



afrialliance
Africa-EU Innovation Alliance for Water and Climate



The Handbook on **Data Collection**

Introduction to the AfriAlliance Handbook on Data Collection

The AfriAlliance Handbook on Data Collection provides guidance on how to best develop a data collection project to ensure high quality data is collected and maximum impact is achieved.

Focusing on the development sector and the collection of data, mainly by people, the Handbook covers the main elements to consider when designing and implementing a data collection project. Where applicable, it will also point you to more detailed resources.

The Handbook provides guidance for projects to:

- Focus on achieving impact.
- Only collect the data that is needed to achieve that impact.
- Build on existing data.
- Use the most efficient method of data collection.
- Make sure data benefits the communities it is collected from.
- Share data whenever possible to ensure others do not have to do the same work.

The growing importance of data

The 21st century is the century of data¹: ninety percent of the data ever created was created in the last two years (2015, 2016). The use of data fundamentally transforms the way we communicate, do business and make decisions.

Focusing on the development sector, the transition from the millennium development goals (MDGs) to the post-2015 sustainable development goals (SDGs) poses new challenges in the realm of data for development and drives a massive demand for high quality data. There is a pressing need for data on specific topics, such as water quality, and in general for data and evidence to enable better monitoring and evaluation of development results and to increase accountability, effectiveness and efficiency. Governments and supporting organisations will be required to generate unprecedented amounts of high quality data in order to monitor and report on progress in implementing the SDGs. It is crucial that this data is collected using consistent methodologies and, when possible, is openly available to all stakeholders, while taking into account that some data may be sensitive and can't be shared in raw format.

Data collection for the development sector

In the development sector, a large part of data is collected by people going into the field to perform surveys and conducting other measurements, such as water quality. The Handbook focuses on these data collection projects, providing guidance on the journey from data to decision

according to these five stages: **prepare – design – capture – understand & share – act**. The guidance will describe how to design such projects; how to implement the data collection process; how to combine the collected data with other data sources; how to analyse the data to gain insights and make informed decisions; and how to make the data openly available.

The importance of a good setup

Too often, data is collected using inconsistent methodologies which lead to data that is not useable and/or comparable. Sometimes, additional data is collected just because there is the opportunity to do so, which leads to data-fatigue, both in the collecting organisations who see no concrete results from their large efforts and the people from which the data is collected who see no tangible results from the large amount of data they have given. Lastly, data is often not shared, causing other organisations to collect the same data instead of building on existing data. Data is only useful if it is high quality and collected with a certain goal in mind.

Set up of the Handbook

The Handbook consists of ten phases each data collection project should go through (see Table 1). These phases are not by definition consecutive or prescriptive, there may be some overlaps, and it may be necessary to go back to a previous phase due to findings at a later stage.

1. <https://www.mediapost.com/publications/article/291358/90-of-todays-data-created-in-two-years.html>

TABLE 1 TEN PHASES OF THE HANDBOOK

1	Prepare project	Prepare the project proposal.
2	Design project	Design the project in detail, together with (local) stakeholders.
3	Research data	Research what data is already available and what this means for the data to be collected.
4	Design samples and surveys	Design the sample in which data will be collected and the survey to be used.
5	Prepare data collection	Prepare for data collection in the field.
6	Collect data	Manage the data collection in the field.
7	Analyse and visualise data and gain insights	Analyse and visualise the gathered data, combine with other data sources and gain insights.
8	Share data and insights	Share the gathered data and insights publicly.
9	Make informed decisions	Make informed decisions based on the insights developed.
10	Evaluate and apply lessons learned	Evaluate the project and apply any lessons learned in the follow up.

When divided over the five stages mentioned above, the setup of the Handbook is as shown in Figure 1 below.

Follow up

This is the first version of the Handbook. In the AfriAlliance project, we will continue to build on this version by improving the guidance and updating the resources. We will also provide practical examples of projects based on this guidance.

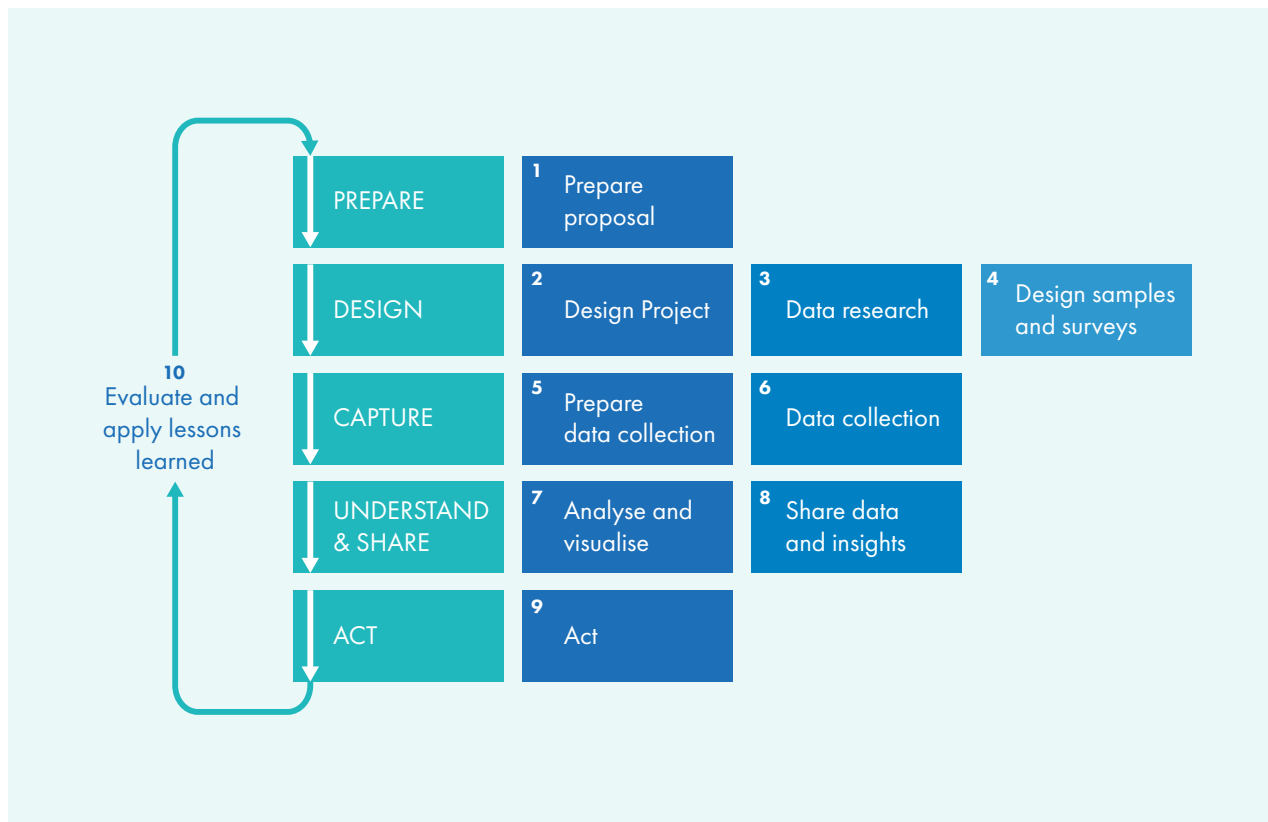
AfriAlliance



The Africa-EU Innovation Alliance for Water and Climate (AfriAlliance), is a 5-year project funded by the European Union's H2020 Research and Innovation Programme. It aims to improve African preparedness for climate change challenges by stimulating knowledge sharing and collaboration between African and European stakeholders. Rather than creating new networks, the 16 EU and African partners in this project will consolidate existing ones, consisting of scientists, decision makers, practitioners, citizens and other key stakeholders, into an effective, problem-focused knowledge sharing mechanism. This is coordinated by means of an online innovation platform: *the Africa-EU Innovation Alliance for Water and Climate*.

<https://afrialliance.org>

FIGURE 1 SETUP OF THE HANDBOOK



EU funded



AfriAlliance is led by the IHE Delft Institute for Water Education (Project Director: Dr. Uta Wehn) and runs from 2016 to 2021. The project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 689162.

Akvo



This handbook was written by Akvo in collaboration with the partners in the AfriAlliance consortium. Akvo works with governments and non-governmental organisations that strive for equal access to public services, reliable infrastructure and a safer environment. Good data is critical for effective decision making, collaboration, and accountability. Akvo provides partners with the data platform to help them capture, clean, visualise and share data. That data platform is supported by a series of services to build local expertise and ensure success from data to decision. With five regional hubs in five continents, Akvo has supported over 20 governments and 200 organisations in more than 70 countries. More information can be found at:

 www.akvo.org.

01

Handbook phase one: Prepare project

Handbook phase one

Prepare project

Author

Harro Riedstra (Akvo.org)

Contributor

Edouard Boinet (International Network of Basin Organisations, INBO)

Most projects start with a project proposal, presented either internally or to an external project funder. The level of detail required for such a proposal varies. Internal proposals are often kept to a minimum, while proposals to government donors sometimes have to be extremely detailed.

In general, a good project proposal for a data collection project should cover the main points outlined in the Handbook, from design to act, with your plans and expectations described for each element. In addition, a project proposal should include elements like schedule, budget and, if presented to an external funder, why your organisation is the best to implement this project.

Below is a list of the main elements of a data collection project proposal. For each element, the main questions are listed (see Table 2). The level of detail for each element depends on the requirements of the project proposal. If you'd like more detail on any particular topic, you can read the relevant Handbook article in full.

TABLE 2 MAIN ELEMENTS OF DATA COLLECTION PROJECT PROPOSAL

Project element	Content	Handbook article
Context	Describe the context in which your project will operate, including which stakeholders will be included.	Two
Objectives	Define the impact you want to reach with your project and think backwards: which outcomes do you need to achieve in order to reach that impact? What data needs to be collected? In a detailed project proposal, a draft Theory of Change (ToC) could be included.	Two
Data needs	Which methodology will be used to assess what data is already available? Which data sources are already known?	Three
Methodology	Which methodology will be used for the data collection, what is the scope of the collection and how will the surveys be developed? How will the data collection take place in practice? Who will collect the data, what tools will be used, what contingencies are built in your plan to deal with unexpected delays, etc.?	Four Five Six
Analysis and visualisation	What information would you like to extract from the data? How are you planning to analyse and visualise the data, and how will you involve communities/stakeholders/decision makers in the process to make sure the end product is useful to them?	Seven
Data sharing	How do you plan to share your data? How will you deal with privacy and sensitivity of collected data?	Eight
Data to decision	How will you ensure that the collected data will contribute to better decision making? Will the intended end users of the data be actively involved in the project?	Nine
Evaluation and application of lessons learned	How will you evaluate your project? What is your draft monitoring framework? Are there any draft targets/How will you set your targets? How will you ensure lessons learned will be used in new projects?	Two Ten
Project timeline	Timeline of planned activities	n/a
Project budget	Budget proposal	n/a

Handbook phase two: Design project

Handbook phase two

Design project

Authors

Anita van der Laan (Akvo.org),
Annabelle Poelert (Akvo.org)

Contributors

Marten Schoonman (Akvo.org),
Karolina Sarna (Akvo.org),
Tarryn Quayle (Local Governments
for Sustainability Africa (ICLEI Africa))

Designing a project or programme in a constantly changing environment can be challenging. The context in which a project operates is often complex, with many different stakeholders and factors involved. For that reason, it's important to base the design of a project on an understanding of the context and choose an approach that allows for flexibility instead of stasis. That way, adaptations can be made when needed. Even a relatively straightforward data collection project needs to keep in mind which stakeholders will be involved and which problems and opportunities exist, at the start and throughout. The methods explained in this article are not just about optimising the outputs of your project design and implementation, but also emphasise the importance of the process in itself, which helps stakeholders to align and feel ownership of the project. The process should be carefully documented in order to capture lessons learned that can be shared within the sector and can be used to feed into future projects. This approach to project design is Theory of Change (ToC) based, a methodology which helps you to structure reality and understand how your project can start a process of change.¹

This article gives suggestions on the three steps of designing a project: how to do an analysis of the context (step one), how to design a ToC (step two), and how to build a monitoring framework (step three). Before going into detail, this article gives a brief outline of the three steps and an introduction to the Theory of Change methodology. Consult the Handbook glossary for definitions of each level of development.

Introduction to the Theory of Change based project design

Creating a Theory of Change (ToC) helps you to structure reality and understand how your project can start a process of change. If you have a certain impact in mind that you want to contribute to as a project, a ToC helps you to understand which different outcomes you need to achieve in order to reach your envisioned impact and how these outcomes are interrelated. While the word impact refers to the ultimate change that your project aims to contribute to, the outcomes are changes that need to happen beforehand. Designing a ToC together with all stakeholders will result in a common understanding and co-ownership of the project and will facilitate the planning of your activities in a participatory way. It will also help you to discover what you collectively want to learn, and therefore to decide what you want to monitor during the project.

Although in this article the steps are presented in a sequence, in reality, the three often overlap and are circular in nature. For example, in order to be able to map all relevant stakeholders, there needs to be an awareness of the context and of what the problems and opportunities are. You might realise after the context analysis that some important stakeholders were overlooked during the analysis. Each step in a ToC based project design can make you realise that something was missing or not clear enough in a previous step and may lead to revisions.

1. The size, or scope of a project influences the time required to conduct a context analysis, design a ToC and build a monitoring framework.

Designing a project in three steps

Step one: Conduct context analysis

Before designing the ToC, you need to have a thorough and common understanding of the context in which your project is operating. Therefore it is good practice to start with a context analysis, which consists of a factor, issue and stakeholder analysis, and a map of the findings.

Step two: Design Theory of Change

Define an impact that you want to reach or contribute to with your project and think backwards. Which outcomes need to be realised to reach the impact, and how are they interconnected? Which strategies will help to achieve these outcomes? Make sure that all your underlying causal assumptions are recorded and made explicit.

Step three: Build monitoring framework

From the ToC, collectively agree on what key expected outcomes (and impact) all stakeholders want to monitor. For those, design a planning, monitoring, evaluation and learning (PMEL) framework, with indicators and means of verification. In addition, monitor the causal assumptions that you are unsure about. Based on your monitoring findings, the ToC should be revised on a yearly basis and adjusted accordingly.

Note: all three steps require a highly participatory approach, to ensure relevance and co-ownership from the start

Step one: Performing a context analysis

Before designing a project, it is crucial to understand the context in which it is operating to ensure that everyone involved has a similar understanding of the situation and that the project is designed to address the relevant issues with the right stakeholders. Context analyses are often outsourced to external consultants. However, our experience shows that some of the knowledge, understanding and connections that the external consultant acquires during the analysis may be lost in the transfer of information. Therefore, we suggest having the context analysis done in a participative

way with skilful insiders. This way, the exercise can result in a deeper understanding and higher usefulness to the design of the project. The context analysis usually consists of two interlinked exercises: mapping and analysing the stakeholders and the factors. Naturally, existing analyses need to be made use of for this step to prevent repetition.

Stakeholder analysis and mapping

The participatory stakeholder analysis helps you to identify and map all relevant actors and their roles, responsibilities, relationships, interests, and relative influence/power. It is

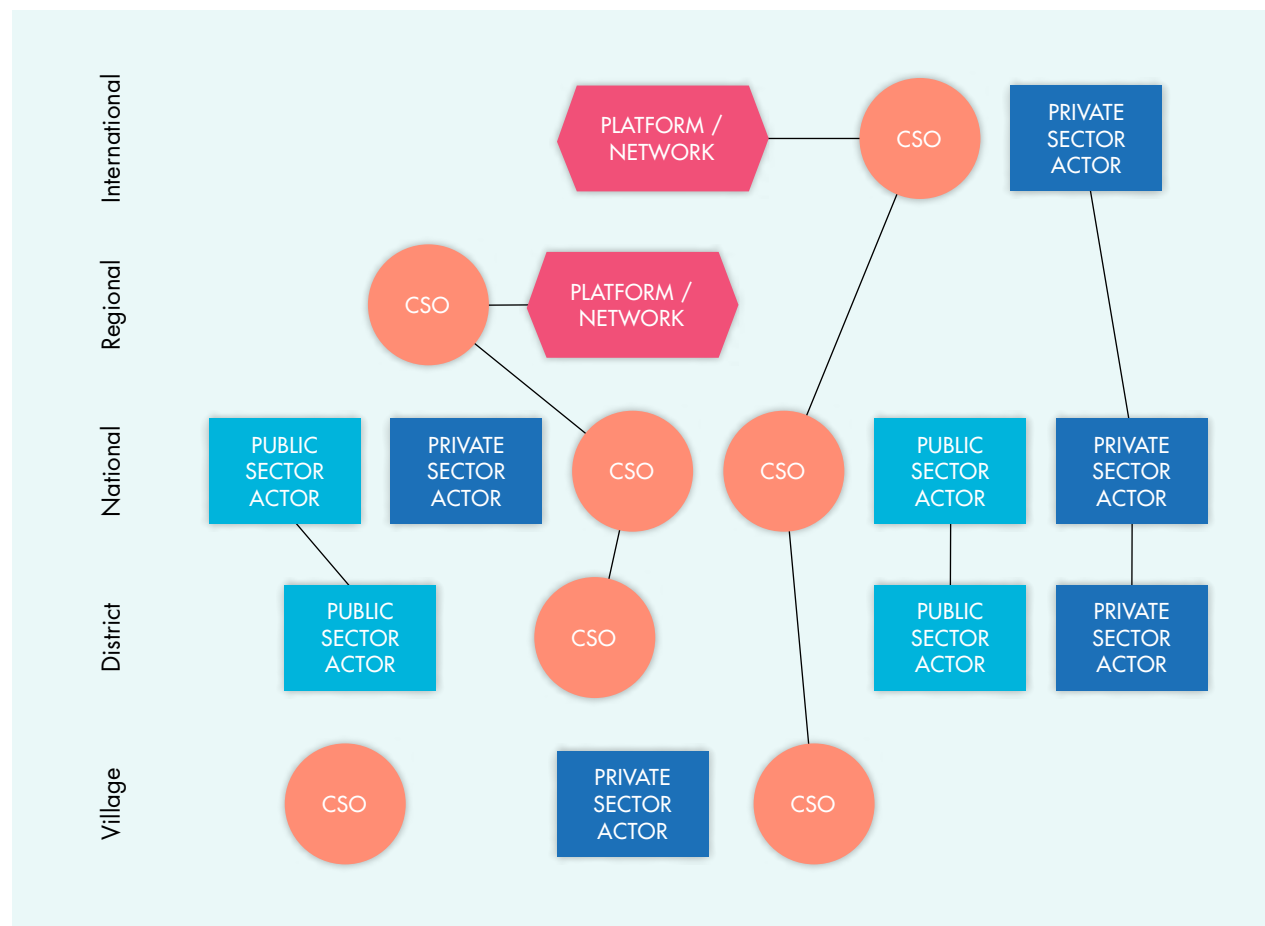
important to make sure all actors are taken into account, including vulnerable and underprivileged groups who may otherwise be overlooked. This exercise is of most value when performed in a group or workshop setting, and will help identify which strategic stakeholders need to be involved in your project, and in what way. During the project, power relations may change and new stakeholders may appear or knowledge gaps regarding existing stakeholders may be filled. It is therefore advisable to review your stakeholder analysis on a regular basis.

Stakeholders² can be categorised into three types, all of whom need to be considered in the stakeholder analysis and mapping:

1. Communities: The people who experience the problem directly, and interact with problem solvers.
2. Problem solvers: The civil servants, non-governmental organisation (NGO) staff, frontline responders, and others on the ground.
3. Policy and decision makers: The people who have access to resources and control allocation, or can influence decision making.

Helpful methods in performing a stakeholder analysis are making a stakeholder map and an interest and influence matrix. In a stakeholder map, all actors who are relevant for the project's success are noted down on colour-coded cards (different colours for public sector, civil society organisations, private sector, platforms/networks), and organised according to the level at which they are most active (international, national, regional, district, village). The nature and strength of relationships between actors can be indicated with lines of different colours and thickness. See the stakeholder map below (Figure 2), intentionally generic and simplified to be useful as an example for any programme or project.

FIGURE 2 EXAMPLE OF A STAKEHOLDER MAP



2. In the next phase, data research, we use the same categorisation of stakeholders.

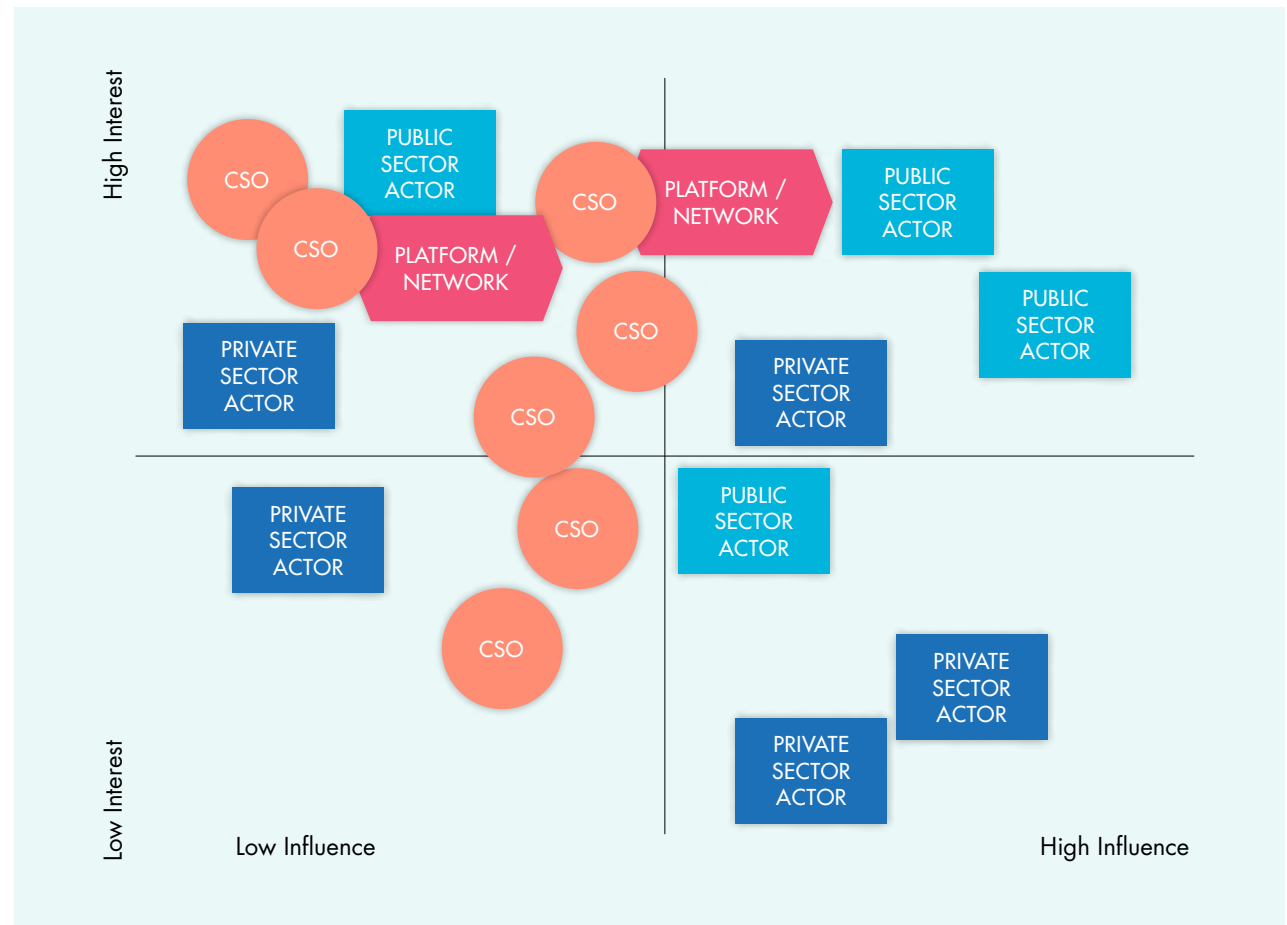
The influence-interest matrix builds on the stakeholder map, see the generic example below (Figure 3). All of the actor cards can be plotted on four quadrants of the y-axis (interest) and x-axis (influence), according to how interested they are in the success of the specific project, and how much influence they have in making it happen. Powerful stakeholders can also have a strong negative effect on the success of the project, and it is important to keep this in mind from the design phase onwards.

While doing both exercises, keep in mind that there are both formal and informal power structures to take into account.

These exercises produce the richest insights when done in a participatory way. The discussions generated by the exercises help to bring out the different perspectives at a time when these can be taken on board, and helps to create co-ownership and a shared focus for all involved.

In the data research phase, which comes after design, the stakeholder analysis is mentioned once again. In practice, the stakeholder analysis will usually be done once, but will consist of two levels of reflection: first as a context analysis exercise to make explicit all stakeholders, and subsequently to zoom into the key stakeholders and discover more about their interest in and influence on the data collection and data-based decision making in the project. The following questions are asked to address the second reflection:

FIGURE 3 EXAMPLE OF AN INFLUENCE – INTEREST MATRIX



- What is their influence on the problem?
- How might this person benefit from the project?
- What could this person do with better data on the problem?
- How does data support this person's decision making now?
- What could they do to undermine the project?
- What is the best way to keep them engaged?
- How can they contribute to a solution?

Factor analysis and mapping

Apart from stakeholders, external factors also need to be taken into account when designing a project. Are there any environmental, historic, political, cultural or socio-economic factors that are likely to have an effect on the success of the project, and in turn, on that which the project can have an effect? Identifying these factors will help to determine the problems and opportunities that need to be addressed. Documenting all factors at play will help to

justify decisions on the project's scope and focus. Helpful methods in performing a factor analysis are interviews with key stakeholders and data research. Read more on data research in phase three of this Handbook.

After performing a stakeholder analysis and a factor analysis, the relevant stakeholders need to reach a common understanding of the problem that the project is trying to address. What are the issues that lead to the overall problem and how are they interrelated? It can help to visualise the issues and their causal relationships. Pay attention to issues that may be overlooked by identifying gaps in your map. The mapping of issues will help to align everyone involved in the project on the most important factors to focus on.

A helpful method for mapping out issues is creating a problem tree³ or a conceptual model (Figure 4). A conceptual model is made by writing down on cards, in a participatory way, all the factors (problems and opportunities) related to the desired impact: what is hampering the achievement of this impact, why is it not happening now? The next step is to cluster the cards according to topics on a wall, and then organise the cards in cause and effect relationships on a map. Such a map, or conceptual model, helps to create a common

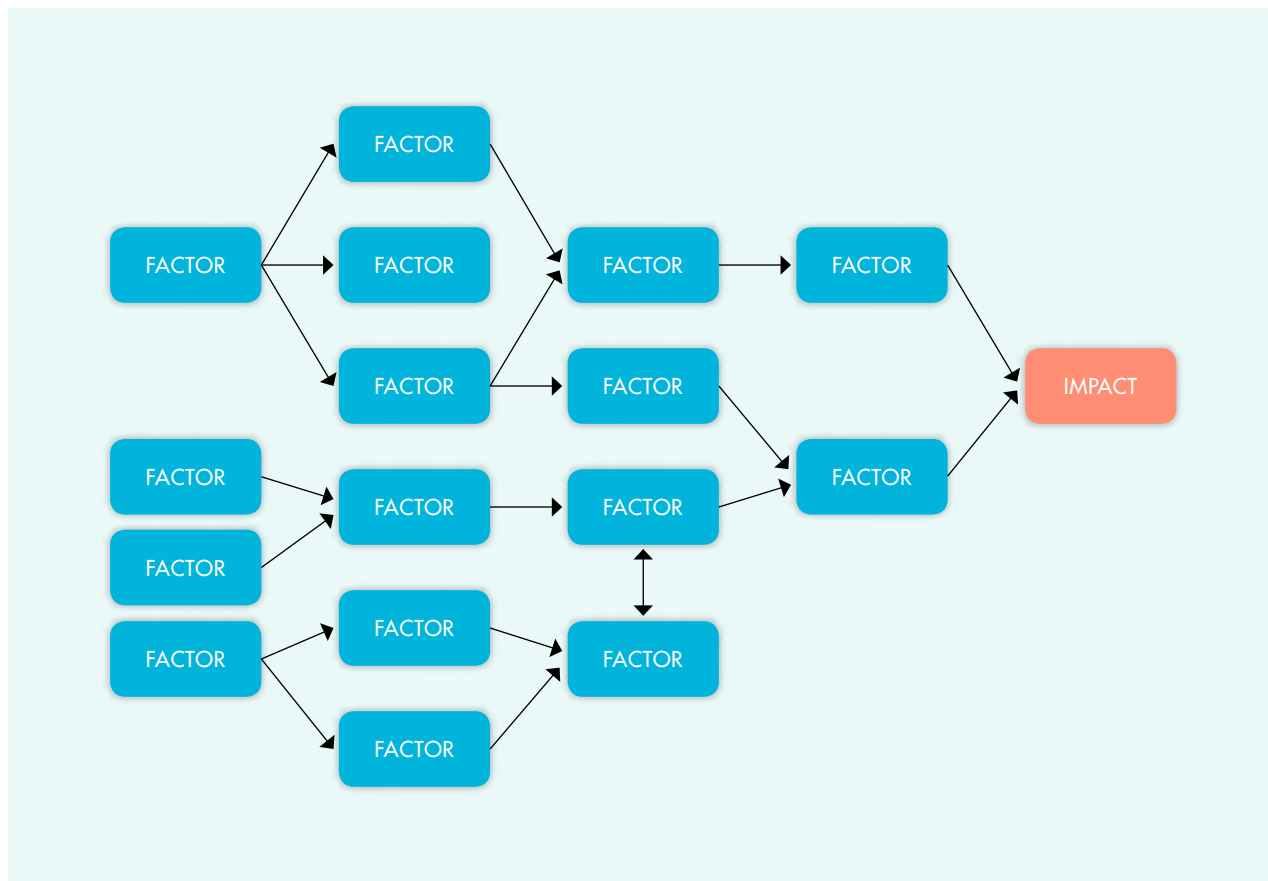
understanding of the problems, how they are interrelated, and what the root causes are. It can also help to distil what interventions should be used to address the problems that the project is trying to address, and what the scope of the project should be.

In the factor analysis part of the context analysis, there is a close link with the research phase of data collection, phase three of the Handbook, in which we zoom into the more specific data collection and data-based decision making questions:

- **Context:** What is the key problem that you are trying to address and why is it important?
- **Environment:** What factors contribute to this problem?
- **People:** Who does it directly affect?
- **Current problem solvers:** Who is already working on this issue?
- **Time:** At what time intervals are decision being made about this problem?
- **Existing data:** What data relevant to this problem already exists?

3. More guidance on how to do a problem tree: <https://www.odi.org/publications/5258-problem-tree-analysis>

FIGURE 4 EXAMPLE OF A CONCEPTUAL MODEL



Step two: Designing a Theory of Change (ToC)

Designing a ToC is the central part of any ToC based approach to project design. As the name suggests, a Theory of Change is a hypothesis of how we think change occurs. It consists of a visual diagram and a narrative with causal assumptions – “if we do X, Y happens because we believe Z.” ToCs are also referred to as Intervention Logics or Results Chains. When designed with an understanding of factors and stakeholders, a ToC helps to make sense of and navigate the complex environment in which the project is operating. A ToC is a perception of reality which is shaped by the norms, values, experience and beliefs of the people who create it. It’s therefore important to involve different stakeholders, to be sure different perspectives are captured in the ToC. Designing a ToC together results in a shared understanding of the partners involved, and co-ownership of the results which need to be achieved to contribute to the desired impact.

When designing a ToC, the first thing that needs to be identified is the desired long-term impact the project wants to achieve or contribute to (see Figure 5). An impact statement without targets is enough. After identifying the intended impact, the participants can determine which outcomes need to be achieved in order to reach that long-term impact. In order to phrase an expected outcome, it helps to use the following mnemonic: “who should be doing what differently?”

As defined in the glossary, an outcome is a change in the behaviour, relationships, actions, activities, policies, or

practices of an individual, group, community, organisation, or institution.

Once all expected outcomes have been identified on cards, they can be organised on the wall in logical cause and effect relationships. When a project has several topics or issues, separate “pathways of change” can be built for each of them, with small teams who can zoom into each specific topic. The resulting pathways of change can then be collectively connected at the end to have the full Theory of Change.

Once the stakeholders have identified one or multiple pathways of change, they can come up with strategies to set into motion the causal chains of events. Strategies are a general description of what the project needs to do to make the expected outcomes happen. Every strategy will have a pathway of change. Strategies are a general description of what needs to be done; the more specific activities in the strategies will be defined later. In reality, strategies may already be determined before the expected outcomes are mapped and an impact is defined. In that case, the ToC design exercise will help to identify the expected outcomes and their causal relationships and understand how the strategies will lead to the envisioned impact.

Once the strategies, expected outcomes, impact(s), and their linkages are identified, the underlying causal assumptions should be made explicit. Trying to document these assumptions can lead to the identification of weak spots in the Theory of Change and, at the same time, results in stakeholders becoming aware of each other’s visions

of reality. When phrasing causal assumptions, it can be tempting to start a circular reasoning.

Avoid phrases like “A leads to B, because B is the result of A”. Instead, try reasoning as follows: “If we do action X, we will contribute to outcome Y because we believe that Z.”

For example, if we support the ministry of water with data collection and analysis for water point mapping, then they will use the map for decision making on investment priorities, because they were actively involved in the identification of the problem (no updated information on functionality of water points) and feel co-ownership of the solution (data collection for evidence-based decisions).

Co-creating a ToC with all stakeholders involved will lead to a common understanding of how change happens, create awareness on different norms and values between stakeholders, generate co-ownership of the project, help to decide on the scope of the project, and support decision-making on what interventions should be pursued to achieve the biggest impact by whom. It can also expose gaps in your activities or show you where there is an overlap with the activities of other actors or projects.

FIGURE 5 FIRST STEPS TO DEVELOP A THEORY OF CHANGE

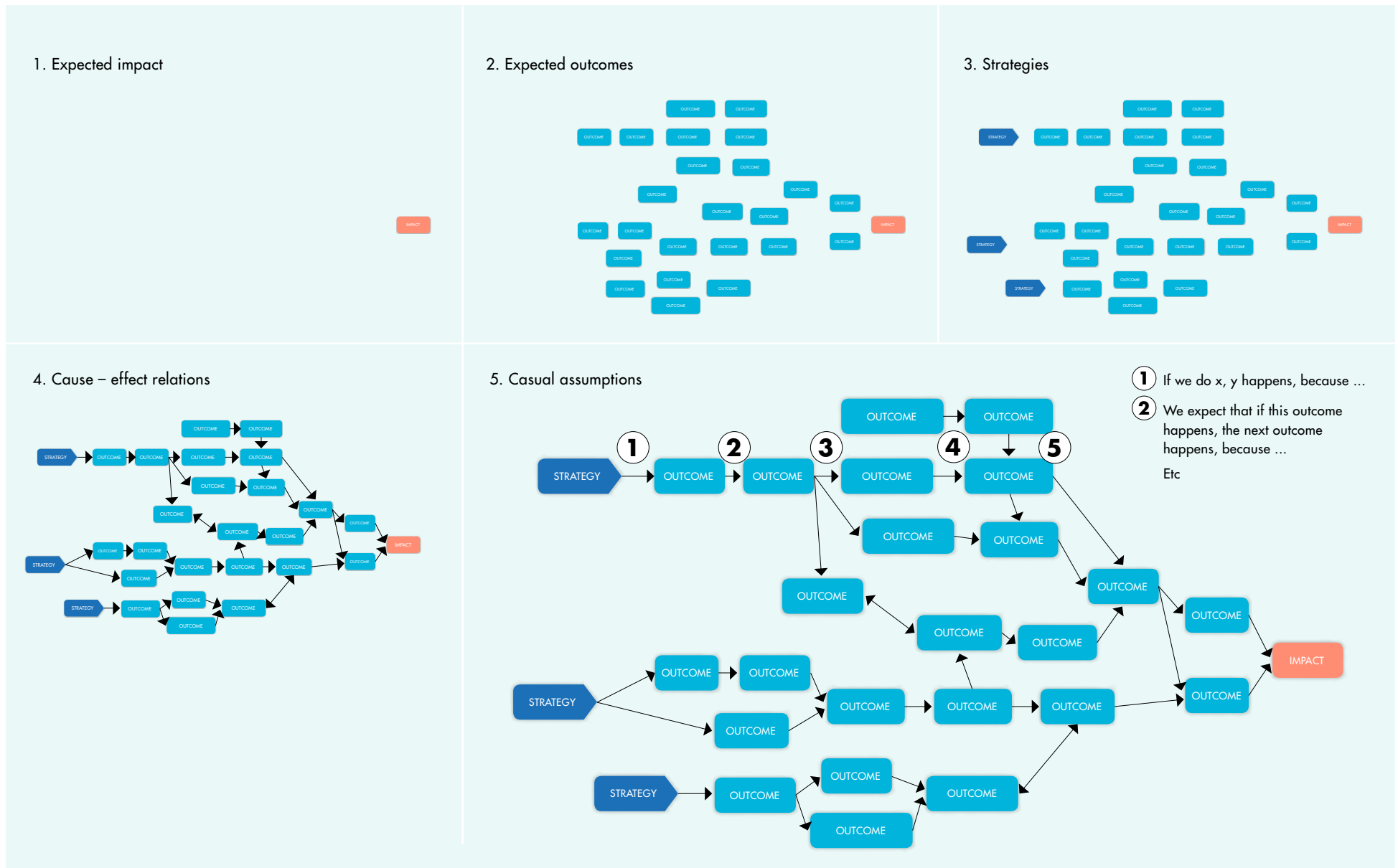
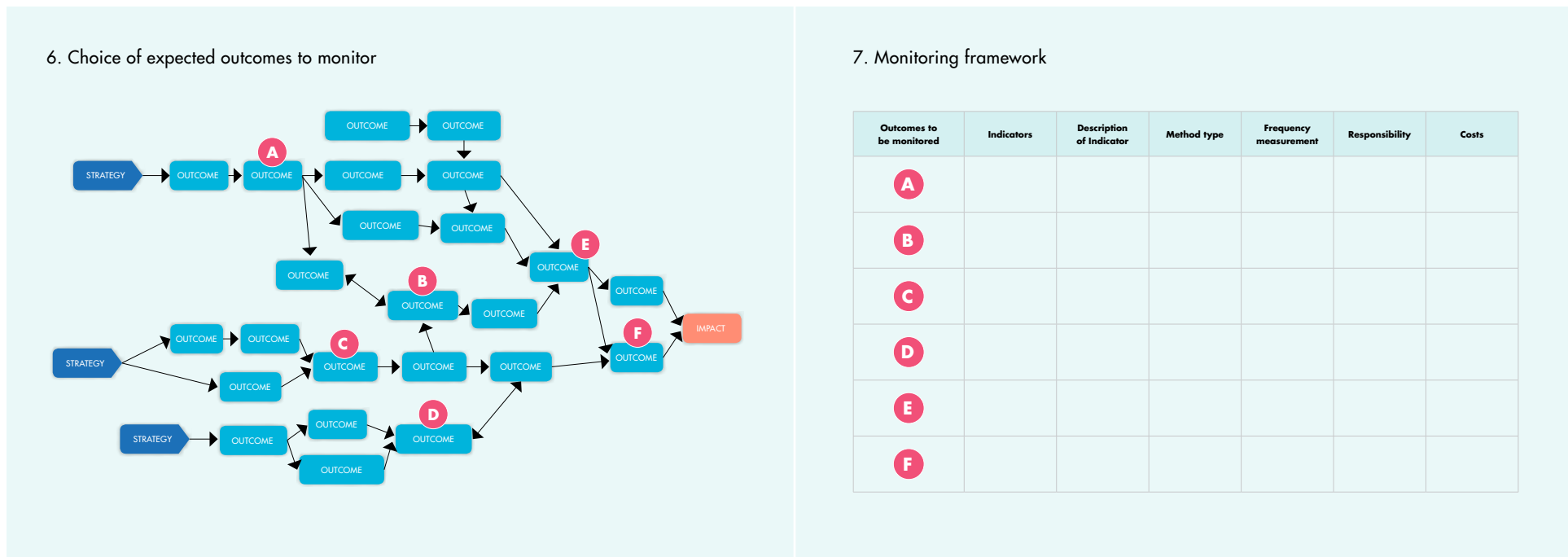


FIGURE 7 LAST STEPS TO DEVELOP A THEORY OF CHANGE



Step three: Designing a flexible planning, monitoring, evaluation, and learning (PMEL) framework

A Theory of Change which has been designed together with the relevant stakeholders and which is based on an understanding of the context is the foundation for a relevant and useful PMEL framework (see Figure 7). All steps between the strategy and the impact are expected outcomes that in theory can be monitored, but in practice should not. Only

a small selection of the outcomes should be selected to be monitored. For example, out of 25 expected outcomes in the ToC, monitor four or five. Together, the stakeholders can define which outcomes are the most important, as well as practical, to monitor across the project.

Each expected outcome can have one or more indicators that make it possible to measure the progress towards this outcome. For example:

Outcome: “The local government effectively monitors WASH infrastructure in district X”

Indicator: “Number of water points monitored on functionality and water quality in district X”

Outcome: “The national government allocates more budget to WASH”

Indicator: “Amount of budget that is allocated to the ministry of water on a yearly basis”

Together, these results and indicators constitute the project's monitoring framework. For each indicator, the method used to measure it needs to be determined. Is the information that is needed quantitative or qualitative in nature? Is the information provided by end users, data collectors, sensors, or secondary sources? At what interval should each indicator be measured? Is there baseline data available? What are the best tools to collect the data? For more information on how to design a survey and sample group, see phase four of the Handbook.

Conclusion

The three steps of a Theory of Change based approach to project design are an effective way of working collaboratively towards a common understanding of what the project should achieve and what activities the project should focus on to contribute to impact. It is important to note that a ToC by nature is subject to constant change. During the implementation of a project, the context may change, resulting in the need for an adaptation of the ToC. During implementation, you may realise that outcomes are missing, causal assumptions need to be adapted, or interrelations changed. Theories of Change therefore need to be reviewed at least once per year, based on monitoring findings and a context analysis update. At the same time, a ToC revision workshop can form an excellent basis for writing an annual activity plan, keeping the ToC at the heart of your project at all times.

The monitoring framework of a project shows what data needs to be collected to describe or score the indicators which are used to monitor progress, as well as which methods will be used. This is how Theory of Change based project design links to the following phase, data research, where we assess whether any of the necessary data is already available from secondary sources, and what data still needs to be gathered by the project.

Suggested reading

A useful manual which is easily accessible is the Hivos Theory of Change Guide.

 https://hivos.org/sites/default/files/publications/hivos_toc_guidelines_final_nov_2015.pdf

Handbook phase three: Conduct data research

Handbook phase three

Conduct data research

Authors

Annabelle Poelert (Akvo.org),
Karolina Sarna (Akvo.org)

Contributors

Anita van der Laan (Akvo.org),
Rajashi Mukherjee (Akvo.org),
Rob Lemmers (Faculty of Geo-Information
Science and Earth Observation (ITC) of
University of Twente)

In phase two of the Handbook, we discussed the design of a project, which helps you to determine what data you need in order to inform stakeholders and help them make data-based decisions. Before rushing out into the field to collect this data, however, it is advisable to take a step back and assess what data is already out there, how reliable it is, and how you can actively involve your stakeholders in making sure you collect relevant data and you communicate it in the right way.

Data research is a method that helps you to systematically assess existing data, allowing you to identify where there are gaps in the data and where you can add value with your data collection project. At the same time, data research gives you the tools to think about your stakeholders and audience. In this article, you'll find an overview of four consecutive steps that will help you in conducting data research. These steps build on the two earlier steps described in phase two: 'Defining a clear and answerable question/problem you are trying to solve', and 'mapping all stakeholders involved in the problem'. The four consecutive steps are:

Data research in four steps

Step one: Make an inventory of existing data/evidence

Avoid double work and explore what data is readily available for use in your organisation.

Step two: Evaluate existing data

Assess the reliability of existing data and determine its usefulness.

Step three: Perform a gap analysis

Identify the data that is not readily available and will need to be collected.

Step four: Understand who will use your data

Think about which stakeholders to actively involve in your data collection and how to present the results to them.

Step one: Make an inventory of existing data/evidence

Once you have identified what data needs you have within your project, you will need to start gathering it. Some data may be readily available, while other data may still need to be captured. You can start off by making an inventory of existing data.

First of all, look into the data resources of your own organisation, including what is gathered in reports and stored in databases. Consider both quantitative data, expressing a certain quantity, amount or range, and qualitative data, which is more descriptive, resulting from small scale surveys, focus group discussions, observations and interviews. You can then think about what data may be available and easily accessible outside of your organisation. Are there any data sharing platforms or other organisations that deal with the same problem or try to answer the same question? What data do they have on this problem? Is it open access? Even if data is not openly accessible, it might be possible to persuade this organisation to share its data.

Step two: Evaluate existing data

Once you have created an inventory of existing data sources, it is important to evaluate the existing data on its accessibility, granularity, credibility and relevance. The following questions can help you understand whether the existing data is available for usage, detailed enough and

has the right scale, and reliable enough for you to use in your programme:

- Is the data openly available, or does it require special permission to access? (Accessibility)
- Is the data structured in a way that is useful for your project? (Relevance)
- How often is the data collected? (Granularity)
- How granular or detailed is the data geographically? (Granularity)
- How granular or detailed is the data demographically? (Granularity)
- When was the data collected? How long has it been retained? (Relevance and Granularity)
- Do the current problem solvers use it for decision making, evaluation, or something else? (Credibility)
- Who collected the data? What was the purpose of their data collection? Has the data been cleaned and/or analysed? And if so, in what way? (Credibility)

Step three: Perform a gap analysis

Now that you have identified the data sources that are available to you and what data you can use for your project, you need to think about what data you still need to collect to answer your questions. To do so, it helps to

ask the following: “what data do I need to answer my questions or describe my indicators?” It’s important that, in the first instance, you don’t think about restrictions that might be apparent in collecting this data. Only after identifying the data you need should you start considering potential restrictions, such as time, (financial) resources and feasibility. It might turn out that data you initially deemed infeasible to collect isn’t as difficult to gather after all.

Once you have identified all the data gaps, take a critical look at the data you have identified as necessary. Do you really need to collect all that data? And what are you going to use all the different elements for? Although it’s tempting to collect data that you may think will be useful in future, a general rule of thumb is that less data is more. It’s better to focus on the things that really matter and minimise complexity. It is less expensive, less time consuming, and you don’t run the risk of collecting the wrong data.

Step four: Understand who will use your data

If you are collecting data to contribute to solving a problem, or to underline the importance of addressing a certain problem, keep in mind that it is crucial to involve all relevant stakeholders from the start of the data research process. This will create ownership of the data, ensure relevance and usefulness of the data, result in communities feeling represented by the data, and avoid decision makers turning a blind eye or questioning the credibility of the data. Start your data collection exercise with an inventory of what the

different stakeholders want to know and how you are going to reach them. Sharing the data with the people directly involved in the problem empowers them to take action. However, this involves thinking about how to share the data in an understandable and accessible way. In remote communities, accessing the data online may prove to be difficult, and radio stations or distribution of offline materials may be a better mode of dissemination. You might want to consider making a data dissemination plan, in which you identify your stakeholders and their respective communication channels. For more information on how to reach your target audience refer to phase eight of the Handbook.

Conclusion

Data research is an approach that will help you to create focus in your project. Thinking in this structured way about data gathering will avoid collection of duplicate and encourage everyone involved to determine the quality and usefulness of available data. This method also allows you to assess whether the data you are collecting is truly relevant to your project, and the different stakeholders involved, and forces you to think about how to disseminate the data to them before the data collection has actually started.

Handbook phase four: Design survey and sampling method

Handbook phase four

Design survey and sampling method

Author

Rajashi Mukherjee (Akvo.org)

Contributors

Camille Clerx (Akvo.org),

Hans van der Kwast

(IHE Delft Institute for Water Education),

Nikki Sloan (Akvo.org),

Stefan Kraus (Akvo.org)

Once you have completed phases two and three of the Handbook, you will have identified the data requirements for your project and the location/s of data collection. In this section, you will learn how to collect your data, from how to frame and ask the questions to whom and how many people you should ask.

Sample design

As outlined in phase three, you will need a thorough overview of the data that is already available and the data that is missing before you start your data collection.

Existing data from external sources is called secondary data, whereas data that you collect yourself by conducting a survey is termed primary data. This chapter will outline ways in which primary data is collected using an example in which a survey on characteristics (tax, education levels, etc.) is collected on residents in five towns. The towns are of different sizes and have a total of 3,200 households. These 3,200 households make up the target population for your survey.

Step one: Define your sample and target population

At times, your survey may require you to cover the entire target population, as in the case of mapping or population studies. That's usually referred to as a census survey. However, target populations are generally large and expensive to survey. In our example, it may not be feasible to visit all 3,200 households of the five towns. Instead, you'd want to choose a smaller sample that would be representative of the population and reflect its characteristics. A survey that is done on a smaller number of the target population is referred to as a sample survey. You can infer your findings for the entire population based on this representative sample. In the following sections, we'll describe the different terminologies that are associated with sample surveys, such as sample size and survey design. These concepts will enable you to determine the number of surveys needed to accurately reflect the true characteristics of a population and to choose the best method of selecting a sample from that population.

The HOW of data collection in three parts: sample design, survey design and data collector selection.

Sample design

1. Define your sample and target population
2. Define your sample size
3. Define your sampling technique
4. Minimise sampling error

Survey design

1. Outline the design of your survey
2. Adopt good practices while designing your questionnaire
3. Design your questionnaires according to the data type
4. Choose the question type to match your data needs
5. Build in steps for quality control

Data collector selection

1. Select data collectors
2. General advice for data collectors

Before you begin designing your survey you should have:

- Clearly defined your research questions
- Identified the characteristics / parameters you want to collect information about
- Completed the secondary data review and identified data gaps AND
- Identified the target population and the geographical location/s for your survey

Step two: Define your sample size

The first step in your sampling exercise will be to decide on an appropriate sample size. There are no strict rules for selecting a sample size. You can make a decision based on the objectives of the project, time available, budget, and the necessary degree of precision.

In order to select the appropriate sample size, you will need to determine the degree of accuracy that you want to achieve. For this, you'll need to establish the confidence interval and confidence level of your sample.

The confidence interval, also called the margin of error, is a plus or minus figure. It is the range within which the

likelihood of a response occurs. The most commonly used confidence interval is ± 5 . If you wish to increase the precision level of your data, you would further reduce the error margin or confidence interval to ± 2 . For example, if your survey question is "does the household pay tax?" and 65% of your sampled households say "yes," then using a confidence interval of ± 5 , you can state with confidence that if you had asked the question to all 3,200 households, between 60% (i.e. $65-5$) and 70% (i.e. $65+5$) would have also responded "yes."

The confidence level tells you how sure you want to be and is expressed as a percentage. It represents how often the responses from your selected sample reflect the responses of

the total population. Thus, a 95% confidence level means you can be 95% certain. The lower the confidence level, the less certain you will be.

Most surveys use the 95% confidence level and a ± 5 confidence interval. When you put the confidence level and the confidence interval together, you can say that you are 95% sure that, if you had surveyed all (3,200) households, between 60% and 70% of the households of the target population would have answered "yes," to the question "does the household pay tax?".

The size of your sample may be determined using any standard sample size calculator such as Survey Monkey or Raosoft. Using a standard sample size calculator (as can be

seen in Table 3) for our example of 3,200 households in five towns, we can examine the difference in sample sizes based on different confidence levels and intervals.

Option A

If you decide on a 5% confidence interval and want to achieve a 95% confidence level, the sample size will be 345 households.

Option B

If you wish to have higher accuracy and increase the confidence level to 99%, the recommended sample size would be 551.

Option C

For an even higher accuracy you could choose a 2% confidence interval and 99% confidence level and arrive at a sample size of 1807.

If time and resources permit, you could opt for larger samples and choose option C, to survey 1807 households. However, the quality of your findings are likely to only be marginally better than with option A or B, as the rate of improvement of accuracy gradually diminishes with the increase in sample size. The size of your sample should therefore be decided by the objectives of the study and resources available.

TABLE 3 CALCULATE YOUR SAMPLE SIZE

Factors	Factor description	Option A	Option B	Option C
Population (no.)	The total population that your sample will represent	3,200	3,200	3,200
Confidence level (%)	The probability that your sample accurately represents the characteristics of your population	95%	99%	99%
Confidence interval (%)	The range that your population's responses may deviate from your samples	5	5	2
	Sample size calculated	345	551	1807

Step three: Define your sampling technique

Once you have chosen the sample size for your survey, you will need to define which sampling technique you'll use to select your sample from the target population.

The sampling technique that is right for you depends on the nature and objectives of your project. Sampling techniques can be broadly divided into two types: random sampling and non-random sampling.

Random sampling:

- Simple random sampling with or without replacement
- Stratified random sampling
- Systematic random sampling

Non-random sampling:

- Purposive or subjective sampling
- Convenience sampling
- Judgment sampling
- Quota sampling
- Snowball sampling

TABLE 4 CALCULATE STRATIFIED RANDOM SAMPLE

1	2	3	4
<i>Location</i>	<i>Population size</i>	<i>Proportion (%) of population</i>	<i>Stratified sample size</i>
Town 1	1200	38%	129
Town 2	900	28%	97
Town 3	800	25%	86
Town 4	180	6%	19
Town 5	120	4%	13
Total	3200		345

Random sampling

As the name suggests, random sampling literally means selection of the sample randomly from a population, without any specific conditions. This may be done by selecting the sample from a list, such as a directory, or physically at the location of the survey. If you want to ensure that a particular household does not get selected more than once, you can remove it from the list. This type of sampling is called simple random sampling without replacement. If you choose not to remove duplicate households from the list, you would do a simple random sampling with replacement.

Systematic sampling is the most commonly used method of random sampling, whereby you divide the total population by the sample size and arrive at a figure which becomes the sampling interval for selection. For example, if you need to choose 20 samples from a total population of 100, your sampling interval would be five. Systematic sampling works best when the population is homogeneous, i.e. most people share the same characteristics. In our example, the sampling interval would be nine ($3200/345=9$ for a 95% confidence level and 5% confidence interval). You would thus select every ninth household in a town.

However, populations are generally mixed and heterogeneous. To ensure sufficient inclusion of all categories of the population, you will need to identify the different strata or characteristics and their actual representation (i.e. proportion) in the population. In such cases, you can use the stratified random sampling technique, whereby you first calculate the proportion of each strata within the population and then select the sample in the same proportion, randomly or systematically, from all the strata.

If we take our earlier example of five towns, to calculate a stratified random sample, you will need to calculate the proportion of each town within the sample size of 345 as shown in Table 4. Column three gives the proportion of each town of the total population (3,200). In column four, the sample size (345) is proportionately divided across the five towns. For example, town three, which is 25% of the total population, will select 86 households with a sampling interval of nine (i.e. $800/86$) in the same manner as was done for systematic sampling.

Non-random sampling

In non-random sampling, the sample selection follows a particular set of conditions and is generally used in studies where the sample needs to be collected based on a specific characteristic of the population. For example, you may need to select only households which own a car, or have children under six years of age. For this, you would consciously select only the 345 or 551 households that have those characteristics. Also termed purposive or

subjective sampling, non-random sampling methods include convenience, judgment, quota and snowball sampling.

Step four: Minimise sampling error

It is normal to make mistakes during sample selection. Your efforts should therefore always be to reduce the sampling error and make your chosen sample as representative of the population as possible. The robustness of your sample depends on how you minimise the sampling error. The extent of errors during sampling varies according to the technique or method you choose for sample selection.

For samples selected randomly from a target population, the results are generally prefixed with the +/- sampling error, which is the degree to which the sample differs from the population. If your study requires you to know the extent of sampling error that is acceptable for the survey, you can select a random sampling technique. In random sampling, you will be able to regulate the survey design to arrive at an acceptable level of error. In a non-random sample selection, the sampling error remains unknown. Thus, when your sample survey needs to infer the proportion of a certain characteristic of the target population, you can select a random sampling method. But if you want to know the perceptions of residents regarding taxation laws or the school curriculum, you would want to capture as many perceptions as possible, and therefore select a non-random method in situations where sampling errors or sampling for proportionality are not of concern. Non-random sampling techniques can be very useful in situations when you need to reach a targeted sample with specified characteristics very quickly.

Survey design

Step one: Outline the design of your survey

After you've defined your sample, you can decide how you'll collect the data from the respondents. Your survey design will include a survey format with a list of questions which correspond to your data needs and the frequency at which the data will be collected. The frequency of data collection depends on the type of data being collected. Baseline or mapping studies are one time surveys while tracking/monitoring surveys are conducted at time intervals. The periodicity of data collection is determined by the project goals and objectives and the related set of indicators listed in your Theory of Change.

Step two: Adopt good practices while designing your questionnaire

A questionnaire is likely to be most effective if you KISS: Keep it short and simple. If you don't have a satisfactory answer to what you will do with the answer to a question, leave it out. Avoid the temptation to add a few more questions just because you are doing a questionnaire anyway. If necessary, place your questions into three groups: must know, useful to know and nice to know. Discard the last group, unless the previous two groups are very short.

Start your questionnaire with an introduction or welcome message clearly stating who you are and why you want the information in the survey. A good introduction or welcome message will encourage respondents to cooperate and

participate. In case of sensitive or private information, reassure your respondent that their responses will not be revealed. In some cases e.g. child/underage surveys, it may be mandatory to seek the consent of the respondent or guardian.

While designing a questionnaire, it's important to reflect on how the order of the questions can impact the results of your survey. Ideally, you should:

- Place the easiest and most pleasant to answer questions at the beginning of your survey.
- Group together questions on the same topic.
- Leave difficult or sensitive questions until near the end of your survey.
- Address the data collector observations, validation issues, GPS readings, photographs, testing (e.g. water quality testing) at the end of the survey, and avoid breaks while interviewing the respondent.
- Use a logical or natural order to answer choices, presenting positive to negative or excellent to poor scales and agree-disagree choices in that order.

Step three: Design your questionnaire according to the data type

Broadly speaking, there are two types of data: quantitative data and qualitative data. Quantitative data is collected with a structured questionnaire which may have closed ended

questions (i.e. with a list of options to choose from) and/or open ended questions, depending on the type of information you need.

Qualitative data is often essential to understanding the context and explaining the quantitative data. It is generally collected as free text, which may be translated into numbers by classifying the information or assigning codes. It is recommended that qualitative information be used sparingly, where the possible responses are not known in advance and will add value to your survey. This is because qualitative answers tend to take longer to check, clean and process.

Most questionnaires will gain value using combination of both types of data. Your questionnaire will depend on the objectives of your project. If you need data to monitor the status of water points across a city, you are likely to ask the following questions:

- What type of waterpoint is it? Respondents can select from a list of options, e.g. hand pump, well or tap.
- Where is the waterpoint located? Respondents can provide the name of the city/village and a GPS reading.
- Is the waterpoint functioning? You'll need to clearly define functionality to ensure a common understanding for all data collectors.
- Is the water safe to drink? For this, you may need to test the water for certain parameters, document perceptions on water safety from respondents, or collect healthcare information from existing records.
- Who owns or is responsible for this water point and is it maintained? You can ask whether it is publicly owned (government) or privately owned, what type of repair (major or minor) has been done in the last few years, how much it cost and who paid for it. Again, define what constitutes major and minor repairs.

In the above example of the waterpoint monitoring, questions (a) and (b) are examples of (structured) questions to collect quantitative data. Questions (c), (d) and (e) could be framed to collect both quantitative and qualitative information. Question (c), where you ask about

perceptions of safe water, is an example of qualitative information, where you will record the responses verbatim and enter the data as free text.

Step four: Choose the question type to match your data needs

There are three basic ways in which questions are designed in surveys:

1. Multiple choice (e.g. *Have you watched this movie?*)

Yes No

2. Numeric open ended (e.g. *How many times have you watched the movie?*)

3. Text open ended (e.g. *What did you like about this movie?*)

Rating Scales and Agreement Scales are two common types of questions also used to qualify multiple choice questions.

1. How would you rate the movie?

Excellent Good Fair Poor

2. On a scale, where '10' means you have enjoyed this aspect the most, how would you rate the movie?

Acting _____ Editing _____ Casting _____

Direction _____ Production _____ Cinematography _____

3. How much do you agree/disagree with the following statements?

	Strongly Agree	Agree	Disagree	Strongly Disagree
a) <i>The movie has a strong social message</i>				
b) <i>Children should not watch the movie</i>				
c) <i>There is unnecessary violence in the movie</i>				
d) <i>The storyline of the movie is weak</i>				

While designing a closed ended questionnaire, you should try to include the maximum possible list of relevant alternatives as answer choices. This helps to systemise and categorise respondent's answers and saves time on text entries. However, this also reduces the scope for capturing detail and you will need to decide how flexible you want to be and to what extent the additional detail will improve the findings of the survey. Choosing a question type is largely based on how you want the data to come out, which depends on how you want to use the data.

You can pre-test the questionnaire before the survey if you want to generate a list of alternative question types. When you're unsure about the possible answer choices, use an open ended format by adding "other (specify)" as one of the alternatives. Also allow a "don't know" or "not applicable" response to all questions, except to those in which you are certain that all respondents will have a clear answer.

In your survey, some questions may be dependent on responses to other questions. For example, in our case of the moviegoer, if the respondent's answer to question one is "no," i.e. they have not watched the movie, the rest of the questions would be redundant. In this case, you would add an instruction "continue survey only if the response is 'yes' in question one."

Step five: Build in steps for quality control

Pretesting your questionnaire is recommended as it will help you to ensure that your survey design is in sync with your data needs. During a pre-test, you get to know whether the questions have been worded properly and are soliciting the expected responses. You may need to edit/remove/add questions and explanations after a pre-test. It also gives you a chance to check the quality of the data collectors.

Digital technology allows survey designs in which you can factor in quality checks. Questions can be marked as mandatory or optional and submission of forms can be made dependent on completion. Also, data quality can be periodically checked for timely feedback and course correction.

Data collector selection

Step one: Select and prepare data collectors

The next step will be to define who will collect your data. Emails, Internet and telephone conversations are popular methods of collecting data. However, in this chapter, we will talk about data gathering with data collectors using mobile/tablet based surveys or paper surveys.

Data collectors should be selected with care and should be sufficiently briefed/trained on the questionnaire before data collection. A guide with a set of instructions and an explanation for each question is always advisable. The guide should also provide definitions for each alternative to ensure a common understanding across data collectors. For example, if the questionnaire requires the data collector to list the type of a water source as improved or unimproved, the guide should clearly mention what constitutes an improved source (i.e. hand pump, tap) vis a vis and unimproved source (open well, pond).

Step two: Guidelines for working with data collectors

Guidelines	Importance
Identify and train more data collectors than you need.	This will enable you to replace data collectors at short notice, without having to train them again.
Select data collectors who have an understanding/familiarity with the local culture and sociology.	This will help to build a rapport quickly with the respondents and demonstrate sensitivity to local cultures.
Select data collectors who are conversant in the local language/dialect.	This will help communication during the survey. By minimising the use of interpreters/translators, you'll also save time and resources.
Brief your data collectors well before the survey. Pre-test the questionnaire through sample visits or simulation exercises.	You need to ensure that the data collectors have understood the questionnaire properly. During field tests or simulations you will be able to judge their capacities and train/guide them further, if needed.
Advise data collectors to be sensitive and value the time being given by the respondent.	Demonstrating sensitivity is the best way of thanking a respondent for the time they are voluntarily giving you.
Advise data collectors to be courteous and respectful to the respondents.	Always remember that a respondent is not obliged to answer your questions. You will need to reassure them that their responses matter.
Advise data collectors to select time slots when the respondent is free to talk.	The timing of your survey is important because if respondents are busy or preoccupied, they are more likely to give a wrong response.
Pretest the questionnaire through sample visits or simulation exercises.	You need to ensure that the data collectors have understood the questionnaire properly. During field tests or simulations you will be able to judge their capacities and train them if needed.

Conclusion

Regardless of the field of study, the goal of all data collection projects is to capture convincing and credible evidence. A variety of accepted data collection techniques are available. You can use your judgement to select the best method, based on the objectives of your research and the available resources. The key to robust data collection is to select the most appropriate method that strikes a balance between improving accuracy while protecting the credibility and reliability of your data. A clean and representative data set is essential in analysing data and helps to reduce time spent on preparing the data before analysis, which you can read more about in phase five of the Handbook. The choice of survey methods, questionnaire design and data collectors' capacities largely contribute to this.

Suggested reading

Sampling techniques and sampling types

 <https://www.slideshare.net/hafizahhajimia/research-method-sampling>

Statistical Methods – 13 Sampling Techniques

 <http://www.statstutor.ac.uk/resources/uploaded/13samplingtechniques.pdf>

What is Qualitative Research? Definition, Sources & Examples

 <https://study.com/academy/lesson/what-is-qualitative-research-definition-sources-examples.html>

05

Handbook phase five: Prepare data collection

Handbook phase five

Prepare data collection

Authors

Camille Clerx (Akvo.org),

Nikki Sloan (Akvo.org),

Stefan Kraus (Akvo.org)

Contributors

Hans van der Kwast

(IHE Delft Institute for Water Education),

Harro Riedstra (Akvo.org),

Rajashi Mukherjee (Akvo.org)

Phase four of this Handbook focuses on how to collect data by looking step by step at survey design and sample design. Once that's done, it's time to put all the work you've done into action in the field. Preparation is key to ensuring the high quality of the end result and the efficiency of the data collection exercise. Before going out into the field to collect data, it's important to adequately plan and prepare all aspects of the process.

Preparing data collection in six steps

Step one: Map processes and logistics

Step two: Recruit team and set responsibilities

Step three: Prepare survey

Step four: Train participants

Step five: Pilot data collection process

Step six: Finalise and launch data collection

Following these steps will ensure that your data collection goes as smoothly as possible and yields high quality and timely data.

Practical suggestions to prepare data collection

- Test the survey thoroughly before data collection.
- Ensure data collectors have a consistent understanding of terminology, survey logic, and the purpose of each question.
- Define clear roles, responsibilities, and timelines, and communicate these to the team members.
- Build contingencies into the data collection plan – things rarely go as planned and it is important to have a plan B (and C, D, E).

Step one: Map processes and logistics

Mapping out the data collection process from start to finish will assist in determining the logistics that need to be prepared. This includes creating a list of all necessary tasks and roles throughout the data collection process, and considering the necessary skill sets for successful completion. Doing this early helps to identify potential gaps or issues that may arise, and roles that need to be filled. The following steps should be taken when mapping out the process.

Seek approval and inform the right people

In many locations, it is important to inform local authorities and the local population (interviewees) before proceeding with data collection. In some cases, you may need a certificate or formal approval from local authorities in order to conduct the surveys. You may need to organise a local information session, or assign a community liaison who will be a point of contact and responsible for handling incoming questions. Besides the formal procedures, it's also advisable to get informed about the attitude of the local population towards the data collection itself as well as any tools you might use to collect the data. If you have to do water quality tests for example, make sure the people know what you are doing to avoid misconceptions about the materials you are using.

Finally, it's important that you gain the informed consent of the participants interviewed. This may require a signature or verbal consent, so be sure to keep a copy of the consent form and information about the project at the ready.

Clarify logistics and budget constraints

Logistics and budget will take up a large part of this planning stage. During field work, data collectors can be exposed to different risks. Knowing how to deal with those risks and creating a safe environment is part of everyone's role, and having responsibilities clearly defined is essential. It is necessary to define what equipment will be used, who owns it, and what happens if something is broken or lost.

For field work logistics and budgeting the following points need to be considered:

Equipment: What equipment will data collectors need for their field work and travelling?

- What technical features does the material require, especially phones for digital collection systems?
- Does one of the parties have administrative constraints in ordering equipment? Government bodies often have protocols to follow which can delay the procurement process for several months.

- Is there any equipment that needs to be shipped from another country and, if yes, what are the procedures and timeframes for clearing the shipment? To whom will the shipment be addressed? Is the help from custom brokers needed and, if yes, who will be responsible for starting the procedure?
- For paper based surveys, you will need to arrange printing of the final survey form. Who will be responsible? How many forms will need to be printed?

Local context: What is the safety situation in that area and are special precautions necessary? At what time of the year are the data collectors travelling, and are there any constraints related to climate? What means of transportation is most suitable for the local context? Are the data collectors and their equipment insured and, if not, how will any accidents be handled? What official documents do the data collectors need to carry with them?

- Accommodation and infrastructure: Take care to arrange accommodation and means of communication in advance, especially in rural areas. Is there network reception available and do you need internet connection? Is electricity readily available at the destination, will power packs and battery charging packs need to be purchased?

- **People:** Are there travel schedules available? Take holy days and holidays into account during which data collectors or interviewees might not be available. Seasonality may also have a major impact on the availability of them both. For instance, try to plan the data collection outside planting and harvest seasons.
- **Budgeting:** Who will cover the costs of the aforementioned elements? How will the payment of the data collectors be calculated and what will the rate be? Will meal costs be covered? Is the whole sum given in advance or will it be paid in instalments?

All of the above also applies to supervisors who perform field visits.

Check communication system

Communications will need to work in two directions – from the data collectors to the data managers, and back from the data managers to the data collectors. In case any answers need to be cross checked, ensure that a clear identification system is in place that links data collectors with each survey that they collected. If using a mobile-based data collection system, ensure that each device used for mobile data collection is linked to a data collector.

Also ensure support materials and a support system are in place in case data collectors and other users require assistance while in the field.

Step two: Recruit team and set responsibilities

Now that roles and responsibilities have been set for practical matters, the same has to be done for the processes within the data collection chain. This will allow the whole team to work well together and focus on evaluating and improving the quality of the data. For example, who will be responsible if there are problems with the survey, who will check and clean the data, and who will oversee the performance of data collectors?

Defining roles

Below is an outline of general roles that appear during the data collection process. When assigning roles and responsibilities consider the skillsets already available – who may already have experience that would make them suitable for a specific role? Who will be responsible for survey creation and data checking, cleaning, and analysis?

Role	Purpose	Tips
Data collector	Data collectors collect the data and interview participants.	The field work can be very demanding and it is needed to make data collectors aware of what it entails during training.
Team lead	The team lead will supervise small groups of data collector, and act as contact points between the data collectors and the coordinator/data team.	
Data collector coordinator	The data collector coordinator will ensure that the logistics of the training sessions and the data collection are prepared in advance.	Should be a centralised person, and data collectors should know who they are and how to contact them. If using a digital system, this person doesn't have to be the same as a dashboard manager.
Data processor/ digitiser	Data processors will digitise the surveys if they are paper based so that the data is aggregated and ready for analysis. This can be the scanning/typing up of responses, or the coding of any qualitative answers.	Depending on the size of your sample, this role may be filled by multiple people. Therefore, have a simple system set up and communicated before processing begins.
Data quality auditor	The data quality auditor will check incoming data for quality, and ensure the sampling is evenly spread	If using a digital system this can be done daily. If using paper based system, it may need to be done in batches as data arrives. (Can be more than one person in this role). Only when doing this from the start and on a regular basis, data quality can be ensured as issues are detected in an early stage (Also see phase six).
Dashboard user	The dashboard user performs specific duties, such as checking incoming data quantity and making sure the data collectors are meeting their quotas.	
Dashboard manager (digital system)	The dashboard manager will oversee the data collection process and ensure that all runs smoothly.	
Data analyst	To perform analysis and necessary visualisations (See phase seven for more details)	It would help for this person to be familiar with the survey content, programme objectives, the sample design being used.

Step three: Prepare survey

Once the questions and order have been designed (see phase four for guidance), it's important to ensure that the survey has been translated into the relevant languages during preparation for data collection, and prior to the training.

Digital systems

Digital or mobile based data collection systems are recommended as they provide a faster, lighter data collection, with the capacity to collect a broader range of data types through one tool.¹

If you are using a digital system, such as a smartphone, ensure that the survey has been digitised and tested on an appropriate device to confirm that the in-built logic of questions work in practice.

A final and vital step in preparing for digital data collection is to ensure that appropriate devices and relevant accessories (power banks, screen protectors, adaptors, sensors and/or test strips) have been procured well in advance of any training or pilot exercises, to minimise confusion and user error. When selecting hardware for smartphone-based data collection, it is important to be clear about the necessary functionality and specifications required for your data collection platform.

When procuring your devices make sure you consider the following:

Hardware:

- Internal memory of the phone and possibility for adding external memory.
- Camera, battery and screen requirements.
- USB On-The-Go connection (OTG) and GPS.

Software:

- Which operating system is the phone running on?
- What software version is required as a minimum for the app to run?

Step four: Training

Nothing can replace thorough training sessions, ensuring all members of the team understand their roles and responsibilities as well as the tools and survey being used. These factors will significantly affect the efficiency of the data collection process and the quality of data that is collected.

In order to get the most out of the training sessions, it is a good idea to consider the following:

Venue and logistics

The venue will need to be appropriately equipped, including (but not limited to) presentation material, air conditioning, internet connection, projector, flip boards, microphones, pens, markers, paper and multiple plug sockets. Also think about size, accessibility and location (within or outside of the city). Will transport need to be organised?

Who will cover the renting costs, costs of materials, costs of meals and costs of participants (both for travelling and participation of data collectors and trainers)? Also consider the administration needed to justify and keep track of expenses, for example keeping a list of participants for receipt of their daily allowance, their equipment, etc.

Training materials

Develop presentation materials, guides and tutorials and prepare all necessary data collection tools for use during the training session (for example, configure smartphones, equip them with the right software and content, print the survey if using paper based forms).

Training methodology and agenda

Training objectives for data collectors are often threefold:

1. Master data collection tools.

1. <http://sustineo.com.au/news/moving-with-the-changing-landscape-of-field-data-collection-benefiting-from-transition-to-table>

2. Master the content of the questions on paper and their translation in the field.
3. Have an understanding of the concept behind the monitoring and evaluation (M&E) system in which the survey is integrated.

This means training always involves teamwork. A clear methodology should be agreed upon and all decisions relating to the agenda should be defined once all parties have been consulted.

To ensure that data collectors have a clear understanding of the logic which underpins the survey, it's advisable to go through each question as a group to ensure consistent understanding and then test the process as a group. Data collectors often have field knowledge and can add valuable contributions.

Lastly, the training should inform the data collectors sufficiently regarding workload, their roles and responsibilities (for example regarding materials used in the field) and the risks of doing fieldwork. This will enable you to filter out participants who do not want or are not able to meet these expectations. It is highly advisable to clearly describe this in the data collector's contract.

Step five: Pilot data collection process

It is best practice and strongly recommended to pilot a survey before commencing the full data collection round. This will help to resolve any practical issues before data

collection has begun, which is far easier than making changes during data collection.

A short pilot serves a number of purposes, including:

1. Testing the survey (and any tools/software) for kinks, faults or user errors.
2. Providing a chance to run the entire 'workflow' in a controlled environment, ensuring everybody knows what they are meant to be doing, when and how.
3. Checking if the data collectors understand the survey questions, and test them in a realistic context.

While the pilot of the survey is occurring, dashboard managers should use this as an opportunity to practice monitoring incoming data for quality, and to cross check that the questions asked in the survey are generating the right information. If there are any questions with strange answers, make sure to clarify with the data collectors. This is best done as a group activity, in order to discuss questions that may need revision and to get feedback from the data collectors on any problems they encountered, either with the survey content or the technology being used.

Step six: Finalise and launch data collection

After incorporating any feedback from the pilot and making final adjustments, the survey can be finalised. You will need to ensure that all devices to be used for data collection have

the correct survey version present, or that the correct survey version has been printed and distributed.

Finally, to avoid confusion, ensure that all pilot data collected during the training sessions has been removed from your database, or at least clearly marked.

Suggested reading

Track & Trace guide on using Akvo Flow mobile data collection tool for tracking and tracing in the agricultural sector (contact Akvo for access).

Mural by Akvo

 <https://tinyurl.com/yb4pvqqw>

Ten misconceptions about mobile data collection

 <https://akvo.org/blog/ten-misconceptions-about-mobile-data-collection/>

Handbook phase six: Collect data

Handbook phase six

Collect data

Authors

Nikki Sloan (Akvo.org),
Camille Clerx (Akvo.org),
Stefan Kraus (Akvo.org)

Contributors

Annabelle Poelert (Akvo.org),
Karolina Sarna (Akvo.org),
Rajashi Mukherjee (Akvo.org),
Uta Wehn
(IHE Delft Institute for Water Education)

In phase four of the Handbook, we address the design of the survey content and sample. Phase five addresses the preparation necessary for data collection in the field, including logistic considerations, role allocation, and training of data collectors. In this phase, we focus on the process and considerations for data collection in the field. As phase five makes clear, it is essential that the process of data collection goes smoothly and the questions and purpose are well understood by the data collectors. The quality of the data is established from the moment the data collectors introduce themselves to the interviewees. This article will go through the five steps of data collection which are summarised below.

Collecting data in five steps

Before

Step one: Conduct final checks before entering the field

During

Step two: Collect Data

Step three: Submit, process and maintain data

Step four: Check data quality as it is collected in the field

After

Step five: Finalise data collection

Step one: Conduct final checks before entering the field

If you've followed the steps covered in phase five of the Handbook, you should be close to sending your data collectors into the field. Here we will touch on a few final considerations to check off beforehand.

Communication

Crosscheck the communication system put in place in phase five.

- How are field issues communicated within the project team, from data collector to supervisor to data analyst?
- Is it clear which surveys are assigned to which data collectors?
- Is the communication flow between data collectors, team leaders and supervisors clear?

Information

Make sure all information is understood and accessible to everyone involved in data collection.

- Consider creating an FAQ document for the data collectors if you haven't already. Having a simple two pager summarising important and relevant information for the survey and data collection process, and relevant project team contact details, can be extremely useful. This two pager can also include survey protocols, definition

lists, sampling guides, a day-to-day timetable of activities, and a checklist for equipment.

- Review the scheduled list of activities.
- Make sure that all data collectors understand their role and responsibilities, as well as the risks related to field work, and agree to them.
- Consistency in the data collected and in sampling is essential for good quality data. In the field, data collectors need to be able to know what to do in case of an issue with sampling, for example if the household head is not at home or if a water point is inaccessible, and who to notify. As mentioned in phase five, data collectors will need to have a common understanding of the questions in their survey so that the data collection can be consistent across the team.

Logistics

Have transportation, agreements on daily service allowance, and accommodation been finalised? Make sure data collectors and supervisors have:

- Clear ID, document or badge showing the organisation and its mission, preferably with local government endorsement.
- The survey available on their smartphone, or a generous number of surveys printed and stored in an envelope.
- A letter from the relevant authorities permitting them to administer the survey.

- The FAQ two pager outlined in step 1.2 above.
- Appropriate mobile phone credits and/or data packages.
- A list of phone numbers to call in case of an emergency or when support is needed.
- Information to give the interviewee upon request, or any other party concerned by the data collection, and to ensure informed consent has been attained.
- A clear understanding of the survey and sample design, and the purpose of both. Please see phase five for further details.

Step two: Collecting data

There are many factors that determine the appropriate methodology for data collection. Identifying how, where, when and by whom will be influenced by the unique context of the project. There is a range of geographic, meteorological, cultural, political and linguistic factors that may influence how the data is collected and these, alongside the questions above, should guide any data collection methodology.

Infrastructure

Ensure that global positioning system (GPS) points are taken when mapping out the site for the first time. Try to get as close to the water point as possible. When taking a photo of the water point, try to capture it in such a way that another person would be able to find it without you in future, for

example by including the pump in context, or with a familiar landmark in the background. Ask the following questions:

- How is data on the water point currently captured and recorded? Who should you speak to about the water point?
- What is the most appropriate method for capturing information about water inputs and outputs? Is it through observation and testing the water itself, or through talking with local users about their experience of extracting and using the water?
- Is there a community or other group responsible for the upkeep of the water point?

Interviews

General rules for data collectors conducting interviews:

- Wear culturally appropriate clothing. I.e. some places it may be inappropriate for women to show their legs, or wear trousers.
- Be respectful of an interviewee's time and answers, as well as possible restrictions on their answers.
- Prompt without leading the responses. Try rephrasing questions to get the information needed.
- Record the answers that the interviewee gives, not the answers that the data collector feels they should have given.

Interviewing the right person

Who will you need to interview to gather all the required data? For instance, your survey has a question about how long it takes to collect water for the household. You are interviewing the household head, however they are not responsible for the collection of water. Instead, you will need to find and ask the person who does collect the water, if it is culturally appropriate to do so.

Restrictions on the interviewer

When collecting data, there may be a number of cultural limitations on when and who the data collector will be able to interview. Gender can come up as a potential restriction. For instance, it may be inappropriate for a male data collector to interview a female interviewee, and vice versa. A potential practical solution to this is to have mixed genders in the data collection team.

Securing the interview location

Be conscious of the location and length of the interview. For instance, if you are recruiting in a market place, but need to ask some personal questions, it might be wise to find a secluded location to conduct the survey. This will again be determined by cultural norms.

Timing

Choose the timing of your survey that would be most convenient for the respondent. Be sensitive to the cultural and practical elements that may affect this. For instance,

there may be times where the interviewees are attending religious ceremonies. If you want to interview farmers, they are unlikely to have time to fill in a long survey during harvest. If a participant is willing to participate in the survey, but unavailable when the data collector arrives, try to find out when they will be available. What are potential times or days that your interviewees may be unavailable? What other factors could affect their availability or participation in the data collection?

Step three: Submit, process and maintain data

It is important to set definitive timelines for the collection, uploading and checking of data. A common setup is for each data collector in a team to hand in their completed surveys to the team leader, who checks that they are complete at the end of the day before approving them. This also gives the team leader an opportunity to check the sampling progress and discuss any issues with the data collector during collection.

Database entry procedures

For the entry of data from the paper based surveys, it is good to make sure that those entering the data have a clear format to follow. This commonly takes the format of a spreadsheet with variables (questions) as column headings, and each survey respondent as a separate row. This is a stage where many data errors can occur, so it is essential to have a way of linking a specific paper based survey respondent with the spreadsheet entry. This can take the

place of an identifier (i.e. gender, country, age could look like F-NL-30).

In the case of mobile data collection, setting a time limit for the data collectors to upload data will be determined by a number of factors.

- Is this a monitoring survey? If yes, they may need to sync regularly throughout the day to ensure the data available in the device is up to date.
- Will the data collectors have access to Internet via mobile data packs, or only WiFi? If using mobile data, then uploading will be more or less continuous. If not, then uploading periodically at hotspots will be necessary.
- When should they do the data syncing and uploading? Making sure that the timing of uploading the data is clear and consistent that will make it easier to coordinate data quality checks.

Data management protocol

Be mindful of data privacy and security considerations when collecting data. Consider the following questions:

- How will you secure the completed surveys?
- Will the data collectors carry the surveys with them, or hand them into the supervisor?
- How does the data get from the field to the processing location?

How will the original data be stored (i.e. paper storage, digitally or both)?

For paper based surveys, you'll need to consider how to store the paper so that it is secure and won't be affected by environmental factors, such as bugs or damp. It may be preferable to scan the paper version to store digitally.

When storing data digitally you can consider whether to use Cloud-based systems, as copies on computers or in external hard drives. All of these are options, but ensuring their security and encryption is also essential. If multiple agencies are involved in the data collection project, you'll need to decide who will host the original data, and which other organisations will need copies.

Backing up data

It is important to ensure that the original data is still available in its unedited form as a backup. This can be because someone may need to check the data and code questions differently. Accident and errors do also occur, and ensuring that there are regular back-ups that are clearly named and stored means that you'll always be able to go back. It's generally recommended to back up at least weekly during the initial data entry and cleaning phase, and reduce the frequency later on.

Track changes

Setting up a process to record any errors or changes made to the data allows for the process to be transparent. This is commonly referred to as a Log Book. This can also include

creating field reports on difficulties encountered during data collection, for example concerning the survey, data entry or field experiences. During cleaning it's also vital to ensure there is a consistent system for the naming and controlling of different versions.

Step four: Check data quality as it is collected in the field

Ensuring the quality of your data is an essential part of the data collection process, and includes checking that everything is running smoothly logistically, tracking the collected data, and checking field work activities. Checking data quality should occur simultaneously with your data collection in the field.

When should a field supervisor intervene?

The field supervisor should ideally check the surveys at the end of each day and initiate course correction immediately. Once a team leaves a survey location it may be difficult to resurvey. Supervisors should review the sample/surveys answers where possible to ensure that all answers have been filled in. Regularly checking in with the data collectors themselves to ensure that they are keeping to quotas is crucial. If issues occur either with the survey answers, or the data collector's performance, the supervisor should look for why this might have occurred and provide guidance accordingly. The table below shows a list of important things to consider to ensure the quality of the data.

Regular data checks	Data/dashboard managers need to check the incoming data at regular and defined intervals (daily, weekly, etc.).
Data quality assessment skills	A skilled data quality auditor is required to isolate inconsistencies and ensure that the sampling breakdown is correct.
Data tracking system	A data tracking system is used to check what comes in per data collector. For digital data collection, the dashboard manager can easily track the last time a data collector was connected to the dashboard. By setting up these systems during data collection, you will be able to track sample size and distribution and infer data quality during the data collection process.
Data triangulation and quality checks	<p>Data triangulation is a powerful technique that uses a combination of questions and question types to ensure data quality and validity. For example, by having a series of questions with region, district and village, coupled with a GPS location question, you can ensure validity and minimise the risk of data entry error.</p> <p>Consider using a tool such as Akvo Lumen, Carto, or another mapping tool to check the points are in the correct locations.</p> <p>You can also use tools such as Microsoft Excel to filter and sort answers to find outliers and inconsistencies in the data. Checking data quality during data collection is a skill set on its own, to find out more check this FAO online article. This will also be discussed in more depth in phase seven of the Handbook.</p>
Feedback and cross checking process	<p>Set up a process for feedback and the cross checking of data collected with the data collector on any errors so that they can notify the data managers ahead of time.</p> <p>Important: avoid changing the survey during data collection. It can be tempting to make small adjustments, especially with mobile data collection. This will endanger consistency.</p>
Documentation	<p>For reporting purposes, keep track of errors found with data entry, difficulties encountered during field implementation, and steps taken to resolve them.</p> <p>Additionally, it is best practice to maintain a codebook to keep track of the editing and/or coding of data. This is also touched upon in phase seven of the Handbook.</p>

Step five: Finalising data collection

In order to finalise the fieldwork, you should perform some final checks on the activities and create reports to document field operations.

- Check that all data collectors have handed in or uploaded their survey forms and that quotas been met.
- Check the sampling and numbers of data points collected per locality. Has anything been missed and will there need to be further data collection?
- For paper-based data collection, archive all survey forms consistently and digitise them. For digital data collection, download raw data reports and store them in your system.
- Create reports on difficulties encountered during data collection, initial results of the data collection, considerations regarding data quality, and the actual sampling in the field against the original sampling method.

Conclusion

The better you have prepared your survey, planned the data collection process, trained the data collectors, and resolved impurities in the data on the go, the easier data cleaning will be and the more valuable your analysis and visualisations will be. Without the appropriate data, you will not be able to calculate the indicators you need, endangering the whole purpose of the data collection project. Not checking data

quality on the go will impede getting results out fast and can lead to a long, cumbersome process of data cleaning.

The next phase of the Handbook will focus on how to clean, analyse and creating meaningful visuals from your data.

Suggested reading

A useful guide which is easily accessible is the UNDP Pulse Lab (Enumerator Guide)

07

Handbook phase seven: Analyse and visualise data

Handbook phase seven

Analyse and visualise data

Authors

Annabelle Poelert (Akvo.org),
Karolina Sarna (Akvo.org)

Contributors

Bettina Genthe (Council for Scientific
and Industrial Research, CSIR),
Camille Clerx (Akvo.org),
Lars Heemskerk (Akvo.org),
Marten Schoonman (Akvo.org),
Nikki Sloan (Akvo.org),
Stefan Kraus (Akvo.org)

Every data collection project should include data analysis in order to extract learning. In essence, data analysis aims to inspect, clean, transform and visualise data with the goal of reaching certain conclusions about a problem at hand. Through data analysis, you want to describe and summarise the data, identify relationships, and find similarities and differences between the variables. You want to create insights that you can use and share with others in a visual and accessible format. This will allow you to make better decisions based on an understanding of the data you have.

Preparation

Step one: What do you want to know?

Before you start solving any problem, you should know and understand what the problem is. Get involved with the problem, not the solution, they say. As trivial as this may sound, asking the right question(s) is crucial to creating good insights from data. Based on the previous phases, specifically the design of your project, this step should be straightforward and indicate clear steps to be taken in the data analysis process.

Analysing and visualising data in six steps

We can divide data analysis into two phases: preparation and execution. Each of the phases consists of three steps.

Preparation

Step one: What do you want to know?

Step two: Which data sources are you planning to use/combine?

Step three: Is your data clean and ready to be used?

Execution

Step four: Understand your data.

Step five: Explore your data and finding relationships.

Step six: What story do you need to communicate? To whom? Why?

Step two: Which data sources are you planning to use/combine?

You now know what the problem is. But to arrive at a solution, you need data. After a successful data collection phase that you went through you might want to consider supplementing the information you gathered with some external sources. There is a wealth of open data resources available (you can read more about different ways of sharing data in phase eight), but linking different datasets in the right format can be complicated and requires careful consideration. It needs to be the right data. Say you want to connect household data on water and sanitation facilities with data about public spending on water, sanitation and hygiene (WASH). Both datasets are collected at different levels: household level and district, region and country. Can you connect the two, and if so, how do you do it? What are the challenges? You can find more information about the data collection process in phases five and six.

It is also important to consider the type of data you need for your analysis. Sometimes, data may be available, but in a format that makes it hard to perform the analysis you want to. You can ask yourself if you can transform the data into a different format or change the type of analysis to match the data format.

Step three: Is your data clean and ready to be used?

One would think that once you get all the information you need you can just jump right in and start doing cool things,

like visualisations and regressions. The truth is that no matter where you get your data from you will always need to clean it. Data cleaning very often takes a big chunk of every data analysis project. When you clean data, you try to understand every single column in your dataset, identify the missing or corrupt values, the outliers and make sure the data is consistent, for example, different spellings of the same thing.

Execution

Step four: Understand your data

Now that your data is clean it's time to look at it and understand what is in that dataset and the characteristics of the data. These characteristics can include size or amount of data, completeness of the data, and correctness of the data. When you first look at the clean dataset you want to understand all the different variables you have. It is easier to understand the information in your data when you visualise it and summarise it in tables.

Step five: Explore your data and find the relationships

You can spend a lot of time just looking at the data, but what you want is to quickly find the relevant patterns that will lead you to the answer you want to find. Most of the time when you have to report on the project you will look at different indicators that were chosen during the design phase and report on the amounts they represent. That's definitely needed, but data analysis should go beyond

counting. When you explore data you should try to compare and contrast different variables. This way, you can identify relationships that might not have been obvious. This process of checking different characteristics of datasets will get easier and more intuitive the more you do it, so just give it a try and have fun with it.

In some projects, you want to go even further and consider the wealth of information that can be obtained by applying more complicated data science techniques. You can look for hidden patterns, find explanations and try to predict future outcomes of the specific indicators.

Step six: What story do you need to communicate? To whom? Why?

In order to make the whole data analysis process worthwhile, you need to communicate the insights you gained in a compelling and digestible way. There is no point in producing information that no one will understand. It is also a waste of resources put into data collection and analysis. You need to create a story that combines the information you obtained with the domain knowledge of your organisation to answer the question you asked. Communicating the results of data analysis is the crucial point. However, it's also very important how you communicate. If the message is not clear you risk no action being taken. You have to make sure that the message you are communicating is clear and targeted at a specific audience. If you make complicated graphics that require a lot of explanation, you risk your audience not giving it enough attention and your message getting lost.

Understand the message you are trying to communicate

When it comes to data visualisation, you are trying to show one of the four things with your data: relationships between data points, a comparison of data points, a composition of data, or a distribution of data.

When you want to show relation, you are focusing on showing connection or correlation between two or more variables. For example, you may want to show how the functionality of a waterpoint relates to its age.

When you want to compare data points, you try to show what distinguishes one set of variables from another. You want to display how the two variables interact, such as the number of water points in the five biggest districts in the country.

When you want to show the composition of data, you collect different types of information that make up a whole and display them together. For example, you may want to show the different founders of water points in the whole country.

When you want to show the distribution, you try to lay out a collection of related or unrelated information simply to see how it correlates. You may want to display the number of reported water point malfunctions over a long time period and see if you can find seasonal patterns in the functionality.

Conclusion

After you have collected the data, it's crucial to clean the data from outliers and unreliable data points to arrive at useful insights. In order to transform the data into insights, the person analysing the data needs to take time to understand the dataset, which will help to uncover patterns in, and relationships between, the data. However, remember that to become really effective in using data for decision making, the last step of communicating the data effectively to your target audience is as important as the data analysis itself. You can read more about that in the next phase: share data and communicate insights.

Suggested reading

A handy guide to the basic data analysis can be found here:

 <https://tinyurl.com/y8q29czg>

A handy help with picking the right type of chart can be found here:

 http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html

You can find a comprehensive guide on how to document your data here:

 <http://www.medicine.mcgill.ca/epidemiology/joseph/pbelisle/CodebookCookbook.html>

You can find out more about data science and its potential in the international development in this blog:

 <https://blogs.worldbank.org/opendata/will-your-project-fail-without-data-scientist>

Here is a link to the data cleaning checklist that will guide you through the process of cleaning:

 <https://www.thedatabank.com/wp-content/uploads/2010/02/Data-Cleanup-Checklist.pdf>

08

Handbook phase eight:
Share data and communicate insights

Handbook phase eight

Share data and communicate insights

Authors

Lars Heemskerk (Akvo.org),

Marten Schoonman (Akvo.org)

Contributors

Arun Kumar Pratihast (Akvo.org),

Annabelle Poelert (Akvo.org),

Beatriz Medina (Water Environment
and Business for Development, WE&B)

After analysing and visualising your data to make it useful (phase seven), you can share it with others. By sharing data, you allow others to access and use it. Sharing data could lead to the following mutual benefits:

- Enable new data uses.
- Lead to new collaborations between various stakeholders.
- Improve transparency and accountability.
- Empower communities to act based on the data.
- Reduce the cost of new data collection projects.
- Increase the impact and visibility of the work.
- Provide resources for education.
- Encourage enquiry and debate.

Preparing the data sharing process

To ensure your communication is effective and that you reach the right target group, answer the five **W** questions.

Why: What are your motives for sharing the data? What do you hope to achieve with your content? Translate these motives into concrete goals and write them down.

Who: Who is your audience? Your answer will determine what the most effective approach is to sharing your materials. Without it, the sharing effort may be fruitless. Are the end users included in your sharing effort? They are best placed to enrich and validate the findings, and they are probably the audience that can benefit the most.

What: Deciding which data to share and how can be challenging. Do you share aggregated data and conclusions, or raw data? And are there risks you need to take into account regarding confidentiality or potential misuse which require anonymization?

When: When did the data collection take place and when is the target group most likely to engage?

Where: When you know why, with whom, and when you are going to share the content, you can decide where the content should be shared.

But before you start, do you know who will be interested and which format and channel will work for them? Are any other steps required to ensure that the data is useful? In the design phase of your data collection project, phase two of the Handbook, many of these questions will already have been addressed and pave the way for effective data sharing. Making use of the prep work you've already done, you should invest time in preparing the data sharing process to ensure that your data is communicated in a productive way. It can be frustrating if you create a great report, well formatted, and share it on the third or fourth menu level of a website, where it isn't found or read by anyone.

Target audience

A good start can be to question who you'll be sharing the data or main indicators with, and to identify these groups according to the type of organisation or community they belong to, their position within the organisation or community, their main purpose for using the data, and their preferred channels.

Where the target audience is located may be an important aspect in relation to which channels should be used, depending on which forms of communication are available.

Channels and formats

To reach an audience, you can use multiple channels and formats. Your choices determine the success rate of the sharing. Examples of channels are:

- In person
- Mobile phone
- Internet
- Social media
- Broadcasting: radio and television

Of course, these can be specified further. The Internet channel has a wide diversity of (social) media platforms, and also includes websites and email. So the question is, which channels do your target audience use? If you pick the wrong channel, sharing is unlikely to be successful.

There is a large diversity of formats, including:

- **In person:** At conferences, networking events, roadshows, workshops, focus groups or webinars.
- **In writing and online:** Websites, newsletters, contact databases, articles, presentations, policy briefs, factsheets, brochures, posters, E-learning platforms, photos or videos.

When looking at the specifics of the data sharing, there are four key types:

- **Story:** For data to be meaningful to a general audience, it is important to find meaning in the numbers. Without a storyline, the output may be limited to just a description of numbers.
- **Tables:** Using tables effectively helps minimise the number of data values in your text. It also eliminates the need to discuss less significant variables that are not essential to the storyline.
- **Charts:** A chart is a visual representation of statistical data, in which the data are represented by symbols such as bars or lines. It is a very effective visual tool, as it displays data quickly and easily, facilitates comparison, and can reveal trends and relationships within the data.
- **Maps:** Maps are the most effective tools to visualise spatial patterns. When carefully designed and presented, they are more than just decorative features in a statistical presentation. They can help people identify and highlight distributions and patterns that might not be apparent from tables and charts.

Data for the development sector is very diverse in nature. It can be, for example, infrastructure related, services related or policy related. Each type of data requires its own format in order for it to be optimally shared with the chosen audience using the preferred channel. Often, a combination of the above formats works best. For example, to compare a baseline water point inventory with a second inventory of water points you may want to use a table. You can use text to explain the differences, and illustrate key elements which are location specific using maps. You can then summarise and conclude with text.

An important consideration is language. What is the literacy level of the target audience and which languages do they speak? Which solutions can you develop to ensure everyone understands? In Kenya, for example, does the target audience understand English, Kiswahili, or would a vernacular language be more useful?

Measuring success

In many cases of online sharing you can measure success. For example, you can use website analytics to record the number of visitors, visits, downloads and shares. This information may be just an indication. For more depth, you can ask for feedback using evaluation forms to find out how people have made use of the data and whether there are any expected or unexpected results. A thorough evaluation can ensure success in future data sharing efforts.

Feedback loop

Sharing data back to the very people and communities involved is often overlooked. There are various reasons to build this into the project plan, including the validation of the data collection and analysis, and to involve them in decision making regarding potential changes to their environment and lives. There are simple ways to do this, for example by printing the data and results on a poster and feeding this into a discussion on location, as illustrated in this example from India: <https://akvo.org/blog/partnering-with-fxb-india-suraksha-for-a-water-security-program-in-india/>. It is likely that new insights will be gained by examining the materials together with the people who know the context best.

Processed versus raw data

When you share with others, you would probably rather share processed data; the results of data analysis. The analysis may be quite straightforward or it may be complex, consisting of advanced statistical or geospatial formula. Certain choices are made in the process of data analysis, and the results (insights, interpretation, conclusions) may not fit the needs of the audience. Taking potentially sensitive data into account, it may be more useful to share the raw data as well. This allows the consumer of the data to either verify the analysis or perform a different type of analysis. Another option is that the shared data is used as a data source for other projects, which is typically the case in rural waterpoint mapping and monitoring.

Open, shared and closed data

Sharing data does not necessarily mean that you should put everything on the Internet just like that. As posted above, the key question is what purpose it serves. This video, made by the Open Data Institute, differentiates between open shared and closed data <https://vimeo.com/125783029>.

Open data is “data that anyone can access, use and share.” This usually refers to raw data. For data to be considered open, it must be accessible (which usually means published on the Internet), available in a machine-readable format, and licensed so that anyone can access, use and share it, commercially and otherwise.

Closed data is “data that can only be accessed by its subject, owner or holder.”

Shared data can be of three types:

Named access: “Data that is shared only with named people or organisations.”

Attribute-based access: “Data that available to specific groups who meet certain criteria.”

Public access: “Data that is available to anyone under terms and conditions that are not ‘open’.”

See the Open Data Institute article “Closed, shared, open data: what’s in a name?” for more information on the above.

<https://theodi.org/topic/data-infrastructure/>

Example of open data sharing

SMARTerWASH in Ghana

SMARTerWASH is a project of the Community Water and Sanitation Agency (CWSA) in Ghana, which sets out to monitor water and sanitation in six regions of Ghana. The SMARTerWASH project aims to strengthen the information and communication technology (ICT) infrastructure and by ensuring interoperability between systems.

Following data collection using Akvo Flow in 131 (of a national total of 216) districts, data from 23,001 handpumps, 938 piped schemes, almost 15,000 Water and Sanitation Management teams (VWSMTs) and 131 service authorities were collected. This data was processed and made available in offline factsheets (regional and district level) and in an online water atlas using an Akvo Site template <https://cwsawateratlas.org/>. Everyone who is interested water, sanitation and hygiene (WASH) data in Ghana can access this information in different formats.

Two examples of open data standards

Water point data exchange (WPDx)

The global rural waterpoint data repository Water Point Data Exchange (WPDx) is an example of a combined raw, open and shared data source <https://www.waterpointdata.org/>. Basic details are shared such as location, date and type of infrastructure. Those who want more data related

to these records can request them directly from the relevant organisation, enabling a sharing-on-request system.

International Aid Transparency Initiative (IATI)

IATI is a worldwide standard for structuring and sharing project and programme data. At the centre is the IATI Standard, a format and framework for publishing data on development cooperation activities, intended to be used by all organisations in development, including government donors, private sector organisations, and national and international NGOs. The IATI registry acts as an online catalogue and index of links to all of the raw data published to the IATI Standard.

To share openly or not to share openly?

It may be complicated to judge whether data should or can be shared openly. As highlighted above, you may also consider sharing data on request. A general rule you can apply is to “share, unless...”. This refers to the fact that there are more (potential) benefits of sharing data than in keeping it to yourself or your organisation. “Unless” refers to reasons for not sharing, which may relate to private, confidential or high risk data. The latter may relate to, for example, data regarding a post-war situation. Private data must always remain closed unless explicit approval is given by the persons or communities involved. There are various ways to anonymise data, which you can about in this quick overview <http://www.ethicsguidebook.ac.uk/Anonymising-your-data-309> or detailed guide <https://ico.org.uk/media/>

[for-organisations/documents/1061/anonymisation-code.pdf](https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf). Some key terms in this realm are data ethics and responsible data, which we have defined in the Handbook glossary. The message here is that you should check whether sharing data or insights poses a risk and to take measures accordingly. A good practice is for an organisation to develop its own data policy. An example of this is the five-point responsible programme data policy by Oxfam https://www.oxfam.org/sites/www.oxfam.org/files/file_attachments/story/oxfam-responsible-program-data-policy-feb-2015-en.pdf.

Data rights

When sharing data openly, you need to consider what the conditions and requirements are for the user of the data; in other words, the licence which applies to the data. Some examples include Creative Commons licenses, Open Government licenses and bespoke or custom-made licenses. To support you in making a choice which license to use for sharing data openly, you can consult Choose a License <https://choosealicense.com/>.

Challenges of data sharing

The growing availability of shared data ensures operational and collaborative opportunities. However, there are some risks and challenges around data sharing:

- Publishing certain data that may violate legislation:
Sharing certain data may be prohibited by law or infringe upon someone's rights or freedoms.
- Publishing data that can be interpreted in different ways:
Users might intentionally or unintentionally misinterpret the data (to cause scandal, to get a competitive advantage, to cause harm to other subjects, etc.). Some data could be misused or wrongly interpreted.
- Finding the right audience: There might not be an audience for your data because it's not possible to locate the dataset or because nobody knows it's available.

Conclusion

Thinking and developing your sharing strategy can ensure the other steps in the process bear fruit. At all times, it is important to be mindful of the target audience you are sharing the data and insights with. Does the format and channel you choose fit their needs? Did you choose the right licence for your data and consider data ethics? And is it possible to gather information about the usage of the shared data and insights?

Suggested reading

Closed, shared, open data: what's in a name?

 <https://theodi.org/topic/data-infrastructure/>

Share your failure

 <https://www.ictworks.org/wellthat-was-a-failure/#.W7sW7H7Lgvp>

RRi (Responsible Research Innovation)

 <https://www.rri-tools.eu/open-access>

"Implementing Responsible Research and Innovation in research funding and research conducting organisations – what have we learned so far?"

 <https://www.rri-practice.eu/wp-content/uploads/2017/09/RRi-Practice-paper-Rome-sept-2017.pdf>

09

Handbook phase nine: *Make informed decisions*

Handbook phase nine

Make informed decisions

Authors

Arun Kumar Pratihast (Akvo.org),

Marten Schoonman (Akvo.org)

Contributors

Harro Riedstra (Akvo.org),

Lars Heemskerk (Akvo.org),

Mamohloding Tlhagale

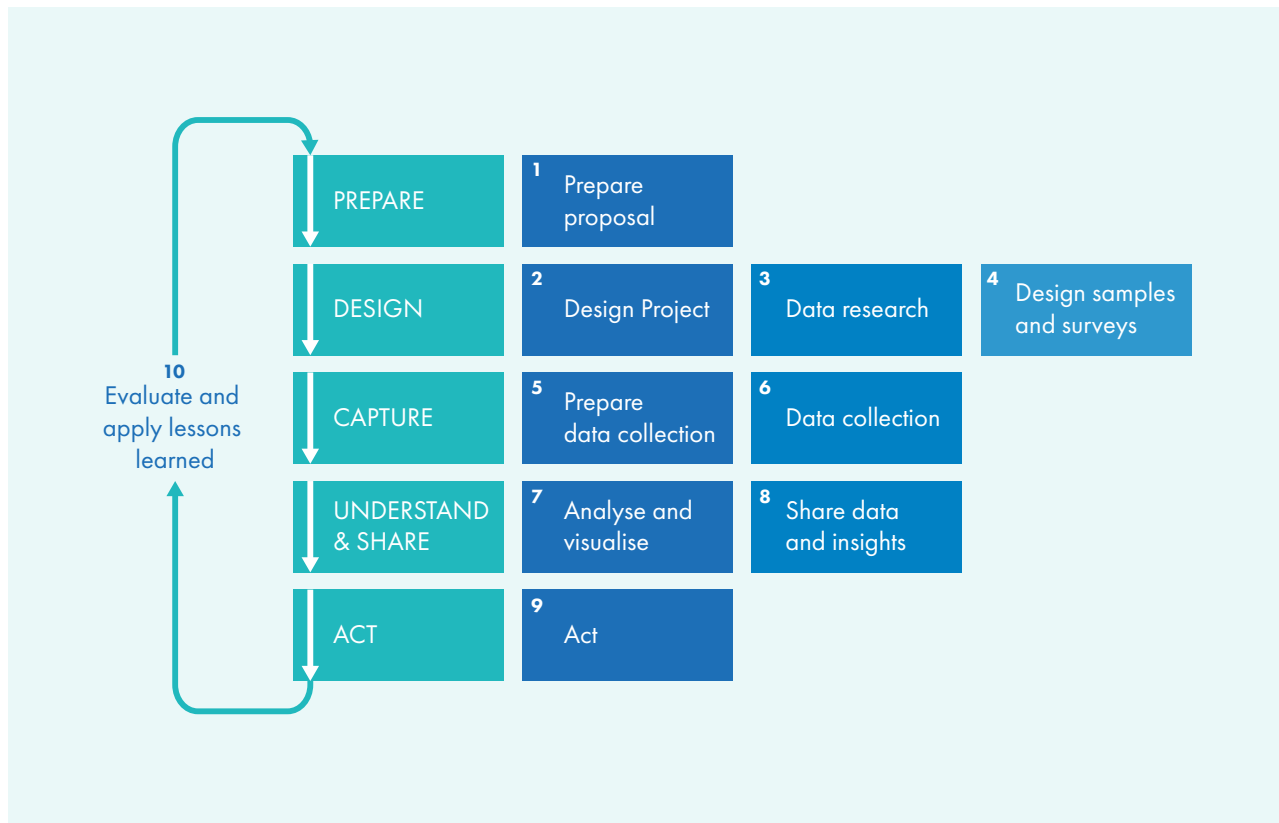
(Water Research Commission, WRC)

Growing opportunities in data collection and analysis have led to changes in the way people make decisions. Data to decision (D2D) refers to the practice of basing decisions on the analysis of data rather than on preconceptions, opinions and other influencing factors. Using data as the foundation for decision making allows for transparency and accountability towards the people that are affected by the decisions.

To improve the decision-making process, several types of D2D models have been developed over the years, such as fact-based decision making, rational decision making and ethical decision making. Fact-based decision making is one of the most commonly used to make informed decisions, because it places emphasis on the gathering of facts, figures, data and evidence. Practice of this model allows decision makers to monitor the outcomes of policies and the national progress towards the sustainable development goals.

Each phase of the Handbook is designed to prepare those involved in the project for data-informed decision making. These phases are outlined below:

1	Prepare project	Prepare the project proposal.
2	Design project	Design the project in detail, together with (local) stakeholders.
3	Research data	Research what data is already available and what this means for the data to be collected.
4	Design samples and surveys	Design the sample in which data will be collected and the survey to be used.
5	Prepare data collection	Prepare for data collection in the field.
6	Collect data	Manage the data collection in the field.
7	Analyse and visualise data and gain insights	Analyse and visualise the gathered data, combine with other data sources and gain insights.
8	Share data and insights	Share the gathered data and insights publicly.
9	Make informed decisions	Make informed decisions based on the insights developed.
10	Evaluate and apply lessons learned	Evaluate the project and apply any lessons learned in the follow up.



Above: Flow diagram of different phases of the data to decision making process¹.

1. A. Acharjee, "Data, big data and data driven decision making strategy: Part 1," 2008.

Informed decision making is only possible when the appropriate data or evidence is available. Thoroughly researching the data needs to be done in phases two and three, followed by the collection of data in phases four to six. The data processing in phase seven is done with the target audience in mind so that the data and the conclusions drawn are relevant for the decision makers. Data and insights are then shared in phase eight so that any relevant parties can make use of it.

The opportunity D2D presents

The success of the D2D process is influenced by three main factors. First, the types of decisions that can be made (policy changes, organisational restructuring, funding decisions) based on the facts, the supported statistics, and the tested intuition. Secondly, the communication of decisions to stakeholders, and lastly, the execution of the decisions. This process provides higher levels of transparency and better awareness about decision making. By leveraging the use of data across the organisation, decision makers can enable effective risk analysis, enhancing organisational resilience and risk management. Of course, the data must be valid and the organisation must have the human resources with the skills and capabilities to analyse the data in a transparent way.

Best practices

Below are some best practice examples in health and water sector.

Health sector

The overall aim of World Health Organization (WHO) is to articulate specific interventions that can improve the demand for and use of data in decision making, so that improvements in health systems can be realised. There are several examples in which WHO have used D2D to more effectively target limited surveillance, prevention, and control resources and to reduce vector-borne and other infectious diseases such as mosquito-borne dengue and malaria.²

Water sector

Higher temperatures and changes in extreme weather conditions are predicted to affect availability and distribution of rainfall, river flows and groundwater, and further deteriorate water quality. Adaptation to climate change is closely linked to water and its role in sustainable development. During the droughts of 2016 and now in 2017, Akvo has worked together closely with the Ethiopian government and other key WASH partners such as UNICEF to map water infrastructure in the most affected drought-prone areas in order to build up a national database of water point functionality and to monitor these water points

at regular intervals. This information has been used to address food security and humanitarian response to minimise the risks and vulnerabilities generated by serious drought.³ Another example is from Tanzania where the government is implementing D2D to increase access to water from 53% in 2005 to 90% by 2025.⁴

Challenges of D2D

Lack of organisational capacity to be data driven

The statistics above demonstrate a growing appetite among organisations for data and data-driven decision making, despite their struggles with the several issues. The first being that many organisations and governments working in the development sector do not have the resources and capacity to design data collection schemes, or to clean, analyse and use data. Furthermore, data is often collected using inconsistent methodologies and located in fragmented sources, which do not integrate well with existing government systems. This means that data and insights are not stored in a structured way, which makes it hard to access for future use and hampers building an institutional memory.

Unwillingness to use facts

The second is the willingness to accept data-driven decision making. Especially in political processes, conclusions drawn from facts may contradict the interests of certain groups, parties or individuals. This influence may be stronger than anticipated and the data or insights may be manipulated to fit specific interests. It is therefore important to be aware of this risk, to detail the risk and develop strategies to reduce the risk. A method that can be used is stakeholder mapping, which is detailed in phase two of the Handbook.

Data quality challenges

There is a range of issues regarding data quality which can stop it from being used. It may be incomplete, and therefore unfit for analysis. Without analysis, gathering insights for decision making isn't possible. As described in phase three (data research), there are four key factors to consider when evaluating data: accessibility, granularity, credibility and relevance. Being aware of these factors and the associated challenges allows you to plan for high quality data collection in the design phase of the project.

2. L. M. Barat, "Four malaria success stories: how malaria burden was successfully reduced in Brazil, Eritrea, India, and Vietnam," Am. J. Trop. Med. Hyg., vol. 74, no. 1, pp. 12-16, 2006.

3. <http://akvo.org/blog/government-of-ethiopia-and-unicef-compile-national-water-sanitation-inventory/>

4. R. Giné and A. Pérez-Foguet, 'Sustainability assessment of national rural water supply program in Tanzania', Nat. Resour. Forum, vol. 32, no. 4, pp. 327-342, 2008.

Involving communities in decision making

There are multiple advantages and disadvantages of community participation in a decision-making process. According to research conducted by R. A. Irvin and J. Stansbury, the advantages include interactive learning from citizen and government; building trust between citizens and project leaders; gaining legitimacy of decisions made; and gaining help with the implementation of decisions. Disadvantages include cost; possibility of process backfiring and creating hostility; opposing interest groups; and government feeling that they have lost control over the decision making.⁵

Conclusion

Data-driven decision making is necessary for good governance. It allows organisations to take prompt action, which may avoid or reduce financial risk and enhance transparency in the use of information. In this article, we describe the importance and challenges of the D2D process. We also outline four phases of the D2D process: establishing the organisational need; using the right tool for data collection and pre-processing the data; data analysis, interpreting and deriving the insights from the data; and making a final judgment. The case study results from the health and water sector confirm that D2D has been successful in improving services.

Suggested reading

Informed Decision-making: You Should Be Getting More Value From Your Data

 <https://pages.sisense.com/rs/601-OXE-081/images/Aberdeen-Report-Impact-of-Data-Informed-Decisions.pdf>

Using Data to Influence Government Decisions: Opportunities and Challenges for User-Centered Design to Improve Governance Data Impact

 <https://tinyurl.com/yb9ut2se>

5. R. A. Irvin and J. Stansbury, 'Citizen Participation in Decision Making: Is It Worth the Effort?', *Public Adm. Rev.*, vol. 64, no. 1, pp. 55-65, 2004.

Handbook phase ten: Evaluate and apply lessons learned

Handbook phase ten

Evaluate and apply lessons learned

Author

Marten Schoonman (Akvo.org)

Contributors

Anita van der Laan (Akvo.org),

Annabelle Poelert (Akvo.org),

Arun Kumar Pratihast (Akvo.org),

Harro Riedstra (Akvo.org),

Natacha Amorsi

(Office International de L'Eau, OIEau)

For any project implementation, it is important to consistently check progress during the implementation as well as evaluate where things can be improved to achieve more of the expected results, more inclusively, more sustainably, and more efficiently. But how to go about such an evaluation and how to draw lessons learned?

Practical suggestions on evaluation and applying lessons learned

1. Before you start, determine what the purpose of the evaluation is, who the audience is, what kind of information will be needed, and what the added value will be.
2. Make use of the online resources available, detailed in this article, to guide your evaluation and make it as useful as possible.
3. Converting the findings of your evaluation into lessons learned and applying them to current or future projects is where the value is created. This article lists clear steps for you to follow.

What do we mean by evaluation?

In the management of programmes, we use the terms monitoring and evaluation (M&E) and planning, monitoring, evaluation and learning (PMEL) to describe the methodology used to improve the effectiveness and sustainability of development projects. An evaluation aims to capture what the real effects were and what could be improved next time. It is a way of identifying lessons that can be learned and used in similar, up-scaled or next phases of a project or programme.

Planning and using evaluations

When planning for evaluations there are a number of things to consider:

- Do you do the evaluation yourself or do you involve an external party to evaluate? The risk of bias is overcome by involving an external party. Most funders require the end evaluation of the project to be conducted by an external party.
- When should you evaluate? If you perform an evaluation before the end of a project, e.g. mid-term, it allows for not only drawing conclusions and lessons but also to immediately implement them in the project before it ends. This is referred to as a “formative evaluation” as opposed to a “summative evaluation” at the end of the project or programme.

- Will you share the results of the evaluation? And with whom exactly? It is usually the funder who requests the evaluation, but you may decide to share evaluation reports more widely or even publicly. This can be a valuable resource for other programmes, while promoting transparency and accountability and instilling trust within the (funding) partners.
- There is a wide range of evaluation approaches, from appreciative inquiry to social return on investment. You can find a comprehensive list including descriptions of each at BetterEvaluation.org: <https://www.betterevaluation.org/en/approaches>.

Lessons learned

An (external) evaluation usually has a number of evaluation questions. As described on BetterEvaluation.org, these usually cover three main areas of a project:

- Appropriateness/relevance
- Effectiveness
- Efficiency

If the evaluation is aimed at finding what works, what doesn't, and why, then the evaluation findings will result in lessons learned: "what would we do differently next time? And what would we do the same?"

Key tips for formulating lessons learned:

- Lessons learned should consist of a generalised principle that can be applied to, or is potentially relevant for, other situations.
- Do not write the lesson only as an observation, description or recommendation that lacks justification. Justify the lesson with proof of why it is valid.
- Explain the lesson in the context of the project. For it to be useful to others, they need to understand the situation in which it occurred to know if it might be appropriate or useful for them.

If you are looking for a format for lessons learned, you can consider this template: If you do X and Z conditions are present then Y will happen. This is further described and provided with examples in this guide. You can draw lessons from implemented activities at any time during the project. An approach that can be applied for that purpose is the After Action Review. Lessons learned can be converted into recommendations for a new project or the next phase of a project.

Where to start

As you have noticed, BetterEvaluation.org is a great source of information, materials and examples on evaluations. A good place to start with if you are not familiar with evaluation is the Manager's guide, which is also available in French. It's really useful for people who are managing an evaluation or contracting a party to perform one, including examples of terms of references.

Conclusion

Having clarity on the purpose of your evaluation is key. Will it benefit your work and/or is it to comply with funders' requirements? For enhancing future work, drawing lessons learned can be very valuable.

Suggested reading

If you still feel very new to the subject, you can refer to Evaluation resources for newbies.

<https://www.betterevaluation.org/blog/resources-for-newbies>





The content of this handbook may be re-used under the Creative Commons 3.0 Attribution 3.0 Unported (CC BY 3.0) license

Design and layout: Jared Farrell and Anke van Lenteren – www.designi.se



AfriAlliance is led by the IHE Delft Institute for Water Education (Project Director: Dr. Uta Wehn) and runs from 2016 to 2021. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689162.

