How to conduct a secondary data review

Explanatory Note

Conducting a systematic secondary data review for assessments or data collection exercises need not be complex or unwieldy. This guidance note gives some practical advice to facilitate the process. It is designed for individuals considering a review and does not require specific information management skills.

Table of Contents

1. Why conduct a secondary data review? ................................................................. 2

2. Step by step secondary data review ....................................................................... 3
   A. Step 1 Plan ........................................................................................................... 3
      1.1 Set clear objectives ......................................................................................... 3
      1.2 Detail information needs .................................................................................. 3
      1.3 Determine resources required .......................................................................... 4
   B. Step 2 Collect and collate ................................................................................... 4
      2.1 Locate data ........................................................................................................ 4
      2.2 Organize your data .......................................................................................... 6
      2.3 Determine usability and trustworthiness ......................................................... 7
      2.4 Consolidate and describe ................................................................................ 8
   C. Step 3: Draw conclusions .................................................................................... 9
      3.1 Jointly analyse .................................................................................................. 9
      3.2 Identify information/ data gaps ....................................................................... 9
   D. Step 4: Share information .................................................................................... 10

Key resources ............................................................................................................... 10
1. Why conduct a secondary data review?

A secondary data review is a process of examining and analysing information or data that has been previously collected by another actor, either outside of UNHCR, or within UNHCR, often with a different purpose in mind.

A review is an essential component of all data collection exercises as it avoids a duplication of efforts and saves time and resources. It can provide information that cannot be collected first hand, for instance on the situation before a crisis and enables a broader understanding than primary data collection can provide. Secondary data reviews vary in scope and duration, but this note is applicable to all types of reviews.

Key tips for reviewing secondary data

- **Know what you are looking for:** Not everything is useful or relevant, so collate only the information you will use. Secondary data review content should reflect the specific information priorities of the context only. Set parameters for data relevance, such as outdated data or specific geographic areas and topics. What you don’t know is as important as what you do know, so ensure that information gaps are identified.

- **Balance time and importance:** Decide whether the importance of the data justifies the time required to find it and know when to stop. Do not overextend the review, especially in the early phases of a crisis. Make it sufficiently broad to capture the full situation but narrow enough to be manageable. Balance the importance of the data versus the time needed to find or process it.

- **Do no harm:** Store, process and share personal or sensitive data in line with UNHCR data protection and security principles.

- **Judge the quality of the data:**
  
  Check the methodology, including the sampling, date of data collection, and appropriateness of data collection tool(s), and review the level of consensus amongst the data. Data collected by different actors and for different purposes may use different methodologies and definitions.

  Note any agreement or disagreement between data sources, as this may inform the assessment exercise’s information needs. This could also include a discussion on which data sources are officially endorsed or accepted by UNHCR and stakeholders in the context.

- **Maintain updates:** Regularly update your secondary data review to track developments and save time for more in-depth reviews when necessary.
2. Step by step secondary data review

A solid secondary data review consists of four main steps, the four main steps commonly used on data collection process, see figure 1 below.

Figure 1 - Secondary data review process

| Step 1: Plan | Step 2: Collect and Collate | Step 3: Draw Conclusions | Step 4: Share |

A. Step 1 Plan

1.1 Set clear objectives

A review of secondary data requires a clear objective that guides the data collation and analysis process. Common objectives of a secondary data review include:

- Define what is known and unknown about a current situation.
- Compare the baseline situation to the current conditions to define the impact of any changes in the context, such as a new crisis.
- Identify lessons learned from similar crises impact and common coping mechanisms.
- Define limitations of available information to inform the design of primary data collection exercises.
- Complement and triangulate primary data.

A clear objective includes reference to the type of decision to be informed, the deadline, and the topics, population groups and geographic areas to be covered.

1.2 Detail information needs

Based on the objective, define the specific information needs. Consider the following parameters of the review:

- Geographic scope and level of detail required (national, sub-national, village etc.)
- Time period covered (current, since the start of the crisis, pre-crisis etc.)
- Themes/sectors included (multisector, onesector etc.)
- Population groups (refugees, hostcommunities, age, gender, other diversity factors etc.)
- Categories of analysis (before and after the start of the crisis, male/female, urban/rural, socio-economic background etc.)

Following this, define the specific information/data required for the assessment. You can refer to the following sources that have listed the common information needs of humanitarian decision makers during an emergency:

- The Coordinated Data Scramble list of information needs in the first phase after a natural disaster.
- The Needs Assessment for Refugee Emergencies (NARE) checklist.

Record these information/data needs on the secondary data review analysis plan. This plan is the manual for the secondary data review of your assessment and helps you to focus on the information/data that the assessment exercise is looking for. The purpose of the secondary data review is to identify and record which
of this information/data has been already collected and which are the information/data gaps that will need to be collected through primary data collection.

Develop a report template based on the information needs from the secondary data analysis plan, with headings covering the information needs identified and key notes on the methodology used for the secondary data review. Don’t spend too much time on the lay-out and formatting as the template will likely change during the process. For instance, findings presented per geographic area can be replaced with a breakdown per affected group, or sectors, if these categorizations are deemed more important.

1.3 Determine resources required

The analysis team should comprise staff with contextual knowledge, and quantitative and qualitative analysis skills. To determine the size and profile of the team required, consider the following:

- **Volume**: In light of the scope of the review and situation, how much data is expected? For example, a middle-income country often has a lot of information available, including household studies and monitoring systems. Likewise in a protracted crisis, countries with a large humanitarian presence and/or a country where disasters occur regularly, will likely have lot of information. Reliefweb post frequency can serve as an initial indicator of data volume.

- **One or multiple languages**: What are the most important languages to include within the review?

- **Raw or processed data**: Is there are lot of raw data that still requires processing?

- **Type of data**: Is the data available mostly quantitate, qualitative or a combination of both? Is geospatial analysis required?

- **Public vs non-public data**: Will the exercise primarily include data that is already available, or should time and resources be dedicated to obtain data that is not yet public? Do partnerships or data sharing agreements have to be set up?

Have a look at what already exists – although a secondary data review used to be an overlooked component of assessments, multiple actors are now regularly providing secondary data analysis. This includes, but is not limited to ACAPS, Clusters and sector working groups… If the required skills or capacity are not available within the team, consider outsourcing the secondary data review to specialized actors or individuals.

B. Step 2 Collect and collate

2.1 Locate data

Locate and track reports, datasets, and analysis products with pre- and in-crisis information. An assessment registry is a good starting point, as it provides a list of data collection exercises relevant to the context. The Raw Internal Data Library (RIDL), a UNHCR data library for maintaining operational raw data from monitoring and needs assessments, can provide an overview of the internal relevant data. For examples of other sources of information, see figure 2 below.

Figure 2 – Examples of information sources

<table>
<thead>
<tr>
<th>SOURCES OF INFORMATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National data sources</strong></td>
<td>Statistics about a specific country’s economy, society and population (census data)</td>
</tr>
<tr>
<td>- National statistics offices</td>
<td>Specific national information about a specific area, such as education, environment, or agriculture.</td>
</tr>
</tbody>
</table>
## UNHCR Resources

<table>
<thead>
<tr>
<th>Resource</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNHCR Country Operation Plans</td>
<td>A clear and concise summary of UNHCR’s programme goals, objectives, and priorities for each of its country operations.</td>
</tr>
<tr>
<td>Registration systems (ProGres)</td>
<td>UNHCR’s corporate, centralized, web-based case management software application, with registration data (demographic and protection data).</td>
</tr>
<tr>
<td>UNHCR Operational Data Portal</td>
<td>An information and data sharing platform to facilitate coordination of refugee emergencies.</td>
</tr>
<tr>
<td>UNHCR’s Refugee Data Finger</td>
<td>The database contains information about forcibly displaced populations spanning more than 70 years of statistical activities. It covers displaced populations such as refugees, asylum-seekers, and internally displaced people, including their demographics. Stateless people are also included, most of whom have never been displaced.</td>
</tr>
<tr>
<td>UNHCR data</td>
<td>UNHCR main reports as Global trends, Global Appeal and Global Report.</td>
</tr>
</tbody>
</table>

## International Organizations

<table>
<thead>
<tr>
<th>Organization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Bank development indicators</td>
<td>It is the primary World Bank collection of development indicators, compiled from officially recognized international sources.</td>
</tr>
<tr>
<td>World Health Organization profiles</td>
<td>WHO country profiles present selected data, statistics and information to provide national health profiles at given points in time.</td>
</tr>
<tr>
<td>UNICEF MICS data</td>
<td>The Multiple Indicator Cluster Surveys programme is a source of statistically sound and internationally comparable data on children and women worldwide.</td>
</tr>
</tbody>
</table>

## Geospatial data

<table>
<thead>
<tr>
<th>Organization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNOSAT</td>
<td>UNOSAT provides satellite image analysis during humanitarian emergencies related to disasters, complex emergencies and conflict situations.</td>
</tr>
</tbody>
</table>

## Humanitarian Networks and Platforms

<table>
<thead>
<tr>
<th>Organization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACAPS</td>
<td>ACAPS provides data analysis on relevant topics that enable a better understanding of the dynamic driving humanitarian needs across the globe and in specific contexts, as well as live updates of countries with existing humanitarian crises or prone to disasters.</td>
</tr>
<tr>
<td>Humanitarian Practice Network (ODI)</td>
<td>Direct access to up-to-date humanitarian reports and publications. Covering the humanitarian situations affecting people across the world.</td>
</tr>
<tr>
<td>Assessment registries</td>
<td>An assessment registry record planned, on-going and completed assessments lead by different agencies in a specific country or region.</td>
</tr>
<tr>
<td>Cluster reports and inter-cluster</td>
<td>Reports with cluster or inter-cluster information for a specific geographic area in a specific period of time.</td>
</tr>
</tbody>
</table>

Other sources of data, such as national, international and social media, community networks, and local informants and community members, may also be considered.
To encourage the sharing of politically sensitive data that has not been made public, set up a data sharing procedure. The entity sharing the information can decide if the data can be attributed, anonymized, or only used for analysis (see Figure 3).

**Figure 3 - Example data sharing procedure**

<table>
<thead>
<tr>
<th>Open</th>
<th>Open – data can be quoted and attributed to the organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted</td>
<td>Can be quoted and attributed to ‘an international NGO’ or ‘a national NGO’ etc.</td>
</tr>
<tr>
<td>Protected</td>
<td>Can be quoted and attributed to ‘a trusted source’</td>
</tr>
<tr>
<td>Confidential</td>
<td>Cannot be quoted directly but can be used for analysis and published without attribution.</td>
</tr>
</tbody>
</table>

Make sure you store the data in a way that makes it easy for retrieval. Use standardized file naming, which includes information on the author, report title, date of the report and/or how the data should be treated in line with any data sharing agreement. At this point, include all reports and data that could be relevant, the selection of the most useful content will be undertaken during the next step.

### 2.2 Organize your data

Find and organize the most relevant content from all reports collected. To avoid getting lost in the often-large amounts of qualitative data, it’s crucial to have a structured way of labeling and storing these pieces of information through tagging.

For the tagging process, you can use categories to help organize the data. Here are some common categories that can be used (more details can be found in the secondary data review analysis template):

- **Source of information**: Date, author, data collection technique.
- **Scope**: Geographic areas, sectors and affected groups covered.
- **Subject**: Subject categorization can be done organically, with categories being assigned throughout the process, depending on the content. However, to streamline the process, it is recommended to use a pre-existing categorized structure, based on the analytical framework and information/data needs. This is an essential requirement if more than one analyst is working on the categorization.
- **Data Use**: In line with the data sharing procedure (see Figure 3), provide details on whether the data is open, restricted, protected, or confidential.
- **Reliability of the source**: Define how trustworthy the source is based on factors like its track record for accuracy, technical expertise in the subject, and potential for bias. You can use a reliability scale (see figure 4) to assess this.

**Don’t skip categorization!** Categorizing your data might seem like an extra step but it’s essential when dealing with large amounts of qualitative information, or a team of analysts. Skipping this step can lead to chaos down the road that will take substantial time to untangle during the processing and analysis phase. Categorize all qualitative data and aggregated quantitative data (labelling according to categories such as sector, geographical area, risk, or problem identified). This ensures the information can be:

- Easily retrieved
- Grouped by topics of interest (sector, geographical area etc.), simplifying analysis
- Processed by multiple analysts, as a categorized database can merge inputs from different staff
- Used for different secondary data review projects
**Train the analysts.** If multiple analysts are involved in categorization, training is required to ensure categories are consistent. Define each category clearly, with examples relevant to the context, to ensure information is categorized consistently across sources, time, and staff. A good example of category definition can be found via the GIMAC Analytical Framework Guide.

**Data categorization or tagging** of large amounts of data needs to be done in a specific software or platform. MS Excel spreadsheets are often used for that. A platform specifically developed for humanitarian actors to process secondary data, DEEP, enables customizable analysis frameworks for secondary data review, and tagging of information contained in large amounts of data and documents. DEEP works by projects, so several secondary data review exercises can be implemented at the same time. Consider the capacity of the secondary data review staff regarding their ability to use specialized software such as DEEP, or to use simpler software such as MS Excel.

### 2.3 Determine usability and trustworthiness

After information has been categorized and systematized, select the information that is most useable and trustworthy. The following criteria can be used to select the information that is most useful:

- **Relevancy:** Does it cover the geographic area, topic, population group, time period of interest?
- **Granularity:** Does it provide the level of detail required?
- **Comparability:** Does it allow for comparison with other datasets important to your review?
- **Reliability:** Looking at the source of the information and the method used to collect the information, is the information reliable? Be wary of including data that comes without a detailed description of the methodology and questionnaire.

**Figure 4 - Example Reliability Scale**

<table>
<thead>
<tr>
<th>Reliability level</th>
<th>Track record accuracy</th>
<th>Expertise</th>
<th>Motivation for bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Reliable</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2 Fairly reliable</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3 Fairly unreliable</td>
<td>No</td>
<td>No</td>
<td>Possible</td>
</tr>
<tr>
<td>4 Unreliable</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>0 Cannot be judged</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quantitative research (e.g. household surveys with a representative sample) and qualitative research (e.g. focus group discussions with a purposive sample) are structurally different in design. Evaluating their usefulness and reliability therefore requires a different approach. Figure 5 describes three errors to look out for in quantitative and qualitative humanitarian assessments.

<table>
<thead>
<tr>
<th>Design</th>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Does the <strong>sampling frame</strong> include all population groups of interest?</td>
<td>Could the <strong>participant selection procedure</strong> have influenced the findings?</td>
</tr>
<tr>
<td></td>
<td>E.g. A random sample based on an outdated list with IDP households is not representative of the conditions faced by new arrivals.</td>
<td>E.g. If all participants were selected by male community leaders, the participant selection might be skewed</td>
</tr>
</tbody>
</table>
### Measurement Error

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are all the terms, theories and concepts mentioned in the study clearly defined?</td>
<td>How were dissenting opinions encouraged and captured?</td>
</tr>
<tr>
<td>Is it likely key terms and concepts were interpreted in the same way by the respondent, assessment teams and analyst?</td>
<td>There is a high risk of powerful participants leading the discussion. Were other opinions captured and welcomed or are only dominant views expressed?</td>
</tr>
<tr>
<td>Specifically review difficult to measure issues, such as SGBV, negative coping mechanisms and illegal activities.</td>
<td>Are conclusions extrapolated to a wide population without highlighting the limitations of the non-representative sample?</td>
</tr>
<tr>
<td>E.g. Assessment results will be erroneous if respondents’ definition of a child differs from the age range used by the analyst interpreting the responses.</td>
<td>E.g. Results of focus group discussions with a small number of women, selected because of their participation in a specific relief project, can be indicative but not representative of the situation for other women.</td>
</tr>
</tbody>
</table>

Figure 5 - Common errors to look out for while judging quantitative or qualitative studies

The objective is not to discard all that is imperfect, as that may be all there is. A careful evaluation aims at collecting a body of data that is most useful according to the objectives and to obtain insights into the limitations of the available information.

### 2.4 Consolidate and describe

Consolidate the information by summarising findings by geographical area, population groups of interest, and/or topics. Start with describing the largest, most reliable datasets, describing the general situation. Afterwards look for more detail.

In case of inconsistent or conflicting information, there are two options:

- Only include the most reliable and useful information
- Include all information and explain possible reasons for the divergence between findings.

Provide comparisons, using the pre-defined categories of analysis. For example, how do the findings differ between sites, population groups or over time? Use common standards (e.g. Sphere standards, UNHCR Emergency Handbook) to put findings into perspective.

Afterwards, the secondary data review analysis plan can be completed, by adding the information that is most useful and reliable. Multiple data sources often have to be combined to cover all the information required.
C. Step 3: Draw conclusions

3.1 Jointly analyse

Once all data has been consolidated, analyze your data by:

- **Explaining** relationships between concerns, looking at possible cause-and-effect and underlying factors
- **Interpreting** the findings by prioritizing geographical areas, groups, needs and protection concerns based on an assessment of severity or scope
- **Anticipating** what might happen next by looking at the likely evolution over time.

Analysis is better done in groups. A **joint analysis session** with subject-matter experts from different backgrounds and representatives of the affected population is an effective way to review the findings, select what is surprising and draw main conclusions. For one example of the different types of analysis see figure 6 below.

Figure 6 - Simplified Analysis Flow

<table>
<thead>
<tr>
<th>Describe</th>
<th>250 families in Zone A have access to 2 liter water per person a day, compared to the emergency standard of 15 liters water per person a day.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explain</td>
<td>There is only one handpump in zone A and there has been no distribution of water containers. Families have to que for hours and are not able to fetch sufficient water.</td>
</tr>
<tr>
<td>Interpret</td>
<td>The water situation in Zone A is more severe compared to other zones, with neighbouring zones showing at least 7 liters water per person a day.</td>
</tr>
<tr>
<td>Anticipate</td>
<td>The rainy season will not start for another 4 months and, unless additional support is provided, the water shortages will turn catastrophic.</td>
</tr>
</tbody>
</table>

3.2 Identify information/data gaps

Identify information gaps by comparing available information against the information required within the secondary data analysis plan. There are four main types of information gaps:

- Geographical (i.e. there is no information on an affected area because of a lack of coverage)
- Thematic (e.g. specific sectors or topics are not covered by the existing data)
- Time (e.g. historical trends are available but there is not up-to-date information on population displacement)
- Or detail: (e.g. there is information available on the food security situation, but it is not possible to disaggregate this data by age, gender, or diversity considerations).

Because information needs often require combining data from various sources, turn to the "limitations and gaps" sections on the secondary data review analysis template to pinpoint any gaps in assessment information. When resources permit and it's a significant concern, you can consider filling in these gaps through primary data collection.

To streamline this process and avoid duplication, rely on the data analysis plan template. It helps you sort out which data should be collected anew through primary data collection and which data has already been identified through the secondary data review. This approach ensures a well-organized and efficient approach to gathering the necessary information. Additional information on this can be found on the Assessment Methodology outline document.

Provide recommendations to the audience, which can include recommendations for (urgent) action and additional data collection to address information gaps.
D. Step 4: Share information

The same guiding principles that apply to all publications also extend to the outcomes of a secondary data review. It's crucial to tailor these outputs to your target audience, leveraging visuals to convey crucial messages. Consider translating the information to ensure it reaches a wider audience, and make sure to share the findings promptly.

To support the overall humanitarian response, go beyond sharing just the findings report. Also share:

- A repository housing non-confidential situation reports (ensure contributors consent to dissemination).
- The organized database of data (with personal, sensitive, or confidential information removed or anonymized). For instance, if the secondary data review used the "DEEP" platform and there's a technical audience interested in the full analysis process, you can grant access to the DEEP project and the categorization guidelines (tagging process).
- The secondary data review report, complete with the methodology employed and key findings. Properly crediting sources is particularly vital in a secondary data review. This practice acknowledges the data collectors and maintains transparency about the information used.

Remember to acknowledge the review's limitations. For example, highlight situations where detailed exploration was constrained due to time constraints. Don't shy away from communicating uncertainty. When presenting significant findings and recommendations, it's valuable to include information about the reliability of sources and any assumptions made to arrive at the final conclusions. This level of transparency enhances the credibility of your insights.

Key resources

- ACAPS, Secondary data review Sudden Onset Natural Disasters
- ACAPS, Free available datasets
- ACPAS, Analysis Workflow
- ECB/ACAPS, The Good Enough Guide to Assessments
- Food Security Cluster and OCHA, Field Guide to Data Sharing
- UNHCR, Needs Assessment for Refugee Emergencies (NARE)
- UNHCR, Needs Assessment Handbook
- UNHCR, Policy on the Protection of Personal Data of Persons of Concern
- WFP, Comprehensive Food Security & Vulnerability Analysis Guidelines, Chapter 3: Desk Study Literature Review and Secondary Data
- UNHCR Raw Internal Data Library (RIDL)