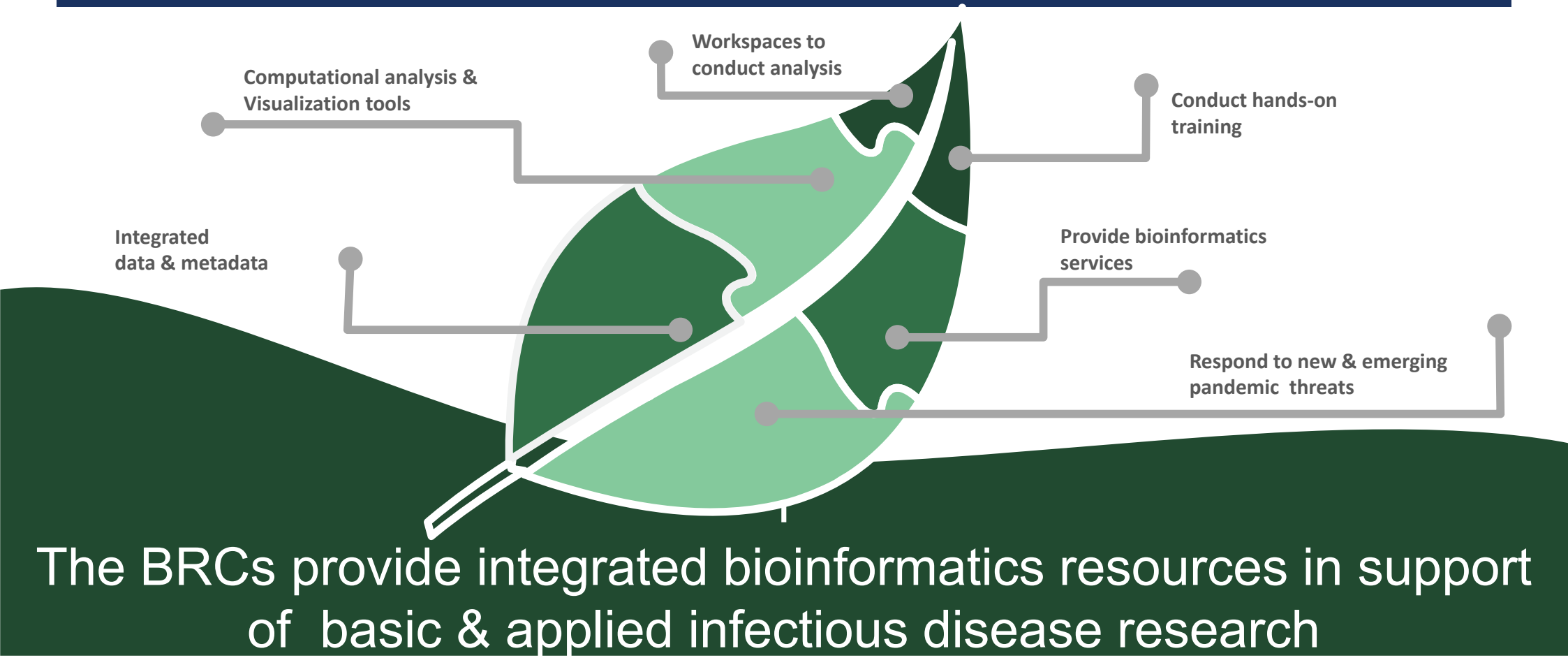
BIOINFORMATICS RESOURCE CENTERS

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

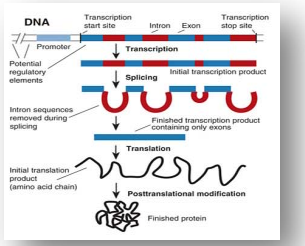
THE CENTERS ORGANIZED BY PATHOGEN GROUPS

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

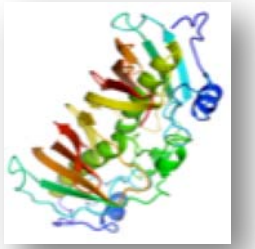
COLLABORATE TO INTEGRATE DATA ACROSS NIAID PROGRAMS

Sequencing

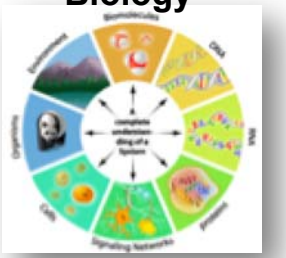
Functional Genomics



Structural



Systems Biology



Bioinformatics

Genomic Research Resources

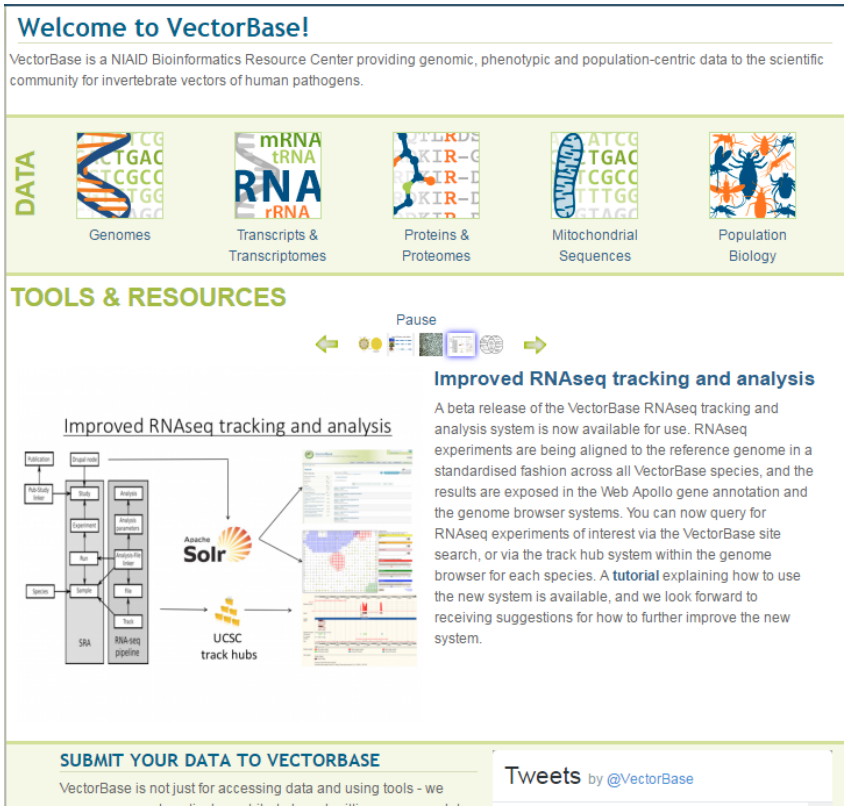
Genomic/Omics Data Sets, Databases, Bioinformatics Tools, Biomarkers, 3D Structures, Protein Clones, Predictive Models

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



EUPATHDB
BRC FOR INVERTEBRATE VECTORS OF HUMAN PATHOGENS

An unique resource for protozoan & fungal pathogens

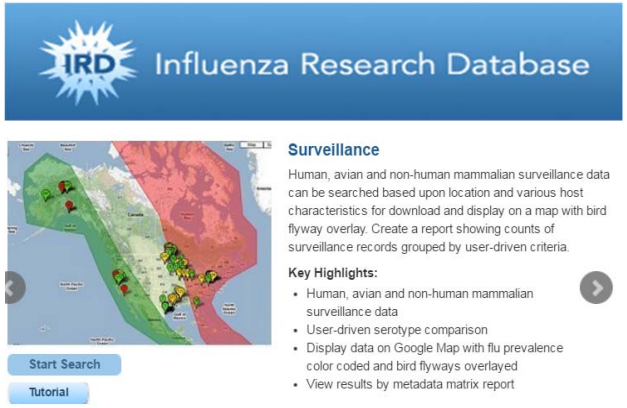
- A comprehensive collection genomic and functional genomic data, isolate data and phylogenomics
  - genome sequences and annotation, strand-specific RNA-seq data, splice junction predictions, phosphoproteomic data, high-throughput phenotyping data, single nucleotide polymorphism, and expression quantitative 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 orthlogy:*
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 (FLUDB)

- **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

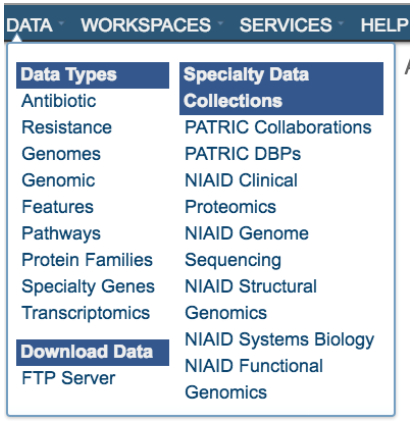


PATHOGEN RESOURCE INTEGRATION CENTER (PATRIC)
BRC FOR BACTERIAL SPECIES

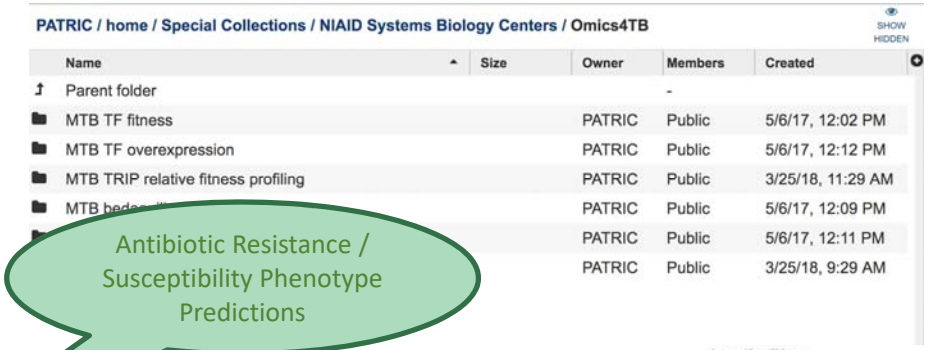
Integrated Data Collections

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

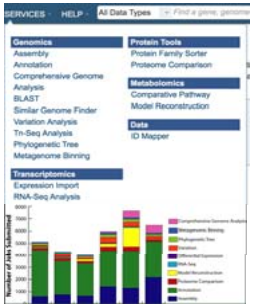


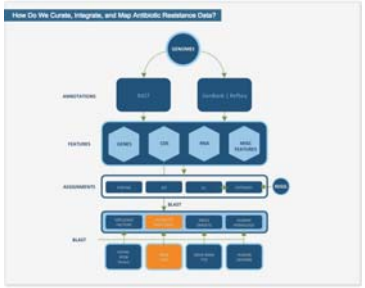
Personal Workspace for analysis of User data and "Virtual Integration" of user data in the context of all the public datasets

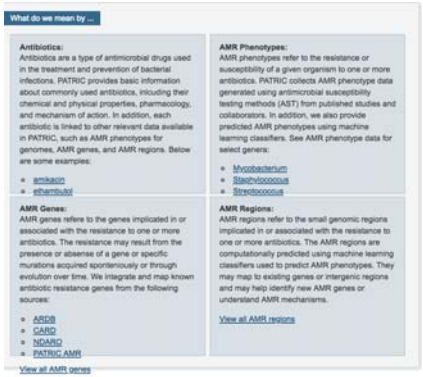


Antibiotic Resistance / Susceptibility Phenotype Predictions

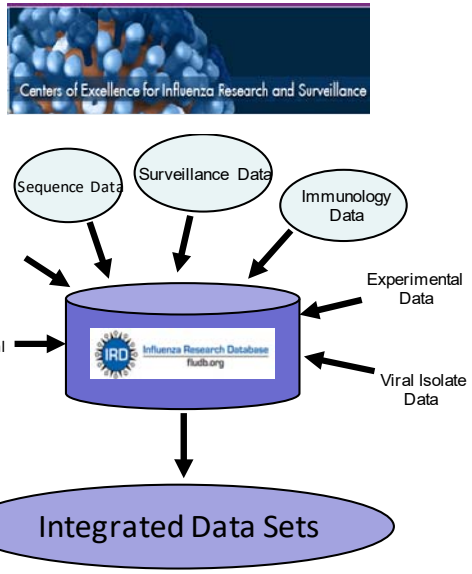
Computational Analysis Services

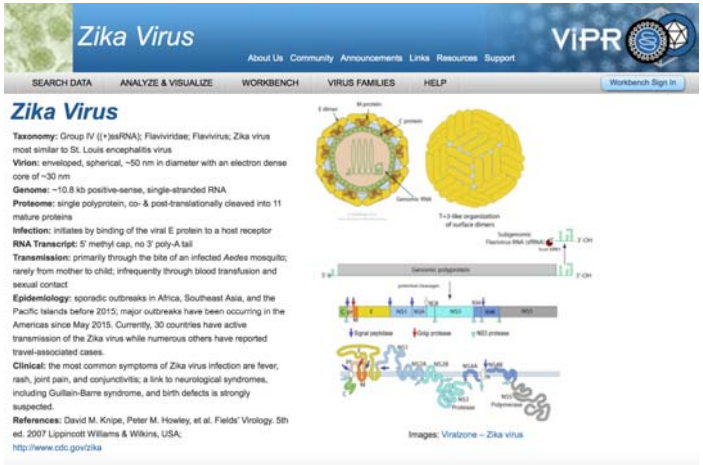


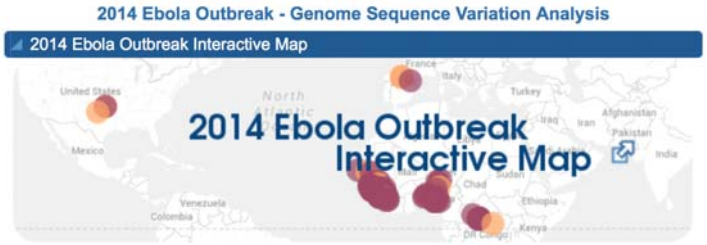


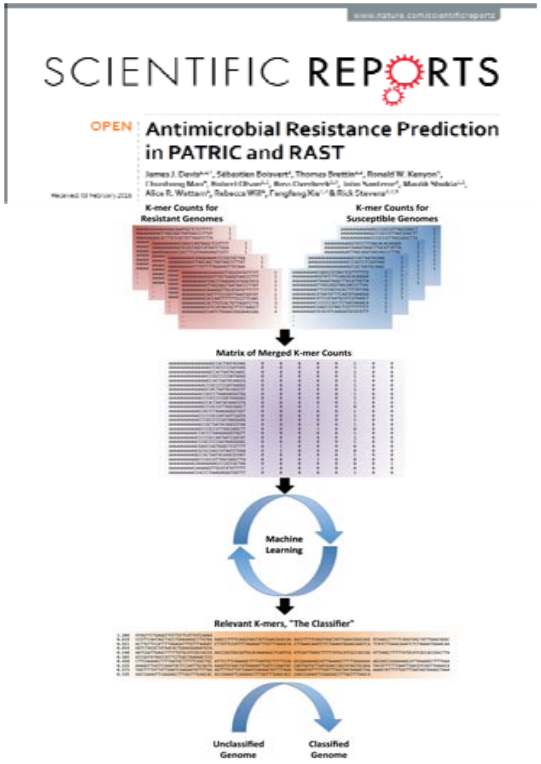


RESPOND TO NEEDS OF ID RESEARCHERS







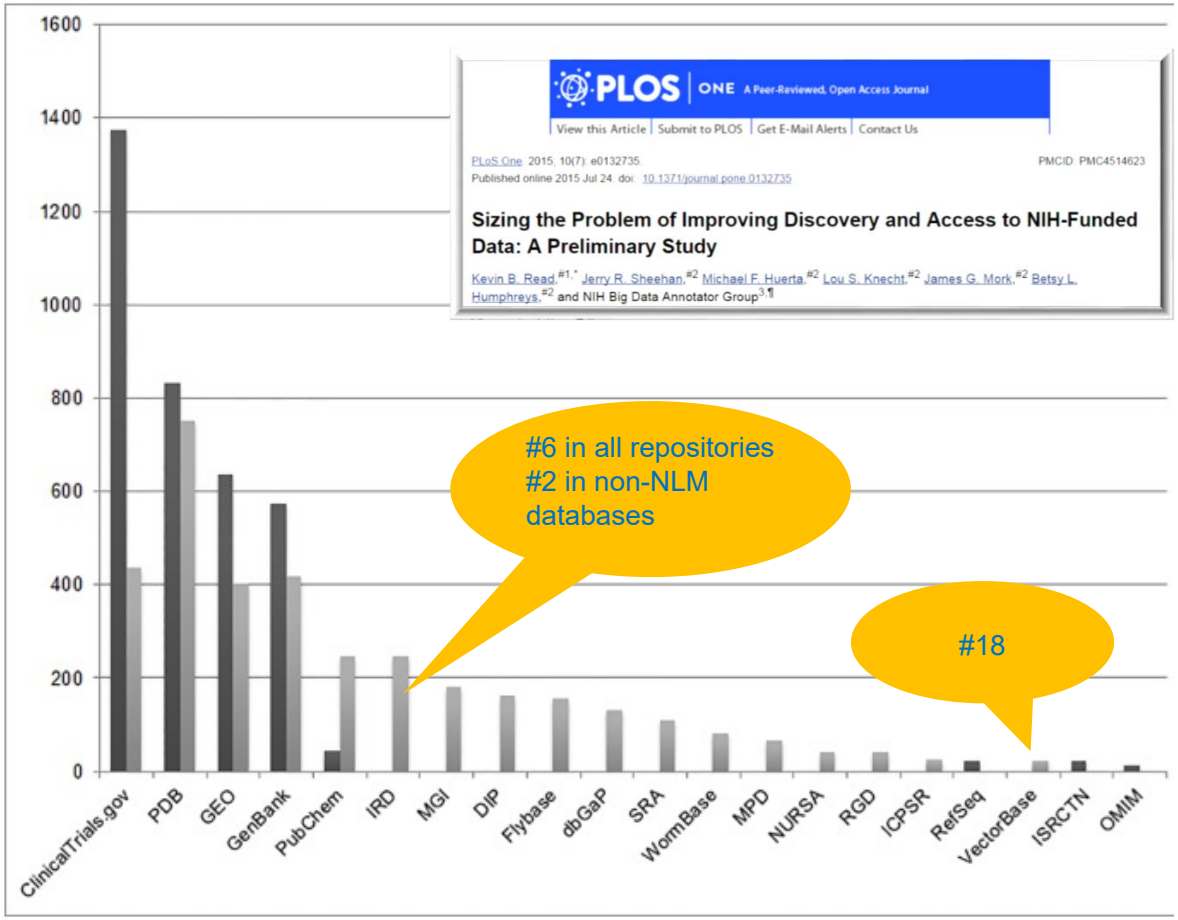


- **Clinical & Epidemiological Data**
- **PopBio Data & Tools**

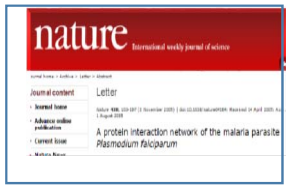
- **AMR data collection**
- **AMR prediction tools**

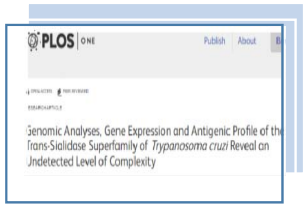
HANDS-ON TRAINING WORKSHOPS GLOBALLY

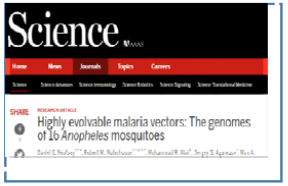
TO INCREASE USAGE & IMPACT

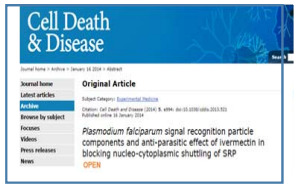


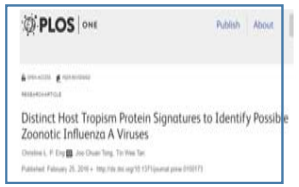
SCIENCE ENABLED BY THE BIOINFORMATICS RESOURCE CENTERS

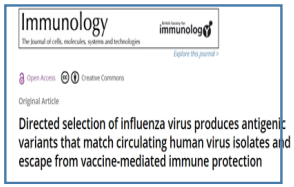


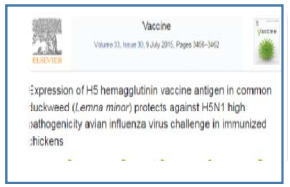


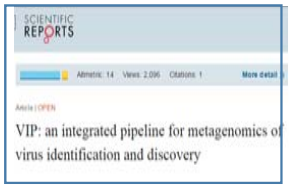




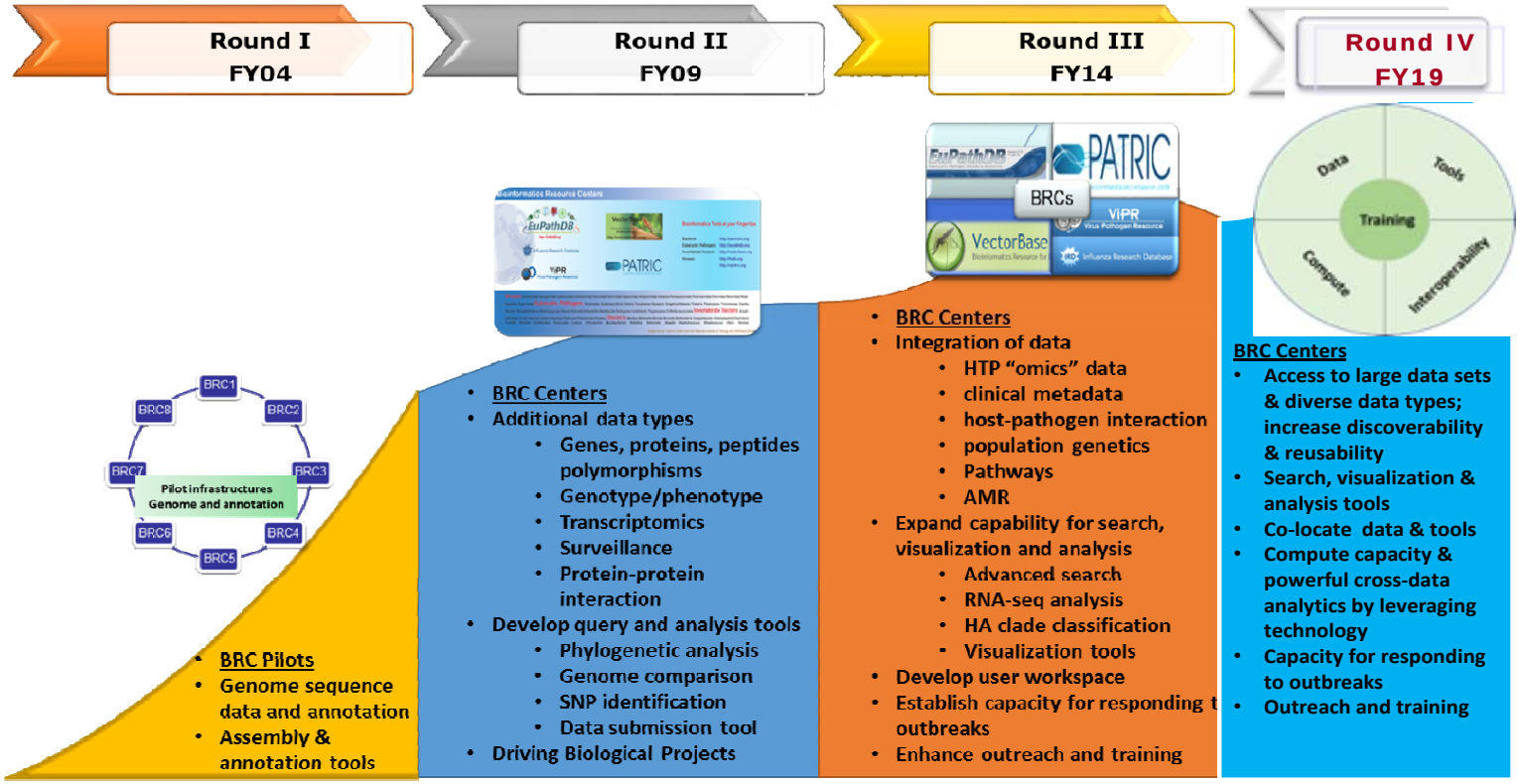








EVOLUTION OF THE RESOURCES



THE NIAID DATA CHALLENGE

CEIRS

ARLG

CHAVI-ID

ITN

VTEUs

HIPC

Functional Genomics

Structural Genomics

Systems Biology

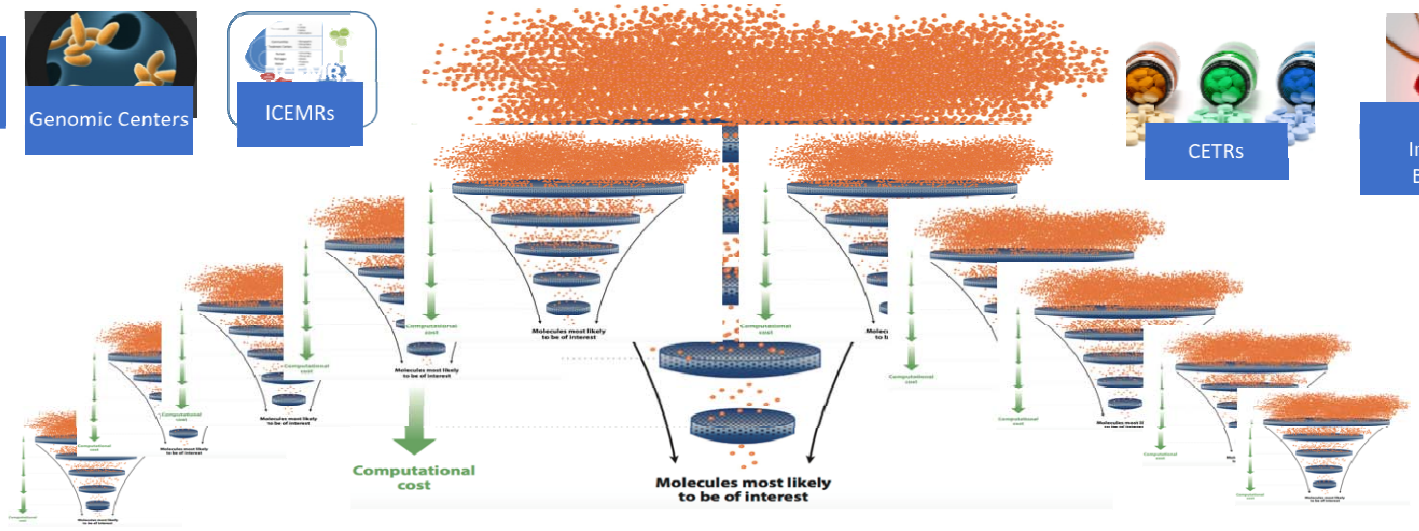
Genomic Centers

ICEMRs

CETRs

Modeling Immunity in Biodefense

NIAID DIR/VRC



RCSB PDB PROTEIN DATA BANK

db GaP GENOTYPES and PHENOTYPES

National Center for Biotechnology Information NCBI

IMMPORT

Influenza Research Database

Immune Tolerance Network TrialShare Clinical Trials Research Portal

PATRIC

ImmuneSpace

VectorBase Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

EuPathDB Eukaryotic Pathogen Database Resources

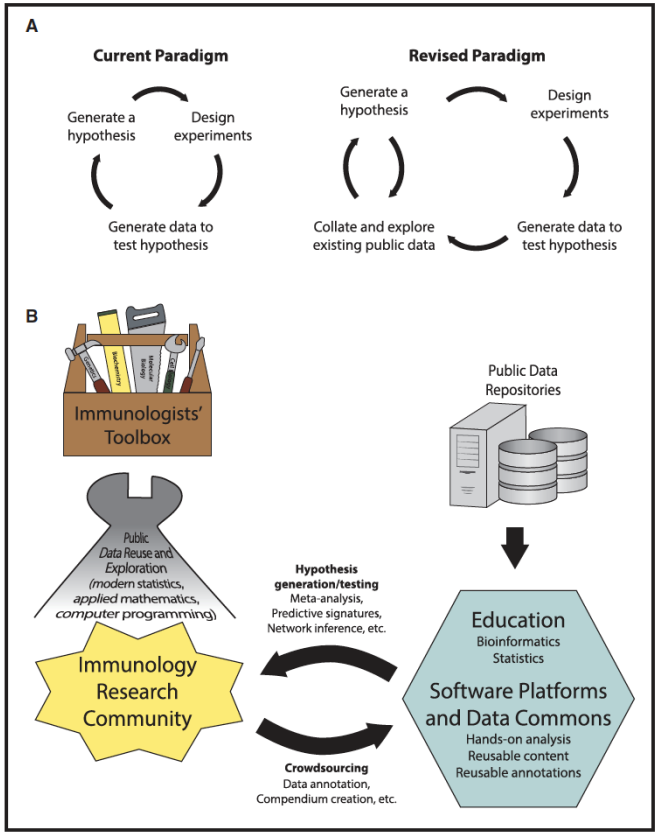
Image Courtesy: Dr. Emily Erbelding (modified)

NEW MODELS OF DATA-DRIVEN HYPOTHESIS GENERATION

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

Rachel Sparks,[1] William W. Lau,[2] and John S. Tsang[1,*]
[1]Systems Genomics and Bioinformatics Unit, Laboratory of Systems Biology, National Institutes of Allergy and Infectious Diseases, National Institutes of Health
[2]Office of Intramural Research, Center for Information Technology, National Institutes of Health
*Correspondence: 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

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

Findable

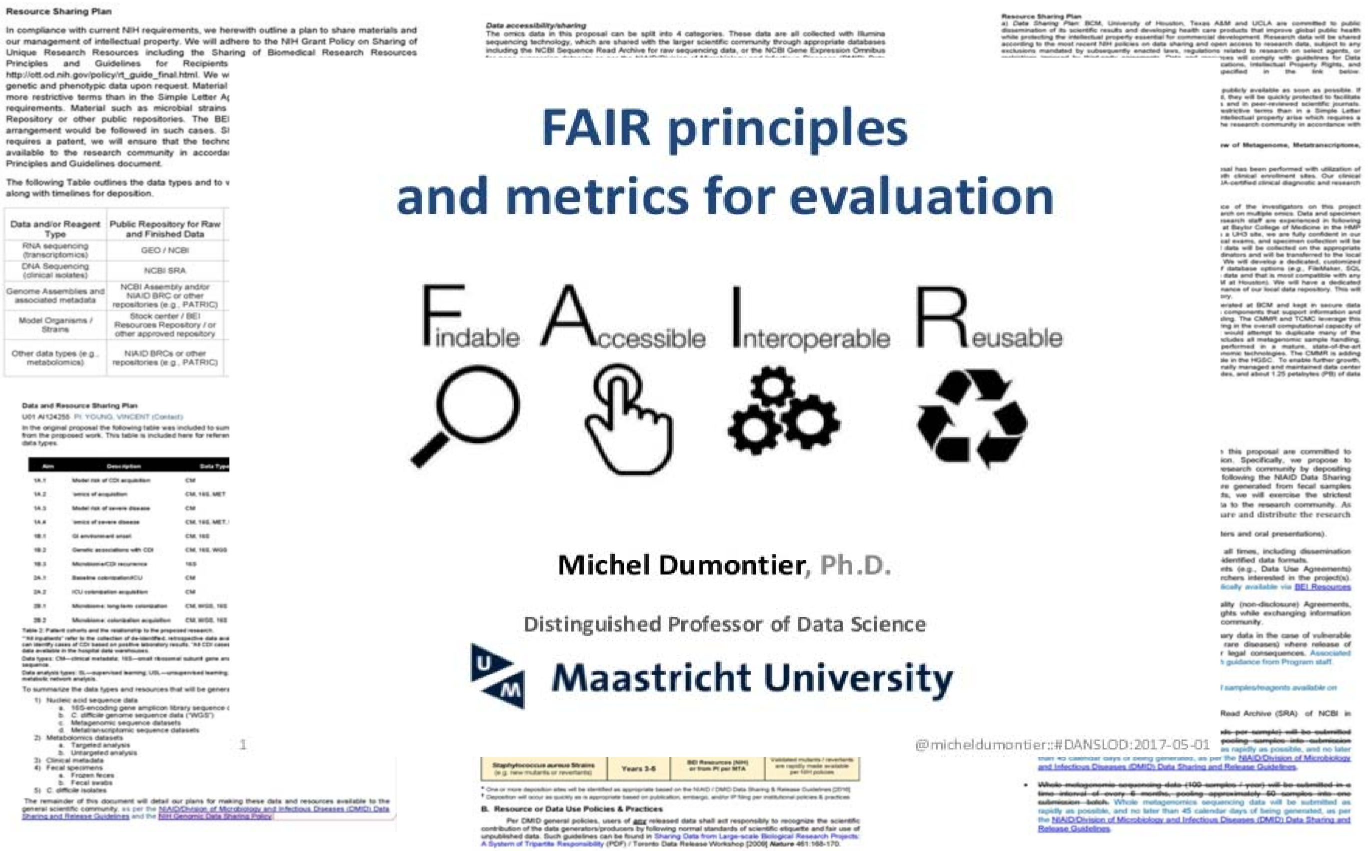
Accessible

Interoperable

Reusable

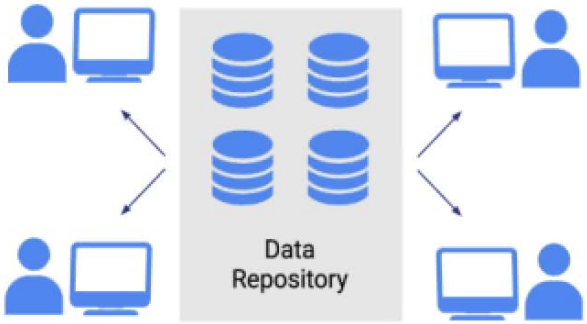
Image Courtesy: Dr. Patti Brennan

REVISIT DATA MANAGEMENT & SHARING PLANS



CO-LOCATE DATA & TOOLS ON CLOUD

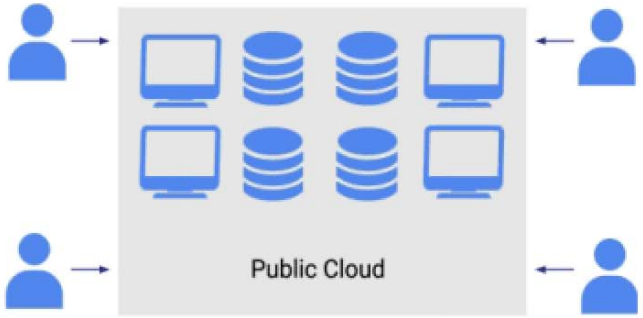
Traditional Way: Bring data to the researchers



Problems

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

Cloud Way: Bring researchers to the data



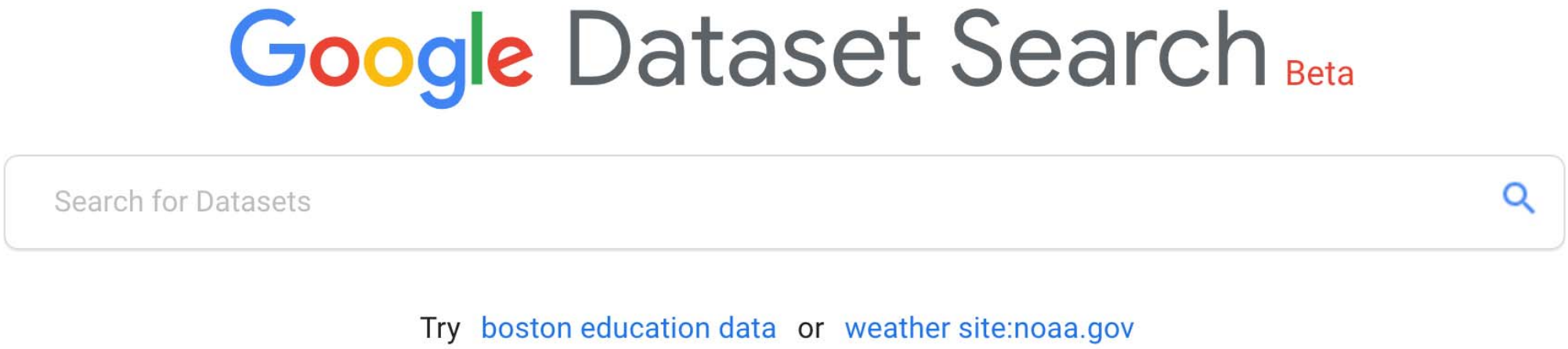
Solutions

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

Image Courtesy: Broad Institute

MINT DIGITAL OBJECT IDENTIFIERS

DATA INDEXING



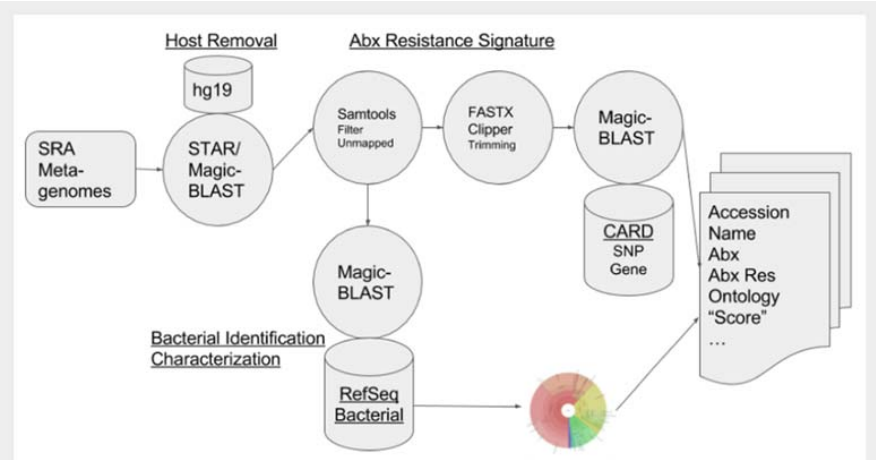
Not an endorsement but an example

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

BROWSE GATEWAYS HOW TO PUBLISH ABOUT BLOG

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[1], Matthew Moss[2], Greg Fedewa[3], Sharif Farag[4], Daniel Quang[5], Alexey V. Rakov[6], Ben Busby[7]

Author details

Grant information

This article is included in the Hackathons collection.

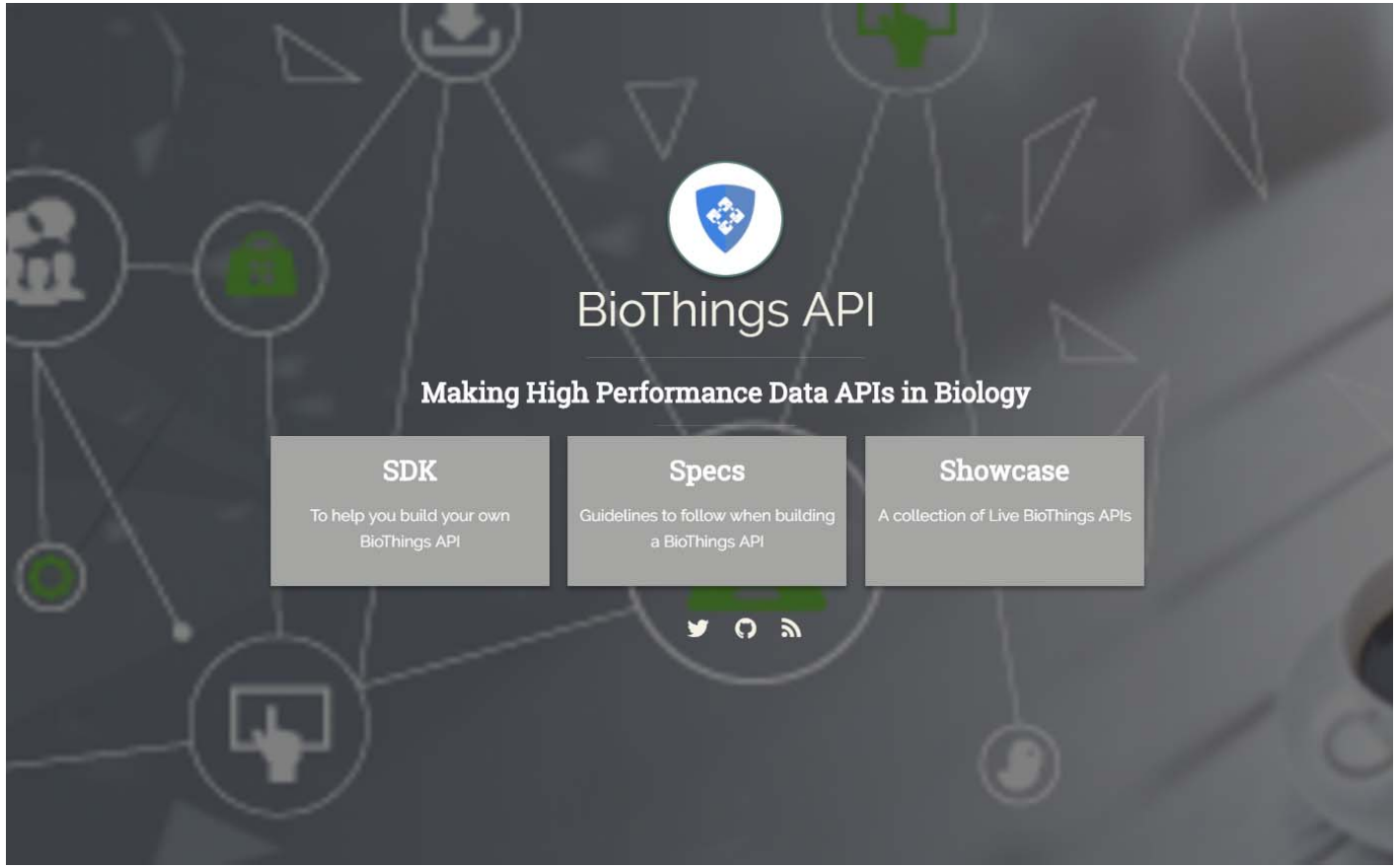
METRICS
462
VIEWS
185
DOWNLOADS

Get PDF
Get XML
Cite
Export
Track
Email
Share

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.

USE OF APPLICATION PROGRAMMING INTERFACES



Not an endorsement but an example

USE OF SMART-IRB, AUTHN, AUTHZ SOLUTIONS

SMART IRB

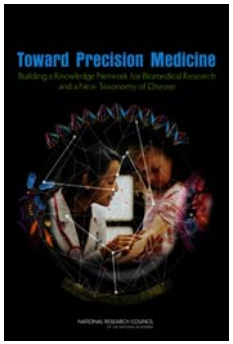
Supporting single IRB review
Advancing collaborative research

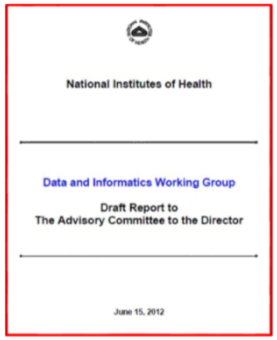
DUOS

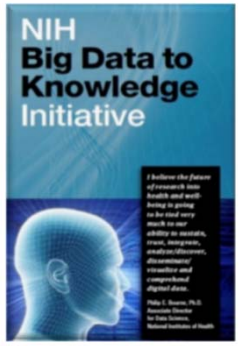
Not an endorsement but an example

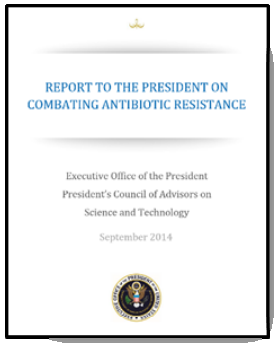
REALIZE DATA SHARING

- 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

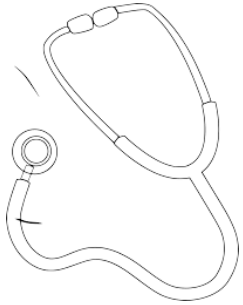


Image Courtesy: Dr. Sheila Reynolds, ISB

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/