

A database for the ENIGMA working group – Organization of Imaging Genomics in Infancy (ORIGINs)

Ann Alex

Institute for Quantitative Health Science and Engineering

ORIGINs

Growing network of researchers actively engaged in neuroimaging research in infancy and early childhood

Overarching goal:

Determine how genetic and environmental factors influence the development of brain morphometry, anatomical and functional connectivity, and cognitive and emotional function from birth to age 6

Currently: 20 sites across 6 countries

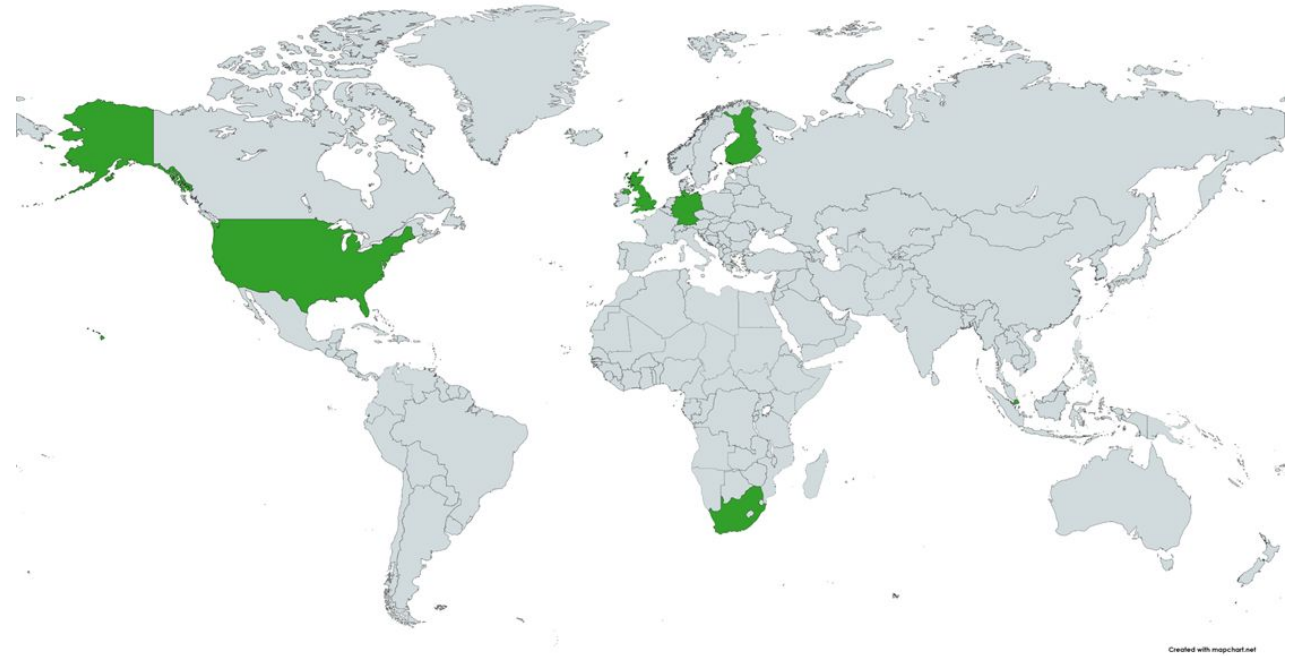
Contribute:

Neuroimaging data

Demographic data

Genetic data

Behavioral/Cognitive data



Aim

Create a database for storing ORIGINs data – understand the complexities associate with data management in cloud-based system

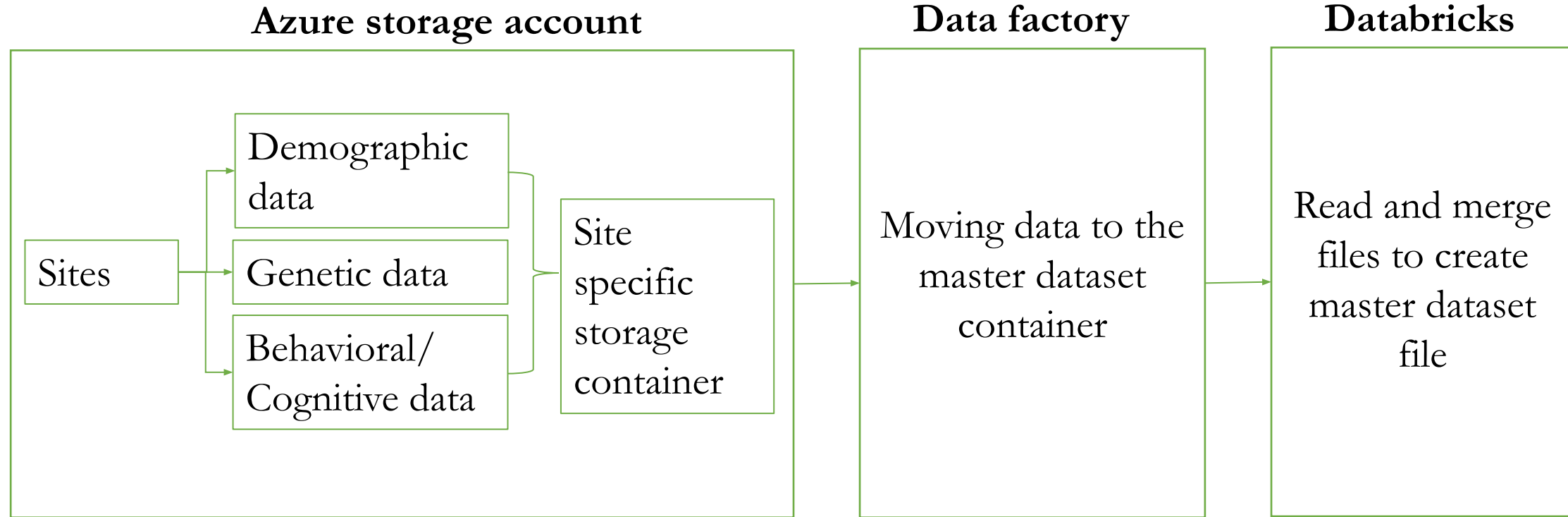
- Site specific access to folders for authorized personnel
- Compiling dataset

Need for cloud-based resources

Efficient storage and access to the data for the collaborating institutions

Developing data repository for serving as a platform for future collaborations and sharing of data with other researchers.

Cloud Resources Employed



- Success and ○ Challenges

Azure Storage Account

Organization of data:

- Site specific storage containers
- Directories/sub folders for storing different types of data
- Restricted access to containers by use of SAS keys – restrict time, IP address

Azure Data Factory

Moving of data:

- Copying different types of files from site specific containers to master data container
- Organization of data in master data set - Moving subfolders instead of individual files
- Triggering copy action when a file is added to any of the containers

Azure Databricks

Compiling and analysis of data:

- Creating cluster and notebook for pipeline
 - Connection to storage container, reading files and merging using SQL and python
-
- Lack of familiarity with SQL and python for apt execution
 - Problem in reading files when switching to R in the notebook

Future Work

Create a database which will enable fast query and extraction of data

And from this experience, for future.....

Better equipped to include a component for cloud-based storage and sharing data in future projects

Use of other cloud-based resources for analysis of big data that are recommended by other data repositories

Acknowledgements

MSU ICER Cloud Computing Fellowship
Mentors: Patrick Bills and Mahmoud Parvizi
Cloud Computing Fellows



Thank you....