

Data Management Planning Considerations – Checklist

Plan

Is there a specific format of data management plan that must be used?
When do you expect the project to start?
When do you expect to complete the project?
What is the schedule and budget for data collection?
Is funding available?
Who is responsible for managing the data and the data management plan?
Where will the data and data management plan be stored?
Who is responsible for creating the metadata?
Who is responsible for updating the metadata?
What data are you collecting? Does the data already exist?
What format will be used to collect the data?
Are there any data standards for this data type?
What format will the data be stored in?
What file type will be used to store the data?
Are there any specific requirements for sharing or storing the data?
Have you checked all possible sources for your data needs? (Other field offices? Other Federal, State, or local agencies? FGDC Metadata repository?)
Are there any access restrictions for the data you are collecting or will serve?
How do these impact the needs to store, access, and protect the data?
What are all the likely uses for the data? Who will use it and what kinds of outputs will be needed? (Electronic or paper? Alphanumeric or spatial? Available to the public?)
Will a different format be used to share or publish the data?
What naming conventions are being used for files? What about table columns and rows?

Acquire

How will data be collected? (Electronically or on paper? In-house or by contractor?)
If you are obtaining data from an outside source, will you be able to store a local copy of the data?
If you are accessing someone else's data, did you document the steps you used in order to obtain the data? (This is particularly important if you are querying a database.)
If the source data is being updated, will you be notified or will you have to manually check?
What will you do if the data source is no longer available? (Website no longer exists, project is no longer funded so data is no longer being updated, data has been replaced with a newer version, etc.)

Preserve

How will non-digital data be preserved, such as field notebooks and maps?
Is someone responsible for migrating data sets that are in old/outdated formats?
If files are being migrated or converted, is someone performing a check on both files (original and converted file) to make sure the data has not been altered and that data has been lost?

Publish/Share

Are there any security or sensitivity issues that might preclude you from sharing the data?
Are there any restrictions on sharing data? How about sharing data that you have acquired from other sources?
When publishing a report, must the data be published as well? If so, which data should be included – raw data, the entire database, model outputs, only the final data set?
Is there a deadline or schedule for sharing your data as required by the funding agency?
How long are you required to provide access to your data?
Does your project or program have a specific repository for your data?
Are there any policies regarding where you should post your data?
If your data is stored elsewhere, how long will they provide access to your data?

Describe/Metadata

What format will the data and metadata be in?
What metadata standard will be used?
What format will be used to collect the data?
Are there any data standards for this data type?
Will a different format be used to share or publish the data?
Will you be using a standard naming format and version control?
Have you captured the workflow (which conceptualizes the data inputs, transformations, and analytical steps to achieve the final data output)?

Manage Quality

How will the data be checked and certified? In-house or by contractor? Note that in-house certification should be done by an individual other than the person collecting the data.
Are the QA/QC steps being documented?
Is QA/QC occurring throughout the data lifecycle?
Is data that is transcribed or copied checked for errors against the original data set?
Have you reviewed the data, such as selecting a random sample, looking for outliers, graphing and plotting the data?

Backup and Secure

How is the data being stored and backed up?
Who is handling the backups?
If your data is being stored elsewhere, do they have a backup policy in place?
Are the data and backups being stored in multiple places and on different media types to protect against a single-point failure?
Is there someone checking to ensure that backups are being done properly?
Does more than one person know where the data are being stored and how to access them?
How long will backups be kept?