Facilitating a Data Review

Acknowledgements

This content was created by Fauna & Flora for external audiences with support from Margaret A. Cargill Philanthropies. *Last update 2025*

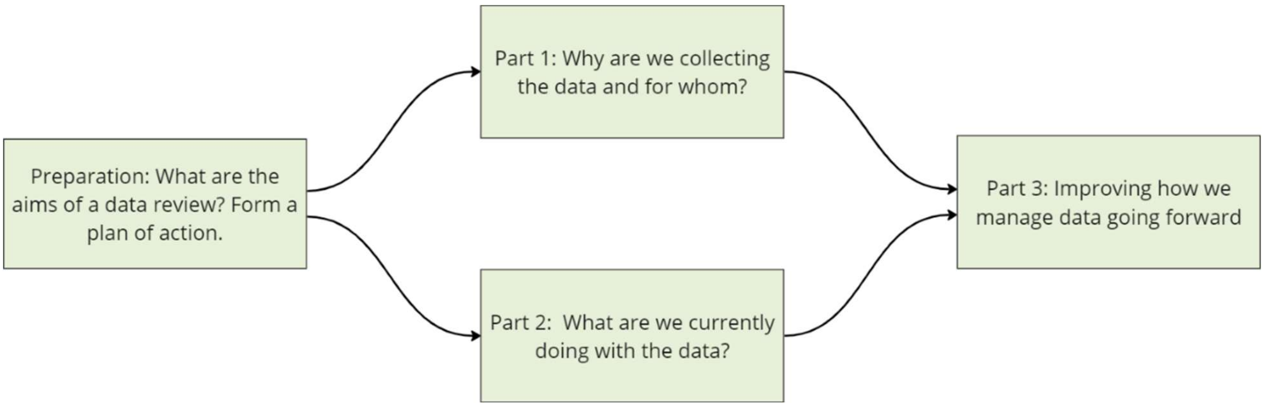
Contents

The purpose of a data review is:

- **to streamline overall data practices***
  - **to ensure relevant audiences have timely access to information**
    - **to make informed decisions and understand impact. i.e., to make information as useful as possible, and processes robust for the future.** ** See "Glossary of key terms" below.*

Structure of a data review

A data review can be broken down into the following sections;



Using the resources

The data review approach suitable for each scenario will vary, and these resources can be adapted and will continue to evolve.

Key considerations

For each conservation project, the dynamics of the data systems and the scale at which a data review is needed will be different. It may not be necessary to apply all of the steps of a data review to every situation. It is possible to jump straight into developing and implementing new data practices and systems. However, taking the time to engage with the questions within each section should help to ensure that the data practices are suitable for the various needs across the project and robust to the future.

Facilitation: This process can be led by someone internal to the team or an external facilitator. These sessions frequently cover challenges or sensitive topics, such as loss of information. To ensure this process enables the full participation and perspectives of the team to be considered it's important to ensure that the facilitator/s are conscious of fostering an open and safe space, free from judgement and aware of power dynamics. Unbiased facilitation is key.

Ownership. While you may be facilitating, it's important that the results (resources, systems, information) are "owned" by the team, so that they will be suitable for their needs, adopted and implemented. Communicating this, encouraging the team to complete sections independently, co-facilitating with core team members with data responsibilities, or any other techniques that can be used to encourage those responsible for the data practices within a team to take ownership is vital for the new practices to be fully adopted.

Scale: Scale has a large influence on the time and people required to conduct a data review. Reviews could be conducted at a range of different levels; country/programme, project, fund or dataset. While higher-level (country/programme level) reviews will take longer, the result is likely to be a more robust data system that meets various needs of a whole team into the future. If it's not suitable or possible to conduct a review at a higher-level, then while completing one for a specific strand of work (e.g., a project or a fund) take a step back and consider if any developments will align with the overall data systems incorporating the whole portfolio of work. Exploring what's working well and the challenges within the preparation stage will help to identify the scale a data review should be conducted.

Who should be involved at each stage? The data review is likely to result in data systems that will be adopted and influence a broad range of team members. Therefore, the overall process should include as many project members or relevant stakeholders as possible. It will not be suitable for everyone to be involved in each stage. However, inclusion of the broader team within the reflection components can be extremely valuable to enable understanding of the "bigger picture" and of each individual's role to strengthen the process and the team. By the end of the process, everyone involved with the data should understand; Why is the data collected? Where does it go? How is it used? What is everyone's role?

Realistic timescales. The time required is dependent on the scale of the data review. In some cases, it can take a much longer time, if very complex or there's a lot of steps to work through with the data identified. Remember, at the end there will be a plan of action to develop and implement any changes to the data management practices of the project, which could potentially take several months.

Part 1: Normally two sessions that are between 2-3 hours.

Part 2: Normally it will take multiple 2-hour or half-day sessions.

Part 3: Group sessions to agree the priorities, a way to move forwards and define roles and responsibilities normally takes 2-3 hours.

There will then be on-going work to implement changes and develop protocols.

At the end of the data review, there will be a plan of action and it could take several weeks or months to fully apply any changes to data management practices.

Part 1: Why are we collecting the data and for who?

The aim of the first part of a data review is to understand what data we need to collect and why. This is through:

- identifying all of the different audiences that need access to information,
- clarifying their uses for the information and any specific data outputs needed, and
- identify the level of priority for the project.

The information gathered through this section will be very informative when looking at the overall data management practices to ensure they are fit for purpose and the demands on the project team to provide various information can be as streamlined as possible.

Some of this information may already exist or have been collated through previous MEL processes, such as high-level indicators and questions relative to a Theory of Change and/or in a Design and Impact Overview, or log-frames for specific donors. We recognise there will be additional MEL related to specific donors or other key stakeholders of the project (e.g., communities, local government etc). **Adapt these sessions to utilise the information already held by the project.**

Outputs

a. List of key questions, uses and data outputs for each audience
b. List of indicators to address the needs of each audience

Step 1: List the audiences. Who are the audiences? Audiences can include any individuals, organisations, groups or stakeholders that have a use or need for the data or it's outputs. This could include funders, the project team, community members, partner organisations, government, senior management team or the board of directors. List the audiences.

Step 2: Understand how they will use the data. How will each audience use the data/information? What are the uses or final outputs of the information by each audience? How frequently do they need it? Are there additional data sharing considerations, e.g., open data? High-level capture of information which can then be used to help guide the next sections. E.g., reporting, learning, workplans, case studies, proposals.

- Consider the Design and Impact Overview, MEL Framework, log-frames or reporting requirements already in place, and take this opportunity to step back and think overall for the project and its conservation goals.

Step 3: What key questions need to be answered? Being more specific, what are the key questions each audience needs to be able to answer or address? Group these into essential and desirable.

Consider **WHY** the audience needs this information:

- What action, questions or decisions are they trying to enable or inform?
- Will this inform the understanding of their impact? (Extent of impact, positive/negative impact, impact persisting over time)
- Who do they want or need to communicate it with? What do they want to communicate?

Step 4: Reflect on the first three steps. Pause and reflect on the information gathered.

- Is there an overlap or similarities between the purpose and uses for each audience?
- Group or align the key questions where possible.
- Are the data outputs for the audiences the same? For each output consider what the format is and the data they will need to contain. You can have two outputs that are in the same format but contain different data for different audiences (e.g., output is an excel sheet of patrols but your team wants a more detailed version than your funder), or you can have two outputs that are in different formats (e.g., from this data the team wants an excel and a dashboard? It can be useful to group or align output format and/or the data they will contain across the audiences, but it's not essential to narrow down the list.

  Tip: It's useful to consider the information above to begin to understand the demands that will fall on the project team managing the data.

Step 5: Make it operational. For the key questions for each audience, consider the data that's needed to provide the "answers" to this question in the format required. i.e., define the indicators. Will this information enable the audience to be able to meet or fulfil its goal?

This step is essentially linking what data audiences need with what data we are currently collecting and then seeing if there are any gaps. Are there data that are not useful for any audience and are data missing which are needed to answer key questions?

> *Tip: It's useful to have knowledge of all of the data variables already collected or have them collated in a way that's easy to reflect on together (e.g., post-it notes or a list), these can be cross referenced with the key questions, and you can begin to consider if the data we are collecting already provides this information or if additional/different information is needed. There is a section below with some information about collating data variables.*

Consider the following:

- What data is needed to address the key questions for each audience? (Data, format and frequency needed). i.e., defining indicators, the specific information that needed.
- What level of detail is needed by each audience? (Ideal and essential)
  - Reflect on how much is enough? Is this essential to be able to understand or address the question?
- Do we have this data already? Reflect on the data variables already held: see Part 2, section c.
  - If so, how much do we have?
  - Where is this data?
  - Is the data ready to be used? Does it need processing, cleaning or organising?
  - What needs to be done to get the data into a useable format (e.g., level of analysis required)? What is needed to produce this information? E.g., analysis software or skills required. Can the team currently provide this? If not, what's needed?
- If we don't already collect this information, what do we need to collect? Can this be added to a current data collection system or is something new needed? What are the implications on the project team? How practical is this? (If not practical/realistic, go back to reflect on the purpose

for the audience, and consider if there is another option suitable that's practical for the project team).

- When reflecting on the two points above it can be useful to align this with the data variables already within datasets/databases. Note where or if any additional variables are needed within a dataset or database. See the "understanding data variables" section of part 2.

Tip: capture notes and action points as you go along.

Step 6: Pause and reflect

- Pull out the priority uses for the information and audiences. What are essential and what are desirable?
- Begin to consider the demands that the various audience needs put on the project team throughout the data management process. Is it possible to streamline this in anyway? Can anything be aligned, i.e., could the same data answer more than one question or be used by more than one audience? This will be reflected on more throughout the next stages of the data review.
  a) For example, can indicators be aligned across audiences, or if the data manager needs to provide information to different audiences can this be streamlined in anyway, e.g., production of dashboards for various audiences from same data sources to ease production and provision of information.

At the end of part 1, information should be saved somewhere that's accessible to the whole project team. This could be captured within the Design and Impact Overview, MEL Framework or Information Management Plan. The content should be documented within the relevant documents to ensure its accessible and implemented by teams, and updated in the future.

Reflection: would the information captured provide sufficient information to a new team member within the project?

Part 2: What are we currently doing with the data?

The aim of this session is to reflect on the management practices in place for all of the data within the project (historical, current and planned), identify if the data are meeting their purposes and if the systems in place can be streamlined in anyway. There are several different elements to understand the full picture of the data management practices and systems in place, more information can be found below.

Outputs:

a) Data registry
b) Mapping of data flows
c) List of data variables collected within the project

Data registry

Aim: Gain an understanding of and make a record of all the data the project holds. This could be overall or for a specific strand of work (the scale should be determined with the team before engaging with the process). Completing a data registry should ensure core information about the data is captured, accessible and secured for the project team, and available to inform reflections to streamline overall data practices that are robust and caters to various needs across the project.

- Check with the data manager if there is a record of the data held by the project? Is this accessible to the team? If the team changes, if the person (or people) charged with looking after the project's data left tomorrow, would another person be able to take over the data management? How easy would that be?

- If the project team already has information captured take some time to reflect and see if it would enable the recommended questions below to be addressed.

- If the project doesn't have a data registry already, work through the recommended questions for each of the datasets/databases held by a project.

- The information for a data registry can be collated in a range of formats (excel, word documents etc.). It's recommended that the final product be an excel, making the information accessible for the long term and with agreed standards for the use of and frequency of updating the document. It can take a lot of time to collect the information for a data registry, but it's a valuable process and resource to capture knowledge and a key stepping stone for developing robust data processes. It's recommended to identify the key variables that should be discussed and those that relevant team members could populate independently.

- The information captured within a data registry can vary relative to each project. The following information is recommended to collect and discuss with the project team:
  - What's the data set or database? (e.g., habitat restoration and management database, or fisheries biodiversity monitoring database)
  - Summarise/describe the subject or content of the data.
  - Who is the data manager? Ideally, they will be involved in these discussions. If not, can they contribute to the information gathered?
  - Who holds the various roles and responsibilities across the data management practices?
  - Source of the data, primary and/or secondary, i.e., who collected the data or who owns it. This could simply be the project team/organisation. It's particularly important to document this if it's a secondary source. Primary: collected directly by the project. Secondary: collected by someone else and made available online e.g., online data repositories or project partners.
  - What method is used for collection (as suitable)? Is there a link to the method protocol followed?
  - When did collection start? If collection has ended, when did it end?
  - How often is it collected?
  - What tool or software is used?
  - What type of data is it? (e.g., quant, qual, spatial)
  - Does the dataset contain sensitive data or personal data?
  - What format is it in?
  - Where is it stored? Can a link or file route be added?
  - Who has funded data collection? Capture if this has changed over time.

- Is there a information management plan?
- Is the methodology documented?
- Is there a data sharing agreement?
- Who owns the data?
- Do you have consent for data storage or sharing? How will you anonymise data?
- Who are the audiences and what are the uses? (link/align with information already collected)
- What are the data outputs required?

Reflection/discussion points for each data set/database:

- What works well?
- What are the challenges or problems?

When this information has been gathered, reflect on all information gathered overall.

- Are there common themes in what is working well or the challenges and problems?
- Where will the data registry information be stored? Is it accessible to the project team? Who is responsible for updating and managing it?

Add a link to the data registry into the Design and Impact Overview, Information Management Plan or other relevant documents so that it's accessible by relevant team members.

Mapping data flows

Aim: visualise the pathway that the data moves through (from design through to its uses), to aid understanding of the processes and people involved within the project team, and enable streamline, robust and reliable practices to be put in place.

- For each data set map out the processes the data goes through to reach its end point (i.e., the data lifecycle). The various stages that can be considered include;
  - Design/planning
  - Piloting
  - Collection
  - Receiving: Data download/received digitally
  - Management and storage
  - Data processing: additional steps that need to be taken ready for analysis, like combing of data sets (not always an additional step to "cleaning" but can be worthwhile to consider this element and ensure there's a plan to complete this)
  - Data visualisation and analysis
  - End use/s: sharing and communicating

Mapping the flows of this visually can aid understanding of complex situations. At each stage consider; where information is stored, who is involved, who is responsible, what is involved within this process, what tools are used or needed, are there any challenges or what is working well, is it sensitive, who has access permissions, does it need to be shared and how?

- Reflect overall for the whole project, identify challenges and what is working well and pull out the priorities.
  - How can processes be aligned or streamlined across the different datasets to ease demands on the project team and increase efficiency?

- Do these align with any priorities highlighted when gaining clarity and understanding of the purpose of the data? E.g., if there is a priority question for the project to address, is this also a priority when reflecting on the data flows? Make an action plan to address the priorities.
- The information captured within this process can be documented (as text, diagram, table) within an Information Management Plan or Methodology document for a specific dataset conservation data plan to ensure that there can be consistency within data practices. Treat these plans as living document and take note of any changes in circumstances. Clearly documenting this information should hopefully make the reliability and consistency of data robust to project or team changes, aid making the data comparable over time, and ensure team members can be held accountable to their roles and responsibilities at different stages within the data flow.

Understanding data variables

Aim: gain a detailed understanding of all the information gathered by the project, to enable reflection on if this can meet the project needs

- The project may be collecting information through various methods or sources to contribute to data. Mapping of the variables currently collected within datasets can be extremely useful to understand if they can align with the indicators required by the project, and aid the planning of analysis. For example, can the variables collected within biodiversity monitoring surveys be analysed or processed to deliver the information needed by the various audiences? OR do the variables within a MEL tracker enable the analysis needed?
- It can be useful to make the variables from a team's datasets accessible for reflection by having datasheet available, such as printing a dataset's headers. These can then be reflected on alongside the audience data needs and/or indicators identified in part 1.

- Reflections on databases:
    - How do the variables align with the various uses and needs for the information identified? E.g., can one variable be used in different ways to produce two indicators (e.g., % change, and count)
    - Do we need to collect more information to address the questions we need to for each audience? Specifically, what variables would be needed. Add these into the relevant datasets.
    - What are the implications on the project team? E.g., what is the data processing or analysis steps required?
    - On what scale and in what format is the information needed? Does the structure of the dataset/ database enable the analysis to be conducted?
        - Does the data format align with all the data outputs required by the audiences?
        - On what scale is the information needed? E.g., how many rows per set of variables (year/month, per entity)
        - If in excel, what information is needed in each sheet to align with the information above?

Throughout consider the practicality for the project team, and the bigger picture (how this can be streamlined or aligned ease demands on the project team). Do the project team have the software or skills needed to deliver the end use?

Part 3: Improving how we manage data going forward

Overall reflections. Bringing it all together.

Outputs

a) List of priorities to ensure there's alignment between the data practices to fulfil the various purposes
b) Roles and responsibilities detailed and captured within the Design and Impact Overview, Information Management Plan, Data Registry or methodologies (see below).

Action plan

Step 1: Actioning Priorities. Do the priorities for the purpose and audiences align with the data management priorities? Overall, what are the actions that need to be delivered to ensure data can fulfil its purpose and ensure the data practices are robust.

Within this section it can be useful to consider the following:

- Scale and compatibility across the project overall. When thinking about adopting a new system, take a step back and reflect on the overall situation. It's important to balance the robustness of data with the practicalities for the conservation team delivering the work on the ground.
    - What needs to happen to reach a point that the information can be used? What's essential? Does the team have the capacity to implement the change – is a consultant needed?
    - Does it align with the current processes?
    - What impact will this have on the team (time, training needs)?
    - How much information is needed to address the question? Does a technical specialist need to be consulted?
    - What is needed overall to streamline the processes for the project team? For example, if the person (or people) charged with looking after the project's data needs to provide information to different audiences can this be streamlined in anyway e.g., production of dashboards for various audiences from same data sources to ease provision of information.
- Do the different datasets need to be comparable, or can they stand alone? This will influence the collection and management systems, and if the format of data should be the same or different.
- What needs to happen to make the ideal a reality?
    - What is the team capacity, are there training needs?
    - Are there resource or equipment needs?
    - Step by step, don't forget the "menial tasks" e.g., consider how long it might take to organise historical data (or new data) and who will do this.
    - It's important to balance the robustness of data with the practicalities for the conservation team delivering the work on the ground.

Step 2: Document and understand the roles and responsibilities within the team.

- Step by step what is needed to ensure the data reaches its end point. E.g., if data cleaning is needed, or formatting files and folders into a uniform storage system, how much time is needed and does the team have the time and capacity to conduct that. Consider what is needed within design, collection, data download/received digitally, data validation, data processing, data analysis and end use.
- Plan for the future and if any additional roles or training are needed.
- Capture and detail the roles and responsibilities of the team somewhere that's accessible to the whole team.

Recommendations:

a. Design & Impact Overview: High-level who is overall responsible for data practices
b. Data registry includes data manager per dataset, and a list of all the various roles and responsibilities with a dataset
c. Methodologies: include more detailed timeline of who has been responsible for various steps with data
d. Information Management Plan: include who is responsible for the information management plan being implemented.

- It can be useful to conduct a RACI Responsibility assignment matrix for the team (responsible, accountable, consulted, and informed), this can aid people management and ensure there's clarity and responsibility shared within the team. Below is an example template. Which can be included in a suitable document as agreed with the team.

	Responsible	**Accountable**	**Consulted**	**Informed**
Activity				
Collecting data				
Checking data				
Processing data				
Analysis of data				
Making decisions based on data				
Writing annual reports				
Producing and sharing lessons learnt				
Producing and sharing learning to external audiences				
Theory of Change (ToC) review				
MEL fund proposal support (including ToC and Logframe development)				

Step 3: Step back and reflect on all of the steps taken within the data review

The following are key things to reflect on, consider and balance when making decisions around the priorities for actions and decisions around the practices put in place:

- **Risk.** Overall where are the biggest risk factors? For example, for the loss of information or information not able to be used for a priority needs or accountability. Identifying the biggest risks may help to prioritise what should be tackled first.
- **What happens if a team member leaves tomorrow?** Is sufficient information documented about the overall process so someone else can continue this role and ensure

information collected will be robust, fulfil its purposes and be comparable? Where is this information documented, e.g., Design & Impact Overview, Information Management Plan, methodologies, MEL Framework? Is anything else needed? (consider all steps of the data lifecycle)

- **Team capacity.** Does the team have the resources (time, skills, funding) to implement all the planned activities?
  a. Who will implement each step of the data lifecycle?
     - Do they have the necessary skills to conduct it fully? If they need additional training or there is a need for additional support who will provide this? Note: before selecting a data lifecycle that will need additional support from another team, like the conservation technology team, ensure you contact them first, include them in the planning and ensure they are included in the budgets.
     - How much time is actually required for data to move through each step of its lifecycle? For example, how much time will it take to process data monthly so it's available for analysis? How much time will analysis take? Does this align with staff availability and other commitments?
     - Long-term is there sufficient funding to maintain this data lifecycle? E.g., replacement camera traps, paying for additional data storage.

Prioritisation exercise and maintaining momentum:

- What are the priorities to be implemented?
  - It can be useful to reflect on the learning from part 2 alongside anything documented from part 1 for the audiences and the outputs needed.

For example:

- What are the priorities for the audiences and data outputs?
- Can these be delivered in a streamline and robust manner? (looking at data flows or notes made while developing the data registry can be useful).
- What are the priorities identified within part 2?
- Do the priorities from parts 1 and 2 align? If not, what's feasible to be implemented and what's the overall priority given the situation with the team?

Allocate action points to specific people, and who will hold the team accountable to these actions.

.

- Schedule your next meeting/check-in.

Tip: these actions are important but slip in priority relative to delivery of other work. Scheduling regular catch-ups (monthly/bi-monthly) or a presentation date for the end result to share learning and/or on-board others has been key to maintaining momentum.

Reflection. Organise a reflection in 6 months (or as suitable) to see if the team are making progress to deliver the action points identified, and to reflect on the situation and see if there are additional tasks that need to be completed.

Glossary of key terms

Data cleaning: (or checking in some cultures), the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

Data management: a process to ensure that diverse data sets can be efficiently collected, integrated/processed, labelled/stored, and then easily retrieved through time by people who want to use them.

Data Practices: include design of methods, planning and data management, collection, storage, processing, and visualisation and analysis of information for interpretation and use.

Indicator: A sign, clue or marker that helps answer our monitoring and evaluation questions and indicates if we are on the pathway to progress.

Variable: A value collected or measured within a survey or held within a dataset or database. A variable is any characteristics, number, or quantity that can be measured or counted.