NDAR Webinar | Data Formatting and Validation

June 2010
Agenda

- NDAR Background, Vision and Implementation (10 minutes)
- NDAR Data Formatting and Validation (40 minutes)
- Questions and Answers (10 minutes)
Who is NDAR?

- **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 Coordinator
- 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, 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 Summer 2010
NDAR Data Process

1. Collect
2. Format
3. Validate
4. Submit
5. Share

Researchers

NDAR
NDAR Data Process

Researcher

Format

Collect

Validate

Submit

Share

NDAR
Format Data

- Data Dictionary
  - Standards and conventions for data-sharing within the ASD research community
  - Define, refine, and standardize data definition for autism research

- Data Types Supported
  - Clinical Assessments
  - Imaging
  - Genomics
  - Other
Data Dictionary – Clinical Assessments

- Autism Diagnostic Observation Schedule - Modules 1 - 4 (2001)
- CHARGE Family Characteristics Questionnaire
- CHARGE Physical Exam
- Karyotype
- Modified CHARGE Family Medical
- CHARGE Medical History
- NIMH Medical History Questionnaire
- Vineland II - Survey (2005)
- Vineland II - Parent and Caregiver Rating (2005)
- Aberrant Behavior Checklist — Community
- PDD Behavior Inventory — Parent Rating
- Social Communication Questionnaire — Lifetime
- Social Communication Questionnaire — Current
Data Dictionary – Clinical Assessments (cont’d)

- Social Responsiveness Scale — Current (2005)
- Peabody Picture Vocabulary Test — Form A (Third and Fourth Edition)
- Peabody Picture Vocabulary Test — Form B (Third and Fourth Edition)
- Differential Ability Scales — Preschool
- Differential Ability Scales — School Age
- Wechsler Intelligence Scale for Children — IV (2005)
- Wechsler Abbreviated Scale of Intelligence (1999)
- Wechsler Adult Intelligence Scale — IV

150 measures are currently defined
Data Dictionary – Imaging

- Data Dictionary provides metadata definition for imaging
- Supported submissions
  - Raw brain images in DICOM format
  - Processed images in a variety of formats including DICOM, MINC 1.0 and 2.0, Analyze, NIfTI-1, AFNI and SPM
- MIPAV (Medical Image Processing, Analysis, and Visualization)
  - MIPAV XML format for gathering metadata on images
  - Enables quantitative analysis and visualization of images for numerous modalities such as PET, MRI, CT, or microscopy
  - Contains a module to process images and generate the MIPAV XML required for imaging data submission to NDAR
Data Dictionary – Genomics

- **Supported Submissions**
  - SNP and gene expression microarray
  - Associated descriptions such as pedigree, biological samples, experiment design, experiment samples, reagents, and protocols
  - NDAR does not store any specimen or samples, but may provide reference data for the storage location

- **Genomics Definition**
  - Set of templates for accepting various types of genomics data
    - `bioSamples.xls`
    - `experiments.xls`
    - `experimentSamples.xls`
    - `protocols.xls`
    - `reagents.xls`
    - `subjectsHuman.xls`
Data Structure Definition - Genomics

Genomics Templates on NDAR Web site:
- NG_Biological_Samples.xls
- NG_Experiments.xls
- NG_Experiment_Samples.xls
- NG_Protocols.xls
- NG_Reagents.xls
- NG_Subjects_Human.xls

Genomics Webinar: Wed Jun 16 at 2:00pm ET

Webinar slides:
Data Dictionary – Location

- Data Dictionary Tool – NDAR Portal, accessed through an approved NDAR Account
- Published Data Structures – NDAR Web site: http://ndar.nih.gov/ndarpublicweb/DataStructures.go
## Data Structure – Sample

### Aberrant Behavior Checklist - Community

**Download CSV**

<table>
<thead>
<tr>
<th>ElementName</th>
<th>DataType</th>
<th>Size</th>
<th>Required</th>
<th>ElementDescription</th>
<th>ValueRange</th>
<th>Aliases</th>
</tr>
</thead>
<tbody>
<tr>
<td>subjectkey</td>
<td>GUID</td>
<td>20</td>
<td>Required</td>
<td>The NDAR Global Unique Identifier (GUID) for subjects which identifies a subject in NDAR</td>
<td>NDAR*</td>
<td></td>
</tr>
<tr>
<td>src_subject_id</td>
<td>String</td>
<td>20</td>
<td>Recommended</td>
<td>The site's subject identification</td>
<td></td>
<td>ABC_Comm_SRC_Subject_ID;</td>
</tr>
<tr>
<td>interview_date</td>
<td>DATE</td>
<td></td>
<td>Required</td>
<td>Date on which the interview or test was completed</td>
<td>MM/yyyy,MM/dd/yyyy</td>
<td>ABC_Comm_interview_date;</td>
</tr>
<tr>
<td>interview_age</td>
<td>Integer</td>
<td></td>
<td>Required</td>
<td>age in months at the time of the interview/test.</td>
<td>0 to 1200</td>
<td>ABC_Comm_interview_age;</td>
</tr>
<tr>
<td>comments_misc</td>
<td>String</td>
<td>255</td>
<td>Optional</td>
<td>&quot;Miscellaneous comments on study, interview, methodology relevant to this form data&quot;</td>
<td></td>
<td>ABC_Comm_comments_misc;</td>
</tr>
<tr>
<td>inter_relationparent</td>
<td>String</td>
<td>50</td>
<td>Recommended</td>
<td>Relationship to client - Parent</td>
<td>Yes;No</td>
<td>ABC_Comm_inter_relationparent;</td>
</tr>
<tr>
<td>inter_relationteacher</td>
<td>String</td>
<td>50</td>
<td>Recommended</td>
<td>Relationship to client - Teacher</td>
<td>Yes;No</td>
<td>ABC_Comm_inter_relationteacher;</td>
</tr>
<tr>
<td>inter_relationtrainer</td>
<td>String</td>
<td>50</td>
<td>Recommended</td>
<td>Relationship to client - Trainer/Supervisor</td>
<td>Yes;No</td>
<td>ABC_Comm_inter_relationtrainer;</td>
</tr>
</tbody>
</table>

**Short Name:** abc_community

**Version:** 01
Data Structure Definition – Template

- NDAR Data Structure Definition Template
Data Structure – Components, part 1

- Element Name – name of the data element
- Data Type – data type for the data element
- Size – maximum number of characters if the data type is “string”
- Unit – unit of measure for the data type [OPTIONAL]
- Required – level of requirement for validation
- Element Description – description for the data element
Data Structure – Components, part 2

- Value Range – acceptable data that will pass validation
- Keywords – allows data elements to be searched [OPTIONAL]
- Notes – allows the data element to be annotated [OPTIONAL]
- Section – allows data elements to be organized [OPTIONAL]
  - Especially useful for clinical assessments
NDAR Data Elements – Data Types

- Integer - a whole number (not a fractional number) that can be positive, negative, or zero
- Float – a number that contains a floating decimal point
- String - a string is a contiguous sequence of symbols or values, such as a character string (a sequence of characters) or a binary digit string (a sequence of binary values)
- Date - a date ranging from January 1, 0001 through December 31, 9999
  - Format = MM/yyyy;MM/dd/YYYY
NDAR Data Elements – Data Types

- GUID – contains a prefix, GUID pattern, & check character
  - Prefix = NDAR (Valid GUID) or NDAR_INV (Pseudo GUID)
  - GUID pattern is AANNNAA, where A represents alphabetic and N represents numerical characters
  - Check character can be numerical or alphabetic.

- File – file(s) linked to data being submitted
  - Multiple files can be submitted separated by semi-colon
  - A zip file is allowed

- Accession – can be used to indicate other rows of data, such as Protocols, already uploaded into NDAR
  - Generated by the system
NDAR Data Elements – Required Levels

- **Required Data Elements**
  - GUID
  - Subject ID
  - Interview Date
  - Age

- **Optional Data Elements**
  - Comments
  - Notes

- **Recommended**
  - Item Level Detail
  - Summary Scores
NDAR Data Elements – Value Range

- Value Range – acceptable values for the data element
  - Multiple values must be separated by a semicolon
    - Example = 1;3;7
  - Ranges of values should be separated with the word “to”
    - Example = 0 to 1200
  - Valid value must match the data type of the data element.
    - Example = If the data element is specified as an integer, it is not appropriate to have a value range of 1;3;NK as the integer type accepts only numbers
Data Dictionary – Data Element Names

What If……We have not yet created a data dictionary?

- If the clinical assessment is already defined by NDAR
  - Use the data structure from NDAR

- If the clinical assessment is NOT already defined by NDAR
  - Create your own data structure using NDAR formatting standards
  - Provide your data structure to NDAR to be uploaded into the NDAR data dictionary
Data Dictionary – Data Element Names

What If…….We have already created a data structure and our data element names are different from NDAR?

- Rename your data elements
  OR
- Create a Mapping File

NDAR Data Structure Mapping Instructions:
http://ndar.nih.gov/ndarpublicweb/Documents/NDAR_Data_Structure_Mapping_Instructions.ppt
Data Dictionary – Mapping File

- Maps locally used data element names to data element names used by NDAR
  - Structure Name – Data Structure short name and version number
  - Alias – data element name used locally
  - NDAR – data element name used by NDAR

```xml
<?xml version="1.0" ?>
<map_file>
  <map_structure name="abc_community01">
    <map_element alias="ID" ndar="src_subject_id" />
    <map_element alias="test_date" ndar="interview_date" />
    <map_element alias="subject_age" ndar="interview_age" />
    <map_element alias="notes" ndar="comments_misc" />
  </map_structure>
</map_file>
```
NDAR Data Process

1. Collect
2. Format
3. Validate
4. Submit
5. Share

Researchers send data to NDAR for validation and formatting.
### Data Set - Sample

<table>
<thead>
<tr>
<th>SubjectKey</th>
<th>SRC_SubjectID</th>
<th>Charge_medicalhist_interview_date</th>
<th>Charge_medicalhist_comments_m misc</th>
<th>Charge_medicalhist_diag_a_ofc</th>
<th>Charge_medicalhist_diag_a_ofcnck</th>
<th>Charge_medicalhist_diag_a_concern</th>
<th>Charge_medicalhist_diag_a_diagnosis</th>
<th>Charge_medicalhist_diag_a_geatdiag</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDARDEMOZY465AYN</td>
<td>6410000</td>
<td>11/29/2008</td>
<td>Take home</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDARDEMOZY069TN6</td>
<td>6480002</td>
<td>11/29/2008</td>
<td>Completed</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDARDEMOZV577NGG</td>
<td>6550005</td>
<td>11/29/2008</td>
<td>Take home</td>
<td>60</td>
<td>Deafness</td>
<td>Downs Syi</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>NDARDEMOZR719MCB</td>
<td>6410000</td>
<td>11/29/2008</td>
<td>Completed</td>
<td>13</td>
<td>Self Stimu</td>
<td>Aspergers</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>NDARDEMOZP984JFH</td>
<td>6480002</td>
<td>11/29/2008</td>
<td>Take home</td>
<td>36</td>
<td>Completed</td>
<td>Delayed S</td>
<td>Aspergers</td>
<td>36</td>
</tr>
<tr>
<td>NDARDEMOZM905KH3</td>
<td>6550005</td>
<td>11/29/2008</td>
<td>Completed</td>
<td>60</td>
<td>Deafness</td>
<td>FragileX</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>NDARDEMOZH938JT4</td>
<td>6410000</td>
<td>11/29/2008</td>
<td>Take home</td>
<td>13</td>
<td>Self Stimu</td>
<td>Down Syndrome</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>NDARDEMOZC697MZN</td>
<td>6550005</td>
<td>11/29/2008</td>
<td>Take home</td>
<td>60</td>
<td>Deafness</td>
<td>Aspergers</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>NDARDEMOZC077NEY</td>
<td>6410000</td>
<td>11/29/2008</td>
<td>Completed</td>
<td>13</td>
<td>Self Stimu</td>
<td>FragileX</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>NDARDEMOYZ244TJ5</td>
<td>6480002</td>
<td>11/29/2008</td>
<td>Take home</td>
<td>36</td>
<td>Completed</td>
<td>Delayed S</td>
<td>Down Syndrome</td>
<td>36</td>
</tr>
<tr>
<td>NDARDEMOYW853RUH</td>
<td>6550005</td>
<td>11/29/2008</td>
<td>Completed</td>
<td>60</td>
<td>Deafness</td>
<td>PDD</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>NDARDEMOYV776FMG</td>
<td>6410000</td>
<td>11/29/2008</td>
<td>Take home</td>
<td>13</td>
<td>Self Stimu</td>
<td>Aspergers</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>
NDAR Data Process

Researcher → Format → Collect → Validate → Submit → Share → NDAR
Validation Tool - Overview

- Verifies that submitted data conform to the required format and range values defined in the NDAR Data Dictionary
- Imports the NDAR Data Dictionary and validates the metadata associated with the files identified by the NDAR user for submission against the data dictionary
- Provides a report of any data discrepancies and warnings
- Creates a submission package
- Uses a Java Web Start application that runs locally on a user's computer
  - Requires the installation of the Java runtime environment
Validation Tool - Preparation

- Convert raw, locally maintained data to supported file type
- File Types Supported
  - Comma Separated Value
  - Tab Delimited
  - XML(CDISC)
- Create New Folder
- Save all files for submission in one folder
Validation Tool - Steps

- Open Validation Tool from NDAR Portal
- Browse to folder that contains data files
- Click ‘Load Files” button
- Exclude files as necessary
- Validate Files
- Fix Errors
  - Errors – Prevent user from continuing with process; Must be fixed
  - Warnings – Do not prevent user from continuing with process
- Validate again, if necessary
- Create Submission Package
Validation Tool – Access Tool on Portal

NDAR Query Tool
The NDAR Query Tool provides a way for the research community to query into the shared data environment for data discovery, processing, and analysis.

NDAR Data Validation Tool
The NDAR Data Validation Tool is an application that validates data to be submitted to NDAR. The tool also creates the submission package (2 files) that can be submitted. Only the submission package created by this tool can be submitted to NDAR.

Version 3.0
February 23, 2010
Validation Tool – Access Tool on Portal

Welcome to the National Database for Autism Research (NDAR) Portal

NDAR Query Tool
The NDAR Query Tool provides a way for the research community to query, processing, and analysis.

NDAR Data Validation Tool
The NDAR Data Validation Tool is an application that validates data to be submitted to NDAR. The tool also creates the submission package (2 files) that can be submitted. Only the submission package created by this tool can be submitted to NDAR.
Validation Tool – Access Tool on Portal
Validation Tool – Browse to Data File Folder
## Validation Tool – Load Files

### Files Loaded

<table>
<thead>
<tr>
<th>File Name</th>
<th>File Type</th>
<th>Structure Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC_Community.xml</td>
<td>UNKNOWN</td>
<td>abc_community01</td>
</tr>
<tr>
<td>Karyotype.csv</td>
<td>CSV FILE</td>
<td>karyotype01</td>
</tr>
<tr>
<td>Karyotype.xml</td>
<td>CDISC FILE</td>
<td>karyotype01</td>
</tr>
<tr>
<td>ADOS4_2001.csv</td>
<td>CSV FILE</td>
<td>ados4_200101</td>
</tr>
<tr>
<td>Karyotype.txt</td>
<td>TAB FILE</td>
<td>karyotype01</td>
</tr>
<tr>
<td>Charge_Physical_Exam.txt</td>
<td>TAB FILE</td>
<td>charge_physicalexam01</td>
</tr>
</tbody>
</table>

### Result Details

<table>
<thead>
<tr>
<th>Detail Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Validation Tool – Exclude Files

NDAR Data Validation [v3.1.1.328]

Utilities

Data Source

Please specify the directory containing the files you would like to validate for submission to NDAR:

Browse  YYARRA Demo Data.

Files Loaded

<table>
<thead>
<tr>
<th>File Name</th>
<th>Structure Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC_Community.xml</td>
<td>Multiple Structures</td>
</tr>
<tr>
<td>Karyotype.csv</td>
<td>ados4_200101</td>
</tr>
<tr>
<td>Karyotype.xml</td>
<td>karyotype01</td>
</tr>
<tr>
<td>ADOS4_2001.csv</td>
<td>charge_physical_exam01</td>
</tr>
<tr>
<td>Karyotype.txt</td>
<td></td>
</tr>
<tr>
<td>Charge_Physical_Exam.txt</td>
<td></td>
</tr>
</tbody>
</table>

Validation Results

<table>
<thead>
<tr>
<th>Result</th>
<th>Summary</th>
</tr>
</thead>
</table>

Exclude Selected Files  Include Selected Files  Mapping Details  Validate Files

Result Details

<table>
<thead>
<tr>
<th>Detail Type</th>
<th>Description</th>
</tr>
</thead>
</table>

Validation Tool – Validate Files

<table>
<thead>
<tr>
<th>File Name</th>
<th>File Type</th>
<th>Structure Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC_Community.xml</td>
<td>UNKNOWN</td>
<td>charge_physicsexam01</td>
</tr>
<tr>
<td>Karyotype.csv</td>
<td>CSV FILE</td>
<td>CDISC FILE</td>
</tr>
<tr>
<td>Karyotype.xml</td>
<td>CSV FILE</td>
<td>TAB FILE</td>
</tr>
<tr>
<td>ADOS1_2001.csv</td>
<td>CSV FILE</td>
<td>charge_physicsexam01</td>
</tr>
<tr>
<td>Karyotype.txt</td>
<td>TAB FILE</td>
<td>charge_physicsexam01</td>
</tr>
<tr>
<td>Charge_Physical_Exam.txt</td>
<td>TAB FILE</td>
<td>charge_physicsexam01</td>
</tr>
</tbody>
</table>

**Validation Results**

<table>
<thead>
<tr>
<th></th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Result Details**

<table>
<thead>
<tr>
<th>Detail Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Validate Files** 15%

Validating Files

Progress: 14%

Cancel
# Validation Tool – Validation Output with Errors

**NDAR Data Validation [v3.1.1.32b]**

### Data Source

Please specify the directory containing the files you would like to validate for submission to NDAR:

<table>
<thead>
<tr>
<th>Browse</th>
<th>MARRA Demo Data</th>
<th>Load Files</th>
</tr>
</thead>
</table>

### Files Loaded

<table>
<thead>
<tr>
<th>File Name</th>
<th>File Type</th>
<th>Structure Names</th>
<th>Result</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC_Community.xml</td>
<td>UNKNOWN</td>
<td>Multiple Structures</td>
<td>Failed</td>
<td>[3 errors + 1190 warnings]</td>
</tr>
<tr>
<td>Karyotype.csv</td>
<td>CSV FILE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karyotype.xml</td>
<td>CSV FILE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADOS4_2001.csv</td>
<td>CSV FILE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karyotype.txt</td>
<td>TAB FILE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge_Physical_Exam.txt</td>
<td>TAB FILE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Result Details

<table>
<thead>
<tr>
<th>Detail Type</th>
<th>Description</th>
</tr>
</thead>
</table>

### Validation Results

**Validation results**

```
Meta-data file(s) contain error(s)
```

**OK**
Validation Tool – Read and Fix Errors

<table>
<thead>
<tr>
<th>Files Loaded</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>File Name</td>
<td>File Type</td>
</tr>
<tr>
<td>ABC_Community.xml</td>
<td>UNKOWN</td>
</tr>
<tr>
<td>Karyotype.csv</td>
<td>CSV FILE</td>
</tr>
<tr>
<td>Karyotype1.csv</td>
<td>CSV FILE</td>
</tr>
<tr>
<td>Karyotype.txt</td>
<td>TAB FILE</td>
</tr>
<tr>
<td>Charge_Typical_Example</td>
<td>TAB FILE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation Results</th>
<th>Result</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Failed</td>
<td>[3 errors + 190 warnings]</td>
</tr>
<tr>
<td></td>
<td>Valid</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Failed</td>
<td>[3 errors + 190 warnings]</td>
</tr>
</tbody>
</table>

**Result Details**

<table>
<thead>
<tr>
<th>Detail Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>Public: src_subject_of Data Element is missing and required</td>
</tr>
<tr>
<td>Error</td>
<td>Public: Meta-data field &quot;head_accnt&quot; has an invalid value of &quot;Missing&quot; with an expected value range of &quot;Yes|No|Missing|None&quot;</td>
</tr>
<tr>
<td>Warning</td>
<td>No rules exist for Data Element &quot;Charge_physiological_eyes_irispattern&quot;</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;height_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_rst&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_mkt&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_vmt&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>No rules exist for Data Element &quot;Charge_physiological_eyes_irispattern&quot;</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;height_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_rst&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_mkt&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_vmt&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>No rules exist for Data Element &quot;Charge_physiological_eyes_irispattern&quot;</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;height_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_rst&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_mkt&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_vmt&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
</tbody>
</table>
## Validation Tool – Read and Fix Errors

### Result Details

<table>
<thead>
<tr>
<th>Detail Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>Rules: src_subject_id Data Element is missing and required</td>
</tr>
<tr>
<td>Error</td>
<td>Rules: Meta-data field &quot;head_occiput&quot; has an invalid value of &quot;Nrmal&quot; with an expected value range of &quot;Flat;Normal;Prominent;NE&quot;</td>
</tr>
<tr>
<td>Error</td>
<td>Rules: Meta-data field &quot;skin_hair&quot; has an invalid value of &quot;Course&quot; with an expected value range of &quot;Fine;Coarse;Thick;Fragile;Kinky;Curly&quot;</td>
</tr>
<tr>
<td>Warning</td>
<td>No rule exists for Data Element Charge_physicalexam_eyes_irispattern</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;height_mett&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;height_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_mett&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;weight_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>No rule exists for Data Element Charge_physicalexam_eyes_irispattern</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;height_mett&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
<tr>
<td>Warning</td>
<td>You did not provide an entry for &quot;height_std&quot; element. NDAR best practices recommend that you provide a value for this field.</td>
</tr>
</tbody>
</table>
Validation Tool – Read and Fix Errors

1. Rules: src_subject_id Data Element is missing and required
2. Rules: Meta-data field "head_occiput" has an invalid value of "Nrmal" with an expected value range of "Flat;Normal;Prominent;NE"
3. Rules: Meta-data field "skin_hair" has an invalid value of "Course" with an expected value range of "Fine;Coarse;Thick;Fragile;Kinky;Curly"
Validation Tool – Read and Fix Errors

1. Rules: src_subject_id Data Element is missing and required
2. Rules: Meta-data field "head_occiput" has an invalid value of "Nrmal" with an expected value range of "Flat;Normal;Prominent;NE"
3. Rules: Meta-data field "skin_hair" has an invalid value of "Course" with an expected value range of "Fine;Coarse;Thick;Fragile;Kinky;Curly"
1. Rules: src_subject_id Data Element is missing and required

2. Rules: Meta-data field "head_occiput" has an invalid value of "Nrmal" with an expected value range of "Flat;Normal;Prominent;NE"

3. Rules: Meta-data field "skin_hair" has an invalid value of "Course" with an expected value range of "Fine;Coarse;Thick;Fragile;Kinky;Curly"
Validation Tool – Validate Again
Validation Tool – Validation Output with Warnings
Validation Tool – Create Submission Package
Validation Tool – Create Submission Package
Validation Tool – Submission Package Created
Validation Tool – Submission Package

- Submission Package process creates two files
  - Data for Submission
  - Ticket file for authentication of data validation
NDAR Data Process

- Format
- Collect
- Validate
- Submit
- Share

Researcher → NDAR
Webinars - Submit Data to NDAR

- **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
NDAR Technical Assistance

- **E-mail:** [NDARhelp@mail.nih.gov](mailto:NDARhelp@mail.nih.gov)
- **Web site:** [http://ndar.nih.gov](http://ndar.nih.gov)
  - Standard Operating Procedures (Policy & Procedures page)
  - Webinars (Training page)
  - Online Tutorials (Training page, Summer 2010)
- **Listservs:** [http://ndar.nih.gov/ndarpublicweb/Listserv.go](http://ndar.nih.gov/ndarpublicweb/Listserv.go)
  - NDAR General: System enhancements, Web updates, training schedules, meetings and special events
  - NDAR Data: Timely technical updates, system releases, new data structures, data management-related tools and trainings
- **Demonstration Environment:** [http://ndardemo.nih.gov](http://ndardemo.nih.gov)
  - Request an Account
  - Practice setting up data structures and submitting data