Interagency Ecological Program

Digital Datasheet Best Practices

version 2.0

Status: Effective

Effective Date: 11/9/2023

Interagency Ecological Program
for the
San Francisco Bay-Delta Estuary

A Cooperative Program of:

California Department of Fish and Wildlife

California Department of Water Resources

National Marine Fisheries Service US

State Water Resources Control Board

US Army Corps of Engineers

US Bureau of Reclamation

US Environmental Protection Agency

US Fish and Wildlife Service

US Geological Survey

Contents

Acronyms & Abbreviations

CDFW – California Department of Fish and Wildlife
DWR – California Department of Water Resources
ID – Identification
IEP – Interagency Ecological Program
PST – Pacific Standard Time
QA – Quality Assurance
QC – Quality Control
SME – Subject Matter Expert(s)
SOP – Standard Operating Procedure(s)
USBR – United States Bureau of Reclamation
USFWS – United States Fish and Wildlife Service

Scope and Application

This best-practices document outlines general considerations and recommended procedures applicable to use of digital datasheets. This document applies to any IEP agencies and programs that are using digital datasheets for documenting data collection. The scope of this document is limited to development of data sheet templates, general use of digital data sheets, Quality Control (QC), archiving and data reporting.

This document is designed to provide guidance and recommendations to staff involved use of digital data sheets (primarily form-based software). There are many different software applications for digital data sheets, and these will vary based on program and agency. Therefore, the recommendations in this document are kept high level for the broadest application. It is intended for users of this document to take what recommendations are most applicable to the digital data sheets in use in their respective program and apply as appropriate. Not all recommendations included in this document may be applicable or appropriate to every application.

Personnel Recommendations

- **Digital Datasheet development-** Staff may require support or training in developing templates for digital datasheets. Based on software, this may require some initial support from IT, technical support from the software company, or additional skills in programming languages such as R, SQL, or Python.
- **Data Entry & Editing** - New staff should be trained by subject matter experts (SME) and/or crew leads in proper data recording, entry, and review procedures. Data entered by new staff should undergo more thorough QA/QC by crew leads and other more experienced staff until they are fully trained and comfortable performing the tasks with less oversight. All data entry must be QC'ed by staff other than the staff that entered the data unless alternate staff are not available. If possible, the QC staff should be a field crew member who has historical experience in the environment and may be better able to gauge erroneous data entry.
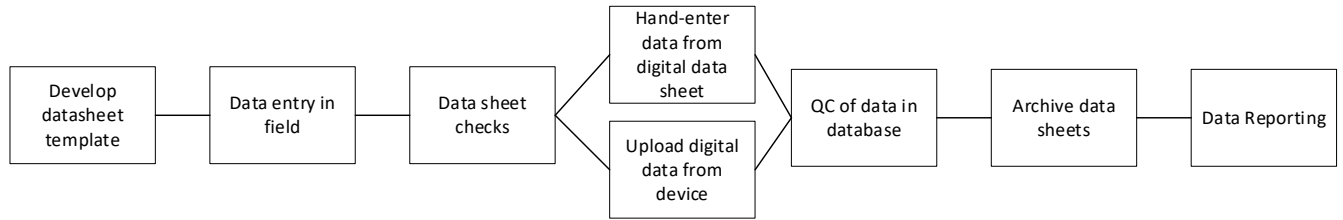
Figure 1 Digital Datasheet workflow

Digital Data Sheet Templates

The following best practices are recommendations in developing templates for ease of use both in the field and in QC.

- If possible, engineer so that the data can go directly from the instrument to the database itself. This eliminates to potential for data entry error by the user.
- If the software allows, build in reporting requirements
  - Specific units or formats, such as Pacific Standard Time (PST) vs watch time
  - Requirements for number of digits recorded, such as requiring the tenths place for dissolved oxygen
  - Automated errors if there are missing data, or erroneous data (such as negative values, out of possible range for pH or GPS)
  - Consider locking down fields for specific formats (such as only numeric for numeric data types, or specific date and time formats)
  - Consider limiting the number of characters in specific data fields
- Enable the ability to edit the datasheets in the field/office (prior to entering to database and archiving). It is highly recommended to have an audit trail (version control) so that the changes made to the digital data sheet over time are tracked.
- Set up templates in the order of field process. For example, if data environmental condition data is collected first (such as weather observations and tide), then include those fields at the top of your sheet.
  - For water quality data collected by sondes, it is highly recommended to order the water quality data fields in the same order as the sonde display.
- In data fields, include drop downs organized by the most frequently used terms on top. This is recommended for categorical data such as species, indices, or codes. See Appendix I in IEP Fish Data QC BP for additional examples.
  - For efficiency in collection of fish species data, consider having designated fields for the most common fish collected, and additional fields with drop downs for the less frequently collected species.
- For resources on data fields that should be included in digital data sheet templates, and examples of field data sheet templates, see the recording data section and appendices in IEP Fish Data QC BP.

- Include a checkbox for doing the QC check so that it is known if the data have been preliminarily reviewed in the field (mostly for completeness- all notes entered, all data entered.)
- Organize digital datasheets to keep data from the individual sampling locations and runs together in order, and easy for others in your program to locate.
- Beta test the template datasheets prior to using them in production.

Data Entry

Data Sheet Best Practices

The clarity and usability of a data sheet can be enhanced by keeping the following in mind:

1. **Clarity:** When filling out a data sheet, ensure all information is clear enough so someone who is not familiar with past conditions can interpret the data accurately (i.e., no slang, abbreviations, or omitted data).

2. **No Data and Zeroes:** If data are not recorded for a relevant parameter enter "N/A" in the data field in permitted by the data sheet or database or leave blank and write an explanation why the data were not taken in the comments section. If using a term other than "N/A" be sure all staff are consistently using that term and that use of the term is identified in protocols and the metadata. If no fish were caught, indicate with a code or description that sample was taken, but no fish were caught. It is important to notate this distinction, as zero catch data are used in effort calculations, while a non-sampling event would not be used in effort calculations.

3. **Comments:** Comments should be written so that they are specific and are easily understood. Use the comment field to document any unusual situations or changes to the SOP.

4. **Corrections:** Corrections to a data sheet and justification for the corrections that are made in the field should be explained in the comments section. Include an initial and date in the comments if there is no version control that tracks who and when changes were made.

5. **Disagreements:** Determine as a program how to resolve disagreements or uncertainty in data types, such as fish identification. Use field guides and send photos or bring fish back to experts for identification. Record any counts or measurements. Notate uncertainty of any data parameter on the data sheet, and once identified, modify species code on data sheet and make a note in the comments with initial and date.

6. **Estimated information**: It is not recommended to estimate information, but in special cases it may be useful. A few examples are: estimating fork length of Delta Smelt since they can't be handled, estimating depth in shallow waters, estimate the last digit for water quality readings if the sensor is fluctuating. If any estimated data is collected, it needs to be clearly documented in the comments.

Database Data Entry Best Practices

1. Digital datasheets should be saved to network drives as soon as reasonably possible from returning from the field.

2. Digital datasheets that can be directly uploaded to databases should be uploaded as soon as reasonably possible after returning from the field. Ideally within a week or following completion of sampling.

3. For data that is hand entered from the digital datasheet to the database:

    a. If data are missing for a field on the field data sheet enter N/A in the data field if allowed by the database. If this is not allowed, leave the field blank and make a note in the comments field. Include an initial and date if the database does not have version control. Be consistent if recording start and end values (e.g., flowmeter counts) for approach when missing a value. For example, if missing a start flowmeter value, you could enter the value of "N/A" and the actual end meter or you could enter both as N/A on the data entry screen. Your approach will be informed for how you contend with missing data, especially from a calculated field (e.g., end flowmeter count – start flowmeter count = total counts).

    b. A single leading zero is recommended to be typed in appropriate numeric fields, providing that a decimal point is used where appropriate.

    c. If there is a comment on the field data sheet, type it into the comment data field on the data entry screen. The only exception is if the comment is not relative to the study, such as a comment that relates to a different study at the same site.

    d. *Catch detail*. For ease in QC, enter fish data, notably fish length values, in the data entry screen in exactly the same order as it appears on the field data sheets, with the exception of plus counts. It is recommended to enter plus counts last.

Quality Control /Quality Assurance (QA/QC)

Digital Data Sheet Checks

Once the digital data sheets are completed, they should be checked by someone (other than the data recorder) who is trained on the process prior to leaving each field site. Some of these checks may be built into the digital data sheets. Examples of these automations are historical range checks (such as temperature range, or species names), theoretically possible range checks (such as bounds for pH or GPS), requiring paired data (such as a beginning and end flow meter reading), and data types (date versus numeric). These automated checks may also be categorical, such as if only accepting specific station codes or names. Whether automated, or done manually, these checks should include, but are not limited to:

1. Missing data (blanks on the data sheet)

2.   Plus counts not calculated or recorded
3.   Incorrect species codes
4.   Incorrect station codes
5.   Data entered in the wrong place
6.   Addition or transfer errors on the summary sheet
7.   Inconsistencies (time, meter and/or gear serial numbers, temperature)
8.   Sample bias (if fish lengths have a pattern such as decreasing, it can indicate that the fish subsampled from large catches were not chosen at random)
9.   Flow meter readings copied incorrectly from one tow to the next
10.   EC meter units not defined

Document any errors found in the datasheets and how they are corrected. Include an initial and date if the datasheet lacks version control that tracks who and when changes are made.

If time and staffing constraints allow, "real time" data QC checks may be done. For these QC checks, crew members conduct real time data entry each day or following day from digital datasheets that are uploaded or saved to network drives. It is highly recommended that these staff be experienced staff. Crew performing "real time" entry immediately address any abnormalities to field crew leads or supervisors. Crew entering data should ask field and lab crew about any questions while the data are still recent and document any changes made to the digital data sheets.

Data Entry checks

Data entry checks are critical to maintaining a high-quality dataset with minimal errors. Many of these checks can be performed with Microsoft Excel, Microsoft Access, GIS software (e.g., ArcGIS), and/or a programming language such as R. If feasible, it is ideal to upload the data from the digital datasheets directly to the data base. In lieu of that capability, the data are first entered into database forms that should have customized error-checking and data validation checks if this is not already done through the digital datasheets (*e.g.*, required fields that must be filled, number formatting rules, etc.). This helps ensure that simple entry errors are avoided. A second staff member compares the data in the database to digital sheets to ensure accuracy. Then, each data field is sorted and/or summarized based on unique records to highlight missing values and outliers for inspection.

Double-Data Entry

Another approach for ensuring data entry accuracy is double data entry. This process works best for digital data sheets that still require hand entry into the database. In this process, the same data are entered from the field sheets into the database twice. After double entering the data, compare the duplicate data entries. This step can be accomplished via Access queries or R code. Performing this step first will save time later by reducing the number of errors present during later QA/QC processes. This process follows the following steps:

1. Compare the tables from the 1st and 2nd entries.
2. If discrepancies are detected, consult the field data sheet to find the correct value.
3. If the field datasheet is wrong or unclear, it should also be corrected with a note or annotation, even in the copy in the digital archives. The incorrect data value should NOT be removed from the datasheet, but the correct value should be indicated via an annotation.
4. Additional QC checks can be done using an R script. From this, a list of errors can be generated in the station data and catch data that need to be corrected.

QC Following Data Entry

For QC of data once it has been entered into the database (by automatic upload or hand entry), see the "QC Following Data Entry" section in the IEP Fish Data QC BP document.

Record Keeping/Archiving

Digital field sheets should be archived on network drives. Consistent file naming conventions within a program should be used for digital datasheets for ease of access.

A Data Management Plan should be created to document how the data is being managed. This should include information about where the data is backed up, and frequencies of backing up and archiving.

Data Reporting

Include information on data reporting processes, both internal and external. If there are requirements for how frequently the data needs to be reported, include those here. For compliance with AB1755, include information on the process and frequency of posting to CNRA data portal. If you are posting your data to the Environmental Data Initiative (EDI), include the appropriate references for connecting the EDI datasets to the CNRA Open data Portal which are available through the DUWG Github.

Hyperlinked URLs Within The Document

IEP Fish Data QC BP - https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=212668&inline

CNRA data portal – https://data.cnra.ca.gov/

Environmental Data Initiative (EDI) – https://edirepository.org/

DUWG Github - https://github.com/InteragencyEcologicalProgram/Open-Data-and-Data-Publishing/blob/master/Resources/Connecting%20EDI%20to%20CNRA_20200626.docx

Revision History

Revision	Effective Date	Section	Description of Change	Justification of Change
1.0	9/1/2023	All	New document	Creation of new document
2.0	11/9/2023	Data Reporting, Hyperlinked URLs Within the Document	EDI link updated	The link in the original document was not directing to the correct webpage

Past Editors & Collaborators / Acknowledgements

First Name	Last Name	Organization	e-mail address	Author/ editor
Rachel	Pisor	DWR	Rachel.Pisor@water.ca.gov	Author
Traci	Treleaven	DWR	Traci.Treleaven@water.ca.gov	Author
Sonia	Miller	DWR	Sonia.Miller@water.ca.gov	Collaborator
Morgan	Battey	DWR	Morgan.Battey@water.ca.gov	Collaborator
Jonathan	Speegle	USFWS	Jonathan_Speegle@fws.gov	Collaborator
Nicholas	Bertrand	USBR	nbertrand@usbr.ca.gov	Collaborator
Emma	Davidson	CDFW	Emma.Davidson@Wildlife.ca.gov	Collaborator
Daniel	Ellis	CDFW	Daniel.Ellis@wildlife.ca.gov	Collaborator