### The NIMH Data Repositories

### November 5, 2014









### **Expansion to Other Scientific Areas**

NIMH recently decided to expand NDAR to include:

- 1) Data from Clinical Trials (NOT-MH-14-005)
- 2) Data related to the Research Domain Criteria (RDoC) initiative
- Both databases now exist and data are being loaded.
- The database is shared with NDAR, so queries across all data infrastructures will be easy.



### The NIMH Data Repositories



- 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 over 77,000 subjects
- ~500TB 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 First Building Block

- 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									
<u>File Edit View History Bookmarks Tools H</u> elp									
National Database for Autist	m Research +								
🗲 🛞 www.ndar. <b>nih.gov</b> /nc	lar_data_dictionary.html?type=All&sourc	e=All&category=All		☆ マ C 8 - G	oogle	₽ ♣ ♠	<b>1</b>		
	National Database								
$\frac{2}{2}$ Serving the autism	n research community								
HERO Serving the datish	research community								
Home Query Harr	nonization Tools Contribute R	equest Access Policy Tut	orials Abou	it FAQ		login			
Data Dictionary	<b>Resolve Subject Identifiers</b>	Harmonization Stand	lards						
	_								
Type:	Source:	Category:	-						
	SHORT NAME	All	SOURCE	CATEGORY	SURMISSION	CHANGE HISTORY	× 1		
ACE Family Medical	ace_fammedhist01	Acoustics	ACE	Med History	Allowed	CHANGE HISTORI			
History new		Behavior	Common						
		DTI	V2, NDAR						
ACE Subject Medical	ace_subjmedhist01	Demographics	ACE	Med History	Allowed	show changes			
HISLOTY new		Diagnostic	Measures						
ACT Cubicat Division	01	ERP	V2, NDAR	Dhua Guan	Allawad				
Exam new	ace_physexam01	Exposure	Common	Phys Exam	Allowed				
		IQ	Measures						
AGRE ADOS Module 1	agre ados1 200102	MEG	AGRE	Diagnostic	Not				
2001		MRI Mod History	1005						
AGRE ADOS Module 2 2001	agre_ados2_200102	Phys Characteristics	AGRE	Diagnostic	Not				
AGRE ADOS Module 3	agre_ados3_200102	Phys Exam	AGRE	Diagnostic	Not				
2001 AGRE ADOS Module 4	agre ados4 200102	Questionnaire Resolve Identifiers	AGRE	Diagnostic	Not				
2001		Social Responsiveness	-						
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:	📕 <u>N</u> ext 👚 <u>P</u> revious 🖉 High	light <u>a</u> ll 📃 Mat <u>c</u> h case							

🥹 NDAR - DataStructure	es cs_ados_g_4	102 - Mo	zilla Firefox								х
File Edit View Histor	n Bookmark	r Tools	Help								
	ry <u>b</u> ookmark	s <u>1</u> 00is	Пеір								
NDAR - DataStructur	res cs_ados_g_4	402	+								
🗲 🞯 www.ndar.nih	.gov/ndarpubl	icweb/Da	ataStructures.go?sho	rt_name=cs_ados_g_402		☆ ⊽ C'	<mark>8</mark> ▼ Google	م	•	<b>^</b>	<b>1</b>
Serving the autism research community											
Home Query	Harmoniza	ation To	ols Contribute	Request Access Policy Tutoria	als About	FAQ					
Back to: <u>Home</u> > <u>Data</u>	Structures			CPEA STAART ADOS G	Module 4			Short Nam Version:	e: cs_a	dos_g_	_4 ≡
ADOS G Mod	ule 4 as def	fined by	the CPEA STAA	RT project							
Download Definitions											
ElementName	DataType	Size	Required	ElementDescription	ValueRan	ge	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_DA	TE_ST	D
cycle	Integer		Recommended	Timepoint information							
patid	String	20	Recommended	src_Subject_id			A Participant ID pr DM-STAT must be process form	ovided by present to			
network	String	20	Recommended	Network			CPEA or STAART N	etwork			
site	String	20	Recommended	Site			Study Site				
study	String	100	Recommended	Study			CPEA or STAART S	tudy			
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 0;1;2;3 to accept	e of: outlier			Ŧ

🛛 🕂 <u>N</u>ext 👕 <u>P</u>revious 🖉 Highlight <u>a</u>ll 🔲 Mat<u>c</u>h case

#### National Database for Aut ×

→ C

ndar.nih.gov/ndar\_data\_dictionary.html?short\_name=ados1\_200102



# **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.
   <u>http://www.youtube.com/watch?v=Tb6euC</u> <u>Voous</u>







- At this point, data has been received. Each subject has a GUID or a pseudo-GUID and the data have been defined in a data dictionary
- How does a user find data?







<sup>1</sup> Numbers reported are subjects by age

Select Data: All Basic P	henotype Ne	uroimaging omicSl	Earch: E	Experiment Results	Show Re	sults	Reset All		
Showing results 1 - 50 of 108,386 Results per Page: 50 🔻 Showing page 1 of 2,168 Jump to Page: 1									
DATA SOURCES	SUBJECT ID	INTERVIEW AGE	GENDER	NDAR CATEGORY		CLINICAL	DIAGNOSIS	ADI SCORE	
NDAR	1	424	FEMALE	AUTISM SPECTRUM MILDLY	AFFECTED	AUTISM SPE	CTRUM MILDLY AFFECTE	D	
NDAR	2	426	FEMALE	AUTISM SPECTRUM MILDLY	AFFECTED	AUTISM SPE	CTRUM MILDLY AFFECTE	D	
NDAR SFARI	3		FEMALE	SIBLING CONTROL		SIBLING CO	NTROL		
NDAR	4	30	MALE	NOT DEFINED					
NDAR SFARI	5		MALE	SIBLING CONTROL		SIBLING CO	NTROL		
NDAR SFARI	6	536	MALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
NDAR SFARI	7	537	MALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
Interactive Autism Network	8	11	MALE	SIBLING CONTROL		SIBLING CO	NTROL		
NDAR SFARI	9	136	MALE	AUTISM SPECTRUM SEVERE	ELY AFFECTED	AUTISM SPE	CTRUM SEVERELY AFFEC	TED NEUROLOGICAL CONTROL	
NDAR SFARI	10		FEMALE	SIBLING CONTROL		SIBLING CO	NTROL		
Interactive Autism Network	11	32	FEMALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
NDAR	10		EEMALE.	NOT DEELNED					

### An Example of Data Associated with a Particular Laboratory

onal Database for Autism Re	esearch - Data - Data from Labs - Mozilla Firefox				
ional Database for Autism R	esearch +				
www.ndar. <b>nih.gov</b> /data_f	rom_labs.html	☆ ⊽	C Soogle	م	
ollection Title:	Biological and Information Processing Mechanisms Underlying Autis	sm Geno	mics Neuroir	naging Phenotype	Collapse
vestigators:	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	)			
llection Description	This center focuses on elucidating fundamental information processing and neurobiological mechanisms causing autism with studies of infant siblings, diagnosed toddlers, and groups of children, adolescents, and adults with a without autism. Project I: Development of Categorization & Facial	first- nd	85	91 790 466	
Download Data					
Grant Information:					
Project Number	Title	Start Date	End Date	Organization	
P50HD55748	iological and Information Processing Mechanisms Underlying Autism	08/06/2007	07/31/2012	UNIVERSITY OF PITT AT PITTSBURGH	SBURGH
Pu	blications (Showing 3 of 39) Show All 💿				
Bi	shop-Fitzpatrick, Lauren; Minshew, Nancy J; Eack, Shaun M "Joy Psychosocial Interventions for Adults with Autism Spectrum Disorders.	autism and de	evelopmental dis	orders" A Systematic	Review
Du vis	Indas, Eva M; Best, Catherine A; Minshew, Nancy J; Strauss, Mark S " ual field bias when individuals with autism process faces.	Journal of autis	sm and developn	nental disorders" A la	ck of left
Ma an As	azefsky, Carla A; Oswald, Donald P; Day, Taylor N; Eack, Shaun M; Min d adolescent psychology : the official journal for the Society of Clinical Child a sociation, Division 53" ASD, a psychiatric disorder, or both? Psychiatric diag	<b>shew, Nancy</b> and Adolescent gnoses in adole	<b>J; Lainhart, Ja</b> Psychology, Amescents with high	net E "Journal of clini herican Psychological h-functioning ASD.	cal child
Data Structures:					
Title		Туре		Number of	Subjects
Autism Diagnostic	Interview, Revised (ADI-R)	Clinical Asse	ssments		212
Autism Diagnostic	Observation Schedule - Module 1	Clinical Asse	ssments		97
Autism Diagnostic	Observation Schedule - Module 2	Clinical Asse	ssments		67
Autism Diagnostic	Observation Schedule - Module 3	Clinical Asse	ssments		105
Autism Diagnostic	Observation Schedule - Module 4	Clinical Asse	ssments		200

Clinical Assessments

Clinical Assessments

Clinical Assessments

Clinical Assessments

Clinical Accordenate

341

399

312

200

33

Benton Facial Recognition Test

CHARGE Medical History

CHARCE Devoiced Even

CELF-4 Clinical Eval of Lang Fundamentals, 4th ed

CHARGE Family Characteristics Questionnaire



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 information on search see our Methods.



<sup>1</sup> Numbers reported are subjects by age

Select Data: All Basic P	henotype Ne	uroimaging omicSE	Earch: E	Experiment Results	Show Re	sults	Reset All		
Showing results 1 - 50 of 108,386 Results per Page: 50 🔻 Showing page 1 of 2,168 Jump to Page: 1									
DATA SOURCES	SUBJECT ID	INTERVIEW AGE	GENDER	NDAR CATEGORY		CLINICAL	DIAGNOSIS	ADI SCORE	
NDAR	1	424	FEMALE	AUTISM SPECTRUM MILDLY	Y AFFECTED	AUTISM SPE	CTRUM MILDLY AFFECT	ED	
NDAR	2	426	FEMALE	AUTISM SPECTRUM MILDLY	Y AFFECTED	AUTISM SPE	CTRUM MILDLY AFFECT	ED	
NDAR SFARI	3		FEMALE	SIBLING CONTROL		SIBLING CO	NTROL		
NDAR	4	30	MALE	NOT DEFINED					
NDAR SFARI	5		MALE	SIBLING CONTROL		SIBLING CO	NTROL		
NDAR SFARI	6	536	MALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
NDAR SFARI	7	537	MALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
Interactive Autism Network	8	11	MALE	SIBLING CONTROL		SIBLING CO	NTROL		
NDAR SFARI	9	136	MALE	AUTISM SPECTRUM SEVER	ELY AFFECTED	AUTISM SPE	CTRUM SEVERELY AFFE	CTED NEUROLOGICAL CO	NTROL
NDAR SFARI	10		FEMALE	SIBLING CONTROL		SIBLING CO	NTROL		
Interactive Autism Network	11	32	FEMALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
NDAR	12		EEMALE	NOT DEETNED					

National Database for Aut ×				
← → C D ndar.nih.gov/study.html?id=294			\$	•]
NIH NDAR National Database for Autism Research Serving the autism research community				
Home Query Harmonization Tools Cloud Contribute Request Acces	s Policy Tutorials About	FAQ	login 🗲	
Owner: Christopher Walsh Owner E-mail: <u>christopher.walsh@childrens.harvard.edu</u> State: Shared	Using Whole-E	xome Sequencing to Ident	ify Inherited Causes of Autism #294	
Summary Cohorts (3) Measures Types Data	Analysis			
<ul> <li>Investigators: Walsh CA; Yu TW, Chahrour MH, Coulter ME, Jiralersp Ataman B, Schmitz-Abe K, Harmin DA, Adli M, Malik A Sanders SJ, Mochida GH, Partlow JN, Sunu CM, Felie J Ware J, Joseph RM, Hill RS, Kwan BY, Al-Saffar M, Mu</li> <li>Abstract: Despite significant heritability of autism spectrum disc extreme genetic heterogeneity has proven challenging Studies of primarily simplex families have implicated or changes and point mutations, but are not optimally de Results: Results published in Neuron, Jan 2013 Readme file for study results □</li> <li>Documents:</li> </ul>	ong S, Okamura-Ikeda K, N, D'Gama AM, Lim ET, M, Rodriguez J, Nasir RH, kaddes NM, et al. orders (ASDs), their g for gene discovery. de novo copy number esigned to identify	Cohorts: Measures:	Control - Parents and siblings (411) Age: 30 to 1200 months Gender: Both Test - ASD (176) Age: 17 to 1200 months Gender: Both Test - ASD-ID (100) Age: 4 to 1200 months Gender: Both Primary Measures (7) Genomics: Genomics Sample - v03 Secondary Measures (9) Clinical Assessments: Neurodevelopmental Phenotype - v01	
		Data Analysis:	v02 Genotyping/NGS	
			Type: SNP, CNV, Exome Sequencing Software: BWA, GATK	
Return Edit				



NIH

Ç.



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 information on search see our Methods.



<sup>1</sup> Numbers reported are subjects by age

Select Data: All Basic P	henotype Ne	uroimaging omicSl	E <b>ARCH</b> : E	Experiment Results	Show Re	sults	Reset All		
Showing results 1 - 50 of 108,386 Results per Page: 50 🔻 Showing page 1 of 2,168 Jump to Page: 1									
DATA SOURCES	SUBJECT ID	INTERVIEW AGE	GENDER	NDAR CATEGORY		CLINICAL	DIAGNOSIS	ADI SCORE	
NDAR	1	424	FEMALE	AUTISM SPECTRUM MILDL	Y AFFECTED	AUTISM SPE	CTRUM MILDLY AFFECTED		
NDAR	2	426	FEMALE	AUTISM SPECTRUM MILDL	Y AFFECTED	AUTISM SPE	CTRUM MILDLY AFFECTED		
NDAR SFARI	3		FEMALE	SIBLING CONTROL		SIBLING CON	NTROL		
NDAR	4	30	MALE	NOT DEFINED					
NDAR SFARI	5		MALE	SIBLING CONTROL		SIBLING CON	NTROL		
NDAR SFARI	6	536	MALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
NDAR SFARI	7	537	MALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
Interactive Autism Network	8	11	MALE	SIBLING CONTROL		SIBLING CON	NTROL		
NDAR SFARI	9	136	MALE	AUTISM SPECTRUM SEVER	RELY AFFECTED	AUTISM SPE	CTRUM SEVERELY AFFECTED	NEUROLOGICAL CO	ONTROL
NDAR SFARI	10		FEMALE	SIBLING CONTROL		SIBLING CON	NTROL		
Interactive Autism Network	11	32	FEMALE	PARENTAL CONTROL		PARENTAL C	ONTROL		
NDAR	12		CEMALE	NOT DEEINED					

### ← → C 🗋 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 <u>Modeling the Autism Spectrum Disorder</u> <u>Phenotype</u> (McCray et al) as an initial implementation of ontological concepts. For changes or additions to the current model, contact us at <u>ndarhelp@mail.nih.gov</u>.

Available Concepts (1 selected)	Clear Selections Collapse Al	General Parameters
Personal Traits         Image: Cognitive Ability         Image: Executive Function         Image: Executive Function      <	Personal T Concep avior Rule Routines	Age in Months From:       0       To:       1200         Gender       Both       Image: Comparison of the second seco
<ul> <li>Insistence on Order</li> <li>Insistence on Routine</li> <li>Repetitive Actions</li> <li>Excessive Repetitive Actions</li> </ul>	ctions (i)	<ul> <li>i) rbsr_q21 between (1::3)</li> <li>i) rbsr_q22 between (1::3)</li> <li>i) scl65 in (moderately; quite a bit; extremely;)</li> <li>i) rbsr_q5 between (1::3)</li> </ul>
No Repetitive Actions (	0	OK
Apply Filters		Results in 750 subjects being discovered
Contact Us   Privacy   Disclaimer   Accessibility	FOIA   OIG   Government Warning Not	

# 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.



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



