

# Secondary Data Review

## Purpose

A review of secondary data is sometimes seen as a difficult and unwieldy exercise. However, it doesn't have to be. This guidance notes gives some practical advice on how to undertake a systematic secondary data review in an emergency setting. The note is for all those considering a review, and no specific information management skills are required to understand or use it.

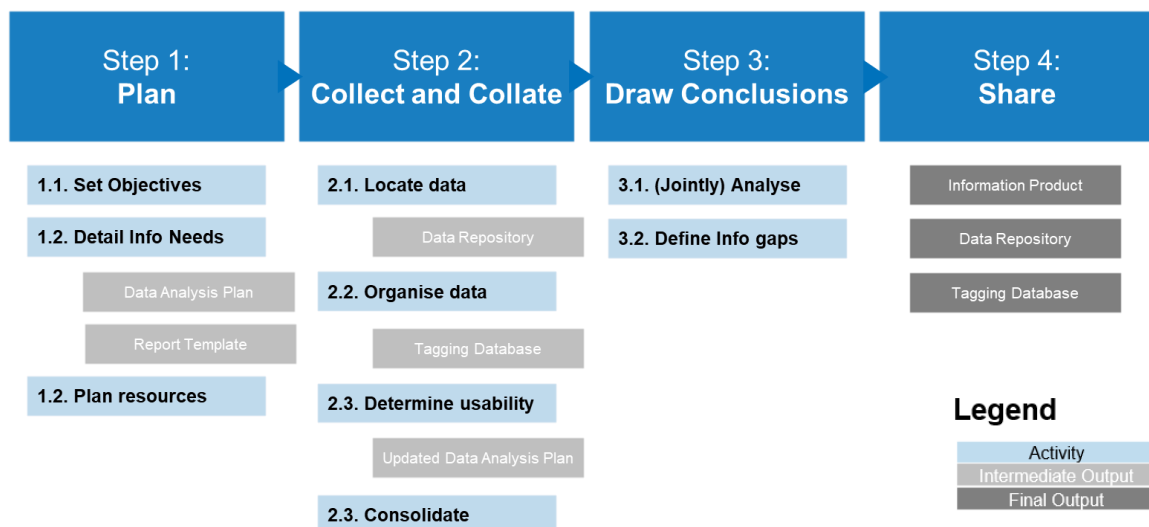
Secondary data is defined as information that has been collected by a different actor or, if it was collected by colleagues, with a different objective in mind. A review is an essential component of all data collection exercises as it avoids a duplication of efforts and saves time and resources. It can provide information that cannot be collected first hand, for instance on the situation before the crisis. And it facilitates a much broader understanding than primary data collection will be able to provide. Secondary data reviews come in many shapes and forms. Some reviews are finalised within day, while others take multiple months. Examples of outputs of secondary data reviews include Factbooks, Briefing Notes and Situation Analysis Reports. This note is relevant to all types of reviews, regardless of the intended output or resources available.

## Principles

- **Prepare for volume:** There is always more data than is foreseen, so plan for a system which can manage and process at least double the data originally expected. It is much easier to scale down processing practices, than increasing them half-way into the review.
- **Dedicate sufficient resources:** A thorough secondary data review requires qualified staff and time. A review can be undertaken remotely as long as there is a clear link to field operations.
- **Do no harm:** Store, process and share personal or sensitive data in line with data protection and security principles.
- **Not an on-off exercise:** Maintaining an updated secondary data review is an efficient way to keep track of important developments and save time when a more in-depth review is required.
- **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.

Figure 1 - Step by step Secondary Data Review

A solid secondary data review consists of four main steps. This method is similar to other commonly used approaches to data collection and analysis, including the UNCHR Needs Assessment Process:



## STEP 1: PLAN

### 1.1 Set objectives

As with all types of data collection, a review of secondary data starts with the definition of a clear objective of the review. Common objectives of a secondary data review include to:

- Define what is known and unknown about a current situation
- Compare the pre-crisis situation to the current conditions to define the impact of the crisis
- Identify lessons learned from similar crises on crisis 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 (currently, since the start of the crisis, pre-crisis etc.)
- Themes/sectors included (multisector, one sector etc.)
- Population groups (refugees, host communities, 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.)

Afterwards, define the specific information needs. Several actors have listed what humanitarian decisionmakers commonly need to know during an emergency. This work can be used to help design the list of information needs;

- The Coordinated Data Scramble list of information needs in the first phase after a natural disaster.
- The Needs Assessment for Refugee Emergencies (NARE) checklist
- A list developed by ACAPS on information needs of emergency decision makers
- Sector specific standards and assessment tools, such as the Rapid Protection Assessment Checklist Data Review.

Record the information needs in a **data analysis plan** (see figure 2). This data analysis plan is the manual for your review and helps to focus the data collection on what is of key importance, highlights remaining gaps and clarifies the objectives. Part B, the information available, will be populated once data has been collected and reviewed (step 2 – Collect and Collate) Keep the data analysis plan as light as possible to facilitate regular updating.

Describe as well what will **not be included**. If certain geographic areas, topics or groups are for instance excluded from the research, include this in the plan to help focus data collection and align expectations.

Based on the information needs, develop a **report template**, with headings covering the information needs identified. Examples of templates for multi-sector reviews are the UNHCR Situation Analysis, Secondary Data Review Template and the JIPS Desk Review. At this stage, do not spend too much time on the lay-out and formatting as the template will likely change during the process. Findings presented per geographic area can for instance be replaced with a breakdown per affected group or sectors if it turns out these categorisations are more important.

Figure 2 : Data analysis plan

A. INFORMATION REQUIRED							B. INFORMATION AVAILABLE	
Topic	Specific Question	Sub-topic	Indicator	Timing data	Disaggregation	Priority	Source	Etc.
Humanitarian Access	What are the main constraints for the population to accessing humanitarian aid?	Access constraints	Access constraints in order of priority, by vulnerable group	< one month	By district, per vulnerable group	Medium		
		Coverage response	% of people in need who are not receiving aid	< one month	By district, per vulnerable group	High		

### 1.3 Determine resources required

The analysis team should combine staff with contextual knowledge, 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 assessment and characteristics of the crisis, how much data is expected? A middle-income country with a functioning Government usually has a lot of information available, including household studies and monitoring systems. In addition, a protracted crisis, countries with a large humanitarian presence and/or country where disasters occur regularly, generate a lot of possibly useful information. The number of posts on Reliefweb can provide an initial indication of the volume of data to be processed. As an example, by November 2017, the countries with most entries posted on Reliefweb in 2017 were Syria (2,108 documents) Peru (2,007) and South Sudan (1,919).
- **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 (e.g. the Child Protection Secondary Data Reviews) If required skills or capacity are not available within the team, consider **outsourcing** the secondary data review to specialised actors or individuals.

Other resources required for the review include a data base to store and tag the data and licenses for any software used (see step 2.1.)

## 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 other sources of information, see Figure 3.

Figure 3 - Secondary Data Resources

PRE-CRISIS INFORMATION	LESSONS LEARNED	IN-CRISIS INFORMATION
<ul style="list-style-type: none"> <li>• National statistics offices, censuses, and relevant line ministries</li> <li>• UNHCR resources include Country Operation Plans, registration systems (ProGres), UNHCR data portal and map portal, popstats.unhcr.org</li> <li>• World Bank development indicators</li> <li>• World Health Organization country epidemiological profiles</li> <li>• UNICEF Multiple Indicator Cluster Surveys (MICS) data</li> <li>• Remote sensing</li> <li>• Civil society, think thanks and UN agency reports</li> <li>• Index for Risk Management (INFORM)</li> <li>• Academic journals and books (accessible through INASP or Research for life)</li> <li>• UNHCR RIDL</li> </ul>	<ul style="list-style-type: none"> <li>• UNHCR Evaluation Reports</li> <li>• ALNAP lessons learned</li> <li>• Civil society, think thanks, and UN agency reports on lessons learned, evaluations</li> <li>• ACAPS Disaster Summary Sheets</li> <li>• Previous appeals, Humanitarian Needs Overviews and Humanitarian Response Plans</li> <li>• Humanitarian Practice Network (ODI)</li> </ul>	<ul style="list-style-type: none"> <li>• National and international Media</li> <li>• ReliefWeb</li> <li>• Humanitarian Data Exchange (HDX)</li> <li>• IASC Common Operational Datasets (CODs)</li> <li>• IRIN News</li> <li>• UNHCR data portal and map portal</li> <li>• Civil society organizations, government, and UN agency situation reports</li> <li>• European Media Monitoring News Explorer</li> <li>• Cluster and inter-cluster reports, websites, and meetings</li> <li>• Remote sensing</li> <li>• Social media, other media, blogs, crowdsourcing</li> <li>• Personal networks</li> <li>• Funding appeals</li> <li>• 3, 4 or 5Ws</li> <li>• UNHCR RIDL</li> </ul>

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, anonymised, or only used for analysis (see Figure 4).

Figure 4 - Example data sharing procedure

<b>Open</b>	Open – data can be quoted and attributed to the organisation
<b>Restricted</b>	Can be quoted and attributed to 'an international NGO' or 'a national NGO' etc.
<b>Protected</b>	Can be quoted and attributed to 'a trusted source'
<b>Confidential</b>	Cannot be quoted directly but can be used for analysis and the analytical deduction published without any attribution.

**Store** the data in a way that allows for easy retrieval. Use standardised file naming, which includes information on the author, report title, date of the report and how the data should be treated in line with the data sharing procedure (e.g. *20171104\_UNHCR\_SituationReport\_Protected*). At this point, include all reports and data that could be relevant, the selection of the most useful content will be undertaken during step 2.1.

## 2.1 Organise your data

The next step is to find and organise the most relevant content from all reports collected. To avoid getting lost in the often-large amounts of qualitative data, a systematised way of labelling pieces of information and storing these in a **tagging database** is required.

Common tagging categories include:

- **Source of information:** Date, author, data collection technique
- **Scope:** Geographic areas, sectors and affected groups covered
- **Subject:** Subject tagging can be done organically, with tags being assigned throughout the process, depending on the content. However, to streamline the process, it is recommended to use a pre-existing tagging structure, based on your data analysis plan. This is an essential requirement if more than one analyst is working on the tagging.
- **Data Use:** In line with the data sharing procedure (see Figure 4), provide details on whether the data is open, restricted, protected or confidential.
- **Reliability of the source:** Define the reliability of the information source, based on the track record for accuracy (does the source have a reputation of providing accurate data), Technical expertise (the source has a specific expertise in the topic it

reports on), motivation for bias (source does not have an obvious motive to provide information that is misleading or incomplete). (See figure 6)

**The case for tagging:** Some analysts decide to skip step 2.1. to save time or resources. However, a review that deals with large volumes of qualitative data or multiple analysts will quickly turn into a mess unless a systematic tagging database is used. A mess that will take substantial time to untangle during the processing and analysis phase. Tagging all qualitative data and aggregated quantitative data (for instance by labelling according to sector, geographical area, risk, or problem identified) ensures the information can be:

- Easily retrieved
- Grouped by topics of interest (sector, geographical area etc.) thereby simplifying analysis
- Processed by multiple analysts, as a tagging database can merge inputs from different staff
- Used for different secondary data review projects

**Train the taggers:** If multiple analysts are involved in tagging, training is required to ensure tagging is consistent. Define each tag clearly, with examples relevant to the context, to ensure the information is tagged consistently across sources, time and staff.

Regular calibration exercises can improve consistency over time. During such an exercise, analysts are requested to tag the same pieces of information. This reveals possible tagging discrepancies and aligns analyst's tagging practices.

The training should touch on good practices for quickly reviewing sources: there is for instance no need to read every page, analysts should instead review the summary and look for key words.

Also train taggers on identifying and tagging 'latent content' — hidden meanings in pieces of information. For example, local media reporting on an unconfirmed and unlikely HIV outbreak following a refugee influx is not only a health concern, but the rhetoric is also emblematic of possible tensions with host communities and refugee scapegoating.

Figure 5 - Example tagging structure

Source	Date	Data Collection	Geographic Area	Narrative	Sector	Reliability	Confidentiality
OCHA	August 2017	Incident reporting	All districts	Access restrictions by all parties to the conflict continue to hamper humanitarian outreach. Of the 20.7 million people in need of assistance, approximately 1.7 million people live in districts with the highest access constraints.	Access	Reliable	Open

**DEEP Platform:** Excel spreadsheets are often used to store and tag data. However, more user-friendly platforms have emerged in recent years. The DEEP was specifically developed by and for humanitarian actors to process secondary data. It is a free, open source software for collaborative secondary data review and managing unstructured data.

Users can upload a variety of sources (news articles, PDFs, Word documents etc) and tag/categorize them using custom analytical frameworks. Catalogued information can then be exported into Excel or Word for further analysis.

For more information on the DEEP, contact the UNHCR Needs Assessment Help Desk.

- **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 6 - Example Reliability Scale

Reliability level		Track record accuracy	Expertise	Motivator for bias
1	Reliable	Yes	Yes	No
2	Fairly reliable	Yes	No	No
3	Fairly unreliable	No	No	Possible
4	Unreliable	No	No	Yes
0	Cannot be judged			

**Determine usability and trustworthiness**

After all information has been systematised, 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?

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<sup>1</sup>. Evaluating their usefulness and reliability therefore requires a different approach. Figure 7 describes three errors to look out for in quantitative and qualitative humanitarian assessments.<sup>2</sup>

Keep in mind that the main objective of this evaluation is not to discard all that is imperfect. In many settings imperfect is all there is. A careful evaluation instead aims at collecting a body of data that is most useful for your purposes and obtain insights into the limitations of the available information.

<sup>1</sup> For a clear description of the difference between quantitative and qualitative research strategies and the implications of the different approaches see Alan Bryman, Social Research Methods, Oxford University Press.

<sup>2</sup> For more detail on how to judge quality and usability of data collected during rapid needs assessments review, see ACAPS, How sure are you?

Figure 7 - Common errors to look out for while judging quantitative or qualitative studies

	Quantitative	Qualitative
Design Error	Does the <b>sampling frame</b> include all population groups of interest? E.g. a random sample based on an outdated list with IDP households is not representative of the conditions faced by new arrivals.	Could the <b>participant selection procedure</b> have influenced the findings? E.g. If all participants were selected by male community leaders, the participant selection might be skewed towards supporters of the leadership.
Measurement Error	Are all the terms, theories and <b>concepts</b> mentioned in the study clearly defined? Is it likely key terms and concepts were interpreted in the same way by the respondent, assessment teams and analyst alike? Specifically review difficult to measure issues, such as SGBV, negative coping mechanisms and illegal activities. 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.	How were <b>dissenting opinions</b> encouraged and captured? There is a high risk of powerful participants leading the discussion. Were other opinions captured and welcomed or are only dominant views expressed?
Processing and Analysis Error	Are the differences highlighted actually <b>significant</b> ? If there is no statistician around, use online confidence interval calculators such as this one to check ranges. E.g. the difference between 49% and 51% is unlikely to be significant as it is a reflection of the sample design, instead of an actual difference in the situation.	Are conclusions <b>extrapolated</b> to a wide population without highlighting the limitations of the non-representative sample? 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.

Afterwards, the **data 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 (see Figure 8).

Figure 8 - Data analysis plan part II: Information available

A. INFORMATION REQUIRED				B. INFORMATION AVAILABLE				
Etc.	Indicator	Timing data	Disaggregation	Source	Timing data	Disaggregation	Reliability	Location
	Access constraints in order of priority, by vulnerable group	< one month	By district, per vulnerable group	OCHA	August 2017	By district	Reliable	URL
				WFP	March 2017	Per vulnerable group	Reliable	URL

## 2.4 Consolidate and describe

Now that all relevant information is tagged, structured and stored, it is time to synthesise the quantitative and qualitative data. 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. How do the findings for instance differ between sites, population groups or over time? Use common emergency standards (e.g. Sphere standards, UNHCR Emergency Handbook, WASH Standards), to put findings into perspective.

## Step 3: Draw conclusions

### 3.1. (Jointly) analyse

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

- Explaining relationships between concerns, looking at possible cause-and-effect and underlying factors
- Interpreting the findings by prioritising 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.

### 3.2. Identify information gaps

Afterwards, identify the information gaps by comparing the information available to the information required within the 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).

#### Simplified Analysis Flow

**Describe:** 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.

**Explain:** 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.

**Interpret:** 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.

**Anticipate:** The rainy season will not start for another 4 months and, unless additional support is provided, the water shortages will turn catastrophic.

Provide **recommendations** to the audience, which can include recommendations for (urgent) action and additional data collection to address information gaps.

## Step 4: Share information

The principles relevant to all publications also apply to the outputs of a secondary data review, including the need to tailor the outputs to the audience, use visuals to channel key messages, consider translation to broaden the audience and share the information in a timely fashion.

To support the overall humanitarian response, share not only the report with the main findings, but also the:

- Repository with all non-confidential situation reports (make sure all those who have contributed information agree to its dissemination)

- The tagging data base (make sure all personal, sensitive or confidential data is removed or anonymised)

Correct **referencing of sources** is of particular importance in a secondary data review, to credit those who have collected the data and be transparent on the information used.

Highlight the **limitations of the review**, for instance if it there was not enough time to go into detail or review a certain body of information.

**Communicate uncertainty:** When developing key findings and recommendations, include the reliability of the sources and any assumptions that had to be made to reach the final conclusions.

## Key Tools

- Assessment Registry Template
- Situation Analysis Template
- Secondary Data Review Template
- Data Extraction and Exploration Platform, (DEEP)
- Raw Internal Data Library, (RIDL)<sup>3</sup>

## Key resources

- ACAPS, Technical brief: Estimation of Affected Population Figures, 2012
- ACAPS, How Sure Are You, 2013
- ACAPS, Secondary Data Review, Sudden Onset Natural Disasters, 2014
- ACAPS, Sources of error in humanitarian assessments, 2017
- ACAPS, Spotting dubious data, 2015
- Carr-Hill, Missing millions and measuring development progress, 2013
- IASC, Humanitarian Population Figures, 2016
- ECB/ACAPS, The Good Enough Guide to Assessments, 2014
- Food Security Cluster and OCHA, Field Guide to Data Sharing (DRAFT)
- Global Education Cluster, Guide to Education in Emergencies Needs Assessments, 2016
- Oxfam, Reviewing the Existing Literature, 2012
- Sphere Project, Sphere for Assessments, 2014
- UNHCR, Age, Gender and Diversity Policy, 2011
- UNHCR, Needs Assessment for Refugee Emergencies (NARE)
- UNHCR, Needs Assessment Handbook, 2017 and accompanying tools:
- UNHCR, Policy on the Protection of Personal Data of Persons of Concern, 2015
- WFP, Comprehensive Food Security & Vulnerability Analysis Guidelines, Chapter 3: Desk Study, Literature Review and Secondary Data, 2014

---

<sup>3</sup> The UNHCR Raw Internal Data Library (RIDL) is a secondary data repository to store UNHCR's raw operational data. It is expected to be launched in 2019.