Towards Automatic Generation of Portions of Scientific Papers for Large Multi-Institutional Collaborations Based on Semantic Metadata

MiHyun Jang¹, Tejal Patted², Yolanda Gil², Daniel Garijo², Varun Ratnakar², Jie Ji², Prince Wang¹, Aggie McMahon³, Paul M. Thompson³, and Neda Jahanshad³

¹ Troy High School,
² Information Sciences Institute, University of Southern California,
³ Imaging Genetics Center, University of Southern California

@yolandagil, @dgarijov

{gil,dgarijo}@isi.edu
Increasing Complexity of Scientific Collaborations


LHC Atlas: 4,000 authors
Massive Multi-Institutional Self-Organizing Collaborations: Neuroimaging Genomics in ENIGMA

PIs formulate joint studies with brain data collected for large populations (cohorts) for specific purposes.
Growth of the ENIGMA Collaboration: Working Groups

2012

- GWAS
  - Hippocampal Volume
  - Intracranial Volume
- Schizophrenia
  - Subcortical Volume
  - Case/Control
- Diffusion Imaging
  - Protocol development
  - Reliability
  - Heritability

2014

- Genetics
  - CNVs
- Imaging
  - Whole genome
  - Subcortical Volume
  - Cortical thickness
  - Diffusion imaging
  - Connectomics
- Computational
  - Machine learning?
  - Voxelwise GWAS
  - 22q deletion
- Diseases
  - Addictions
  - ADHD
  - Autism
  - Bipolar
  - Depression
  - HIV
  - OCD
  - Schizophrenia

Case/control

Complexity of the ENIGMA Collaboration:
Projects of the Schizophrenia Working Group

- Projects focus on a specific goal and analysis, publish a final joint paper
- Analysis is done collaborative with many researchers contributing differently
- Projects are often joint with other working groups

- Subcortical Shape (Wang, Gutman et al. NU, USC)
- Cortical Thickness/Surface (Turner/Van Erp et al., GSU, UCI)
- Negative / Positive Symptoms (Walton et al., Germany)
- Normal Variation with Aging (Dima/Frangou et al., Great Britain)
- Vertexwise Thickness/Surface (van Erp/Turner et al., GSU, UCI)
- Hippocampal Subfields (van Erp/Turner et al., GSU, UCI)
- First-order Relatives (van Haren et al., the Netherlands)
- First-Episode, Longitudinal (Roiz-Santiañez et al., Spain)
- Cannabis (Koenders et al., AMC)
- Diffusion Tensor Imaging (Kelly et al., USC)
- Connectomics (Kelly et al, USC)
- Deficit Schizophrenia and DTI (de Rossi/Spalletta et al., Rome)
- Aggression (Nickl-Jockschat/Gur et al., Germany/USA)
- Early Onset Psychosis (Agartz/Gurholt/Raballo et al., Norway)
- Sulci (Jahanshad/Pizzagalli et al., USC)
- Laterality (Tuulio/Clyde/van Erp/Hashimoto/Gur et al.)
- Motion (van Erp et al.)
- Cross Disorder (SZ/BD/MDD)
- Genetics (many-PIs)

Challenges in Managing Information in ENIGMA

1. Working Group Leader
   • Tracking projects, datasets available

2. Project Leaders
   • Tracking tasks, contributors, datasets, progress

3. Cohort PI
   • Tracking all tasks, delegating, awareness of new projects

4. Managing overall collaboration
   • Who has data on adolescents across all disease groups?
   • What project(s) is a site involved in?
   • What diseases are we studying?
   • Did we already have a group to study cerebral ataxia?
Approach: Organic Data Science Framework Provides Semantic Repository for ENIGMA

A Controlled Crowdsourcing Approach to Scientific Ontology Development and Data Annotation. ISWC-17 presentation will describe approach with application to climate collaboration

Dynamically generated content based on queries

Tracking contributions

Crowdsourced data and metadata annotation

Dynamic visualizations of wiki contents
ENIGMA Data Model

- Datasets are collected by a funded project
  - Follows a very precise acquisition procedure (*protocol*)
    - What brain scanner, how it was set up, flip angle, voxel size, etc.
- Participants in a study are selected based on phenotype
  - *Inclusion criteria* (e.g., ADHD, aged 12-24)
  - *Exclusion criteria* (e.g., no smokers)
Current Contents

Total: 400 pages
• 3 projects
• 89 cohort groups
• 54 cohorts
• 4 acquisition protocols
• 8 scanner types
• 112 persons

• Ongoing work:
  • Reorganizing ontology
  • Populating site

Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex

Tulio Guadalupe1,2, Samuel F. Mathias3, and Thea G. M. van Erp4

1 Department of Neurology, Southern Connecticut State Medical School, New Haven, CT 06511, USA
2 Department of Psychiatry and Psychotherapy, University of Ghent, Belgium
3 Department of Psychology, University of Bonn, Germany
4 Department of Clinical Neurology, Universiteit van Amsterdam, Amsterdam, Netherlands

Abstract: The two hemispheres of the human brain differ functionally and structurally. Despite over a century of research, the extent to which brain asymmetry is influenced by sex, handedness, age, and genetic factors is still controversial. Here we present the largest ever analysis of subcortical brain asymmetries in a humanized multi-site study using meta-analysis methods. Structural symmetry of seven subcortical structures was assessed in 15,847 MRI scans from 72 Americas.

Keywords: brain asymmetry, handedness, age, genetics, subcortical structures

Electronic supplementary material: The online version of this article (doi:10.1007/s11682-016-9629-z) contains supplementary material, which is available to authorized users.

Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex

Tulio Guadalupe1,2, Samuel F. Mathias3, and Thea G. M. van Erp4

1 Department of Neurology, Southern Connecticut State Medical School, New Haven, CT 06511, USA
2 Department of Psychiatry and Psychotherapy, University of Ghent, Belgium
3 Department of Psychology, University of Bonn, Germany
4 Department of Clinical Neurology, Universiteit van Amsterdam, Amsterdam, Netherlands

Abstract: The two hemispheres of the human brain differ functionally and structurally. Despite over a century of research, the extent to which brain asymmetry is influenced by sex, handedness, age, and genetic factors is still controversial. Here we present the largest ever analysis of subcortical brain asymmetries in a humanized multi-site study using meta-analysis methods. Structural symmetry of seven subcortical structures was assessed in 15,847 MRI scans from 72 Americas.

Keywords: brain asymmetry, handedness, age, genetics, subcortical structures

Electronic supplementary material: The online version of this article (doi:10.1007/s11682-016-9629-z) contains supplementary material, which is available to authorized users.

Contributions:

TG and SRM designed the project. TGMV, CDW, and MPZ contributed cohorts. ...

The image contains a page from a scientific paper discussing human subcortical brain asymmetries and their effects of age and sex. The authors present the largest ever analysis of subcortical brain asymmetries in 15,847 MRI scans from 72 Americans. The paper highlights the importance of considering sex, handedness, age, and genetic factors in understanding brain asymmetry.
How ENIGMA Information is Used in Papers: (II) Supplementary Information

- Tables to describe cohorts
  - Demographics
  - Inclusion/exclusion criteria
  - Acquisition protocols

<table>
<thead>
<tr>
<th>Site</th>
<th>Sequence</th>
<th>Field Strength</th>
<th>Acquisition Direction</th>
<th># of Slices</th>
<th>Slice Gap</th>
<th>Voxel Size (mm³)</th>
<th>TI</th>
<th>TE</th>
<th>TR</th>
<th>Flip Angle</th>
<th>Citation</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amsterdam</td>
<td>3D T1-weighted turbo field echo (TFE)</td>
<td>3T scanner Philips Gyroscan Intera</td>
<td>coronal</td>
<td>182</td>
<td>0mm</td>
<td>1x1x1.2</td>
<td>0ms</td>
<td>4.6ms</td>
<td>9.621ms</td>
<td>8</td>
<td>1, 2</td>
<td>FreeSurfer (5.0)</td>
</tr>
<tr>
<td>Barcelona (Site 1: FIDMAG)</td>
<td>3D T1-weighted enhanced fast gradient echo (EFGRE3D)</td>
<td>1.5T GE Signa</td>
<td>Axial</td>
<td>180</td>
<td>0mm</td>
<td>0.47x0.47x1</td>
<td>710ms</td>
<td>3.93ms</td>
<td>2000ms</td>
<td>15</td>
<td>3, 4</td>
<td>FreeSurfer (5.3.0)</td>
</tr>
</tbody>
</table>
Using ENIGMA Metadata: (II) Automated Generation of Tables for Papers


Generated image acquisition protocol table:

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Data Type</th>
<th>Scanner</th>
<th>Acquisition Direction</th>
<th>Sequence</th>
<th>Data Acquisition Matrix</th>
<th>Flip Angle</th>
<th>Number of Slices</th>
<th>Scan Time</th>
<th>TE</th>
<th>TI</th>
<th>TR</th>
<th>Voxel Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLING</td>
<td>T1-weighted MRI</td>
<td>3T Magnetom TIM Trio</td>
<td>Sagittal</td>
<td>MPRAGE sequence</td>
<td>256 x 256</td>
<td>9</td>
<td>192</td>
<td>8 min 26 sec</td>
<td>3.26 ms</td>
<td>900 ms</td>
<td>2250 ms</td>
<td>1 mm³</td>
</tr>
<tr>
<td>HMS</td>
<td>T1-weighted MRI</td>
<td>1.5T Magnetom Sonata</td>
<td>Sagittal</td>
<td>MPRAGE sequence</td>
<td>256 x 256</td>
<td>15</td>
<td>176</td>
<td>5 min</td>
<td>4.0 ms</td>
<td>700 ms</td>
<td>1900 ms</td>
<td>1 mm³</td>
</tr>
</tbody>
</table>

Generated demographics table:

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Total</th>
<th>Control Total</th>
<th>Patient Total</th>
<th>Male Patients</th>
<th>Female Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLING</td>
<td>372</td>
<td>323</td>
<td>49</td>
<td>36</td>
<td>13</td>
</tr>
<tr>
<td>HMS</td>
<td>101</td>
<td>55</td>
<td>46</td>
<td>32</td>
<td>14</td>
</tr>
</tbody>
</table>
Conclusions and Future Work

• Problem: capture information about multi-institutional collaborations
  • Working groups and projects
  • Datasets
    • Acquisition protocols
    • Inclusion/exclusion criteria
  • People participation in projects and dataset collection

• Approach: Semantic repository using Organic Data Science framework
  • Core ontology reflects main information to be captured
  • Crowd extensions to account for new properties for specific projects
  • See talk in ISWC in-use track!

• Semantic repository used to generate portions of multi-institutional publications
  • Author lists and acknowledgements

• Ongoing work:
  • Populate repository from current idiosyncratic spreadsheets kept by projects
  • Evaluate use of system for generating portions of future publications
Towards Automatic Generation of Portions of Scientific Papers for Large Multi-Institutional Collaborations Based on Semantic Metadata

MiHyun Jang\textsuperscript{1}, Tejal Patted\textsuperscript{2}, Yolanda Gil\textsuperscript{2}, Daniel Garijo\textsuperscript{2}, Varun Ratnakar\textsuperscript{2}, Jie Ji\textsuperscript{2}, Prince Wang\textsuperscript{1}, Aggie McMahon\textsuperscript{3}, Paul M. Thompson\textsuperscript{3}, and Neda Jahanshad\textsuperscript{3}

\textsuperscript{1}Troy High School, \\
\textsuperscript{2}Information Sciences Institute, University of Southern California, \\
\textsuperscript{3}Imaging Genetics Center, University of Southern California

@yolandagil, @dgarijov

\{gil,dgarijo\}@isi.edu

First Workshop on Enabling Open Semantic Science (SemSci 2017)