



Certification Guidelines for Inventory and Monitoring Data Products

Background

The primary goal of the Inventory and Monitoring Division (IMD) is to collect scientifically credible data for use in science-based management of NPS resources. It is also expected that data use will extend beyond parks and networks to include the broader scientific community. For these reasons, all data that are produced and published by IMD must include metadata that provides enough information that an end-user can independently determine the data's suitability for use, including methods for data collection, analysis, and the amount, types, and results of any quality assurance or quality control procedures performed.

Definitions

- **Quality Assurance (QA) procedures:** those implemented to ensure collection of data with minimal errors
- **Quality Control (QC) procedures:** those implemented to flag or modify data to ensure that data meet established data quality standards
- **Data Certification:** confirmation that all planned QA/QC procedures have been implemented in accordance with a quality assurance plan
- **Certification Status:** Description of the amount of QA/QC processing that has occurred

Quality Assurance Plans

Guidance is forthcoming from the IMD Central Office regarding detailed quality assurance plans (QAPs) that will be required for any protocols developed or revised after the guidance is issued (expected on or around May 1, 2016).

The requirement for QAPs, and associated changes in the designation “certified,” are aligned with broader changes being planned for the basic IMD data model. The revised model moves monitoring data away from a configuration of stand-alone databases, to a “pipeline” model that comprises the full data management lifecycle for a protocol, including applications for data collection; a central repository to cumulatively house, manage, and track data procedures and QA/QC; and tools for disseminating, reporting, and viewing data.

The revision of protocols and development of associated QAPs will be a phased-in process extending over multiple years. The guidance in this document applies to data that have been or are currently being

collected under legacy protocols that may not have the desired level of documented QA/QC procedures. This guidance will start moving IMD towards consistency in standards and terminology regarding data certification, and help ensure that all data users are fully informed of data provenance, processing, and suitability for use.

Data Processing Levels

Data processing levels provide a standardized way of defining the degree of data verification or transformation that has been performed on a data set. For the purposes of monitoring data certification, the following levels are based on the documented quality assurance and quality control procedures applied to a data set, and consider whether those procedures are defined in the protocol narrative and SOPs or in a QAP.

Table 1. Source of quality standards and allowable processing levels

Processing level	Source of Quality Standards	
	Standards documented in Protocol and SOPs	Standards documented in Quality Assurance Plan (QAP)
Raw	allowed	allowed
Provisional	allowed	allowed
Accepted	allowed	not allowed
Pre-certified	not allowed	allowed
Certified	not allowed	allowed

The following levels, raw and provisional, apply to data sets either with or without an associated QAP.

Raw

Data in their state of original collection, either in the form of paper or electronic field forms, electronic sensor files, or digital media (photo, video, or audio). These data generally have been collected following some QA procedures, but have not undergone any QC procedures.

Provisional

Data that have undergone all planned QA procedures, have been entered into the system of record, and have passed verification procedures to ensure that data in the system of record match source data (field forms, sensor data files, etc.). Provisional data have not undergone QC procedures.

The following level, accepted, applies only to data sets without an associated QAP.

Accepted

Data have undergone all QA and QC procedures as defined in the published protocol and associated SOPs.

The following levels, pre-certified and certified, apply only if a QAP is in place:

Pre-certified

Data within a system of record that have undergone defined minimum levels of QC as defined in the QAP, rendering the data suitable for use for a limited set of designated or defined purposes only. The QC procedures that have been performed are communicated clearly to the end user.

Certified

Data have been through all QC procedures as outlined in the published protocol and the associated published QAP.

- Data either meet the documented quality standards, or data that do not meet the standards are appropriately identified at the record level
- Procedures for data collection, processing, and transformation have been documented in appropriate metadata formats
- Data are suitable for use

Note: Only data that have an approved and published QAP can be designated as pre-certified or certified. When a QAP is in place, data review and certification take place at the record (i.e., observation) level as opposed to the data set level.

Network Guidelines

These guidelines apply to all IMD protocol data sets that do not yet have a QAP. These guidelines will be replaced once networks are managing their monitoring data using the “pipeline” model (see above) with an associated QAP and web services enabled.

1. Data sets that have already been designated as “certified” and that do not have a QAP, should be re-designated as raw, provisional, or accepted based on the level of QA/QC described in the protocol.

To comply with Director’s Order 11B, there must be a description of the overall data quality of the data set. Documentation should be a metadata file (text or pdf format only) that accompanies the associated data set. This description must indicate whether data were estimated as opposed to measured. Because of the wide variety in current QA/QC procedures across IMD protocols and data sets there are no templates for documenting data quality; however, enough information must be provided to the end user so that he or she can independently determine data that are suitable for use.

2. Monitoring data, at the highest level of processing, must be published to the IRMA Data Store either annually, or at the end of a data collection cycle as specified in the protocol.

These data can be published in the Data Store in one of two ways:

Option 1 (preferred): a cumulative data set and metadata file that contain all data collected (e.g., 2010-2015 data). If this option is used, create a Data Store record for the first version of the data set, and then use the Data Store versioning tool to create a new Data Store record for each updated version of the data set. There should be one data set file per reference. A data set file consists of the data set and associated metadata in a single zip file.

Option 2: multiple data sets that each contain data for a specific data collection period (e.g., 2011, 2012, 2013 data). If this option is used, create one Data Store record for each zipped protocol data set and metadata file (e.g., one reference for the 2011 data set, one reference for the 2012 data set, etc.). These should then be grouped together by a Data Store project record.

The following Data Store guidance applies to both Option 1 and Option 2:

- Upload only one version of a data set per Data Store reference. In general, this also means there is only one file per reference.
- Recommended reference type: tabular data set or geospatial data set (as appropriate).
- Once uploaded to a reference, digital files may not be removed or modified, and no files should be added at a later date.
- The *Date of Issue* is the date the data set was finalized.
- The *Content Begin Date* and *Content End Date* fields should match the field data collection dates for the data set.
- Once a data set has been published (i.e., the reference has been activated), neither the data set nor the metadata file may be altered or removed. A new reference version is the only way data changes can be published. The Data Store reference record, however, may be edited.
- In addition to versioning, any changes made to published data must be identified and communicated to end users. At a minimum, a narrative explanation should be included in the data set's metadata file that summarizes what changed, when, and why. Ideally, a table within the database will also track the field, the original value, the new value, when the change was made, who made the change, and why the change was made.
- All data sets must have an accompanying metadata file that describes how data were collected, QA/QCed, and (if appropriate) transformed. If data have been changed through versioning (using either option 1 or 2 above), the entire change history must be included in the metadata file, not just changes between single versions.
- Include, in the *Description* field of the reference, links to relevant versions of the protocol narrative and SOPs used for data collection, processing, and/or transformation.

- Enter the data set processing status and definition in the *Use Constraints* fields with the following text: “The processing status of this data set is [status] - [definition]” (e.g., “Provisional - Data have undergone all planned QA procedures.”)
- Published data must be at the highest processing level (i.e., accepted or certified) if used for reporting or disseminated to the public. Data at the pre-certified level may disseminated to parks to support science-based decision making in limited cases where uses and minimum QC procedures are defined.
- Regardless of the processing status, all I&M monitoring data sets should be marked as “High” in the *Quality Evaluation* field.

3. Links to monitoring data published in the Data Store should be included as part of Annual Administrative Work Plan reporting. If data have not been published, an explanation should be provided.

Data should be published to the Data Store within one year of data collection¹

4. Sensitive data and information holds

Data sets that contain sensitive information should be published in redacted form, or in total with limited access. In the Data Store, the reference should always be visible to the public; however, the associated data sets can be limited to internal distribution to appropriate audiences. Access should not be restricted beyond what is required—data must be available for science-based decision making by park personnel.

Data sets (sensitive or not) may be published to the Data Store and restricted to “internal use only” for a period of up to two years. This provides time for IMD staff to professionally publish on the data before their public release.²

¹ Note that all data should undergo QA/QC procedures to a state of provisional within one year of data *collection*, but we realize this may not possible in cases where lab or post-processing of data is required.

² This is subject to change--we are still determining an appropriate holding time. Note, however, that any data used as the basis for published reports or park management decisions are subject to FOIA even if not yet publicly released.