

HDX Dataset Quality Assurance Checklist

Background

HDX quality assurance (QA) officers should check each new or updated dataset against this checklist to ensure datasets shared on HDX meet the minimum HDX quality standards for datasets. The checklist is in three parts: a) a checklist for dataset resources, b) a data responsibility checklist and c) a metadata quality checklist. Each of these checklists are presented below.

A. Dataset Resources Checklist

Check each resource in the dataset against the following checklist

Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The download link is working	-Place the resource under review. -Engage the data contributor about the broken link.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The download link is not password protected	-Place the resource under review. -Engage the data contributor about the password protected file.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The file is in useful data format e.g. XLS, CSV, SHP etc and not PDF or Word unless the inaccessible file contains additional metadata.	-Place the resource under review. -Engage the data contributor about the inaccessible data format.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	There is NO PII in the data file	-Place the resource under review. -Immediately complete the data responsibility checklist.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	There is NO DII in the data file	-Place the resource under review -Immediately complete the data responsibility checklist
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	There is NO individual survey data in the data file	-Place the resource under review. -Immediately complete the data responsibility checklist.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The file type is consistent with the file's actual file type	-Attempt to fix the inconsistency and engage the data contributor if attempts unsuccessful.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The file contains a zipped shapefile and the file type is correctly set as "zipped shapefile"	-Attempt to fix the issue so geo preview can work on the dataset.

B. Data Responsibility Checklist

Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>		The dataset is safe to share and is not potentially sensitive.	- Place the dataset under review - Inform the data contributor - Report the dataset for SDC review
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the dataset is potentially sensitive; I have placed the dataset under review	-Place the dataset under review
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the dataset is potentially sensitive; I have informed the dataset contributor by email that the dataset has been placed under review.	-Inform the data contributor using the email template
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the dataset is potentially sensitive; I have made an entry in the SDC for Microdata Log	- Make an entry in the SDC for Microdata Log for the dataset
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the dataset is potentially sensitive, I have informed the HDX SDC taskforce to kick off the disclosure control process	- Inform the HDX SDC taskforce about the dataset

C. Metadata Quality Checklist

Title Checks				
Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>		The dataset has a title	-Make the dataset private -Engage the data contributor about the missing dataset title.
<input type="checkbox"/>	<input type="checkbox"/>		The dataset title is consistent with the data actually contained in the dataset.	-Engage the data contributor about the inconsistent title. Suggest a title if possible.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The dataset title is not in camel case	-Edit the dataset title into camel case
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The dataset title does not start with the name of the (place) country, region or place to which the data refers.	-Edit the dataset title so it starts with the place name -Inform the data contributor of edit(s)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The dataset title does not includes a relative time reference such as current, latest, new, old	-Edit the dataset title and remove the references. -Inform the data contributor of edit(s)
Description Checks				
Yes	No	N/A	Check	Action if check fails

<input type="checkbox"/>	<input type="checkbox"/>		The dataset has a description	-Make the dataset private -Engage the data contributor about the missing description metadata.
<input type="checkbox"/>	<input type="checkbox"/>		The description accurately describes the data in the resources.	-Engage the data contributor to edit the description
<input type="checkbox"/>	<input type="checkbox"/>		The description is free of personal information like the name of a contact or their email address. Generic email addresses and academic citations are fine.	-Make the dataset private -Engage the data contributor to remove the PII from the description.

Source Checks

Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>		The dataset has a source	-Make the dataset private. -Engage the data contributor about the missing source metadata.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the source is listed as Multiple Sources; the multiple sources are clearly listed elsewhere in the dataset (such as in the description, comments or additional metadata file)	-Make the dataset private -Engage the data contributor about the missing source metadata.
<input type="checkbox"/>	<input type="checkbox"/>		The source is a valid source for the data	-Make the dataset private. -Engage the data contributor about the invalid source.

Date of Dataset Check

Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>		The dataset has a Date of Dataset	-Make the dataset private -Engage the data contributor about the missing date of dataset metadata.
<input type="checkbox"/>	<input type="checkbox"/>		The Date of Dataset is NOT a future date	-Engage the data contributor to check whether the date of dataset is indeed in the future.
<input type="checkbox"/>	<input type="checkbox"/>		The Date of Dataset is consistent with the freshness of the data. In other words, the data is as up-to-date as is suggested by the date of dataset.	-Engage the data contributor to check the out-of-date data.

Expected Update Frequency Checks

Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>		The dataset has an Expected Update Frequency.	-Make the dataset private. -Engage the data contributor about the missing Expected Update Frequency metadata.

<input type="checkbox"/>	<input type="checkbox"/>		The data is fresh according to the Expected Update Frequency.	-Engage the data contributor about the non-fresh data.
Location Checks				
Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The dataset has at least one location	-Make the dataset private -Engage the data contributor about the missing location metadata.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The location(s) specified in the dataset correspond to the location(s) that the data actually refers to.	-Fix the locations if possible. -Inform the data contributor of edit(s)
Licence Checks				
Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>		The dataset has a licence	-Make the dataset private. -Engage the data contributor about the missing Licence metadata. Suggest CC-BY.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the license is "OTHER"; the text in the LICENSE OTHER field describes clear and meaningful licence terms.	-Make the dataset private. -Engage the data contributor about the missing licence terms or request the data contributor to clarify any unclear licence.
Methodology Checks				
Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>		The dataset has a methodology	-Make the dataset private. -Engage the data contributor about the missing methodology metadata.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the methodology is "OTHER", the text in the METHODOLOGY OTHER box describes a clear and meaningful methodology.	-Make the dataset private. -Engage the data contributor about the missing methodology or request the data contributor to clarify the methodology.
Caveats/Comments Checks				
Yes	No	N/A	Check	Action if check fails
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The Caveat/Comments text is clear and meaningful	-Engage the data contributor about the unclear comments/caveats metadata.
Sub-national Flag Checks				
Yes	No	N/A	Check	Action if check fails

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the dataset contains subnational data, does the dataset have the SUB-NATIONAL flag	-Add the SUB-NATIONAL flag to the dataset. -Inform the data contributor of edits.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If the dataset does NOT contain sub-national data, does the dataset NOT have the SUB-NATIONAL flag	-Remove SUB-NATIONAL flag from the dataset. -Inform the data contributor of edits.

For more information

Should you have any questions, comments, or concerns, please contact:

- The HDX team at hdx@un.org

Last updated 23 November 2020