Transitioning Bioinformatics Core to Support Biomedical AI/ML Research: Lessons Learned

Yang C. Fann, Ph.D.

Director, IT and Bioinformatics Program

NINDS/NIH/HHS

Disclaimer

- The opinions or assertions contained herein are the private ones of the speaker and are not to be construed as official or reflecting the views or policies of the speaker's employer (NINDS/NIH/HHS) nor the US government. Therefore, they can not be liable or assume any responsibility for any consequence of using information presented.

- This presentation may contains forward-looking statements regarding future plans that are based on current projections and milestones which may change overtime and are only intended for educational and knowledge sharing at this meeting.

- The speaker makes no claims of any conflict of interest with any party mentioned in this presentation.

Neuroscience Bioinformatics Research


NINDS Bioinformatics

DNA-Seq	Biological Function Enrichment
RNA-Seq	Canonical Pathway Enrichment
miRNA-Seq	Network Analysis
Exome-Seq	Mechanism Deconvolution
ChIP-Seq	BioMarker Discovery
SNP/DIP Discovery	Diagnostic Development
HLA Analysis	GWAS Analysis
Promoter Modeling	Theranostic Development
Gene Expression	Prognostic Development
Q-PCR Analysis	RNA Structure
Protein Structure	Methylation Analysis

Oncoming of AI/ML/DL in Biomedicine@NIH

- Clinical applications
  - Pathology diagnostics
  - Dermatology, ophthalmology diagnostics
  - Radiology –mammograms, CXRs, CT, etc…
  - Inferring treatment options for cancer, complicated diseases
  - Robotic guided surgery
  - Natural language processing of EHR data

- Basic science applications
  - Interpretation of images: cryo-EM, confocal, etc.
  - Neuroscience and the BRAIN initiative
  - Genomics: variants and risk of disease, gene structure
  - Microbiome/metagenomics
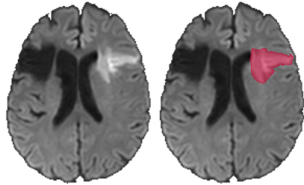 - Epigenomics: histone marks, TF binding, enhancers, DNA methylation

From Bioinformatics to Data Science & A.I. Research
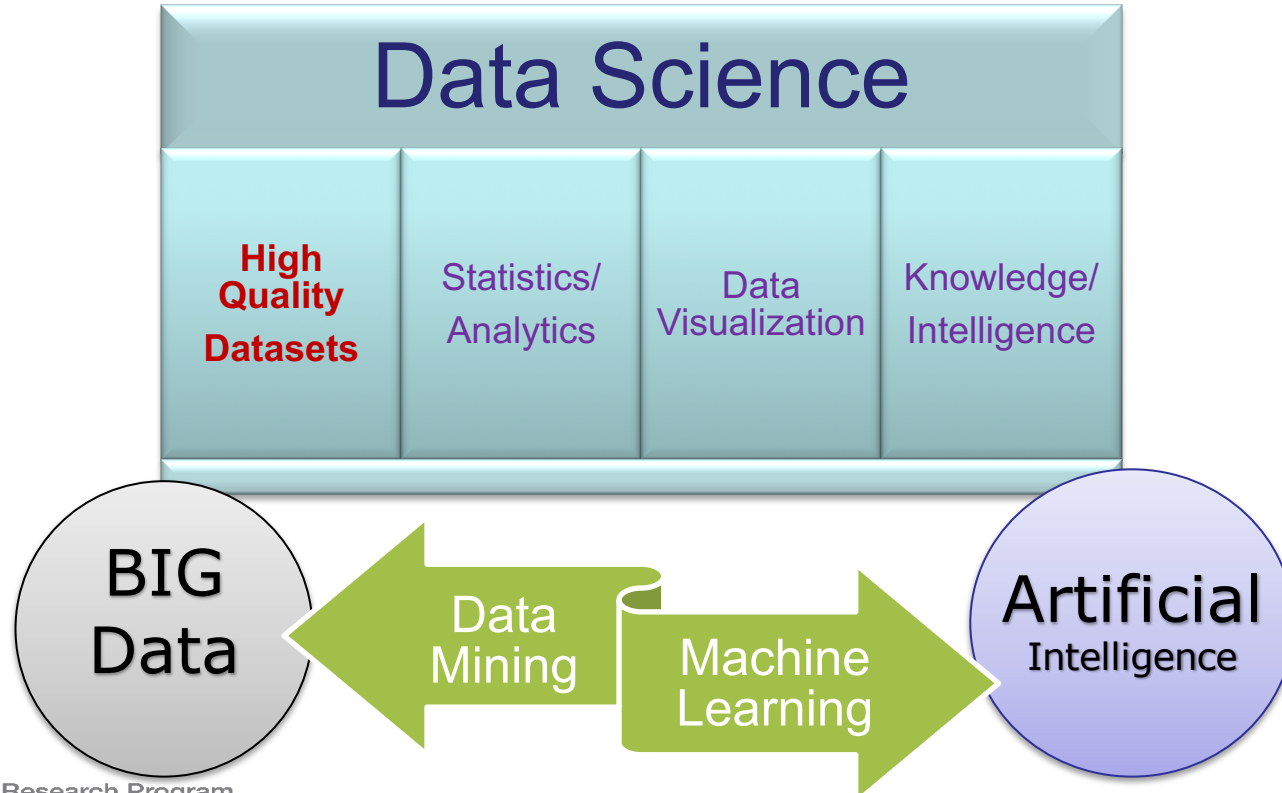
BIG Data Analytics/A.I. Research for Stroke

From Recognition, Classification, Prediction to Inferring & Simulation

Acute ER --> Diagnosis --> Treatment --> Outcome --> Recovery

Brain Lesion

- ER (vital, lab, etc.)
- Imaging (CT/MRI)
- **NIHSS/mRS**
- Clinical Assessment
- Treatment/ medication
- Follow-ups

Aim: Could we predict and treat for best outcomes of Stroke patients to raise the standard Stroke care across the nation?

Medical profile, diagnosis, EHR, images, outcomes, genomics, etc.

Prediction/Prevention Health Management A.I.

A.I. for Diagnosis & Clinical Decision Support
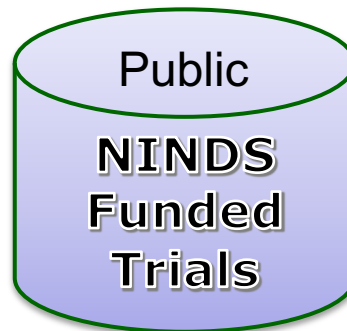
Rehab and Recovery Care Management A.I.

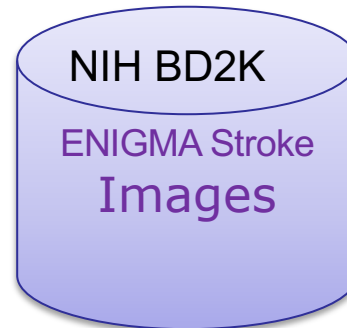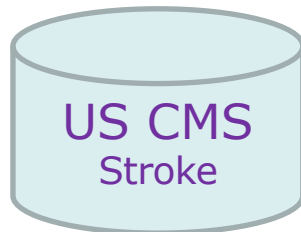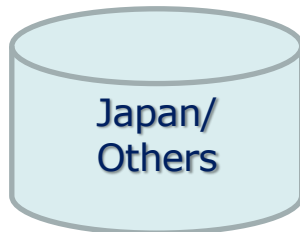# What Stroke Datasets do We Currently Have?

What Do You Need to Start a (Biomedical) A.I. Team?

- What Data Science (A.I.-ready) projects to you have?
  - What are the A.I. challenges and end goals (tool, product, publication or else)?

- What information (datasets) or tools do you have or need?
  - **Quality/quantity** and knowledge of datasets (BIG data is not always better)
  - *What **algorithms and tools** to test/use? (Dataset and end goal dependent!)*

- What are the potential barriers to overcome (institution dependent)?
  - Self-sufficient or with dependency (e.g. needs more data sources, integration or new tool development?)
  - Regulatory (potential IRB/patent/FDA) for datasets
  - ***What skillsets do you or your team currently have?***
  - ***Inter-discipline domain knowledge team (postdoc/collaboration)***

- Resources (infrastructure, staff and $$!)

NINDS IT and Bioinformatics Program

Neuroscience Bioinformatics

Biomedical Informatics
Data & Tool team

NIH Intramural
Projects/Initiatives

AI/ML
Data Scientists
Domain Knowledge

NIH › U.S. National Library of Medicine

Databases | Find, Read, Learn | Explore NLM | Research at NLM | NLM for You

NLM Customer Support

NIH › Trans-NIH BioMedical Informatics Coordinating Committee (BMIC)

BMIC Home | CDE Resource Portal

Home

NIH Data Sharing Repositories

This table lists NIH-supported data repositories that make data accessible for reuse. Most accept submissions of appropriate data from NIH-funded investigators (and others), but some restrict data submission to only those researchers involved in a specific research network. Also included are resources that aggregate information about biomedical data and information sharing systems. The table can be sorted according to name and by NIH Institute or Center and may be searched using keywords so that you can find repositories more relevant to your data. Links are provided to information about submitting data to and accessing data from the listed repositories. Additional information about the repositories and points-of-contact for further information or inquiries can be found on the websites of the individual repositories. Are we missing a data sharing repository? Contact us.

Show 50 entries

Search:

IC	Repository Name	Repository Description	Data Submission Policy	Access to Data
NCI	Cancer Nanotechnology Laboratory (caNanoLab)	caNanoLab is a data sharing portal designed to facilitate information sharing in the biomedical nanotechnology research community to expedite and validate the use of nanotechnology in biomedicine. caNanoLab provides support for the annotation of nanomaterials with characterizations resulting from physico-chemical, in vitro, and in vivo assays and the sharing of these characterizations and associated nanotechnology protocols in a secure fashion.	How to submit your data to caNanoLab	How to access caNanoLab data
NCI	The Cancer Imaging Archive (TCIA)	The image data in The Cancer Imaging Archive (TCIA) is organized into purpose-built collections of subjects. The subjects typically have a cancer type and/or anatomical site (lung, brain, etc.) in common.	How to submit data to TCIA	How to access TCIA data
NCI (NHGRI, NIGMS)	PeptideAtlas	PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences.	How to submit data to PeptideAtlas	How to access Peptide Atlas data
NEI	EyeGENE®	The eyeGENE® Biorepository and corresponding Database contain family history and clinical eye exam data from subjects enrolled in eyeGENE® Program coupled to clinical grade DNA samples. This data and samples are submitted by collaborators throughout the US and Canada and the data is available on a controlled access basis to researchers world-wide.	How to submit data to EyeGENE®	How to access EyeGENE® data
NHGRI	FlyBase: A Drosophila Genomic and Genetic Database	Drosophila Genomic and Genetic database that includes proteomics data, microarrays and Tiling BAC's.	How to submit data to Flybase	How to access Flybase data
NHGRI	The Zebrafish Model Organism Database (ZFIN)	ZFIN serves as the zebrafish model organism database. It aims to: a) be the community database resource for the laboratory use of zebrafish, b) develop and support integrated zebrafish genetic, genomic and developmental information, c) maintain the definitive reference data sets of zebrafish research information, d) to link this information extensively to corresponding data in other model organism and human databases, e) facilitate the use of zebrafish as a model for human biology, and f) serve the needs of the research community.	How to submit data to ZFIN	How to access ZFIN data
NHGRI	WormBase	WormBase is an international consortium of biologists and computer scientists dedicated to providing the research community with accurate, current, accessible information concerning the genetics, genomics and biology of C. elegans and related nematodes.	How to submit data to WormBase	How to access WormBase data
NHGRI/NIGMS	The Universal Protein Resource (UniProt)	The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc).	How to submit data to UniProt	How to access UniProt data
NHLBI	Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC)	The goal of BioLINCC is to facilitate and coordinate the existing activities of the NHLBI Biorepository and the Data Repository and to expand their scope and usability to the scientific community through a single web-based user interface.	How to submit data to BioLINCC	How to access BioLINCC data

https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Question?

Email: fann@ninds.nih.gov