NeuroBridge: Connecting Big Data for Reproducible Clinical Neuroscience

A National Institute on Drug Abuse funded project for data sharing, discovery and selection of Neuroscience datasets

Howard Lander
Renaissance Computing Institute
The University of North Carolina at Chapel Hill

renci
RESEARCH \ ENGAGEMENT \ INNOVATION

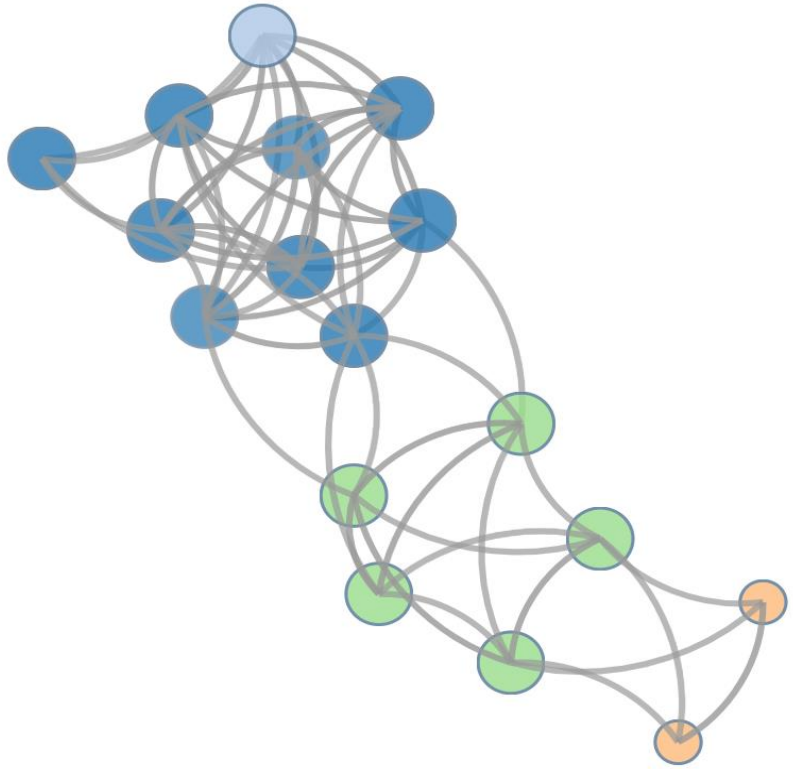
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

What is the NeuroBridge?

- A collaborative data sharing project for reuse of neuroscience data
  - Northwestern University Feinberg School of Medicine (L. Wang)
  - Georgia State University (J. Turner)
  - University of Southern California (J-L. Ambite)
  - UNC Chapel Hill (A. Rajasekar, H. Lander, H. Xu, Y. Wang)
  - Washington University (D. Marcus)
- Based on previous collaborations
  - DataBridge for Neuroscience (L. Wang, Turner, Rajasekar, Lander)
    - NSF EAGER (South BD Spoke IIS 1649397)
  - SchizConnect (L. Wang, Turner, Ambite, Marcus)
    - Midwest Big Data Hub (ACNN Spoke 1636840, 1636846, 1636893, 1636850, and 1550320)
- Funded by the NIH National Institute on Drug Abuse (NIDA)
  Grant Number 1R01DA053028-01 Start date 9/1/2020

Why NeuroBridge?

The NeuroBridge will assemble metadata for a large number of neuroimaging datasets and provide scientists the ability to efficiently and accurately discover appropriate data



- Replication, mega-analysis, meta-analysis are critical in neuroimaging research
- Creating new data is expensive and time consuming; better to reuse existing data
- Different neuroimaging databases have large collections but aren't linked
- Data is often created but only used in papers and never published
- Too much data is as big a problem as too little.

NeuroBridge Architecture

Objective 1: The WebPortal will be the interface for the user to query the system

Objective 2: The Mediator will gather metadata from known neuroimaging databases, while the Miner will use machine learning techniques to extract relevant metadata from journal articles

Objective 3: The NeuroBridge DataBridge processes the system metadata in order to return suggestions of relevant datasets to the user

Objective 4: The initial use case leverages the XNAT platform to acquire metadata and data quality metrics from existing datasets