SHO SYRIA 2013
META-EVALUATION

March 2016

Commissioned by UNICEF Netherlands on behalf of Giro555,
Samenwerkende Hulp Organisaties (SHO)
info@unicef.nl, info@giro555.nl

Table of Contents

Meta-evaluation of the 2013 Syria action

1) Introduction

The organisations participating in Samenwerkende Hulp Organisaties (SHO), an alliance of (in 2013) ten Dutch aid organisations, cooperate in response to major humanitarian crises. They are jointly accountable for the expenditure of the funds, through their own respective mechanisms. On March 25, 2013 the SHO launched a national campaign in order to scale-up the individual organisations' relief operations in Syria and/or in neighbouring countries. The SHO received more than 5 million EURO for this action through their Giro555 campaign.

Each member of SHO that spends more than EURO 250.000 per emergency operation has the obligation to conduct an external evaluation for its own operation[1]. The evaluation reports are used to verify the impact, effectiveness and general quality of the members' operations. The purpose of this meta-evaluation is to assess the quality and validate the conclusions of the individual evaluation reports. The quality of the implemented projects of SHO member organisations is not assessed by the meta-evaluation, as this should already have been accounted for in the individual evaluation reports.

The organisations participating in this action were CARE Nederland, Cordaid mensen in nood, ICCO en Kerk in Actie, Nederlandse Rode Kruis, Oxfam Novib, Save the Children, Stichting Vluchteling, Terre des Hommes, UNICEF Nederland, World Vision. Out of these, Save the Children, Terre des Hommes and World Vision did not have to be evaluated, since they spent less than Euro 250,000 in the 2013 Syria action. Total funds received on Giro555, the SHO's bank account, were more than 5 million Euro. The Terms of Reference (TOR) for this meta-evaluation has been included in Annex 1.

2) General observations

A quality assessment grid (see Annex 2 for the full grid) was used to assess the quality of the evaluation report, in order to standardise the analysis. Since the approaches of the participating organisations differed considerably, it remained difficult to maintain a standardised approach. Care, Cordaid and Oxfam, had used the money for a project based operation with implementing partners, and had reported against the achievement of this project, which made assessing of the quality of the evaluation under consideration easier. ICCO, het Nederlandse Rode Kruis and Stichting Vluchteling had incorporated the funds into a larger programme with others or as part of an alliance, and UNICEF had included it into their own (regional) programme, leading to the quality assessment focusing on the evaluation of the overarching program, or even of the alliance or network that they were part of, which may have quite different and broader strategic demands.

The quality of the TORs appeared to strongly influence the quality of the evaluation reports. The quality of the TOR is the basis for the quality of the evaluation, and

[1] Due to the security situation in Syria, the Board has decided that a desk evaluation is the minimum requirement for the participating organisations.

therefore, the assessment has already started at that stage, though evaluating the TOR was not required in the TOR of this meta-evaluation. In the case of more strategic evaluations, the requirements and evaluation questions were also fairly different from project evaluations, leading to a different type of report.

When analysing question 1 of the assessment grid (Annex 2): “The report adequately meets the information needs expressed in the terms of reference in a way that reflects the stated priorities” it immediately appeared that priorities were far from similarly phrased for each of the participants.

Therefore, together with the evaluation reports, the TORs of each evaluation have also been submitted, and the quality of the TOR and its linkage to the evaluation report has been made part of this assessment.

Finally, the budget and number of days available for the evaluation have probably influenced the quality. Standards on the budget (as a percentage or an absolute amount) were not provided and budget amounts available were not part of the information package, so this aspect could not be taken into account.

This report starts with providing an overview of assessments per SHO participant, and finishes with a comparative section, conclusions and recommendations.

3) Findings of the evaluation analysis per participating SHO member

A short version of the full grid has been used to reflect the results of the quality assessment. A score of 1 means unacceptable, 2 is weak, 3 is good and 4 is excellent. The overall rating is presented as a numerical score; though often close to the average, it is not always the exact arithmetic average of the scores, but an overall rating of the quality of the report.

a. Cordaid

Cordaid implemented part of the activities in Syria (Euro 400.000 or 675.000, implemented by Jesuit Refugee Services (JRS)) and part in Lebanon (Euro 675.000 or Euro 475.000, implemented by Caritas’ partner Caritas Lebanon Migrant Centre (CLMC)). The TOR and the evaluation report mention different amounts for each of the partners. The evaluation in Syria is a desk study, conducted in Lebanon. The evaluation of the Lebanon component contains also field visits.

The main objective of the evaluation is “to provide insight in the appropriateness, effectiveness and timeliness of the intervention; also, to evaluate the quality of the process for this intervention at field level. The evaluation is supposed to provide conclusions and recommendations to SHO, Cordaid and local partners. SHO is clearly mentioned in relation to the action.

Five key questions are phrased, linked to the objective, namely on challenges and dilemmas, on involving program beneficiaries, relevance in relation to strategy development and humanitarian needs, effectiveness and timeliness, and appropriateness.

The TOR is very concise and not well structured. Though some of the DAC criteria[2] were taken into account, others were not; the criteria are not brought up as a

[2] The *DAC Principles for the Evaluation of Development Assistance*, OECD (1991), Glossary of Terms Used in Evaluation, in 'Methods and Procedures in Aid Evaluation', OECD (1986), and the *Glossary of Evaluation and Results Based Management (RBM) Terms*, OECD (2000).

coherent group of criteria to be assessed, but are scattered throughout the text. Where it is understandable that in the light of the output focused type of humanitarian activities and time frame, not all criteria are equally useful, no further explanation was provided. Partnership as criterion for instance was only brought up in view of implementation and downward accountability, but coordination and cooperation were not addressed. The consultant has to some extent made up for the limitations by designing an extensive questionnaire. Roles of Cordaid's coordinator and the reference group established for the purpose of the evaluation have been outlined, though in a very short manner.

18 days were available, 4 for inception, 11 for travel and field visit, 2 for writing the draft report and 1 for finalisation of the report, which is rather little.

1. Though the evaluation is supposed to assess the interventions against relevance, appropriateness, effectiveness and timeliness in response to the humanitarian needs of the Syrians in Syria and Lebanon, the focus of the evaluation was on outputs and project management. For the Syria based intervention, the response to the question about the strategies to reach the most needy groups has been elaborated, and in the case of Lebanon considerable attention was paid to downward accountability. The part of the report concerning the intervention in Syria is very short.

2. The method is described reasonably clearly, but is rather limited in quality. The evaluation team was supposed to answer the questions based on a two-day desk study – which enabled only interviews with the JRS staff from Syria, and just from two out of three locations. Two days is very little to collect sufficient data, especially when the funds under consideration are a poorly demarcated part of an overall programme. Moreover, the TOR requires evaluation of the process "especially at field level" which in the case of Syria is very difficult to address, unless a good creative method is developed. The report does briefly touch upon the limitations. The development and use of questionnaires are clearly described. Lack of time is brought up as a limitation as well.

3. The evaluation in Lebanon is conducted with a "co-evaluator" of the implementing partner of Cordaid, who is supposed to be the subject of evaluation, which may shed doubts on the independency of the approach. Other than that, the approach was sufficiently independent, though independency was not emphasized in the text.

4. As a result of the type and length of the desk review, the team was only able to conduct interviews with a few stakeholders; this hampered the possibility of crosschecking and validation; triangulation is not brought up. When even outputs according to the evaluator are "a best guess", the question is how reliable the assessment of more complex outcomes and criteria is. The overall reasoning in the analysis especially related to outputs was sufficiently sound.

5. In most of the cases, the findings look valid, especially since the majority of the observations are output based, but for few exceptions. In Lebanon, though UNHCR is brought up as the cause of a gap in response and risk of duplication, the team has not succeeded to interview UNHCR.

6. The conclusions seem impartial and related to the findings. As conclusions are presented after each and every section though, they are mainly about operational and smaller issues. The conclusions may help the implementers to adapt their implementation, and at best to some extent help Cordaid to coordinate with their partners, but no lessons of wider interest emerge using this approach.

7. The recommendations are simple and sufficiently detailed to enable follow-up. As they follow the conclusion after each section, the same observation may be made as to their linkage to operations and activities. The recommendations are not at political

or strategic level, but mainly about project management and monitoring. On the other hand, the quality of the TOR has not led to focusing on strategic issues.

8. The report is reasonably clear, though repetitions and inconsistencies were found here and there as well as mistakes in grammar and style. There is an executive summary, but it does not contain all of the most important information of the report and does not follow its flow, and thus does not do full justice to the report. Table 1 provides the assessment results.

Table 1:
Assessment Cordaid, report date 18 February 2014

	1	2	3	4
1. Meeting needs		X		
2. Appropriate design		X		
3. Reliable data		X		
4. Sound analysis		X		
5. Valid findings			X	
6. Impartial conclusions			X	
7. Useful recommendations		X		
8. Clear report		X		
Overall assessment		**2.25**		

In general, the evaluation does not follow the usual path of analysis, which would look at context, assessing the activities and achieving conclusions by using the DAC criteria, at least the criteria required by the TOR. Relevance for instance was not linked to the problems brought up in the context and gaps were not identified and analysed. Effectiveness was not discussed based on the questions whether objectives were achieved.

Conclusions and recommendations were reflected throughout the report, instead of at the end based on the overall analysis of the intervention. The fact that the two partners had addressed the planning of the intervention very differently (one did not even have objectives) and that the method was different, made it more difficult for the evaluator to follow the usual evaluation path. The evaluation report would have been interesting as a progress or an end-of-project report, but less so as an evaluation report.

b. ICCO Kerk in Actie

This evaluation concerns the external evaluation of the ACT[3] Appeal SYR 121 in Syria, Lebanon and Jordan, an operation that seems much bigger than ICCO's share under the SHO Syria 2013 Action, though no financial data have been included. International Orthodox Christian Charities (IOCC), Lutheran World Foundation (LWF), Finn Church Aid (FCA) and the Department of Services to Palestinian Refugees/Middle East Council of Churches (DSPR/MECC) are the organisations, which were responsible for the implementation of the action. ICCO has not been involved in the direct implementation (and therefore not mentioned in the report), but is contributing to strategic issues and technical assistance.

Probably as a result of the size of the operation, there was sufficient budget to conduct a good quality evaluation, even though here the evaluators still bring up limited time availability as a constraint.

[3] Action by Churches Together for Development, a global alliance of churches.

The TOR is of very good quality and focuses on management and operational issues as well as more strategic areas and follows the DAC criteria. Lessons learned are part of the requirements and they are broken down into areas of interest. The evaluation questions are clearly phrased, understandable and relevant; they focus on results, strategic issues and DAC criteria and less on outputs.

1. The report fully meets the information needs as per the TOR. The experts had even further broken down the questions of the TOR into sub-questions, which enabled a step-by-step explanation and enhances the reader's understanding of how far the intervention had met the requirements.

2. The design is appropriate for the circumstances. As in most analysed evaluations, the evaluators could not travel to Syria to conduct a field mission, and therefore had to change the methodology. The process and reasons behind this change of approach are clearly described, and the report highlights how the evaluators have tried to make up for this (postponement and finally remote methods). The report also describes the limitations of the change in approach on the quality of the findings (the risk was brought up that some findings will be more indicative than definitive) and on the conclusions.

3. Within the limitations as described under 2, the data seem sufficiently reliable. There are enough data reflected and the analysis is sufficiently well described to draw conclusions, as brought up by the evaluators. Sources are quoted throughout and text boxes have been added, singling out feedback from certain respondents, which adds to the interest value of the report. In one case, a possible bias has been brought up and the underlying reason is described.

4. The analysis is sound. The various sections are broken down per ACT member to give the reader an impression about the background, the type of intervention and the achievements of each member, discussed as well per country; this also enables a certain amount of comparison. The analysis follows the rules of good programming and discusses also issues like advocacy - how well the advocacy has been included and what is still needed from ACT level – and coordination and partnership. The report describes clearly how criteria such as effectiveness are defined and evaluated and the report follows the described process. The evaluators have also looked at the use of a participative approach. External factors are defined and described in each country and at various stages. The work and role of the alliance is discussed separately and is made sufficiently clear.

5. Though the report focuses on results and not on outputs, like the TOR requires, the findings also reflect what has been achieved per activity and per ACT member in each country. The findings are described in sufficient detail and the description is linked to the needs analysis from the context and the appropriateness from the side of the implementer and partners.

6. There is a logical structure going from background of the intervention through roles and strategies, needs analysis, programme results and implementation and coordination, finally leading to conclusions. The conclusions are clearly stated per section as well as in the end of the report and the conclusions are based on the findings as reflected in the report. The TOR required lessons learned for the intervention and for the appeal itself, and it also required an emphasis on lessons learned throughout. The evaluators have honoured that request; additionally, an annex displays lessons learned as identified by the member organisations during the implementation of the action.

7. The report provides practical recommendations to the implementers as well as strategic recommendations to ACT as a group and to its individual members. The recommendations are forward looking and feasible.

8. The report is well written and clear even to a reader without a relevant background; the use of jargon has been confined to the minimum. There is a clear structure and linkage is made between various sections, allowing the reader to follow the analysis of the evaluators. There was no executive summary provided. Some of the acronyms were missed in the list.

Table 2:
Assessment ICCO, report date 21 February 2014

	1	2	3	4
1. Meeting needs				X
2. Appropriate design				X
3. Reliable data				X
4. Sound analysis				X
5. Valid findings				X
6. Impartial conclusions				X
7. Useful recommendations				X
8. Clear report			X	
Overall assessment				3.9

The report is well written and well balanced. The structure of the report is clear and there is a logical flow between the sections. Implicitly, the reader gets information on how to coherently plan a similar programme. Whilst sufficient attention is paid to the findings and observations, analysis and conclusions are also well elaborated and clearly based on the findings. The evaluators appear knowledgeable and experienced. The report, notwithstanding its professional quality, also manages to display a practical approach. Recommendations are provided at various levels, and they seem practical and feasible. The evaluators have evaluated the implementing organisations as well as ACT and its role and possibilities for the future. The report contains no executive summary. It is not clear how many days have been spent, apart from three weeks in the field. As ICCO or SHO are not mentioned in the report, it is unclear how the evaluation links to the 2013 Syria action or how much has been contributed financially by ICCO to this action.

c. Care

Care reports about its own implementation of an intervention with Syrian refugees in Jordan. SHO is clearly mentioned, together with other co-funders; the amount coming from SHO is US$ 167,693, which is 4.5% of the total of almost US$ 4 million.

The TOR contains useful information, but it is quite short and has a number of gaps. The background description is very short (one paragraph) and general. A lot of emphasis is given to CARE's list of outputs in the intervention under evaluation, but a structured overview of the programme is missing. The TOR describes briefly the required methodology, but evaluation questions have not been included.

The main purpose of the evaluation is phrased as to assess the performance of CARE, to develop lessons learned and ensure accountability for the use of the donors' funds. The DAC criteria of effectiveness, efficiency, impact and sustainability are assessed. The latter two may be questioned in the context of a refugee camp. The TOR gives a short explanation that sustainability should be seen in the context of longer-term needs, but more explanation is needed for the deviation from the usual meaning of sustainability. The TOR does not require assessing relevance though, and the absence of this criterion has not been justified.

Prospect respondents have been described. Strangely enough, in the TOR CARE wants the team to exist of an independent Team leader and three CARE staff. Though the TOR stipulates, that the Team Leader's responsibility for the content and the quality of the report would guarantee the independency, this does not come across as a sufficient precaution. There are many other points in the evaluation process where these team members will be able (even without the explicit purpose to do so) to influence the evaluation and findings.

The TOR contains a detailed time frame, though it does not prescribe the number of days, which are foreseen as necessary for this assignment. It is left to the discretion of the prospect Team Leader, who will hand in a proposal. The TOR provides a basic table with selection criteria and their weight.

1. It is difficult to say whether the report responds to the information needs of the TOR, since the TOR does not stipulate the questions and remains vague about information needs. The consultant has made an effort to overcome this gap by developing evaluation questions, one-on-one linked to the DAC criteria. Especially regarding effectiveness and sustainability, the questions are not very relevant. Though in the TOR another set of criteria was proposed, the change is not further explained. The TOR clusters timeliness/effectiveness and sustainability/connectedness, but the evaluation uses effectiveness and sustainability.

2. The consultant has conducted a 10-day field visit, but it is not clear how long the reporting has taken. The concepts are not always clear, and it looks as if the consultant does not fully comprehend the meaning and purpose of the DAC criteria. Under sustainability, the consultant strives to assess the level of participation and the influence of external factors, whilst leaving untouched the question how sustainable the intervention really is. One would question anyhow whether support to refugees in a camp can be easily assessed for its sustainability. The method of data collection has been fairly adequately described.

3. Whilst the TOR described the involvement of 3 CARE staff members as participating in the team, this is not further discussed in the report and it is therefore unclear whether these staff members were part of the team, and if there was reason to suspect bias. The presence of the monitoring and evaluation officer from CARE as translator is brought up but potential consequences of this involvement are not highlighted either. The report discusses another possible bias, where the evaluators did not get access to refugees without others in the room. Also the fact that many events were taking place and hampering the process of evaluation was brought up in the report. Sources of the findings are quoted, but their reliability is not sufficiently assessed.

4. No clear systematic analysis has been used throughout. Even the context analysis lacks structure. It starts with a paragraph about CARE's involvement and history in Jordan, followed by a one-page context description, almost entirely copied from the TOR. The next part already reflects findings from the field mission. Under the "purpose of the evaluation" the report describes some circumstance of the action, before reflecting the purpose as stipulated in the TOR.

5. All findings relate to activities and outputs. The description reads like a word-by-word story of someone, who visited a refugee camp and talked to the refugees, without going beyond that stage. All details of activity are reflected, but without linking them to the analysis.

6. It is not always clear how the consultant arrived at the conclusions. As for effectiveness, objectives and outcomes of the intervention have not been included or clearly phrased in the report and it is therefore difficult to assess how far

effectiveness has been achieved. Furthermore, the writer continuously confuses the meaning of effectiveness in the context of an evaluation with effectiveness related to effectively conducting an activity. The meaning of efficiency has been better understood.

Conclusions are brought up after each group of activities, and they are strictly linked to the replies on the evaluation questions, which were brought up as if part of the DAC criteria. As a result, the pathway of arriving from findings to conclusions becomes entirely mixed up with the assessment of the intervention against DAC criteria. The final assessment of the DAC criteria is presented at the end of the evaluation report as a kind of arithmetic sum of the evaluation question replies per activity sector.

In a number of cases, conclusions are very weak. The report describes that “with 3,000 male and female unique participants, assuming the somewhat unlikely zero overlap between psychosocial and recreational activities’ participants”, but subsequently just continues the analysis based on 3,000 unique participants – which is not only somewhat, but most probably highly unlikely.

Conclusions are often also obvious and not sufficiently specific, such as where case management and referral was found very relevant “especially in the case of new arrivals”. One hardly needs to visit the refugee camp to come up with such a conclusion. Moreover, such a conclusion demonstrates, that the writer may not have understood the full meaning of DAC criteria.

7. The recommendations are also very much activity based and may be only useful for CARE and its partner UNHCR at field level. The advice is sometimes based on commonplaces, such as in “continuous and targeted support has a high potential of impact and sustainability” and “any effort should be based on a collaborative and consultative process” without actually elaborating, what that means for the current (or a similar or future) intervention. The level of detail in the recommendations, which are clearly tied to the current intervention, renders them almost unsuitable to other development actors. In a last section, called “Final considerations – strategic discussion going forward” the reader is presented with some thoughts, which could have been worked out to useful recommendations.

8. The report is very short. Of the total of 25 pages, 3 pages are taken up by the executive summary and one page by a table of achievements against programme outputs. Whilst these have been discussed already at length in the preceding sections, also usually such a table would be captured in an annex. Moreover, there are many blank spaces and repetitive tables, and the same conclusions are repeated throughout all sections. Though the report provides information on whether the activities have been conducted, there is no clear structure and no linkage between the various sections. The TOR did not require the consultant to submit a proposed layout of the report before starting the evaluation, which might have at least partially prevented these quality and content issues.

Table 3:
Assessment CARE, report date 10 June 2015

	1	2	3	4
1. Meeting needs		X		
2. Appropriate design		X		
3. Reliable data		X		
4. Sound analysis	X			
5. Valid findings		X		
6. Impartial conclusions	X			
7. Useful		X		

recommendations				
8. Clear report		X		
Overall assessment		**2**		

The quality of the report is rather poor and information is limited to output level. The consultant does not demonstrate a large experience, especially with regard to strategic issues, and his or her writing skills need improvement. The sentences are all very short with many style and grammar errors. The only conclusion one can draw after having read the report is that CARE has conducted more or less the activities that have been planned. Whether the intervention was appropriate or suitable, and how coordination with the main other funder (UNHCR) and other stakeholders took place, is not addressed. The budget of the intervention is almost US$ 4 million, and one might have expected a better quality evaluation of such an intervention.

d. Stichting Vluchteling

Stichting Vluchteling (SV) implemented their intervention in Iraq and Syria, but the evaluation report focuses on Syria only. SHO is mentioned a couple of times in the text of the TOR. SV works for this intervention under a two-year framework with the International Rescue Committee (IRC); SV provides funding to IRC, and the Emergency Preparedness and Response Unit (EPRU) of IRC is responsible for the implementation in Syria.

The TOR is fairly clear. There is no section containing background on the context, but the cooperation between IRC and SV is explained. The purpose and specific objectives of the evaluation are given. A number of interventions under the SV/IRC framework have been evaluated, and the report under evaluation focuses on Syria.

The TOR provides the evaluators with the eight DAC criteria to be assessed, and breaks them down into practical sub-questions. The methodology is also proposed, but the evaluators had to deviate from that methodology. Since field visits to Syria were not possible. The TOR also presents a work plan to be followed and a table displaying stakeholder involvement per type of stakeholder. Evaluation outputs and time frame are stipulated.

The TOR is completely finalised and the consultants will be identified and selected under the assumption that they will work with this TOR.

1. The report replies to all the questions, that have been raised in the TOR under each criterion, in a sufficiently elaborate manner; these questions as such were found relevant and exhaustive, leading to the report covering all information needs. The report is very short, but the project in Syria is also rather small (Euro 250,000, SHO funded) and for the major part consists of hardware procurement. Moreover, the project duration is only 3 months, and therefore no progress reports have been handed in. The evaluation report follows the criteria and sub-questions as provided by the TOR. The TOR has also requested coverage and sustainability to be assessed, but this was replaced by connectedness and proportionality. The report justifies this by referring to the emergency response nature of the project, which makes these criteria more suitable.

2. The design was appropriate and the method used the only one possible in the current context of Syria, namely a desk review. The evaluators are not very creative or innovative in their approach, but they do explain clearly the underlying reasons. The evaluators feel that this does not hamper the quality of the report and conclusions. Moreover, they bring up the safety of people involved, which is the most important priority.

3. The data seem reliable, even if they have only been derived from a limited number of documents; it is quite an accomplishment how the evaluators managed to come to

such a clear analysis based on such limited resources. The evaluators do not specifically highlight consequences for conclusions and recommendations and they do not address whether the reports are reliable, though they do compare report data with surveys and need assessments.

4. The analysis is sound, and the evaluators have managed to draw in information from every source they could identify, including identification reports and needs analysis. Based on these, their analysis shows that SV and IRC may have been a bit ambitious on one hand, but that they have still overlooked a number of issues. The report even quickly assesses the expenditure, which most of the evaluation reports have left out. The evaluators furthermore highlight some inconsistencies in statistics and ambiguity in accountability. Lobby and advocacy have been assessed as well, and thus a fairly comprehensive overview has been provided.

5. The findings look valid as far as one can expect from a desk study, and the evaluators have tried to triangulate the findings by interviews with former staff. The report contains no context analysis, but parts of context characteristics are scattered throughout the report. Both the TOR and the report assume, that the readers are aware of the context, which may not be necessarily the case, depending on the audience.

6. The conclusions look impartial and are derived using a clear line of reasoning. The conclusions are also based on comparison of the achievements against the goal and objectives of the intervention; goal, objectives, result and project characteristics have been described before moving on to what has been achieved. The evaluators assess the content, quality and feasibility of the original proposal, and bring up the limitations. Following this, they assess the achievements, highlighting what has and what has not been achieved and compare this with the original plan and their observations. Though the intervention is straightforward, the evaluators succeed in deriving a few interesting conclusions.

7. There are only three short recommendations, but they are based on the findings and conclusion and thus relevant to the intervention and SV/IRC. They are clearly formulated and sufficiently strategically and generally phrased, to be of use for other development actors who work in similar conditions or plan to design such an intervention.

8. The report is well written, it is informative, readable and interesting, and it has a clear logical structure. It contains an executive summary, which does not leave out any important information. The report contains a few typing errors, but not to the extent that it disturbs the reader.

Table 4:
Assessment Stichting Vluchteling, report date December 2014

	1	2	3	4
1. Meeting needs				X
2. Appropriate design				X
3. Reliable data			X	
4. Sound analysis				X
5. Valid findings			X	
6. Impartial conclusions				X
7. Useful recommendations				X
8. Clear report				X
Overall assessment				3.9

This evaluation provides proof that a short report is not necessarily an unsuitable report. Though it contains only 14 pages including the executive summary, all aspects have been covered also outside the area of activities. The report presents the reader with an overview of what has been implemented as well as with an insight into the question, whether funds have been used well. Also, the report manages to make clear to any reader, that IRC and SV have done a fairly good job, though without forgetting to mention the shortcomings.

e. Het Nederlandse Rode Kruis/ IFRC

The TOR is of good quality, probably based on an existing format. The evaluation is a real time evaluation, meant to inform future global emergency response operations and approaches. The duration of the consultancy is 30 days with 15 days in the field.

The TOR is very informative. It starts with a summary, proceeding with the current and historic involvement of the International Federation of Red Cross and Red Crescent Societies (IFRC) in the area and the surrounding countries, the appeals that have been launched and the goal and content of the current appeal.

The TOR explains that real time evaluations are usually done in an early stage, that this was not possible in the current circumstances and how this has influenced the contents and prospect use of the evaluation.

The evaluation is supposed to look at three areas: how well the IFRC secretariat coordinated its response, what were the challenges and opportunities of a multi-country response and how has IFRC dealt with the long-term nature of the crisis. The evaluation is highly strategic in nature but is not focused on the intervention as such and does not link in any way to the part of the intervention under the 2013 Syria action. The only relevant conclusion for SHO would be about the suitability of IFRC/RC as strategic partner.

The TOR provides guiding questions, which are clear, but are not linked to the DAC criteria. The guiding questions are all sub-questions to the three areas of interest, mentioned in the previous paragraph. The roles of the evaluation management team, the team leader and the evaluators are also described in a detailed manner. The evaluation methodology lists desk review, field visits and focus group discussions. In case field visits are impossible, Skype or phone conversations are foreseen; this is mentioned for Iraq, but not for Syria. In the end, Iraq was not included into the evaluation.

The evaluation deliverables, time frame and quality standards have all been included in the TOR.

1. The report highlights the questions asked in the TOR, and provides elaborate answers to these same questions. Judging by the TOR, the information provided by the evaluators was exactly what IFRC was after. There is however not much clarity about how the intervention was conducted, as that was not among the questions of the TOR.

2. The design is appropriate and the method is detailed and well described, including desk review, interviews, an online survey, field visits and feedback and consultations. The team managed to go to Syria, though they only went to the capital and could not interview beneficiaries. The limitations and reasons of that approach are not discussed. It is hard to compare the evaluation with the other evaluations, as it has a totally different approach and does not make use of the DAC criteria. It does not look at the management of interventions either, like most of the other evaluations do.

3. The evaluators have ensured the reliability of data by adhering to the standards, which are clearly outlined in the TOR, and by collecting data through a large number

of interviews and an online survey. Based on these data, they also have been able to share observations on security and how staff members deal with security, on gender and gender based violence and on humanitarian diplomacy.

4. The analysis seems sound and coherent, and follows a logical path throughout the report. The analysis links the intervention and the position of IFRC to the existing situation and needs.

5. The findings describe the scale of the operations and relate this scale to the existing needs, the appeals of IFRC and the use of IFRC standard humanitarian tools, coordination between IFRC and ICRC and Movement Advisory Platform issues. Whilst this is all very interesting, the reader needs a large amount of background knowledge to understand the text and it is difficult to say whether these findings are valid or not. They are however linked to the requirements of the TOR and seem to be based on what was identified with the data collection tools.

6. The conclusion section is well elaborated. There are 10 conclusions and all are based on the findings and background of the report and supported by a sound analysis. There are a handful, about resource mobilisation, scale variability for instance and innovation, which may be of interest to people and organisations outside IFRC's working sphere, but most conclusions are really only useful to IFRC. It is not possible to find any link to the Netherlands Red Cross and its implementation of the SHO Syria funds, since the report addresses the intervention based on the use of pooled funds.

7. The recommendations are sensible, well described and link to the findings and conclusions. The recommendations at country specific level may be interesting to others, but again most recommendations are very strategic and entirely related to IFRC and therefore only useful to IFRC. They are about the Movement Advisory Panel Process, the internal coordination within the secretariat, IFRC strategic planning and resource mobilisation, internal cooperation and issues related to Host National Societies. Little of that is of actual use to other humanitarian actors.

8. The report contains an executive summary, which lays out a short description of the context, all the findings, the constraints and the summary recommendations. The recommendations are very extensive, having a length equal to the remainder of the executive summary. The executive summary is only suitable for readers with a considerable knowledge of the approach and background of IFRC. The report provides a country context analysis, which is clearly written and gives a good insight into the situation and how the humanitarian support of IFRC can be relevant to that. Overall, the report is well written and informative.

Table 5:
Assessment Rode Kruis/IFRC, report date 10 December 2014

	1	2	3	4
1. Meeting needs				X
2. Appropriate design			X	
3. Reliable data				X
4. Sound analysis				X
5. Valid findings			X	
6. Impartial conclusions				X
7. Useful recommendations				X
8. Clear report				X
Overall assessment				3.8

The evaluation report is informative, based on a thorough in-depth research and written by someone with excellent writing skills. Though very strategic and specific to IFRC, to some extent it is still made understandable to others, but most of the findings and recommendations are not useful to the wider humanitarian community.

f. Oxfam Novib

Oxfam Novib has implemented a project for Syrian refugees and Palestinian refugees in Lebanon and gatherings together with the partners Association Nadjeh (AN) and Popular Action for Relief and Development (PARD), meant to improve the wellbeing of refugee and host families.

There is no reference to SHO or to the budget size of the project.

The quality of the TOR is mixed. Whilst there is a reasonable country context description and an elaborate part on the project's activities, result and objectives, the method prescribed is very limited. Though two questions with sub-questions are phrased, the plan still needs to be further developed. The two consultants get a total of six working days to conduct the entire mission, including developing an evaluation plan, have a half-day debriefing meeting and write a 30-page report. Even for the most experienced and knowledgeable consultants, this would be a labour of Hercules and it is therefore no surprise that they did not succeed.

1. The evaluators have responded to some extent to question 1 of the TOR, which was to assess the results of the project and three DAC criteria; the second question about constraints and dilemma's and the participation of beneficiaries in design and management is less well responded to.

The evaluation partly uses DAC criteria to assess the performance, but the criteria differ from the ones used by other evaluations. Effectiveness and timeliness are common, but others have not brought up adequacy, and partnership is usually included in another criterion though it may be an important issue to look at. Under adequacy, the evaluators assess appropriateness and relevance, which then turns out to be the most interesting section, highlighting a number of needs of the target population.

The evaluators have made an effort to break down the questions under the criteria, and at least in that section, a coherent approach can be identified, but after section 2 this gets lost.

2. It was decided that no beneficiaries would be interviewed, but only implementing partners and staff members, since the project had already been finalised one year before. The methodology mentions, that a desk study has been conducted, but the text only refers to interviewees as sources and hardly any reference is made to what was found in context related documents or progress reports.

3. On the one hand, it is very clear where the data come from, as interviewees are cited at length, with their full statements and often even with their full designation; whilst this is questionable in view of source protection, all the reflected statements are positive, so the respondents probably were happy to appear in the reports. The approach gives the report a highly anecdotal character. It is not clear how data were triangulated with documents or other sources; moreover, only implementers were interviewed; the possible bias was brought up in the methods section.

4. There is no systematic analysis and underlying assumptions or external factors hardly have received any attention. There is no evidence that crosschecking of data has taken place. The assessment of the operational management and the conducting of activities are mixed with the assessment of the DAC criteria. Under effectiveness, instead of assessing the achievements against the objectives or even

results, the evaluators again look at achievements at activity level. The evaluators seem to mix concepts of output, results and objectives, though finally under the section of adequacy they do analyse the results. As the logframe apparently contains objectively verifiable indicators, this could have been done more consistently throughout the report.

5. Following the method used, the findings could not really be called evidence based. Though the implementing partner staff confirms that all activities have taken place exactly as planned and outputs have been delivered as per plan and better, no effort has been made to check for supportive evidence.

6. Though the evaluators, under the section about goals of the evaluation, bring up that they will further explore the action in the conclusions and recommendations section, the conclusions seem to have escaped their attention and only recommendations have been included. Unless the conclusions are meant to be implicit part of the findings section, which is difficult to check for the reader, there are no conclusions whatsoever in the report. Without the explicit phrasing of these conclusions, the validity of the recommendations is also at stake.

7. The recommendations do not always succeed in bringing up useful or valuable advice. The report recommends “addressing challenges in reaching the most needy in emergencies”, which is rather generic as a recommendation, and then further explains “as refugees’ dependence on assistance is likely to continue, increase amount and/or frequency of relief assistance”; the two statements do not seem to be logically linked, nor is one a solution for the other. The evaluators recommend involving program beneficiaries in more stages of the intervention, whilst they themselves have failed to do so. As a consequence one cannot derive from the report, that this intervention is what beneficiaries want and need. The last recommendation on partnership is rather generic as well, but may be of some interest at least to other development actors.

8. The report is at best moderately clear. It loses itself in word-by-word statements of people interviewed and as a consequence, the level of detail is much too high. The structure and flow are sub-optimal. The report does not contain an executive summary, a context analysis or an acronym list. It contains many repetitions, including the repeating of the planned results of the project, which could have been included in an annex. The same goes for the entire list of interviewees including their names, which has been included in the main body of the text. As a result, only limited space was left for the ultimate evaluation related text. The writing style is a bit sloppy, with a strange use of capital letters and acronyms are not always explained.

Table 6:
Assessment Oxfam Novib, report date September 2014

	1	2	3	4
1. Meeting needs		X		
2. Appropriate design		X		
3. Reliable data		X		
4. Sound analysis		X		
5. Valid findings	X			
6. Impartial conclusions	X			
7. Useful recommendations		X		
8. Clear report		X		
Overall assessment		**1.75**		

The report reads more like a newspaper article about the intervention than like an evaluation report. It contains many details, which are not always necessary or useful. Though it uses some DAC criteria, it does not analyse these criteria along the lines of identifying gaps to be addressed and how well the organisation has achieved this by the mentioned intervention. The analysis does not go beyond the services delivered, and the main finding is that all services were delivered and all was in order.

g. UNICEF

UNICEF was the last to finalize the evaluation related to the SHO response. UNICEF has implemented its intervention with implementing partners in Syria, Jordan, Lebanon and Turkey. The report was severely delayed, as dissemination had been foreseen from July 2015.

The evaluation is a comprehensive region-wide assessment of UNICEF's overall response to the Syria crisis. With the total funding of the programme under evaluation at US$ 659 million, the SHO funding is a very small part of the funding pool and no reference has been made to it.

The quality of the TOR is good; its follows the regular format of UNICEF and complies with the UNICEF evaluation standards for TORs. The TOR clearly outlines to the evaluators, what is expected from them, for what purpose the information is collected and what should be the extent of detail. The programmatic focus is clear and the questions have been well and extensively elaborated. Even the methodology is fully provided to the evaluation team.

A team of 7-8 consultants has worked on the evaluation for a period of 8 months, but the total number of working days has not been reflected in the TOR.

1.The report is well written and answers all of the questions, raised in the TOR. UNICEF's programmatic response in each of the sectors is brought up, its role and strategy and its engagement with others, all as per requirement of the TOR. The information is structured in such a way, that it is easy to get an overview of the results as well as the constraints, which prohibited those from being achieved.

2. The methodology and approach of the evaluation are adequately elaborated and the task distribution is clear, including that of UNICEF's Evaluation Office. Though the description of method and tools in the report itself is rather concise, an appendix depicts the entire approach at length, illustrated with schematic representations of the approach and analysis. The limitations and constraints are discussed, including the limited access to Syria for the purpose of primary data collection.

3. As far as appears from the report, the evaluators have adhered to the standards of the TOR; they have collected data through a large number of interviews with different respondents in various locations and a desk review, and triangulated these data to avoid inconsistencies. Inconsistencies between UNICEF internal and external data were brought up as a limitation though, which could not be eliminated.

4. The analysis looks sound. Since the programme is so large, covers a large geographic space and multiple interventions and UNICEF tools, it is at times difficult to present the flow of thinking, but taking into account the constraints, the evaluation team has done what was possible.

5. The findings look sufficiently valid. The evaluators have measured what they intended to and what was asked from them. There is sufficient mention of specific sources of information where that is relevant and necessary. A large part of the evaluation was qualitative, but indicators were used throughout, which helped quantifying the findings; moreover, sufficient quantitative data are also reflected in the report.

6. The conclusions are grouped in five sections, and highlight what went well and what was more difficult. At the start of the conclusions chapter an overview is provided, which helps the comprehension of the reader as to what are the most important conclusions and whether UNICEF has been able to achieve its core objectives. The conclusions follow a logical path starting from the context analysis through findings and lessons learned. The structure of the findings is reflected in a similar structure of the conclusions, which are phrased in a clearly understandable and seemingly impartial manner.

7. The recommendations are preceded by an advice of what steps should be undertaken to enable proper implementation and by a concise overview of the recommendations themselves. There are six recommendations, which display a good balance between strategic, management, operational and accountability related recommendations. The recommendations look practical and pragmatic and written in an understandable manner. It is also clear, who should be responsible for implementing them; therefore the recommendations were found very useful.

8. The report is well written and informative. It contains a good quality executive summary, reflecting all of the important findings from the report. It is clear that the report follows UNICEF evaluation quality standards. The findings, conclusions and recommendations are well balanced and well linked.

Table 7:
Assessment UNICEF, report date December 5, 2015

	1	2	3	4
1. Meeting needs				X
2. Appropriate design				X
3. Reliable data			X	
4. Sound analysis				X
5. Valid findings				X
6. Impartial conclusions				X
7. Useful recommendations				X
8. Clear report				X
Overall assessment				3.9

Overall, the quality of the report is good. Knowing that the evaluation covers a broad area geographically as well as subject-wise and has been conducted by a large group of experts, the end product has become cohesive and consistent, and easy to read even for those that do not have a strong background in development. The logical path is strong and the conclusions and recommendations useful for the current and future programmes.

4) Discussion

Considerable variations were found in the quality, contents and approach of the evaluations, in the reports as well as in the TORs. Some TORs are used as a basis for tender or to procure services from consultants, and in that case consultants sometimes are required to come up with a further elaborated methodology, the quality of which then contributes to their selection. Other TORs are already fully developed (to the extent that they provide the full methodology) and approved and do not require any further input from the consultants, except possible comments. In such a mechanism, the evaluators are identified differently, for example through a roster, based on their qualifications and daily fee.

The number of days available and the number of experts involved varied considerably as well. It was not possible to thoroughly analyse this, first of all because some of the TORs did not reflect the number of days available or the number of experts involved at all, whilst others either presented only the days for field visits or the total number of days. Furthermore, the size of the evaluations differed considerably between the organisations, which gave rise to different needs. It did become clear though, that experts at times availed of too little time to write a good quality report.

The meta-evaluation confirms, that the quality of the evaluation reports is very much depending on the quality and content of the TORs. Beyond that, the writing skills, experience and strategic insight of the evaluators obviously also contribute to the quality of the report. Since no further information on the consultants was provided, this statement cannot be further substantiated.

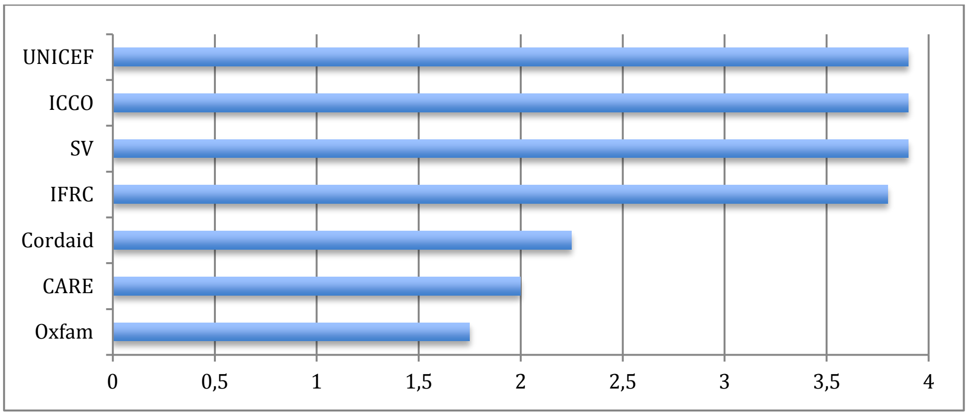
The scoring per category and organisation contained a large variation. Table 8 below gives an overview of the scoring of the organisations per category.

Table 8:
Scores of the meta-evaluation per organisation and assessment category

	Meeting needs	Appropriate design	Reliable data	Sound analysis	Valid findings	Impartial conclusions	Useful recommendations	Clear report
UNICEF	4	4	3	3	4	4	4	4
ICCO	4	4	4	4	4	4	4	3
SV	4	4	3	4	3	4	4	4
IFRC	4	3	4	4	3	4	4	4
Cordaid	2	2	2	2	3	3	4	4
CARE	2	2	2	1	2	1	2	2
Oxfam	2	2	2	2	1	1	2	2

From table 9, it becomes clear that scores were either rather good, or rather poor. The scores ranged from 1.75 to 3.9; the average was 3.07, and 4 of the organisations scored very well (3.8 or above) whilst the remaining 3 scored poorly (2.25 or below).

Table 9:
Overall scores per organisation



In the interventions, where the organisations had been participating in an alliance or network[4], or where the intervention was part of a broader programme[5], a standard

[4] *ICCO, het Nederlandse Rode Kruis and Stichting Vluchteling*
[5] UNICEF

approach to evaluations was used, often with a predefined set of criteria and sub-questions being readily available. The organisations operating on a project base[6] with implementing partners designed a TOR related to the specific evaluation, though it may be assumed that they use existing examples of previous evaluations as a basis. As these organisations probably not always avail of staff with the specific designation to guide evaluations, the quality may turn out less. The size of the available budget may also contribute to this. Most of the evaluations, conducted by alliances, were of good quality, whilst the NGOs, which conducted the evaluation in a more independent manner, produced a somewhat lower quality. A positive deviant was Stichting Vluchteling, which operated as part of a framework, but produced a very good quality report independently, notwithstanding the fact that it was comparatively short.

Working through an alliance may enable individual organisations to enhance their impact and evaluations and TORs of larger networks of organisations were usually of better quality. On the other hand, it was more difficult to distinguish how the member of SHO had participated in the initiative, let alone what its achievement was. From the contents of such reports, it is mostly impossible to make any judgement about the work of the member organisation under the Syria 2013 action. If evaluations were conducted for the work of the SHO member only, though the quality of the report was less, it was at least easier for the reader to understand what had been done with the funds and how well the funds had been spent. Conclusions and recommendations looked more feasible and practical for other SHO members working together on the same action.

The larger evaluations were often entirely or mostly focused at strategic level, whereas the projects evaluations looked more at activity and output level, though the better ones also assessed the achievement of results and objectives. The approaches were not always equal or even similar.

Many larger organisations change their evaluation approach on a more or less continuous basis with the obvious aim of improvement. Currently, there is a trend towards evaluating on the basis of sets of evaluation questions, with less focus on DAC criteria, or sometimes incorporating only a select number of criteria. DAC criteria are not abolished though, but they often are no longer the main or sole basis of the evaluation.

Table 10 below demonstrates that in this meta-evaluation, apart from IFRC, which had a standard approach of its own, four organisations used a minimum of three DAC criteria (relevance, effectiveness, efficiency, impact and sustainability). All organisations except IFRC used 3-5 predefined criteria, though these were not necessarily DAC criteria. All reports assessed effectiveness, but other criteria were less commonly used. Half of the criteria were only used by one organisation, namely adequacy, coherence, connectedness, coverage, impact, partnerships and proportionality. The use of sustainability as a criterion is debatable in the light of the emergency context and the short duration of the interventions; the same can be said for some extent about impact, unless it is defined very clearly. Not all organisations clearly defined the criteria and only a minority justified the use of a specific subset of criteria and non-use of other criteria.

Table 10:
The use of DAC criteria

	CORDAID	ACT/ICCO	CARE	SV	IFRC	Oxfam	UNICEF
Adequacy						X	

[6] Care, Cordaid and Oxfam

Appropriateness/Relevance	X	X	X	X			X
Coherence							X
Connectedness				X			
Coverage							X
Efficiency		X	X	X			X
Effectiveness	X	X	X	X		X	X
Impact			X				
Partnerships						X	
Proportionality				X			
Sustainability		X	X				
Timeliness	X	X				X	

Though participants of the SHO 2013 Syria action were required to spend a predefined percentage of the funds in certain geographic areas, this information could not be found in the evaluation report, since often the programmes were conducted by a group of partners, which implemented the action in different geographic locations. The geographic locations as per the reports are reflected in the table below.

Table 11:
The geographic location of the interventions

	CORDAID	ACT/ICCO	CARE	SV	IFRC	Oxfam	UNICEF
Syria	X	X		X	X		X
Jordan		X	X		X		X
Lebanon	X	X			X	X	X
Others				X	X		X

CARE and Oxfam did not work inside Syria at all, whilst SV worked inside Syria only. All evaluators brought up the difficulty or impossibility to conduct field visits inside Syria for data collection. They have different solutions though; some just carry out a few interviews with a subgroup of people, whereas others have made an effort to also interview beneficiaries. Some evaluators justify the use of different methods, whereas others don't. IFRC has visited only Damascus to conduct interviews.

5) Conclusion

The quality of the evaluation reports, methods and approaches vary to a great extent. Not all evaluations have achieved sufficient quality of approach and reporting. The aggregation of the singular evaluations therefore does not provide sufficient evidence to estimate accountability related to SHO funds. Still, most of the reports are interesting to read, and the findings, conclusions and recommendations are very useful for SHO, since they give an indication of the quality of the work of the organisations, and provide learning opportunities for other SHO members, even if the reports have been produced for a different reason and audience.

The influence of the protracted nature of the emergency and the ongoing difficult security situation should be remarked as severely constraining the evaluations. Not only was the collection of data hampered, but also in a number of cases the operations are still ongoing, even if SHO funds have been spent.

SHO stipulates that “each member that spends more than Euro 250.000 per emergency operation, has the obligation to conduct an external evaluation for its own operation” but many of these evaluations have no link whatsoever to the own operation of the organisation. If SHO wants evidence of the money being well spent and the action being well programmed, another form of evaluation would be required. Moreover, if no reference to SHO or the action under consideration is made in such an evaluation report, it is difficult for any reader, including donors, to understand that link.

6) Recommendations

1. Define the objectives of the evaluation
SHO should clearly define the goals of the evaluation, which it requires from its members. If the funds are transferred to a network and the subsequent evaluation does not describe any relation to the SHO action, the evaluation may be useful for learning, but its value regarding accountability for SHO funds is more limited, independent of its quality. Clearer guidelines would be useful to ensure that such an evaluation provides information on the mentioned funds and enables using the results in communication to external parties.

2. SHO to conduct the evaluation
Even if goals and objectives of the evaluations of members were to be defined more clearly by SHO as under recommendation 1, having a number of participants in a campaign evaluate their campaigns whilst others don’t, may still lead to the evaluation results and coverage being of a fragmented nature. Ideally, a percentage of the total budget would be dedicated to conducting an evaluation, to be carried out by SHO. SHO would draw up a TOR, to be peer reviewed by the participating organisations in the campaign (including those with a relatively small budget). All organisations should provide the necessary information and documentation preceding the evaluation. SHO would conduct or commission the evaluation, which would be finally peer reviewed just like the TOR. Such an evaluation would not only provide a more structured overview of the achievements and accountability for the SHO funds, it would also be more user friendly for future fundraising.

3. Ensure a minimum standard for evaluation methods and approach
SHO should advise and/or require that TORs be shared before commencement of an evaluation for peer review among the relevant members of SHO, to ensure its quality and allow submitting comments and suggestions for consideration. Since the actions will bound to be conducted in an emergency context, SHO should require that member organisations come up at an early stage with a solution for data collection in areas that are difficult to access. It may be easier for the organisations to develop the approach for the evaluation at an earlier point in time, when they are still conducting their intervention in the field.

Note about the author by UNICEF Netherlands:
The name and CV of the consultant who conducted the meta-evaluation is available upon request through UNICEF Netherlands or SHO Bureau. Among the consultants that applied and met the minimum requirements as per the TOR in Annex 1, the consultant with the lowest daily rate was selected.

Annex 1: Terms of Reference

Terms of Reference

Meta-evaluation of the 2013 Syria action

Background

On 15 March 2011, large-scale protests started in Syria. They echoed the protests in North African countries, which had become known as the Arab Spring. A significant part of the Syrian people revolted and demanded the resignation of the incumbent president, Mr. Assad, and his ruling party. The regime violently suppressed the protests. Within months the situation escalated into a civil war, which has entered its fourth year in March 2015. The conflict, in which more and more parties became involved, has led to an ever growing humanitarian disaster.

At the start of the SHO campaign, two years after the first protests, the United Nations reported that 4 million Syrians needed assistance. During the campaign this number increased to 5.5 million. By the end of 2014, approximately 12.2 million people were in urgent need of humanitarian assistance; amongst them were 5.6 million children.

THE SHO campaign

The Samenwerkende Hulp Organisaties (SHO) is an alliance of ten Dutch aid organisations. In response to major humanitarian disasters (both natural disasters and conflicts) the SHO raises funds in order to provide humanitarian assistance to the victims of the disaster. In addition, the SHO cooperates with media partners such as television, radio, newspapers and various online media to inform the general public about the particular disaster. The aid agencies of the SHO are jointly accountable for the expenditure of the funds, through their own respective mechanisms. This report is an example of joint accountability.

SHO participants, their local partners and/or their umbrella organisations have been working in Syria and/or in neighbouring countries since the start of the conflict. On March 25, 2013 the SHO launched a national campaign in order to scale-up the individual organisations' relief operations. Total funds received on Giro555, the SHO's bank account, were more than 5 million Euro.

Organisations that participated in the SHO action

CARE Nederland, Cordaid mensen in nood, ICCO en Kerk in Actie, Nederlandse Rode Kruis, Oxfam Novib, Save the Children, Stichting Vluchteling, Terre des Hommes, UNICEF Nederland, World Vision.

Main reasons and purposes of the evaluation

The SHO member organisations jointly raise funds for emergency operations. They spend the raised funds separately, through their respective partners, networks and alliances, with coordination on the ground through the cluster system. The results and expenditure of the emergency operations, funded by SHO, are then jointly reported back to the public.

Since the SHO members do not operate as one organisation during the emergency operation, they need a mechanism to ensure that the individual members perform up to the high standards that the SHO has set for itself. Each member that spends more than EURO 250.000 per emergency operation, has the obligation to conduct an external audit for its own operation[7]. The evaluation reports are used to verify the impact, effectiveness and general quality of the members' operations. The SHO conducts this meta-evaluation to assess the quality and validate the conclusions of the individual evaluation reports.

The meta-evaluation is thus focused on the quality of the evaluation reports. The quality of the implemented projects of SHO member organisations is not assessed by the meta-evaluation, as this has already been accounted for in the individual evaluation reports.

SHO follows the ALNAP[8] definition for meta-evaluations, adapted from Lipsey (2000): systematic synthesis of evaluations providing the information resources for a continuous improvement of evaluation practice.'

The meta-evaluation report will be shared with the SHO Board of directors to monitor the internal adherence to quality standards. It will also be used by individual members as input to improve evaluation practices.

The meta-evaluation takes into account the agreements made with regard to evaluations as stated in the Organisation Protocol ("Organisatie reglement SHO, October 2013", chapters 1.14 – 1.19). This refers to:

- The OECD DAC criteria for disaster response evaluations, in accordance with the ALNAP Quality Proforma;
- The Sphere minimum standards in Humanitarian Response during the intervention;
- The Code of Conduct for The International Red Cross and Red Crescent Movement and NGOs in Disaster Response Programmes during the intervention.

Methodology and deliverables

The quality check of the evaluation reports will be carried out through a desk study of the reports of external evaluations and reviews. Annex 1 provides a quality assessment grid that will be used to assess the quality of these evaluations. This quality assessment grid includes elements such as meeting of objectives indicated in the ToR, quality of the analysis and reliability of the data. The consultant will provide the rationale for scores given and indicate strengths and weaknesses. The quality assessment grid will be used for all project evaluation reports shared. Based on the outcomes a synthesis is written with specific attention for each component in the assessment grid.

[7] Due to the security situation in Syria, the Board has decided that a desk evaluation is the minimum requirement for the participating organisations.

[8] ALNAP is the Active Learning Network for Accountability and Performance in Humanitarian Action and was established in 1997. ALNAP is a collective response by the humanitarian sector, dedicated to improving humanitarian performance through increased learning and accountability.

Based on this desk study a draft report will be developed that will be discussed during a meeting with SHO members. This draft report, no longer than 20 pages (including tables and annexes), should consist of:

I. An analysis on the quality of the evaluation reports based on the assessment grid mentioned under objective 1;
II. A section with the conclusions, and lessons learned based on the above analysis;

The meta-evaluation report should be written in **English**

During a meeting with the SHO Board of directors, the consultant will present the outcomes of the research.

After being selected, the consultant will develop a detailed work plan at the beginning of the process. All materials collected in the undertaking of the evaluation process should be lodged with the SHO Secretariat prior to termination of the contract.

Responsibilities participating SHO organisations

The process of developing and finalising the meta-evaluation is steered by representatives of the SHO quality group. UNICEF Nederland is responsible for the overall coordination of the meta-evaluation, supported by Terre des Hommes, through a Reference Group. This Reference Group will:

- Select the evaluator(s) and facilitate the communication with the evaluator(s);
- Facilitate the administrative steps around this evaluation in coordination with SHO back office (contract, payment etc.);
- Support the consultant in assembling the relevant documentation (i.e. evaluation reports and reviews).

The members of the SHO Quality Group will

- Provide relevant information within their own respective organisations and send it to the evaluator(s);
- Provide input into draft report;
- Agree on the final report.

Participating SHO agencies are required to submit the following materials as input for the meta-evaluation (preferably both in hard copy and electronic format to the SHO secretariat)

- Reports of external and internal evaluations and reviews

Expected workload

Activity	*Number of days*
Preparation and reading (basic documents)	1
Desk study (analysis different evaluations)	7
Writing meta-evaluation report	2
Discussing Report with SHO members and presentation to board, final adjustments	1-2
Total working days	**11-12 days**

Timeline

Activity	*Date*
Deadline for application	August 25, 2015

First meeting Reference Group / Consultant	Week of August 31, 2015
Draft meta-evaluation report	September 30, 2015
Adjustments after quality workgroup has read the meta-evaluation*	By October 15, 2015
Finalisation of the process **	October 27, 2015

*Between September 30 and October 15 the SHO quality workgroup will read the meta-evaluation and make comments. Adjustments have to be made before October 27, 2015.
** The meta-evaluation will be presented during the Board meeting on 27 October 2015(afterwards some small adjustments might have to be made).

Minimum requirements

The evaluator should have relevant skills and a proven background in humanitarian emergency project implementation and/or evaluations. No specific sector expertise is required. The selection criteria are:

- Experience in humanitarian aid
- Knowledge of the context of the Syria conflict
- Fluent in both Dutch and English
- Availability

The consultant should not have evaluated one or more of the individual evaluation reports of the participating SHO member organisations.

Applications

Applications can be sent to **mroelfsema@unicef.nl** by **August 25 2015 (17:00)** the latest. The application should contain a short motivation (max 1 page), past experience (max 1 page), a Curriculum Vitae and the consultants' daily rate. Only short-listed candidates will be contacted.

Annex 2: Full quality assessment grid

Criterion	Unacceptable	Weak	Good	Excellent
1. Meeting needs The report adequately meets the information needs expressed in the terms of reference in a way that reflects the stated priorities. The demands, which were made during the evaluation process are mentioned, and satisfied when possible.				
2. Appropriate design Key concepts and criteria are precisely defined. The method is described clearly. It is adequate for addressing the questions. Methodological limitations are explained, as well as their consequences on the strength of conclusions, and on the substance of recommendations.				
3. Reliable data Data are sufficiently reliable with respect to the conclusions that are derived from them. Data collection tools have been applied in accordance to standards. Sources are quoted and their reliability is assessed. Potential biases are discussed.				
4. Sound analysis Data are cross-checked, interpreted and analysed systematically and appropriately. Underlying assumptions are clarified. The main external factors are identified and their influence taken into account.				
5. Valid findings The findings are based on evidence through a clear chain of reasoning. The limitations to validity are clearly stated.				
6. Impartial conclusions The conclusions are based on explicit criteria and benchmarks. They are free of personal and partisan considerations. Points of disagreement are reported truthfully. Lessons of wider interest are identified.				
7. Useful recommendations Recommendations stem from conclusions. They are applicable and detailed enough to be implemented by the addressees. The level of recommendations (political, strategic, managerial, etc.) reflects that of the questions.				
8. Clear report The style of the report is interesting for and accessible to the intended users. A short summary stresses the main findings, conclusions, lessons and recommendations in a balanced and impartial way.				
Overall assessment Taking into account the contextual constraints on the evaluation, the report satisfies the above criteria				

Annex 3: Acronyms

AN	Association Nadjeh
CLMC	Caritas Lebanon Migrant Centre
DAC	Development Assistance Committee
DSPR	Department of Services to Palestinian Refugees
EPRU	Emergency Preparedness and Response Unit
FCA	Finn Church Aid
GBV	Gender Based Violence
ICRC	International Committee of the Red Cross
IFRC	International Federation of Red Cross and Red Crescent Society
IOCC	International Orthodox Christian Charities
IRC	International Rescue Committee
JRS	Jesuit Refugee Services
LWF	Lutheran World Foundation
MECC	Middle East Council of Churches
SHO	Samenwerkende Hulp Organisaties
PARD	Popular Action for Relief and Development
SV	Stichting Vluchteling
TOR	Terms of Reference