THE CHALLENGE
There is a wealth of data collected on the conditions of the Rohingya refugees and their hosts. The available information comes in all shapes and forms, from structured registration data captured in databases, to WhatsApp messages from field staff. The quantitative data, for instance survey results, is processed and often shared through HDX and HumanitarianInfo.HR. However, as much of the qualitative data is unstructured, data processing and analysis requires a significant investment in terms of time and resources. As a result, a large part of the data remains unprocessed or is not shared beyond specific teams. Valuable information from focus group discussions, media analysis, field staff communication is not consistently fed into an overall analysis of the situation.

THE JOINT APPROACH
Several organisations, including UNHCR and ACAPS, are reviewing how to make better use of the qualitative data generated by the humanitarian community. If these efforts are aligned, the resulting output will support a more joint review of secondary data. Within an aligned system, data processed by UNHCR can be easily merged with data reviewed by ACAPS. While each actor is planning to set up a tailored project, with restricted user access, data that’s not confidential can be exported from that project and used for collective analysis.

There are several software solutions available to manage and process qualitative data, including Envivo, Mxeg and DEEP. Both ACAPS and UNHCR are piloting a project in DEEP, a platform specifically developed by and for humanitarian actors to process substantial amounts of 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.

DATA FLOW

Step 1: Upload information sources to the DEEP software
Step 2: Tag data using an analytical framework, by topic, geo area etc.
Step 3: Export data to Word/Excel and use it to create analysis products

DATA ARCHITECTURE
Secondary data processing involves a review of all secondary data, selecting interesting information snippets and tagging these by geographic location, information source, topic, population groups and other categories of interest. The output can answer specific questions such as: ‘what do we know about protection concerns in camp X’ or ‘what information do we have on the food security situation of Rohingya refugees arriving before 2016’.

The overall objective of a joint approach towards this secondary data processing is twofold:

- Increases sharing of qualitative data currently not processed and/or shared
- The processing and tagging of data by geographic location and topic facilitates use of data for analysis.

Using the same analytical framework is the most useful approach as it ensures datasets can be easily merged. However, even if a different framework is used, if the minimum requirements are met, the
resulting data will be comparable. This means that project data is compatible, even if differing levels of detail are captured. The minimum requirements are to have common tags and definitions for:

- Sector/topics described
- Characteristics of the group described
- Data sharing classification
- Administrative levels
- Type of source

The Analytical Framework and attributes can be found in the ‘Data Architecture Rohingya Refugee Crisis’. While the propose architecture is set-up to inform the design of DEEP projects, it is sufficiently generic to be compatible with other software solutions, including a simple Excel sheet. Other sectors and agencies are encouraged to adopt a similar structure, or to share their data with the joint project for processing.

DATA SHARING
All data is to be shared in a safe and responsible matter, in line with principles outlined within Protection Information Management guidance. Each document and, if necessary, specific information snippets are classified as ‘confidential’ or ‘unprotected’. Only information tagged as ‘unprotected’ will be exported and become publicly available. Information classified as ‘confidential’ data is only accessible to the restricted number of users with access to the specific project.

<table>
<thead>
<tr>
<th>Category</th>
<th>When to use</th>
<th>Disclosure protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidential</td>
<td>Disclosure would cause serious damage to the respondent, community or organisation. Data contains personally identifiable information and is highly sensitive.</td>
<td>Cannot be shared publicly</td>
</tr>
<tr>
<td>Unprotected</td>
<td>All information not classified as confidential. Disclosure is associated with little or minimal risk. Data is insensitive or almost insensitive. Data can be disclosed for public, e.g. on open websites, and is subject to standard copyright rules.</td>
<td>Can be shared publicly</td>
</tr>
</tbody>
</table>

On a monthly basis, the unprotected, structured information will be exported and merged into one Word document and one Excel sheet. These outputs will be shared on HumanitarianResponse.Info, clearly crediting the organisations who have supported data processing and sharing. Specific data requests, e.g. data for a specific period of time or for a specific sector, can be made directly to the Focal Point.

DIVISION OF LABOUR

ACAPS
- Processing of publicly available qualitative data in the joint project. If the data to be processed exceeds capacity, information sources will be prioritized based on expected relevance of the content.
- Processing of qualitative data requested by agencies/sectors that is relevant to the Rohingya humanitarian analysis and can be shared publicly
- DEEP capacity building for interested actors

---

1 This is the philosophy behind the set-up of the CODs/FODs – using common data standards strengthens interoperability and use of datasets.

2 The DEEP team is currently exploring whether one project can host different user rights. Once that has been implemented, different groups of users can have access to different types of information within the same project.
UNHCR
- Processing of publicly available qualitative data in the joint project. If the data to be processed exceeds capacity, information sources will be prioritized based on expected relevance of the content.
- Sharing processed, non-confidential data from the protection specific project to be merged with the outputs of the joint project.

IFRC
- Processing of publicly available qualitative data in the joint project. If the data to be processed exceeds capacity, information sources will be prioritized based on expected relevance of the content.
- Sharing processed, non-confidential data from the protection specific project to be merged with the outputs of the joint project.

ISCG
- Support to identification of common analytical framework and tags, roll-out of project and training of staff
- Liaison with stakeholders to expand information sharing and use.
- Until dedicated resources are in place:
  - Support tagging
  - Serve as a DEEP focal point, including merging and dissemination of outputs and facilitation of specific data requests.

Contact details: ISCG Information Management Unit, IM5@ISCGCXB.ORG