SOP: Loading/Linking Data for Application Inventory Registry (AIR) Compliance

Version 2.0

Revision History Log:

Previous Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #
NA	9/13/2017	Dratch, Dieffenbach, Kingston	NA	Original version	1.0
1.0	10/5/2017	Kingston	Added instructions for using Protocol Tracker for Step 3. Changed "Protocol Tracker" to "Protocol Tracking" throughout. Changed use of keyword from suggestion to directive.	Protocol Tracker dev. completed. Mistake on app. name. Recommendation from Data and Systems Officer.	2.0

Background

See AIR Brief (https://irma.nps.gov/DataStore/Reference/Profile/2243908)

Purpose:

Define:

- … what data "counts" (i.e. scope)
- … how to organize the data
- … ways to get data on an approved DOI/NPS Data Center server
- … how to tie the data to the application described in Protocol Tracking

Part One: Planning Ahead addresses a few common questions related to the AIR process.

Part Two: Getting your Application Data on an approved DOI/NPS Data Center server provides explicit instructions for each option given to help networks comply with AIR.

Part Three: Associate Application Data with Protocol Tracking Application ties it all together so that applications recorded in Protocol Tracking now point to the associated data on an approved DOI/NPS Data Center server.

Part One: Planning Ahead

How do I determine what data I need to include for AIR compliance?

- This effort applies to data (hereafter referred to as application data) in applications that are identified as "Master Storage" in the Protocol Tracking application (https://irma.nps.gov/protocoltracking/)
- We want the most complete (hopefully within the last year), most processed, best available data that you have
- We are not tackling non-tabular data (e.g., data photos, audio, video) in this effort

What access level should you allow in Data Store?

There is no requirement that references and files associated with application data be broadly discoverable for this effort.

> Unless you're **absolutely** certain that your data are ready for discovery by the **public**, we recommend that you **restrict access** to your application data

Access should be set using the following guidelines.

ACCESS	
References	**Publicly accessible or internal NPS only?** References are just metadata and should not be sensitive. It is acceptable to make them internal NPS only.
Files	**Publicly accessible, internal NPS only, or restricted access to only a few individuals?** For application data that are sensitive, interim, or 'not-ready-for-prime-time', no matter the reason, restricted access is recommended.
Data Harvesting	**Will restricted application data be harvested by other systems (e.g. data.doi.gov)?** **NO** references designated as internal will be harvested. Our permissions model will *always* enforce access restrictions when a Data Store reference or file URL is exposed.

Should I create a reference for each file, or can I link multiple files to one reference (if more than one file is related to a protocol)?

Protocol Tracking allows pointing back to more than one reference for each protocol, but consider the following:

- **All** files linked to a single reference inherit the same file access level
- Reference metadata and type should describe **all** the files linked to it
- Versioning, when used, occurs at the reference level, NOT the file level

Recommendations:

- Create a separate reference for each file that should be both described and discoverable independent from other files.

EXAMPLE	A geospatial dataset and a tabular dataset should have different references because they are different 'reference types.'

- Link multiple files to a single reference *if* the individual files are closely related (either parts of a whole, necessary as a suite for interpretation, or broken up for practicality of upload/download because of file size)

Should the file be in a certain format?

- **NO**.
  For now, upload application datasets in whatever form they currently exist.
  - For file-based databases, post copies of the data files.
  - For enterprise databases that are not already on an approved DOI/NPS Data Center server (Denver or Reston), use a backup file (e.g., for SQL Server, use a .bak file).
- Remember that we are *not* dealing with non-tabular data for this effort. It will be addressed when we work on Remodeled Data Store.
- Format decisions may be influenced by file size.
  Data Store can handle very large files, but some file types may not be practical to upload/download. Consider zipping or otherwise compressing large files.
- If you are unable to upload application data files from the network (large files or bandwidth constraints), create the reference(s) and the central office will work with you to provide assistance to upload the file(s).

Should the dataset be cumulative as of the date of upload?

- The dataset(s) should cover the ***entire*** period of collection for the protocol. Whether this is accomplished in separate files by field season/year or a single file with all of the data is not important.
- Ideally, we'd like the data to be current within the last year.

EXAMPLES	Network 'A' manages annual datasets as separate files and has ten years of data. The database schema has changed slightly over the years. ***Recommendation:*** Make one Data Store reference. Attach all data files to that reference.
	Network 'B' has a single cumulative database for all years of data. ***Recommendation:*** Make one Data Store reference. Attach the cumulative data file to that reference.

Do I need to organize/link the individual references with a special Data Store structure?

- **NO.**
  The first priority is to get the application data on an approved DOI/NPS Data Center server! However, if you prefer to organize the references using one of the Data Store constructs, you may.
- There is no requirement at this time for doing anything else to organize these references. The Remodeled Data Store will address this further at a later date.

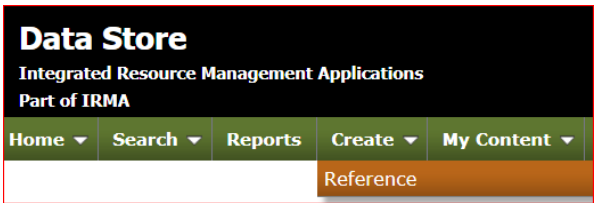
Part Two: Getting your Application Data on an Approved DOI/NPS Data Center Server

Select the option that matches the case for your application data and complete the steps as outlined.

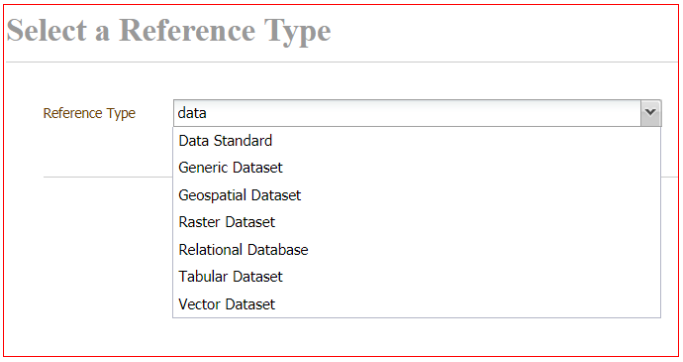
Option 1: Creating a New Data Store Reference

1. **Go to the Data Store (https://irma.nps.gov/DataStore/)**

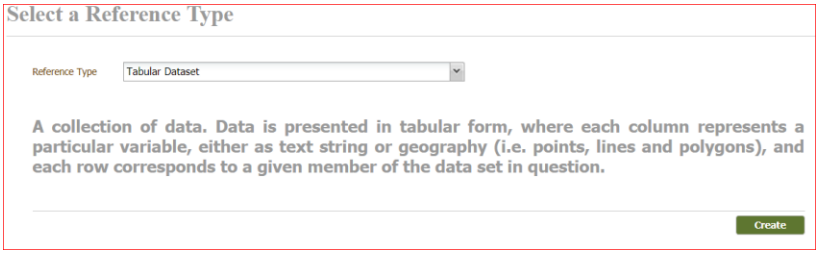
2. **Select Create > Reference**



3. **Select a Reference Type** from the list (start typing and the list will filter).



4. The definition of the Reference Type will display. Click **Create** to continue.



Most common tabular reference types used for tabular data:

Reference Type	Data Description
Generic Dataset	Spatial or tabular data that cannot be more specifically defined (e.g. as a vector dataset).
Geospatial Dataset	Inherently geospatial data; can be natively rendered using geospatial software.

Raster Dataset	Storage format for managing the location of geospatial information as grids and their associated attribute information. Raster datasets are natively geospatial.
Relational Database	A database containing more than one table, with the tables sharing data.
Tabular Dataset	A collection of data. Data is presented in tabular form, where each column represents a particular variable, either as text string or geography (i.e. points, lines and polygons), and each row corresponds to a given member of the data set in question.
Vector Dataset	Storage format for managing the location of geospatial information (points, lines and polygons) and their associated attribute information. Vector datasets are natively geospatial.

5. Fill in metadata

The create/edit page will have a series of tabs with elements to complete the metadata.
There is no requirement to fill in every field, especially if it creates redundancy

* - required field

Core (*available fields and labels may vary slightly between reference types*)

a. **Title** *
   Should reflect that it is data and indicate what is being monitored.
   The network isn't necessary in the title if it is expressed elsewhere in the core metadata (e.g. creator or publisher).

EXAMPLES	NCBN Estuarine Nutrient Enrichment Seagrass Monitoring Database
	SCPN Aquatic Macroinvertebrate Monitoring Database
	NCCN Landbird Monitoring tabular data

b. Dates, and person/organization responsible for the application data (creators, stewards, publisher) *recommended*
c. Description *highly recommended*
d. **'Created by and/or for the NPS'** check the checkbox*

Permissions

a. **Reference Owners***
   Owners may edit the reference and download restricted files.
   Creator of the reference is automatically added as an owner. Add any other owners.
b. **Reference Access Level** *
   Choose Public or Internal
c. **File Access Level** *
   Choose Public, Internal, or Specific Individuals.
   For specific individuals: you can add file access permission for non-reference owners.
d. **Justification for File Access level** *
   Choices will change depending on file access level. Defaults are Not Sensitive, None (no copyright or distribution restrictions), Quality unknown.

Files and Links

a. Add (upload) files *
b. Add a description for the uploaded files (optional)
   Most important when there are multiple files.
   Description can be added at the time of upload, or added/edited after upload.

Units and Geography

a. Add the network code associated with the network. *
b. Check the box if you want to include all linked units in the network, or add individual parks if that is more appropriate.

Subjects and Keywords

a. Add keyword phrase *Application Inventory Registry* to identify this as being linked in Protocol Tracking as part of AIR. *

Save and Close

a. Click on Activate
   The reference will be 'published' with the reference and file access level as indicated in the permissions tab.
b. Save as draft *if* you want to finish creating the record later, *before* you publish.
c. Copy the Reference Code and go to Part 3.

Edit

- If you own the reference, edit at any time by selecting Edit from the Actions dropdown in the upper right corner of the reference profile.

Option 2: Using an Existing Data Store Reference

1. Add the keyword phrase *Application Inventory Registry* identified in Option #1 above. *
2. Copy the Reference Code and go to Part 3.

Option 3: Refer to Data on an Approved DOI/NPS Data Center Server

Note that this includes the following:

- Data in WRD EQuIS
- Data in WRD STORET
- Data in WRD Aquarius
- Data on an approved DOI/NPS Data Center SQL Server instance
- Data on a file share on an approved DOI/NPS Data Center server
- Data in an application on an approved DOI/NPS Data Center server

1. Copy the appropriate information:
    - application name (EQuIS, STORET, or Aquarius)
    - instance name (SQL Server)
    - path (file share)
    - URL (application or other approved DOI/NPS Data Center storage)

    and go to Part 3.

Unable to access the application data with a URL?
You may need to pursue Option #1.

Part Three: Associate Application Data with Protocol Tracking Application

The Protocol Tracking application has been updated (as of 10/5/2017) and is the preferred method for associating application data with the protocol applications.

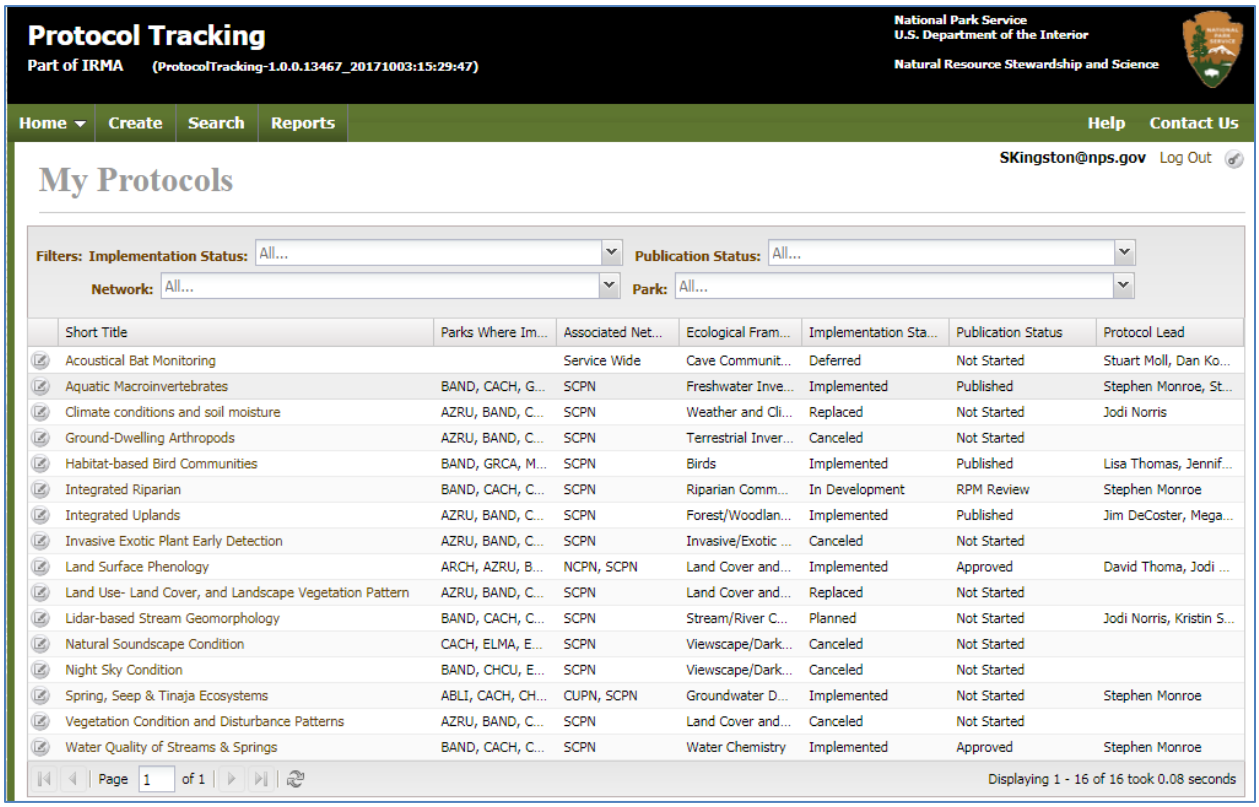
Instructions are also included from the original September, 2017 release of this SOP when the Protocol Tracking application was in the midst of updates that prevented its use for AIR compliance. Instead, records from Protocol Tracking were copied into a Google Sheet (on 9/11/2017 at 2:10 PM Mountain Time) to facilitate AIR compliance.

Networks may use either method, but direct entry into Protocol Tracking is preferred.

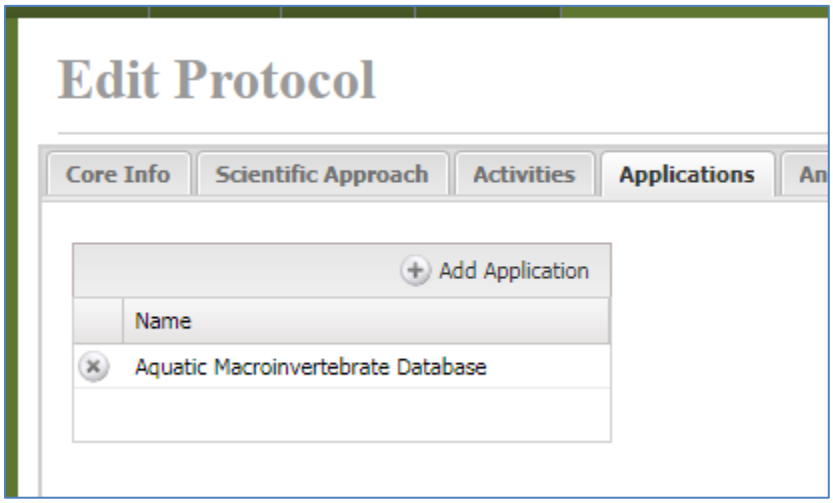
Option 1 (Preferred): Use the Protocol Tracking Application

Option 1, Scenario #1: Happy Path

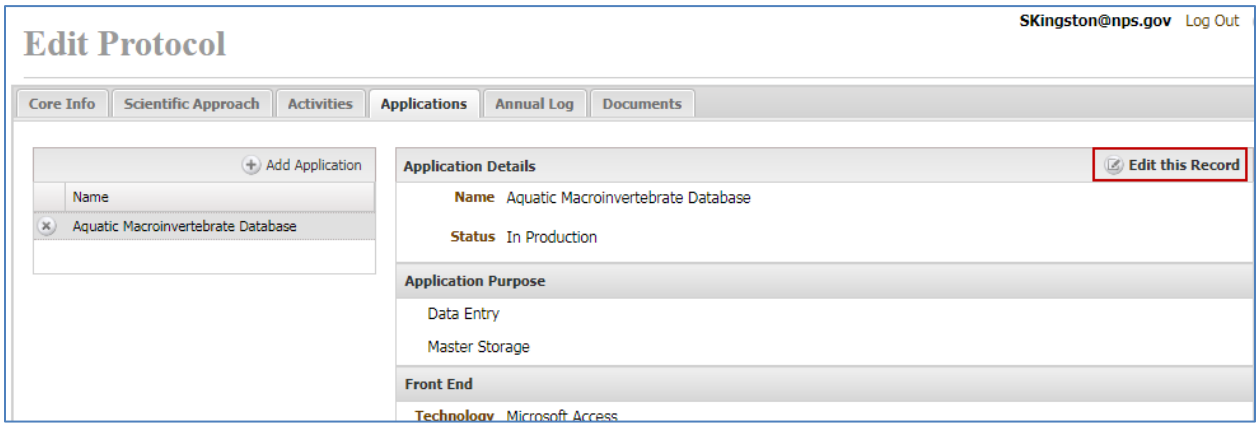
1. Open the Protocol Tracking Application (https://irma.nps.gov/ProtocolTracking/). You should see all of the protocols associated with your network (SCPN is used in this example image). If you don't, then you don't have a role in Protocol Tracker and should contact Joe DeVivo.

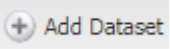


2. Click the Edit button (small pencil and pad icon) next to the protocol you wish to update. This will open the Edit Protocol page.
3. Click on the Applications tab.

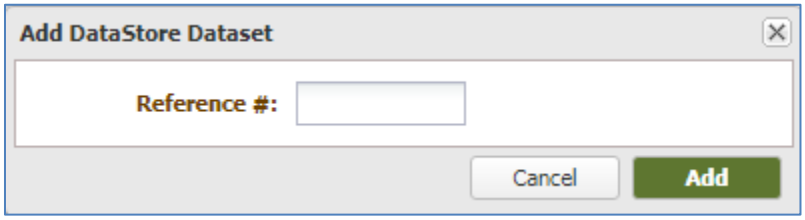


4. Click on the name of the application that contains the master data to show the details for the application.
5. Click on the Edit this Record button at the top right of the Application Details.



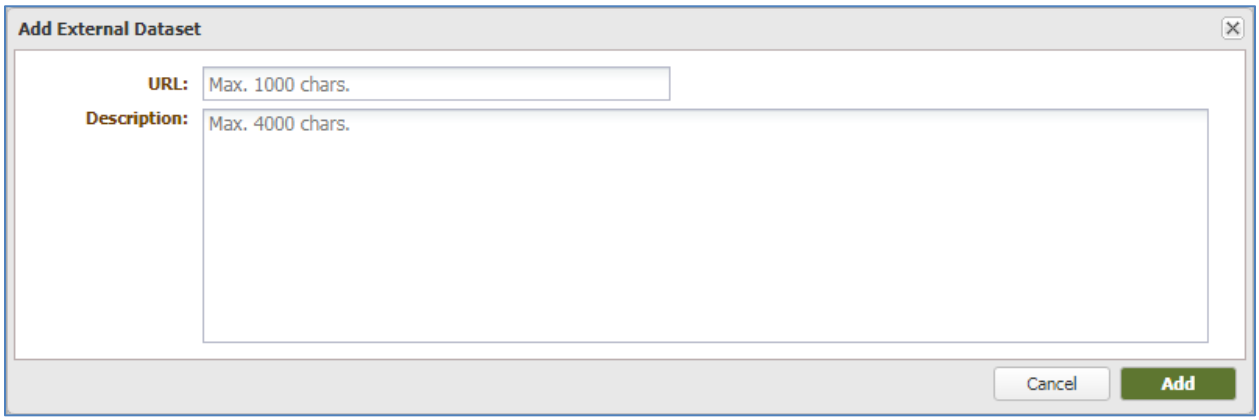
6. Scroll down to the bottom of the Add/Update Application form to see the Application Database Datasets section and the External Datasets section.
7. If the data are posted to Data Store, click on the Add Dataset button 

on the Application Database Datasets section and then paste in the Reference Code and click the Add button.



8. If the data are posted to another location that can be referenced by URL, click on the Add Dataset button

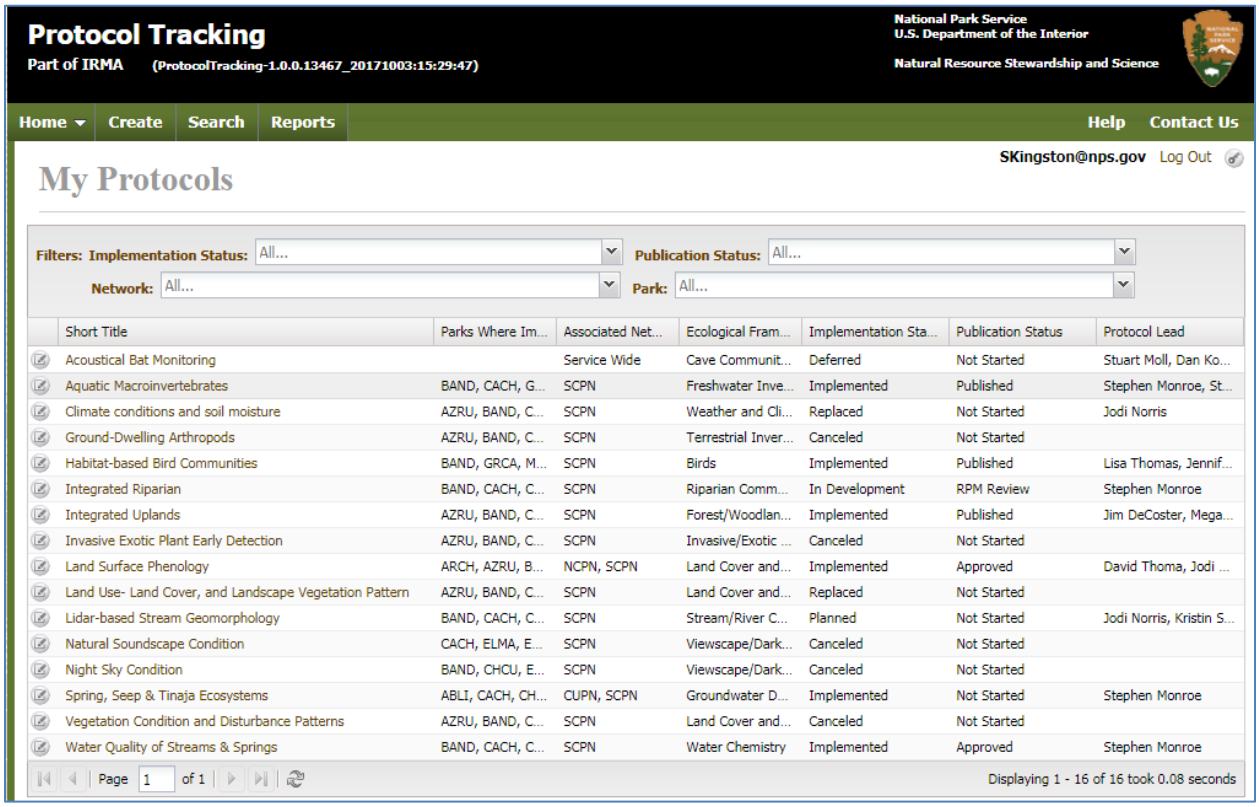
on the External Datasets section and then paste in the URL, fill out the Description, and click the Add button.

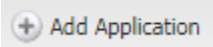


9. If you have any notes to add, put them in the "Data Storage Notes" field.

Option 1, Scenario #2: The Application is not in Protocol Tracking

1. Open the Protocol Tracking Application (https://irma.nps.gov/ProtocolTracking/). You should see all of the protocols associated with your network (SCPN is used in this example image).



2. Click the Edit button (small pencil and pad icon) next to the protocol you wish to update. This will open the Edit Protocol page.
3. Click on the Applications tab.
4. Click the Add Application button 

.
5. Complete the application information.
6. When you get to the Application Database Datasets and External Datasets at the end of the form, proceed with Option 1, Scenario #1, Step #7.

Scenario #3: Two Protocols/Applications have Data in the Same Place

It's OK to use the same location in the Application Database Datasets / External Datasets for more than one protocol/application. Just make sure that each relevant application has an entry.

Scenario #4: I have Data in more than One Place for an Application

1. Follow the instructions for Option 1, Scenario #1 and repeat Steps #7-9 for each dataset.

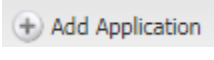
Option 2: Use the Google Sheet

Option 2, Scenario #1: Happy Path

1. Open the AIR Data Reporting Google sheet (https://docs.google.com/a/doi.gov/spreadsheets/d/1nom2S8nd4GKRcDDYU_8jzqHPYVj8J0pc8rnRxPhjwCc/edit?usp=sharing).
2. Scroll down to the relevant application for your network (should have TRUE in the column Application is Master for Stored Data, but some networks don't have any application listed as Master for their protocols).
3. Paste the information (Data Store reference code/application name/instance name/path/URL) from the end of Part 2 into the "AIR DataStore Code / URL of Uploaded Data" column.
4. If you have any notes to add, put them in the "AIR Upload Notes" column.

Option 2, Scenario #2: The Application is not on the AIR Data Reporting Google Sheet

We downloaded data from Protocol Tracking on 9/11/2017 at 2:10 PM Mountain Time. If records were added/edited/deleted after that time, it will not be reflected on the AIR Data Reporting Google Sheet.

1. Open the Protocol Tracking Application (https://irma.nps.gov/ProtocolTracking/).
2. Click the Edit button (small pencil and pad icon) next to the protocol you wish to update.
3. Click on the Applications tab.
    a. If you find a record for the application, contact Simon Kingston to copy the information into the AIR Data Reporting Google Sheet.
    b. If you do not find a record for the application, click the Add Application button 

, complete the application information, and then contact Simon Kingston to copy the information into the AIR Data Reporting Google Sheet.
4. Once the AIR Data Reporting Google Sheet is updated with the new record, proceed with Option 2, Scenario #1.

Option 2, Scenario #3: The Information on the AIR Data Reporting Google Sheet needs to be updated

We downloaded data from Protocol Tracking on 9/11/2017 at 2:10 PM Mountain Time. If records were added/edited/deleted after that time, it will not be reflected on the AIR Data Reporting Google Sheet.

1. Open the Protocol Tracking Application (https://irma.nps.gov/ProtocolTracking/).
2. Click the Edit button (small pencil and pad icon) next to the protocol you wish to update.
3. Click on the Applications tab.
4. Click on the name of the application that needs to be updated.
5. Click on the Edit this Record button at the top right of the Application Details.
6. Correct the information as necessary on the Status and Data Management tabs and click the Save button.
7. Contact Simon Kingston to copy the updates into the AIR Data Reporting Google Sheet.
8. Once the AIR Data Reporting Google Sheet is updated, proceed with Option 2, Scenario #1.

Option 2, Scenario #4: Two Protocols/Applications have Data in the Same Place

It's OK to use the same location in the "AIR DataStore Code / URL of Uploaded Data" column for more than one protocol/application. Just make sure that each relevant application has an entry.

Option 2, Scenario #5: I have Data in more than One Place for an Application

1. Contact Simon Kingston to create duplicate rows for your application in the AIR Data Reporting Google Sheet.
2. Once the AIR Data Reporting Google Sheet is updated, proceed with Option 2, Scenario #1, adding entries for each data location.

References

- AIR Brief (https://irma.nps.gov/DataStore/Reference/Profile/2243908)
- Protocol Tracking application (https://irma.nps.gov/ProtocolTracking/)
- Data Store (https://irma.nps.gov/DataStore/)
- AIR Data Reporting Google sheet (https://docs.google.com/a/doi.gov/spreadsheets/d/1nom2S8nd4GKRcDDYU_8jzqHPYVj8J0pc8rnRxPhjwCc/edit?usp=sharing)

Questions/Comments/Confusion?

Contact:

Simon Kingston (simon_kingston@nps.gov, 970-225-3551) – General questions

Kathy Dratch (kathy_dratch@nps.gov, 970-267-2125) – Data Store questions