

# NKN: Data Management Decision Key

Follow the steps in this guide to help you plan how to **document, share, and get credit** for your data & research products

## A few definitions:

*raw*= data as it was collected with no alterations

*processed* = data that is used in subsequent analyses, after initial quality checks

*derived data products* = data products that result from analysis of processed data

*research products* = unique intellectual contributions that can not be defined as research data, such as models, tools, scripts, code, & visualization products

*digital object identifier (DOI)* = a unique code used to create a citable record for data or products and to enable stable online links using this code

<b>Question 1:</b> Are you conducting research that you plan to publish in a peer-review journal?	
<input type="checkbox"/> Yes	... Go to Step 1
<input type="checkbox"/> No	... You can stop reading.
<b>Step 1:</b>	
Write a Data Management Plan; <a href="#">Get tips</a>	
<b>Question 2:</b> Are you collecting raw data (biophysical and/or social)?	
<input type="checkbox"/> Yes	... Go to Question 3
<input type="checkbox"/> No, I only use existing data (publicly available, from a colleague).	... Skip Question 3 & Steps 2-3, Go to Question 4
<b>Question 3:</b> Does your raw data include sensitive personal information and/or did you have to get IRB approval for your study?	
<input type="checkbox"/> Yes	... Go to Step 2
<input type="checkbox"/> No	... Skip Step 2, Go to Step 3
<b>Step 2:</b>	
It is your responsibility to perform any data cleaning and de-identification steps required to protect sensitive personal information collected as part of your research. Examples of techniques commonly used to protect personal information are: deletion/anonymizing, recoding/anonymizing, top coding, collapsing categories, and statistical disclosure control. Be sure to follow protocols and human subjects risk mitigation plans throughout the conduct of your study and in sharing your data. Some best practice for managing data during your project include using data entry tools to ensure fidelity and securing data/databases with appropriate read/write permissions and passwords. The need to protect sensitive personal information and follow IRB approvals supersedes any funding agencies requirement to share all data. All data should be documented with metadata and only data from which personal information has been removed is required to be shared publicly. Data or versions of data that contain sensitive personal information should be archived in restricted repositories and documented with publicly available metadata records.	

### **Step 3:**

Track the steps you used to quality check, clean, and process your raw data to produce your processed data (ready for subsequent analyses). At this stage use a system that works for you in your process, e.g., comments in your R script, 'meta' worksheet in your Excel file, NKN's Intermediate Metadata documents ([biophysical](#), [social](#)). These Intermediate Metadata documents will give you a good idea of what information you should be tracking in whatever system you choose to use.

### **Question 4:** How are you analyzing your data?

- I am analyzing my data using existing tools, software, etc. ... Skip Question 5 & Steps 4-5, Go to Step 6
- I am analyzing my data using new tools/models/scripts that I am creating. ... Go to Question 5
- I am using existing data to create a visualization product (e.g., City Engine scene, Unity virtual world). ... Skip Question 5 & Step 4, Go to Step 5

### **Question 5:** Are you using GitHub for development of your new tool/model/script?

- Yes ... Go to Steps 4
- No, I am using a different version control system or I am NOT using a version control system. ... Skip Step 4, Go to Step 5

### **Step 4:**

Get a DOI for your tool/model/script when it is ready for release using [Mozilla Science Lab's 'Code as a Research Object'](#) process to archive your GitHub code repository to [figshare](#) and receive a DOI for it.

### **Step 5:**

Use [NKN's metadata editor](#) to write a Dublin Core metadata record for your research product (i.e., new tool/model/script or visualization product). If you completed **Step 4**, include the DOI link as an online resource in your metadata record. Publish this record to NKN's Portal.

-- End of data management steps for new tools/models/scripts & visualization products. --

*... go back to Question 4*

*(You may also need to work through the key for the first choice in Question 4)*

### **Step 6:**

During your analysis, track all your steps, versions of existing data you used (source, when downloaded, url, etc.), software, programming languages, and parameters, used, etc. At this stage use a system that works for you in your process, e.g., comments in your R script, 'meta' worksheet in your Excel file, NKN's Intermediate Metadata documents ([biophysical](#), [social](#)). These Intermediate Metadata documents will give you a good idea of what information you should be tracking in whatever system you choose to use.

### **Step 7:**

I am done with my analysis and am ready to prepare materials for data upload to a public repository.

### **Step 8:**

Assemble a folder of materials related to your final data (do this for each set of data; group data as makes sense for your project). Read some helpful tips on how to prepare your data package [here](#). Make sure your original tabular raw data and your final data are in CSV format. If you used Excel, data that

represent intermediate steps (calculations, etc.) can be included in multi-worksheet Excel format. If you used a script written in an open-source programming language (e.g., R) to process and analyze your data, include your raw data and any other data files you read into the script, as well as the script file. Other files that should be included as applicable: original survey instrument, codebooks, steps taken to remove sensitive personal information, IRB-related protocols, human subjects risk mitigation plan, Intermediate Metadata documents, etc.

**Question 6:** Do you wish to detail more of your social science survey information in standards-based metadata? (compare to entry form in [NKN's metadata editor](#))

- |   |                                |
|---|--------------------------------|
| <input type="checkbox"/> Yes  | ... Go to Step 9, Skip Step 10 |
| <input type="checkbox"/> No, or my data do NOT include social survey data | ... Skip Step 9, Go to Step 10 |

#### **Step 9:**

We recommend you use the [DDI standard](#). [Nesstar](#) is a downloadable program you can use to create these metadata. At this time NKN does not support DDI metadata in our Portal, but we plan to offer upload, validate, and publish functionality for this standard in the future.

#### **Step 10:**

Use [NKN's metadata editor](#) to write an ISO 19115 metadata record for your new data. If your data will NOT be stored in NKN's data repository, please include a link to where the data are stored or served as online resources in your metadata record.

**Question 7:** Do your data materials contain GIS files (standalone shapefiles, ArcGIS geodatabases, ArcGIS files offered as services through ArcServer (sde)?

- |                              |                                 |
|------------------------------|---------------------------------|
| <input type="checkbox"/> Yes | ... Go to Step 11               |
| <input type="checkbox"/> No  | ... Skip Step 11, Go to Step 12 |

#### **Step 11:**

Make sure metadata are attached to your GIS files' native xml so that this important information will travel with the GIS files. If you use ArcGIS, you can export metadata records from [NKN's metadata editor](#) to ESRI-compatible ISO and import this file within ArcCatalog. You can also export the ISO xml and you may be able to use this in other program. If you are publishing a number of GIS files that are components of your final data package, it may be appropriate to write a simpler metadata document using the editor native to the GIS program you use.

#### **Step 12:**

Publish your metadata record to NKN's data repository and [options]:

- 1) upload the data materials folder you assembled in **Step 8**,
- 2) request a DOI for your data (if you upload your data with NKN, we will take care of ensuring permanent links to your metadata record AND data, if you only share your metadata with us, we will take care of ensuring a permanent link to your metadata record but it will be up to YOU to update the link to your data provided in the metadata record if it changes),
- 3) request an embargo period if you would like public access to your data to be delayed for a period of time, and
- 4) request permanent restriction of access to versions of the data that contain sensitive personal information.

... You're DONE! Celebrate it!