National Database for Autism Research

NDAR Data Validation and Submission User Guide

7/14/2008

Revision

Version	Date	Description
1.5.4	7/14/2008	Original

Table of contents

Introduction

The NDAR portal hosts a collaborative environment offering tools for the greater autism research community, enabling researchers to store and share clinical assessment (CA), medical imaging, and genomics data using the NDAR Central Repository to provide cooperation on multi-institutional investigations.

The NDAR Central Repository requires users to upload both data and metadata. Metadata consist of information that describes data and are used to provide documentation for data products. Metadata is a way of characterizing what is in a file without actually needing to open the data files.

CA submission files should be configured in Clinical Data Interchange Standards Consortium (CDISC) XML format using the data definitions described in the NDAR Clinical Assessments and Imaging Codebook. Imaging submission files are generated using the NDAR Image Submission Package Creation tool within the MIPAV {http://mipav.cit.nih.gov/} application. Metadata for genomics data are contained within a set of MS Excel spreadsheets.[1]

For more information about creating metadata, refer to Appendices 1−3 on page 12.
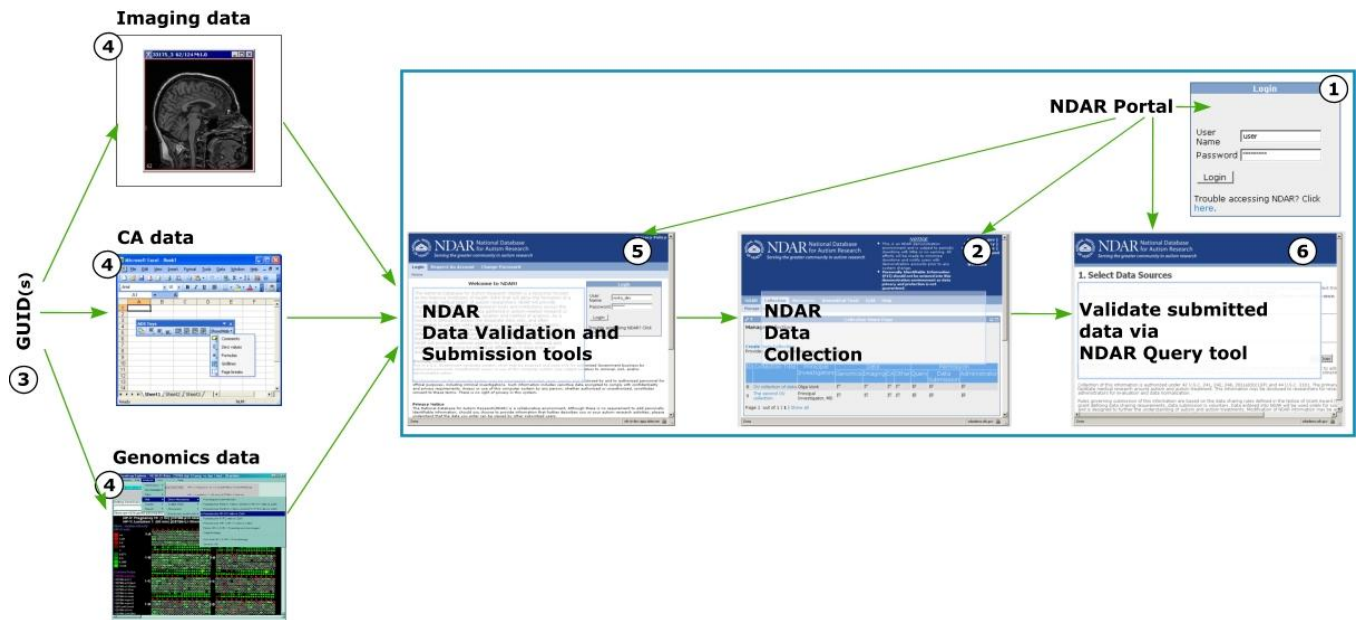

Figure 1. Uploading data to NDAR

---

[1] CDISC models can be reviewed at http://www.cdisc.org/publications.html.

Data Validation and Submission Process Overview

1   Before submitting data to NDAR, obtain a user account to the NDAR portal. In order to do that, connect to the NDAR portal {https://ndarportal.nih.gov}. Then, open the Request an Account tab and follow the instructions provided by the tab.

2   On the NDAR portal, create your own private data collection(s), which is a reserved space on the NDAR Central Repository where you will store your data. In order to do that, follow the instructions provided in the *Creating a Data Collection* Section of this document.

3   For all your data intended for submission, obtain Global Unique Identifiers (GUID).

4   Prepare data and metadata for submission as described in Appendices 1−3.

5   Run the NDAR Data Validation tool to validate your metadata and create the submission package.

6   Submit your data to NDAR by uploading the submission package to the appropriate collection.

7   Run the NDAR Query tool to query data stored in the NDAR Central Repository once your submission has been processed and loaded by NDAR. See Figure 1.

The NDAR public website has additional information with respect to the data submission process, refer to {http://ndar.nih.gov/ndarpublicweb/datasubmission.go}.

Creating a Data Collection

Prior to submitting data to NDAR, you must create a data collection where your research data will be stored. This section outlines the steps necessary for creating a data collection.

   Before creating the collection, read, complete, and sign the NDAR Central Repository Intuitional Certification to Submit form.

To create a data collection:

1   Log in to the NDAR portal.

2   Navigate to the Collection tab.

3   In the Collection tab, click on the Create New Collection link.

4   Fill out the Create Collection form. Note that fields with asterisk (*) are the required fields.

5 Upload the signed Institutional Certification to Submit (ICS) file. In order to do that,

  ▪ First, click Add;

  ▪ Then, use the Browse button to open the file.

6 Press Create.

The new collection appears on the portal in your account. To view the collection, click on the Manage Collection link. See Figure 2.
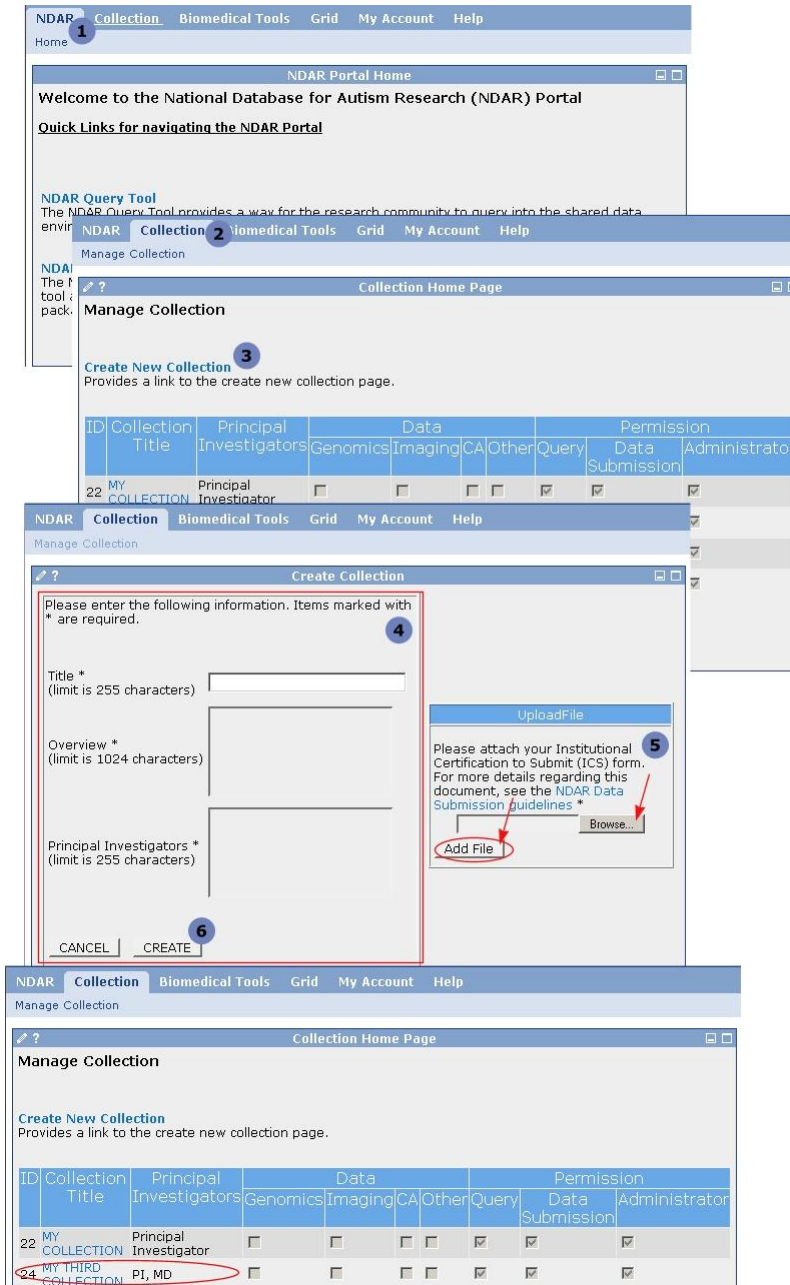

Figure 2. Creating a new data collection

NDAR Data Validation

This section of the document provides help for the NDAR Data Validation tool, which is designed to assist NDAR users with the submission of data into the NDAR Central Repository.

Before submitting data, the metadata associated with the data must be validated for accuracy of information using the NDAR Data Validation tool. The tool validates metadata using the NDAR codebooks to ensure that all supported data conform to NDAR standards.

Prerequisites

Before you run the NDAR Data Validation tool, the following conditions must be met:

- Java 1.5 must be installed on your PC.
- All identifying information (e.g. names, Social Security Numbers, etc.) must be removed from the data.
- Required dataset meta-information and NDAR GUIDs for each dataset should be obtained and written into the metadata file(s) associated with your data file.
- You must have an NDAR Portal account.

Running the NDAR Data Validation tool

To run the tool,

1 Go to the NDAR Portal, and then navigate to the NDAR Data Validation tool link.
2 Click on the link, the NDAR Data Validation tool starts running, and the NDAR Data Validation tool dialog box appears.

In the dialog box,

3 Specify the directory where all of the metadata files you wish to validate are located. You can either use the Browse button or enter the whole path to the directory. See Figure 3.
Note that, if your metadata files are stored in different subdirectories of the same directory, you only need to specify the upper level directory and the tool will recursively browse and validate metadata in all subdirectories.
4 Click Validate Files.
5 The validation process starts and the following information appears in the Validation Results window:
- The metadata file name including the directory name;

- The file type and the form name;
- The validation result – Valid or Failed;
- The validation details that show how many validation errors appear in the file that failed validation and also how many data files are attached to the metadata file that pass validation. See Figure 3.
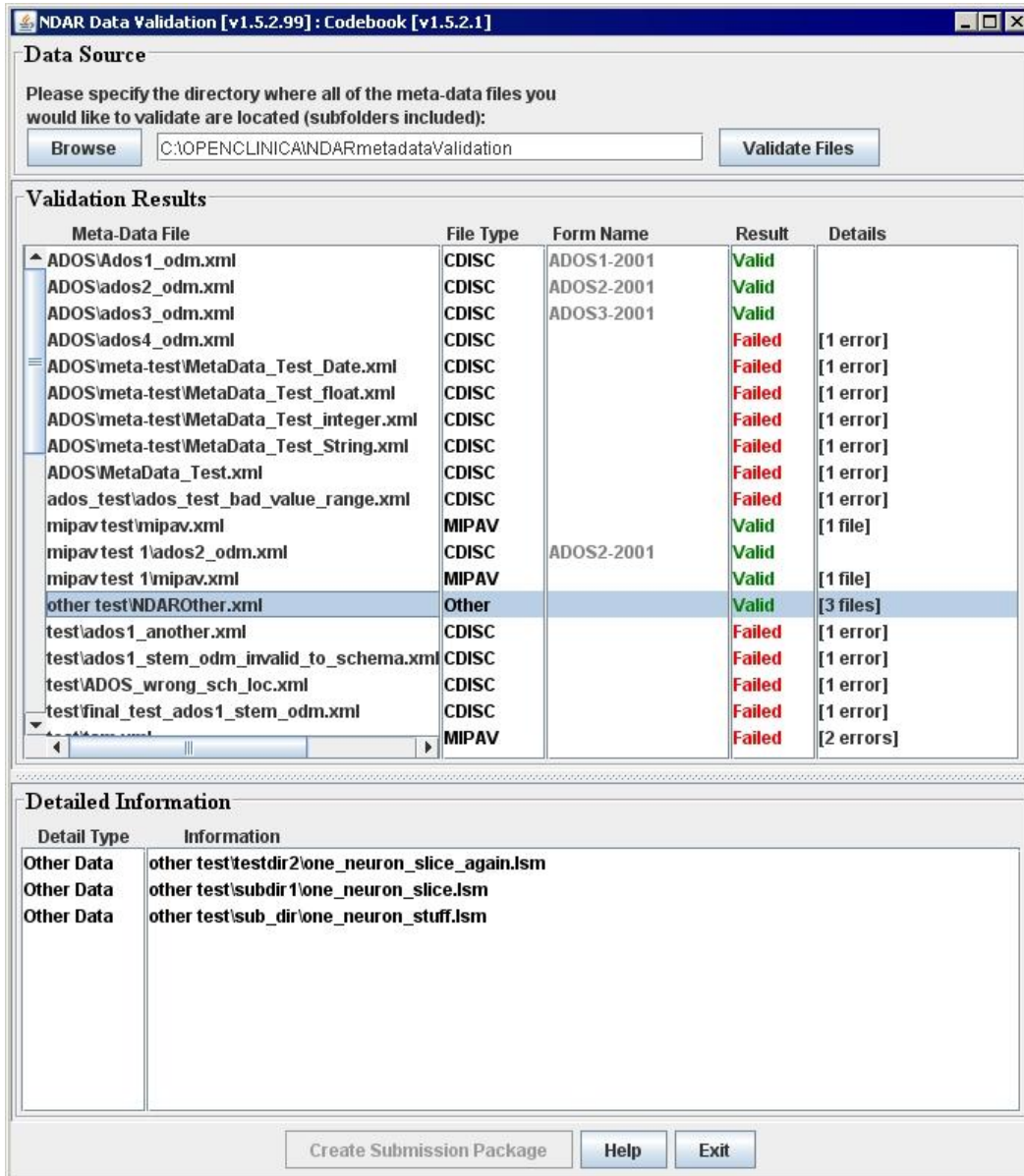

Figure 3. The NDAR Data Validation dialog box

Detailed information

To view more detail, such as data file names and/or explanations of validation errors, highlight the file of interest in the Validation Results window and navigate to the

Detailed Information window. For a given metadata file, the following information appears in the Detailed Information window:

- If the metadata file has errors, for each of them, the Error keyword appears in the Detail Type column and the explanation of error appears in the Information column. See Figure 4.
- If the metadata file has data files associated with it and has no errors, the list of data files appears in the Information column. See Figure 5.
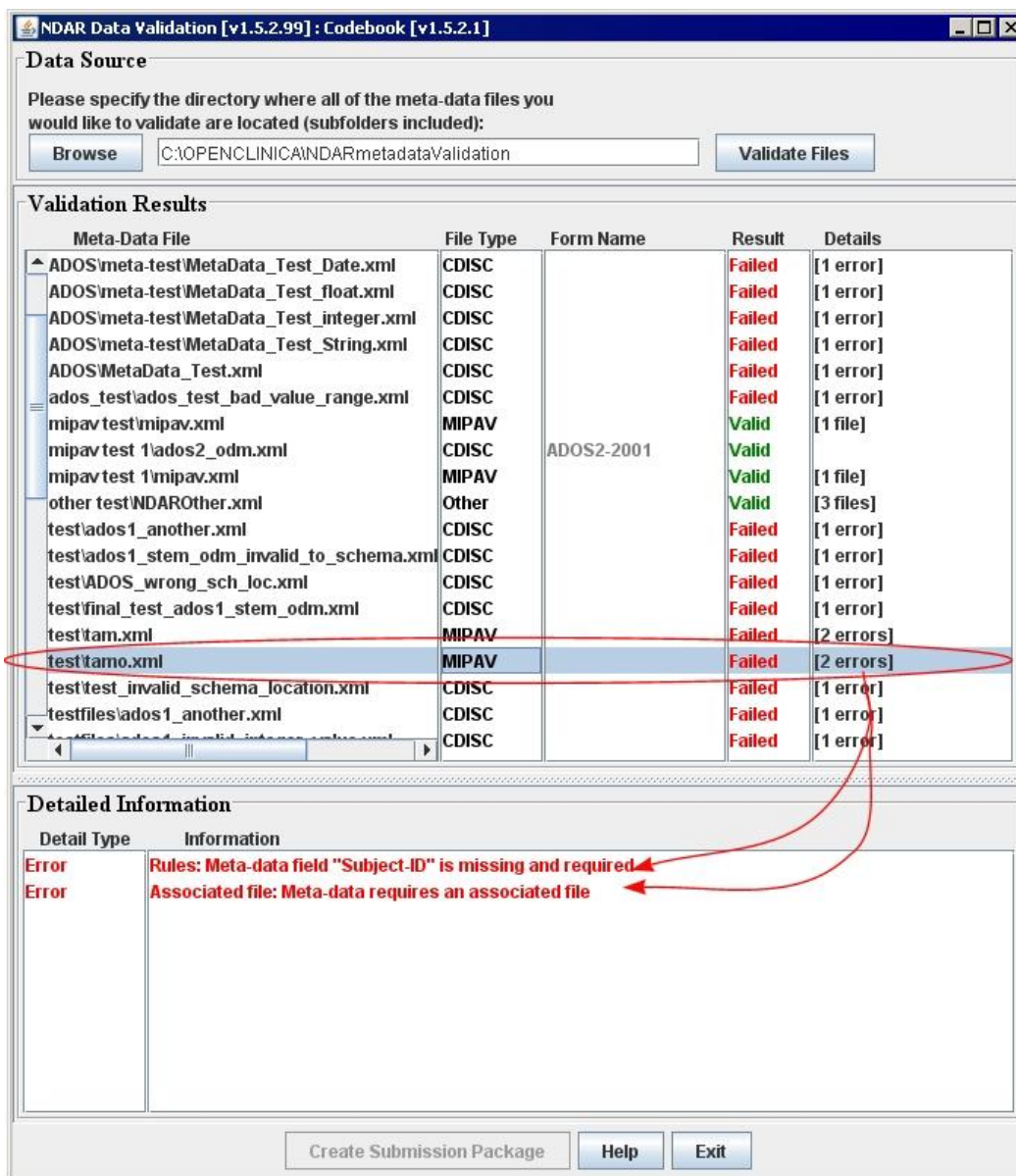

Figure 4. This file has validation errors, thus the information about them appears in the Detailed Information window
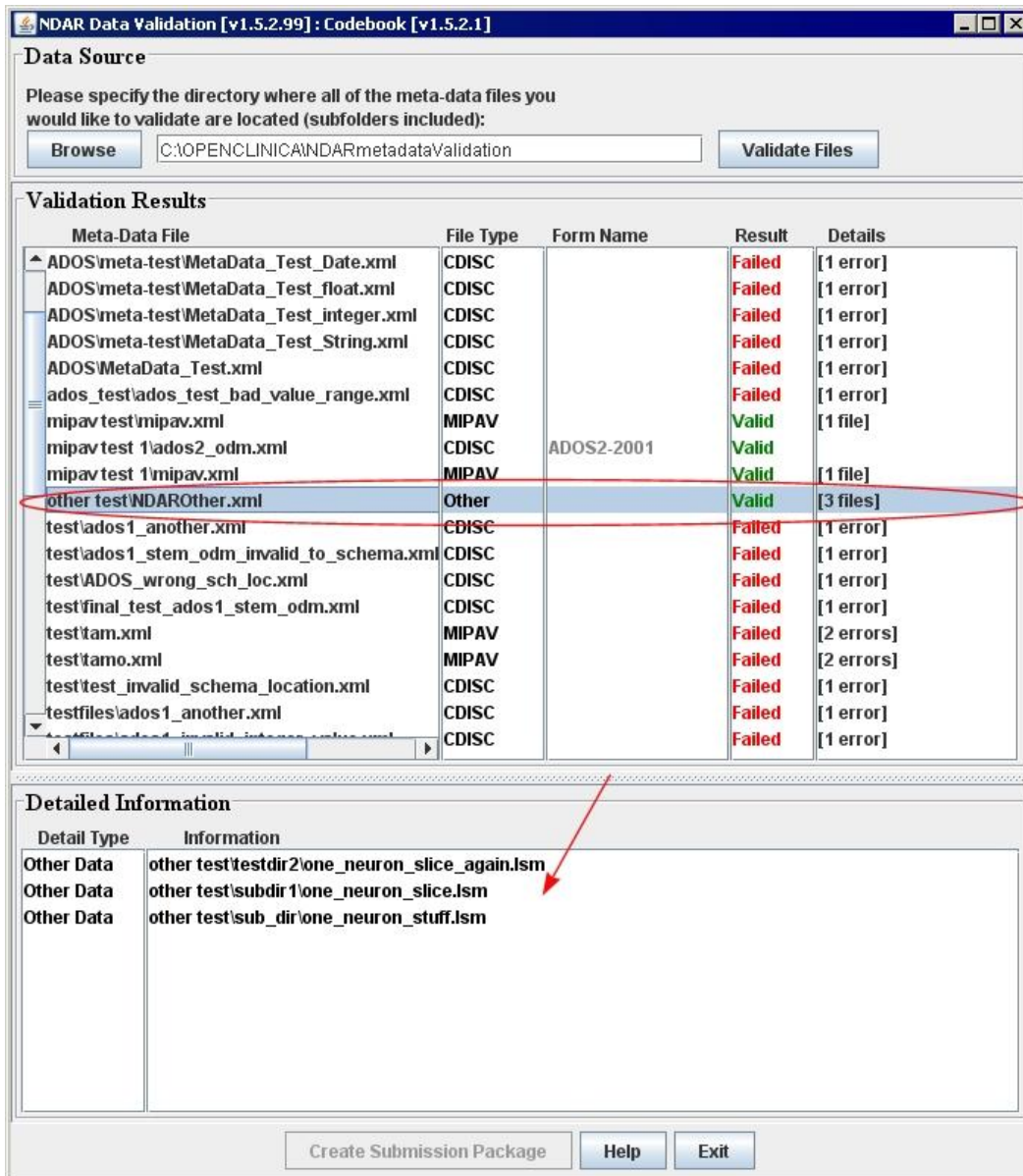
Figure 5. This metadata file has no validation errors. In the Detailed Information window, the list of associated data files appears

Creating the submission package

The Create Submission Package button becomes active only if all metadata files in the directory you provided pass validation. If your metadata pass validation, the message "All metadata files are valid" will appear in the Information window and the word Valid will appear for all files in the Result column.

To create a submission package:

1   Click the Create Submission Package button.

2 In the dialog that appears, specify the directory where you wish the submission package to be saved, and click OK.

3 The tool starts running and the Information window appears displaying the paths to the submission package and submission ticket. Click OK to complete the process.

The submission package (a ZIP archive) and submission ticket (an XML file) appear in the designated directory. See Figure 6.
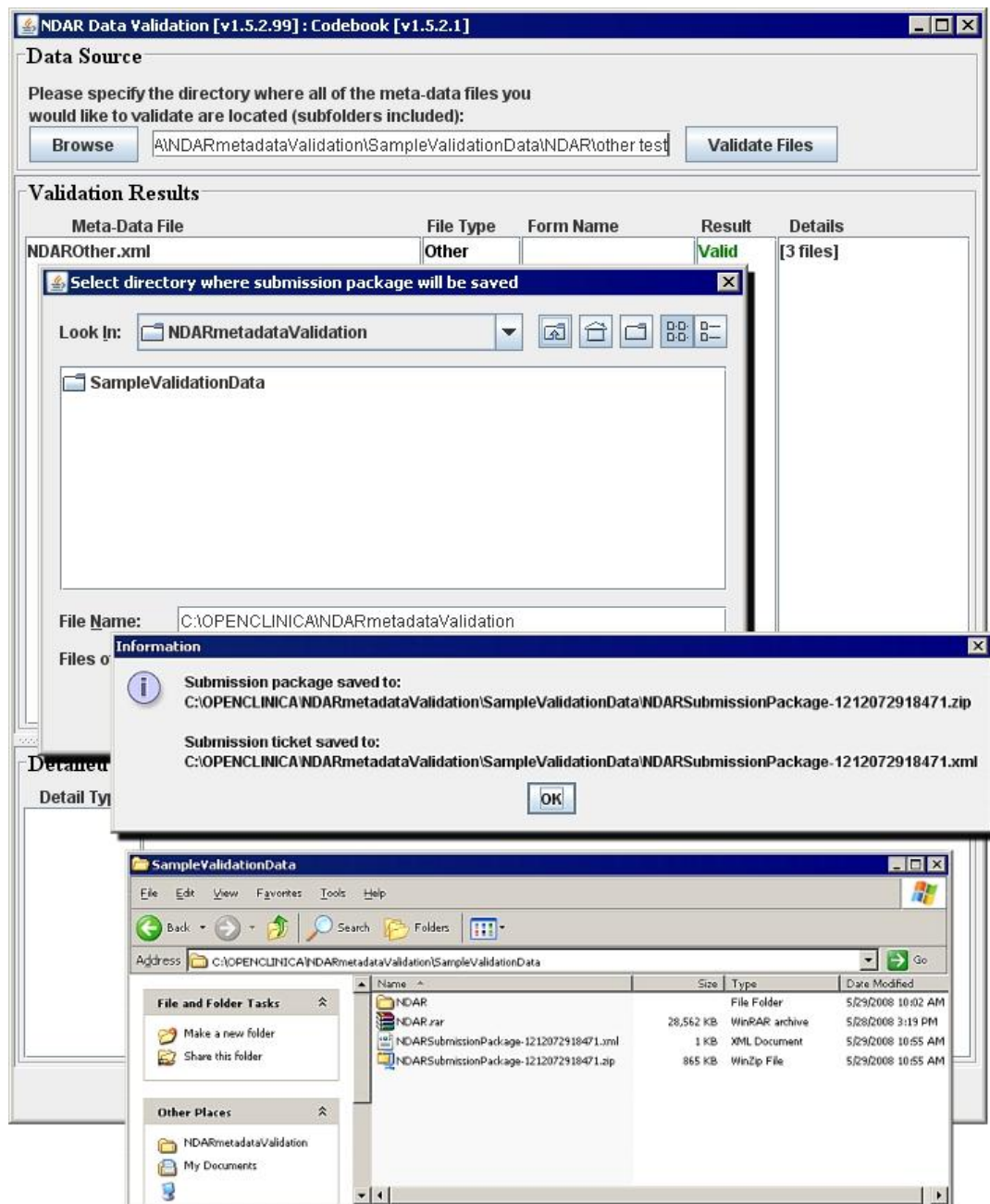
Figure 6. Creating a submission package

Submission package and submission ticket

The submission package is a ZIP archive that contains both the data and metadata files from the validated directory. The submission ticket is an XML file that serves as the upload instruction for NDAR. It lists all files included in the submission package and also includes security information to ensure that the submission package has not been modified or corrupted after passing validation.

Figure 7. The submission ticket

Both the submission package and submission ticket are required for submission using the data submission link for the appropriate collection.

NDAR Data Submission

The NDAR portal allows submission of data to the NDAR data repository via a link for each collection defined in NDAR.

Prerequisites for data submission

To submit data to NDAR, the following is required:

- The user's data collection(s) must be created on the NDAR portal.
- The user's account must have been granted Data Submission privileges to the collection by the Principal Investigator.
- A submission package ZIP file and its corresponding XML ticket file. Please refer to the "NDAR Data Validation" section of this document for more details.[2]

Submitting data to a collection

To submit your data to NDAR, do the following (see also Figure 8):

1   On the NDAR portal, navigate to the Collection tab.

2   Select the collection, and then click on the collection's name. The Collection Home page appears.

3   In the Collection Home page, click the View Submissions button. The second Collection Home Page appears.

4   Click the Upload Data Submission Package link. The Submit Data to NDAR page appears.

5   Read the NDAR submission agreement and then check "I agree" box and also press the "I agree" button.

6   Select and upload the submission ticket file (an XML file). Select and upload the submission package ZIP file. Note that both files must have the same name.

If submission was successful the following message appears: "The submission package has been successfully uploaded".

Once submitted, dataset IDs for the data types submitted will appear on the Collection Home page. The submitted data will also be retrievable using the NDAR Query tool.

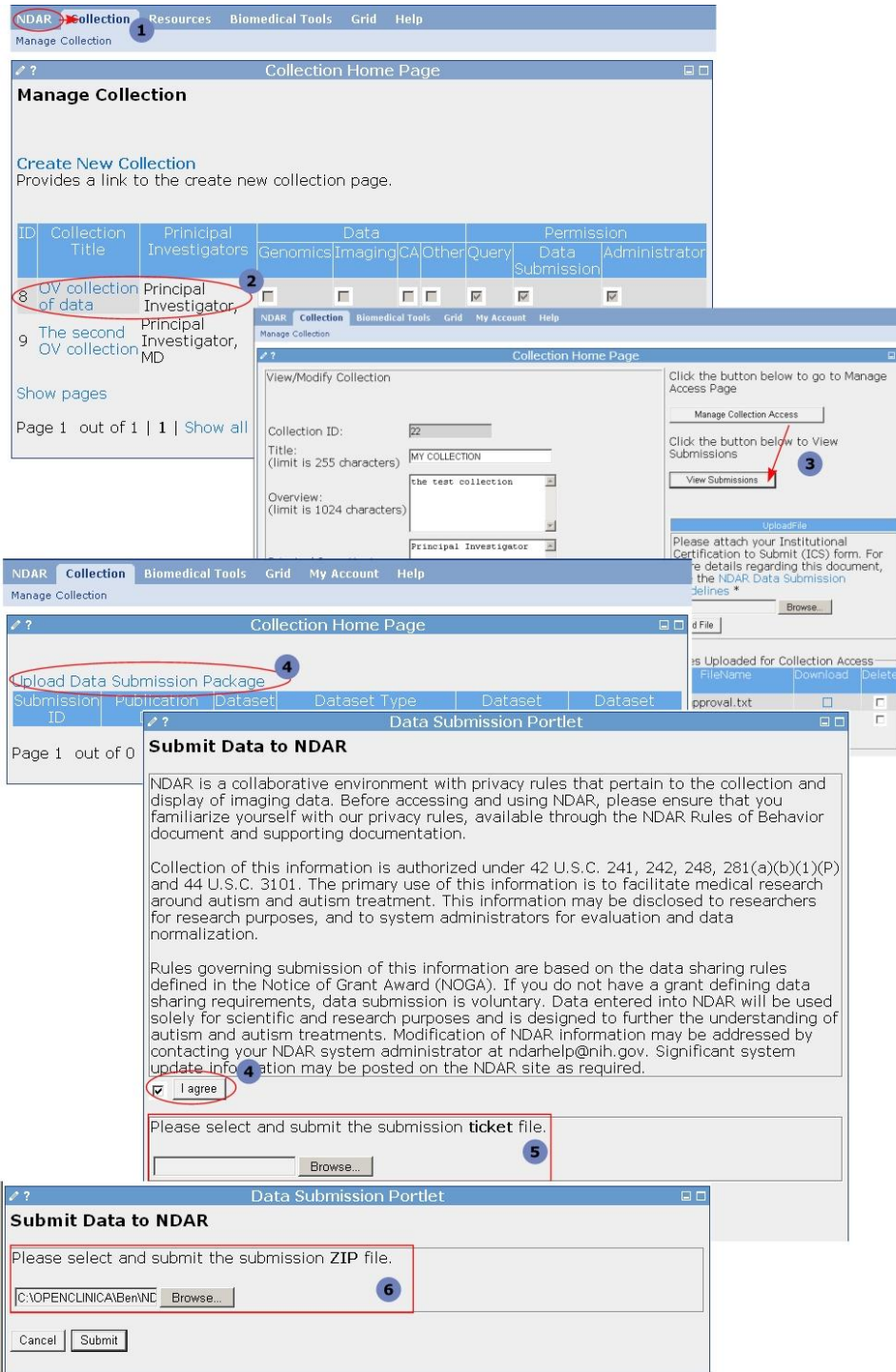[2] The Genomics Codebook is currently under development.

Figure 8. Submitting data to NDAR

Appendix 1: Creating Clinical Assesment metadata

NDAR has adopted the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM) XML format which all submitted clinical assessments metadata must adhere to in order to be loaded in the NDAR Central Repository. In order to help the users with using that format, the NDAR Clinical Assessments & Imaging Codebook has been established to identify the autism related instruments and forms supported by NDAR. The codebook defines the NDAR field names that correspond to the instruments and forms data entries and can be queried by the NDAR Query tool.

When submitting data to the NDAR Central Repository, the data field names must match the field names in the corresponding form in the codebook. Fields marked as "Required" must be included in the data being submitted. The Data Validation tool will display an error if required fields are missing, and the file will fail validation.

Fields marked as "Recommended" are recommended for submission, but are not required. The Data Validation tool will display a warning when a "Recommended" field is missing. However, the file will still pass validation if the recommended field is missing.

"Optional" fields may be left out if the data is not available.

Sample CDISC formatted clinical assessments metadata can be viewed in the NDAR Clinical Assessments and Imaging Codebook in the CDISC XML Sample File spreadsheet.

This appendix covers creating the metadata for CA data using the NDAR Clinical Assessment and Imaging Codebook and provided XML template.

1 From the NDAR portal, download the NDAR Clinical Assessments & Imaging Codebook. The Clinical Assessment XML template can be found in the codebook spreadsheet named CA CDISC XML Template File, see Figure 9.

2 Open the codebook and navigate to the spreadsheet that describes the form for which you would like to create metadata, see Figure 10.

3 Then, navigate to the CA CDISC XML Template File spreadsheet. Create an XML file replicating the spreadsheet content to that file. See Figure 11.

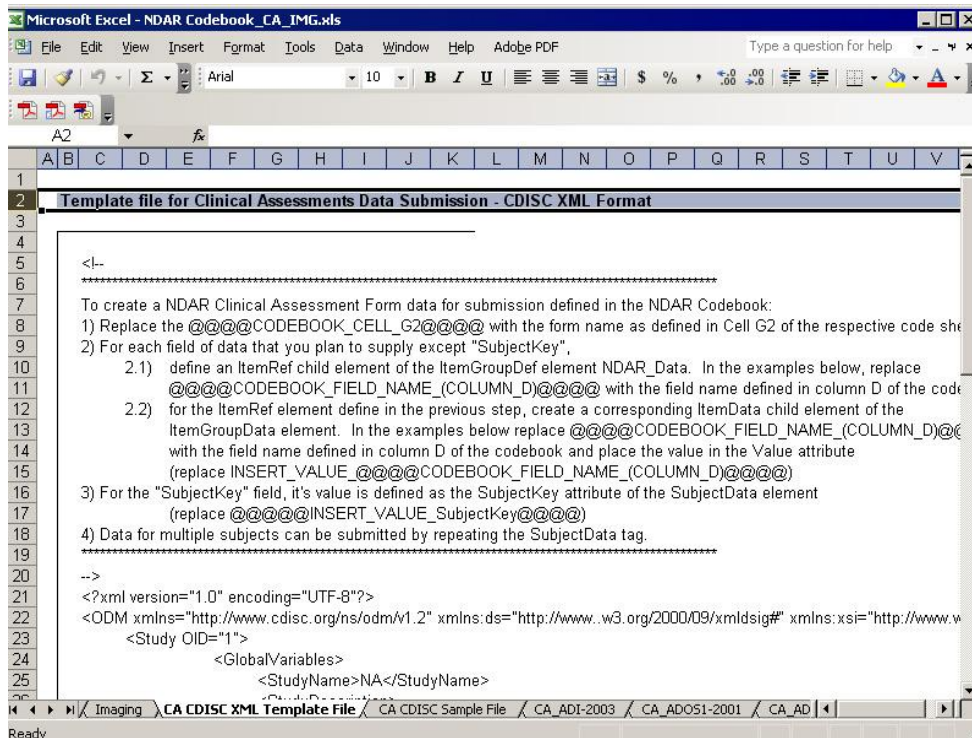4 Follow the steps outlined in the template file spreadsheet to create the dataset, see Figure 11—Figure 13.

Figure 9. The CA CDISC XML Template spreadsheet with instructions

Figure 10. The NDAR Clinical Assessments & Imaging Codebook

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.2"
xmlns:ds="http://www..w3.org/2000/09/xmldsig#"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2
http://www.cdisc.org/models/odm/v1.2.1/ODM1-2-1-foundation.xsd"
Description="ADOS1_CRF" FileType="Snapshot" FileOID="5" CreationDateTime="2008-01-
21T17:10:06">
<Study OID="1">
<GlobalVariables>
<StudyName>NA</StudyName>
<StudyDescription>N/A</StudyDescription>
<ProtocolName>N/A</ProtocolName>
</GlobalVariables>
<MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0">
<Protocol>
<StudyEventRef StudyEventOID="SE.1" Mandatory="No" />
</Protocol>
<StudyEventDef OID="SE.1" Name="NDAR" Repeating="No" Type="Scheduled">
<FormRef FormOID="NDAR" Mandatory="No" />
</StudyEventDef>
<FormDef OID="FORM.1" Name="@@@@CODEBOOK_CELL_G2@@@@" Repeating="No">
<ItemGroupRef ItemGroupOID="IG.1" Mandatory="No" />
</FormDef>
<ItemGroupDef OID="IG.1" Name="NDAR_Data" Repeating="No"
SASDatasetName="FormData">
<ItemRef ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
Mandatory="No" />
<ItemRef ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
Mandatory="No" />
<ItemRef ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
Mandatory="No" />
</ItemGroupDef>
</MetaDataVersion>
</Study>
<ClinicalData StudyOID="1" MetaDataVersionOID="v1.0.0">
<SubjectData SubjectKey="@@@@@INSERT_VALUE_SubjectKey@@@@">
<StudyEventData StudyEventOID="SE.1">
<FormData FormOID="FORM.1">
<ItemGroupData ItemGroupOID="IG.1" TransactionType="Insert">
<ItemData ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
Value="INSERT_VALUE_@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@" />
<ItemData ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
Value="INSERT_VALUE_@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@" />
<ItemData ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
Value="INSERT_VALUE_@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@" />
</ItemGroupData>
</FormData>
</StudyEventData>
</SubjectData>
</ClinicalData>
</ODM>
```

Figure 11. The CA CDISC XML Template

For example, to create the metadata for the "ADOS - Autism Diagnostic Observation Schedule - General Module 1 - Pre-Verbal/Single Words (2001 or later)" form:

1   Open the NDAR Clinical Assessments & Imaging Codebook and navigate to the CA_ADOS1-2001 spreadsheet.

2   Open the CA CDISC XML Template.

3   In the CA_ADOS1-2001 spreadsheet, copy the form name to the Clipboard. The form name is defined in the cell G2 (column G row 2 in the spreadsheet). See Figure 12.

4   Then, go to the CA CDISC XML Template and go to the FormDef OID tag and replace the @@@@CODEBOOK_CELL_G2@@@@ with the form name as it was defined in the cell G2. See Figure 12.

5   In the CA CDISC XML Template for each field of data that you plan to supply (except SubjectKey), define an ItemRef ItemOID child element of the ItemGroupDef OID element. In order to do that,
    ▪ In the CA_ADOS1-2001 spreadsheet, navigate to the column D and copy the field name to the Clipboard;
    ▪ Then, return to the CA CDISC XML Template and replace @@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@ with the field name from the Clipboard.
    ▪ Repeat these steps for as many field names as you need.

6   For each ItemRef element defined in the previous step 6, create the corresponding child element ItemData ItemOID of the ItemGroupData element.

7   In the CA CDISC XML Template, navigate to the <ItemGroupData ItemGroupOID="IG.1" TransactionType="Insert"> line and replace @@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@ with the field name defined in column D of the codebook.

8   Then, enter the value in to the Value=" tag. In order to do that replace INSERT_VALUE_@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@ with the value from your form. See Figure 12.

9   The SubjectKey field value (NDAR GUID) from the CA_ADOS1-2001 spreadsheet must be defined in the SubjectData SubjectKey tag. In order to do that, copy the NDAR GUID received for the subject; then navigate to CA CDISC XML Template to the SubjectData SubjectKey tag, replace

@@@@@INSERT_VALUE_SubjectKey@@@@ with the value you copied to the Clipboard. Data for multiple subjects can be submitted by repeating the SubjectData tag. CA CDISC Sample File spreadsheet in the codebook has sample CA ADOS1-2001 for multiple subjects.

10   Save the metadata XML file.

11   Login to the NDAR portal and run the NDAR Data Validation tool.

12   Validate the metadata XML file (created in steps) using the NDAR Data Validation tool.

13   Resolve errors, if any and run the validation tool again.

14   If validation was successful, submit your CA data to your collection on the NDAR Portal.
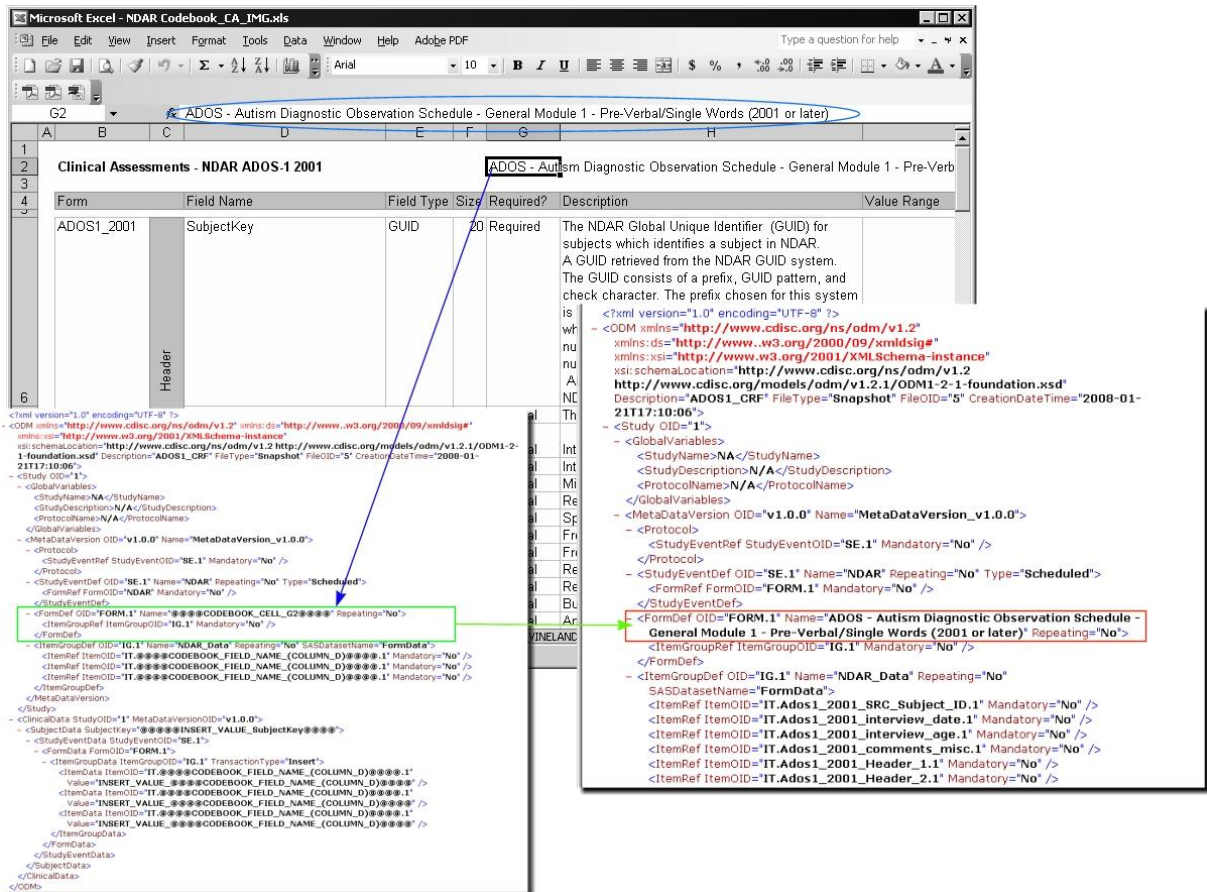
Figure 12. In the CA_ADOS1-2001 spreadsheet, copy the form name to the Clipboard. The form name is defined in the cell G2. Then, go to the CA CDISC XML Template and go to the FormDef OID tag and replace the @@@@CODEBOOK_CELL_G2@@@@ with the form name as it was defined in the cell G2

Figure 13. In the CA CDISC XML Template, for each field of data that you plan to supply (except "SubjectKe y"), define an ItemRef ItemOID child element of ItemGroupDef OID: in the CA_ADOS1-2001 spreadsheet, navigate to the column D and copy the field name to the Clipboard; then, return to the CA CDISC XML Template and replace @@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@ with the field name from the Clipboard. Repeat these steps for as many field names as you need

ADOS - Autism Diagnostic Observation Schedule - General Module 1 - Pre-Verbal/Single Words (2001 or later)

C	D	E	F	G	H	
Coding G					Stereotyped/Idiosyncratic Use of Words or	
	Ados1_2001_codinga_stereo	integer		Optional	Phrases	0;1;2;3;8
	Ados1_2001_codinga_uothr	integer		Optional	Use of Other's Body to Communicate	0;1;2;8
	Ados1_2001_codinga_point	integer		Optional	Pointing	0;1;2;3
	Ados1_2001_codinga_gest	integer		Optional	Gestures	0;1;2;8
	Ados1_2001_codingb_ueye	integer		Optional	Unusual Eye Contact	0;2
	Ados1					0;1;2;3
	Ados1				thers	0;1;2
	Ados1				haviors during	
	Ados1					0;1;2;3
	Ados1					0;1;2
	Ados1					0;1;2;3
	Ados1					0;1;2;3
	Ados1					0;1;2
	Ados1					0;1;2
	Ados1				tention	0;1;2
	Ados1					0;1;2;3
	Ados1					0;1;2;3
	Ados1					0;1;2;3
	Ados1					0;1;2;3

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.2" xmlns:ds="http://www.w3.org/2000/09/xmldsig#"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2 http://www.cdisc.org/ns/models/odm/v1.2.1/ODM1-2-
  1-foundation.xsd" Description="ADOS1_CRF" FileType="Snapshot" FileOID="5" CreationDateTime="2008-01-
  21T17:10:06">
  <Study OID="1">
    <GlobalVariables>
      <StudyName>NA</StudyName>
      <StudyDescription>N/A</StudyDescription>
      <ProtocolName>N/A</ProtocolName>
    </GlobalVariables>
    <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0">
      <Protocol>
        <StudyEventRef StudyEventOID="SE.1" Mandatory="No" />
      </Protocol>
      <StudyEventDef OID="SE.1" Name="NDAR" Repeating="No" Type="Scheduled">
        <FormRef FormOID="NDAR" Mandatory="No" />
      </StudyEventDef>
      <FormDef OID="FORM.1" Name="@@@@CODEBOOK_CELL_G2@@@@" Repeating="No">
        <ItemGroupRef ItemGroupOID="IG.1" Mandatory="No" />
      </FormDef>
      <ItemGroupDef OID="IG.1" Name="NDAR_Data" Repeating="No" SASDatasetName="FormData">
        <ItemRef ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1" Mandatory="No" />
        <ItemRef ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1" Mandatory="No" />
        <ItemRef ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1" Mandatory="No" />
      </ItemGroupDef>
    </MetaDataVersion>
  </Study>
  <ClinicalData StudyOID="1" MetaDataVersionOID="v1.0.0">
    <SubjectData SubjectKey="@@@@INSERT_VALUE_SubjectKey@@@@">
      <StudyEventData StudyEventOID="SE.1">
        <FormData FormOID="FORM.1">
          <ItemGroupData ItemGroupOID="IG.1" TransactionType="Insert">
            <ItemData ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
              Value="INSERT_VALUE_@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@" />
            <ItemData ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
              Value="INSERT_VALUE_@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@" />
            <ItemData ItemOID="IT.@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@.1"
              Value="INSERT_VALUE_@@@@CODEBOOK_FIELD_NAME_(COLUMN_D)@@@@" />
          </ItemGroupData>
        </FormData>
      </StudyEventData>
    </SubjectData>
  </ClinicalData>
</ODM>
```

```xml
<ClinicalData StudyOID="1" MetaDataVersionOID="v1.0.0">
  <SubjectData SubjectKey="NDARVN966EA6">
    <StudyEventData StudyEventOID="SE.1">
      <FormData FormOID="FORM.1">
        <ItemGroupData ItemGroupOID="IG.1" TransactionType="Insert">
          <ItemData ItemOID="IT.Ados1_2001_Header_1.1" Value="John" />
          <ItemData ItemOID="IT.Ados1_2001_Header_2.1" Value="Parent" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_fvoc.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_gest.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_iecho.1" Value="2" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_inton.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_OLANG.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_point.1" Value="3" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_stereo.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingA_uothr.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingB_faceo.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingB_give.1" Value="1" />
          <ItemData ItemOID="IT.Ados1_2001_CodingB_gzsov.1" Value="0" />
          <ItemData ItemOID="IT.Ados1_2001_CodingB_qsov.1" Value="2" />
          <ItemData ItemOID="IT.Ados1_2001_CodingB_req.1" Value="2" />
          <ItemData ItemOID="IT.Ados1_2001_CodingB_rjnt.1" Value="1" />
```

Figure 14. Entering the value into the Value= tag

Appendix 2: Using Image Submission Package Creation tool to create imaging metadata

NDAR has adopted the MIPAV XML format for gathering metadata on images to be submitted to NDAR. The MIPAV (Medical Image Processing, Analysis, and Visualization) application enables quantitative analysis and visualization of medical images of numerous modalities such as PET, MRI, CT, or microscopy. The MIPAV application contains a module that will process images and generate the XML formatted files required for imaging data submission to NDAR. The NDAR Clinical Assessments & Imaging Codebook defines the field names and acceptable values for image metadata which the MIPAV XML data file(s) must adhere to. The codebook also has a sample imaging MIPAV XML file for reference. This appendix provides the help for the NDAR Image Submission Package Creation tool, which is designed to assist uses in creating imaging metadata.

Prerequisites

Before you run the NDAR Image Submission Package Creation tool, the following conditions must be met:

- MIPAV version 4.1.0 or later must be installed on your computer, refer to {http://mipav.cit.nih.gov/}.
- All indentifying information must be removed from the image (i.e., research participant metadata in the image header).
- The image file format must be one which can be read by MIPAV, for example: Analyze, AFNI, DICOM 3.0, MINC, MIPAV XML, RAW, TIFF, among many others. For more information about file formats supported by MIPAV, refer to the MIPAV User Manual Volume 1, APPENDIX C: Supported formats, refer to {http://mipav.cit.nih.gov/documentation.php}.
- Required dataset meta-information and NDAR GUIDs for each image dataset, must be obtained.

Running the NDAR Image Submission Package Creation tool

To run the tool, run MIPAV. Then, call the File > NDAR > NDAR Image Submission Package Creation menu. The NDAR Image Submission Package Creation tool dialog box appears. The dialog has four tabs: Main, Images, GUID and Log.

Main tab

1 In the Main tab of the dialog box, read the statement, and then check the "I agree to the above statement" box. Refer to page 24 for the MIPAV privacy statement.

2 The Next button becomes available; press it to proceed to the Images tab.
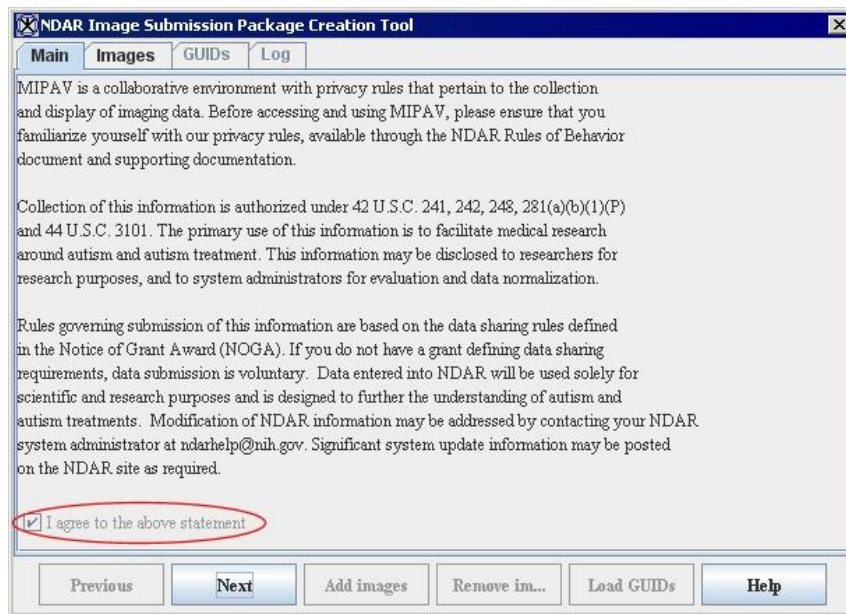
Figure 15. The Image Submission Package Creation Tool dialog box - Main tab

Images tab

3 In the Images tab, use the Add Images button to select image datasets from a disk. Use the SHIFT+left mouse button combination of keys to select multiple images. If adding DICOM datasets, in the Open dialog box, check the Open as Multifile option, and then select the first image from the dataset. The application will automatically add all files from the dataset. See Figure 16.

4 When you finish adding images, press Next. In the confirmation window that appears, press Yes to confirm adding datasets. The GUIDs tab appears.
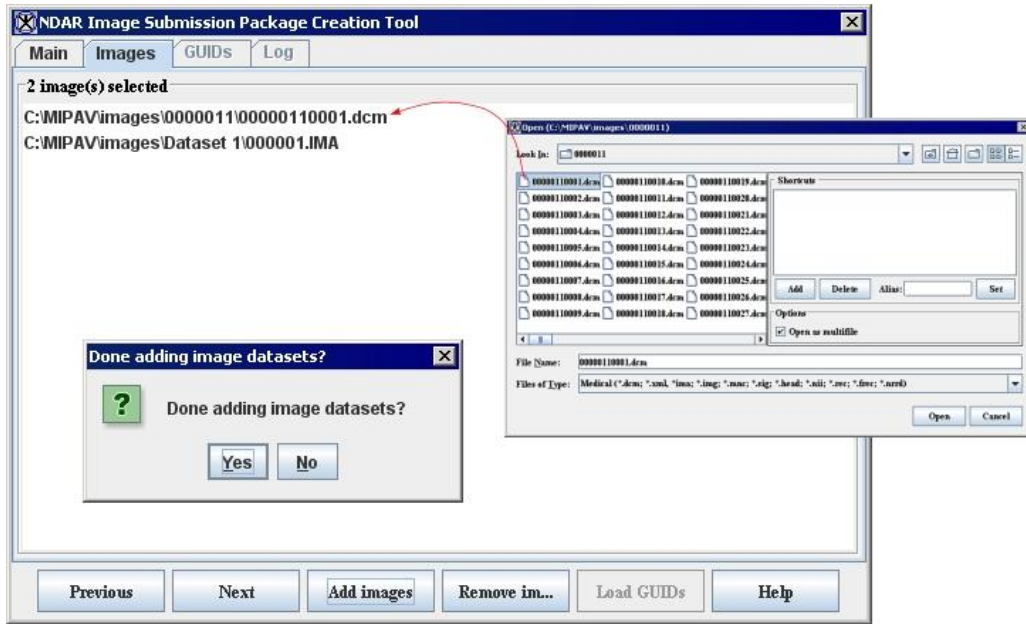
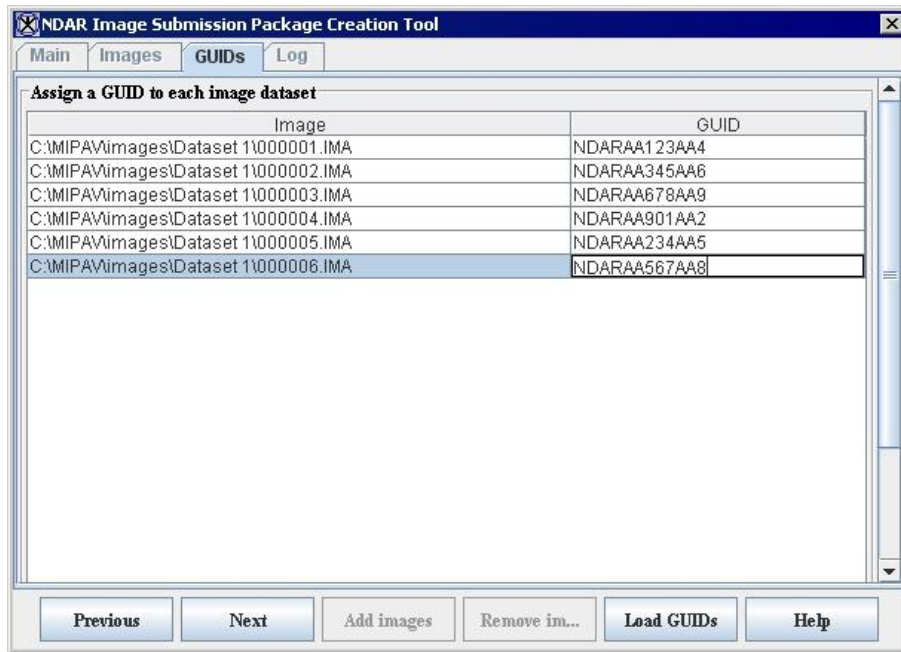Figure 16 The Image Submission Package Creation Tool dialog box - Images tab

GUIDs tab

Figure 17. The Image Submission Package Creation Tool dialog box - GUIDs tab

5   In the GUIDs tab that appears next, assign a GUID to each dataset. You can either do
that manually or use the Load GUIDs option to load GUIDs from a file. Note that if

the image dataset has a valid NDAR GUID in its name, the software will recognize it, and it will appear in the GUID textbox automatically.

6   Press Next. For each image dataset, the software will create a submission package (a ZIP file) and submission ticket and store them in the following directory:

```
C:\Documents and Settings\username\mipav\NDAR_Imaging_Submission
```

Log tab

If the submission package creation was successful, the following information appears in the Log tab:

▪ The full path to the directory where the submission packages are stored

▪ The list of files added to the submission packages

▪ The statement "Submission package processing complete"
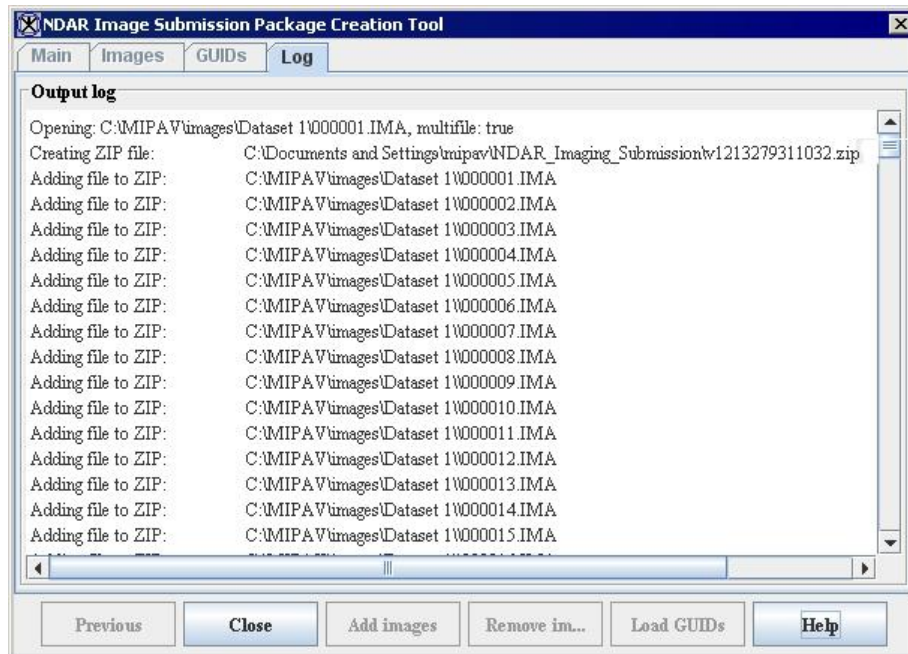Press Close to close the application.

Figure 18. The Image Submission Package Creation Tool dialog box - Log tab

Output of the NDAR Image Submission Package Creation tool

For each dataset selected for submission, the tool reads in the dataset from a disk and creates two files:

▪ A ZIP file that contains the image data;

- An XML file containing the NDAR GUID in the Subject ID tag and basic dataset metadata such as image dimensionality, resolution, type, and others from the original image. See Figure 19.

By default, both files are stored in the

`\mipav\NDAR_Imaging_Submission` subdirectory of user directory, for example,

`C:\Documents and Settings\username\mipav\NDAR_Imaging_Submission`

For both files the file names are as follows:

```
username_time stamp.zip
username_time stamp_image name.xml
```


Figure 19. The MIPAV XML file that contains the NDAR GUID in the Subject ID tag and basic dataset metadata

Submitting data to NDAR

The output file of the NDAR Image Submission Package Creation tool is the accepted format for NDAR imaging metadata and data submission. Launch the NDAR Data Validation tool (refer to page 4 for more information) and use it to validate the imaging metadata. Then, navigate to the appropriate collection to upload your data to NDAR (refer to page 10 for more information).

MIPAV Privacy Statement

- MIPAV is a collaborative environment with privacy rules that pertain to the collection and display of imaging data. Before accessing and using MIPAV, please ensure that you familiarize yourself with our privacy rules, available through the NDAR Rules of Behavior document and supporting documentation.

- Collection of this information is authorized under 42 U.S.C. 241, 242, 248, 281(a)(b)(1)(P) and 44 U.S.C. 3101. The primary use of this information is to facilitate medical research around autism and autism treatment. This information may be disclosed to researchers for research purposes, and to system administrators for evaluation and data normalization.

- Rules governing submission of this information are based on the data sharing rules defined in the Notice of Grant Award (NOGA). If you do not have a grant defining data sharing requirements, data submission is voluntary. Data entered into NDAR will be used solely for scientific and research purposes and is designed to further the understanding of autism and autism treatments. Modification of NDAR information may be addressed by contacting your NDAR system administrator at ndarhelp@nih.gov. Significant system update information may be posted on the NDAR site as required.

Appendix 3: Using NDAR Genomics Templates to submit genomics information

This appendix covers submission of genomics data in Tab-delimited text format. In order to describe your data, you should fill out the NDAR Genomics Templates (ngTemplates), which are available for download from the NDAR web site. This involves providing information about your experiment using a spreadsheet program such as Microsoft Excel, because these transfer templates are .xls files.

There are six templates and a results file available for providing of the experiment information:[3]

subjectsHuman.xls – used to provide information about research subjects including GUID(s);

experiments.xls – used to provide information about the experiment such as hypothesis and protocols;

bioSamples.xls – used to provide information about the biological samples used in the experiment, such as tissue, blood, cells, DNA, protein, and any others;

reagents.xls – used to provide information about key reagents used in the experiment such as manufacturers and catalog numbers;

protocols.xls – used to provide information about the protocols used in the experiment such as sample extraction and scanning;

experimentSamples.xls – used to provide information about the experiment samples such as names of files containing measurements;

results file or folder - a file or folder with your experiment results. This file is not necessary in MS Excel format, because the raw data formats depend on the technology. In experimentSamples.xls, a single experiment sample ID is linked to the results file or folder, providing the information for each submitted microarray. See Figure 20 and Figure 27.

---

[3] http://ndar.nih.gov/ndarpublicweb/datarepository.go
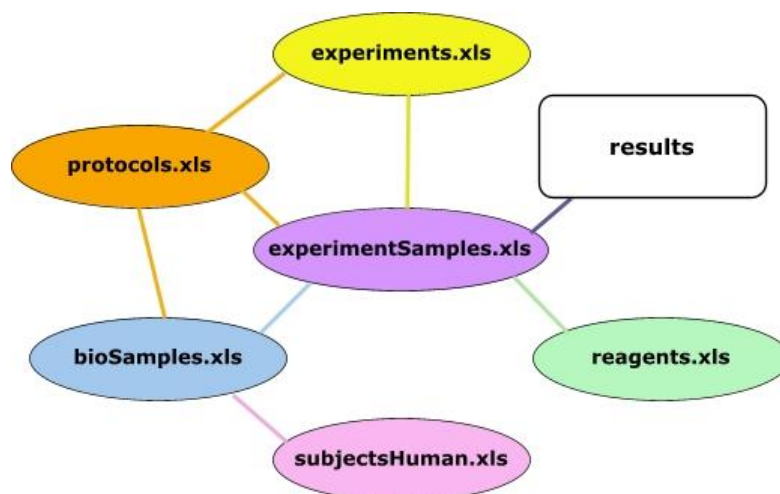
Figure 20. This diagram illustrates the relationships between experiment components that may be included in the experiment data submission package

For user convenience, the document includes screenshots taken from each template. *Note that the templates may change from one version of NDAR to another, so do not be discouraged if actual templates look slightly different from screenshots presented in this document.*

All ngTemplates possess similar features such as:
- You should not edit or delete the very first column in each template.
- Each template has a white area and a shaded or grey area. You must enter metadata in the columns in the white area. This is the required information. The grey area is optional; however, it would be very helpful for querying data, if you submit as much optional metadata as possible. Pedigree information is in the subjectsHuman.xls file. Pedigree information is required for family based studies.
- The required fields are also marked with a red asterisk (*).
- The MS Excel worksheets (or templates) connected to each other via key linkages which are also color coded for your convenience. For example, the Protocol IDs defined in protocols.xls are also referenced in bioSamples.xls when you indicate which protocol was used for a particular biological sample. It is important to make sure all of these key linkages match up when they are used. See Figure 20.
- A pull-down menu appears in those cells where you must use a term from the list.
- A red triangle in the column name indicates that there are additional instructions in the form of comments available upon mousing over.
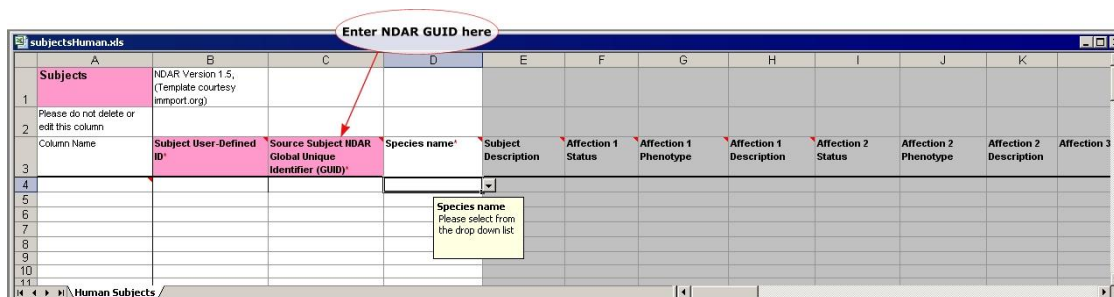
subjectsHuman.xls

Figure 21. subjectsHuman.xls

This template has three required fields: Subject User-Defined ID, Source Subject NDAR Global Unique Identifier (GUID), and Species Name. Three additional pedigree fields are required for family based studies.

In the Source Subject NDAR Global Unique Identifier (GUID) column, enter the NDAR GUID for each research subject mentioned in the file. Enter identifiers for a subject's family, mother, and father, if available (required for family studies). In the Species Name field, use the list box to select Homo Sapiens for each subject. Also, for each subject, complete all columns that contain additional information which is available, such as Affection Status, Gender, and others. Save the file in Tab-delimited format.

Note that the subjectHuman.xls template has the orange color coded fields (which are not shown in Figure 21): Subject Treatment Protocol User-Defined ID and Subject Treatment Protocol NDAR Accession. The Subject Treatment Protocol User-Defined ID field refers to protocols.xls that described any treatments for the chosen Subject. The values entered into Subject Treatment Protocol User-Defined ID are used as links to the protocols.xls file. After you have completed protocols.xls, copy and paste values from the Protocol User-Defined ID field to the corresponding Subject Treatment Protocol User-Defined ID field. If you have already uploaded protocols.xls to NDAR, you can use the protocol accession number to refer to that protocol. In that case, enter accession numbers into the Subject Treatment Protocol NDAR Genomics Accession field. For both fields one or more values can be entered per row. To separate values, use a semi-colon (;).

Remember that in the metadata transfer file templates, the column headers are color coded by their category of metadata.

experiments.xls

Figure 22. experiments.xls

This is a template intended to describe the goal and methods used in an experiment. For each experiment, enter the information in the Experiment User Defined ID and Experiment Name columns. Then, use the drop-down lists to select the Experiment Type and Measurement and Analysis Techniques. Save the file in Tab-delimited format.

Note that the experiments.xls template also has the orange color coded fields that refer to protocols.xls. The fields are Protocol User-Defined ID and Protocol NDAR Accession. After you have completed protocols.xls or obtained the protocol accession numbers, enter the values from the Protocol User-Defined ID field in protocols.xls into the corresponding field of experiments.xls, or enter the accession number into Protocol NDAR Genomics Accession.

bioSamples.xls

Required fields are: Biological Sample User-Defined ID, Biological Sample Type, Biological Sample Description, Source Biological Sample Provider Accession, Source Biological Sample Provider, Source Subject NDAR Global Unique Identifier (GUID).

In the Biological Sample User-Defined ID column, enter the sample ID which should look like "Sample_from_NDAR GUID". Then, in the Source Biological Sample User-Defined ID column, enter the NDAR GUID for each subject. This will link the information from the bioSamples.xls file to the information stored in the subjectsHuman.xls file. Note that these fields are pink color coded (they are not shown in Figure 23 below).

Figure 23. bioSamples.xls

In the Biological Sample Type column, use the list box to select the proper type from the list. Save the file in Tab-delimited format. If the biological samples that you used are derived from other biological samples mentioned in the submission package, enter the source sample IDs or accession numbers for these samples into the Source Biological Sample User-Defined ID and Source Biological Sample Accession Number fields.

reagents.xls

Figure 24. reagents.xls

Since different assay reagents have different characteristics, reagents.xls has several reagent-specific worksheets to capture important aspects of the reagents. For example, in Figure 24 the worksheet tabs are Reagents_Array, Reagents_Illumina, and Other. To fill out the reagents.xls template, select a worksheet that corresponds to your experiment. In the Reagent User Defined ID column, enter your local ID that should also be referenced into the experimentSamples.xls template. Also, complete the other required (white) and optional (shaded) column(s). For example for the Array worksheet, the required fields are: Array Reagent User-Defined ID, Array Platform Name, Manufacturer, and Catalog Number. Save the file in Tab-delimited format.
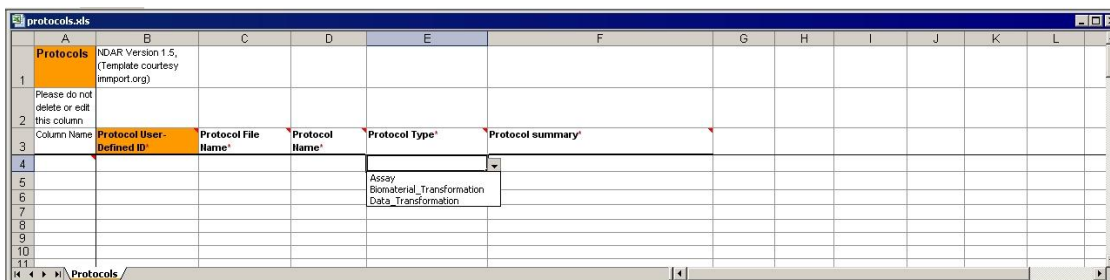
protocols.xls

Figure 25. protocols.xls

In the protocols.xls template, fill out the following required fields: Protocol User-Defined ID, Protocol File Name, and Protocol Name. The Protocol User-Defined ID should be also referenced in subjectsHuman.xls, experimentSamples.xls, and experiment.xls. The Protocol File Name field must contain the file name of the protocol document (if any) included into the submission package.
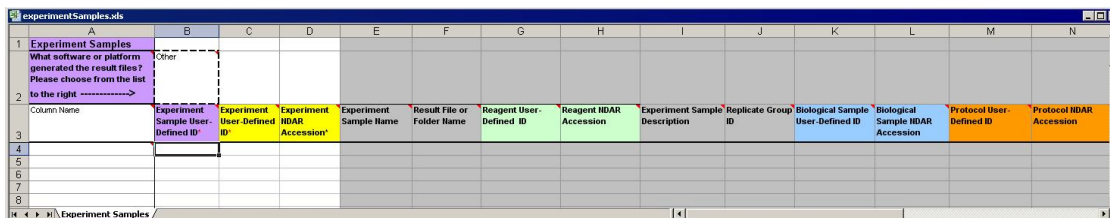
experimentSamples.xls

Figure 26. experimentSamples.xls

experimentSamples.xls is the last file you need to fill out in order to put all your submission data together. It links the following files: bioSamples.xls, reagents.xls, and experiments.xls. While filling out that template, you should, first, indicate an assay method that you used to generate results. In order to do that, go to the cell B2 and use the list box to select the assay method from the controlled vocabulary (see Figure 27). If you cannot find the proper value, choose "Other". Then, fill out the Experiment User-Defined ID field. [4]

---

[4] If "Other" is chosen, a request may be sent to NDAR to review addition of the assay to the controlled vocabulary list.

Figure 27. Filling out experimentSamples.xls

In experimentSamples.xls, a single experiment sample ID is linked to a result file, which means that the value you have entered into the Experiment User-Defined ID field will be linked to the value that you put into the Result File or Folder Name field (column F). Note that there are scenarios where, instead of a single file, a set of files is associated with an experiment sample ID, for example, Affymetrix genotype results. In that case, put all required files into a folder, and then enter a folder name in the Result File or Folder Name field (the column F). See Figure 26.

To store a dataset as used in a publication with specific transformations and normalizations, a separate column Published Result File or Folder Name is provided.

In order to accurately link data from bioSamples.xls, experiments.xls, reagents.xls, and results.xls to experimentSamples.xls, you should do some copy and pasting of data entered earlier into other metadata transfer templates.

Filling out experimentSamples.xls:

1   In the Experiment User-Defined ID field, copy and paste the Experiment User-Defined ID that you entered into the experiments.xls template cell B4. This will link a new experiment sample to the experiment you defined earlier in experiments.xls. To add data to an existing experiment, in the Experiment NDAR Genomics Accession field, enter the accession numbers for those experiments that have already been uploaded to NDAR. See Figure 26.

2   In the Reagent User-Defined ID field (column H), copy and paste a value from the reagents.xls column B (Array Reagent User-Defined ID).

3   In the Biological Sample User-Defined ID field (column L), copy and paste a value from the bioSamples.xls column B (Biological Sample User-Defined ID). If you have more than one biological sample, separate the sample IDs using a semi-colon (;).

4   Save the file in Tab-delimited format.

What to do next?

After you have filled out the metadata templates, launch the NDAR Data Validation tool (refer to page 4) and use it to validate genomics metadata. Then, launch the NDAR Data Submission tool (refer to page 10) to upload your data to NDAR.