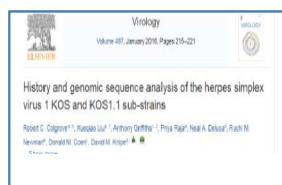
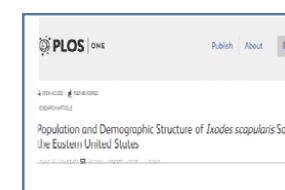
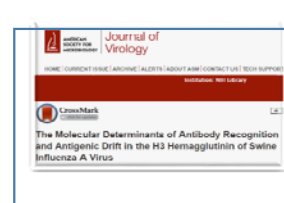
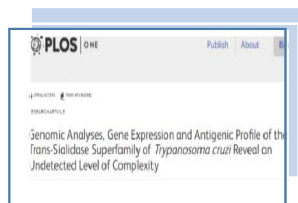
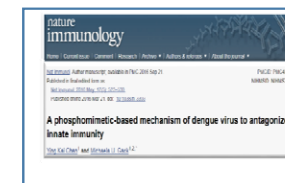
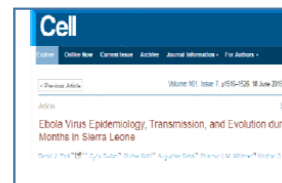
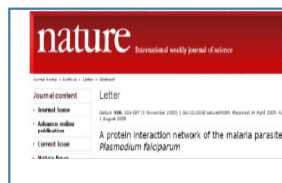
# HANDS-ON TRAINING WORKSHOPS GLOBALLY



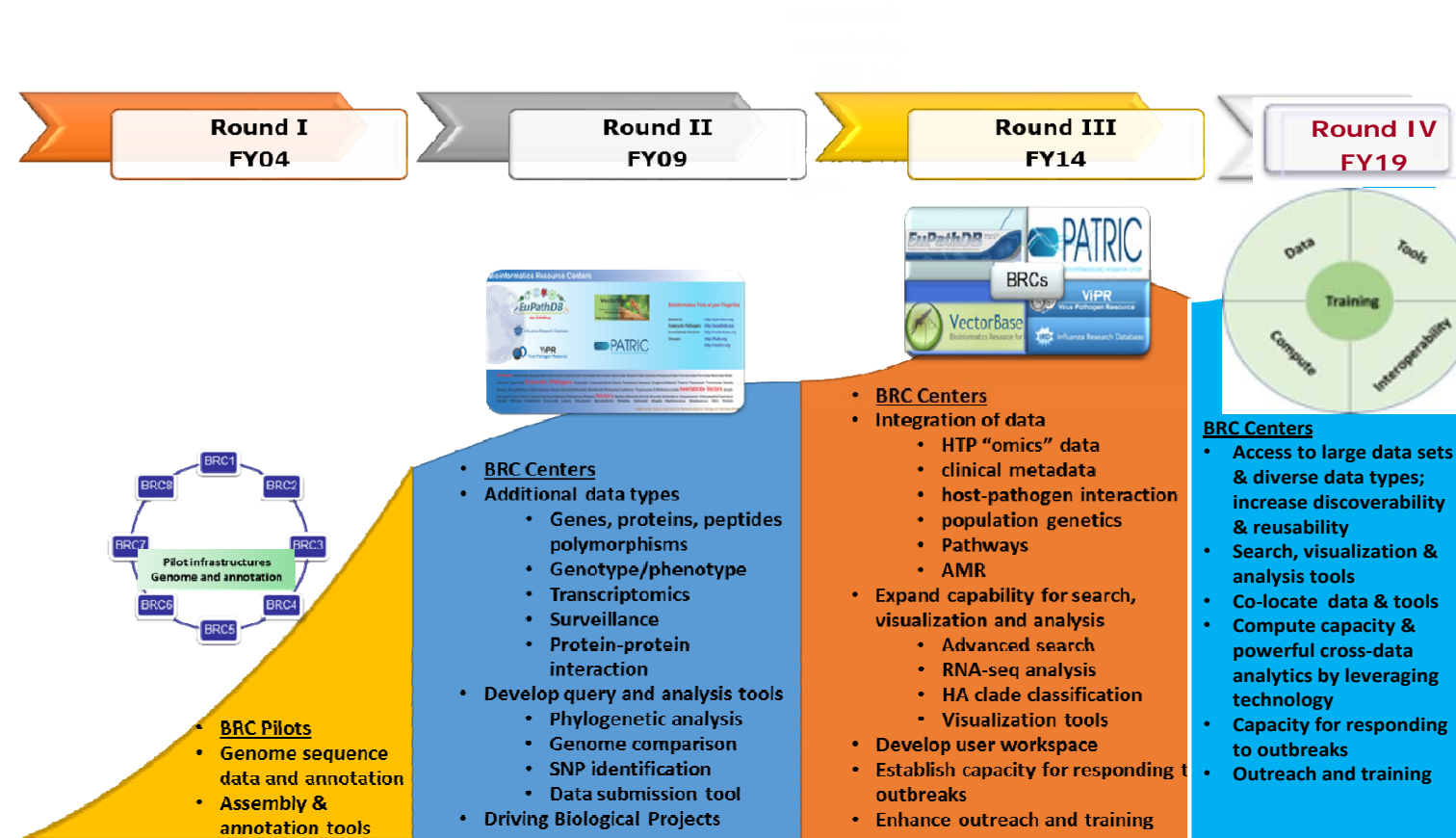
# TO INCREASE USAGE & IMPACT



# SCIENCE ENABLED BY THE BIOINFORMATICS RESOURCE CENTERS



# EVOLUTION OF THE RESOURCES





# THE NIAID DATA CHALLENGE

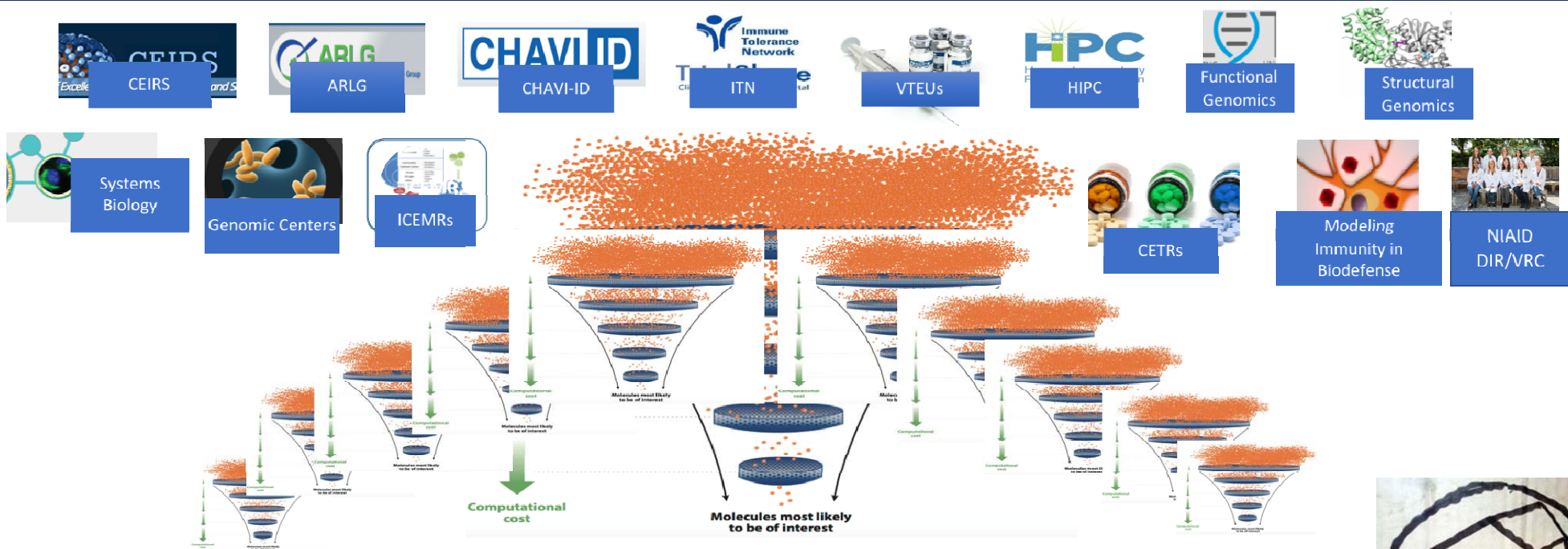


Image Courtesy: Dr. Emily Erbelding (modified)

# NEW MODELS OF DATA-DRIVEN HYPOTHESIS GENERATION

## Expanding the Immunology Toolbox: Embracing Public-Data Reuse and Crowdsourcing

Rachel Sparks,<sup>1</sup> William W. Lau,<sup>2</sup> and John S. Tsang<sup>1,\*</sup>

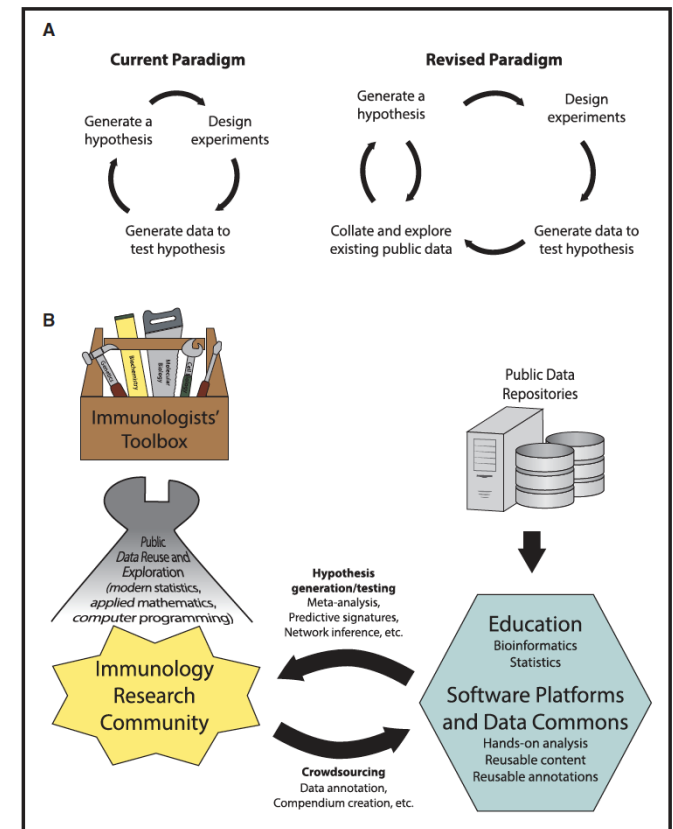
<sup>1</sup>Systems Genomics and Bioinformatics Unit, Laboratory of Systems Biology, National Institutes of Allergy and Infectious Diseases, National Institutes of Health

<sup>2</sup>Office of Intramural Research, Center for Information Technology, National Institutes of Health

\*Correspondence: [john.tsang@nih.gov](mailto:john.tsang@nih.gov)

<http://dx.doi.org/10.1016/j.immuni.2016.12.008>

New technologies have been propelling dramatic increases in the volume and diversity of large-scale public data, which can potentially be reused to answer questions beyond those originally envisioned. However, this often requires computational and statistical skills beyond the reach of most bench scientists. The development of educational and accessible computational tools is thus critical, as are crowdsourcing efforts that utilize the community's expertise to curate public data for hypothesis generation and testing. Here we review the history of public-data reuse and argue for greater incorporation of computational and statistical sciences into the biomedical education curriculum and the development of biologist-friendly crowdsourcing tools. Finally, we provide a resource list for the reuse of public data and highlight an illustrative crowdsourcing exercise to explore public gene-expression data of human autoimmune diseases and corresponding mouse models. Through education, tool development, and community engagement, immunologists will be poised to transform public data into biological insights.





# FAIR GUIDELINES

**F**indable  
**A**ccessible  
**I**nteroperable  
**R**eusable

Image Courtesy: Dr. Patti Brennan



[Comment](#) | [OPEN](#) | Published: 15 March 2016

## The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons [✉](#)

*Scientific Data* **3**, Article number: 160018 (2016) | [Download Citation](#) [↓](#)

## REVISIT DATA MANAGEMENT & SHARING PLANS

In compliance with current NIH requirements, we herewith outline a plan to share materials and our management of intellectual property. We will adhere to the NIH Grant Policy on Sharing of Unique Research Resources including the Sharing of Biomedical Research Resources Principles and Guidelines for Researchers and Reviewers available at [http://ottd.nih.gov/policy/ottd\\_guide\\_final.html](http://ottd.nih.gov/policy/ottd_guide_final.html). We will deposit all genetic and phenotypic data upon request. Material will be deposited in the Sample Bank A requirements. Material such as microbial strains will be deposited in the Microbial Strain Repository or other public repositories. The BE arrangement would be followed in such cases. If requires a patent, we will ensure that the techs available to the research community in accordance with the research community in techs Principles and Guidelines document.

Data and/or Reagent Type	Public Repository for Raw and Finished Data
RNA sequencing (transcriptomics)	GEO / NCBI
DNA Sequencing (clinical isolates)	NCBI SRA
Genome Assemblies and associated metadata	NCBI Assembly and/or NIAID BRC or other repositories (e.g., PATRIC)
Model Organisms / Strains	Stock center / BEl Resources Repository / or other approved repository
Other data types (e.g., metabolomics)	NIAID BRCs or other repositories (e.g., PATRIC)

## U01 A134255 PI: YOUNG, VINCENT (Contact)

Age	Description	Data Type
1A.1	Median risk of CVD acquisition	CIM
1A.2	Service of acquisition	CIM, 100, MET
1A.3	Median risk of severe disease	CIM
1A.4	Service of severe disease	CIM, 100, MET
1B.1	GI environment stress	CIM, 100
1B.2	Genetic associations with CVD	CIM, 100, 9000
1B.3	Mitochondrial CVD occurrence	100
1A.5	Baseline colonization (ICU)	CIM
1A.6	ICU colonization acquisition	CIM
2B.1	Mitochondrial long-term colonization	CIM, 9000, 100
2B.2	Mitochondrial colonization acquisition	CIM, 9000, 100

To summarize the data types and resources that will be generated:

- 1) **Raw sequencing data**
  - a. 150-endpoint gene amplicon library sequencing
  - b. C. difficile genome sequence data (WGS)
  - c. Metagenomic sequence datasets
  - d. Metatranscriptomic sequence datasets
- 2) **Metadatasets**
  - a. Targeted analysis
  - b. Untargeted analysis
- 3) **Clinical metadata**
- 4) **Fecal specimens**
  - a. Frozen feces
  - b. Fecal swabs
- 5) **C. difficile isolates**

The omics data in this proposal can be split into 4 categories. These data are all collected with Illumina sequencing technology, which are shared with the larger scientific community through appropriate databases including the NCBI Sequence Read Archive for raw sequencing data, or the NCBI Gene Expression Omnibus.

**TABLE 1**

F<sub>indable</sub> A<sub>ccessible</sub> I<sub>nteroperable</sub> R<sub>eusable</sub>

Distinguished Professor of Data Science



<sup>2</sup> One or more deposition sites will be identified as appropriate based on the NND / DMO Data Sharing & Release Guidelines (DNRG).

For data generators/producers, users of [Data Release](#) releases take strict responsibility to recognize the scientific contribution of the data generators/producers by following normal standards of scientific etiquette and fair use of unpublished data. Such guidelines can be found in [Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility](#) (PDF) / Toronto Data Release Workshop [2006] *Nature* 441:168-170.

a) Data sharing: Participants, University of Houston, Texas A&M and UCLA are committed to public dissemination of its scientific results and developing health care products that improve global public health while protecting the intellectual property essential for commercial development. Research data will be shared according to the most recent NIH policies on data sharing and open access to research data, subject to any applicable laws, regulations, and/or institutional policies.

publicly available as soon as possible. If it, they will be quickly protected to facilitate a need to more systematic, scientific research.

review of Metagenome, Metatranscriptome,

ce of the investigators on this project

erated at BCM and kept in secure data components that support information sharing. The CMAMR and TICAM leverage existing in the overall computational capacity of would attempt to duplicate many of the studies all heterogeneous sample handling, performed in a mature state-of-the-art genomic technologies. The CMAMR is adding to the HGGC. To elaborate further, genotypically managed and maintained data, currently, and about 1.25 petabytes (PB) of data

1 this proposal are committed to  
ion. Specifically, we propose to  
research community by depositing  
following the NAID Data Sharing  
re generated from local samples  
fta, we will exercise the strictest  
a to the research community. As  
are and distribute the research

all times, including dissemination  
identified data formats.  
rds (e.g., Data Use Agreements)  
others interested in the project(s).  
locally available via [DEI Resources](#)

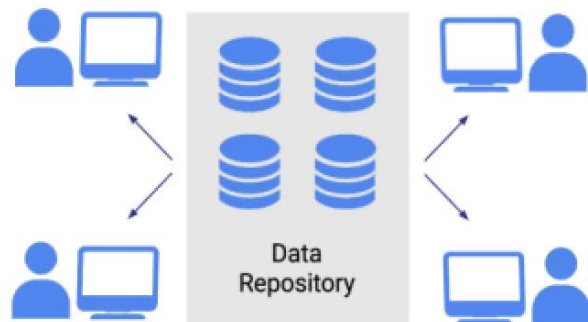
† guidance from Program staff.

dis-*per-sample*) will be evaluated

- Whole-metagenomic sequencing data (100 samples / year) will be submitted in a time interval of every 6 months, online accessible 50 samples into one

# CO-LOCATE DATA & TOOLS ON CLOUD

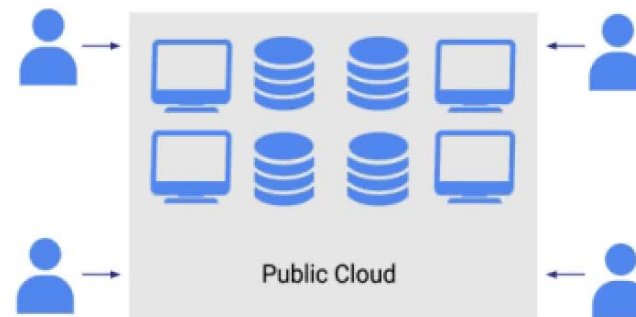
Traditional Way: Bring data to the researchers



## Problems

Data sharing = data copying  
Limits Access  
Largely fixed compute  
Individual security implementations

Cloud Way: Bring researchers to the data



## Solutions

True data sharing  
Democratizes access  
Elastic compute and storage  
Centralized security implementation

# MINT DIGITAL OBJECT IDENTIFIERS



# DATA INDEXING



**Google**  
Big Query

**Google** Dataset Search Beta

Search for Datasets



Try [boston education data](#) or [weather site:noaa.gov](#)

Not an endorsement but an example

# USE OF JUPYTER NOTEBOOKS



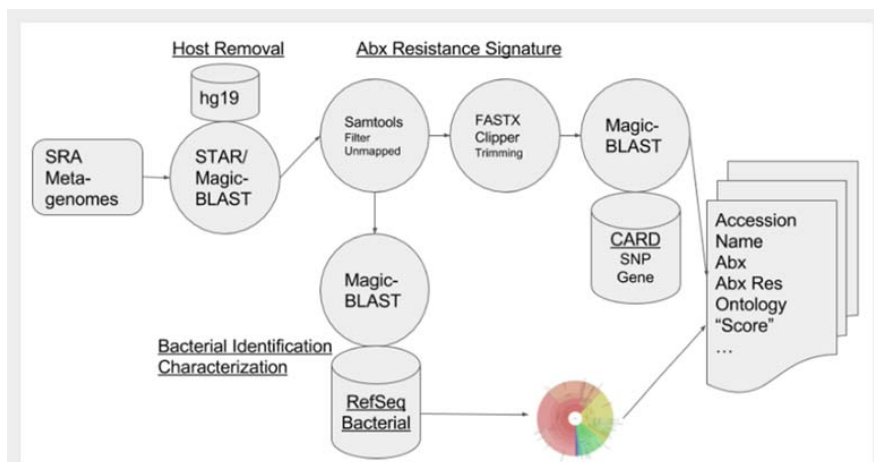
Not an endorsement but an example

# CONTAINERIZATIONS OF TOOLS

## NIH Data Science Learning and Testing Hackathon

February 23, 2018!

The NIH will host a Learning and Testing Data Science hackathon on February 23rd, 2018 on the main campus in Bethesda, MD. Learners will test alpha and beta code that have been generated in full, collaborative development hackathons for a wide range of scientific problems, including general bioinformatics and genomic analyses in addition to text, image, and sequence processing. This event is for researchers who are in the early stages of their data science journey, including students and postdocs. Other non-scientific developers, mathematicians, or librarians in a similar educational place are also welcome! Learning in this event will be primarily hands-on and self-driven, but will also include short workshops on topics such as Docker and GitHub. Mentors will also be available to assist with questions.



**F1000Research**  
Open for Science

Search

BROWSEGATEWAYSHOW TO PUBLISHABOUTBLOG

Check for updates

SOFTWARE TOOL ARTICLE

**NastyBugs: A simple method for extracting antimicrobial resistance information from metagenomes [version 1; referees: 2 approved with reservations]**

✉ Hsinyi Tsang<sup>1</sup>, Matthew Moss<sup>2</sup>, Greg Fedewa<sup>3</sup>, Sharif Farag<sup>4</sup>, Daniel Quang<sup>5</sup>, Alexey V. Rakov<sup>6</sup>

✉ Ben Busby<sup>7</sup>

Author details

Grant information

This article is included in the Hackathons collection.

**Abstract**

Multidrug resistant bacteria are becoming a major threat to global public health. While there are many possible causes for this, there have so far been few adequate solutions to this problem. One of the major causes is a lack of clinical tools for efficient selection of an antibiotic in a reliable way. NastyBugs is a new program that can identify what type of antimicrobial resistance is most likely present in a metagenomic sample, which will allow for both smarter drug selection by clinicians and faster research in an academic environment.

**METRICS**

462

VIEWS

185

DOWNLOADS

Get PDF

Get XML

Cite

Export

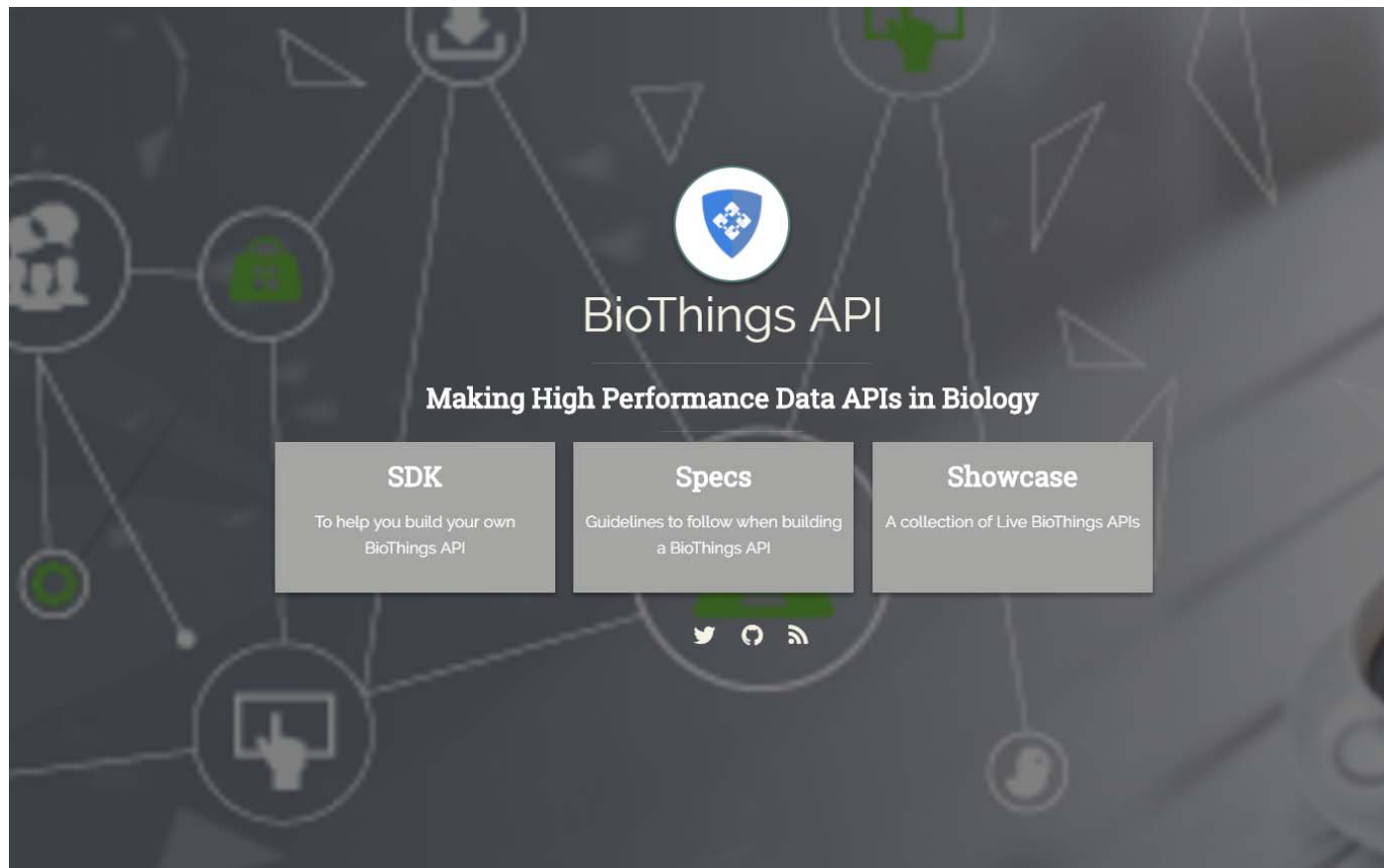
Track

Email

Share



# USE OF APPLICATION PROGRAMMING INTERFACES



Not an endorsement but an example

# USE OF SMART-IRB, AUTHN, AUTHZ SOLUTIONS



Supporting single IRB review  
Advancing collaborative research



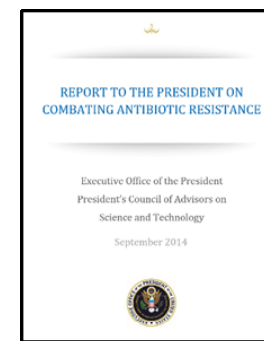
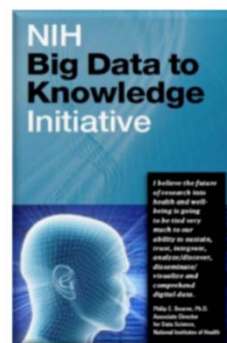
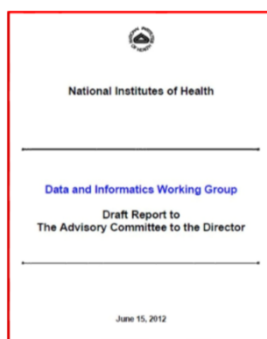
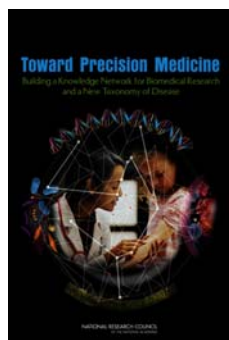
**DUOS**



Not an endorsement but an example

# REALIZE DATA SHARING

- Aligns with multiple federal and NIH Missions:
  - 2011: Precision Medicine Report
  - 2012: Data and Informatics Working Group (DIWG) of the Advisory Committee to the NIH Director
  - 2013: OSTP White House Memo - Increasing Access to Federally Funded Scientific Research Results
  - 2013: NIH Big Data to Knowledge (BD2K) Initiative
  - 2014: National Strategy for Combating Antibiotic-Resistant Bacteria (CARB)
  - 2018: NIH Strategic Plan for Data Science



# EXPAND USER BASE - INTERACTIONS OF USERS & CONTRIBUTORS



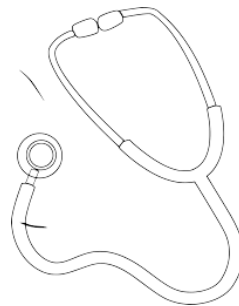
**PI / Biologist**  
web access



**Computational  
Research Scientist**  
Python, R, SQL



**Algorithm Developer**  
ssh, programmatic  
access



## ANTICIPATED OUTCOMES – DEMOCRATIZE RESOURCES

Discover & access relevant generated data, tools, & standards

Identify/provide home for research data that needs a home

Enable new paradigm/model of data-driven hypothesis-driven science

Continuous assessment and optimization of data science research resources

Reduce informatics costs in the long run

# FUNDING OPPORTUNITY ANNOUNCEMENTS (FOA)

## Secondary Analysis of Existing Datasets for Advancing Infectious Disease Research (R21 Clinical Trial Not Allowed)

R21 Exploratory/Developmental Research Grant

New

None

PA-19-068

## Informatics Methodology and Secondary Analyses To Explore Shared Immunology Study Data in ImmPort

Program Announcement With Special Receipt, Referral, and or Review Considerations—proposed FY 2020 initiative

# Webale

Ishwar Chandramouliswaran  
Office of Data Science & Emerging Technologies  
NIAID, NIH, DHHS



@ishwartweets  
<https://www.linkedin.com/in/ishwarc/>

