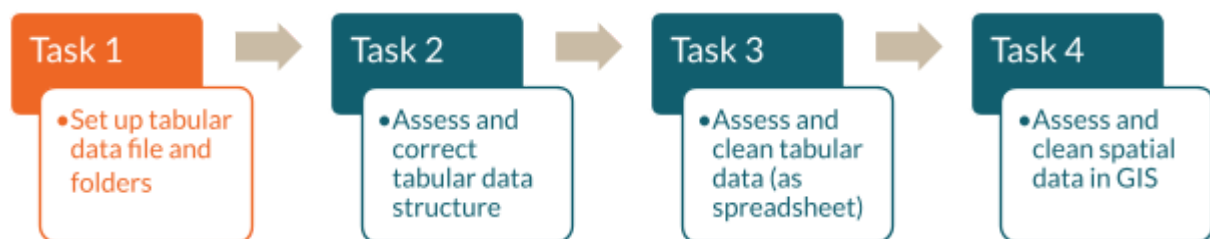


GIS DATA PREPARATION AND MANAGEMENT

Practical Task 1: Initial data setup

Task description

This is the first of four practical tasks that collectively constitute a “data preparation and cleaning” workflow. The objective of this task is to understand best practices for the initial setup of your data to facilitate effective review and allow for any necessary corrections.



Once you have downloaded, or received, a raw data file, the first step is to a) save a copy of the data in a readily accessible, fixed directory on your computer and b) adapt a consistent and standardized naming convention. This will help other users (and yourself in the future) understand the version of each data file. It is also very important to check/create metadata to describe the data.

Data rules covered in this task:

- Always **save a copy** of the raw data before making any edits
- Any **updates or changes** to a data file should be recorded in the **metadata**
- If a data file comes to you with **no metadata**, you should **create** metadata with as much detail as possible



Part 1: Initial file management

For this course, you have been provided with an existing folder structure, which was determined and developed by your data team to ensure everyone working with the data knows where to find the information they need.

1_SourceData

This folder serves as the point of entry for all data and information related to the project. Data files are placed here before any processing has been done. NO edits should be made to files in this folder.

2_RawData

Before making any edits to the data file, make a **copy** of the original file from the 1_SourceData folder and save it in this 2_RawData folder.

3_IntermediateData

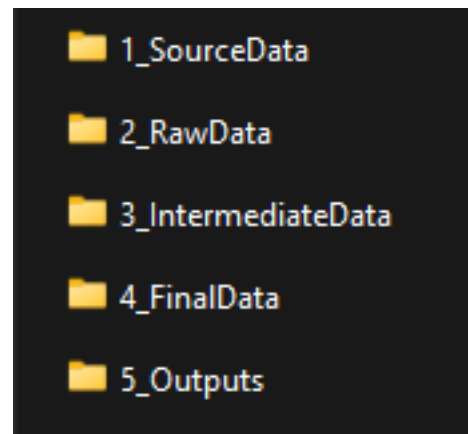
Any edited versions to the data files from the 2_RawData folder should be saved in this 3_IntermediateData folder.

4_FinalData

Once the spreadsheet is ready to be used elsewhere, it can be saved in this folder with a version number.

5_Outputs

This folder is for files created from the spreadsheets, such as maps, graphs, and statistical reports.



File management example

Step 1: Open the following spreadsheet:

...\\GRID3DataMng\Data\11_SourceData\NGA-HFs Kano LGAs sample.xlsx

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2	uniq_id	timestamp	latitude	longitude	wardname	wardcode	lganame	lgacode	statename	statecode	updated_on	func_stats	category	ownership	type	source	prmy_name
3	8345	18/12/2018	12.13267	8.32125	Marke	KN1009	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Primary Health Center	State Primary Healthcare Development Agency	Primary	eHA Polio	Kunya Primary Health Care
4	8370	18/12/2018	12.03477	8.39608	Danguguwa	KN1001	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Primary Health Center	State Primary Healthcare Development Agency	Primary	eHA Polio	Babawa Model Primary Health Center
5	8375	18/12/2018	12.09308	8.30719	Dawaki West	KN1003	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Primary Health Center	State Primary Healthcare Development Agency	Primary	eHA Polio	Gayawa Primary Health Center
6	13041	15/11/2019	12.05056	8.37907	Danguguwa	KN1001	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Primary Health Center	Private	Primary	eHA Polio	Yada Kunya Leprosy Hospital
7	13104	15/11/2019	12.14019	8.24805	Marke	KN1009	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Zagon Mata Health Post
8	13107	15/11/2019	12.13114	8.43112	Tumfafi	KN1011	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Dan Jaka Health Post
9	13108	15/11/2019	12.16253	8.487	Tumfafi	KN1011	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Yama Kanawa Dispensary
10	13115	15/11/2019	12.14638	8.49576	Tumfafi	KN1011	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Gambawa Health Post
11	13116	15/11/2019	12.13461	8.48474	Tumfafi	KN1011	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Doga Health Post
12	13119	15/11/2019	12.15612	8.44435	Tumfafi	KN1011	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Shawa Health Post
13	13124	15/11/2019	12.1017	8.31783	Dawaki West	KN1003	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Kuru Dispensary
14	13125	15/11/2019	12.09577	8.30629	Dawaki West	KN1003	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Zabairawa Dansudu Health Post
15	13126	15/11/2019	12.12022	8.3258	Dawaki East	KN1002	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Jirgawa Health Post
16	13127	15/11/2019	12.11772	8.35416	Kwa	KN1010	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Geranya Dispensary
17	13128	15/11/2019	12.10098	8.35663	Dawaki East	KN1002	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Kartama Babba Health Post
18	13129	15/11/2019	12.11782	8.37578	Kwa	KN1010	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Garfi Health Post
19	13130	15/11/2019	12.08283	8.33082	Dawaki West	KN1003	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Taskiya Dispensary
20	13131	15/11/2019	12.14852	8.3788	Gargari	KN1006	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Kwarkiya Health Post
21	13132	15/11/2019	12.15363	8.32747	Gargari	KN1006	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Danbawa Health Post
22	13133	15/11/2019	12.1375	8.40336	Dawanau	KN1004	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Tunkunawa Health Post
23	13134	15/11/2019	12.09304	8.36999	Danguguwa	KN1001	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Primary Health Center	State Primary Healthcare Development Agency	Primary	eHA Polio	Gandirawa Basic Health Center
24	13135	15/11/2019	12.10455	8.41731	Dawanau	KN1004	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Abakaya Health Post
25	13136	15/11/2019	12.13363	8.42455	Dawanau	KN1004	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Gurjiya Dispensary
26	13137	15/11/2019	12.1785	8.4041	Kwa	KN1010	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Sarbi Health Post
27	13138	15/11/2019	12.13388	8.37468	Kwa	KN1010	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Kankarawa Dispensary
28	13139	15/11/2019	12.1126	8.30309	Marke	KN1009	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Geza Health Post
29	13140	15/11/2019	12.12862	8.35319	Dawaki East	KN1002	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Garke Health Post
30	13152	15/11/2019	12.14206	8.46874	Tumfafi	KN1011	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Yumbu Dispensary
31	13151	15/11/2019	12.05604	8.41108	Danguguwa	KN1001	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Zango Dispensary
32	13156	15/11/2019	12.09229	8.40487	Dawanau	KN1004	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Wasai Health Post
33	13157	15/11/2019	12.15963	8.41997	Tumfafi	KN1011	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Zangi Health Post
34	13161	15/11/2019	12.098035	8.3175049	Dawaki West	KN1003	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Gwariji Health Post
35	13175	15/11/2019	12.04658	8.394779	Danguguwa	KN1001	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Gabari Health Post
36	13177	15/11/2019	12.196126	8.473915	jalli	KN1007	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Madaci Health Post
37	13216	15/11/2019	12.14704	8.2999	Gargari	KN1006	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Jemomi Health Post
38	13236	15/11/2019	12.07065	8.37619	Danguguwa	KN1001	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Primary Health Center	State Primary Healthcare Development Agency	Primary	eHA Polio	Koya Basic Health Center
39	13247	15/11/2019	12.07477	8.35915	Danguguwa	KN1001	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Sanbauna Basic Health Center
40	13249	15/11/2019	12.128773	8.271503	Marke	KN1009	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	State Primary Healthcare Development Agency	Primary	eHA Polio	Marke Health Post
41	13245	15/11/2019	12.1376957	8.3260754	Marke	KN1009	Dawakin Tofa	20021	Kano	KN	01/03/2019	Functional	Dispensary	National Primary Healthcare Development Agency	Primary	eHA Polio	Kaleku Health Post

This spreadsheet contains information on health facilities in eight local government areas (LGAs) in Kano State, Nigeria. The data have been collated from various sources and require cleaning before the data points can be mapped in QGIS.

Note: The data used for this course have been amended for training purposes.

Step 2: Immediately before proceeding, save a copy of the original file in the 2_RawData folder. It is very important that NO edits are made to the original file.

- **File > Save As**
 - Navigate to the course data folder on your computer
- **GRID3DataMng > Data > 2_RawData**
 - Save a copy of the data as an Excel file with the following name (change the date to reflect today's date):

Kano_HealthFacilities_31Mar2024_raw.xlsx

File naming tips

Keep file names concise and informative. Key information such as location (Kano), data type (health facilities), date, and version should always be included. Adopting a naming convention within your organization or team will promote effective data sharing and management.



When preparing data for GIS use, some geoprocessing tools or platforms will not work if the file names do not follow the correct format. Do not use spaces, hyphens, parentheses, brackets, or other special characters (such as \$, %, and #). Using underscores (_) instead of spaces solves this problem. This rule can also apply to the entire file path; folder names should not contain spaces or special characters, either.

Examples of **INCORRECT** file names

HealthFacilities_31032024

- ✗ Geographic extent of data not specified
- ✗ The date format (DDMMYYYY) may not be clear to all data users

NGA_Kano_HF data

- ✗ Not clear what "HF" means
- ✗ Space used between "HF" and "data"
- ✗ No date indicated

Kano.Helth_Facilities-31March2024

- ✗ Spelling mistake ("Health" not "Helth")
- ✗ Full stop (.) used may cause issues for geoprocessing tools in GIS
- ✗ Hyphens(-) used may also cause problems in some programs

Example of **CORRECT** file name

Kano_HealthFacilities_31Mar2024_raw

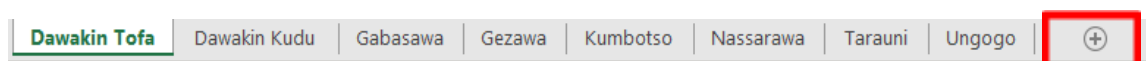
- ✓ Correct spelling
- ✓ Location and date indicated in the correct format
- ✓ No use of special characters



Part 2: Creating or updating metadata

Metadata are data about data. It is crucial to maintain clear, informative, and up-to-date metadata about all the data you work with. As well as providing crucial information about how and when the data was collected, metadata help identify known limitations of the data by recording known errors and accuracy issues. Complete and up-to-date metadata provide an audit trail of the data collection, processing, and analysis journey.

The raw data file currently has no recorded metadata, so you need to **create** the information from scratch. Create a new tab and call it *Metadata*.

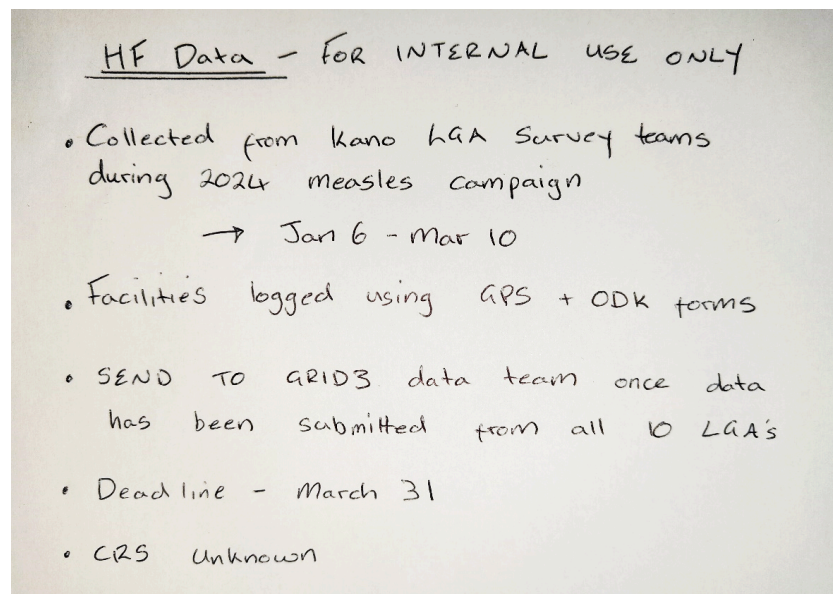


Create the following metadata headers in column A (you do not need to include the text in brackets):

- **Date created** (the date when this file was created)
- **Created by** (include name, job title, organization, and contact details)
- **Details** (information on the data included in the file)
- **Date updated** (date of any data updates or edits)
- **Edits made** (details of any edits made to the data)
- **Version** (include a version of the data file)
- **Data source** (indicates where the data come from; include details of primary data collection or a link to where the data were downloaded)
- **CRS** (indicates whether the data already have a known coordinate reference system)
- **Terms of use** (any licensing or data sharing restrictions)
- **Known errors** (include any known errors in the data)
- **Additional notes** (add any other information about the data that may be useful for others to know)



Fill out the metadata tab as thoroughly as possible using your existing knowledge of the data and the information on the note below, which was provided by a colleague.



Ensure you have included today's date and your name (and email address, if applicable). So far, no edits have been made to the data; those fields can be left blank (or use N/A).

The metadata tab should now look similar to the table below:

Date created	31 March 2024
Created by	A. Learner, Junior data scientist, GRID3
Details	Health facility data for 8 LGAs in Kano State, Nigeria (Dawakin Tofa, Dawakin Kuda, Gabasawa, Gezawa, Kumbotso, Nassarawa, Tarauni, Ungogo). Data collected between 6 January and 10 March 2024.
Date updated	N/A
Edits made	N/A
Version	1
Data source	Collected by LGA survey teams during 2024 measles campaign
CRS	Unknown
Terms of use	For internal use only
Known errors	Data expected from 10 LGAs but only received for 8
Additional notes	GPS and ODK forms were used during data collection. Metadata tab created by A. Learner on 31 March 2024



It is also important to add a data dictionary to your metadata tab so that other users can easily understand the data fields. Add a list of data field headers and their descriptions to your metadata tab below the information you have already entered.

Data dictionary	
uniq_id	Identification number of each health facility
timestamp	Date of data collection
latitude	Latitude of health facility
longitude	Longitude of health facility
wardname	Name of ward the health facility is in
wardcode	Code of ward the health facility is in
lganame	Name of local government area (LGA) the health facility is in
lgacode	Code of local government area (LGA) the health facility is in
statename	Name of state the health facility is in
statecode	Code of state the health facility is in
updated_on	Date that the data were last updated
func_stats	Status of the functional statistics
category	Health facility category
ownership	Health facility ownership (public or private)
type	Health facility type
source	Organization in charge of collecting the data
prmry_name	Health facility name

Now that you have processed the raw data file and created its metadata, you are ready to move on to Task 2 (in which you will start to look at the contents of the data file). Ensure you have saved your file in the *2_RawData* folder before moving on to the next task!



Learning checklist

By following the above directions correctly, you will have demonstrated good practice in data preparation and management for GIS by having:

- Saved a raw data file in a readily accessible, fixed directory with a consistent naming convention
- Navigated a pre-defined data management folder structure
- Created detailed metadata via previous knowledge of the data

In the next task, you will create an intermediate-level data file after addressing any anomalies with the table structure that would generate errors when importing the data into GIS.

