The National Database for Autism Research

April 8, 2014

Greg Farber, Ph.D.
Director
Office of Technology Development and Coordination
National Institute of Mental Health
National Institutes of Health
NDAR Overview

- Joint initiative supported by NIMH, NICHD, NINDS, and NIEHS
  - Federal data repository
  - Contains data from human subjects related to autism (and control subjects)
  - Data are available to the research community through a not too difficult application process
  - Summary data are available to everyone with a browser
- Begun in late 2006, and first data was received in 2008
- The data types include demographic data, clinical assessments, imaging data, and –omic data
- Currently has data available from nearly 70,000 subjects
- ~400TB of imaging and –omic data is securely stored in the cloud
NDAR Implementation

- NDAR has deep federation with the following data repositories. This federation allows NDAR to query data in those repositories and to return data to the user from multiple repositories simultaneously.
  - Autism Tissue Program
  - Autism Genetic Resource Exchange
  - Interactive Autism Network
  - Simons Foundation Autism Research Initiative

- NDAR has two key features to allow data standardization and aggregation: data dictionaries and the Global Unique Identifier (GUID)

- Generally, NIH funded investigators are expected to share their data via NDAR. Investigators with funding from other sources are welcome to deposit their data.

- Over 150 studies have registered data.
The NDAR data dictionary is one of the key building blocks for this repository. It provides a flexible and extensible framework for data definition by the research community.

- 500+ instruments, freely available to anyone
  - 50,000+ unique data elements and growing
  - A research community platform for defining the complex language characterizing autism research
    - Clinical
    - Genomics/Proteomics
    - Imaging Modalities
- Accommodates any data type and data structure
- Extended and enhanced by the ASD research community
- Curated by NDAR
- Allows investigators to quickly perform quality control tests of their data without submitting data anywhere.
<table>
<thead>
<tr>
<th>TITLE</th>
<th>SHORT NAME</th>
<th>TYPE</th>
<th>SOURCE</th>
<th>CATEGORY</th>
<th>SUBMISSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE Family Medical History</td>
<td>ace_fammedhist01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE Subject Medical History</td>
<td>ace_subjmedhist01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE Subject Physical Exam</td>
<td>ace_physexam01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGRE ADOS Module 1 2001</td>
<td>agre_ados1_200102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGRE ADOS Module 2 2001</td>
<td>agre_ados2_200102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGRE ADOS Module 3 2001</td>
<td>agre_ados3_200102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGRE ADOS Module 4 2001</td>
<td>agre_ados4_200102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGRE ADOS-G Module 1 (2000 or earlier)</td>
<td>agre_ados1_pre200002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGRE ADOS-G Module 2</td>
<td>agre_ados2_pre200002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Element Name</strong></td>
<td><strong>Data Type</strong></td>
<td><strong>Size</strong></td>
<td><strong>Required</strong></td>
<td><strong>Element Description</strong></td>
<td><strong>Value Range</strong></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------</td>
<td>----------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>subjectkey</td>
<td>GUID</td>
<td></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>interview_date</td>
<td>Date</td>
<td></td>
<td>Required</td>
<td>Date on which the interview/genetic test/sampling/imaging was completed</td>
<td>Required field</td>
</tr>
<tr>
<td>cycle</td>
<td>Integer</td>
<td></td>
<td>Recommended</td>
<td>Timepoint information</td>
<td></td>
</tr>
<tr>
<td>patid</td>
<td>String</td>
<td>20</td>
<td>Recommended</td>
<td>src_Subject_id</td>
<td>A Participant ID provided by DM-STAT must be present to process form</td>
</tr>
<tr>
<td>network</td>
<td>String</td>
<td>20</td>
<td>Recommended</td>
<td>Network</td>
<td>CPEA or STAART Network</td>
</tr>
<tr>
<td>site</td>
<td>String</td>
<td>100</td>
<td>Recommended</td>
<td>Site</td>
<td>Study Site</td>
</tr>
<tr>
<td>study</td>
<td>String</td>
<td></td>
<td>Recommended</td>
<td>Study</td>
<td>CPEA or STAART Study</td>
</tr>
<tr>
<td>ados_alg_cmsit</td>
<td>Integer</td>
<td></td>
<td>Recommended</td>
<td>Communication Social Interaction Total</td>
<td>0 : 24</td>
</tr>
<tr>
<td>ados_alg_cmmt</td>
<td>Integer</td>
<td></td>
<td>Recommended</td>
<td>Communication Total</td>
<td>0 : 10</td>
</tr>
<tr>
<td>ados_alg_imgr</td>
<td>Integer</td>
<td></td>
<td>Required</td>
<td>Imagination Creativity</td>
<td>0 : 2</td>
</tr>
<tr>
<td>ados_alg_sbrit</td>
<td>Integer</td>
<td></td>
<td>Required</td>
<td>Stereotyped Behaviors and Restricted Interests Total</td>
<td>0 : 6</td>
</tr>
<tr>
<td>ados_alg_sbrits</td>
<td>Integer</td>
<td></td>
<td>Required</td>
<td>ADOS: Stereotyped Behaviors-Restricted Interests Severity Score</td>
<td></td>
</tr>
<tr>
<td>ados_alg_sclit</td>
<td>Integer</td>
<td></td>
<td>Required</td>
<td>Social Interaction Total</td>
<td>0 : 14</td>
</tr>
<tr>
<td>ados_alg_sclits</td>
<td>Integer</td>
<td></td>
<td>Required</td>
<td>ADOS: Social Severity Score</td>
<td></td>
</tr>
<tr>
<td>ados_date</td>
<td>Integer</td>
<td></td>
<td>Recommended</td>
<td>ADOS Date</td>
<td></td>
</tr>
<tr>
<td>ados_dia_class</td>
<td>String</td>
<td>50</td>
<td>Recommended</td>
<td>ADOS Classification</td>
<td></td>
</tr>
<tr>
<td>ados_dia_diag</td>
<td>String</td>
<td>75</td>
<td>Recommended</td>
<td>Overall Diagnosis</td>
<td></td>
</tr>
<tr>
<td>ados_dx</td>
<td>String</td>
<td>15</td>
<td>Recommended</td>
<td>ADOS Diagnosis</td>
<td></td>
</tr>
<tr>
<td>ados_lac_afig</td>
<td>Integer</td>
<td></td>
<td>Recommended</td>
<td>Asks for Information</td>
<td>Nulled value range of: 0;1;2;3 to accept outlier data</td>
</tr>
</tbody>
</table>
Distribution for Data Structure: ados1_200102 and Element: scoresumm_btotal

Description
Social Interaction Total

Value Range
No Restriction

Notes
None Provided

*112 subjects have no value provided for scoresumm_btotal

No filters currently applied. Click bar on chart to add filter.
Global Unique Identifier – the Other Building Block

- The NDAR GUID software allows any researcher to generate a unique identifier using some information from a birth certificate.
- If the same information is entered in different laboratories, the same GUID will be generated.
- This strategy allows NDAR to aggregate data on the same subject collected in multiple laboratories without holding any of the personally identifiable information about that subject.
- The GUID is now being used in other research communities and can be made available to you. We have created a video to help with informed consent issues. [http://www.youtube.com/watch?v=Tb6euCVoous](http://www.youtube.com/watch?v=Tb6euCVoous)
The National Database for Autism Research (NDAR) is an NIH-funded research data repository that aims to accelerate progress in autism spectrum disorders (ASD) research through data sharing, data harmonization, and the reporting of research results. NDAR also serves as a scientific community platform and portal to multiple other research repositories, allowing for aggregation and secondary analysis of data.

Data Distribution
98,239 subjects by age, 69,719 individuals

Gender
- Female: 60,244
- Male: 34,537
- Not Reported: 4,458

Phenotypic
- Autism Sx: 27,727
- Autism Sx: 8,710
- Autism Sx: 14,324
- Typical Cx: 9,006
- Parental Cx: 23,729
- Sibling Cx: 13,329
- Neurological: 8,710
- Fragile X: 7,709
- Not Defined: 9,935

Neuroimaging
- DTI: 1,423
- fMRI: 1,423
- MRI: 1,423
- Spectroscopy: 1,423
- EEG: 1,718

Genomic
- SNP and indel: 9,935
- Next Gen: 7,709
- Gene expr: 1,716
- Gene reg: 7,709
- Stem cells: 7,709
- Gene exp: 7,709
- STR Genes: 7,709
An Example of Data Associated with a Particular Laboratory

**Collection Title:** Biological and Information Processing Mechanisms Underlying Autism

**Investigators:** Nancy Minshew, M.D., Mark Strauss, Ph.D., Kevin Pelphrey, Ph.D., Marcel Just, Ph.D., Thomas Mitchell, Ph.D., Diane Williams, Ph.D. (Owner: Minshew, Nancy)

**Collection Description:** This center focuses on elucidating fundamental information processing and neurobiological mechanisms causing autism with studies of infant siblings, first-diagnosed toddlers, and groups of children, adolescents, and adults with and without autism. Project I: Development of Categorization & Facial...

**Download Data**

**Grant Information:**

<table>
<thead>
<tr>
<th>Project Number</th>
<th>Project Title</th>
<th>Start Date</th>
<th>End Date</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>P50HD55748</td>
<td>Biological and Information Processing Mechanisms Underlying Autism</td>
<td>08/06/2007</td>
<td>07/31/2012</td>
<td>UNIVERSITY OF PITTSBURGH AT PITTSBURGH</td>
</tr>
</tbody>
</table>

**Publications** (Showing 3 of 39) Show All

- Dundas, Eva M; Best, Catherine A; Minshew, Nancy J; Strauss, Mark S. "Journal of autism and developmental disorders" A lack of left visual field bias when individuals with autism process faces.
- Mazefsky, Carla A; Oswald, Donald P; Day, Taylor N; Eack, Shaun M; Minshew, Nancy J; Lainhart, Janet E. "Journal of clinical child and adolescent psychology : the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53" ASD, a psychiatric disorder, or both? Psychiatric diagnoses in adolescents with high-functioning ASD.

**Data Structures:**

<table>
<thead>
<tr>
<th>Title</th>
<th>Type</th>
<th>Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism Diagnostic Interview, Revised (ADI-R)</td>
<td>Clinical Assessments</td>
<td>212</td>
</tr>
<tr>
<td>Autism Diagnostic Observation Schedule - Module 1</td>
<td>Clinical Assessments</td>
<td>97</td>
</tr>
<tr>
<td>Autism Diagnostic Observation Schedule - Module 2</td>
<td>Clinical Assessments</td>
<td>67</td>
</tr>
<tr>
<td>Autism Diagnostic Observation Schedule - Module 3</td>
<td>Clinical Assessments</td>
<td>105</td>
</tr>
<tr>
<td>Autism Diagnostic Observation Schedule - Module 4</td>
<td>Clinical Assessments</td>
<td>200</td>
</tr>
<tr>
<td>Benton Facial Recognition Test</td>
<td>Clinical Assessments</td>
<td>341</td>
</tr>
<tr>
<td>CELF-4 Clinical Eval of Lang Fundamentals, 4th ed</td>
<td>Clinical Assessments</td>
<td>33</td>
</tr>
<tr>
<td>CHARGE Family Characteristics Questionnaire</td>
<td>Clinical Assessments</td>
<td>399</td>
</tr>
<tr>
<td>CHARGE Medical History</td>
<td>Clinical Assessments</td>
<td>312</td>
</tr>
<tr>
<td>CHARGE Physical Exam</td>
<td>Clinical Assessments</td>
<td>200</td>
</tr>
</tbody>
</table>
Results in 750 subjects being discovered.
In the past 2 years, NDAR has accumulated significant imaging and genomics data.

Both of these data types are harder to query and make easily useful than the clinical and demographic data in NDAR.

We are very interested in working with anyone who is interested to collaborate on ways to query the data or on ways to create data processing pipelines that can work on the data we have in the cloud.

Current collaborators: David Kennedy and Jack Van Horn for imaging, Evan Eichler in genomics.
How is NDAR being used?

- With biological databases, it is not true that if you build it they will come.
- More than 270 users have been granted access to NDAR. Data access is separate from those who are depositing data.
- David M. Richman and colleagues have published a study, “Predictors of self-injurious behavior exhibited by individuals with autism spectrum disorder” where all of the data in the paper came from NDAR (J. Intellect. Disabil. Res. (2013), 57:429-439).
- Vinod Menon and colleagues have published a paper, “Brain hyperconnectivity in children with autism and its links to social deficits” (Cell Rep. (2013), 5(3), 738-747. where some of the data is from NDAR and some is newly measures.
- Many are using data from NDAR as part of NIH grant applications.
Summary

NDAR, is a useful data archive that makes autism data:

A) Discoverable – federation, useful queries, XML web services
B) Useful to Others – data access, data QC, data analysis pipelines
C) Citable – data from labs, data from papers
D) Linked to the Literature – data link in PubMed