National Database for Autism Research

Strategies for Sharing Heterogeneous Data

February 6, 2012

Greg Farber, Director
Office of Technology Development and Coordination
National Institute of Mental Health
Take Home Messages

- The maturity of a field will determine the data sharing mechanisms that have a chance to work.
- Young fields often need a “nudge” from either the government, the journals, or both
- If there is any possibility of a subject participating in experiments in different labs, a GUID should be used.
- If the data sharing infrastructure is not easy for the investigators to use, they won’t.
National Database for Autism Research (NDAR)

- Joint initiative supported by NIMH, NICHD, NINDS, and NIEHS and the NIH Center for Information Technology
  - Federally controlled system
  - Organizationally NDAR resides in the NIMH Office of Technology Development and Coordination
  - Data access committee is made up of NIH Program Officers/Human Subjects Experts
  - Trans-NIH program officer advisory committee
  - Relevant ICDs act as the Board of Governors for the project
- Begun in late 2006
- First data received in 2008 (NIH ACEs)
NDAR Implementation:

- **Global Unique Identifier (GUID)** – universal subject ID
- **Data Dictionary** – flexible framework for harmonizing data across studies and institutions
- **Data Sharing Regimen** – community and NIH input
- **Federation** – deep linkage to other relevant repositories
NDAR Data Dictionary

- Flexible and extensible framework for data definition
- 200+ instruments, image definitions, genomics defined
  - 29,000+ unique data elements and growing
  - A research community platform for defining the complex language characterizing autism research
    - Clinical
    - Genomics/Proteomics
    - Imaging Modalities
    - Eye Tracking
- Accommodates any data type and data structure
- Extended and enhanced by the ASD research community
- Curated by NDAR
Funded by the National Institutes of Health, the National Database for Autism Research (NDAR) is a secure research data repository promoting scientific data sharing and collaboration among autism spectrum disorder (ASD) investigators. NDAR’s goal is to help accelerate scientific discovery in ASD through data sharing, data harmonization, and the reporting of research results. Data from over 25,000 research participants are available to qualified investigators through the NDAR portal. Summary information on the research projects sharing data, the type of data shared and data definitions are provided below.

## Available Records: 106,210

### Select Data Source(s):
- [ ] NDAR
- [ ] Pediatric MRI Data Repository

### Select Gender

### Age From (in months)

### Select Phenotype
- Select Phenotype
- Autism-like Developmental Disorders
- Fragile X
- Non-ASD Control
- Not Enough Data Provided

### Select Sub Phenotype
- Select Sub Phenotype
- Mildly Affected
- Severely Affected

### Available Data Type(s)

#### Clinical Assessments
- Aberrant Behavior Checklist - Community
- Autism Diagnostic Interview - Revised (ADI-R) - (2003)
- Autism Diagnostic Interview Rev. (ADI-R) Toddler 2004
- Autism Diagnostic Observation Schedule (ADOS) Toddler
- Autism Diagnostic Observation Schedule (ADOS) Module 1
- Autism Diagnostic Observation Schedule (ADOS) Module 2
- Autism Diagnostic Observation Schedule (ADOS) Module 3
- Autism Diagnostic Observation Schedule (ADOS) Module 4
- Autism quotient
- CELF-4
- CHARGE Family Characteristics Questionnaire

#### Imaging
- MRI Spectroscopy
- MRI DTI
- MRI Structural
- MRI

#### Genomics
- Drug metabolism analysis / microarray
- Gene expression / PCR array
- Gene expression / microarray
- Gene expression / sequencing
- Gene expression / microarray
- Gene regulation / mass spectrometry
- Gene regulation / microarray
- Genetic analysis / high throughput sequencing
- Metabolomic analysis / chromatography
- Metabolomic analysis / mass spectrometry
- Protein interaction / biosensor systems
- Protein interaction / Immunoassay

Summary Data Approach - pdf
Casanova, Manuel F; El-Baz, Ayman; Elnakhk, Ahmed; Switala, Andrew E; Williams, Emily L; Williams, Diane L; Minshew, Nancy J; Conturo, Thomas E, "Autism : the international journal of research and practice," Quantitative analysis of the shape of the corpus callosum in patients with autism and comparison individuals.

Conturo, Thomas E; Williams, Diane L; Smith, Charles D; Gultepe, Eren; Akbudak, Erbil; Minshew, Nancy J, "Journal of the International Neuropsychological Society : JINS," Neuronal fiber pathway abnormalities in autism: an initial MRI diffusion tensor tracking study of hippocampo-fusiform and amygdalo-fusiform pathways.

Damarla, Saudamini Roy; Keller, Timothy A; Kana, Rajesh K; Cherkassky, Vladimir L; Williams, Diane L; Minshew, Nancy J; Just, Marcel Adam, "Autism research : official journal of the International Society for Autism Research," Cortical underconnectivity coupled with preserved visuospatial cognition in autism: Evidence from an fMRI study of an embedded figures task.

(14) publications found, displaying 3 record(s)
### NDAR Data Dictionary

#### Clinical Assessments
- Aberrant Behavior Checklist - Community v01
- AGRE ADOS Module 1 2001 v01
- AGRE ADOS Module 2 2001 v01
- AGRE ADOS Module 3 2001 v01
- AGRE ADOS Module 4 2001 v01
- AGRE ADOS-G Module 1 (2000 or earlier) v01
- AGRE ADOS-G Module 2 (2000 or earlier) v01
- AGRE ADOS-G Module 3 (2000 or earlier) v01
- AGRE ADOS-G Module 4 (2000 or earlier) v01
- AGRE Autism Diagnostic Interview - R (ADI-R) 2001 v01
- AGRE Autism Diagnostic Interview - R (ADI-R) 2002 v01
- ATP Brain Donor Genetic Tests v01

#### Table: AGRE ADOS Module 1 2001

<table>
<thead>
<tr>
<th>ElementName</th>
<th>DataType</th>
<th>Size</th>
<th>Required</th>
<th>ElementDescription</th>
<th>ValueRange</th>
</tr>
</thead>
<tbody>
<tr>
<td>subjectkey</td>
<td>GUID</td>
<td>0</td>
<td>Required</td>
<td>The NDAR Global Unique Identifier (GUID) for subjects which identifies a subject in NDAR</td>
<td>NDAR*</td>
</tr>
<tr>
<td>src_subject_id</td>
<td>String</td>
<td>20</td>
<td>Required</td>
<td>The site's subject identification</td>
<td></td>
</tr>
<tr>
<td>src_record_id</td>
<td>Integer</td>
<td>0</td>
<td>Optional</td>
<td>Research site ID for record</td>
<td></td>
</tr>
<tr>
<td>interview_date</td>
<td>Date</td>
<td>0</td>
<td>Required</td>
<td>Date on which the interview/genetic test/sampling/imaging was completed</td>
<td>MM/yyyy; MM/dd/yyyy</td>
</tr>
<tr>
<td>interview_age</td>
<td>Integer</td>
<td>0</td>
<td>Required</td>
<td>Age in months at the time of the interview/test/sampling/imaging.</td>
<td>0 to 1200</td>
</tr>
<tr>
<td>gender</td>
<td>String</td>
<td>50</td>
<td>Optional</td>
<td>Gender</td>
<td>M; F</td>
</tr>
<tr>
<td>untestable</td>
<td>String</td>
<td>50</td>
<td>Optional</td>
<td>Was the patient untestable?</td>
<td>Yes; No; N/A</td>
</tr>
<tr>
<td>untestable_reason</td>
<td>String</td>
<td>255</td>
<td>Optional</td>
<td>what was the reason he/she was untestable?</td>
<td></td>
</tr>
<tr>
<td>olang</td>
<td>Integer</td>
<td>0</td>
<td>Optional</td>
<td>Overall Level of Non-echoed Language</td>
<td>0; 1; 2; 3; 8</td>
</tr>
<tr>
<td>f voc</td>
<td>Integer</td>
<td>0</td>
<td>Optional</td>
<td>Frequency of Vocalization Directed to Others</td>
<td>0; 1; 2; 3</td>
</tr>
<tr>
<td>inton</td>
<td>Integer</td>
<td>0</td>
<td>Optional</td>
<td>Intonation of Vocalizations or Verbalizations</td>
<td>0; 1; 2; 8</td>
</tr>
</tbody>
</table>
The REDCap Consortium is comprised of 307 active institutional partners from CTSA, GCRC, RCMI and other institutions, and it supports a secure web application (REDCap) designed exclusively to support data capture for research studies. The REDCap application allows users to build and manage online surveys and databases quickly and securely, and is currently in production use or development build-status for more than 25,010 studies with over 36,140 end-users spanning numerous research focus areas across the consortium.
**Social effects of oxytocin in humans: context and person matter**

Bartz JA, Zaki J, Bolger N, Ochsner KN
Mount Sinai School of Medicine, Seaver Autism Center, Department of Psychiatry, New York, New York, 10029, USA. jennifer.bartz@mssm.edu


**Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of ASD**

Ingelhalikar M, Parker D, Bloy L, Roberts TP, Verma R
Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA.


**Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations**


**Link to Publication**

**Link to Data**
NDAR Implementation:

- **Global Unique Identifier (GUID)** – universal subject ID
- **Data Dictionary** – flexible framework for harmonizing data across studies and institutions
- **Data Sharing Regimen** – community and NIH input
- **Federation** – deep linkage to other relevant repositories
Data Sharing Regime

- Policies, procedures, & agreements
  - NDAR Policy
  - Data Submission Agreement
  - Data Access Agreement
  - NDAR Standard Operating Procedures
  - NDAR Data Federation MOU
- Data sharing expectations in FOAs
  - PA-10-158 R01s for ASD Research
  - PA-10-159 R03s for ASD Research
  - PA-10-160 R21s for ASD Research
  - Other RFAs
- Cost model for data sharing expected with applications
- Data sharing terms included in the terms of award
  - Descriptive (clinical, baseline, structural MRI, raw genomics/proteomics)
    - Submitted every six months – shared 4 months later
  - Experimental (all other non-identifying data)
    - End of grant + 1 year or publication of primary objectives
NDAR Implementation:

- **Global Unique Identifier** (GUID) – universal subject ID
- **Data Dictionary** – flexible framework for harmonizing data across studies and institutions
- **Data Sharing Regimen** – community and NIH input
- **Federation** – deep linkage to other relevant repositories
The National Database for Autism Research

Pediatric MRI Study of Normal Brain Development

ATP Federated Clinical Assessments

ATP Brain MRI data and Images

IAN Interactive Autism Network

AGRE Restricted Access Permissions Group

AGRE General Access Permission Group
Take Home Messages

- The maturity of a field will determine the data sharing mechanisms that have a chance to work.
- Young fields often need a “nudge” from either the government, the journals, or both.
- If there is any possibility of a subject participating in experiments in different labs, a GUID should be used.
- If the data sharing infrastructure is not easy for the investigators to use, they won’t.
Contact Information

- Email: ndarhelp@mail.nih.gov or halldan@mail.nih.gov
- Watch Short Video Tutorials at: http://ndar.nih.gov/ndarpublicweb/training.go