**Breeding Bird Monitoring Protocol for the Heartland Inventory and Monitoring Network**

**Standard Operation Procedure 6: Data Management**

**Version 3.00 (05/21/2019)**

**Revision History Log:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Previous Version # | Revision Date | Author | Changes Made | Reason for Change | New Version # |
| 1.01 | May 2008 | D.G. Peitz | Entire document | Edited to reflect that bird monitoring has been expanded to ten additional network parks | 2.00 |
| 2.00 | May 2019 | D.G. Peitz  G. A. Rowell | Edited to clarify conversion of our database to SQL Server; Updated throughout to NRR format. | Updates reflect the conversion of the bird database from Access to an online SQL database. Made SOP NRSS compliant. | 3.00 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

This Standard Operating Procedure describes procedures for managing the Heartland Inventory and Monitoring Network database for breeding bird communities. Specifically, this document addresses procedures for data entry, verification, validation, export to outside systems, security, and availability.

**Data Model**

The data object relationships in the Heartland Inventory and Monitoring Network Landbird database are shown in Figure 6.1. The database is a plot count with continuous distance and with detailed vegetation information. Vegetation profile data are collected to the nearest sampling season (year) while bird observations are collected to the nearest minute interval.



Weather

Bird

Observations

Sampling Period and Location

**Bird Data**

**Habitat**

Slope and Aspect



Vegetation Type

Tree Tally

Canopy Height

Canopy Cover

Basal Area

Ground Cover

Foliar Cover

Horizontal Vegetative Structure

Vertical Vegetative Structure

Location Only

(Permanent Features)

Plot

Subplot

Sampling Period and Location

Plot

Subplot

**Figure 6.1**. Object model for Heartland Inventory and Monitoring Network breeding bird database.

**Field Data Preparation**

Quality assurance and quality control procedures related to data recording are important components of any project. Sampling data, for example sampling intervals, site conditions, species plot count data and vegetation profile data (where collected), are recorded and checked for completeness either before leaving a site or within 24 hours of data recording. This will aid in verification and validation of the data after entry into the database. To prevent the complete loss of field form data due to unforeseen circumstances (i.e., fire or flood in the workplace), field sheets are scanned into a computer and electronic copies of the data sheets stored on the Heartland Inventory and Monitoring Network server located at Wilson’s Creek National Battlefield, Republic, MO. This will ensure that at least one copy of the field sheets is available for data entry and verification.

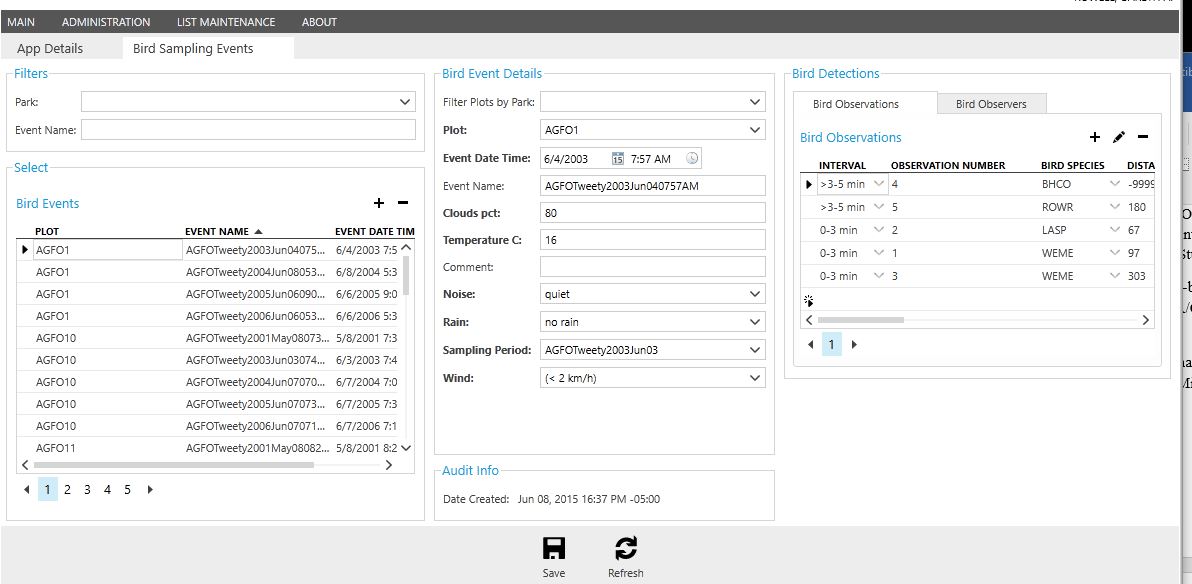
**HTLN\_Landbirds Database**

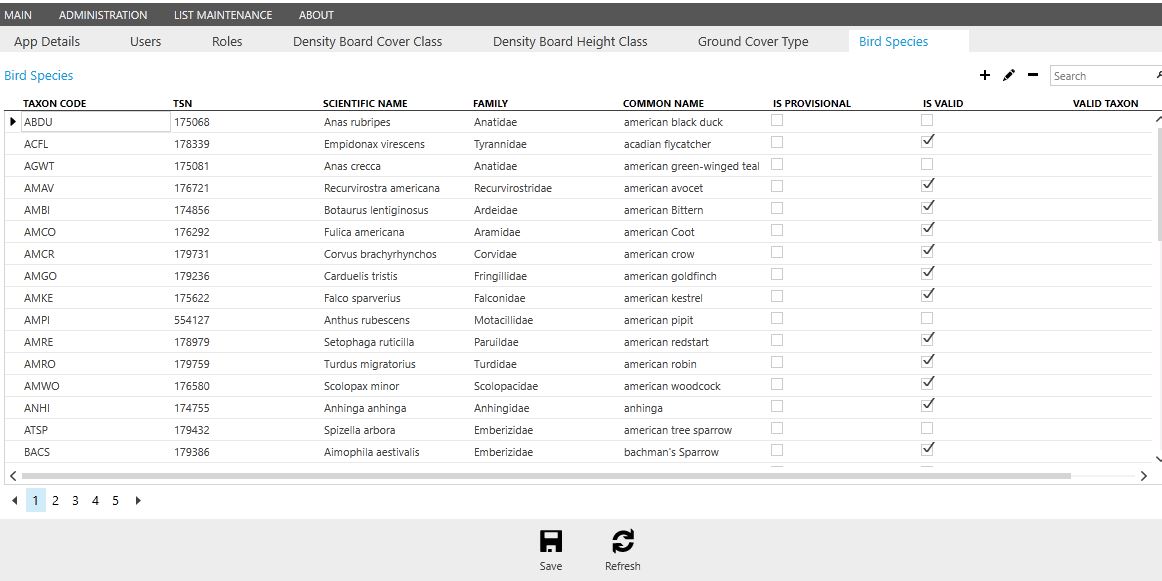
The breeding bird monitoring database is implemented in Microsoft SQL Server. It is called HTLN\_Landbirds. The database is run from the WASO IRMA production server (aka “the IRMA instance”). Database users will need to become familiar with the associated web-based data entry / edit tools (Figure 6.2) while database administrators will need to learn to use web-based tools for look-up tables (Figure 6.3) and the SQL Server Management Studio backend (Figures 6.4 and 6.5).

Web-based user interface

Changes in data-values can *only* be made through the web-based user interface. The web-based user interface for the Heartland Inventory and Monitoring Network database may be reached from the following URL. Users must utilize either Internet Explorer or Microsoft Edge. The web-based interface is not compatible with Chrome or Firefox.

<https://irma.nps.gov/HTLN_Landbirds/DesktopClient/>

**Figure 6.2.** Users can manage bird and habitat data through this web-based interface.

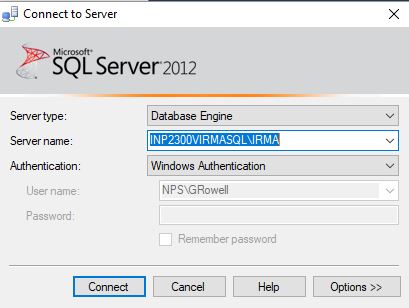


**Figure 6.3.** Administrators can manage look-up tables through this web-based interface.

SQL Server Management Studio

A local copy of SQL Server Express (free version) is necessary to view the database backend (see Appendix 6.1 for installation notes). Once the SQL Server Express installation is successful, go to the Microsoft SQL Server folder under the Start window and open SQL Server Management Studio. The database in SQL Server Management Studio is read-only and can be used to download data from the backend for QA/QC and analysis. The user interface for SQL Server Management Studio is very intuitive but it requires some basic knowledge of SQL in order to download data. Fortunately, the primary and foreign key naming conventions in all of the tables are very consistent, making SQL JOINs straightforward. Furthermore, SQL scripts are saved to the user’s local computer by default so that they can easily be reused. Some examples of SQL scripts and their associated data are given in Appendix 6.2.

When SQL Server Management Studio is opened, it will immediately ask which SQL Server instance should be used:

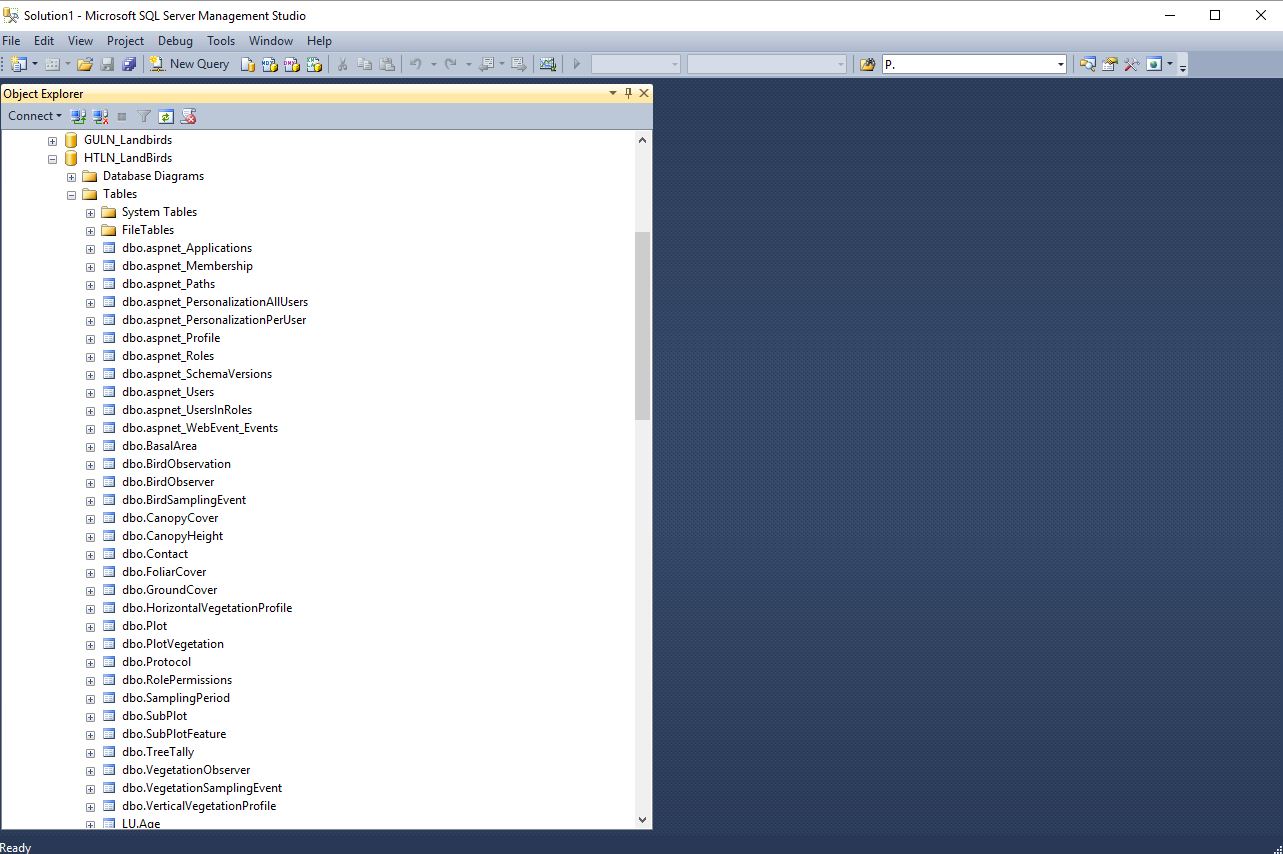


**Figure 6.4.** Connecting to the IRMA production instance from SQL Server Management Studio.

The instance of SQL Server database engine is called:

**INP2300VIRMASQL\IRMA**

Select Windows Authentication and click on Connect. The IRMA instance should connect at this time and the Databases folder can be opened to view all of the databases associated within the IRMA WASO production instance. Look for the HTLN\_Landbirds database and click on it to see the tables.

****

**Figure 6.5**. Tables displayed in the HTLN\_Landbird database from SQL Server Management Studio.

Click on any table and view the columns. The primary key in *every* table is called ID. The foreign keys in every table are called <table>ID. This notation will become extremely useful when writing SELECT scripts.

**Data Verification**

Data verification immediately follows data entry. Computer records are verified for accuracy against paper field data sheets. Hard copy of data records should be used in the verification against field data to minimize proof-reading error. Compare the output directly with original field data sheets to identify missing, mismatched, or redundant records. The verification step should be completed by staff other than those doing data entry, if at all possible. Following verification, the project manager should recheck 10% of the records. The verification and recheck steps should be repeated until no errors are discovered

Procedures:

1. Open SQL Server Management Studio and connect to the IRMA instance.
2. Create a new query or open an existing one to SELECT the data needing verification.
3. Execute the query.
4. Select all from the Results tab and paste it into Excel. Names of columns will need to be designated afterwards. Check the record count in Excel to see that it matches the results in SQL Server Management Studio.
5. Recheck 10% of records. If errors are detected, repeat steps 2 and 3.

**Data Validation**

Data validation involves the application of business rules and checking the accuracy of data against outside controls or specifications. Four types of data validation are used with the breeding bird database. They are:

* + Explicit primary – foreign keys
  + No nulls for any column on all tables
  + Nominal data are restricted to pick lists / combo boxes
  + Reasonable values for continuous attribute data

Referential Integrity

Referential integrity is a property of the relationships between database tables. Referential integrity is created by imposing rules or “constraints” on the relationships between key fields. A key can be either a primary key or a foreign key depending upon what rules are assigned to it. Primary key values must be unique and cannot be null. Each value in a foreign key must be derived from the domain of its related primary key. Referential constraints prevent dangling references between rows of related tables. Furthermore, they reduce the chance of inadvertent record deletions.

Nominal data

Nominal data is controlled during the data-entry process by limiting data-entry to pre-determined values using pick list and combo-box controls on forms. Examples of nominal data include species, ground cover type, foliar cover type and vegetation type.

Continuous data

Validation of continuous data requires the attention of the project manager. Knowledge of the ecological system is necessary to determine what constitutes a reasonable quantitative value for a particular parameter. The data can be exported to a statistical package for quantitative evaluation. Although analysis might be conducted on trimmed distributions, the complete dataset should always be retained to identify possible sources of error.

Procedures:

1. Export the datasets of interest (see Exporting Data below).
2. Import the data into a spreadsheet or statistical package.
3. Calculate descriptive statistics such as mean, median, standard deviation, range, and sample size. Plot a histogram and identify outlier values.
4. Extreme individual values may indicate recording or data-entry errors.

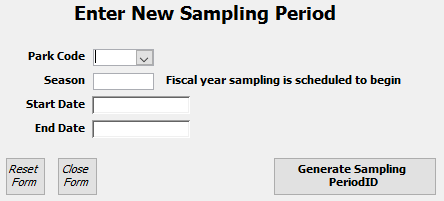
#### Protocol – SOP collections

#### Breeding Bird Sampling Period Data:

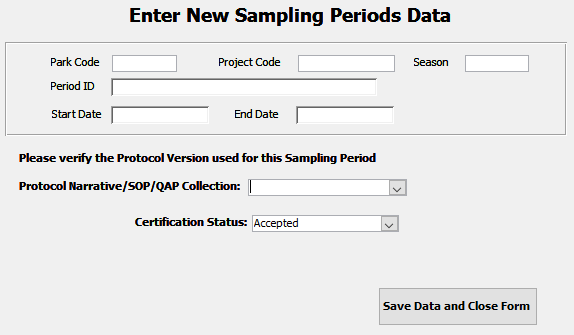
Information that characterizes the entire sampling period for breeding bird data is stored in an Access standalone application called “Bird Protocol Traceability”. This database is located with other Heartland Inventory and Monitoring Network databases on the N: drive:

**N:\HTLN\Databases\BreedingBirds**

Data are entered into the table incrementally throughout the data processing, analysis, and reporting process using master forms and sub-forms. Initially, a new sample period is created for each park using a sampling period data form (Figure 6.6). The specific versions of a monitoring protocol, Quality Assurance Plan, and Standard Operating Procedures default to the most recent collection posted to National Park Service Data Store (Figure 6.7, Appendix 6.3).



**Figure 6.6.** Breeding bird sampling period data entry form.



**Figure 6.7.** Breeding bird protocol traceability data entry form.

**Quality Assurance and Control Procedures and Documentation**

Quality Assurance (QA) includes all activities designed to ensure that data, products, or services meet specified requirements. Quality Assurance focuses on building-in quality to prevent defects. Quality Control (QC) includes procedures for checking whether data meet standards and annotating or qualifying data that do not.

QA and QC procedures and design elements occur throughout data collection, processing, and reporting. The Quality Assurance Plan for monitoring breeding birds in the Heartland Inventory and Monitoring Network (*20xx*) fully describe all QA and QC elements for this protocol.

A primary objective of the QA/QC process is to ensure our monitoring data includes an assessment of its quality. We evaluate and document data quality through three primary means: auditing record counts, assigning a data processing level, and conducting an annual operational review.

Data Processing Level

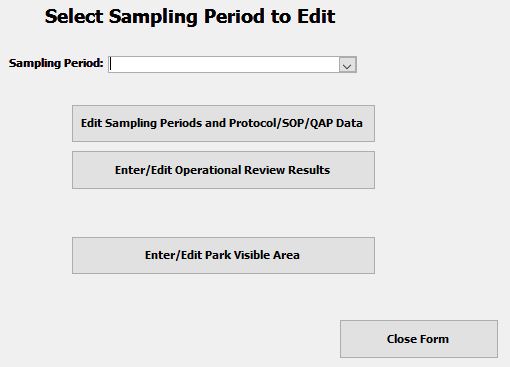
Data processing levels provide a standardized way of defining the degree that QA/QC has been performed on a dataset. Once a dataset has passed the QA/QC procedures specified in the protocol and standards as identified in the breeding bird Quality Assurance Plan, the dataset will be considered “Certified” as defined in Section *5 Data Processing Levels* section of the Minimum Implementation Standards for Network Projects (Frakes and Kingston 2017; Appendix 6.4). All data collected prior to 20xx are designated as “Accepted” given there was no published Quality Assurance Plan for breeding bird monitoring. Data collected in 20xx and thereafter will be considered “Certified” assuming the quality assurance and quality control procedures were followed as described in the Quality Assurance Plan (Heartland Inventory and Monitoring Network *in draft*).

***Operational Review – See SOP #8 “Reporting”***

An annual operational review is required for all active monitoring protocols (Mitchell et al. 2018). Completion of an operational review, a summary of any flagged data, and a link to the review report are stored in the table Tbl\_SamplingPeriods.

QA/QC Data Entry

1. Open form Edit Existing Sampling Periods Data.
2. Select the appropriate data processing level. Default is “certified.”
3. Confirm completion of an annual operational review (Y/N).
4. Did the operational review ‘flag’ any data records? (Y/N). If yes, the reason(s) are documented in the final report which is located on the network server.
5. Enter Data Store Reference ID for final report.

**** 

**Figure 6.7.** Quality assurance and quality control documentation data entry form.

**Data Publication**

Annually, in fulfillment of the Data Analysis and Reporting Requirements (Gallo, K. memorandum dated 4/23/2018), the dataset will be uploaded to IRMA Data Store. The dataset is flagged as “read only” for all users except the Project Leader and Data Manager.

**References**

Frakes, B., S. Kingston, and M. Beer. 2015. Inventory and Monitoring Division database standards: September 11, 2015. Natural Resource Report NPS/NRSS/NRR—2015/1035. National Park Service, Fort Collins, Colorado.

Frakes B and S. Kingston. 2017. Minimum Implementation Standards for Network Projects v.3.0. National Park Service. Last accessed 5/23/2018: <https://irma.nps.gov/DataStore/DownloadFile/588637>

Gallo, K. 2018. Data Analysis and Reporting Requirements memorandum dated 4/23/2018. Last accessed 5/23/2018: <https://irma.nps.gov/DataStore/DownloadFile/600325>

Heartland Inventory and Monitoring Network. *In draft*. Quality assurance plan for monitoring breeding birds in the Heartland Inventory and Monitoring Network. Natural Resource Report NPS/HTLN/NRR—xxxx/xxxx. National Park Service, Fort Collins, Colorado.

Mitchell, B., A. Chung-MacCoubrey, J. Comiskey, L. Garrett, M. MacCluskie, B. Moore, T. Philippi, G. Sanders, and J.P. Schmit. 2018. Inventory and Monitoring Division protocol review guidance. Natural Resource Report NPS/NRSS/IMD/NRR—2018/1644. National Park Service, Fort Collins, Colorado.

National Park Service. 2016. Certification Guidelines for Inventory and Monitoring Data Products. NPS, IMD. Fort Collins, CO. IRMA reference number – 2227397. <https://irma.nps.gov/DataStore/Reference/Profile/2227397>

Rowell, G.A., M.H. Williams, and M.D. DeBacker. Unpublished. Data management plan: Heartland I&M Network and Prairie Cluster Prototype Monitoring Program.

**Appendix 6.1.** Notes on Microsoft SQL Server Express installation

A local copy of SQL Server Express Version 12 or more recent is required in order to connect with the remote IRMA SQL Server instance. The Heartland Inventory and Monitoring Network Landbird Database backend is located on the IRMA SQL Server instance. SQL Server Express software is free. Authentication is based on the Windows user, so no login information is required. Go to the official Microsoft site to download SQL Server Express. There are instructions for installing the software on the Microsoft site, be sure to read them. There are also many tutorials at other sites. Do the install slowly and carefully. *Be sure appropriate administrative permissions are in place before beginning*. Plan to spend an hour or so on the install if this is the first time. There will normally be plenty of disk space and memory, and the prerequisite .net framework software will already be installed. Users will need the version “Express Advanced” which means it has more than just the database engine. Advanced is required to get the SQL Server Management Studio (SSMS):

<https://www.microsoft.com/en-us/sql-server/sql-server-editions-express>

For later versions of SQL Server Express, specify the “Express Advanced” version after the initial install executable is downloaded.

**Appendix 6.2.** Examples of SQL for downloading data.

Example 1. SELECT command for site conditions and bird observations at TAPR and results.

USE HTLN\_LandBirds;

SELECT P.LocationName AS Plot, Right(E.EventName,15) AS EDateTime,

E.Temperature\_C AS Temp\_C, W.Code AS Wind, R.Code AS Precip,

E.Clouds\_pct AS PctCloud, N.Code AS Noise, B.ObservationNumber AS ObsNo, I.Name AS Interval, S.TaxonCode AS AOUCode, B.Distance, D.Code AS DetectType, X.Name AS Sex, A.Name AS Age, B.FlockSize, B.IsPreviousPlot, B.IsFlyover

FROM dbo.BirdSamplingEvent AS E

JOIN dbo.BirdObservation as B

ON E.ID = B.BirdSamplingEventID

JOIN LU.BirdSpecies AS S

ON B.BirdSpeciesID = S.ID

JOIN dbo.Plot AS P

ON E.PlotID = P.ID

JOIN LU.Noise AS N

ON E.NoiseID = N.ID

JOIN LU.Rain AS R

ON E.RainID = R.ID

JOIN LU.Wind AS W

ON E.WindID = W.ID

JOIN LU.Interval AS I

ON B.IntervalID = I.ID

JOIN LU.Sex AS X

ON B.SexID = X.ID

JOIN LU.DetectionType AS D

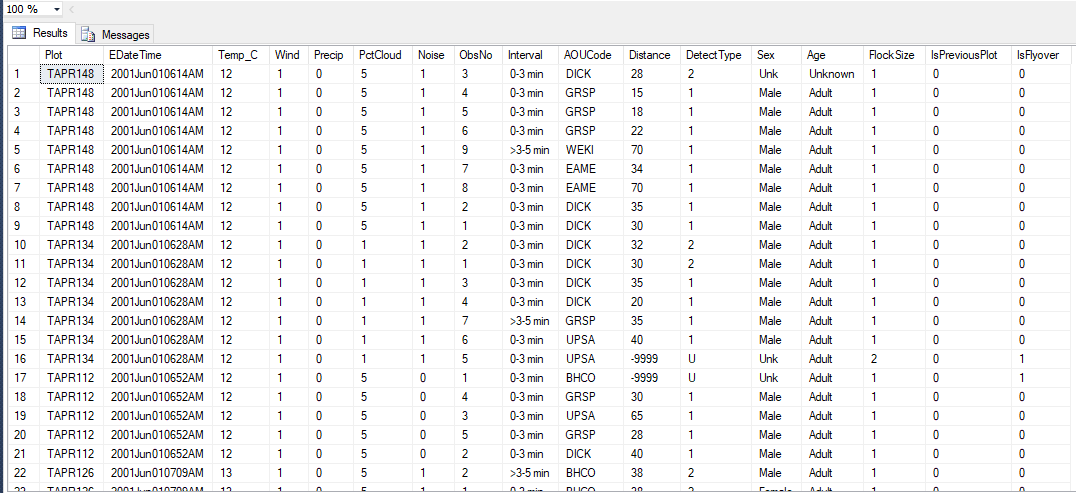
ON B.DetectionTypeID = D.ID

JOIN LU.Age AS A

ON B.AgeID = A.ID

WHERE (LEFT(P.LocationName,4) = 'TAPR')

ORDER Right(E.EventName,15), P.LocationName;



Example 2. SELECT command for Horizontal Distance Profile at TAPR and results of query.

USE HTLN\_LandBirds;

SELECT Right(Left(E.EventDateTime,11),4) AS Year, L.LocationName AS Plot, S.SubPlotNumber AS SubPlot,

X.Name AS Distance, D.Name AS Height, C.Code AS CoverClass

FROM dbo.Plot AS L

JOIN dbo.VegetationSamplingEvent AS E

ON L.ID = E.PlotID

JOIN dbo.SubPlot AS S

ON E.ID = S.VegetationSamplingEventID

JOIN dbo.HorizontalVegetationProfile AS H

ON S.ID = H.SubplotID

JOIN LU.DensityBoardHeightClass AS D

ON D.ID = H.DensityBoardHeightClassID

JOIN LU.DensityBoardCoverClass AS C

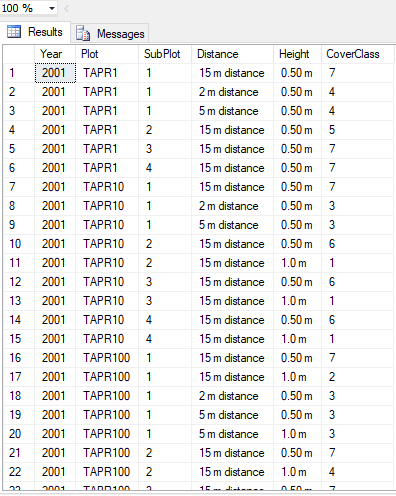
ON C.ID = H.DensityBoardCoverClassID

JOIN LU.DensityBoardDistanceClass AS X

ON X.ID = H.DensityBoardDistanceClassID

WHERE (LEFT(L.LocationName,4) = 'TAPR')

ORDER BY Right(Left(E.EventDateTime,11),4), Plot, SubPlot, Distance, Height

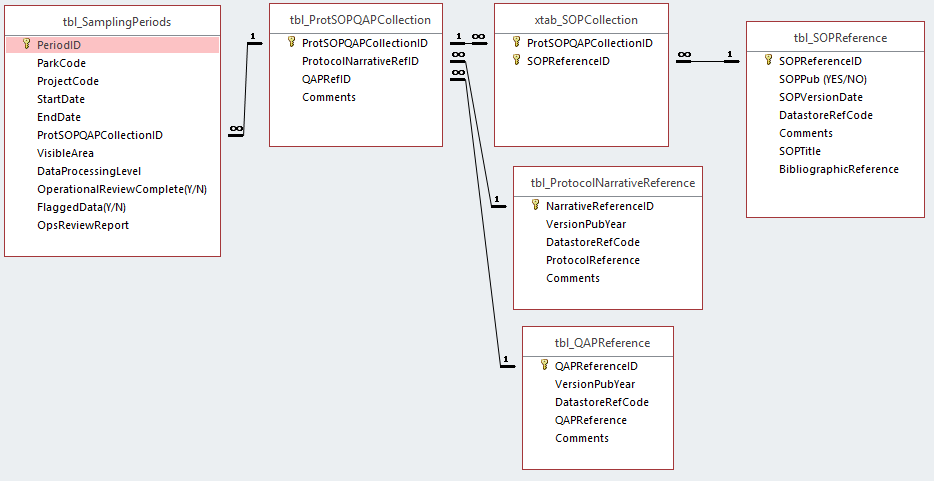


**Appendix 6.3.** Protocol Traceability

Inventory and Monitoring Division Database Standards (Frakes et al. 2015) requires every datum to be unambiguously traceable to a specific version of a monitoring protocol, a Quality Assurance Plan, and suite of Standard Operating Procedures.

The database contains tables that document the assigned data processing level and identifies the protocol narrative, Standard Operating Procedures, and Quality Assurance Plan used for each sampling period. This makes the assigned methods traceable to the record level through the PeriodID (see Figure 6A.1).

As versions change, the project lead works with the data manager to create a new collection of narrative, Quality Assurance Plan, and Standard Operating Procedures. This becomes the new default for the ProtSOPQAPCollectionID field.



**Figure 6A.1.** Breeding bird database tables and their relationship for documenting the specific versions of the protocol narrative, individual SOPs, and Quality Assurance Plan followed for each sampling period. Replace with ERD from Breeding Bird Protocol Traceability Application.

**Appendix 6.4.** Data Processing Level

Natural Resource Stewardship and Science, Inventory and Monitoring Division policy and guidance require a data processing level be assigned to all data. (Frakes, et al. 2015, National Park Service 2016, Frakes and Kingston 2017). Data processing levels provide a standardized way of defining the degree that QA/QC has been performed on a data set. Table 6.A.1 describes certification levels based on the documented quality assurance and quality control procedures applied to a data set.

**Table 6A.1**. Data processing certification levels based on the documented quality assurance and quality control procedures applied to a data set. (taken from Minimum Implementation Standards for Network Projects, Frakes and Kingston 2017).

