

## Our Example

Meta-analysis is usually thought of as a *post hoc* procedure, however this is neither required nor always desirable. The procedure used here is a *prospective* meta-analysis; a meta-analysis designed based on data availability and developed before any analysis is done. It is one of the simplest types of analysis that the COINSTAC architecture supports.

### INTRODUCTION

- One of the central problems across the social, behavioral, and brain sciences is the use of small local samples.
- One solution to this problem is to collect samples from multiple international locations (sites) and combine these into a single analysis:
  - This improves power as well as international representation in the analyses
  - A roadblock to this data-sharing procedure is concern for participant privacy

Two solutions have been created for these problems in the neurosciences.

### ENIGMA

The **Enhancing Neuroimaging Genetics through Meta-Analysis** (ENIGMA) project uses a variety of approaches including meta-analyses across international participating sites:

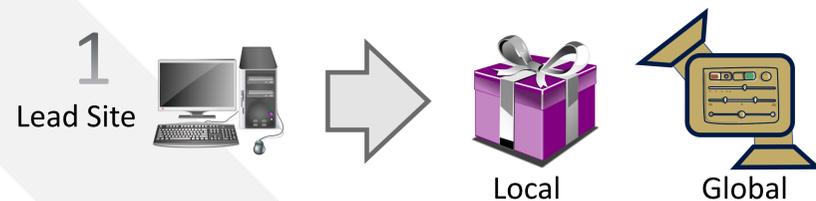
- This meta-analytic approach implements separate but identical analyses on distributed datasets, and then aggregates the results in a preplanned (prospective) meta-analysis
- This achieves some of the power of large samples without requiring data sharing, enhancing research without sacrificing participant privacy
- This approach has proven successful in **many studies** of genetics and neuroimaging
- ENIGMA has moved beyond these simple analysis to more complicated and computationally intensive ones (the approach here supports these analyses, too)

### COINSTAC

The **Collaborative Informatics and Neuroimaging Suite** Toolkit for Anonymous Computation (COINSTAC) project has been developed to:

- Automatically analyze data across multiple sites
- Maintain data privacy and confidentiality
- Transport the derived results to a central location for meta-analysis

Here we consider an example workflow for the COINSTAC system, implementing an ENIGMA-type prospective meta-analysis.



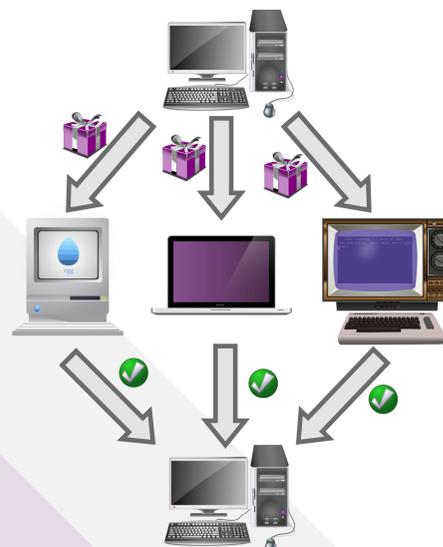
The process begins when the **lead site** defines the local and global analyses.

The **local analysis** is sent out to be run at each site to generate intermediate results that are combined in the **global analysis** that is run at the lead site.

These analyses are defined at the start so that participating sites know what is expected.

The lead site invites other sites to the consortium.

### 2 Participating Sites

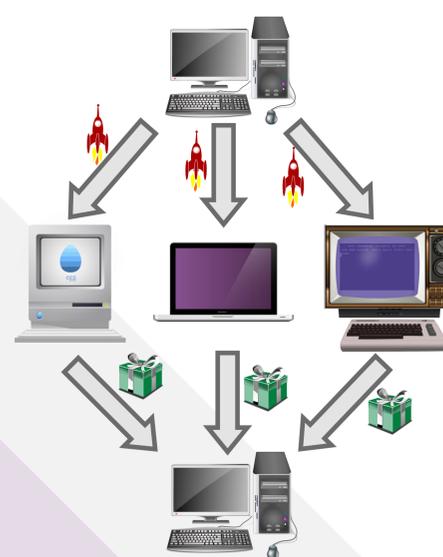


**Participating sites** (collaborating sites with similar data) can see how their data is to be used (**they can see all of the analyses proposed**).

If they agree, they may join the consortium to participate in the meta-analysis, and map their data to the analysis.

You can think of “mapping the data” as pointing the local analysis to the correct columns of a simple spreadsheet, but much more complicated data structures can be accommodated.

### 3



Once the participating sites have joined the consortium and mapped their data, the lead site **launches** all of the local analyses.

The participating sites automatically run the local analysis and **return** the intermediate data to the lead site.

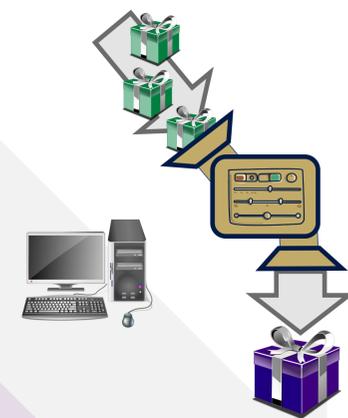
Although not used in this example, COINSTAC implements procedures where this stage of the analysis can be iterative, with the sites repeatedly communicating with the lead site to implement calculations requiring shared information.

This allows computation of much more complex analyses including machine learning techniques.

### NEUROIMAGING EXAMPLE

- We are currently using this framework for a distributed meta-analysis involving 10 sites from the following countries:
  - US
  - Germany
  - Italy, and
  - Australia
- We have neuroimaging data from over 3,000 individuals
- We are examining the relationships between subcortical and cortical brain volumes and negative symptom severity in schizophrenia, using this local/global analysis in a meta-analysis framework
- Future analyses will include decentralized machine learning approaches for classification and prediction of symptom severity based on brain structural and functional patterns

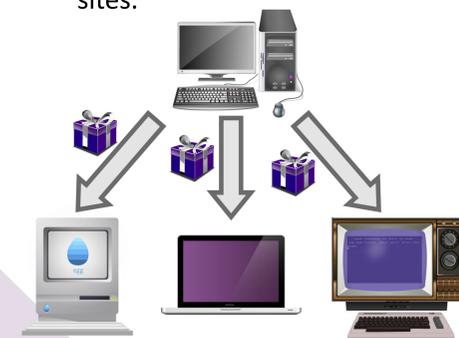
### 4



The lead site **combines** the intermediate data from the local sites into the global analysis, generating the meta-analytic results.

If agreed upon, these results are returned to all the participating sites.

### 5



For more information, contact: [mturner46@gsu.edu](mailto:mturner46@gsu.edu)

<sup>1</sup>Matthew D. Turner, <sup>1</sup>Harshadvan Gazula, <sup>2</sup>Peter Kochunov, <sup>3</sup>Paul Thompson,

<sup>3</sup>Neda Jahanshad, <sup>3</sup>Chris Ching, <sup>3</sup>Fabrizio Pizzagalli, <sup>4</sup>Greg Strauss, <sup>5</sup>Anthony Ahmed, ENIGMA SZ Working Group, <sup>1</sup>Vince D. Calhoun, <sup>6</sup>Theo van Erp, <sup>1</sup>Jessica A. Turner

<sup>1</sup>Georgia State University; <sup>2</sup>University of Maryland; <sup>3</sup>University of Southern California; <sup>4</sup>University of Georgia; <sup>5</sup>Cornell University; <sup>6</sup>University of California, Irvine

# ENIGMA COINSTAC: Increasing neuroimaging data diversity with managed privacy