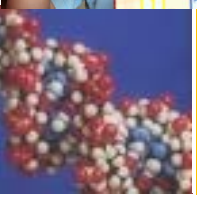
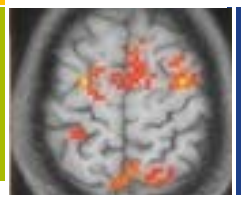
The National Database for Autism Research

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

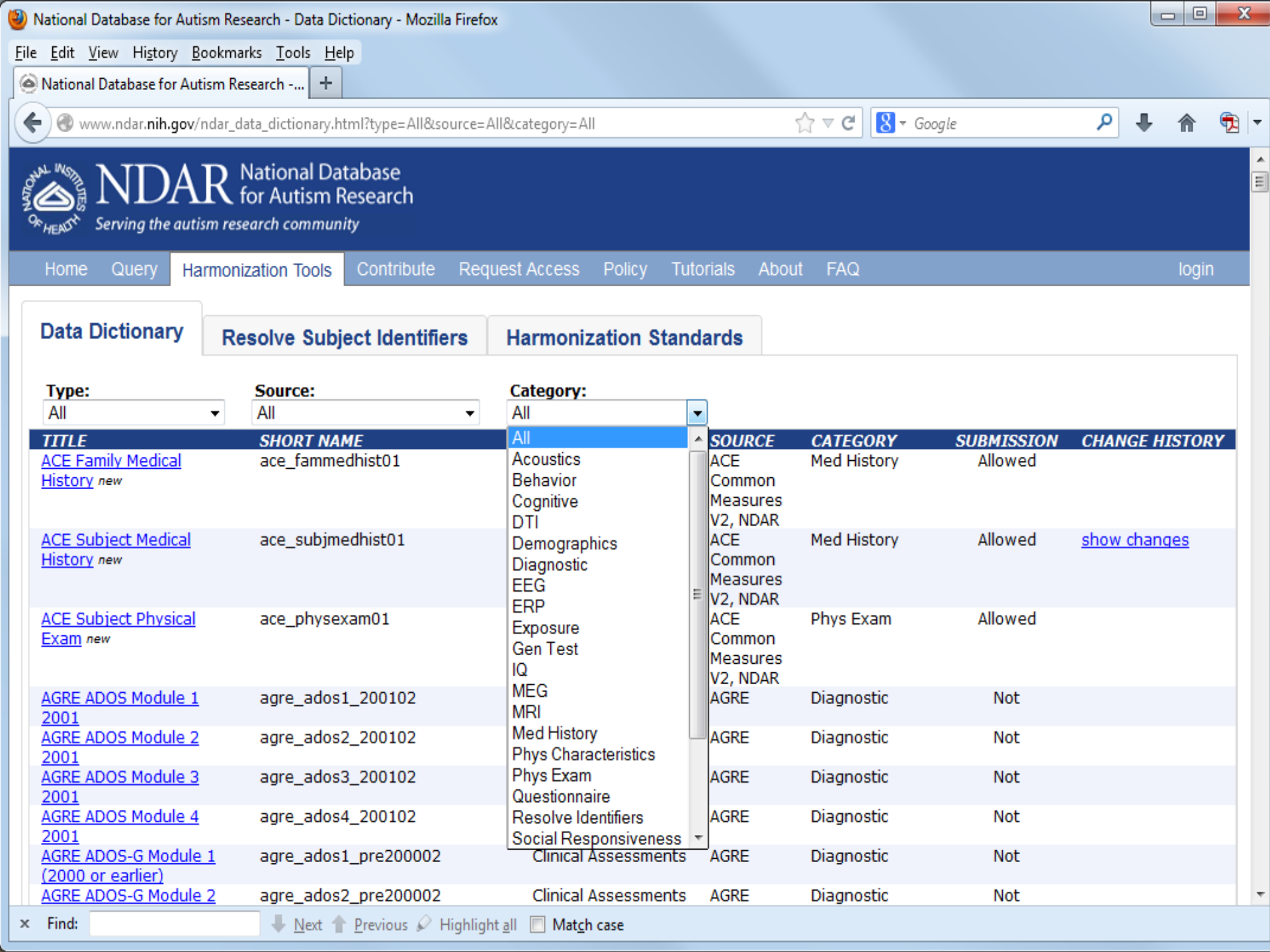
- Joint initiative supported by NIMH, NICHD, NINDS, and NIEHS
  - Federal data repository
  - Contains data from human subjects related to autism (and control subjects)
  - Data are available to the research community through a not too difficult application process
  - Summary data are available to everyone 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.

Data Dictionary

- 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.


National Database for Autism Research - Data Dictionary - Mozilla Firefox
File Edit View History Bookmarks Tools Help
National Database for Autism Research -...
www.ndar.nih.gov/ndar_data_dictionary.html?type=All&source=All&category=All
Google
NDAR National Database for Autism Research
Serving the autism research community
Home Query Harmonization Tools Contribute Request Access Policy Tutorials About FAQ login
Data Dictionary
Resolve Subject Identifiers
Harmonization Standards
Type:
All
Source:
All
Category:
All
All
Acoustics
Behavior
Cognitive
DTI
Demographics
Diagnostic
EEG
ERP
Exposure
Gen Test
IQ
MEG
MRI
Med History
Phys Characteristics
Phys Exam
Questionnaire
Resolve Identifiers
Social Responsiveness
TITLE
SHORT NAME
SOURCE
CATEGORY
SUBMISSION
CHANGE HISTORY
ACE Family Medical History new
ace_fammedhist01
ACE Common Measures V2, NDAR
Med History
Allowed
ACE Subject Medical History new
ace_subjmedhist01
ACE Common Measures V2, NDAR
Med History
Allowed
show changes
ACE Subject Physical Exam new
ace_physexam01
ACE Common Measures V2, NDAR
Phys Exam
Allowed
AGRE ADOS Module 1 2001
agre_ados1_200102
AGRE
Diagnostic
Not
AGRE ADOS Module 2 2001
agre_ados2_200102
AGRE
Diagnostic
Not
AGRE ADOS Module 3 2001
agre_ados3_200102
AGRE
Diagnostic
Not
AGRE ADOS Module 4 2001
agre_ados4_200102
AGRE
Diagnostic
Not
AGRE ADOS-G Module 1 (2000 or earlier)
agre_ados1_pre200002
Clinical Assessments
AGRE
Diagnostic
Not
AGRE ADOS-G Module 2
agre_ados2_pre200002
Clinical Assessments
AGRE
Diagnostic
Not
Find:
Next
Previous
Highlight all
Match case

CPEA STAART ADOS G Module 4

Short Name: cs_ados_g_4
Version: 02

ADOS G Module 4 as defined by the CPEA STAART project

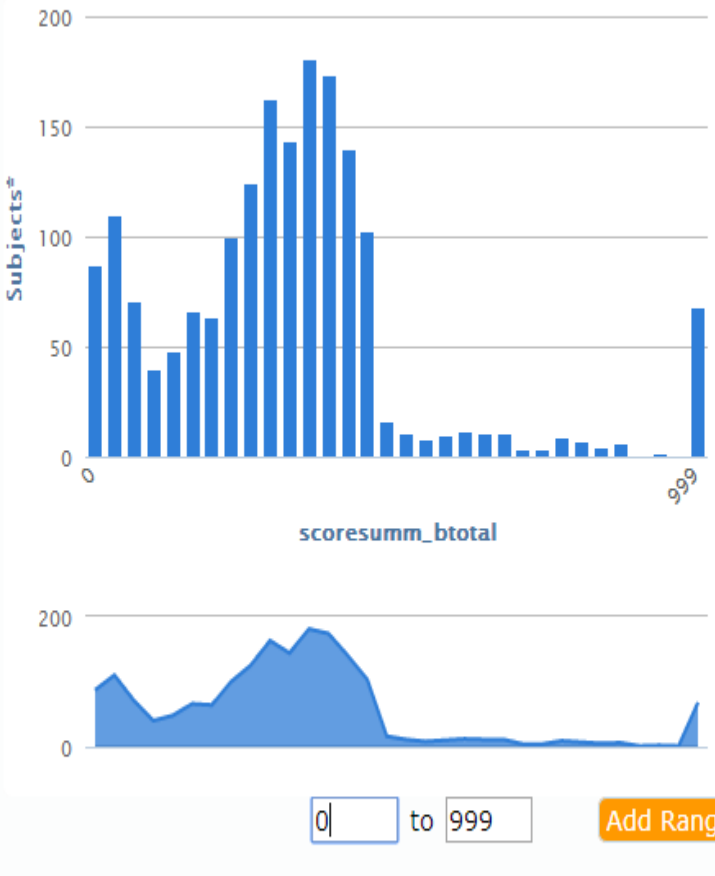
Download Definitions

ElementName	DataType	Size	Required	ElementDescription	ValueRange	Notes	Aliases
subjectkey	GUID		Required	The NDAR Global Unique Identifier (GUID) for subjects which identifies a subject in NDAR	NDAR*		
interview_date	Date		Required	Date on which the interview/genetic test/sampling/imaging was completed		Required field	ADOS_DATE_STD
cycle	Integer		Recommended	Timepoint information			
patid	String	20	Recommended	src_Subject_id		A Participant ID provided by DM-STAT must be present to process form	
network	String	20	Recommended	Network		CPEA or STAART Network	
site	String	20	Recommended	Site		Study Site	
study	String	100	Recommended	Study		CPEA or STAART Study	
ados_alg_cmsit	Integer		Recommended	Communication Social Interaction Total	0 :: 24		
ados_alg_commt	Integer		Recommended	Communication Total	0 :: 10		
ados_alg_imgcr	Integer		Recommended	Imagination Creativity	0 :: 2		
ados_alg_sbrit	Integer		Recommended	Stereotyped Behaviors and Restricted Interests Total	0 :: 6		
ados_alg_sbrits	Integer		Recommended	ADOS: Stereotyped Behaviors-Restricted Interests Severity Score			
ados_alg_sclit	Integer		Recommended	Social Interaction Total	0 :: 14		
ados_alg_sclits	Integer		Recommended	ADOS: Social Severity Score			
ados_date	Integer		Recommended	ADOS Date			
ados_dia_class	String	50	Recommended	ADOS Classification			
ados_dia_diag	String	75	Recommended	Overall Diagnosis			
ados_dx	String	15	Recommended	ADOS Diagnosis			
ados_lac_afinf	Integer		Recommended	Asks for Information		Nulled value range of: 0;1;2;3 to accept outlier data	

National Database for Aut ×

ndar.nih.gov/ndar_data_dictionary.html?short_name=ados1_200102

Distribution for Data Structure: ados1_200102 and Element: scoresumm_btotal



0 to 999 Add Range

*112 subjects have no value provided for scoresumm_btotal

Description
Social Interaction Total

Value Range
No Restriction

Notes
None Provided

Filters

No filters currently applied. Click bar on chart to add filter.

Return

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

National Database for Aut ×

ndar.nih.gov

NIH NDAR National Database for Autism Research

Serving the autism research community

Home | Query | Harmonization Tools | Cloud | Contribute | Request Access | Policy | Tutorials | About | FAQ | login

Quick Navigation

- **Query**
  - General Query
  - Data from Labs
  - Data from Papers
  - By Measure/Element
  - By Concept (beta)
  - omicSearch (beta)
- **Share**
  - Harmonization Standards
  - Prepare and Submit
- **Compute**
- **Resources**
  - GUID Tool
  - Validation Tool
  - Download Manager
  - Data Dictionary
- **Contact Us**
- **Request Account**

NDAR News

Recent news is available from the "About NDAR" page.

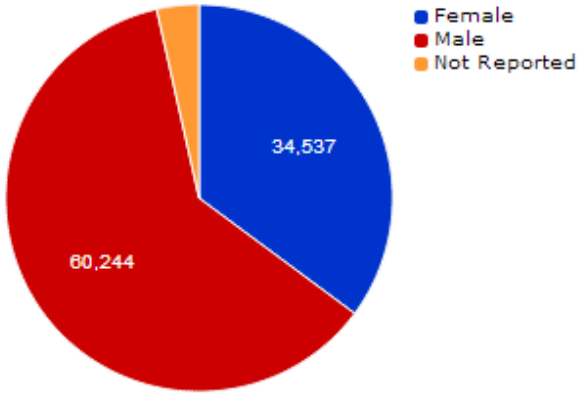
Learn About NDAR

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.

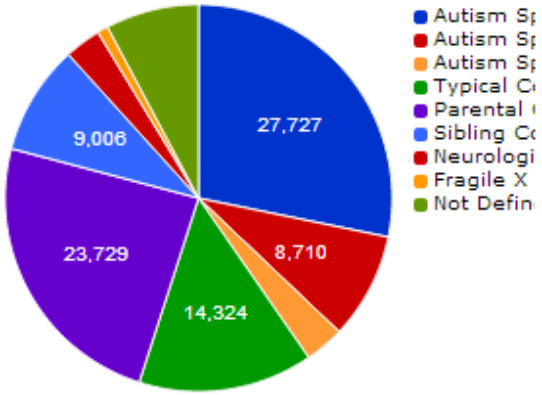
Data Distribution

98,239 subjects by age, 69,719 individuals

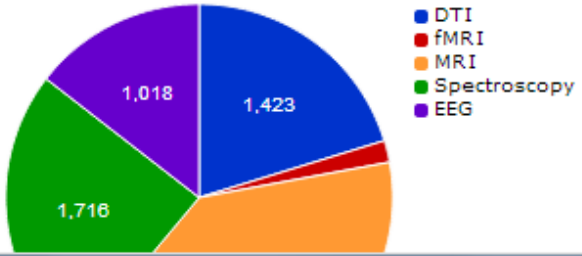
Gender[1]



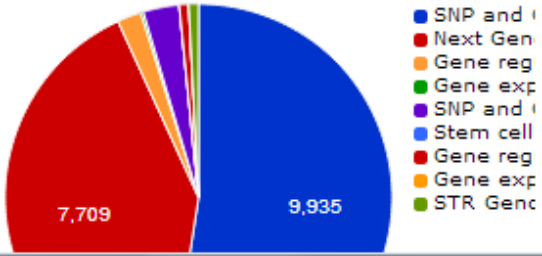
Phenotypic[1]



Neuroimaging



Genomic



An Example of Data Associated with a Particular Laboratory

National Database for Autism Research - Data - Data from Labs - Mozilla Firefox

File Edit View History Bookmarks Tools Help

National Database for Autism Research -...

www.ndar.nih.gov/data_from_labs.html

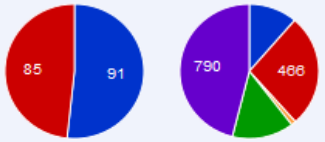
Collapse

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...

Genomics Neuroimaging Phenotype



Download Data

Grant Information:

Project Numbe	t Title	Start Date	End Date	Organization
P50HD55748	iological and Information Processing Mechanisms Underlying Autism	08/06/2007	07/31/2012	UNIVERSITY OF PITTSBURGH AT PITTSBURGH

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

Bishop-Fitzpatrick, Lauren; Minshew, Nancy J; Eack, Shaun M "Jou of autism and developmental disorders" *A Systematic Review of Psychosocial Interventions for Adults with Autism Spectrum Disorders.*

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:

Title	Type	Number of Subjects
Autism Diagnostic Interview, Revised (ADI-R)	Clinical Assessments	212
Autism Diagnostic Observation Schedule - Module 1	Clinical Assessments	97
Autism Diagnostic Observation Schedule - Module 2	Clinical Assessments	67
Autism Diagnostic Observation Schedule - Module 3	Clinical Assessments	105
Autism Diagnostic Observation Schedule - Module 4	Clinical Assessments	200
Benton Facial Recognition Test	Clinical Assessments	341
CELF-4 Clinical Eval of Lang Fundamentals, 4th ed	Clinical Assessments	33
CHARGE Family Characteristics Questionnaire	Clinical Assessments	399
CHARGE Medical History	Clinical Assessments	312
CHARGE Physical Exam	Clinical Assessments	299

National Database for Aut ×

ndar.nih.gov/query_data.html?expandFilters=true

NIH NDAR National Database for Autism Research

Serving the autism research community

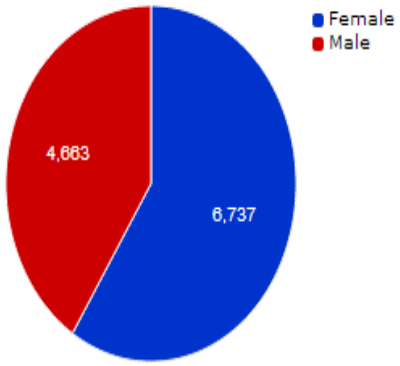
Showing all of the data in IAN

Home | Query | Harmonization Tools | Cloud | Contribute | Request Access | Policy | Tutorials | About | FAQ | login

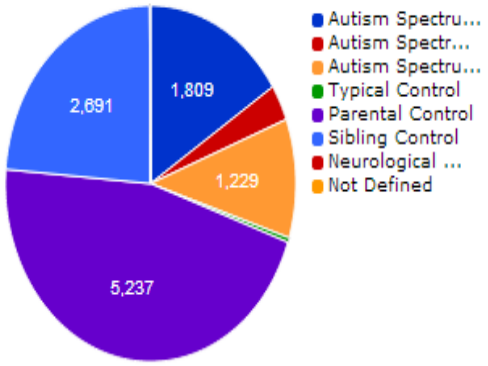
Query Data | Data from Labs | Data from Papers | Query by Data Dictionary | Query by Concept | Query Instructions

Use "Select Data" below to query the data available in NDAR. Then, select download to create a package and download your results. Use the Data tab above to search in other ways. For m
information on search see our Methods.

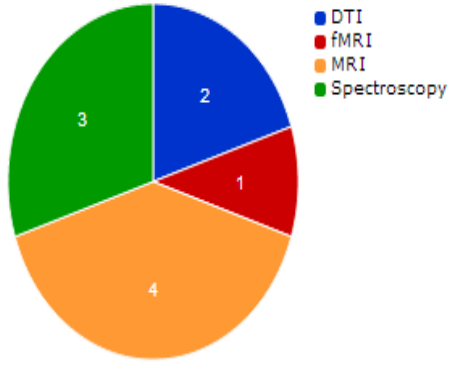
Data Distribution by Gender[1]



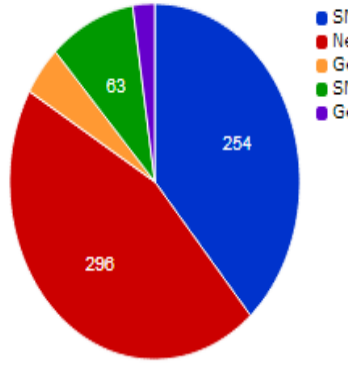
Phenotypic Data Distribution[1]



Neuroimaging Data Distribution



Genomic Data Distribution

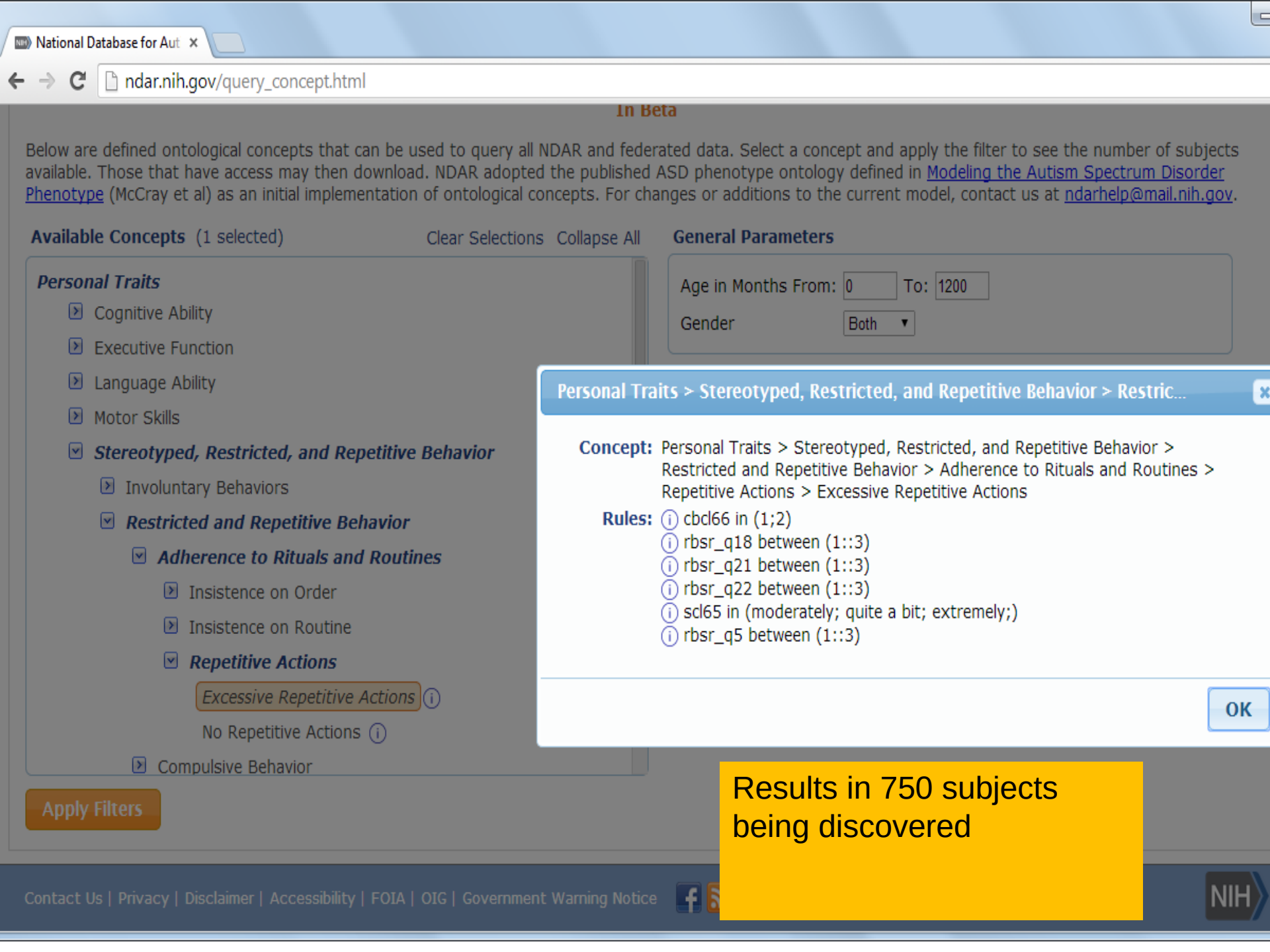


[1] Numbers reported are subjects by age

Select Data: All | Basic | Phenotype | Neuroimaging | omicSEARCH: Experiment | Results | Show Results | Reset All | Download Dat

Showing results 1 - 50 of 11,400 | Results per Page: 50 | Showing page 1 of 228 | Jump to Page: 1

DATA SOURCES	SUBJECT ID	INTERVIEW AGE	GENDER	NDAR CATEGORY	CLINICAL DIAGNOSIS	ADI SCORE	ADOS CLINICAL DIAGNOSIS
Interactive Autism Network	8	132	MALE	SIBLING CONTROL	SIBLING CONTROL		
Interactive Autism Network	9	135	MALE	SIBLING CONTROL	SIBLING CONTROL		
Interactive Autism Network	12	384	FEMALE	PARENTAL CONTROL	PARENTAL CONTROL		
Interactive Autism Network NDAR	14	108	MALE	AUTISM SPECTRUM AFFECTED	AUTISM SPECTRUM AFFECTED		


National Database for Aut
ndar.nih.gov/query_concept.html
In Beta
Below are defined ontological concepts that can be used to query all NDAR and federated data. Select a concept and apply the filter to see the number of subjects available. Those that have access may then download. NDAR adopted the published ASD phenotype ontology defined in Modeling the Autism Spectrum Disorder Phenotype (McCray et al) as an initial implementation of ontological concepts. For changes or additions to the current model, contact us at ndarhelp@mail.nih.gov.
Available Concepts (1 selected)
Clear Selections
Collapse All
General Parameters
Age in Months From: 0 To: 1200
Gender Both
Personal Traits
Cognitive Ability
Executive Function
Language Ability
Motor Skills
Stereotyped, Restricted, and Repetitive Behavior
Involuntary Behaviors
Restricted and Repetitive Behavior
Adherence to Rituals and Routines
Insistence on Order
Insistence on Routine
Repetitive Actions
Excessive Repetitive Actions
No Repetitive Actions
Compulsive Behavior
Personal Traits > Stereotyped, Restricted, and Repetitive Behavior > Restric...
Concept: Personal Traits > Stereotyped, Restricted, and Repetitive Behavior > Restricted and Repetitive Behavior > Adherence to Rituals and Routines > Repetitive Actions > Excessive Repetitive Actions
Rules:
cbcl66 in (1;2)
rbsr_q18 between (1::3)
rbsr_q21 between (1::3)
rbsr_q22 between (1::3)
scl65 in (moderately; quite a bit; extremely;)
rbsr_q5 between (1::3)
OK
Apply Filters
Results in 750 subjects being discovered
Contact Us | Privacy | Disclaimer | Accessibility | FOIA | OIG | Government Warning Notice
NIH

Imaging and Genomic Data

- 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 Hessl and collaborators used NDAR to collect and analyze their data in a private space before publication ("Psychometric study of the aberrant behavior checklist in Fragile X syndrome and implications for targeted treatment", J. Autism Dev. Disord. (2012), 42:1377-1392).
- 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