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

MiHyun Jang[1], Tejal Patted[2], **Yolanda Gil**[2], **Daniel Garijo**[2], Varun Ratnakar[2], Jie Ji[2], Prince Wang[1], Aggie McMahon[3], Paul M. Thompson[3], and Neda Jahanshad[3]

[1] Troy High School,

[2] Information Sciences Institute, University of Southern California,

[3] Imaging Genetics Center, University of Southern California

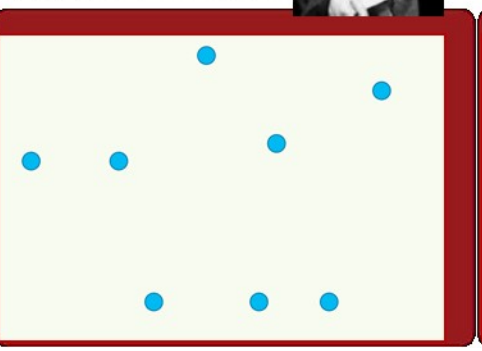
@yolandagil, @dgarijov

{gil,dgarijo}@isi.edu

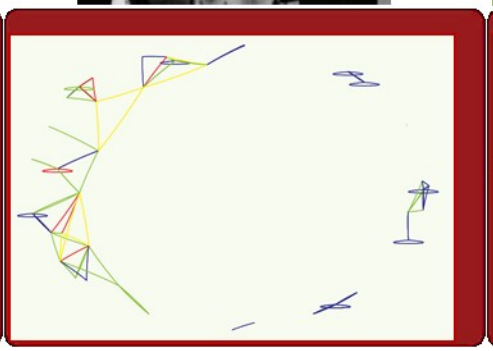
First Workshop on Enabling Open Semantic Science (SemSci 2017)

Increasing Complexity of Scientific Collaborations

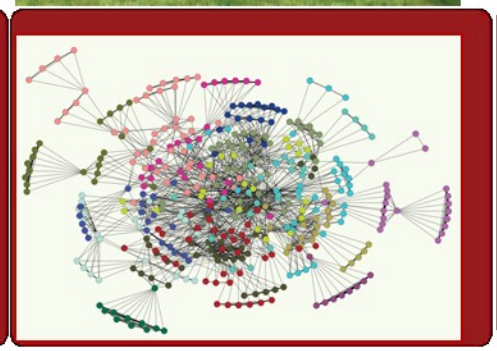
single-authorship



co-authorship



large number of co-authors



community as author

LHC Atlas: 4,000 authors

Evolution of the scientific enterprise from [Barabasi, Science 2005] extended with the ATLAS Detector Project at the Large Hadron Collider [The ATLAS Collaboration, Science 2012].

Massive Multi-Institutional Self-Organizing Collaborations: Neuroimaging Genomics in ENIGMA

PNAS

nature COMMUNICATIONS

OPEN

Molecular Psychiatry (2015), 1–7
© 2015 Macmillan Publishers Limited All rights reserved 1359-4184/15
www.nature.com/mp

ORIGINAL ARTICLE

Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA conso

TGM van Erp[1,39], DP Hibar[2,39], JM Ra
AM Dale[9,10], I Melle[5], CB Hartberg[5,6],
DM Cannon[15], A Corvin[14], MWJ Mac
RE Gur[16], SG Potkin[1], DH Mathalon[1
HJ Bockholt[25,27,28], SR Sponheim[29],
B Crespo-Facorro[32,33], L Wang[34,35],
R Hashimoto[37], PM Thompson[2] and

Molecular Psychiatry

Molecular Psychiatry (20
© 2016 Macmillan Publish
www.nature.com/mp

ORIGINAL ARTICLE

Subcortical volumetric abnormalities in b

P Hibar[1], LT Westlye[2,3], TGM van Erp[4], J Rasmussen[4], CD Leonardo[1], J Faskowitz[1], UK Ha
M Dale[6,7], O Gruber[8,9], B Krämer[8], S Trost[8], B Liberg[10], C Abé[11], CJ Ekman[10], M Ingvar[11,12],
E Bearden[17,18,19], the Costa Rica/Colombia Consortium for Genetic Investigation of Bipolar Endophenotypes, E Sprooten[20,21],
C Glahn[20,21], GD Pearlson
Caseras[26], NS Lawrence[27]
Henry[32,33], UF Malt[34,35,36],
Bilderbeck[41,43], L Abramc
T Hajek[46,47], B Mwangi[48], J
OA And

nature

nature neuroscience

OPEN

Molecular Psychiatry (2015), 1–7
© 2015 Macmillan Publishers Limited All rights reserved 1359-4184/15
www.nature.com/mp

ORIGINAL ARTICLE

Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group

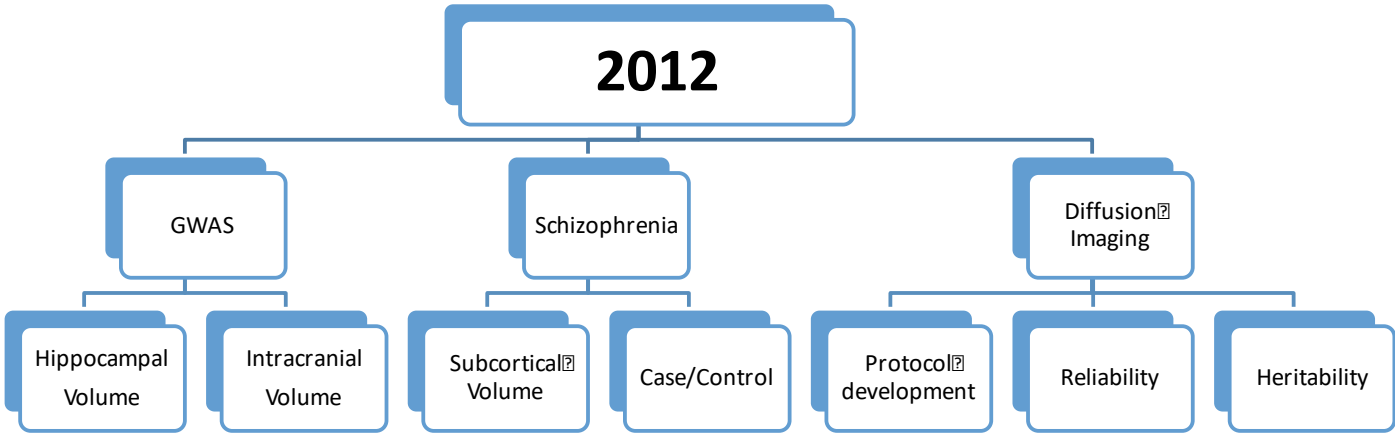
L Schmaal[1], DJ Veltman[1], TGM van Erp[2], PG Sämann[3], T Frodl[4,5], N Jahanshad[6], E Loehrer[7],
MW Vernooij[7,9], MA Ikram[7,9,11], K Wittfeld[12], HJ Grabe[12,13,14], A Block[13], K Hegenscheid[15],
J Lagopoulos[17], SN Hatton[17], IB Hickie[17], B Goya-Maldonado[18], B Krämer[18], O Gruber[18],
LT Strike[19,20,21], NT Mills[
Medland[24], NG Martin[
GM MacQueen[27], EM Fre
Tol[29], NJ van der Wee
E Schramm[34], C Norman
T Nickson[38], AM Mcl
JE Sussmann[38], BR Godl
41,43], BWJH Penninx[1],
ENIGMA-Major Depressiv

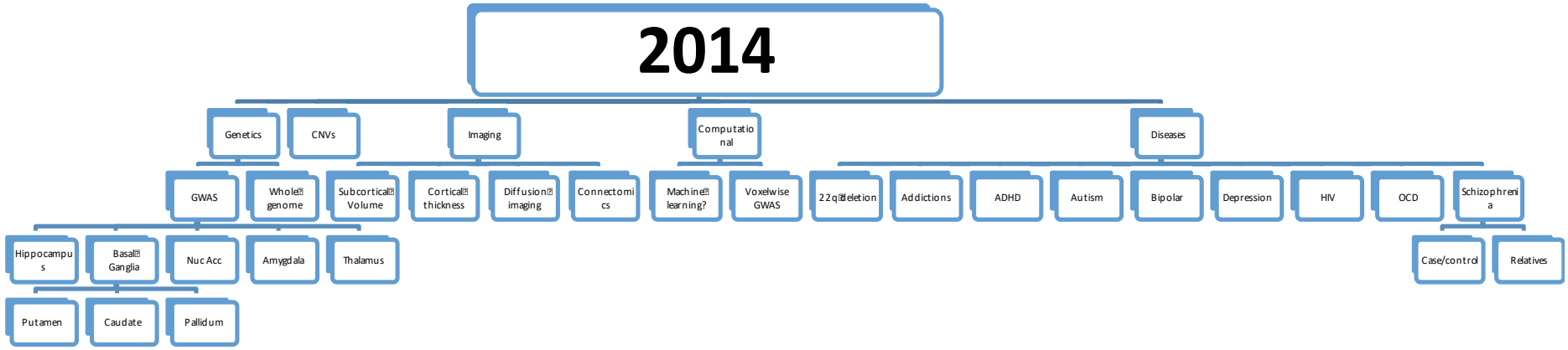
nature genetics

The American Journal of Psychiatry

PIs formulate joint studies with brain data collected for large populations (*cohorts*) for specific purposes

Growth of the ENIGMA Collaboration: Working Groups

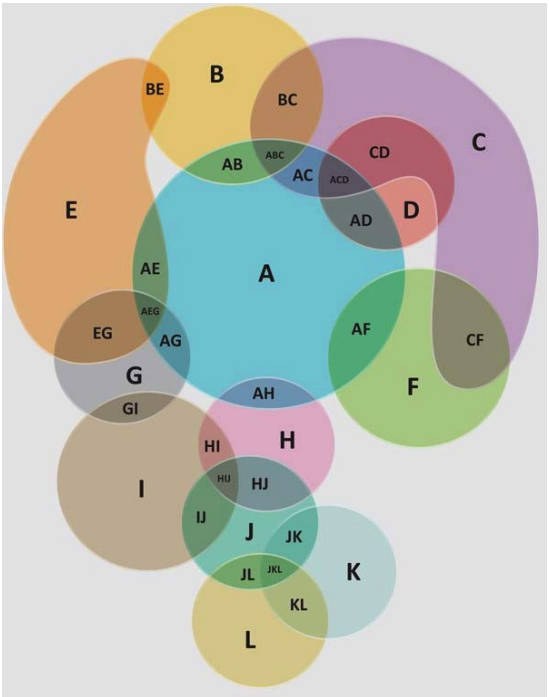




Complexity of the ENIGMA Collaboration: Projects of the Schizophrenia Working Group

- 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 Volume (van Erp/Turner et al., UCI, Mol. Psych. (2015)
- 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

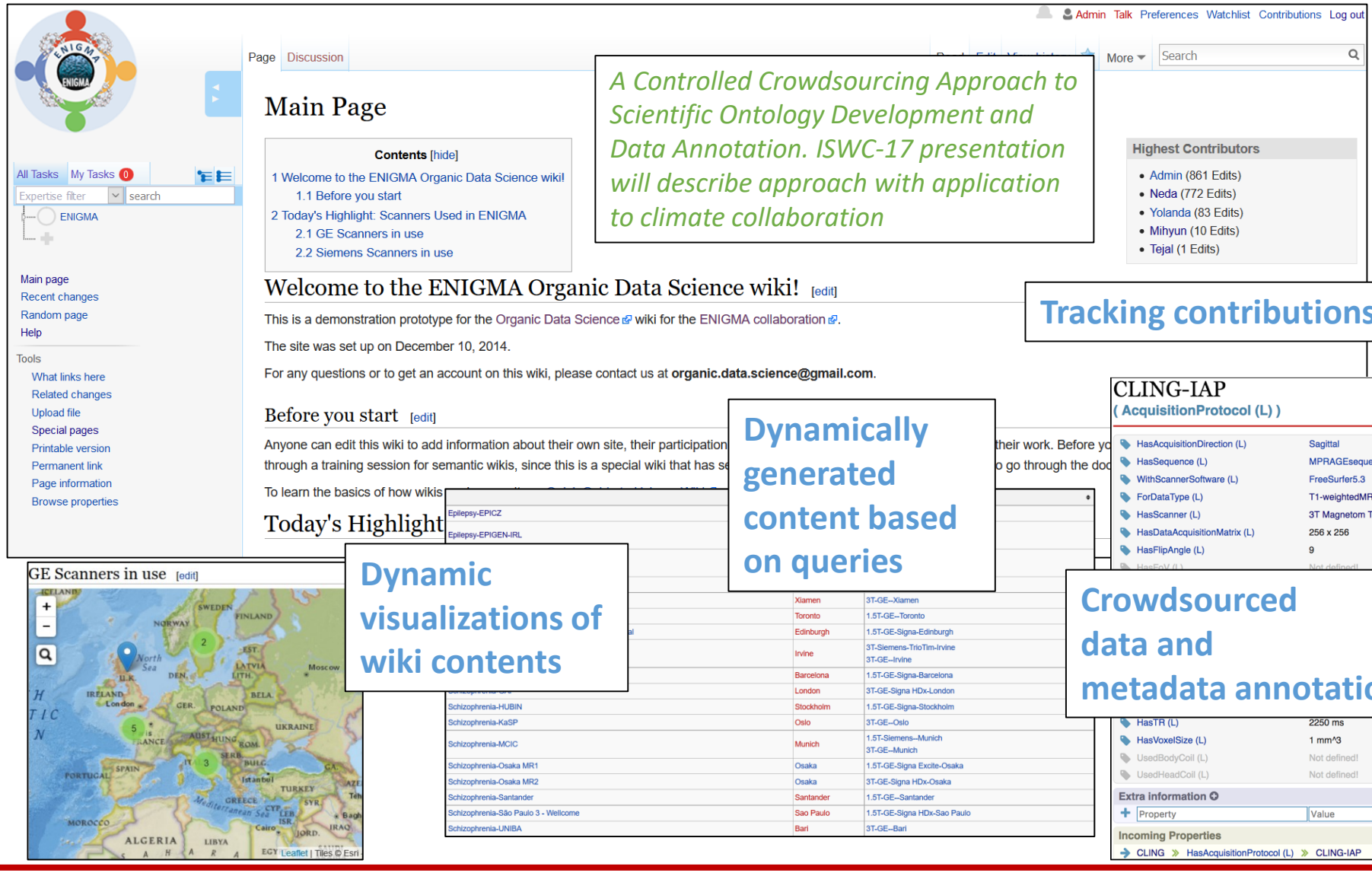
A Controlled Crowdsourcing Approach to Scientific Ontology Development and Data Annotation. ISWC-17 presentation will describe approach with application to climate collaboration

Tracking contributions

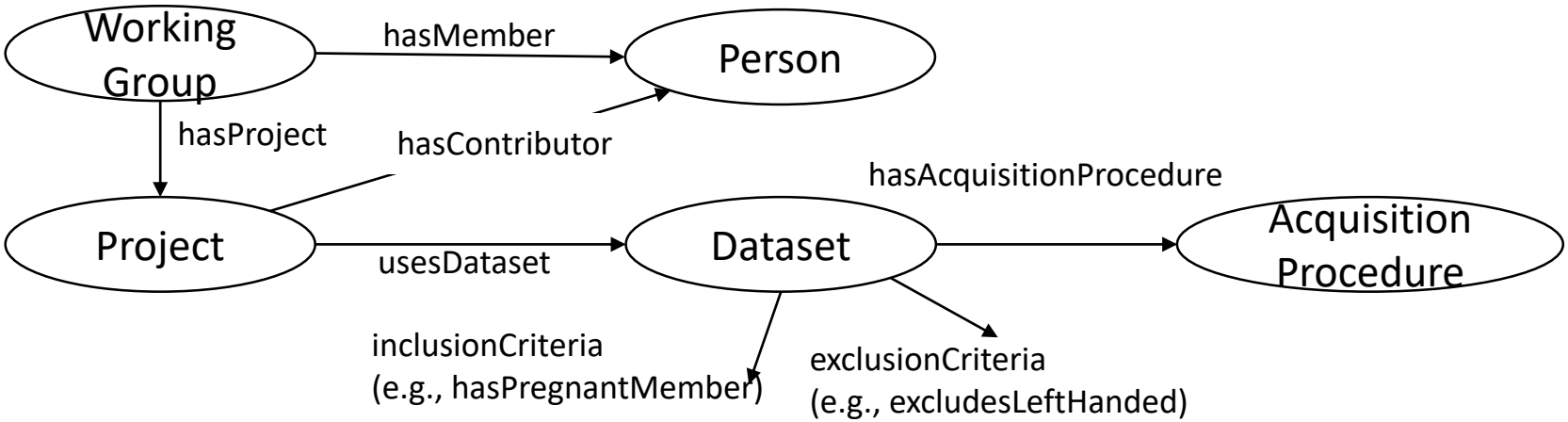
Dynamically generated content based on queries

Dynamic visualizations of wiki contents

Crowdsourced data and metadata annotation



ENIGMA Data Model



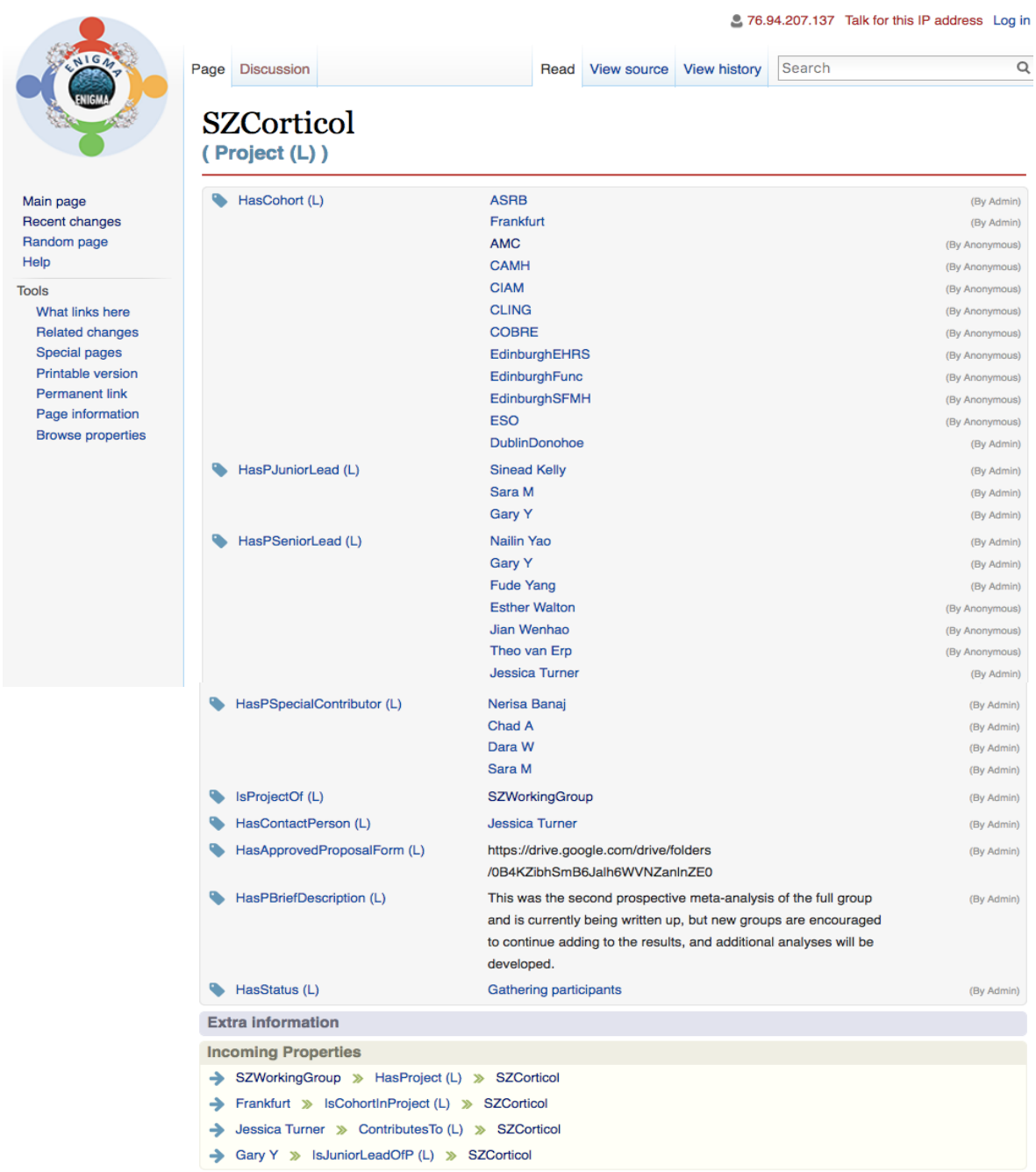
- Datasets are collected by a funded project
  - Follows a very precise acquisition procedure (*protocol*)
    - What brain scanner, how it was set up, flip angle, voxel size, etc.
- Participants in a study are selected based on phenotype
  - *Inclusion criteria* (e.g., ADHD, aged 12-24)
  - *Exclusion 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



How ENIGMA Information is Used in Papers: (I) Author List and Contributions

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

Tulio Guadalupe[1,2] &Samuel R. Mathias[3] &Theo G. M. vanErp[4] &
Christopher D. Whelan[5,6] &Marcel P. Zwiers[7] &Yoshinari Abe[8] &Lucija Abramovic[9] &
Ingrid Agartz[10,11,12] &Ole A. Andreassen[10,13] &Alejandro Arias-Vásquez[14,15,16] &
Benjamin S. Aribisala[17,18] &Nicola J. Armstrong[19,20] &Volker Arolt[21] &Eric Artiges[22] &
Rosa Ayesa-Arriola[23,24] &Vatche G. Baboyan[25] &Tobias Banaschewski[26] &
Gareth Barker[27] &Mark E. Bastin[18,28,29,30] &Bernhard T. Baune[31] &John Blangero[32,33] &
Arun L.W. Bokde[34] &Premika S.W. Boedhoe[35,36,37] &Anushree Bose[38] &Silvia Brem[39,40] &
Henry Brodaty[41] &Uli Bromberg[42] &Samantha Brooks[43] &Christian Büchel[42] &
Jan Buitelaar[16,44,45] &Vince D. Calhoun[46,47] &Dara M. Cannon[48] &Anna Cattrell[49] &
Yuqi Cheng[50] &Patricia J. Conrod[51,52] &Annette Conzelmann[53,54] &Aiden Corvin[55] &
Benedicto Crespo-Facorro[23,24] &Fabrice Crivello[56] &Udo Dannlowski[57,58] &
Greig I. de Zubicaray[59] &Sonja M.C. de Zwarte[9] &Ian J. Deary[28] &Sylvane Desrivières[49] &
Nhat Trung Doan[10,13] &Gary Donohoe[60,61] &Erlend S. Dørum[13,62,63] &
Stefan Ehrlich[64,65,66] &Thomas Espeseth[13,67] &Guillén Fernández[16,44] &Herta Flor[68] &
Jean-Paul Fouche[69] &Vincent Frouin[70] &Masaki Fukunaga[71] &Jürgen Gallinat[72] &
Hugh Garavan[73] &Michael Gill[55,74] &Andrea Gonzalez Suarez[75,76] &Penny Gowland[77] &
Hans J. Grabe[78,79] &Dominik Grotegerd[80] &Oliver Gruber[81] &Saskia Hagenaars[82] &
Ryota Hashimoto[83,84] &Tobias U. Hauser[85,86,87] &Andreas Heinz[88] &Derrek P. Hibar[5] &
Pieter J. Hoekstra[89] &Martine Hoogman[14] &Fleur M. Howells[43] &Hao Hu[90] &
Hilleke E. Hulshoff Pol[9] &Chaim Huyser[91,92] &Bernd Ittermann[93] &Neda Jahanshad[25] &
Erik G. Jönsson[12,94] &Sarah Jurk[95] &Rene S. Kahn[9] &Sinead Kelly[96] &
Bernd Kraemer[81] &Harald Kugel[97] &Jun Soo Kwon[98,99,100] &Herve Lemaitre[22] &
Klaus-Peter Lesch[101,102] &Christine Lochner[103] &Michelle Luciano[28] &
Andre F. Marquand[7,104] &Nicholas G. Martin[105] &Ignacio Martínez-Zalacaín[106] &

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.

* Clyde Francks
clyde.francks@mpi.nl

1 Language & Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

2 International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

3 Department of Psychiatry, Yale School of Medicine, New Haven, CT 06519, USA

4 Department of Psychiatry and Human Behavior, University of California, Irvine, CA, USA

5 Imaging Genetics Center, Institute for Neuroimaging & Informatics, Keck School of Medicine of the University of Southern California, Marina del Rey, CA, USA

6 Molecular and Cellular Therapeutics, The Royal College of Surgeons, Dublin 2, Ireland

7 Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

8 Department of Psychiatry, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan

9 Brain Centre Rudolf Magnus, University Medical Centre Utrecht, Utrecht, The Netherlands

10 NORMENT - KG Jebsen Centre, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

11 Department of Research and Development, Diakonhjemmet Hospital, Oslo, Norway

12 Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, Stockholm, Sweden

13 NORMENT - KG Jebsen Centre, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

Contributions:

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

Jean-Luc Martinot[107,108] · David Mataix-Cols[109] · Karen Mather[19] · Colm McDonald[48] · Katie L. McMahon[110] · Sarah E. Medland[105] · José M. Menchón[106,111,112] · Derek W. Morris[60] · Omar Mothersill[60,55] · Susana Munoz Maniega[82,18,29,30] · Benson Mwangi[113] · Takashi Nakamae[8,114] · Tomohiro Nakao[115] · Janardhanan C. Narayanaswaamy[38] · Frauke Nees[68] · Jan E. Nordvik[62] · A. Marten H. Onnink[14] · Nils Opel[21] · Roel Ophoff[116,9] · Marie-Laure Paillère Martinot[22,117] · Dimitri Papadopoulos Orfanos[70] · Paul Pauli[118] · Tomáš Paus[119] · Luise Poustka[26,120] · Janardhan Y.C. Reddy[38] · Miguel E. Renteria[105] · Roberto Roiz-Santiáñez[23,24] · Annerine Roos[103] · Natalie A. Royle[82,18,29,30] · Perminder Sachdev[19] · Pascual Sánchez-Juan[75,76] · Lianne Schmaal[121] · Gunter Schumann[49] · Elena Shumskaya[14,7] · Michael N. Smolka[95] · Jair C. Soares[122] · Carles Soriano-Mas[106,111,123] · Dan J. Stein[124] · Lachlan T. Strike[125] · Roberto Toro[126] · Jessica A. Turner[127,128,47] · Nathalie Tzourio-Mazoyer[56] · Anne Uhlmann[129] · Maria Valdés Hernández[82,18,29,30] · Odile A. van den Heuvel[130,36,37] · Dennis van der Meer[89] · Neeltje E.M. van Haren[9] · Dick J. Veltman[121] · Ganesan Venkatasubramanian[38] · Nora C. Vetter[95] · Daniella Vuletic[43] · Susanne Walitza[85,131,132] · Henrik Walter[68] · Esther Walton[64,127] · Zhen Wang[90] · Joanna Wardlaw[82,18,29,30] · Wei Wen[19] · Lars T. Westlye[13,63] · Robert Whelan[133] · Katharina Wittfeld[134] · Thomas Wolfers[14,44] · Margaret J. Wright[135,105] · Jian Xu[50] · Xiufeng Xu[50] · Je-Yeon Yun[136] · JingJing Zhao[137,140] · Barbara Franke[14,15] · Paul M. Thompson[5] · David C. Glahn[138,139] · Bernard Mazoyer[56] · Simon E. Fisher[1,44] · Clyde Francks[1,44]

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 harmonized multi-site study using meta-analysis methods. Volumetric asymmetry of seven subcortical structures was assessed in 15,847 MRI scans from 52 datasets

14 Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands

15 Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

16 Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands

17 Department of Computer Science, Lagos State University, Lagos, Nigeria

18 Brain Research Imaging Centre, University of Edinburgh, Edinburgh, UK

19 Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales (UNSW), Sydney, Australia

20 Mathematics and Statistics, Murdoch University, Murdoch, Australia

21 Department of Psychiatry, University of Münster, Münster, Germany

22 Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Sud, University Paris Descartes -Sorbonne Paris Cité, Paris, France

23 Department of Psychiatry, University Hospital Marqués de Valdecilla, School of Medicine, University of Cantabria-IDIVAL, Santander, Spain

24 CIBERSAM, Centro Investigación Biomédica en Red Salud Mental, Santander, Spain

25 Imaging Genetics Center, Institute for Neuroimaging & Informatics, Keck School of Medicine of the University of Southern California, Los Angeles, USA

26 Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim, Germany

27 Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

28 Centre for Cognitive Ageing and Cognitive Epidemiology, Psychology, University of Edinburgh, Edinburgh, UK

29 Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

30 Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) Collaboration, Department of Neuroimaging Sciences, University of Edinburgh, Edinburgh, UK

How ENIGMA Information is Used in Papers: (II) Supplementary Information

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

Site	Sequence	Field Strength	Acquisition Direction	# of Slices	Slice Gap	Voxel Size (mm3)	TI	TE	TR	Flip Angle	Citation	Segmentation
Amsterdam	3D T1-weighted turbo field echo (TFE)	3T scanner Philips Gyroscan Intera	coronal	182	0mm	1x1x1.2	0ms	4.6ms	9.621ms	8	1, 2	FreeSurfer (5.0)
Barcelona (Site 1: FIDMAG)	3D T1-weighted enhanced fast gradient echo (EFGRE3D)	1.5T GE Signa	Axial	180	0mm	0.47×0.47×1	710ms	3.93ms	2000ms	15	3, 4	FreeSurfer (5.3.0)

Using ENIGMA Metadata: (II) Automated Generation of Tables for Papers

Generated image acquisition protocol table:

Cohort	Data Type	Scanner	Acquisition Direction	Sequence	Data Acquisition Matrix	Flip Angle	Number of Slices	Scan Time	TE	TI	TR	Voxel Size
CLING	T1-weighted MRI	3T Magnetom TIM Trio	Sagittal	MPRAGEsequence	256 x 256	9	192	8 min 26 sec	3.26 ms	900 ms	2250 ms	1 mm^3
HMS	T1-weighted MRI	1.5T Magnetom Sonata	Sagittal	MPRAGEsequence	256 x 256	15	176	5 min	4.0 ms	700 ms	1900 ms	1 mm^3

Generated demographics table:

Cohort	Total	Control Total	Patient Total	Male Patients	Female Patients
CLING	372	323	49	36	13
HMS	101	55	46	32	14

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[1], Tejal Patted[2], **Yolanda Gil**[2], **Daniel Garijo**[2], Varun Ratnakar[2], Jie Ji[2], Prince Wang[1], Aggie McMahon[3], Paul M. Thompson[3], and Neda Jahanshad[3]

[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

First Workshop on Enabling Open Semantic Science (SemSci 2017)