**DRAFT Checklist for NPS Data Package Review**

(last updated 04/05/2024)

Intro text sharing the value of data packages and where to find more information and guidance:

<https://doimspp.sharepoint.com/sites/nps-nrss-imdiv/data-publication/SitePages/Data-Package-Specifications.aspx>

The role of the data and metadata reviewer is to ensure the accuracy, completeness, and usefulness of the entire data package. This process should enlist the help of a colleague who can step through this checklist. Note that this checklist provides general guidelines and can be customized, as necessary, for use in specific NPS programs, parks, or offices. A good reviewer should be knowledgeable about the metadata standard being used, and should be given ample time to review a metadata record. Starting the process early is key.

**Title**: Invasive Plant Monitoring for the Heartland Inventory and Monitoring Network: 2006-2022 - Data Package

**DataStore Reference**: Data Package (Code: 2303109)

**Author’s Special Instructions**:

**Reviewer’s Name**: Alexa Ron

| **#** | **Yes** | **No** | **Item** |
| --- | --- | --- | --- |
| 1 |  |  | Many automated checks have been built into the [NPSdataverse DPchecker](https://nationalparkservice.github.io/DPchecker/) tool. Have these been run on the data package and warnings or errors been dealt with? DPchecker version \_\_\_\_\_\_\_\_\_\_\_\_\_\_  Comments: |
| 2 |  |  | Is the data package comprised of only CSV data files and an XML metadata file (thus meeting machine-readable and open format goals)?  Comments: |
| 3 |  |  | Is the title appropriate and sufficiently descriptive (the title conveys the “what, where, when, and (if applicable) scale” of the data)?  Comments: |
| 4 |  |  | Does the abstract provide a short statement that allows users to quickly evaluate the content of the dataset?  Comments: |
| 5 |  |  | Are QA/QC methods appropriate and adequately described and/or referenced?  Comments: DRR would need to be reviewed |
| 6 |  |  | Are the sampling and data-collection methods adequately summarized and referenced such the data can be understood by a 3rd party?  Comments: |
| 7 |  |  | Does the metadata contain definitions of the values in each field such that a 3rd party can understand them?  Comments: verbatimCoordinates is defined as “The *taxonomic* coordinates as it appeared in the original record.” What are taxonomic coordinates?  Reply: It should be “The verbatim original spatial coordinates of the LocationID.” and has been corrected. |
| 8 |  |  | If missing values are present (i.e. data are absent), are appropriate missing value codes included and defined in metadata?  Comments: |
| 9 |  |  | Do filenames provide context, are they reasonably clear, and do they use a consistent naming convention (or example, ‘MeasurementA\_Location1\_Date.csv’, ‘MeasurementA\_Location2\_Date.csv’, ‘MeasurementB\_Location1\_Date.csv’,)  Comments: |
| 10 |  |  | Do column labels provide context, can they be interpreted by humans, and are they consistently formatted?  Comments: |
| 11 |  |  | Does each column contain only one data type (excluding missing value codes) or one data format? For example, date formats should be consistent within a column; numeric data only contains numbers, etc.  Comments: |
| 12 |  |  | Across filenames and column names are capitalization and punctuation used consistently (and where applicable follow standard conventions)?  Comments: |
| 13 |  |  | Have unused columns and rows been removed? For example, .csv files do not have trailing commas at the end of a row.  Comments: |
| 14 |  |  | Do all URLs work and point to the expected web page?  Comments: |
| 15 |  |  | Do keywords accurately represent the data and include terms from standard vocabularies whenever possible?  Comments: The keywords that are in the title seem redundant, as does the network name and abbreviation (a user can select the network in a quick search)  Reply: I will get them removed. |
| 16 |  |  | To the best of your knowledge, If the data package is intended for a public audience, does it have protected data that needs to be removed prior to publication?  Comments: I assume invasive plants would not be protected? Reply: You are correct. |
| 17 |  |  | To the best of your knowledge, does the data package **metadata** (.xml) refrain from sharing any proprietary or protected information, including PII? Metadata files – even for non-public data - will eventually be shared with public audiences.  Comments: |
| 18 |  |  | Do the geographic coordinates in the data and metadata match the locations in titles, keywords, and other text?  Comments: |
| 19 |  |  | Does the data package (metadata and data) have an appropriate level of professionalism (comments are relevant to the protocol and the data; no derogatory, negative, or inflammatory language)?  Comments: In the “Notes” section on DataStore, the sentence, “Coded values for CoverClassInMetersSquared available on page 10 of the aforementioned 'Problematic Plant Monitoring Protocol for the Heartland Inventory and Monitoring. 2022' and duplicated in the tlu\_CoverClass.csv look-up table.” could be rewritten as “Coded values for CoverClassInMetersSquared *are* available on page 10 of the aforementioned 'Problematic Plant Monitoring Protocol for the Heartland Inventory and Monitoring. 2022' and duplicated in the tlu\_CoverClass.csv look-up table.” And “This correction was made with the approval of HTLN's Haack-Gaynor JL.”  Similarly, the sentence, “Added a missing decimal to the LongitudeInDecimalDegrees for Hopewell Culture National Historical Park.” could be rewritten as something like, “During data processing, a missing decimal was added to the LongitudeInDecimalDegrees coordinate for Hopewell Culture National Historic Park.”  But not a big deal! |

Additional Reviewer Comments:

Great job!! Looks so clean. Where is the tlu\_CoverClass.csv look-up table that is mentioned in the “Notes” section on DataStore? Maybe add instructions on where that can be accessed

Reply: The cover class codes were integrated directly into the values of the CoverClassInMetersSquared column in the dataset, so I have amended the note to make that more clear with the following:

“Code values for CoverClassInMetersSquared (available on page 10 of the aforementioned 'Problematic Plant Monitoring Protocol for the Heartland Inventory and Monitoring. 2022' and duplicated in the tlu\_CoverClass.csv look-up table) were directly integrated into the column variables to clarify the meaning of each value.”

Last question – were quality flags left off at the request of the network?

Reply: Yes, Jen opted for no flag columns.