Data Review Checklist

IP Number: IP- Senior USGS Author:

Title:

Data Reviewer’s Name:

Your Title, Affiliation, and Location:

Your Review Rating: Approve Approve with minor changes Approve with Major Changes Reject

Taylor Stewart

IP-073304

Western Basin Trawl Survey, Lake Erie (2013-2015)

Brian C. Weidel

Research Fishery Biologist, USGS GLSC, Oswego NY

x

In April 2014, the Community for Data Integration (CDI) developed a checklist of data quality-check questions that a reviewer can use as reference. Not all questions are applicable or need to be answered; instead this is a guidance document. Included on this template is the checklist and space for the data reviewer to add their comments and recommendations. This document will be collected by the data author and placed in IPDS as part of the data release documentation. [[1]](#footnote-1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Review Checklist** ✓ | | | | | |
| **Questions to Consider:** | | **A** | **MI** | **MA** | **N/A** |
| Are data values reasonable? | | x |  |  |  |
| ⮚ | Are they in a valid range for that measurement? | x |  |  |  |
| ⮚ | Do they display seasonal or daily trends that are applicable? | x |  |  |  |
| ⮚ | Is there consistency between adjacent or otherwise related datasets, within the product? | x |  |  |  |
| ⮚ | Are the geographic locations given for the data, reasonable? (i.e., are ocean data points actually showing in the ocean?) |  | x |  |  |
| ⮚ | Is the accuracy claimed for the data reasonable? | x |  |  |  |
| ⮚ | Are “no data” values accurately defined? | x |  |  |  |
| ⮚ | Are data anomalies or gaps explained in the metadata? | x |  |  |  |
| ⮚ | Do analysis values add up? | x |  |  |  |
| Are the data development methods scientifically sound and well described? | |  | x |  |  |
| ⮚ | Could a scientist or technician recreate the final data set from the descriptions? |  | x |  |  |
| ⮚ | Can the documentation about methodology be easily found and used? | x |  |  |  |
| ⮚ | Are processing software and versions identified? | x |  |  |  |
| Does the product include metadata in a standard endorsed by the Federal Geographic Data Committee (FGDC) for all the data, such as the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) or ISO 19115? | |  | x |  |  |
| Are the coordinate system and datum defined appropriately (both horizontal and vertical)? | |  |  | x |  |
| Does the Product as a whole, through its design or documentation, provide enough information so that the data and metadata can be easily found and used? | |  | x |  |  |
| ⮚ | Is the format of the file identified? | x |  |  |  |
| ⮚ | Is there information about the software required to use the data? | x |  |  |  |
| ⮚ | Is the location of files documented? | x |  |  |  |
| ⮚ | If the data are released on a web page, does the page have useful discovery metadata, for example the web page clearly identifies the contents, keywords and metadata tags are provided, and geospatial attributes are presented? | x |  |  |  |

# Reviewer’s Comments and Recommendations:

Can’t change but had to note:

As a contributor, maintainer, and user of the entire Great Lakes Science Center’s bottom trawling data archives I was disappointed to see this data was not presented or organized in the same table structure and format (RVCAT, Research Vessel Catalog). As one who is familiar with that RVCAT format and now familiar with these Lake Erie data, I can assure these authors that the RVCAT will more than adequately represent these data and in doing so our collective understanding of Great Lakes fishes would be improved. Ok, soap box over, as is the data are important enough to serve that I am willing to overlook this issue and hope that the authors will move to the RVCAT structure in the future.

Changes

Borer: I suggest the authors refer to the Borer et al. 2009 paper in the ESA bulletin that addresses simple rules for data management. They lay out a number of common sense practices, that if followed, help all of science share their data collaboratively. For instance, I’d suggest Stewart et al. consider their column naming strategy and advise they pick a ‘strategy’ and then stick to it. For some columns a unit is provided as “\_unit” sometimes no “\_”, sometimes no unit. Also the capitalization etc, if standardized would help the user.

Catch: in reality the data provided has been analyzed already, in that a total area swept must have been estimated and then catch divided by this unknown data point, to provide a kilograms per hectare. It would be better to provide the number caught, weight of those fish, and the area swept, with some way of determining how that area swept was calculated. In addition, the authors should note whether they include a term for catchability or not ask (Kocovsky and Stapanian, 2011) illustrate how this number might change in Lake Erie.

DateL I found “season” to be a fairly useless an objectively defined term and suggest including a column for dateSampled

In this day in age of spatially-derived questions, I was almost shocked the data did not include any kind of spatially oriented data. Per the checklist: *Are the coordinate system and datum defined appropriately (both horizontal and vertical)* . This suggests at a mimimum Lat, Long, Datum and some indication of fishingDepth.

For depth factor consider using multiple columns represent ‘binMaxDepth’ and ‘binMinDepth’ rather than representing the data as a range/list (0,1]

Small stuff:

In a few of the files the csv contain serial row numbers/names and it is difficult to tell if these are actually referring to something or just an artifact of how saving. (if using write.table you could consider setting row.names=FALSE when saving the csv )

1. \*Note that there is a separate document for metadata review, but that both the data review and metadata review may be completed by the same person depending on supervisorial approval. For more information, review the [checklist description](http://www.usgs.gov/datamanagement/documents/DataReviewChecklist_2014.pdf) provided by the CDI. [↑](#footnote-ref-1)