



NDAR Webinar | Data Submission Planning



Agenda

- NDAR Vision and Implementation (10 minutes)
- NDAR Data Submission Planning (35 minutes)
- Questions and Answers (15 minutes)

NDAR Stakeholders

■ NIH Funding Organizations

- National Institute of Mental Health (NIMH)
- Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)
- National Institute of Neurological Disorders and Stroke (NINDS)
- National Institute of Environmental Health Sciences (NIEHS)

■ NIH Contributing Organizations

- Center for Information Technology (CIT)

■ NIH Advisory Groups

- NDAR Governance Board
- NDAR Implementation Team (NIT)
- Autism Coordinating Committee (ACC)

NDAR Team

- Mike Huerta – Director
- Anne Sperling – Policy Analyst
- Dan Hall – Manager
- Matt McAuliffe – Technical Manager
- B. Lynn Young – Genomics Lead
- Gretchen Navidi – Principal Analyst
- Kristin Mead – ARRA Outreach
- Eric Stanton – ARRA Web Programming

NDAR Vision

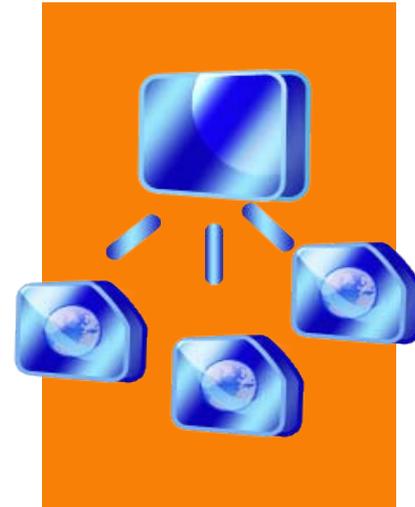
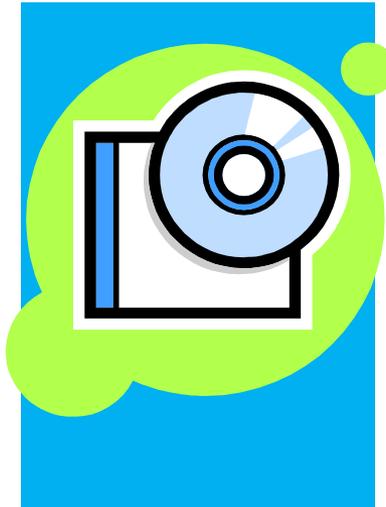
- The NDAR research portal is a secure bioinformatics platform for scientific collaboration and data-sharing for Autism Spectrum Disorder (ASD) investigators
 - Facilitates data sharing and scientific collaboration
 - Provides bioinformatics solutions to address community-wide needs
 - Enables the effective communication of detailed research data, tools, information, and results

NDAR Implementation

- The NDAR research portal offers a single point of access to ASD-relevant research data
- Researchers can query simultaneously:
 - Data stored in the NDAR Central Repository
 - Federated (external) data sources such as Pediatric MRI Data Repository
- Results can include:
 - Phenotypic data – clinical assessments and demographics
 - Imaging data – MRI, fMRI, spectroscopy and volumetric data
 - Genomic data – microarray, expression, SNP, etc.
 - Additional data types such as EEG, MEG, outcomes, etc.
 - Links to appropriate supporting documentation

NDAR Implementation

- NDAR core capabilities link data in a meaningful way
- **Global Unique Identifier (GUID)** – universal subject ID
- **Data Dictionary** – flexible framework for harmonizing data across studies and institutions
- **Federation** – linkage to other relevant data repositories



NDAR Global Unique Identifier (GUID)

- Universal subject ID unique to each participant
- Allows researchers to share data specific to a study participant without exposing personally identifiable information (PII)
- Links multiple types of data collected from a single participant across multiple projects, regardless of where and when that data was collected



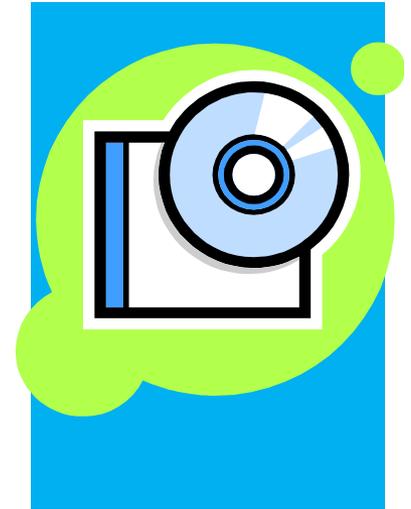
NDAR Global Unique Identifier (GUID)

- Derived from four elements found on a typical birth certificate
- Generated by the research site using software supplied by NDAR
- Double-encryption scheme protects participant confidentiality
- Personally identifiable information remains at the research site
- Approved by NIH Office of General Counsel



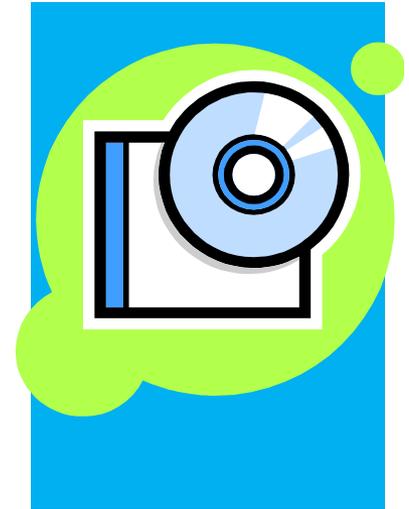
NDAR Data Dictionary

- Flexible and extensible framework for data definition
- Accommodates any data type and data structure
- Alias feature provides linkage between a project's and NDAR data structures
- Extended by the ASD research community
- Curated by NDAR
- Will support ontology integration



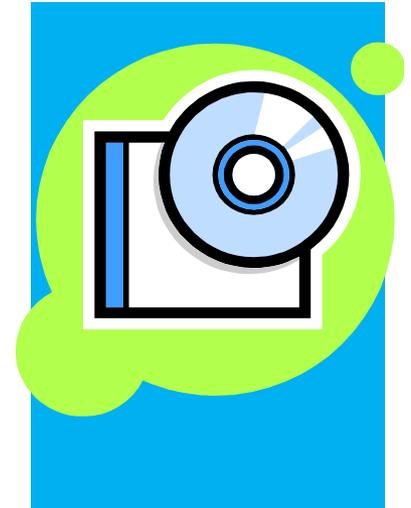
NDAR Data Dictionary

- Works with the NDAR Validation Tool
- Data submitted to NDAR are validated against the Data Dictionary
 - Validation against each data element
 - Errors are reported for rapid resolution
- As data structures are refined:
 - Data Dictionary is updated
 - Validation Tool is updated
 - Updated codebooks are posted to the NDAR public Web site



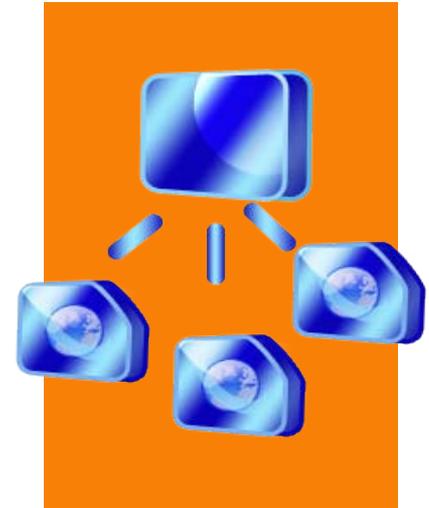
NDAR Data Dictionary

- Brings the scientific community together to establish common data elements
 - NDAR account holders can add keywords and aliases
 - Add comments to a data element and establish a dialog
- Promotes detailed data definition and specificity
- A research community platform for defining and standardizing the complex language that characterizes ASD research



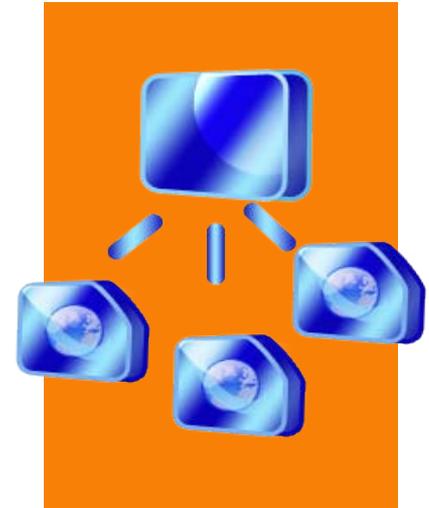
NDAR Federation

- Links the NDAR Central Repository with external data sources
- Provides single query access to all data
- Data elements are aliased through the NDAR Data Dictionary
- Data can be validated with NDAR Validation Tool prior to submission to a federated repository
 - NIMH Genetics Repository
- Supports the data access policies of the federated resource



NDAR Federation

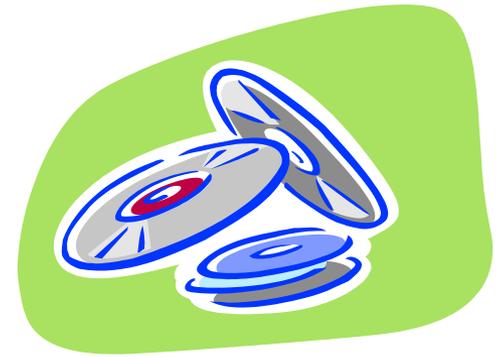
- Federation underway with ASD-relevant repositories
 - Autism Genetic Resource Exchange (AGRE)
 - Interactive Autism Network (IAN)
 - NIMH Genetics Repository
 - NIMH Transcriptional Atlas of Human Brain Development
 - Pediatric MRI Data Repository
- Rich phenotypic, genetic and imaging data from Pediatric MRI Data Repository available Spring 2010



Two-tiered Data Submission

■ Descriptive data

- Includes nonproprietary behavioral and diagnostic data
 - Examples include standard clinical assessments, family medical history, demographics, raw unprocessed images, and genetic data.
- Submitted biannually in January and July
- Making this information available early in the research process allows other investigators to understand the general characteristics of the participants enrolled.



Two-tiered Data Submission

■ Experimental data

- Examples include outcome measures, laboratory results, analyzed images, and volumetric data.
- Submitted 12 months after the primary aims of the research have been achieved, or the time of publication, whichever comes first.
- The submitted data will remain in a private state for an additional 4 months as the investigator performs any final quality control procedures.



Data Submission Timelines

■ First Submission – July 15, 2010

- Expect descriptive data for participants enrolled up until June 1, 2010

	Submission Schedule	Sharing
Descriptive	Semi-annually <ul style="list-style-type: none"> • January 15 • July 15 	4 months after submission
Experimental	within 12 months after each primary aim has been completed or at the time of publication, whichever occurs first.	4 months after submission

Data Submission Planning Process

- Consent participants for data-sharing
- Collect participant demographic data to generate Global Unique Identifiers (GUIDs)
- Prepare data collection systems
- Complete the NDAR Data Submission Agreement
- Request an account on the NDAR research portal
- Create an NDAR Collection
- Generate GUIDs
- Submit data to NDAR



Consent Participants for Data-Sharing

- Investigators must obtain appropriate informed consent from participants whose data will be shared with NDAR
- Sample language is available:
 - For institutions with a certificate of confidentiality
 - For institutions without a certificate of confidentiality
- NDAR data sharing brochure for parents and participants available



Collect Participant Demographic Data

- **Minimum information required:**
 - Full legal name of participant at birth (as it appears on the birth certificate)
 - Participant date of birth
 - Gender
 - Town or municipality of birth (as it appears on the birth certificate)
- **Additional information increases the precision of the creation of the GUID**



Prepare Data Collection Systems

■ NDAR Data Dictionary

■ Data Structures Defined

- ~50 common assessments
- Genomics
- Imaging

■ Data Element Information

- field names
- field types
- descriptions
- range values

■ Provided in Excel spreadsheets

■ Data Dictionary Tool

- Allows for the creations of customized data structures and data elements



NDAR Data Submission Agreement

- Required before data can be submitted
- Specific to a single project
- Must be signed by two parties
 - The Principal Investigator responsible for collecting the data and
 - The NIH-recognized business official at the investigator's affiliated institution
- Reviewed by NDAR Data Access Committee (DAC)
- Once approved, PI will be able to create an NDAR Collection



NDAR Research Portal Account Request

- Required before data can be submitted
- Investigators, Data Managers and Supporting Staff need their own accounts
- Access Levels
 - NDAR portal (PI & Data Manager)
 - GUID Client (Data Manager)
 - Create Collections (PI)
 - Query Shared Autism Data (PI & Data Manager)
- Requests are reviewed by NDAR Data Access Committee (DAC)



NDAR Research Portal Account Privileges

Account Privilege	Provides	Requirement	Personnel
NDAR Portal	Basic Access to NDAR Portal	Portal Account Request	PI, Data Managers, Supporting Staff
GUID Client	Provides the GUID Client to create GUIDs	Software Transfer Agreement (STA)	Data Managers, Supporting Staff
Create Collections	Allows the creation and modification of an NDAR Collection	Data Submission Agreement	PI
Query Shared Data	Allows user to query and retrieve shared data	Data Access Agreement & SF 424 (R&R)	PI, Data Managers

Create an NDAR Collection

- NDAR Collection is a virtual container to hold data submitted to NDAR
- Created by Investigator
- Define Collection
 - Title
 - Administrative Files
 - Investigators
 - Supporting Documentation
 - Grant Number
- Manage Permissions
 - Assign access to other staff members as needed
 - Query – allows user to query data for QA/QC prior to sharing
 - Data Submission – allows user to submit data to Collection
 - Administrator (allows user to manage permissions)



Generate GUIDs

- GUID Client Software
- Specific to a computer
 - Domain Name
 - IP Address and Subnet Mask
- Creates a GUID based on participant data entered into GUID Client Software
 - By individual participant
 - Batch processing
- Note – Personally Identifiable Information is NEVER transmitted from the site to NDAR



Submit Data to NDAR

■ **Data Formatting and Validation**

- Introduction to NDAR data structures and codebooks
- Demonstration of the NDAR Data Validation Tool

■ **Data Submission – Genomics Data**

- Step-by-step procedures for submitting Genomics data
- Demonstration of the NDAR Data Validation Tool

■ **Data Submission – Clinical Assessment Data**

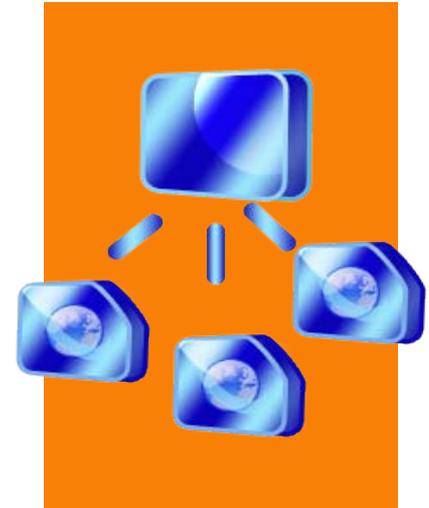
- Step-by-step procedures for submitting Clinical Assessment data
- Demonstration of the NDAR Data Validation Tool

■ **Data Submission – Imaging Data**

- Step-by-step procedures for submitting Imaging data
- Demonstration of the NDAR Data Validation Tool

Data Submission Planning Timeline

- Know your deadlines
 - Data submission biannually (Descriptive Data)
 - January 15
 - July 15
- Understand your tasks
 - Refer to Data Submission Checklist
- Understand the effort required
 - Defines data submission planning tasks
 - Define effort required by staff to complete tasks
- Create a data submission planning timeline
- Monitor your progress



NDAR Technical Assistance

- E-mail: ndar@mail.nih.gov
- Web site: <http://ndar.nih.gov>
 - Standard Operating Procedures (Policy & Procedures page)
 - Webinars (Training page)
 - Online Tutorials (Training page, Spring 2010)
- Listservs
 - NDAR General: System enhancements, Web updates, training schedules, meetings and special events
 - NDAR Data: Updates to Data Dictionary and codebooks