

# **Supporting the Consortium for Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) through the Organic Data Science Framework**

**Neda Jahanshad<sup>1</sup>, Sarah Madsen<sup>1</sup>, Yervand Azatian<sup>1</sup>, Derrek P. Hibar<sup>1</sup>, Paul M. Thompson<sup>1</sup>,  
and Yolanda Gil<sup>2</sup>**

<sup>1</sup> Imaging Genetics Center  
University of Southern California

<sup>2</sup> Information Sciences Institute and Department of Computer Science  
University of Southern California

Contact: [gil@isi.edu](mailto:gil@isi.edu)

## **Introduction**

The Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium is a growing international team of over 300 scientists, spanning over 30 countries. It was formed in 2009 with 3 research groups around the world pooling together their resources in order to discover genetic influences on brain structure. Taking a meta-analytical approach to neuroimaging, a unique direction in the field, meant that research groups would perform harmonized protocols and statistical analyses and only share summary results of statistical tests and not raw data itself. Welcoming all interested imaging groups to join, ENIGMA has now expanded rapidly worldwide, and the opportunity to participate in high-powered collaborative work unbounded by data privacy and exchange policies has become scientifically rewarding with major, reliable, neuroscience and genetic discoveries being made and published in the highest impact journals [Hibar et al., Nature, 2015; Stein et al., Nature Genetics 2012]. All groups within the consortium have a unique set of features yet are bounded by a common few. The Consortium now is involved in many projects beyond genetics, including studies of over 10 diseases with numerous sub-aims. Members with the capacity and desire to lead a collaborative project within the ENIGMA framework, may propose an sub-analysis based on the common features and involve groups with relevant data available (scan modality, patient population, genetic information etc). Projects will start and each participating group has a set of necessary tasks; it will then complete or move towards a follow up project, hopefully after successful publication. The continual formation of these sub-communities, projects, and their termination, enforces the need for an organizational framework where projects can be maintained and the participation of groups as well as member accurately logged. The Organic Data Science (ODS) Framework offers this collaborative environment, using a task-centered approach and drawing design principles from social studies of online communities.

## **Methods and Strategies**

Team science experts have developed ODS to provide a framework where information can be maintained and organized at the individual, group and project level. Within a group, different members contribute to different projects. Informing all members of all starting projects and all updates within the Consortium would be overwhelming due to the specific interests of individual

members. On the other hand, if participants are not knowledgeable of proposed projects that may pique their interest there would be missed opportunities for their contributions. ODS provides means of dynamically creating special interest areas based on properties of the tasks and the participants, and monitor the progress of members in the group by the tasks completed. The many projects and hence tasks required of the groups can become burdensome, and continually contacting busy group leaders regarding individual projects can act as additional chores, thus we have opted for a more centralized organizational core. ODS provides a user-friendly interface for organizing and monitoring all the collaborative activities in ENIGMA. Originally developed for an environmental science project, ODS extends a semantic wiki platform to support the description of users, institutions, projects, tasks, datasets as separate entities with semantic properties that can support structured queries. For example, a property of a project to study Alzheimers in young patients describes what imaging and genomic data are required in order for an institution to participate. This enables the creation of special pages that dynamically aggregate projects, institutions, and participants with different criteria that allow users to easily find ongoing activities of interest.

### **Findings and Conclusions**

ENIGMA has brought together top researchers in the biomedical field and therefore an ever-growing set of data and ideas. First formed with one goal in mind, initial maintenance of the network was possible in a few spreadsheets. ENIGMA grew exponentially out of pure scientific interest and has branched into dozens of projects, some completely independent of others, and others with substantial overlap. Due to the unexpected growth and the unforeseen projects initiated, the consortium needed collaboration support. We have set up the OSD framework and used it to describe the participating groups and ongoing tasks so that the tasks and groups are easy to monitor and accessible to newcomers. One lesson learned in the process is that the representation of tasks and groups must highlight commonalities between the groups that may be the seeds for future collaborative projects.

### **References**

- [Hibar et al 2015] “Common genetic variants influence human subcortical brain structures.” Derrek P Hibar, Jason L Stein, Miguel E Renteria, Alejandro Arias-Vasquez, Sylvane Desrivieres, Neda Jahanshad, et al. *Nature* 20(7546), 2015. doi: 10.1038/nature14101
- [Stein et al 2012] “Identification of common variants associated with human hippocampal and intracranial volumes.” Jason L Stein, Sarah E Medland, Alejandro Arias Vasquez, Derrek P Hibar, et al. *Nature Genetics* 44, 552–561, 2012. doi:10.1038/ng.2250Stein et al., *Nature Genetics* 2012