



UNEG
United Nations Evaluation Group

Evaluation in the SDG era: lessons, challenges and opportunities for UNEG

18 April 2016

This is the final draft report of a review commissioned by the SO3-SDG Working Group, for presentation and discussion at the UN Evaluation Week in Geneva, April 2016.

The study aims at providing evidence and analysis on the key issues, major strategic implications and urgent priorities facing UNEG and its membership in the context of the Agenda 2030. It is intended to be one of the building blocks for the development and implementation of a UN evaluation strategy 'fit' for the SDGs and Agenda 2030

The findings of the report will be presented at the EPE.

The suggestions on the way forward for UNEG's consideration will be the main focus of the AGM presentation.

Table of Contents

TABLE OF CONTENTS	II
ACRONYMS	IV
1 BACKGROUND.....	1
2 PURPOSE, APPROACH AND METHODOLOGY OF THE REVIEW	3
2.1 Purpose and scope.....	3
2.2 The analytical framework	4
2.3 Methodology.....	5
2.4 Limitations	6
3 THE CONCEPT OF EVALUABILITY AND THE SDGS.....	7
3.1 Definitions and principles	8
3.2 The use of evaluability assessments among UNEG members	9
3.3 The debate on the evaluability of the SDGs	13
4 LESSONS LEARNED FROM PAST EVALUATION EXPERIENCES	16
4.1 The selection of relevant evaluations.....	16
4.2 Findings about evaluation approaches, management and methods.....	18
4.3 Findings about monitoring and evaluation systems	21
5 AGENDA 2030 IMPLICATIONS FOR UNEG MEMBERS	24
5.1 National ownership of Agenda 2030 and its review	25
National ownership and the United Nations Sustainable Development Frameworks.....	25
Country-led reviews.....	28
5.2 National evaluation capacity development	29
5.3 Leave no one behind	34
5.4 Paramount focus on human rights, gender equality and sustainable development	37
5.5 Interlinkages across SDGs and targets.....	39
5.6 Universality of the SDGs	43
5.7 Partnerships.....	45

6	IMPLICATIONS OF THE AGENDA 2030 FOLLOW-UP AND REVIEW MECHANISM	49
6.1	Inter-governmental processes of relevance to the UN evaluation system	50
6.2	The regional dimension.....	51
6.3	The global indicator framework.....	52
7	THE WAY FORWARD	54
7.1	On the evaluability of the SDGs and a UNEG strategy	55
7.2	On country-level evaluations and country-led reviews.....	60
7.3	On National Evaluation Capacity Development	61
7.4	On evaluation criteria, methods and approaches	63
7.5	On Evaluability assessments	66
7.6	On Lessons learned from past evaluations.....	68
7.7	On Partnerships.....	69

Annexes

- Annex 1. Terms of Reference
- Annex 2. Bibliography
- Annex 3. List of people interviewed
- Annex 4. Evaluability Assessments by UNEG and its Members
- Annex 5. Some recommendations with relevance for UN organizations, from selected evaluations
- Annex 6. Lessons and recommendations with relevance for the UN monitoring and evaluation system, from selected evaluations
- Annex 7. Agenda 2030, Paragraph 74
- Annex 8. The Bangkok Declaration
- Annex 9. The Global Indicator Framework, March 2016

Acronyms

DaO	Delivering as One United Nations
DfID	Department for International Development, United Kingdom
EA/s	Evaluability Assessment/s
ESCAP	United Nation Economic and Social Commission for Asia and the Pacific
FAO	Food and Agriculture Organization
HI	High Income
HLCP	High-level Committee on Programmes
HLPF	High-level Political Forum on Sustainable development
IAEG-SDGs	Inter-Agency and Expert Group on Sustainable Development Goal Indicators
IAHE/s	Inter-Agency Humanitarian Evaluation/s
IDEAS	International Development Evaluation Association
IFAD	International Fund for Agricultural Development
IFIs	International Financial Institutions
ILO	International Labour Organization
IOCE	International Organization for Cooperation in Evaluation
IR	Inception Report
ISWE	Independent System-Wide Evaluation mechanism
JIU	Joint Inspection Unit of the United Nations
LI	Low Income
LMI	Lower-middle Income
MDGs	Millennium Development Goal/s
M&E	Monitoring and Evaluation
MS	Member State/s of the United Nations
N/ECD	National/Evaluation Capacity Development
NES	National Evaluation Systems
NSOs	National Statistics Organizations
OIOS	Office of Internal Oversight Services
RBA	Rome-Based Agencies
RBM	Results-based management
RC/s	Resident Coordinator/s of the UN Country Team
RDT	Regional Directors Team
RFSD	Regional Forum/a on Sustainable Development
SCP	Sustainable Consumption and Production
SD	Sustainable Development
SG	Secretary General of the United Nations
SO3/SDG-WG	UNEG Strategic Objective3 SDG-Working Group
QCPR	Quadrennial Comprehensive Policy Review
ToC/s	Theory/s of Change
ToR	Terms of Reference
UN	United Nations
UNCDF	United Nations Capital Development Fund
UNCT/s	United Nations Country Team/s
UNCTAD	United Nations Conference on Trade and Development
UN/DESA	Department of Economic and Social Affairs of the UN Secretariat
UNDAF	United Nations Development Assistance Framework
UN/DOCO	UN Development Operations Coordination Office
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNFPA	United Nations Population Fund
UNEG	United Nations Evaluation Group

UNGA	United Nations General Assembly
UN-HABITAT	United Nations Human Settlements Programme
UNHCR	United Nations Refugee Agency
UNICEF	United Nations Children's Fund
UNIDO	United Nations Industrial Development Organization
UNIFEM	United Nations Development Fund for Women
UN/OCHA	UN Office for the Coordination of Humanitarian Affairs
UNODC	United Nations Office on Drugs and Crime
UNRCs	UN Regional Commission/s
UNSDF	UN Sustainable Development Frameworks
UNV	United Nations Volunteer
UN-Women	United Nations Entity for Gender Equality and the Empowerment of Women
UMI	Upper-Middle Income
VOPE/s	Voluntary Organization/s for Professional Evaluation
WFP	World Food Programme

Executive Summary

The Executive Summary will be included in the final version of the report

DRAFT

1 Background

1. In September 2015, the United Nations (UN) Member States unanimously endorsed the adoption of the document ‘Transforming our World: The 2030 Agenda for Sustainable Development’, or Agenda 2030, “*an ambitious framework for human, economic, social and sustainable development spanning 17 goals and 169 targets covering a complex multiplicity of actors, sectors and themes*”.¹ At their core, the 17 Sustainable Development Goals (SDGs) ‘*seek to realize the human rights of all and to achieve gender equality and the empowerment of all women and girls. They are integrated and indivisible and balance the three dimensions of sustainable development: the economic, social and environmental*’.²

2. Every UN Member State is called upon to implement the SDGs, with each being responsible ‘*for its own economic and social development*’³ and for deciding how to progress towards the SDGs and their targets. The Agenda also invites all Member States and stakeholders to participate in a Global Partnership for Sustainable Development, a mechanism of solidarity, wherein also the UN system stands along with other multilateral organizations, in support of the UN Member States in their efforts towards the SDGs at national, regional and global level.

3. Agenda 2030 also commits the Member States to ‘*systematic follow-up and review*’ which will be ‘*voluntary and country-led*’ and based on country-level data, issued by National Statistics Organizations (NSOs) that will track progress towards the SDGs against sets of national indicators. The indicators will be selected by national governments from within the Global Indicator Framework that was developed by the UN Statistical Commission and the related UN Inter-Agency and Expert Group (IAEG-SDGs) and endorsed in March 2016. Country reports will represent the building blocks for regional and global assessment of progress.

4. The SDGs represent a challenge of a tall order, at all levels. The inherent complexity of the Sustainable Development concept, the emphasis on the inter-linkages of the SDGs, the principles underpinning Agenda 2030 and its universality, and all other elements of the new international context, will also have a significant bearing on how evaluations can be credible, reliable and useful. The 2014 United Nations General Assembly Resolution 69/237 on “Building capacity for the evaluation of development activities at the country level”⁴ and the International Year of Evaluation in 2015, achieved also thanks to a strong advocacy campaign wherein UNEG played a central role, have raised the global profile of evaluation among governments and numerous other stakeholders and triggered a debate about how evaluation can contribute to the pursuit of the SDGs.

¹ UNEG SO3-SDG WG: Towards evaluability of the 2015 SDGs: A review of evaluability literature and past evaluation experience to inform UNEG’s contribution to Agenda 2030, Terms of Reference, November 2015.

² Agenda 2030, Preamble.

³ Agenda 2030, paras 41 and 63.

⁴ A/RES/69/237, Resolution adopted by the General Assembly on 19 December 2014.

5. In this context, the United Nations Evaluation Group (UNEG) and its membership,⁵ have a responsibility to contribute to the implementation of Agenda 2030, and have to define how best they can fulfil their mandate and support this global endeavor. This Review was therefore launched to be one, among others to come, of the building blocks for the development and implementation of a UN evaluation strategy ‘fit’ for the SDGs and Agenda 2030.

6. This is the final report of the Review, to be shared with all UNEG members and presented during the UNEG Evaluation Week in April 2016. It is structured as follows:

- Section 2: Purpose, approach and methodology of the Review;
- Section 3: the analysis of the concept of evaluability with respect to the SDGs;
- Section 4: the analysis of relevant past evaluation experience;
- Section 5: the analysis of the overarching principles in Agenda 2030 that have implications on UNEG and its membership;
- Section 6: highlights on key on-going processes that will have a bearing on the evaluation of the SDGs;
- Section 7: options and suggestions for the way forward.

7. Annexes to the report include:

- Annex 1, Terms of Reference
- Annex 2, Bibliography
- Annex 3, List of people interviewed
- Annex 4, Evaluability Assessments by UNEG and its Members
- Annex 5, Some recommendations with relevance for UN organizations, from selected evaluations
- Annex 6, Lessons and recommendations with relevance for the UN monitoring and evaluation system, from selected evaluations
- Annex 7, Agenda 2030, Paragraph 74
- Annex 8, The Bangkok Declaration
- Annex 9, The Global Indicator Framework, March 2016

8. Insofar as possible, conclusions are only drawn in Section 7 to minimize repetitions. Key messages boxes on the findings have also been included at the beginning of all main sections. And in

⁵ UNEG is the voluntary professional network that brings together the evaluation units of the UN system; it was founded in 2003 and as of April 2016, it comprises 46 members and four observers.

Section 7, the various discussion points have been cross-referenced to the respective section/s in the main report.

2 Purpose, approach and methodology of the Review

2.1 Purpose and scope

9. The UNEG SO3 SDG Working Group (hereinafter SO3/SDG-WG), taking into account all the above and in the spirit of making a ‘*modest contribution*’ in support of UNEG members and their partners in tackling the evaluation challenges ahead, agreed in late 2015 to conduct a Review that would provide:⁶

- I. *An analysis of the evaluability challenges, opportunities and issues to consider in Agenda 2030, based on review of the evaluability literature, the 2030 Agenda documentation, stakeholder analysis, and on-going indicator development work;*
- II. *A review of recent evaluation experiences and reports pertaining to MDG themes and/or selected country-level evaluations related to the MDGs.*
- III. *Based on I. and II. derive lessons to support UNEG’s evaluative efforts in respect of Agenda 2030, and provide advice on what UNEG and its members should and might consider when a. developing future evaluation strategies, plans, approaches and methods; and, b. considering contributions to a shared global SDG evaluation agenda, including the potentials and risks of harmonized approaches.*

10. With these purposes in view, the Study focused on what could be learned from the evaluability assessments and evaluations conducted, or not, of the MDGs, and on un-packing the meaning of the Agenda 2030 principles. This with the aim of identifying what UNEG members should take into account in steering their methods, plans, strategies and possibly policies to make their evaluations in the SDG era, credible and reliable, while responding to the principles embedded in Agenda 2030 and the call for support to country-led evaluations of the SDGs, while fulfilling their mandates as evaluation units of UN organizations.⁷

11. Furthermore, the interviews conducted during the Review had indicated, also explicitly, a demand from among UNEG members, for more detailed information about a number of the topics touched upon in the Review itself. The report is also meeting this need too, to enable a more in-depth discussion within UNEG membership on these key issues.

⁶ Terms of Reference (ToR) for the Review.

⁷ In this document, the term ‘UN organization/s’ collectively refers to all UN funds, programmes, specialised agencies, UN Secretariat departments and affiliated organizations.

2.2 The analytical framework

12. The Inception Report (IR)⁸ laid out the analytical framework and the detailed methodology for the Review. The OECD/DAC definition of the term Evaluability was proposed, complemented by the notion that an Evaluability Assessment (EA) should assess Design, Data and Demand for an evaluation. The IR also identified the need to clarify what ‘evaluability’ means for UNEG members, and the use of EAs in their evaluation processes, in order to relate this body of experience and knowledge to the evaluability, and future EAs, of the SDGs.

13. Following a rapid analysis of the official documents and discourse around the approval of the SDGs and Agenda 2030, including the differences between the new global commitments and the Millennium Development Goals (MDGs), the IR identified the overarching principles that underpin the SDGs, and should drive and inform the implementation of Agenda 2030. These were proposed as the main thread of analysis, as follows:

- National ownership of the Agenda 2030 process, including its review and evaluation;
- Leave no-one behind;
- Paramount focus on human rights, gender equality and sustainable development;
- Inter-linkages of and across the SDGs;
- Universality of the SDGs; and
- Partnerships.

14. In addition, the call to the UN for supporting National Evaluation Capacity Development (NECD),⁹ had to be given full attention, considering the focus of the SDGs on Capacity Development and the commitment in Agenda 2030 to country-led follow-up and reviews.

15. The IR also clarified that the primary stakeholders for the Review are UNEG members and accordingly, that the key analytical perspective would be ‘*the view-point of the UN evaluation units*’. In doing so, it was important to take into account that all UNEG Members, through a variety of mechanisms and arrangements, report to the head and/or membership of their respective organizations. This means that their shared core mandate is to evaluate, and therefore enable corporate and system-wide accountability and lessons learning, the work and services delivered by their respective organizations.¹⁰ Therefore, the scope of UNEG Members’ evaluations in the context of Agenda 2030 would depend on two main factors:

- the extent to which the UN as a system, and each UN organization within it, will align its own vision, strategy and structure to contribute to the SDGs; and

⁸ The Inception Report was shared and commented upon by the members of the Review Task Team of the SO3/SDG-Working Group and the Peer Review Panel Members, in January 2016.

⁹ See footnote 4.

¹⁰ See definition of evaluation, UNEG Norms at <http://www.unevaluation.org/document/foundation-documents>.

- the extent to which the principles listed above will inform implementation and delivery by the UN.

16. Taking all of the above into account, the over-arching question to be answered by the Review was defined as follows: *‘How can UNEG members steer and meet their mandate, i.e. the evaluation of the work conducted by their respective organizations, while responding to the Agenda 2030 imperatives of national ownership of the review processes and of National Evaluation Capacity Development.’*¹¹ Also, the IR committed to explore possible options for the role of UNEG in the evaluation of SDGs, both as a network that supports its members and as an institution that has a credibility and profile in the evaluation arena, in its own right.

2.3 Methodology

17. The responsibility for conducting the Review was assigned to a consultant, who worked in close collaboration with the members of the Review Task Team of SO3/SDG-WG. The Review adopted a consultative and transparent approach with stakeholders throughout the process, and followed the UNEG Norms & Standards.¹² Triangulation of evidence and information underpinned the data collection and analysis and supported the formulation of conclusions and suggestions.

18. The Review consisted of research, study and analysis of relevant documents and materials, complemented by a round of consultation with stakeholders. Main tools were:

- Analysis of documents related to the SDGs and Agenda 2030, and of the process that led to their preparation and endorsement, including the work done by the UN Statistical Commission; UN Secretariat publicly-available documents were the majority in this group, followed by papers and comments by research and think-tank organizations;¹³
- Analysis of documents and guidelines discussing the concepts of evaluability and evaluability assessment, as well as their practical application; this also included a call for relevant examples to all UNEG members;
- Analysis of the available completed evaluations of the MDGs, Delivery-As-One (DaO),¹⁴ and Paris Declaration; a small sample of four final United Nations Development Assistance Frameworks (UNDAF) evaluations was also included, selected from among those completed in 2014 and 2015, in four different regions;
- Analysis of all available evaluation policies of UNEG Members;
- Analysis of documents on the evaluability of the SDGs and related topics, including among others, the Rome-Based Agencies Technical Seminar on the Evaluability of

¹¹ Inception report for this Review, February 2016.

¹² <http://www.uneval.org/normsandstandards>.

¹³ See Annex 2 for the complete list of reference documents and bibliography.

¹⁴ This included: the Independent Evaluation of Delivering as One, September 2012; the Delivering As One Country Led Evaluations: Synthesis Report”, October 2011.

SDG2; EvalPartners, and EvalSDGs strategy; the Global Evaluation Agenda 2016-2020; proceedings of the four Conferences on Evaluation Capacity Development;

- Consultation with key stakeholders: this consisted of semi/open-ended interviews with 43 stakeholders and informants, including from SO3/SDG-WG members, UNEG Heads and members, UN and non UN organizations;¹⁵
- Attendance of the event “No one left behind. Evaluating SDGs with an equity-focused and gender-responsive lens” (hereinafter called the No-one left behind event), held in March 2016 in New York.

19. Systematic consultation with SO3/SDG-WG members was an important feature of the Review. All deliverables, namely the Inception Report, a ‘work-in-progress’ zero-draft and a Draft report, were shared with the SO3-SDGWG Task Team and the Peer Review Panel members (all deliverables), the Working Group (Inception and Draft report), for their comments and suggestions. The Advanced draft report will be presented at the UNEG Evaluation Practice Exchange (EPE) seminar and Annual General Meeting (AGM) to be held in Geneva in the last week of April 2016. Comments provided during those events will be integrated into the final report of the Review, by mid-May 2016.

20. The Review process was also supported by a Peer Review Panel, to provide additional perspectives, experience and competence, beyond those of SO3-Working Group on SDGs. Individual interviews were held with each Peer Reviewer, in addition to their comments on the deliverables. The Panel comprised 5 members, identified on the basis of both their institutional role and personal capacity, who together represented the following groups of stakeholders:

- UNEG Heads, both current and former, for additional high-level UN evaluation function institutional and management perspective;
- UNEG members from the UN Regional Commission (UN-RCs), to bring in the regional perspective which is gaining increasing importance in the SDG context;
- Academy and think-tanks, with evaluability and evaluation experience to enrich the Review with an independent and non-UN perspective.

21. As already mentioned, the main stakeholders for the Review are UNEG members. The members of UNEG SO3/SDG-WG are at the same time, key stakeholders and the internal Peer Reviewers for the exercise. Secondary stakeholders were identified in the evaluators and commissioners of evaluation in the international evaluation community, including multilateral evaluation networks,¹⁶ IOCE, EvalPartners and its sub-networks, Voluntary Organizations for Professional Evaluation (VOPEs), independent consultants.

2.4 Limitations

22. The Review was conducted in early 2016, at the very beginning of the implementation phase of Agenda 2030, when all stakeholders, from Member States to the UN to Civil Society and others,

¹⁵ See Annex 3 for the complete list of interviewees.

¹⁶ These include the OECD/DAC Evalnet and the Evaluation Cooperation Group (ECG) of the International Financial Institutions.

were just starting to grapple with the complexity and the challenges posed by the SDGs. This means that no references nor examples could be made of concrete initiatives contributing to the SDGs, let alone their evaluations. In addition, and most importantly, the global framework for follow-up and review of Agenda 2030 was still object of discussions at the inter-governmental level. This entails that any option or suggestion on the future role and contribution of UN evaluations to the national and global reviews, could become irrelevant or impossible to implement due to decisions taken in the political arena.

23. As often the case, the time-schedule and resources for this Review imposed some limitations to the breadth and depth of the analysis. For example, the Inception Report had planned a limited search of articles and analysis by the main international press in English, French and Spanish during and around the 2015 United Nations General Assembly (UNGA). However, the number of official documents and evaluation-related material identified during the exercise, the increase in the number of stakeholders interviewed and the inclusion of the mission to attend the ‘No one left behind’ event, did not allow any space for additional bibliographic search. Further, the initial process foresaw three consultation steps for the draft report with SO3/SDG-WG members; eventually, due to time constraints, the draft report was circulated only twice, the first time in the form of incomplete draft.

24. Other challenges emerged with respect to the Review’s purpose of gaining lessons from previous evaluation experiences: although most of UNEG members make their evaluation reports publicly available, these can only be accessed through each member’s on-line repository, which are all structured along different search criteria. The UNEG repository created with the purpose of facilitating this type of search, is far from complete; and its download function is quite time-consuming, in particular when large numbers of reports are involved.

25. These constraints notwithstanding, the author believes that the richness of information canvassed from interviewees, and the wealth of feedback on the draft reports received from SO3-SDG/WG and the Peer Review Panel members, helped in avoiding major gaps in the evidence base.

3 The concept of Evaluability and the SDGs

26. This section answers the demand in the Terms of Reference, to draw lessons from the literature on evaluability and Evaluability Assessments, for the SDG era. It thus reviews the concept of evaluability and what are the key issues to take into consideration when assessing if the evaluation of a given initiative can be credible and reliable. It also analyses what has been the use of Evaluability Assessments by UNEG members and what lessons can be learned from these. Finally, it synthesizes the current status of the debate about the evaluability of the SDGs.

Key messages

The definition of Evaluability by the OECD/DAC is clear and widely accepted. However, the term in itself is somewhat awkward to grasp and can lead to misinterpretations.

The best known model for Evaluability Assessments, focuses on the analysis of the clarity and coherence of the subject’s Design, and on the availability of Data and the Demand that exist for the evaluation. EAs can be conducted at different points in time in the life-cycle of the evaluation subject, and serve quite different purposes of diverse stakeholders.

Evidence and experience from among UNEG members point to an overall “dearth of evaluability” within the UN system, and suggest that a more systematic use of the ex-ante/early use of the EAs

tools could improve the quality of design of UN strategies and initiatives, and of national policies and programmes, which in turn would allow more credible and reliable evaluations.

The discussion about the evaluability of the SDGs should focus on “how the SDGs can be evaluated” and what the evaluation community can do to ensure that the evaluations in the Agenda 2030 context, will be credible, reliable and useful.

3.1 Definitions and principles

27. The international evaluation community generally accepts the definition of evaluability given by the OECD/DAC:¹⁷ *‘The extent to which an activity or project can be evaluated in a reliable and credible fashion.’* The Glossary also states that Evaluability Assessment (EA) *‘calls for the early review of a proposed activity in order to ascertain whether its objectives are adequately defined and its results verifiable.’* Among the OECD/DAC evaluation standards, evaluability assessment is included to determine *‘whether or not the development intervention is adequately defined and its results verifiable, and if evaluation is the best way to answer questions posed by policy makers or stakeholders’*.¹⁸

28. Important work in further defining what is EA has been carried out in recent years, elaborating on the OECD/DAC definition.¹⁹ This included making some clarity on what an EA is, namely the extent to which something can be evaluated, the likely usefulness of its evaluation and its practicality. An important additional caveat is that evaluability can be assessed *‘In principle, given the nature of the project design; in practice, given data availability to carry out an evaluation and the systems able to provide it.’* Also, an EA *‘should not be confused with an evaluation, which should deliver the evaluative judgements about project achievements’* and should focus on assessing three main issues:

- programme **Design**, to explore the clarity and relevance of the intervention, the validity and robustness of the Theory of Change;
- **Data** availability about the intervention, including progress reports, baseline and monitoring information on participants, counter-factual, level of disaggregation by various criteria, quality and accessibility;
- institutional context, including both the practical feasibility aspects of an evaluation and the broader **Demand** for the evaluation, including stakeholders mapping, specific questions, potential ethical issues and risks.

29. In this perspective, EAs can be applied to single projects, programmes, country and sector strategies and Theories of Change among others, and can have different purposes, depending on the

¹⁷ OECD/DAC, Glossary of Key Terms in Evaluation and Results Based Management.

¹⁸ OECD/DAC, Quality Standards for Development Evaluation, 2010.

¹⁹ The main source of reference for this section, unless otherwise specified, was the work by Rick Davies on EA, and the various papers he authored (see Annex 2), including the DfID-funded “Working paper n.40 Planning Evaluability Assessments-A Synthesis of the literature with recommendations” issued in 2013. In consideration of the thorough mapping and detailed analysis of what exists on EA in Rick Davies’ work, this Review has not dedicated any time to finding out additional examples beyond those provided by UNEG members.

moment in the life-cycle of the evaluation subject, they are conducted. EAs are in fact seen as potentially contributing to:

- a. improve the [initiative] design prior to approval;
- b. inform the design of an M&E framework in the inception period;
- c. decide if an evaluation should take place later on; or
- d. inform the specific design of an evaluation that has been planned for, including options for evaluation timing, questions and methods.

30. Among the recommendations formulated about evaluability and its assessment, the most relevant to this Review appear to be:

- Many problems of evaluability have their origins in weak [initiative] design and EAs can be a cost-efficient approach if they can suggest changes to intervention design and thus, enable ‘better’ and more useful evaluations later on;
- Evaluability Assessments should be carried out by independent third parties, neither project managers nor those commissioned to carry out a subsequent evaluation; this would help safeguarding the independence of later evaluation processes;
- Checklists emerged as the most frequent and useful tool for conducting Evaluability Assessments, as they support consistency of areas of enquiry across different projects, in case of comparison, or phases of the assessment. Check-lists can be used as stand-alone tools along with ratings, or be integrated by comments and analysis or have a more background role informing the coverage of a detailed narrative report.

31. UNEG provides a slightly more restrictive definition of EA with UNEG Norm 7.2.²⁰ This states that an evaluability exercise would be useful prior to undertaking a major evaluation that requires a significant investment of resources, to ensure that there is ‘*clarity in the intent of the subject to be evaluated, sufficient measurable indicators, assessable reliable information sources and no major factor hindering an impartial evaluation process*’. Norm 7.1 clarifies that any contribution by an evaluation function to Management for improving ‘*the ability to evaluate the undertaking*’, and even to ensure evaluation provisions for it, ‘*should be performed in an advisory capacity only*’, so as to safeguard independence of the evaluation function itself.

32. In this perspective, an EA would basically consist of options c. and d. above, and when conducted at approval or during the early implementation phase of a new initiative, there should be measures in place so as not to compromise the impartiality of the actual evaluation conducted later on. This concern is fully in line with Davies’ recommendation that the EA should be conducted by external consultants, with respective evaluation units only in a facilitation role for the process.

3.2 The use of evaluability assessments among UNEG members

²⁰ At the time of writing this report, UNEG was revising its Norms and Standards. All references in this Review to the UNEG Norms and Standards are to the 2005 version.

33. The Review ran a word-search of the 33 UNEG members' evaluation policies available in UNEG Web site.²¹ Half of the policies, 17, include a reference to evaluability and in virtually all of these, it is a step in the evaluation preparatory process. Six evaluation policies also include in their mandate, the EA of projects and programmes at an early stage of formulation, to ensure the consistency and logic of projects' Results-Based hierarchies or Theories of Change. Other four evaluation units clearly state in their policies that ensuring the evaluability of projects and programmes is responsibility of Senior Management and Programme units.

34. A non-comprehensive search of evaluation guidelines and manuals issued by the 17 UNEG members whose policies include a reference to evaluability, suggests that UNEG Norm 7.2 has been broadly adopted and integrated in current evaluation guidance materials. In at least two instances, EA has the additional explicit purpose of contributing to integrate the gender dimension in the evaluation design.

35. A number of UNEG members, including ILO, UNCDF, UNICEF, UNODC and UN-Women, have developed guidance tools on how to conduct an EA, to be also used by Programme Managers in revising project documents at the time of appraisal or early in the intervention cycle.²² Parameters for assessment typically include Indicators, Baselines, Milestones, Risks and assumptions, and Monitoring and Evaluation. In the case of UNCDF and ILO, the assessment is done by Evaluation Office staff; in their views, procedures in place prevent potential conflict of interest at the time of the evaluation. The template by UNODC Independent Evaluation Unit foresees the possibility for an evaluation to be cancelled, however this has not happened so far, neither in UNODC nor in other UN organizations, according to the information available to this Review.

36. To explore how UNEG members apply EA in practice, a call for examples of evaluability assessments was launched to UNEG members at the inception of this Study. This led to five Evaluability Assessments, two of which titled evaluations, of organizations' strategic plans and programmes of work, namely: two by UNEP and one by UNICEF, UNIFEM and WFP evaluation units respectively.²³ The Synthesis of the Evaluability Assessments of the Delivering as One pilots (henceforth called the DaO EA)²⁴ and the 2006 "UNEG Study on the Evaluability of the UN Development Assistance Framework" (henceforth called the UNDAF EA) were also added to the list for this Review.

37. The analysis of these seven EAs/evaluation reports by purpose, timing, scope, objectives, criteria and methods, showed the following:²⁵

²¹ UNEG Web site contains 31 evaluation policies, however, as UNDP evaluation policy is also the policy for UNCDF and UNV, in total 33 evaluation units are represented. Also, twelve UNEG members have no evaluation policies posted in UNEG Web site.

²² See UNICEF Eval News, Issue n.62, September 2015.

²³ ILO EVAL also conducted an evaluability review of the Office's Strategic Policy Framework, however information available on how it was conducted was not sufficient to include here that experience as well.

²⁴ The Synthesis of the DaO Evaluability Assessments comprises the EAs of the eight Delivering as One pilots; here, they are referred to as one single exercise.

²⁵ See Annex 4 for more detail.

- a. All exercises focused on the strategic plans/frameworks (SPs) of the respective organizations; in the case of the DaO EA, the focus was on the DaO results frameworks in each country and in the case of the UNDAF EA, on 35 UNDAFs;
- b. The terminology adopted and the scope of these exercises largely meet Rick Davies' differentiation between EA and evaluation; on terminology, the exception was UNEP, which called its own exercises 'formative evaluations' and considers that this ensured stronger interest and buy-in among Management; and on scope, the exception was the DaO EA, which included in its assessment an initial evaluation of the DaO Pilots;
- c. Purpose of the EAs and evaluations differed: UNDAF, UNEP, UNICEF and WFP EAs and evaluations aimed at improving the Design of the respective agencies' SPs and of the UNDAFs, thus falling into Davies' category a.; the WFP EA also wants to contribute to evaluation priority setting, i.e. category c.; UNIFEM EA aimed at providing feedback on the feasibility of an evaluation, i.e. categories c. and d.; and DaO EAs only aimed at improving the evaluation design, i.e. Davies' category d.;
- d. Four out of seven exercises took place early in the cycle of the subject of analysis, and the other three, towards the end. However, only the UNICEF EA will provide inputs to the same SP, when this will go through the mid-term review. All other exercises did or will feed into the following plan or guideline, one or two years later. The DaO EAs informed the country-led evaluations conducted two years later. In the case of UNIFEM, the EA which had been initially planned to be conducted at mid-point of the SP life, became a final learning exercise due to the merging of the entity into UN-Women;
- e. Clarity, relevance and coherence were the common assessment criteria to all the exercises, although they were defined slightly differently in UNIFEM and DaO EA; the latter two also included other criteria, due to their specific purposes;
- f. With respect to methods and tools, the EAs tended to rely on extensive desk-reviews, supported by rubrics applied to the reviewed documents, and interviews of a selected sample of key stakeholders. UNICEF, UNIFEM and obviously the DaO EAs, included interviews at country level. Benchmarking with other organizations was not done.²⁶ UNEP and UNICEF explicitly adopted a theory-based approach.

38. Timeliness of delivery of the EAs was of paramount importance, at least in the cases of UNICEF and WFP, to capture the window of opportunity offered in the respective SP programming cycles. However, only the UNICEF EA provided information that could be integrated in the same cycle programming documents. In the cases of UNDAF, UNEP and WFP EAs, the assessment of the on-going SPs or policy fed into the next cycle, thus detracting on relevance, efficiency and effectiveness of the whole process. This model is likely to be the most frequent, unless EAs as an early analytical review tool, are fully built into the programming process.

39. With the exception of the DaO exercise which had a very complex management set-up, no specific information was available in the reports about the arrangements for conducting the EAs. All appear to have been conducted by consultants responding to the respective evaluation units, with the

²⁶ Only the DaO EA mentioned benchmarking, against the 2001 and 2004 Triennial Comprehensive Policy Reviews.

exception of UNEP, where Evaluation Office staff conducted both. All evaluation units should ensure, in order to protect their credibility, a high degree of segregation of duties between those who managed or conducted the EAs and those who will design, manage and conduct the evaluations of the SPs. On this specific point, the position taken by the four UNEG members that clearly indicated in their policies that responsibility for the evaluability of strategies and programmes rests with Management, appears highly relevant and more UNEG members may wish to consider following these examples.

40. When it comes to the use of the EAs' findings and recommendations, the DaO EAs were quite extensively referred to, and considered useful in providing substantive and methodological information for, the subsequent evaluations. By the time of this Review, Management Responses were available for the 2011 UNEP evaluation and for UNICEF EA. In the case of UNEP 2011, its findings were referred to in the following Work Plan and the Executive Director's 2013 report to the Governing Council on the design of the 2014-17 Medium term strategy. The UNICEF Management Response was positive about the recommendations; the direct uptake will be seen when the revised Strategic Plan will be issued. Similarly, the UNEP 2015 and WFP assessments will fully prove their worth when the new respective SPs will be finalized or even evaluated. However, it is reasonable to assume that all these exercises did/will contribute to improve the clarity, coherence and completeness of the intended SPs, including their Theories of Change (ToCs).

41. With respect to costs, information available suggests that these varied widely, from approximately two to three months of a mix of senior and junior staff-time in the case of UNEP, to approximately USD 140,000 in the case of UNICEF to an amount possibly in the order of a few hundred thousand dollars in the case of the DaO EAs. The scope and approach of each exercise are the determining factors of cost, as in the case of evaluations. On the same line, no sufficient information is available for a meaningful calculation of savings in evaluations due to previous EAs.

42. The low number of Evaluability Assessments of SPs conducted by UNEG members so far, also suggests that a certain amount of advocacy might be necessary to raise awareness and buy-in for these exercises, as Senior Managers at the receiving end, may not be well informed about the potential usefulness of this tool. Similarly, UN Member States may also not be sufficiently aware of the added value of independent early assessments of their own SDG-related plans and strategies. In this respect, an additional argument in favor of EAs of Strategic Plans comes from the "Independent Thematic Evaluation of UNIDO's Contribution to the Millennium Development Goals", discussed more in detail later in the report, which formulated the two insightful Lessons learned included in Box 1 below.

Box 1. Lessons learned in the "Independent Thematic Evaluation of UNIDO's Contribution to the Millennium Development Goals"

1. Paying close attention to the definitions of MDG targets and indicators helps to better understand, how projects may contribute to the MDGs and actually contribute.
2. Using project indicators that are compatible with the national MDG indicators increases the relevance of UNIDO.

Source: UNIDO Independent Thematic Evaluation of UNIDO's Contribution to the Millennium Development Goals

43. The same argument should be applicable to the national contexts, through the ex-ante or early analysis of SDG-related national policies and programmes, from the point of view of soundness of Theories of Change, internal coherence, indicators and targets. In these cases, the primary stakeholders

promoting this type of exercise should be parliamentarians, VOPEs, Civil Society and UN Country Teams.

44. All the above points to the scope and need for consolidating knowledge and experience about EAs within UNEG; and for UNEG members to advocate for improved evaluability of the work done by their organizations in support of the implementation of Agenda 2030. Section 7.5 discusses this more in detail and includes specific suggestions.

3.3 The debate on the evaluability of the SDGs

45. The discussion within the evaluation community on the specific challenges linked to the evaluability of the SDGs emerged in the evaluation community in connection with the 2014 UNGA resolution on National Evaluation Capacity Development²⁷ and the launching of the International Year of Evaluation 2015. In March 2015, the newly elected Chair of UNEG identified, among other priorities, evaluability studies to ensure that the new Development Agenda be evaluable.²⁸

46. In parallel, the evaluability of the SDGs had found its place in the agendas of UNEG and some of its members. In particular, the evaluation units of the four agencies with headquarters in Rome, Italy (Rome-Based Agencies, RBAs), namely the CGIAR (formerly known as the Consultative Group on International Agricultural Research), the Food and Agriculture Organization of the United Nations (FAO), the International Fund for Agricultural Development (IFAD) and the World Food Programme (WFP), joined forces to organize an international technical seminar focused on the evaluability of “SDG2-End hunger, achieve food security and improved nutrition and promote sustainable agriculture”, fully in line with the mandates of the four organizations which together, cover a wide range of issues associated with this Goal in contexts ranging from humanitarian crisis to development.

47. The event, with title “Enhancing the evaluability of Sustainable Development Goal 2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture” aimed to *‘contribute to a shared understanding of how SDG2 could be evaluated and identify actions needed to enable evaluations of SDG2 through the United Nations system, other international organizations or countries themselves.’*²⁹ The seminar took place in mid-November 2015, at IFAD premises, with attendance in person of 160 invited guests and some 1000 virtual participants who followed the event through web-stream. The discussion was very rich and the event considered a success. The debate stressed the links between evaluation and democracy and the RBAs evaluation units were challenged to *‘do things differently’*. In this respect, there is already a body of evidence that indicates that evaluation units, in their capacity as promoters of change, should, and in fact do contribute to nudging their agencies towards improvements in their modus-operandi.

²⁷ Op.cit.

²⁸ Marco Segone speech, <https://www.youtube.com/embed/7erKmkAbodg>.

²⁹ Concept note, Enhancing the evaluability of Sustainable Development Goal 2 (SDG2): "End hunger, achieve food security and improved nutrition and promote sustainable agriculture" Technical seminar jointly organized by the evaluation offices of CGIAR, FAO, IFAD and WFP Date: 17-18 November 2015 Rome, Italy. See the event on YouTube at: <https://youtu.be/-JGdaIr3cpY>.

48. The proceedings for the seminar issued in January 2016 distils two sets of Next steps, one for the RBAs themselves and one for the four evaluation units.³⁰ As shown in Box 2, the latter identifies potential pathways ahead for UNEG and its members, irrespective of the technical mandate of respective organizations. No specific follow-up was foreseen at the time of writing this report.

Box 2. Next steps for the evaluation functions of the Rome-Based Agencies

To strengthen the evaluability of SDG2 on the demand and supply sides, the evaluation functions of the RBAs should:

- Apply a political economy perspective to evaluation and become an agent of change; do not just conduct evaluations, but also find ways to instill evaluative thinking;
- Pay attention on building country capacities to both evaluate and persuade decision-makers of the value of evaluation;
- To strengthen country capacities, provide tools and guidance on evaluability assessment;
- Develop an evaluation agenda taking into account the needs to localizing SDG2 in response to the new context; consider the multiplicity of goals and players in the approach to accountability and learning;
- Provide functional support to simplify the complexity of monitoring the large number of SDG2 indicators at country level and improve the robustness and credibility of in-country data, for example through:
 - Improved methodologies;
 - Building the demand side;
 - The establishment of quality assurance processes aiming at (i) harvesting the low-hanging fruits; and (ii) reduce the burden associated with the collection of data.

Source: Proceedings, Enhancing the evaluability of Sustainable Development Goal 2 End hunger, achieve food security and nutrition, and promote sustainable agriculture International technical seminar

49. Throughout 2015, the discussion on the evaluability of the SDGs continued in the many events linked to the Year of Evaluation and within UNEG. Notably, the EvalSDG network emerged within EvalPartners, developed a Concept note and was formally launched at the 2015 Global Evaluation Week in November 2015;³¹ and the Bangkok Declaration was issued at the conclusion of the Fourth International Conference on National Evaluation Capacities, in October 2015. Both instances make explicit reference to the evaluability of the SDGs; the former commits to advocate for performance indicators to be evaluable and the latter, suggests that evaluability should inform the priority-setting process at country level in the context of SDG-related national plans.

50. In other, simpler words, these statements aim at ensuring that evaluative evidence be among the tools used by decision-makers in deciding priorities, in designing policies and programmes, and in assessing progress towards the SDGs, including the reasons why progress did or did not occur. These

³⁰ See Proceedings at <http://www.ifad.org/evaluation/events/2015/sdg2/seminar.pdf>.

³¹ See <http://forum.ioce.net/forum/open-forums/evalsdg>.

are fully legitimate goals, which however relate more to the soundness of policies and programmes to the use of evaluation in informing the policy-cycle, than to the actual evaluability of the SDGs.

51. These examples, added to the diversity of exercises that have come under the title of Evaluability Assessment discussed in the previous section, suggest to the evaluation community, which correctly wants to make of “Evaluation” an acknowledged and respected profession, the need to strive for a higher degree of consistency and clarity in the definition and terminology around ‘evaluability’.

52. In addition, the use of the term ‘evaluability’ out of context might even be counter-productive: if an initiative does not appear ‘evaluable’ now because the evaluation community does not have the methods and tools yet while it still has a potential to improve livelihoods or the state of natural resources, as partly is the case with the SDGs, should it not be launched or tested simply because evaluators do not know how to evaluate or measure it? These may appear idle questions, but not distant memories of discussions about linking development assistance exclusively to programmes “measurable through Randomized Control Trials”, make one wary of a certain type of language.

53. Finally, the Inception Report for this Review stated that *‘the SDGs cannot be evaluated per se: the political process that led to their formulation and approval was characterized by such an unprecedented level of participation, inclusiveness and consensus, that their assessment will only be possible through a historical perspective, whenever its time will come’*. The interviews and the document analysis conducted during the Review, partly supported and partly challenged this statement. Support came from those who agreed with the likely lack of political interest for an evaluation of the Goals per se, e.g. of their relevance or comprehensiveness, and of the process that led to them. A number of evaluations discussed in the next section, did for example focus on the ‘evaluation of the effects’ of political declarations or system-wide policy decisions, rather than on the analysis of the commitments per se. Conversely, challengers of the statement observed that “everything can be evaluated” and that declaring the SDGs as non-evaluable would pre-empt any future possibility of evaluative analysis.

54. These discussions showed the risk for the question about the evaluability of the SDGs to remain focused on a theoretical debate about “**should the SDGs be evaluated**”, rather than on a more pragmatic analysis of “**how the SDGs can be evaluated**” and consequently, what the evaluation community should do to ensure that the evaluations in the Agenda 2030 context, will be effective in providing rigorous and useful assessments of the progress made towards the SDGs. This is discussed in the following sections.

4 Lessons learned from past evaluation experiences

55. The Terms of Reference for the Review included among its purposes, the ‘*Review of recent evaluation experiences and reports pertaining to MDG themes and/or selected country-level evaluations related to the MDGs*’. It is important to clarify here that the Review sought out lessons learned or to be learned about the evaluation processes themselves, rather than on the subjects of the selected evaluations. One exception is the OIOS “Thematic evaluation of monitoring and evaluation of the Millennium Development Goals: lessons learned for the post-2015 era” (henceforth called the OIOS MDG evaluation), whose subject of analysis were the UN M&E systems and which is discussed in a separate sub-section.

Key messages
UNEG members did not engage at any significant level in the evaluation of the MDGs.
The single repository for evaluations conducted by UNEG members is not adequately populated
The few existing evaluations of the MDGs and of complex political commitments such as the Paris Declaration and the DaO evaluations, and probably the upcoming ISWE pilot evaluations, provide highly useful and relevant lessons that should be properly taken up in the UN evaluations of the SDGs.
The evaluations analyzed here, and probably a number of the UNDAF and country-level evaluations, contain a wealth of learning opportunities for the Member States and the UN system on joint delivery, coordination and partnerships, that should be extracted, and made available in easy formats suitable for decision-makers “on-the-run”.
The lessons and proposals included in the OIOS MDG evaluation could be considered for the formulation of a UNEG strategy for a future UN evaluation architecture for the SDGs.

4.1 The selection of relevant evaluations

56. The Review carried out a rapid search in UNEG on-line document repository for evaluation reports linked to MDGs.³² This produced 96 records, 76 of which were evaluations from UNDP, 9 from UNICEF and eleven from five other agencies. Four evaluations were tagged as “joint”. In terms of geographical scope, 26 did not have an indication, although from the titles seven were conducted in Turkey, five in China and in the Philippines respectively, and the remaining 66 focused on 30 different countries. Last, 58 records out of 96 were evaluations of the MDG Fund. UNEG repository also includes the DaO evaluation.

57. A similar search in UNDP Evaluation Resource Centre produced 4438 records related to MDGs. Of these, the Independent Evaluation Office was the owner for 223. However, the search engine captures the term MDG even if it is simply included in the report, which one would expect to happen

³² Data downloaded on 26 February 2016, from <http://www.unevaluation.org/evaluation/reports>.

in any case for most UNDP initiatives implemented in the period 2000-2015.³³ Thus, the simple selection of which of the 223 records were MDG evaluations proper, would have absorbed a significant amount of time.³⁴

58. The difficulty of identifying “MDG evaluations” was confirmed by the much more thorough and credible analysis conducted by OIOS, for its “Thematic evaluation of monitoring and evaluation of the Millennium Development Goals: lessons learned for the post-2015 era”. The OIOS evaluation team scanned a wide range of sources for Millennium Development Goals-related evaluations, and reached the conclusion that *‘it proved very challenging to access Millennium Development Goal evaluations in a meaningful, systematic or digestible manner’*. Main challenges included sorting out evaluation reports from lists of documents in the order of hundreds and thousands of records, separating evaluations conducted as per the UNEG Norms and Standards, as distinct from reviews or studies, and the dispersion of sources. Nor were meta-evaluations or meta-analysis found of multiple evaluative assessments of the contribution by the UN to the achievement of one or more MDG.

59. Based on OIOS valuable experience, complemented by a call to UNEG members for evaluations *‘conducted over the last 3 years which included evaluation of the MDGs, the Delivery-As-One process and UNDAF evaluations’* and consultation with the Peer Review Panel members, this Review focused its analysis on the lessons that could be learned from the following evaluations:

- the Independent Evaluation of Delivering as One (henceforth called the DaO evaluation) and the DaO country-led evaluations Synthesis (henceforth called the DaO CLE synthesis);
- OIOS “Thematic evaluation of monitoring and evaluation of the Millennium Development Goals: lessons learned for the post-2015 era”, already mentioned (henceforth called the OIOS MDG evaluation), issued in March 2015 and presented to ECOSOC;
- the Evaluation of the Paris Declaration, Phase 2 (henceforth called the Paris Declaration evaluation);
- the UNDP “Evaluation of the Role of UNDP in Supporting National Achievement of the Millennium Development Goals” (henceforth called the UNDP MDG evaluation);
- the UNDP Global and Thematic Evaluation of the Millennium Development Goals Achievement Fund, issued in September 2014 (henceforth called the MDGF evaluation);
- the Final UNDAF evaluations conducted in Bangladesh, Honduras, Kazakhstan and Mozambique;³⁵

³³ For example, a number of FAO evaluations conducted in the period 2005-2014 included ‘contribution’ to selected MDGs as a criterion; however, this was not done in a systematic manner and they could not be called ‘evaluations of the MDGs’ by any stretching of the term.

³⁴ The Web site of the Evaluation Coordination Group (ECG) contains a number of MDG-related evaluations by ECG members, with limited useful information for the purpose here.

³⁵ These were selected from a universe of eight UNDAF final evaluations conducted in 2015, to have a sample of evaluations conducted in four different regions and with different budgets.

- a review of UNHCR engagement with the DaO process (henceforth called the UNHCR review); and
- the “Independent Thematic Evaluation of UNIDO’s Contribution to the Millennium Development Goals” (henceforth called the UNIDO MDG evaluation).

60. The UNDAF Evaluability Assessment already discussed in Section 3, is included here as well because of its analysis of and recommendations on UN system-wide monitoring and evaluation.

61. The difficulties met by the OIOS in its MDG evaluation and by this Review, in identifying relevant evaluations produced by UNEG members, raise two main points. First, the MDGs were not evaluated by the UN evaluation system, with very few exceptions. Why this was the case, might be an interesting area of enquiry, or at least discussion among UNEG members, if there are relevant lessons to be learned for the SDGs. Second, it would be highly desirable that all evaluations produced by the membership are systematically made available on one single repository, using the existing UNEG database for evaluations. This could also include an “SDG category”, and UNEG Secretariat might consider improving the user-friendliness of the download function from the database.

4.2 Findings about evaluation approaches, management and methods

62. These evaluations are very different from each other, as to be expected, in terms of complexity of their subject of analysis, scope, resources, etc. Key distinctive elements are:

- The DaO, UNDP MDGs and the Paris Declaration evaluations assessed, insofar as possible, how political declarations, decisions and goals agreed at the supra-national level had an effect, through a multiplicity of decisions made by actors in the most diverse circumstances and levels, on national policies, strategies, resource allocations, modus-operandi and people’s lives, and how these effects have in turn contributed to achieving the intended commitments;
- The MDG-F evaluation focused on the Fund as a ‘multilateral mechanism for development cooperation’, on its impact on the UN system-wide coherence, and on the effectiveness and impacts of a long list of projects implemented across a large number of countries;
- the UNDAFs evaluations focused on the respective frameworks, focusing insofar as possible on the immediate results or organizational outcomes;
- The UNHCR review aimed at assessing ‘the extent to which factors internal to UNHCR or the DaO Initiative, including mandate, policies and business practices, influence the effectiveness of UNHCR’s engagement with it’;
- the UNIDO MDG evaluation, basically a meta-evaluation, aimed at assessing ‘UNIDO’s technical cooperation activities [that] have contributed to the achievement of the MDGs or have the potential to do so’ and included an analysis of the clarity and coherence of UNIDO’s strategic plans, based on the evidence gathered and analyzed through past evaluations.

63. Nevertheless, the evaluations do share some similarities. First of all, with the exceptions of the UNHCR review, the UNDAFs and the DaO CLEs, all others used theory-based approaches and

developed Theories of Change (ToCs) at the beginning of the work, to map out what each subject of analysis had aimed at achieving and how. The DaO CLEs typically did not have a model of causality against which performance could be tested, therefore used '*objectives mapping and establishment of historical causality*', which might be proxies for ToCs. The wide use of a theory-based approach does not prove that this is the only possible approach to tackle complex issues such as those assessed in these evaluations, but it definitely says that such an approach allows robust evaluations in a context of multiple stakeholders and streams of action.

64. Contribution analysis was the most frequent approach used and proved its value, though in some cases it could not be applied, e.g. in the UNDAF evaluations. All evaluations largely relied on qualitative evidence. In the view of the authors of the Paris Declaration evaluation, this, when '*transparently sourced, harnessed and presented, is powerful and invaluable evidence, especially when applied to account for changes in complex areas..*'. When reliable quantitative evidence was available (e.g. aid volumes), some evaluations used it too. However, only the Tajikistan UNDAF evaluation explicitly mentioned national statistical data as a source of information for assessing results at the outcome level.

65. All evaluations focused on relevance, effectiveness, efficiency and sustainability. Impact was not included as a criterion, with the exception of the UNIDO MDG meta-evaluation that extracted information from previous evaluations, about how projects had contributed to the MDGs. Some evaluations also explicitly included 'equity', while others analyzed Gender Equality and Human Rights as cross-cutting issues, though not as criteria. Tools tended to comprise desk-reviews and analysis of data-sets, semi-structured interviews, opinion surveys of stakeholders, and all used triangulation to validate the various sources of information.

66. Common identified challenges, that confirm to some extent the low evaluability of programs and policies due to issues related to Design and Data, were:

- lack of baseline data and of disaggregated data to quantify changes, for example in the efficiency of business practices and cost savings;
- fragmentation and overlapping of interventions, multiplicity and diversity of approaches towards the same goals or objectives;
- inconsistencies in parameters and practices of programme management, diversity in monitoring and evaluation, differences in budgetary frameworks, and uneven and incomplete management information systems, particularly with regard to financial information.

67. An additional challenge specific to synthesis and meta-evaluations of country-level evaluations, such as the DaO CLEs and UNDAFs was the lack of common frameworks and benchmarks developed early in the evaluation process, for allowing comparison of progress assessment across countries. This makes any attempt at extracting common lessons very complex and short of potential. The Paris Declaration evaluation was on the contrary a good example of how dispersion of information can be avoided.

68. Last commonality, none of the evaluations used a counter-factual approach: the DaO evaluation regretted that the analysis of 'non-pilot' countries was not part of its Terms of Reference, and the UNDP MDG evaluation stated that 'a 'counterfactual' scenario (such as an identical world without the MDGs in it) was absent. The development of counter-factual in this type of evaluation appears indeed highly complicated and possibly not useful, in particular when it comes to attempts at comparing countries.

Seemingly, the use of a counter-factual in evaluative frameworks within the Agenda 2030 context, will be subject to the same complexities as in the world pre-SDGs.

69. The most detailed and comprehensive example available of how regional, thematic and global evaluations of the SDGs could be developed and conducted is the Paris Declaration evaluation and its complementary documents. Both Phase 1 and Phase 2 evaluations were based on a number of elements or principles, namely:³⁶

- a. the joint nature of the Evaluation, that would be conducted across donor agencies and partner countries, and within a governance structure that reflected this [jointness];
- b. the management and governance structure at national and international level, devised to ensure the joint nature of the process, as well as the participatory and consultative approach adopted, and solid quality assurance;
- c. the voluntary basis of participation, meaning that inclusion in the sample of countries and agencies was essentially on the basis of self-selection;
- d. the primacy of the country as the main arena for evaluation, to allow a focus on how aid effectiveness reform has played out in practice;
- e. for donor headquarter studies, the application of the same Terms of Reference as those from Phase 1 to enhance comparability;
- f. a trilingual exercise; in order to serve the participants and intended users of the Evaluation as fully as possible, the process was organized to operate throughout in English, French, and Spanish.

70. The analysis of the process led to the following conclusions, which also appear particularly relevant for any evaluation of the SDGs at regional and global level:³⁷

- g. A fully participatory approach is essential but can add burdens of its own, to ensure common understandings, to manage consistency and to secure broad-based engagement, though it also carries the risk of an overload of questions / issues to be addressed within the evaluation;
- h. a very clear central framework for analysis is necessary to analyze findings across diverse contexts, often generated through diverse methods; this also facilitates the assessment of both the quality of evidence provided and the substance of the findings at synthesis stage;
- i. a clear common framework with a clear results logic, developed through a participatory approach, are also necessary to ensure consistency of findings whilst allowing for the required flexibility at the local level;

³⁶ The Paris Declaration Evaluation - Process and Methods. Betts, J; Wood, B. in *The Canadian Journal of Program Evaluation* Vol. 27 No. 3 Pages 69–102, 2013.

³⁷ The Paris Declaration Evaluation - Project Completion Report, March 2012. Niels Dabelstein, Paris Declaration Secretariat, March 2012.

- j. where new approaches are required, capacity constraints need to be anticipated and addressed, as support to the technical aspects of individual evaluations is crucial to ensuring consistency and commonality of approach;
- k. independence is critical, particularly where potentially contentious findings are likely to emerge; this should include strong procedures and clear governance systems, articulated from an early stage, reinforced throughout the process, and checked and verified as part of ongoing evaluation management;
- l. for a multi-site evaluation, the synthesis process needs to be clarified from the start and a clear and robust framework for synthesis is necessary as early as feasible in the process, to ensure that evaluation design, data collection and analysis are fully geared towards the synthesis stage;
- m. dissemination activities need to be planned well ahead, including the targeting of stakeholders who are external to the evaluation, such as key decision-makers.

71. It is also important to note here that, as incidentally also suggested by the incumbent UNEG chair in a message to the members,³⁸ all DaO-, MDG- and UNDAF evaluations, and possibly others such as joint evaluations of country or global programmes, have produced significant evidence and formulated interesting recommendations and lessons learned for various groups of stakeholders, from the implementation of DaO and UNDAFs to factors affecting progress towards the MDGs and the Paris Declaration. This constitutes quite a powerful body of evidence and recommendations, for Member States and the UN system, to improve consistency in approaches, collaboration, efficiency and effectiveness of their collaboration. Although these are not directly relevant to the evaluation practice per se, they are relevant to what UN Member States have embarked into by committing to the SDGs and endorsing Agenda 2030.³⁹

4.3 Findings about monitoring and evaluation systems

72. The UNDAF Evaluability Assessment, the DaO Independent Evaluation and the OIOS MDG evaluation, analyzed aspects of the UN monitoring and evaluation systems and formulated important recommendations in this regard. For example, the 2006 UNDAF Evaluability Assessment found that compliance with the UNDG Guidelines for the preparation of the UNDAFs was limited and that *'evaluation of the UNDAFs against the scope suggested [by the Guidelines] would require significant additional investment in addressing issues not adequately dealt with during the design and initial implementation of the UNDAFs'*. Additional key points were as follows:⁴⁰

- absence of clear logic, both internally and between the UN interventions and impact on national goals, and of outcome indicators;

³⁸ Op.cit.

³⁹ A selected sample of such lessons and recommendations are included in Annex 5.

⁴⁰ The full text of Lessons Learned and Recommendations from these evaluations, relevant to this Review, is available in Annex 6.

- duplication of monitoring and reporting efforts between UNDAF and individual agencies;
- UN Country Teams (UNCTs) without the expertise and the resources to commission an evaluation compliant with UNEG norms and standards.

73. Key recommendations of the Study were to limit the UNCTs' mandate to strategic monitoring and review of the joint work, assess the real demand for evaluation of the UNDAFs and transfer the responsibility for the latter, to the UN evaluation functions. Further, national governments should be responsible for monitoring the higher level goals, whereas individual agencies should monitor their own work below the outcome level. The first recommendations were not accepted, and seem to be even more relevant now than in 2006.

74. In 2012, the DaO evaluation formulated a recommendation on the need for independent system-wide evaluations of system-wide approaches. The 2012 UNGA resolution 67/226⁴¹ built upon it, with the Independent System-Wide Evaluation mechanism (ISWE) hosted by JIU, that started operating in May 2014.

75. The OIOS MDG evaluation issued in March 2015 focused on the UN system monitoring and evaluation set-up with respect to the MDGs. It aimed at answering three questions, the third of which took a forward-looking perspective: *'What Millennium Development Goals monitoring and evaluation challenges/ opportunities and good practices will be applicable to maximize the contribution of monitoring and evaluation functions in support of the post-2015 development agenda'*.

76. The evaluation, whose report was presented to the ECOSOC Committee for Programme and Coordination in March 2015, is highly relevant to the future of the evaluation system in the UN and may be considered, similarly to what is the JIU system-wide assessment of the UN evaluation functions,⁴² a baseline for the performance of the UN monitoring and evaluation system, for any future related assessment. The report includes some highly significant lessons for UNEG members, as follows.

77. Lesson 1 states that the current monitoring and evaluation UN system framework consists of a variety of monitoring components which have evolved over time, as also underlined by the DaO Independent Evaluation with respect to the whole UN system. This "original fragmentation" stems from the historical evolution of the UN system from a set of diverse organizations attached only at the high-level of principles to the UN umbrella, with all the consequences this entails. In fact, this still holds true for the governance systems of all UN organizations that are not part of the Secretariat.⁴³ As the ISWE experience appears to confirm so far, the UN being a "system of systems" represents a major challenge also for the evaluation of system-wide issues and of common policies and goals, as was the case for the DaO and the MDGs.

⁴¹ A/RES/67/226, UNGA Resolution on the Quadrennial comprehensive policy review of operational activities for development of the United Nations system.

⁴² JIU/REP/2014/6, Analysis of the Evaluation Function in the United Nations System, United Nations, 2014.

⁴³ This means that the majority of UNEG member organizations do not have a direct reporting line to the UN Secretary General.

78. Linked to the observations underpinning Lesson 7, that identified a gap in the establishment of rigorous evaluations of the progress towards the MDGs, Lesson 2 states that the post-2015 SDGs would benefit from an early as possible definition of a clear, overarching framework of monitoring and evaluation objectives, role/responsibility definitions and coordination mechanisms. It also entails that UNEG and its members should urgently and pro-actively work to ensure that their views and experience are fully reflected in the proposed over-arching framework.

79. Lesson 5 on a strategy for National Evaluation Capacity Development, which should include *‘assigning clear roles, responsibilities and resources to build evaluation capacity for systematic, cyclical input to decision-making by stakeholders involved at the global, regional and national levels’* similarly calls for a rapid definition of roles and responsibilities for UNEG members vis-à-vis other stakeholders, the UN development system, national governments and other international and national partners. The report further notes that more resources will be required at the national, regional and global level, for effective monitoring and evaluation of the SDGs. The same point had been raised in the JIU 2014 analysis of UN evaluation functions.⁴⁴

80. The report also defines evaluation as a bridge between monitoring and accountability at the different levels – national, regional and global – and proposes *‘a monitoring and evaluation framework that integrates both the vertical and horizontal (or temporal) levels, to capture and disseminate digestible key lessons learned for use by decision makers seeking to implement evidence-based corrections in pursuit of effective achievement of the sustainable development goals’*. The framework foresees comprehensive evaluations and reporting every five years, fed from annual monitoring and reviews performed by a variety of organizations and stakeholders.

81. The OIOS report finally formulates a recommendation for the *‘development of a coherent, coordinated sustainable development goals monitoring and evaluation plan to support Member State decision-making’*. To a broad extent, this recommendation has been integrated in the report of the UN Secretary General, issued in January 2016, on the follow-up and review mechanism for the SDGs at the global level.⁴⁵ The discussion at the ECOSOC Committee for Programme and Coordination of the OIOS recommendation showed diverging views among Member States on the need for an overarching monitoring and evaluation framework, and have spilled over into the discussion of the SG’s proposal, as discussed in Section 6.1.

⁴⁴ Op.cit.

⁴⁵ A/70/684, Report of the Secretary-General on critical milestones towards coherent, efficient and inclusive follow-up and review at the global level, 15 January 2016.

5 Agenda 2030 implications for UNEG members

82. This section discusses how the overarching principles underpinning the SDGs and Agenda 2030, including National Evaluation Capacity Development as identified in the Review Inception report, should inform, and should be taken into account in the work and modus operandi of UNEG and its members, so as to strengthen the contribution of the UN evaluation system to the achievement of the SDGs.

Key messages
National ownership and country-level evaluations
The paramount principle of country-ownership for Agenda 2030 achievements will be reflected in the UN modality of work at country level, through the widespread adoption of the Delivery as One modality and the centrality of the UN Country Teams (UNCTs) in delivering support
The complexity inherent in the UN Sustainable Development Frameworks (UNSDF) evaluations, the request for harmonization of country-level evaluations within the UNSDF evaluation and the low-capacity of UNCTs in managing quality evaluations, all point to the need that UNEG members take on the first responsibility for conducting UNSDF evaluations.
UNEG should also act as clearing-house to coordinate UNEG members' support to the Member States' country-led reviews, upon request.
National Evaluation Capacity Development
Time has come to define the mandate and scope of UNEG members for NECD, as well as relevant norms and standards, taking into account the need to maintain evaluation units independent and separate from management and programmatic work. The partnership with EvalPartners and the involvement of UNCTs should be key factors in meeting the demand for NECD.
UNEG has an important role to play in the support of NECD at the global normative level and should invest in it.
No-One left Behind
The evaluation community, including the UN evaluation system, has a moral duty to ensure that the voices of those who are left behind, are listened to and heard 'loud and clear' and for making duty-bearers accountable. This raises also practical challenges, as often the vulnerable and marginalized are not easy to identify and reach. Participatory methodologies, unpacking the assumptions of the power relations in evaluations, opening up to the unexpected, are just a glimpse of the possible approaches to conduct more inclusive evaluations
Cross-cutting issues
Human Rights, Gender Equality and Sustainable Development are considered both goals and cross-cutting issues in Agenda 2030 and will have to be integrated as such in evaluations. The UNEG guidance documents on the integration of Human Rights and Gender equality in evaluations are widely used; a new one, specifically focused on the evaluation of the SDG through an equity-focused and gender-responsive lens, is in preparation.
Sustainable Development is a multi-dimensional concept, each dimension requiring diverse sets of methods, tools and time-frames, to be assessed. There is no UNEG guidance on this theme, a gap to be filled.

SDGs inter-linkages
The inter-linkages across the SDGs add to the complexity of working towards and evaluating the Goals. Explicit recognition of complexity is the ‘new normal’, although reality has always been complex in people’s livelihoods and environments; new ways of conceptualizing evaluation in the context of explicit complexity, will be necessary, as well as new or fine-tuned methods and tools
The Universality principle
For UNEG members, taking into account the Universality principle will mean: <ul style="list-style-type: none"> - Fine-tuning methods and tools for the evaluation of Global Public Goods and Capacity development; - Expand the scope of evaluations, to include all Member States as appropriate; and - Develop partnerships with existing National Evaluation Systems in all Member States.
Partnerships
Partnerships will be a constant feature in the evaluations of the UN system, as a modality in their conduct, as a performance criterion and a goal to be assessed. Of particular relevance to this are the experiences and lessons learned from the Inter-Agency Humanitarian Evaluation group coordinated by OCHA; the ISWE mechanism; the evaluations of Multi-Stakeholder Partnerships such as REACH and SUN
The partnership between UNEG, IOCE and EvalPartners is a crucial achievement; all parties need to better understand and agree on the opportunities this opens up and how

5.1 National ownership of Agenda 2030 and its review

National ownership and the United Nations Sustainable Development Frameworks

83. Agenda 2030 and the SDGs firmly state that every UN Member State is responsible ‘*for its own economic and social development*’ and for deciding how to progress towards the SDGs and their targets. Member States should, in the words of the UN Secretary General, ‘*translate the Agenda into a nationally owned visions and objectives, leading to transformative action*’.⁴⁶ National ownership, including by all national stakeholders, is considered to be critical to the achievement of the SDGs and Agenda 2030. Member States called on the UN for its support in this endeavor, each organization according to its own mandate, as partners advocating for, proposing and implementing initiatives inscribed in the national plans and strategies. Member States have also expressed a strong demand for coordinated UN work at country level, as confirmed by the request of 53 among them, for the Delivering as One approach.

84. The UN development system has been active in getting ready for the challenge. In January 2016, the UNDP Administrator affirmed that ‘*we are following the imperative of national ownership, with our actions firmly determined by country needs and national capacities. Our reform efforts within the UN system must be flexible to adapt to country contexts, allowing UN Country Teams either to scale*

⁴⁶ Ibid.

up efforts, or to change course quickly in light of lessons learned'.⁴⁷ On the same occasion, it was clearly affirmed that the DaO is '*the floor, not the ceiling*' for work at country level, and that no-one agency 'owns' any of the SDGs.

85. Among the foreseen strategic actions, is the improvement in the coherence between the development, humanitarian, human rights and peacebuilding agendas.⁴⁸ Another step is the adoption by the UNDG of the Mainstreaming, Acceleration and Policy Support (MAPS) approach,⁴⁹ that entails mainstreaming the SDGs into national policies and plans, accelerating the identification of obstacles and bottlenecks and related solutions and providing '*co-ordinated policy and technical support available from the UN system to countries on request*'. The UNCTs and the Resident Coordinator system (RCs) will be the centre-piece of the UN support mechanism to Member States. This will have a strong bearing on how all UN organizations should operate at country level, including the non-resident agencies and those that mostly provide normative support.

86. A new generation of UN assistance frameworks, renamed UN Sustainable Development Framework (UNSDF), will define the contents of UN support at country level. In early 2016, UN-DOCO circulated for comments, a draft version (1.0) of the new UNSDF guidelines. A version 2.0 of the document should be circulated in April 2016 and tested in 29 countries later in the same year.⁵⁰ The Guidance complements the Standard Operating Procedures (SOPs) issued by UNDG in 2014,⁵¹ and includes a number of significant steps for enhancing the integration of UN work at country level. The 1.0 version includes three core programming principles, namely: i) Human rights and addressing inequalities and discrimination, including for gender equality; ii) Sustainability, reducing environmental risks and increasing resilience; and iii) Accountability, underpinned by strengthened national capacities, robust data and results-based management. It was suggested that country-ownership be added as fourth core principle. Moreover, the document stresses the need for both the alignment of specific organizations' programming documents with the UNDAF and the linking of indicators, baselines and targets to other reporting processes.

87. These changes in themselves, as key steps in aligning the UN work at country level with the SDGs principles, would be sufficient to require adjustments in UNEG members' country-level evaluations. For example, introducing 'country-ownership' among the principles will require integrating a related criterion in country-level evaluations, as well as tools and rubrics for measuring it.

88. In addition, and most importantly, the guidance document also calls for harmonization of both procedures and timelines of individual organizations' country-level evaluations, with the end-of-life

⁴⁷ <http://www.undp.org/content/undp/en/home/presscenter/speeches/2016/01/18/sustainable-development-goal-implementation-the-un-development-system-and-undp-s-roles.html>.

⁴⁸ ECOSOC Dialogue on the longer-term positioning of the UN development system in the context of the 2030 Agenda for Sustainable Development, Second Phase, Briefing Session, 17 December 2015, at <https://www.un.org/ecosoc/sites/www.un.org.ecosoc/files/files/en/qcpr/ecosoc-dialogue-briefing-summary-17-dec-2015.pdf>.

⁴⁹ MAPS: Mainstreaming, Acceleration and Policy Support for the 2030 Agenda, UNDG Concept Note, 26 October 2015.

⁵⁰ Information provided by one interviewee.

⁵¹ <https://undg.org/home/guidance-policies/delivering-as-one/standard-operating-procedures-non-pilots/>.

UNDAF evaluations, and proposes that ‘...an inter-agency M&E/Data support group is responsible for ensuring a sound monitoring system, coordinated agency data collection, analysis and capacity development activities, planning for management arrangements of the UNDAF evaluation and, where needed, providing M&E advice to inter-agency programme, communications and operations groups’. Last, a suggestion was made that ‘evaluations’ be added as one of the tools contributing to the accountability principle.

89. The supply side from UNEG and of central evaluation units to the UNDAFs evaluations, has been mostly at the normative and advisory level. In 2011/2012, UNEG issued in collaboration with UN-DOCO, three guidance documents on the UNDAF evaluations, on ToR, Management response preparation and Frequently Asked Questions (see <http://www.unevaluation.org/document/>). UNEG role in them was defined as, in collaboration with DOCO, to ‘provide guidance (including the UNEG Norms and Standards for Evaluation in the UN System) and maintains a roster of consultants. In particular, those with expertise in UNDAF evaluation’. Agency-specific monitoring and evaluation units could be called upon by UN Resident Coordinators. In practice, collaboration has consisted of the guidance documents, and over the years the occasional involvement of some UNEG members in providing some support and advice to a few UNDAF evaluations.

90. In the assumption that the Agenda 2030 commitment to accountability and measurement of progress will be actually embraced by the UN Member States and the UN system, UNEG members’ detachment from the UNSDF evaluations, with UNEG only in a normative level, will no longer be sufficient. Main reasons are:

- the complexity of the evaluation subjects, taking into account the need to integrate the SDGs principles discussed in this report;
- although the M&E teams could engage in the planning of the management arrangements as foreseen in the version 1.0 of the Guidance, the weak capacity of UNCTs in conducting evaluations first underlined in the UNDAF Evaluability Assessment conducted by UNEG in 2006 does not appear to have been resolved;⁵² and
- the need for a strong common framework for all UNSDF evaluations, to enable comparability for higher-level analysis, including regional and global reviews.

91. In the Agenda 2030 context, moreover, if UNSDF evaluations will not yet be the only country-level evaluations by the UN, they are likely to become the single visible evaluation wherein all UNCT organizations will want to see their contributions and their relevance to country progress towards the SDGs, duly reflected. This, for their own credibility vis-à-vis the countries themselves and for their respective Governing Bodies. And for this, they may want the UNSDF evaluations to be independent, rigorous and credible. If all this hold true, it is possible that the demand from UNCTs for UNEG members to actively participate in the UNSDF evaluations, will increase.

92. Experience from various UNEG members shows that they could not engage with UNDAFs evaluations, as these were “yet another” addition on already busy agendas. Nevertheless, all UNEG members conducting country-level evaluations are fully aware of the frequent overlap of their organization-specific evaluations, and that only limited coordination on timing to avoid over-burdening

⁵² This was discussed in Section 4.3. It remains to be seen whether this observation will be confirmed by the up-coming ISWE-led meta-analysis of the UNDAFs evaluations.

of governments and country-offices, was attempted. A quite conservative guesstimate, based on the incomplete UNEG repository and members' Web sites, suggests that between 2013 and 2015, UNEG members collectively, conducted at a minimum, 15-20 country-programme evaluations/year, some of which in the same countries, in the same year or at one-year distance. Still allowing for the need to maintain internal capacity for organization-focused exercises, part of the resources currently assigned to the country-level evaluations, both human and financial, could be re-allocated to UNSDF evaluations. This would have clearly consequences on the responsibility and leadership for these evaluations, which, in the view of this Review, should reside with UNEG members. This is discussed more in detail in Sections 7.1 and 7.2.

Country-led reviews

93. Agenda 2030 also commits Member States to systematic follow-up and review of the Agenda implementation, with the purpose of ensuring accountability through a '*robust, voluntary, participatory, transparent and integrated*'⁵³ framework, to which the UN system will provide '*active support*'. Of greatest relevance for UNEG members, Paragraph 74 of the Agenda defines the principles guiding the Follow-up and Reviews, which will:

- Be voluntary and country-led;
- Be rigorous and based on evidence, informed by country-led evaluations;
- Require enhanced capacity-building support for developing countries, including the strengthening of national data systems and evaluation programmes, particularly in African countries, least developed countries, small island developing States, landlocked developing countries and middle-income countries.⁵⁴

94. The recent proposal by the UN Secretary General⁵⁵ on a global follow-up and review mechanism, suggests that each Member State should approximately present two country-led reviews to the High-level Political Forum on Sustainable Development (HLPF) by 2030. The first post-2015 HLPF in July 2016 will receive the national reviews of 21 volunteer countries, representing a balanced sample of regional and economic development representation, which are expected to report on all the 17 SDGs.⁵⁶ The purpose of presenting national reviews will be peer-learning through sharing success stories, lessons learned and, as appropriate, also highlight areas where further support and resources may be required. A strong element of accountability should also inform the reviews. The UN Secretariat is facilitating the process, by issuing templates and detailed guidelines for the preparation of the reviews. UNEG leadership had contacted the Permanent Representatives to the United Nations in New York

⁵³ Agenda 2030, paragraph 72.

⁵⁴ See Annex 7 for the entire text of paragraph 74.

⁵⁵ Op.cit.

⁵⁶ See <https://sustainabledevelopment.un.org/hlpf>. Volunteer countries will be: China, Egypt, Estonia, Finland, France, Georgia, Germany, Madagascar, Mexico, Montenegro, Morocco, Norway, Philippines, Republic of Korea, Samoa, Sierra Leone, Switzerland, Togo, Turkey, Uganda and Venezuela. This Review could not find out whether any of the programme countries in this group, has requested support from the respective UNCT to prepare its review.

about the availability of the Group in supporting them.⁵⁷ As of end of March 2016, none of the 21 countries had taken advantage of the proposal.

95. At the time of writing this report, no information was available on the contents of the 21 reviews and on the process each Member State is following for its preparation. Similarly, it is not known the extent to which the first set of reviews will guide the future ones, considering the on-going debate on the whole mechanism (see section 6.1). What appears certain, is that there will be a huge variability in the type of support that Member States will require from the UN system: groups of countries will not require any support; others will ask for some advice and guidance, to different degrees; others will have to be fully supported in the effort;⁵⁸ in others, namely states in the tragic grip of nation-wide conflicts, country-led reviews will hardly be an option.

96. The role of UNEG in this process could be of a ‘clearing-house’ for these requests, coordinating the contribution by those UNEG members who can do so. Over the next two to three years, it looks like a case-by-case approach will be the only way forward. Factors that will affect the whole process will be: the national capacity in evaluation, including the existence of a National Evaluation System; the existence in each country of VOPEs and other stakeholders with whom to partner; financial resources made available by the countries themselves, or by partner countries on their behalf. One important element to enhance the capacity of UNEG and its members to respond to the requests in a timely and effective manner, would be the possibility to plan country-led reviews with sufficient anticipation, e.g. two years, to allow conducting UNSDF or other organization-specific evaluations that would feed into the country-led reviews. An institution with a UNEG-type mandate at the inter-governmental level, would be instrumental in such a process.⁵⁹

97. The changes in the modality of UN work at country level, and its consequences on the conduct of UNSDF evaluations, as well as the possible future requests for support to country-led reviews, need to be discussed and agreed by UNEG members, with a certain level of urgency given the on-going processes within the UN. Section 7.2 discusses a possible scenario, based on the perceived need for UNEG members to take the leadership of UNSDF evaluations.

5.2 National evaluation capacity development

98. The critical role of capacity development in achieving development objectives had already been stressed in the 2005 Paris Declaration on development effectiveness. More recently, it was also strongly emphasized in the UNGA Resolution 67/226 on the 2012 QCPR, whereby the UN is called to support developing countries ‘*to establish and/or maintain effective national institutions and to support the implementation and, as necessary, the devising of national strategies for capacity-building, including policy advisory support, to deal with national and global challenges*’.⁶⁰ Thus, developing capacities at all levels and in all its mandated areas, is at the core of the support the UN system is expected to provide to Member States to achieve the goals they have established for themselves.

⁵⁷ Letter by UNEG Executive Group to all Permanent Representatives, September 2015.

⁵⁸ See next section on National Evaluation Capacity Development in this respect.

⁵⁹ See sections 6.3 and 7.1.

⁶⁰ A/RES/67/226, UNGA Quadrennial comprehensive policy review of operational activities for development of the United Nations system, December 2012; it strongly emphasized the role of the UN in supporting programme countries to develop national evaluation capacity.

99. The same 2012 UNGA resolution had also called on the UN to strengthen its focus on developing capacities in Member States on evaluation and monitoring. The 2014 UNGA Resolution on National Evaluation Capacity Development,⁶¹ made even more explicit the responsibility of the UN system to “*support, upon request, efforts to further strengthen the capacity of Member States for evaluation, in accordance with their national policies and priorities*”.

100. The demand to the UN and UNEG members on NECD is stated to be significant, and requests in this direction are expected from Member States that see a need for support on institutional set-up and evaluation conceptualization, methods and organizational aspects, to properly engage in country-led reviews. Some valuable mapping work has been carried out by UNDP IEO (see below), but a more systematic analysis is still missing.

101. National Evaluation Capacity Development has thus become a prominent topic for UN evaluation units. The interviews conducted for this Review revealed at the same time, diverging views about what the commitment to NECD should entail for UNEG members, and a widespread level of uncertainty on what it should be. Therefore, the Review analyzed the available 33 UNEG members’ evaluation policies⁶² to identify to what extent, NECD, internal ECD, i.e. directed to staff within the entity itself, and/or national ownership for evaluation, were integrated in the respective mandates. This showed the following:

- five UNEG members systematically do or will engage and dedicate significant resources to NECD;⁶³ these represent 9% of the total UNEG membership;
- twelve UNEG members have a commitment in their policies but are not systematically active; these represent approximately a quarter of the total UNEG membership;
- sixteen UNEG members, half of those with an evaluation policy posted in UNEG Web site and 35% of the total membership, have no reference at all to NECD in their policies.

102. Box 3 below lists the UNEG members whose policies include either some reference to national evaluation capacity development, and/or to evaluation capacity development within the organizations themselves and/or national ownership of evaluation.

⁶¹ Op.cit.

⁶² See footnote 20.

⁶³ WFP commitment is embedded in its new evaluation policy, approved in November 2015.

Box 3. National ownership, institutional and National Evaluation Capacity development in UNEG members' evaluation policies			
Evaluation unit	Policy includes NECD in the mandate	Policy includes corporate-level ECD in the mandate	Policy includes support to national ownership of evaluation in the mandate
<i>The GEF</i>	X		
<i>ICAO</i>		X	
<i>IFAD</i>	X	X	
<i>ILO</i>	X	X	
<i>ITC</i>	X	X	
<i>OCHA</i>		X	
<i>UNCDF</i>	X		
<i>UNDP</i>	X		X
<i>UNESCO</i>	X		
<i>UNFPA</i>	X		X
<i>UN-Habitat</i>		X	
<i>UNICEF</i>	X	X	
<i>UNODC</i>		X	
<i>UN-Women</i>	X	X	X
<i>UNV</i>	X		
<i>WFP</i>	X	X	
<i>WIPO</i>	X	X	
Source: UNEG members' evaluation policies, UNEG Web site			

103. Statements in the policies about commitments to NECD vary, from firm and explicit as in the case of UNDP, UNFPA, UNICEF, UN-Women and WFP, to ‘*as feasible*’ in the case of most others. Active commitment to NECD is also prevalent in evaluation functions with decentralized M&E staff. Reportedly, despite their policies, the resources available to several UNEG members for N/ECD are very limited or non-existent. A rapid analysis has also shown that the work actually done under the banner of N/ECD is very diverse, as follows.

104. The work of UNDP IEO on ECD has mostly comprised the organization of international conferences on National Evaluation Capacities, four so far: starting in 2009 in Morocco, the conferences were then held in South Africa (2011), Brazil (2013) and Thailand (2015). These have been important opportunities for a wide range of stakeholders, the majority coming from evaluation units in national governments, to discuss and exchange experiences and views on various aspects of the evaluation function and profession. Occasionally, ECD per se was the focus of some presentations, but mostly, the Conferences were important events where participants enriched their personal, and by extension possibly institutional capacities as well, on specific themes within the evaluation domain. The third conference also issued 18 Commitments to NEC, including enhancing national evaluation capacities and encouraging accountability by calling on countries and NEC participants to commit to actions and collaboration. In addition, as mentioned above, UNDP IEO very recently issued a publication describing 43 National Evaluation Systems (NES), which will serve as the building block for a more comprehensive baseline of NES world-wide.

105. UNICEF Evaluation Office has been very active over the years, with the production of several key guidance books and material either directly focused on how to conduct N/ECD or contributing through information and experience sharing, to N/ECD. In addition, major efforts have gone into the establishment of EvalPartners, a ‘*global partnership to strengthen national evaluation capacities*’, discussed more in detail in the Partnership section. Suffice here to say that EvalPartners is playing a key role in triggering an increase of demand for evaluation and in supporting NECD in many countries.

106. UNFPA is also initiating a process, of strengthening the capacity of its network of country-level M&E Officers and national partners at the request of the UNFPA Executive Board. UNODC included in its 2016/17 work-plan, a pilot initiative to partner with one national ministry and developing its evaluation capacity. And UN-Women Evaluation Office has also been engaged in NECD, through organization of and participation in training events with modules on Monitoring and Evaluation.

107. At a different scale of action is the process ‘No One Left Behind’, launched by UN-Women and EvalGender+ and supported by UNEG and a host of other partners (see section 5.3), which should provide hands-on evidence on how to conduct evaluations through an Equity-Focused and Gender-Responsive lens country level. Should the process proceed as announced at the New York event, it will entail providing technical assistance at country level, through national VOPEs with support from UN Women Regional Evaluation Specialists. This initiative appears highly interesting and the lessons learned through it will be useful, in addition to its explicit purpose, to inform on: how to run a practical collaboration between UNEG members and VOPEs at country level; how UNEG members can work at country level on NECD without undermining their independence by working in the programmatic arena; and on the role of regional UN Evaluation Group Inter-agency Networks (see Section 6.2). In a learning perspective, a final evaluation included in the pilot experience would be highly valuable.

108. UNEG itself has contributed to global ECD through its normative work, by issuing Norms and Standards and a long list of guidance materials on various aspects of evaluation. Also, the preamble of UNEG Norms mentions the request of UNGA to the UN system to conduct evaluations by contributing to national capacity development. On a similar line, the wealth of guidelines and templates produced by many UNEG members on how to conduct evaluations, to be used by consultant evaluators, as well as M&E officers or programme officers who have responsibilities for managing reviews and evaluations, can also be considered a contribution to global ECD.

109. Other UNEG members, with or without the mandate for N/ECD in their policies, have started to engage more systematically with National Evaluation Systems (NES), as members of Reference Groups or Core Learning Partnerships when conducting country programme evaluations, or through NECD initiatives for staff of counterpart ministries. This would be greatly facilitated if UNEG developed one single repository for all NESs contacts in Member States. Furthermore, many UNEG members have privileged an indirect though effective approach to NECD, by recruiting consultants and/or firms in the countries or regions where evaluations are conducted. In the view of JIU, ‘*[This] is not an insignificant effort with effects on mutual capacity development. But more needs to be done in a more systematic and less ad hoc manner.*’⁶⁴

110. This short overview shows that efforts so far have been mostly pitched at the global and normative level, with the exception of some training events on evaluation led or contributed to by a few UNEG members and the ‘in-service learning’ approach. The diversity of contributions to N/ECD by UNEG members, albeit interesting, reflects to some extent, the lack of clarity and harmonized

⁶⁴ Op.cit.

definitions of what their role and responsibility should actually be. It also appears that none of the initiatives conducted so far has been evaluated.

111. At a broader level, beyond the strict membership of UNEG, the Independent Evaluation Group (IEG) of the World Bank also allocates significant financial resources to ECD at the global level through two major initiatives: the Regional Centres for Learning on Evaluation and Results (CLEAR), a *‘collaborative, global partnership that works to strengthen partner countries’ capacities and systems for monitoring and evaluation (M&E) and performance management (PM) to guide evidence-based development decisions*,⁶⁵ and the International Program for Development Evaluation Training (IPDET), that offers training programs on evaluation. A key feature of IEG’s support to ECD is the strong segregation of duties within IEG, between evaluation and ECD work, and the planned external evaluation of both initiatives.

112. Also the OECD/DAC Network on Development Evaluation has committed attention to ECD, through a Task Team for evaluation capacity development; its focus was, for the period 2014-2015, on establishing a stronger evidence base of ‘what works’ in ECD, sharing and identifying opportunities for collaboration, and developing joint initiatives.⁶⁶

113. Other recent voices, including the Fourth International Conference on National Evaluation Capacities in November 2015 with its Bangkok Declaration,⁶⁷ the Concept Paper issued by EvalSDG in 2015,⁶⁸ and the Global Evaluation Agenda 2016-2020 (EvalAgenda2020),⁶⁹ have all advocated for intensive engagement with VOPEs and other partners to develop national evaluation capacities. The Bangkok Declaration for example, identifies 13 different types of potential initiatives and efforts in support of National Evaluation Capacity Development.⁷⁰ Some among these options for example *‘Developing systems to promote transparent follow-up of evaluations recommendations’* and working with ‘young and emergent evaluators’, are already, or can be easily, integrated in the modus-operandi of UNEG members. Other options, on the contrary, e.g. *‘Establishing national evaluation legal frameworks - legislation and policies’*, basically entail policy support work and require skills and competences that are not necessarily those found among UNEG members staff; and it is questionable that they should be.

114. Furthermore, national evaluation capacity development for supporting country-led reviews will require, in a number of Member States at least, a long-term engagement at the three levels of capacity development, i.e. individual, institutional and enabling environment. This is typically a task of policy support and programme functions in UN organizations, that require specific skills, competences and resources.

⁶⁵ See http://www.theclearinitiative.org/Clear_about.html.

⁶⁶ See <http://www.oecd.org/dac/evaluation/evaluatingcapacitydevelopment.htm>.

⁶⁷ Bangkok Principles on National Evaluation Capacity for the Sustainable Development Goals era, November 2015, at: http://web.undp.org/evaluation/nec/nec-2015_declaration.shtml.

⁶⁸ IOCE/EvalSDG, at <http://forum.ioce.net/forum/open-forums/evalsdg/1565-evalsdgs-overview-page-and-concept-paper>.

⁶⁹ http://www.mymande.org/evalyear/global_evaluation_agenda_2016-2020.

⁷⁰ See Annex 8.

115. An important factor in all of this, thanks to the impressive efforts of UNEG leadership in recent years in developing partnerships (see Section 5.7), is that a good number of actors are already active in NECD, starting with EvalPartners, its members and all the affiliated VOPEs. Other ‘natural’ important partners in this endeavor are the Resident Coordinators’ network and the UNCTs with M&E staff from the various agencies at country level, who have already been contacted by UNEG leadership in this respect,⁷¹ the UN Staff College in Turin, CLEAR and IPDET, just to mention a few.

116. Taking all of this into account, and as confirmed by most of the interviewed stakeholders, time has come to define the mandate and scope of UNEG members for NECD, as well as relevant norms and standards. Also, more work on mapping the actual needs of Member States on NECD is required. These are discussed in Section 7.3.

5.3 Leave no one behind

117. The second core principle of Agenda 2030 is the politically charged commitment to ‘leave no-one behind’, which explicitly raises the Member States’ responsibility and accountability on, among others, human rights, gender equality, empowerment, inclusiveness, equity, democracy, right to information and transparency. The principle is mainstreamed in the Agenda, that includes, among the vulnerable ones who are in need, *‘all children, youth, persons with disabilities (of whom more than 80 percent live in poverty), people living with HIV/AIDS, older persons, indigenous peoples, refugees and internally displaced persons and migrants’*.⁷² The SDGs, in turn, make systematic reference to ‘all’. The same commitment is inscribed in the Global Indicator Framework, which foresees indicators to be disaggregated, as relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics.⁷³

118. Member States expect that the UN will respond to the commitment through more effective engagement in least-developed countries and in those affected by conflict as well as targeting the most marginalized groups in countries that are doing better in average growth and social development.⁷⁴ If it is possible to reach the income-poor, the same does not necessarily apply to those who are marginalised and those who marginalize themselves for fear of repression and violence linked to ethnic, political, religious or other discriminating grounds. The default situation is one of lack of data, and resistance from many sides, to identify vulnerable and marginal populations. More efforts and thinking are necessary to overcome the multiple obstacles, discrimination and barriers faced by excluded groups, and to make sure that people in those groups are listened to, and given voice and opportunities.

119. The principle ‘Leave no-one behind’ is also embedded in the foreseen follow-up and review processes that should be, among others, *‘open, inclusive, participatory and transparent for all people and will support reporting by all relevant stakeholders’* and *‘people-centred, gender-sensitive, respect human rights and have a particular focus on the poorest, most vulnerable and those furthest behind’*.⁷⁵

⁷¹ Letter by UNEG Executive Group to all RCs, September 2015.

⁷² Agenda 2030, paragraph 23.

⁷³ See Annex 9.

⁷⁴ ECOSOC Dialogue, Op.cit.

⁷⁵ Agenda 2030, paragraph 74; see also Section 5.1.

120. The evaluation community, including UNEG members, will have a strong responsibility in addressing, and in meeting the call for, equity and inclusiveness, by listening and bringing to the forefront, the voices of those who should be the primary beneficiaries of the implementation of Agenda 2030, and in contributing, by so doing, to holding the Member States and the U, accountable on this principle. The Global Evaluation Agenda 2020 in particular, makes a strong case for evaluation helping to raise the voice of all stakeholders, in particular that of the marginalized and disadvantaged.⁷⁶

121. This concept was strongly re-emphasized during the RBAs' event on the evaluability of SDG-2 mentioned earlier, and during the high-level event followed by a technical seminar in mid-March 2016, organized by UN-Women Evaluation Office and EvalGender+ in partnership with a host of other organizations, with title "No one left behind. Evaluating SDGs with an equity-focused and gender-responsive lens".⁷⁷ The latter event is part of a process launched in early 2016 with a six-week long on-line consultation on four key themes, that also foresees the preparation of a guidance document on Equity-Focused and Gender-Responsive (EF-GR) evaluation, and the testing of the proposed approach in selected countries.

122. The New York event was attended by approximately 100 people with diverse backgrounds in monitoring, evaluation, gender and equity-focused work and analysis and development, and representing a variety of multi-, bi-lateral, governmental and non-governmental organizations, academia as well as parliamentarians.⁷⁸ Participants had the opportunity to listen, share and discuss about the challenges inherent in the implementation and evaluation of the 'leave no-one behind' principle. The firm commitment of participants to their moral responsibility in making evaluation an effective tool for empowerment and accountability was a striking feature of the event.

123. This document is not the proper place to attempt a synthesis of the very rich discussions that took place during the event. However, a few conclusions of the break-out groups, in particular those grounded in the participatory evaluation tradition, are highly relevant to the purpose of this Review:

- a. Evaluators should be mindful of the power balance in evaluation, and its effects on establishing the frameworks: who establishes the boundaries and who is left out; who participates, who speaks and who is listened to; whose indicators and measures are used. In general, evaluators should be open to capture different perspectives, also beyond scope boundaries;
- b. Evaluation evidence should include data at household- and intra-household level, collected from both the head of the households and other household members, as well as about the latter, children and youth; tools can include surveys, focus-groups, interviews, though data generated from unstructured methods, e.g. "walk and talk" approach and

⁷⁶ See <http://forum.ioce.net/forum/global-evaluation-agenda-2016-2020/1770-global-evaluation-agenda-final-version> and Section 5.7.

⁷⁷ The event took place in New York City on 15-17 March 2016. Partner organizations were EvalPartners, Global Parliamentarians Forum for Evaluation, International Organizations for Cooperation in Evaluation (IOCE), Swiss Development Cooperation, UNEG, United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA), and Coneval Mexico.

⁷⁸ This is an estimate of the author of the Review, who was among the participants.

storytelling, can better capture the perspectives and voices of women and marginalized groups;

- c. Evaluation protocols should be adapted to particular contexts to capture the views of those affected, for example, by natural and man-made disasters, Internally Displaced People, refugees, informal housing settlements, nomads, etc.;
- d. There is a need for balancing aggregation versus specificity, with stronger focus on the individual who is being left behind, and on the last mile, rather than on average and coverage.

124. The evaluation community, collectively, has a solid experience with participatory approaches, tools and techniques, stemming from the action-research and participatory approaches developed in the '80s and '90s by a number of development-oriented organizations across the world. These include participatory monitoring and evaluation systems, wherein goals, indicators and targets are established by the communities and groups concerned, who also organize the systematic measurement of progress and achievements. In recent years, these methods have also been applied to the monitoring of the state of natural resources for their sustainable management, producing data that can be effectively integrated into large-scale indicator measurements.⁷⁹

125. When participatory approaches are adopted, often the monitoring and evaluation process in itself is empowering and part of the outcome. Participatory approaches in evaluation are also valid for assessing impacts '*from the point of view of the participants*', as for example in the Qualitative Impact Protocol (QUIP) developed by an initiative funded by the UK Economic and Social Research Council (ESRC) and the Department for International Development (DFID), that uses semi-structured household interviews and focus group discussions to assess impact based on self-reported attribution.⁸⁰ UNEG and its members have conducted similar exercises, and the related experience was consolidated in a UNEG guidance document on impact evaluation.⁸¹

126. Therefore, reviving and fine-tuning participatory approaches, methods and tools for evaluation and systematically developing evaluation frameworks that include the voice of those who typically are left out, should become the norm for all UNEG members. The author of the Review is also aware that the work of some UNEG members, reflecting the mandates of their organizations, is closer and more amenable to include the marginalized and those without, or with limited, voice, as well as to the adoption of participatory and inclusive approaches. Nevertheless, the point made here is that there is certainly room for greater inclusiveness in the design and scope of all evaluations, across the UN spectrum of work.

127. Finally, evaluations that want to ensure that 'No-one is left behind' will have to be conducted in close partnerships with a range of stakeholders, including National Statistics Offices, national evaluators, Civil Society Organisations that work with minority groups or with difficult-to-reach

⁷⁹ A specific case of community-level monitoring of groundwater levels was developed in the State of Andhra Pradesh, India, see <http://www.water-alternatives.org/index.php/alldoc/articles/vol7/v7issue2/247-a7-2-1/file>.

⁸⁰ See <http://www.bath.ac.uk/cds/projects-activities/assessing-rural-transformations/documents/complete-quip-guidelines.pdf>.

⁸¹ Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management, UNEG, 2013, see <http://www.unevaluation.org/document/detail/1433>.

communities, local communities themselves. Whenever appropriate, joint participatory evaluations bringing together national ministries, UN agencies and bilateral partners and conducted with full attention to ensuring the privacy and security of those whose voices and views are sought, should also contribute to increase the visibility of the marginalized ones and their plights, raise awareness among all concerned and strengthened the likelihood of uptake of recommendations.

5.4 Paramount focus on human rights, gender equality and sustainable development

128. The third over-arching principle in Agenda 2030 is about integrating and at the same time, aiming at the achievement of Human Rights, Gender Equality and Sustainable Development, in the process of attaining the SDGs. The Preamble firmly states that the SDGs *‘seek to realize the human rights of all, and to achieve gender equality and the empowerment of all women and girls. They are integrated and indivisible and balance the three dimensions of sustainable development: economic, social and environmental.’* Nevertheless, neither the SDGs, nor Agenda 2030 include any definition for empowerment, gender equality or sustainable development (SD). If these gaps in definitions may be ascribed to the political nature of the SDGs and their targets, they persist even in the more technical Global Indicators Framework, as for example in the case of women’s empowerment.⁸² This is likely to add to the complexity of both implementation and evaluation, given the range of possible interpretation of what is actually intended by these concepts.

129. The follow-up and review mechanism, and evaluations accordingly, should fully integrate the three objectives as well, which should moreover be systematically integrated as evaluation criteria. With respect to Human Rights and Gender Equality, UNEG members already have a significant experience in terms of definitions and integration in evaluation conceptualization and practice. The UNEG Handbook and the Guidance Document on integrating Human Rights and Gender Equality in evaluation, represent at the same time the consolidated knowledge of the members, as well as a key reference on these issues, and appear to be widely used.⁸³ Also, contribution-based evaluative frameworks, such as Outcome Mapping, Outcome Harvesting, and Most Significant Change, are some effective methods in this context. More cross-thematic tools may be required however, for example, composite indices that combine the assessments of gender and equality. The issues raised in the previous section on inclusive evaluations largely apply to these two cross-cutting issues. The on-going work in the context of the No-one left behind process, to provide guidance on evaluation of the SDGs with an Equity-Focused and Gender-Responsive lens, is expected to provide precious detailed methodological suggestions for the international evaluation community.

130. With respect to Sustainable Development (SD), its integration into evaluation practice has on the contrary lagged behind. The concept has several dimensions, and can typically include environmental, institutional, social, technical, economic and financial aspects, each with different time-frames and stakeholders and requiring different analytical frameworks and tools. The Rio+20 Declaration, also known as “The future we want” expands the original definition by the Brundtland

⁸² Target 5b -Enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women, should be measured through Indicator 5.b.1 - Proportion of individuals who own a mobile telephone, by sex.

⁸³ See UNEG Handbook and guidance document “Integrating Human Rights and Gender Equality in Evaluation –Towards UNEG Guidance”, at <http://www.unevaluation.org/document/guidance-documents>.

Commission⁸⁴ as follows: ‘We also reaffirm the need to achieve sustainable development by promoting sustained, inclusive and equitable economic growth, creating greater opportunities for all, reducing inequalities, raising basic standards of living, fostering equitable social development and inclusion, and promoting the integrated and sustainable management of natural resources and ecosystems that supports, inter alia, economic, social and human development while facilitating ecosystem conservation, regeneration and restoration and resilience in the face of new and emerging challenges’.⁸⁵

131. The key elements to be considered when assessing SD were well synthesized in 1996 by the International Institute for Sustainable Development (IISD):⁸⁶

- consider equity and disparity within the current population and between present and future generations, dealing with such concerns as resource use, over-consumption and poverty, human rights, and access to services, as appropriate;
- consider the ecological conditions on which life depends;
- consider economic development and other, non-market activities that contribute to human/social well-being.

132. Moreover, sustainability, similarly to resilience, requires longer time horizons than other parameters, and this has significant implications on measurement methods and identification of contribution and other interferences. Methods and tools for assessing the various dimensions of sustainability, seem to be less readily available, at least within the UNEG community, including on the environmental dimension despite the experience of the scientific and development community on Environmental Impact Assessments.⁸⁷ This should be an area where UNEG should focus its attention, given its relevance. A rapid Web-based search on ‘Sustainable Development’ suggests that some material exists on the topic, including checklists for sustainability,⁸⁸ which may provide a good starting point for further elaboration. Also, there are other institutions working on this subject, as for example appears to be the case of one Technical Working Group of the European Evaluation Society.⁸⁹

⁸⁴ Our Common Future, Brundtland Commission, 1987. Here, Sustainable Development was defined as “development that meets the needs and aspirations of the present without compromising the ability of future generations to meet their own needs”.

⁸⁵ Resolution 66/288. The future we want (Rio plus 20).

⁸⁶ Assessing Sustainable Development: Principles in Practice, Peter Hardi and Terrence Zdan eds, 1996, International Institute for Sustainable Development; <https://www.iisd.org/pdf/bellagio.pdf>.

⁸⁷ The author could not verify this with UNEG members in environment-oriented organizations; it will be a pleasure to change the statement if incorrect. For example, UN-Women is also working on a Feminist Systems Thinking approach, which may prove useful to UNEG membership in due time, although it is not designed as a UNEG product.

⁸⁸ For example, <http://www.sustainablemeasures.com/node/95> and <http://usaidprojectstarter.org/sites/default/files/resources/pdfs/SEC-revised.pdf>.

⁸⁹ See <http://www.europeanevaluation.org/community/thematic-working-groups/twg-2-evaluating-sustainable-development>.

5.5 Interlinkages across SDGs and targets

133. Agenda 2030 defines the SDGs to be ‘*integrated and indivisible*’ and the interlinkages⁹⁰ among the SDGs to be of crucial importance for their achievement.⁹¹ The indivisibility of the SDGs is increasingly being stressed in the political discourse, also with respect to the UN institutional set-up. The emphasis on the interlinkages among the SDGs, makes explicit and visible the complexity of the challenges embedded in the implementation of Agenda 2030 and in the achievement of the SDGs, and reinforces the concept of interconnectedness among the three pillars, economic, social and environmental, of Sustainable Development.

134. On the same line, the Secretary General’s proposal for the global review system, underlines the need for a holistic perspective on progress and obstacles and aims at a mechanism that ‘*promotes a cross-cutting understanding of the significant interlinkages across the Goals and targets*’.⁹² Similarly, interlinkages were and are being sought among the indicators identified for the global indicator framework and further work in this direction is foreseen for the IAEG-SDGs.⁹³ Corollary to this is that cooperation and partnerships among all stakeholders are a must, and that no organization or institution should claim exclusive ownership or responsibility for implementing, contributing or reviewing any one of the Goals.

135. A useful visualization of the inter-linkages across SDGs comes from a network analysis conducted on the targets established for 16 SDGs,⁹⁴ based on their wording which, as well explained by the author, embeds the political discourse informing the targets. This showed that some thematic areas covered by the SDGs are well connected among one another, whereas other parts of the network have weaker connections with the rest of the system, as can be seen in Chart 1 below. Other analyses, based for example on the scientific knowledge about the interactions within the Climate, Land, Energy and Water nexus, or between the energy system, economy and climate change, have typically produced more densely populated networks.⁹⁵

⁹⁰ The term interconnectedness was also used to a limited extent, including in the Inception report for this Review. More recent publications however seem to privilege ‘interlinkages’, which is therefore adopted here.

⁹¹ Agenda 2030, Preamble.

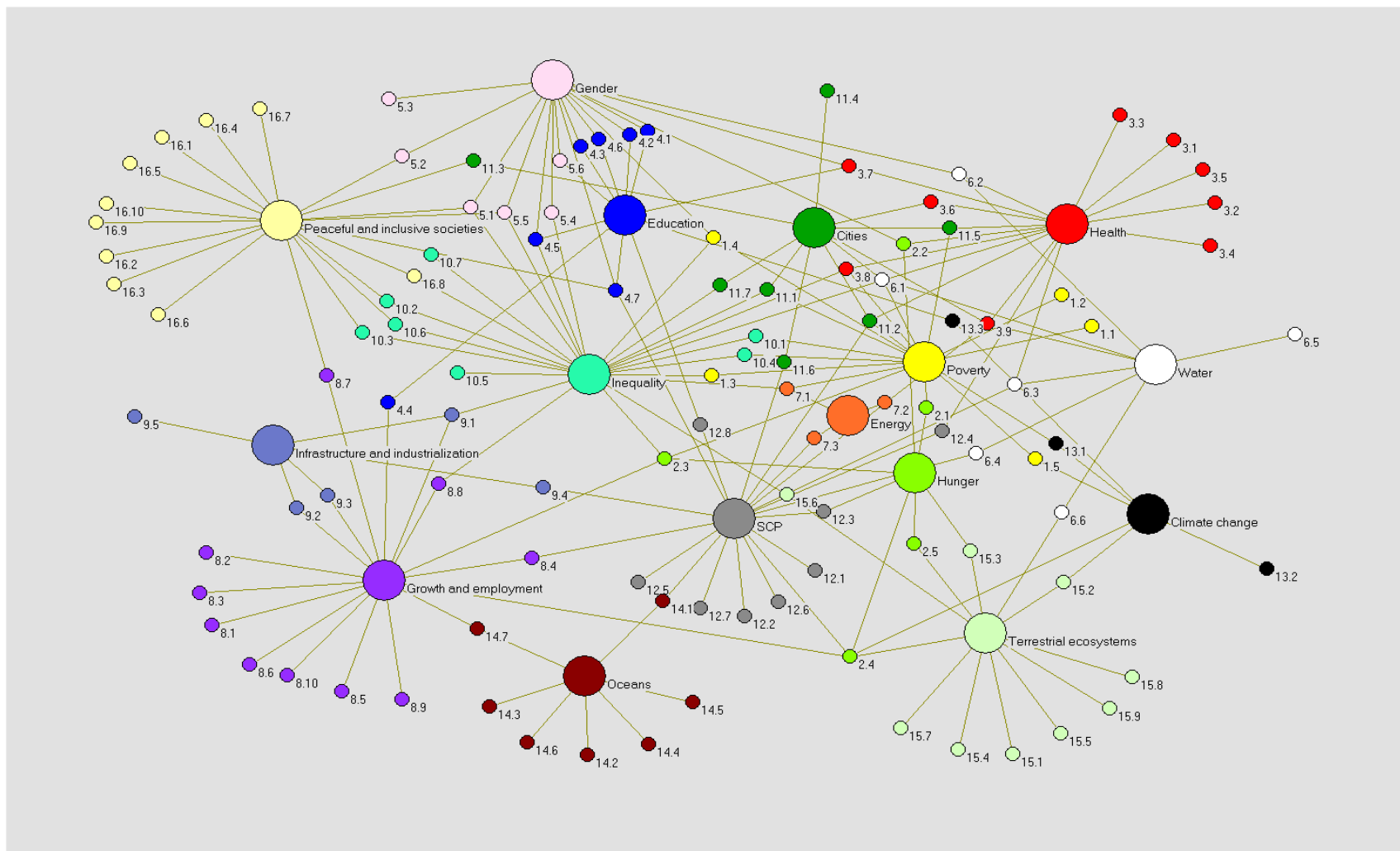
⁹² Op.cit.

⁹³ See Section 6.3.

⁹⁴ Towards integration at last? The sustainable development goals as a network of targets, David Le Blanc, DESA Working Paper No. 141, ST/ESA/2015/DWP/141, March 2015.

⁹⁵ See additional graphic representations of other network analysis in the same paper at http://www.un.org/esa/desa/papers/2015/wp141_2015.pdf; or in <http://peleah.me/category/me-unme/>.

Chart 1 - The SDGs as a network of targets



Source: Towards integration at last? The sustainable development goals as a network of targets, David Le Blanc, UN-DESA Working Paper N. 141, Figure 1.
 Note: targets labels are the numerals which refer to them in the report of the Open Working Group on SDGs.

136. A key conclusion emerging from this, is that the SDGs can only be achieved through actions on, and interactions among, many more parts of the system or elements of the jigsaw, than so far acknowledged in prevailing governmental and UN discourse and structures. It also entails that the understanding and mapping of the interlinkages should go beyond the SDG targets as currently formulated, and explore the actual social, economic and bio-physical interactions at the various level. In other words, it will be necessary to take into account the visible and explicit together with the non-visible and non-explicit contributions and effects, on broader sets of targets than in the Business-As-Usual model.

137. This will be new for organizations that have for long decades, used narrow sectoral frameworks to conceptualize their work, actions and impact pathways. In their perspective, complexity will indeed be the ‘new normal’, as they will have to translate complexity from the conceptual level of the SDGs interactions, into the design, planning and implementation of plans, policies and programmes. Explicit focus on complexity should indeed be welcome; social, economic and natural systems have long been known to be complex and whoever has worked with women and men in households and communities, rural or urban, across the world, or with environmental issues, knows that complexity was artificially left out of policies, programmes and projects.

138. If the UN, and the Member States will indeed embrace complexity in their way of working, which is desirable, then complexity will also squarely “land” in the evaluation domain. The debate about complexity in evaluation is not new. Nevertheless, it is obviously incipient on the specific challenges posed by the SDGs. This Review can only briefly highlight some of the commonly recognized issues so far, and bring attention on possible ways forward.

139. First of all, an observation on the scope of UNEG members’ evaluations. This has been so far mostly pitched on processes and achievements of organization-specific outputs and outcomes, which in any case will retain their importance for corporate accountability and learning even in the Agenda 2030 context. However, UNEG members will have to expand and address through their evaluations, how the work of the UN ‘as One’ will interact - or not - with that of other stakeholders and contribute - or not - to higher level outcomes and impacts on the pathways toward broader sub-sets of multiple Goals and targets. And in doing so, also identify what are the un-intended consequences of the same work.

140. Second, UNEG members have largely embraced theory-based approaches, or are moving in this direction. The Lessons learned by the evaluations discussed in Section 4 also underline the usefulness of developing Theories of Change at the beginning of complex evaluations. However, even single ToCs developed for example for one single target, will be rather complex, as exemplified in the Impact Pathway developed by the global partnership “Girls Not Brides”, for Indicator 3.1 “Proportion of women aged 20-24 years who were married or in a union before age 15 and before age 18” within target 5.3 “Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilation”.⁹⁶ The Rome-Based Agencies seminar on the evaluability of SDG-2 concluded that it will not be possible ‘*to identify a single theory of change against which to evaluate progress on SDG2. Rather, theories of change will come into play only at the level of particular topics, sectors or countries.*’⁹⁷ This means that multiple Theories of Change may be required.

⁹⁶ See the ToC at <http://www.girlsnotbrides.org/wp-content/uploads/2014/12/Girls-Not-Brides-Theory-of-Change-on-Child-Marriage.pdf>

⁹⁷ Proceedings, RBA event on SDG-2, January 2016.

141. In this regard, experts of complexity argue that linear ToCs are not useful to map complex programmes, or to evaluate them. This means for the time being, recognizing that no easy answers or alternatives are available and that more appropriate and accurate tools need to be developed, and in the mean-time, work with ‘second-best’ approaches. For example, if an evaluability assessment indicates that one set of interventions contributing to the same targets or SDGs cannot be evaluated in its entirety, it should still be possible to ‘break it down’ into sizable components for both EA and evaluation. This will require robust evaluation frameworks for all components, that allow sufficient flexibility for data gathering while also enabling consolidation, comparison and cross-cutting analysis at the synthesis phase, to identify the added value of the sum-of-the-parts.

142. Thirdly, evaluations of complex initiatives will clearly require significant reliance on partnerships and a plurality of actors contributing to higher-level evaluations, with evaluations conducted by many parties, in many places, and looking at different things, including processes, partnerships and intermediate outcomes across diverse SDGs. This will require systemic thinking, careful planning and coordination.

143. Evaluation in such a complex environment will be necessarily adaptive, and its concepts and approaches will have to evolve to accommodate the complexity. In addition to what already mentioned above, a few issues and questions that have emerged so far are listed here:

- The SDGs targets and the global indicators framework will be adapted by Member States to national contexts; although a number of similar national frameworks and patterns will emerge, diversity across countries may limit the potential for effectively sharing experiences and lessons learned;
- Up to now, there is no, or very limited, methods and experience in evaluating interconnectedness; in addition to networking analysis for the mapping of inter-linkages;⁹⁸
- Experience so far suggests that monitoring systems have not been systematically used; whenever in place, focus was mostly on performance and delivery measurements; in the SDGs context, harmonization of indicators, development of baselines and systematic monitoring at outcome and impact level over long time-spans, would appear to be of the utmost importance; this would be even more effective, if data so generated could feed into higher levels of aggregation and analysis, and help monitoring changes across the whole spectrum of the SDGs. A recent publication on impact-oriented monitoring and evaluation can provide initial sound guidance in this respect;⁹⁹
- Evaluations at this level of complexity will require significant resources, including financial, expertise and time, that may not be easily available and which may undermine capacity to meet demand.

⁹⁸ See <http://betterevaluation.org/search/site/contribution%2520analysis> and <http://betterevaluation.org/search/site/Qualitative%2520Comparative%2520Analysis>.

⁹⁹ Peersman, G., Rogers, P., Guijt, I., Hearn, S., Pasanen, T., and Buffardi, A. (2016) ‘When and how to develop an impact-oriented monitoring and evaluation system’. A Methods Lab publication. London: Overseas Development Institute.

144. In other words, evaluation of the SDGs and the inter-linkages across them will be complex. It will probably be complicated as well, in particular in terms of coordination and management.¹⁰⁰ Section 7.4 proposes a few key issues to be taken into account in this respect.

5.6 Universality of the SDGs

145. The Universality principle commits all UN Member States to deliver on the SDGs. This is a major departure from the MDGs commitments, where accountability mostly applied to the donor-recipient relationship. It has been stated that this shift towards a collective responsibility for ‘People, Planet, Prosperity, Peace and Partnerships’ and the sharing of the same goals and targets across countries at different levels of progress, may help in blurring the divisions and mark the beginning of a more equal relationship among them.¹⁰¹ This is obviously highly desirable and time will confirm or otherwise; in this, the role of a range of stakeholders, including the UN evaluation system, in “keeping the commitment on track” will be crucial.¹⁰²

146. It has also been stated that the call for a more universal responsibility may lead to a stronger engagement with the UN system, of national domestic sector ministries in High-Income countries.¹⁰³ As all those UNEG members within UN organizations with an important normative function know, national line-ministries already are the decision makers in, or contribute to, the corporate governance mechanisms.¹⁰⁴ Still, the sharing of a single high-level agenda and the strong degree of interconnectedness of the SDGs may possibly contribute to bringing closer to the UN, as well as among themselves and with peer organizations in other sectors, a number of national governmental organizations that previously did not perceive themselves as partners in a global community.

147. This paramount shift in the political vision of what are the global collective responsibilities may have also been facilitated by the large increase since the 2000s, of the number of countries that have

¹⁰⁰ The distinction here between complex and complicated is sourced from Better Evaluation, http://betterevaluation.org/blog/addressing_complexity.

¹⁰¹ Ibid.

¹⁰² The fact that for example, no UN Member State has achieved all the targets on gender equality, may play at the same time as an encouragement for all to do better, but also to do little because in any case, “nobody is there yet”.

¹⁰³ Ibid.

¹⁰⁴ Just to mention a few: FAO, where representatives of national governments elect the Director-General of the Organization at the Conference and sectorial ministries are the decision-makers in its thematic governing bodies; ICAO, that manages the administration and governance of the Convention on International Civil Aviation which applies to all countries; ILO, with its unique tripartite constituency of employers, governments and workers; UNODC, where national ministries actively engage in its Commission on Narcotic Drugs and Commission on Crime Prevention and Criminal Justice; UNEP, where UNEA brings together all Environment Ministries; and WHO, where regional directors for the organization are elected by the ministries of health in each region.

achieved significant improvements in their social and economic development indicators. Box 4 shows the change in the number of countries across the World Bank income categories in 2004 and 2014.¹⁰⁵

Box 4. Number of countries by income category, 2004 and 2014			
Categories, World Bank data	Number of countries, 2004	Number of countries, 2014	Number of programme countries, 2014
<i>Low income (LI)</i>	63	31	31
<i>Lower-middle income (LMI)</i>	54	51	46
<i>Upper-middle income (UMI)</i>	31	52	42
<i>High income (HI)</i>	39	61	11
Source: World Bank data			

148. Among many others, one consequence of this transformation has been the change in the expectations and demands vis-à-vis the UN and the multi-lateral system in general, of many Member States. From the view-point of the UN development system, both the Universality principle and the change in social and economic status, are a challenge to its relevance and role. The increase in the number of UMI and HI countries, and their enhanced institutional absorbing capacity, mean that the normative function of many UN organizations is becoming increasingly more important and useful to a larger share of UN Member States than ever before. The large majority of both LMI and UMI countries are still ‘programme countries’, as reflected by the network of 112 UN Resident Coordinators that currently covers 130 countries, including through multiple-accreditation. Nevertheless, the type of work that UNCTs deliver in the 53 UMI and HI countries is very different from the support provided in LI and in a number of LMI countries, and will progressively transform into the same relationship that UN organizations with a normative mandate have with technical ministries in UMI and HI countries. Increasing attention will thus be required to work conducted on norms and standards and Global Public Goods, capacity development across a broad range of sectors, as well as support to the development of enabling environments. Regional and sub-regional organizations are also gaining in importance in this context.

149. Thus, on the one hand, HI countries may become more engaged with the UN because of the shared goals; on the other hand, the majority of UN Member States have an increasingly converging demand, and most importantly a growing capacity to offer knowledge and experience, for the development, adoption and adaptation of what could be collectively called the Global Public Goods and Capacity Development produced through the UN system. In this, knowledge-dealing and south-south collaboration should increasingly become key modalities of work.

150. Third, as a parallel process, more countries than ever across all income-groups have some form of National Evaluation Systems (NES), at various levels of development and strength, with which UNEG member can interact. The potential demand from MICs to the UN system for strengthening their NES has been discussed earlier in the report and the recently released UNDP/IEO baseline of 43 NES

¹⁰⁵ The difference in the total number of countries is due to both availability of data for some countries and redefinition of a few countries’ boundaries over the period 2004-2014.

provides very valuable information on this, while it does not, and correctly so, assess the evaluation systems in non-programme countries.¹⁰⁶ In many non-programme countries however, NES also exist together with well-structured VOPEs, and these organizations may become important stakeholders in the country-led reviews, beyond and independently from the OECD/DAC Evaluation Group members. Developing good links with NES and VOPEs in all countries should also help in mapping out the evaluation plans at national levels, as well as the gaps.

151. Taking into account these trends, the implications of the Universality principle for UNEG members appear to mostly involve the spheres of methods, scope and partnerships. Within methods, the UNEG guidance document on the evaluation of Normative work issued in 2014 is a very solid basis for evaluations focused on this type of work.¹⁰⁷ Still, there may be a need to develop or fine-tune what is already practiced, to better capture the complexity raised by the focus on inter-linkages and the need to integrate the SDGs principles. With regards to scope, this should be broadened to include in evaluations, the extent to which the relevant UN outputs are used in both programme and non-programme countries and contribute to the achievement of the SDGs. And on partnerships, by establishing these to broaden consultation and collaboration with NES in all categories of Member States.

5.7 Partnerships

152. Partnerships is the last principle identified in the Inception Report as a key feature of the SDG process with relevance for the UN evaluation functions. Grounded in the Rio Declaration and Agenda 21, the concept of Partnership permeates the entire Agenda 2030. It is included among the areas of critical importance; and a revitalized ‘Global Partnership for Sustainable Development’, initially launched in Rio in 1992, is one of the means for the implementation of the Agenda. At its core, is the recognition that Sustainable Development, and the SDGs, can only be achieved through active and collaborative engagement among all stakeholders, including governments, academy, private sector, civil society, citizens at large, the UN, other international organizations and multi-stakeholder partnerships.¹⁰⁸

153. Partnerships are thus a goal in themselves, products of cooperation and solidarity, and a precondition to tackle the complex and multidisciplinary challenge represented by the SDGs. Multi-sectoral and multi-stakeholder collaboration will be required among governmental and non-governmental agencies to make progress in implementing Agenda 2030, as well as cooperation through and across many existing global partnerships.

154. In cascade, complex partnerships and collaborations in implementation will necessarily require “coalitions for evaluation”, that should enable shared and transparent ownership of processes, results and learning; mutual accountability; inclusiveness of diversity of perspectives; and rigorous mapping, interpretation and analysis of stakeholders’ contributions and respective results to progress. A strong partnership-based approach will also be the only possible way forward for UNEG members to meet

¹⁰⁶ There are two exceptions, the Russian Federation and Saint Lucia, which do not host Resident Coordinators; however, both are programme countries for some UN organizations.

¹⁰⁷ UNEG Handbook for Conducting Evaluations of Normative Work in the UN System, <http://www.unevaluation.org/document/download/2014>.

¹⁰⁸ These are partnerships embedded in global frameworks that cut across several SDGs. Examples are the Scaling up nutrition (SUN) initiative and possibly, the Committee on World Food Security.

their responsibility and contribute to the evaluation of the SDGs, from supporting Member States in the conduct of country-led reviews to having a major role in regional, global and thematic evaluations.

155. Partnerships for the majority of UNEG members are not new. Most evaluation policies refer in various ways to partners as internal and external stakeholders for both accountability and learning; a number of policies foresee establishing Core Learning Partnerships bringing together representatives from the organizations themselves, governments, other organizations participating in the initiatives being evaluated, to enhance credibility and ownership of evaluations. Joint evaluations are frequently mentioned as well, as the preferred modality for evaluating joint programmes. Partnerships are also a means to improve implementation, thus a criterion for performance assessment, and goals in themselves of the evaluation subject. Some policies also include specific questions for the evaluation of the partnerships' use and effectiveness.

156. UNEG, an example of professional partnership in itself, has also entered into many partnerships in its own right, on the basis of its independence and credibility. Examples include the evaluation process of Delivering as One, the long-lived joint task force with the OCED/DAC Evalnet on Peer Reviews, and the current engagement in ISWE. To these partnerships, UNEG effectively provided/s the best support that can be collectively generated by the UN evaluation functions.

157. Experience of UNEG and its members in partnerships is thus significant and diverse. In particular, joint evaluations have been more frequent for some UNEG members than others and that typically, although not exclusively, did and will assess 'explicit joint work'.¹⁰⁹ Although most were conducted in the field of humanitarian aid interventions,¹¹⁰ joint evaluations are definitely not uncharted territory for many members. The evaluation of the Paris Declaration discussed earlier in the report, fully built on a partnership framework and provides some important lessons for similar complex endeavors. The Inter-Agency Humanitarian Evaluations (IAHE) coordinated by OCHA have made significant progress, in the view of participants, towards improving evaluative efforts in terms of efficiency, effectiveness, synergies and quality, and is a robust example to get inspiration from. The UNEG Resource Pack on Joint Evaluations developed in 2013, brings together the experience of members on joint evaluations up to that point in time, and it should not be hugely complex to update it, as considered necessary, with the additional experience gained since.¹¹¹

158. There is little doubt that in the context of the UN support to Agenda 2030 implementation, joining forces among UNEG members will become 'a must' for assessing the UN contribution to the progress towards the SDGs, at national, regional, thematic and global levels. Also, as the ToRs for this assignment point out, evaluation resources are scarce, and although coordination and collaboration entail high transaction costs, joint evaluations are an adaptive measure if coordination mechanisms can be designed to operate in a light, transparent and efficient manner. Partnerships with other evaluation stakeholders will also likely be necessary, to strengthen with their inputs the evaluative evidence at the various levels of analysis and thereby, compensate the risk of missing important details, which occurs when evaluations are pitched at a high strategy level.

159. In parallel with joint evaluations, UN system-wide evaluations should increasingly play a role in the assessment of the UN contribution to the Agenda 2030 implementation at regional, thematic and

¹⁰⁹ Joint evaluation of non-joint work presents additional challenges.

¹¹⁰ An assessment of joint evaluations was not part of the scope of this Review.

¹¹¹ Resource Pack on Joint Evaluations, <http://www.unevaluation.org/document/download/2111>.

global level, as urged by the 2012 QCPR. At the time of writing this report, the ISWE mechanism was conducting two pilot system-wide evaluations: a meta-analysis of the UNDAF evaluations already mentioned above, and a Comprehensive System-Wide Evaluation of Capacity Development in Statistics. Lessons learned from both pilots and from the ISWE experience as a whole, will have particular relevance for the future SDG evaluation architecture and modalities.

160. Last but definitely not least in this short review of UNEG and its members' experience with partnerships, EvalPartners emerges as a unique initiative. Launched by UNICEF with the International Organization for Cooperation in Evaluation (IOCE) in November 2012, EvalPartners is a global network that, together with its thematic sub-networks, has become an important voice in the international evaluation community, with a significant convening and outreach capacity. The partnership as a whole brings together more than 50 organizations, including VOPEs, International Non-Governmental Organizations, foundations, bilateral development agencies, among others. The growth of EvalPartners has been rapid and steady also thanks to, in the view of the team of evaluators who assessed its performance in 2014/15, its capacity to meet a growing need for representation of the evaluation profession and to its contribution to both strengthening the demand for professional evaluation and its offer.¹¹²

161. UNEG and some of its members, are among the 14 core partners of EvalPartners. Recently, a UNEG Vice-Chair has become the co-chair of EvalPartners, as a tangible sign of UNEG's commitment and engagement with EvalPartners. This is opening the way to a range of possible collaborations based on the mandates, potentials, comparative advantages and resources from all members, all of particular relevance considering the need for a multiplicity of evaluation efforts at the national and higher levels in the context of Agenda 2030. The "No one left behind" process discussed earlier in the report (see Section 5.3) is a good example of how this partnership can leverage resources and its potential outreach.

162. Always within the EvalPartners umbrella, and in addition to the already mentioned EvalSDGs, the Global Parliamentarians Forum for Evaluation was also launched at the Global Evaluation Week in November 2015, in the wake of a sequence of events that started in South Asia in early 2013 with the establishment of Parliamentarians Forum for Development Evaluation (PFDE), and promises to be an important building block for raising demand for accountability and learning through professional evaluations at the national level. For example, as emerged in the No-one left behind event, the PFDE members can also be active promoters of equity and gender equality issues in the SDGs framework, including in evaluation.

163. In January 2016, EvalPartners also issued, in collaboration with a number of main partners¹¹³ and after more than one year of consultation with evaluators across the world, the already mentioned Global Evaluation Agenda 2020.¹¹⁴ The Agenda is built around four essential dimensions of the evaluation system, each articulated in sub-sets of higher level outcomes: 1) the enabling environment for evaluation, 2) institutional capacities, 3) individual capacities for evaluation, and (4) inter-linkages among these first three dimensions. Each member of the global evaluation community is invited to

¹¹² See [http://www.mymande.org/sites/default/files/files/EvalPartners-Evaluation-Full-Report-\(2015-02-03\).pdf](http://www.mymande.org/sites/default/files/files/EvalPartners-Evaluation-Full-Report-(2015-02-03).pdf).

¹¹³ These were: IDEAS, IOCE, UNEG and the evaluation offices of GEF and UN-Women.

¹¹⁴ Op. cit.

identify what “*pieces of the apple it will chew off.*”, with the hope that the collective effort will lead to the achievement of most, or all the outcomes.

164. Interviews conducted during this Review have, on the one hand, underlined the potential that the close association between UNEG and EvalPartners offers, for both expansion and outreach beyond the current institutional boundaries, and heightened profile and increased visibility. On the other hand, although information on EvalPartners is systematically shared, there is lack of clarity and understanding among UNEG members, of what it is, what it does and how interaction and collaboration can occur. For example, although it is far too early to make any assessment of the extent to which UNEG members have taken on board the Global Evaluation Agenda 2016-2020, the fact that none of the UNEG interviewees for this Review, made any reference to it, is somewhat indicative of the limited awareness about it.

165. Therefore, there is a strong need for a more in-depth debate within UNEG about what the association with EvalPartners can bring to the members, and how. The New York event has been a good start. Nevertheless, more examples, and at a more concrete and practical level, e.g. through joint pilot initiatives, are necessary to enable effective buy-in and collaboration between the members of the two platforms. Areas that appear particularly promising are: i) networking with UNCTs to meet the demand for national evaluation capacity development; ii) raising awareness about evaluation among Parliamentarians; iii) production of joint guidance documents; iv) collaboration with Civil Society through Major Groups members to enhance the use of evaluation for learning, accountability and ‘transformational change’.

166. Two other sets of partners have emerged during this Review, as important players to be taken into account in the evaluation of the SDGs and Agenda 2030. The first are the so-called Multi-Stakeholder Partnerships (MSPs), that comprise global movements, initiatives, programmes and others, who are important players at global and country level on one or more SDGs and on cross-cutting issues. Some of these stem from an initial partnership of UN organizations that opens up to others, for example the “Renewed Efforts Against Child Hunger and undernutrition initiative” (REACH), or originate from a mix of UN, governments, civil society, non-profit and private sectors, for example the Scaling Up Nutrition (SUN), or directly from the highest political level in the UN as is the case for the “Every Woman Every Child” movement. Governance, funding and internal organization of this wide array of MSPs vary hugely. Their importance in the evaluation of Agenda 2030 stems from their role as partners of the UN and of Member States, in contributing to achieving the SDGs, while maintaining separate accountability frameworks.¹¹⁵ The REACH initiative was recently evaluated as a large joint evaluation with a multi-agency management group. It proved to be a successful example of collaboration among of several evaluation offices, based on an EA, clear governance and quality assurance system fully aligned with UNEG N&S and related guidance, all of which helped ensuring its efficiency and effectiveness. Conversely, the SUN initiative, closely linked to the UN at the management level, commissioned an evaluation of its performance without any significant consultation with the evaluation units of the UN organizations contributing to SUN. To avoid similar situations in future, it would be important that UN Senior Management engaged in these MSP, ensured that UNEG and its members as appropriate, were part of the respective Evaluation Management Groups.

167. The second set of partners comprise the Major Groups (MGs), which originated in Agenda 21 and were revitalized in the Rio+20 process. Agenda 2030 identifies nine MGs, namely Women,

¹¹⁵ In this, their role appears to be not too distant from private foundations, e.g. the Bill and Melinda Gates Foundation, that are also fully engaged in pursuing the Goals.

Children and Youth, Indigenous Peoples, Non-Governmental Organizations, Local Authorities, Workers and Trade Unions, Business and Industry, Scientific and Technological Community, and Farmers. MGs bring together organizations and associations at the global, regional and national levels and represent huge memberships, in the order of hundreds of millions. The collaboration with the UN system is for some a long-standing feature of their activity, as for example is the case for the Workers and Trade Unions MG with ILO since the foundation of the Office, or more recent, as in the case of the Farmers MG with the Committee for World Food Security.

168. The MGs actively participated in the process leading to the adoption of the SDGs and Agenda 2030, through direct comments, lobbying and advocacy. They are also expected to report on their own contribution to implementation as well as in the country-led reviews. UN-DESA supports their participation and is currently facilitating their involvement through capacity-development initiatives.

169. Interviews conducted with a few MG representatives during this Review indicated strong experience in advocacy and negotiation and a powerful capacity to reach out to their memberships at the different levels. Expectations vis-à-vis the UN is that it will continue supporting their participation and visibility in the Agenda implementation, including in the monitoring process and country-led reviews, where they need support and capacity development. It would appear that their multi-tier structures, vast memberships, as well as the respective mandates, both thematic and cross-cutting with respect to the SDGs, make the MGs important partners with whom UNEG and its members, as well as EvalPartners and its members, should engage with, in the context of the country-led reviews. MGs could also contribute to the identification of evaluative gaps and themes for evaluations at the regional, thematic and global level.

6 Implications of the Agenda 2030 follow-up and review mechanism

170. This section briefly discusses the most important on-going inter-governmental processes that have a relevance for the future UN evaluation architecture, the enhanced attention to the UN regional dimension and the role of the UN Regional Economic and Social Commissions and the respective Fora on Sustainable Development in this, and the progress made by the UN Statistical Commission on the Global Indicators Framework.

Key messages

Inter-governmental negotiations were on-going as of Spring 2016, among the UN Member States on the global follow-up and review mechanism, including on the country-led reviews, and on the future structure of the UN development system. These are opportunities to firmly position the UN evaluation system in the mechanism.

The regional level will play a key role in the implementation and review mechanism of Agenda 2030; the UN evaluation function should be strengthened at this level, to contribute relevant evaluative evidence to the Regional Fora on Sustainable Development.

The indicator framework at the different levels, should be used as benchmarks for evaluations to validate and/or challenge.

The evaluation community should develop closer links and collaboration with the Statistics System at global, regional and national level, to: pursue joint evaluation capacity development;

fine-tune, add or revise indicators at the various levels; and validate the sources and quality of Data for Development including Big Data.

6.1 Inter-governmental processes of relevance to the UN evaluation system

171. The commitment to country-led follow-up and review processes is a key principle of Agenda 2030, as already mentioned. In January 2016, in compliance with paragraph 90 of the Agenda itself, the UN Secretary General submitted to the General Assembly a proposal for a global mechanism for follow-up and review. This includes an institutional architecture, defines roles and responsibilities and suggests a tentative work-plan for Member States to assess and discuss progress and challenges in the implementation of the Agenda, at the High-level Political Forum on Sustainable Development (HLPF), that will meet under the auspices of ECOSOC every year, and of the General Assembly, every four years.¹¹⁶

172. The SG's proposal reinforces the principles for the reviews enunciated in paragraph 74 of the Agenda, confirms the national ownership of country-level assessments, and lists a number of desirable features for these, including focus on means of implementation, inclusiveness, use of existing platforms, and rigorous, data and evidence-based nature. In addition, the proposal also suggests the alignment of the discussion themes for the HLPF and ECOSOC, and a possible planning approach to in-depth assessments, based on annual sub-sets of SDGs over a four-year period, with SDG 17, Means of Implementation, being under constant monitoring and review. Reporting is suggested to take place through the Secretary General Annual report on the SDGs, and the Global Sustainable Development Report (GSDR), whose frequency was still under discussion.

173. The President of the GA, following an informal discussion with Member States in early February 2016, launched an informal consultation process led by the Permanent Representatives of Belize and Denmark to respond to requests for '*further clarity...on the global follow-up and review framework*'.¹¹⁷ This consultation, on-going at the time of writing with a tentative deadline of end of May 2016, should issue a draft Resolution to be submitted to the HLPF, ECOSOC and UNGA. The consultation is also extended to Stakeholders, including Major Groups,¹¹⁸ and is supported by the UN Secretariat. Key matters of concern appear to be: safeguard of the commitments made in Agenda 2030 including strengthening the indivisibility of and ensuring the balance between the Goals; a need for

¹¹⁶ Op.cit.

¹¹⁷ Letter from the President of the General Assembly on the Secretary-General Report (3 March 2016) at https://sustainabledevelopment.un.org/content/documents/98163-Mar_2030-Agenda-Follow-up-and-review.pdf.

¹¹⁸ Roadmap by the co-facilitators, Denmark and Belize, for the informal consultations on the Follow-up and Review of the 2030 Agenda at the global level, at <https://sustainabledevelopment.un.org/content/documents/9884RevisedRoadmapFURprocess.pdf>.

greater clarity on reviews of thematic and cross-cutting issues; close coordination and harmonization between this and the other on-going high-level processes.¹¹⁹

174. The second important process that may trigger important changes in the UN Development System institutional and operational architecture, and therefore on its evaluation, is the ECOSOC Dialogue initiated in 2015, which will also feed into the UNGA Resolution on the QCPR to be issued in December 2016.¹²⁰ As stated in a recent briefing session of the Dialogue,¹²¹ and as confirmed by a number of interviewees for this Review, there is a need for a change of paradigm for the UN development system, including a complete re-thinking of its role at country level, of the way it collaborates with the Member Countries and across the system itself. Most Member States apparently recognize the need for a major change in the way the UN Development System is structured and operates, and see the interlinkages of the SDGs as a crucial opportunity to act on the fragmentation of the UN system. Differences in views seem to mostly reside in the desired pace for change. The UNGA Resolution on the 2016 QCPR should clarify the pace of the process and its direction.

175. In the light of the above, it is of the utmost importance and urgency that UNEG takes the opportunity of these on-going processes, and in particular of the informal consultation on the global follow-up and review mechanism, to firmly include the UN evaluation function in the agenda for the role that competes to it.

6.2 The regional dimension

176. The regional dimension has been an important building block of the development discourse at least since Agenda 21 was issued in 1992, including in UNGA resolutions. Agenda 2030 and the SDGs make extensive reference to this level, as the logical interface where the global and national levels can more easily interact and be harmonized. Regional reviews and regional-level peer-learning exercises are also envisaged in Agenda 2030, and the regional indicator frameworks should feed into the global indicators framework. Some research institutes have already engaged in conducting trend analyses at the regional level.¹²²

177. The five UN Regional Commissions have an important role to play in supporting Member States in implementing, monitoring and reporting on Agenda 2030.¹²³ Each Regional Commission has established, in 2014 or 2015, a Regional Forum on Sustainable Development (RFSD) as a platform for fostering regional voices on the implementation, follow-up and review of the 2030 Agenda, and

¹¹⁹ See statements by the EU, G77 plus China and Japan, at <https://sustainabledevelopment.un.org/hlpf/follow-up>.

¹²⁰ See http://www.un.org/en/ecosoc/qcpr/pdf/dialogue_roadmap.pdf.

¹²¹ Op.cit.

¹²² Just to mention one, the Overseas Development Institute (ODI 2015) presented a regional SDG Scorecard projecting trends across key dimensions of the SDGs to determine areas in which the fastest acceleration of progress will be required.

¹²³ The five commissions are: ECLAC, Economic Commission for Latin America and the Caribbean; ECE, Economic Commission for Europe; ESCWA, Economic and Social Commission for Western Asia; ESCAP, Economic and Social Commission for Asia and the Pacific; ECA, Economic Commission for Africa.

providing a clear link with the High-Level Political Forum on Sustainable Development, ECOSOC and the General Assembly.

178. The RFSD should also be the platform where regional reviews, built on the national level, should be discussed and prepared, providing inputs for the annual HLPF, and where regional peer-learning should be facilitated. Until now, the case of ESCAP, which established the RFSD in 2014, indicates that discussion has focused on evidence based on statistical data. A proposal to include evaluative evidence in the RFSD discussion of progress in the implementation of Agenda 2030, should be submitted for decision to the Member States.

179. A number of UN organizations have regional offices and have established at that level, monitoring and evaluation positions, usually reporting to the regional directors.¹²⁴ Staff in these posts, together with the evaluation staff in the UN Regional Commissions, have contributed to establish in recent years a number of UN Evaluation Group Inter-agency Networks, to promote an evaluation culture, contribute to UN coherence on evaluation at regional level, and strengthen regional evaluation capacities among UN agencies and their partners. As of March 2016, the UN Evaluation Development Group for Asia and the Pacific (UNEDAP), the Nairobi Interagency Evaluation Network (NIEN) and the UN Inter-Agency Working Group on M&E for Latin America and the Caribbean were active. These regional networks aim at harnessing the core capacity of the UN at the regional level in M&E. This makes them the ideal institutional actor in networking with regional organizations, including the Regional Parliamentary Fora that already exist in four regions; the regional evaluation organizations and the regional members of the Major Groups.

180. Therefore, it appears important that the UN evaluation system becomes more effective at the regional level, and in particular in providing evaluative evidence to the RFSD. The evaluation units of the Regional Commissions and the UNEG regional networks should have a role in this. One useful inputs in this discussion might be the upcoming OIOS evaluation of the UN Regional Commissions.

6.3 The global indicator framework

181. The United Nations Statistical Commission was mandated by the UNGA to develop and implement the SDG Global Indicator Framework. In 2015, at its 46th session, the Commission established, with a view to a long-term engagement, the Inter-Agency and Expert Group on Sustainable Development Goal Indicators (IAEG-SDGs), comprising representatives from National Statistics Offices (NSOs) and, as observers, representatives from NSOs of countries not members of the Expert Group, as well as from regional and international organizations and agencies. The IAEG-SDG worked through a transparent and consultative process to develop a global indicator framework, which was approved by the Statistical Commission in March 2016.¹²⁵

¹²⁴ In the case of UN-Women, the regional evaluation advisors report directly to the Evaluation Office in headquarters.

¹²⁵ This and the following two paragraphs extensively draw from the Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators, E/CN.3/2016/2/Rev.1, 19 February 2016 and supporting documents.

182. The framework includes 230 indicators for the 169 targets, with some indicators linked to two or more targets.¹²⁶ Three tiers of indicators have been established:

- Tier I, for which an established methodology exists and data are already widely available;
- Tier II, for which a methodology has been established but for which data are not easily available;
- Tier III, for which an internationally agreed methodology has not yet been developed.

183. The IAEG-SDGs clearly stated that more work will be required over the years, to refine and improve indicators as knowledge progresses. This will include, for example, the development of indicators to measure global progress in the implementation of the Sendai Framework, as currently the task of by an open-ended working group established by the UN General Assembly.

184. Measurement of the indicators at the global level will rest on regional, thematic, national and sub-national indicator frameworks. Member States, taking inspiration from the global and thematic indicator frameworks, will develop and establish their national indicators and targets, and in the spirit of Agenda 2030, produce data through their NSOs. It is foreseen that significant capacity development is required in this respect. International organizations, based on their respective existing mandates and/or expertise, will compile regional and global aggregates of data against the global indicator framework. These will be consolidated and fed into SDGs reporting system mentioned above (see Section 6.1). Discrepancies between national, regional and global data will be reconciled as possible, or at least explained.

185. A major implication for UNEG members of the indicator frameworks at the various levels, is that these could, and should be used to benchmark progress and accomplishments, as relevant to and feasible depending on each evaluation subject. In addition, in a significant number of cases, UN evaluations should aim at complementing the quantitative measurements conducted by NSOs, or even challenging them, by unpacking and exploring the causal relationships between policies, interventions and results, and by identifying success stories, best practices, gaps in coverage and implementation, unexpected positive and negative effects. In doing so, it will be of the utmost importance that UN evaluations put in practice the commitment to equity and inclusiveness, and give voice to those groups of stakeholders, whose views and perspectives, or indicators, may not have been captured in the official indicator frameworks and data.

186. Another opportunity in this respect is the expansion of innovative approaches to data gathering through mobile technology, to canvass perceptions and views of large numbers of people about, for example, implementation of SDG-related national policies and programmes. The use of ‘unstructured data’ also known as Big Data, will allow access to information at a reasonable price, while raising at the same time, a number of issues about privacy and confidentiality. Also, as pointed out during the New York event, *‘disaggregation is often not granular enough to capture variation within groups and among marginalized people’*.

¹²⁶ See Annex 9.

187. In addition, the use of global, regional and thematic indicators as benchmarks for evaluations pitched at these levels should also be possible, although unpacking and exploring causal relationships at these levels might be more complex than at the national level.

188. The prominent role in Agenda 2030 to the Statistics Systems, at global, regional and national level, makes of partnerships between these Systems and the evaluation community, a ‘must’. The purposes of collaboration are multi-fold, including:

- a. establishing a dialogue to enrich both sides as the overall goal is very similar, i.e. understanding what happens in people’s lives and in the environment;
- b. using statistical data for benchmarking, unpacking, analyzing, providing evidence and if necessary challenging the relevance of the indicators and the data itself; and
- c. assessing the rigour and compliance with quality standards, of the Big Data available through mobile and geospatial mapping technologies.

189. Finally, it is important to mention that in the views of some UNEG members, the structure of an Inter-Agency and Expert Group, such as those supported by the UN Statistical Division on SDG Indicators and on Gender and Statistics, appears to be an interesting and inspirational model for the evaluation function. This is discussed in detail in Section 7.1.

7 The way forward

190. This section draws on the discussion in the report, starting with an analysis of the evaluability of the SDGs, followed by the proposal of options for an overall evaluation strategy and architecture, and by more specific and focused issues. The Review proposes 13 Suggestions to UNEG and its membership. In this respect, it is important to remember that the political decisions on the future follow-up, review and evaluation architecture of Agenda 2030 are still ‘in flux’ and possibly subject to significant changes from what initially proposed. This had a bearing on the scope and precision of what this Review could suggest. Also, evaluation in the SDG era will mean dealing with 193 Member States, each diverse and unique with different priorities, plans and. Thus, a number of legitimate and important questions that were raised by the Working Group members, could simply not be answered here, and will have to wait until some progress is made in the implementation of the Agenda. This clearly does not mean that UNEG and its members should sit and wait.

191. Agenda 2030 and the SDGs exert a strong urgency on the UN system, and therefore on UNEG members, to act in a different manner, and this will require efforts on various fronts. The sub-sections and suggestions follow a ranking of importance, from the most urgent and over-arching, to those that are less urgent in nature and focus on more practical and operational matters. Still, UNEG members should attentively consider all of them, and decide their own priorities also depending on the resources available, as all are necessary inputs and building blocks, in the view of the author, for the efficiency, credibility and usefulness of the UN evaluation function in future.

192. The author of the Review is also aware that in November 2013, UNEG approved its Strategy for the period 2014-2019. This establishes four Strategic Objectives, each articulated through a Theory-of-Change accruing to the overall Vision and Mission for the Group. The Strategy also foresees the possibility of conducting a mid-term review in 2015/16. It will be a decision of UNEG Heads whether what accepted, among the suggestions in this Review, can be accommodated within the existing

framework. Or whether UNEG strategy and operational structure should be re-visited, after taking stock of both the implications of Agenda 2030 on the UN evaluation function, and of the achievements, strengths and weaknesses of the current approach.

7.1 On the evaluability of the SDGs and a UNEG strategy

193. The Review argues that the question about the SDGs evaluability can be answered positively, despite the complexities and challenges inherent in the task, and that the SDGs can and should be evaluated, by focusing on “**how the SDGs can be evaluated**” and on what the UN evaluation system can do, to fulfill its mandate.

194. A first step in this direction, using Davies’ framework for Evaluability Assessments, suggests that the credibility and reliability of the evaluations of the SDGs, will depend, among others, on the following:

- a. the Design of the interventions: how Member States and their partners, including the UN, will explicitly articulate their high-level Theories of Change about the causal relationships between their strategic frameworks and plans, at global, thematic, possibly regional, and country levels, and the achievement of the SDGs; this will include the extent to which these same frameworks will integrate the Agenda 2030 principles and will align targets and indicators (see the following bullet point);
- b. the availability of baseline and monitoring Data for the SDGs targets and indicators at national, regional, thematic and global level, as per the indicator frameworks that Member States will develop, drawing from the regional, thematic and global frameworks approved by the UN Statistical Commission in March 2016, and their future revisions; data for a number of global level indicators will take some years to develop, which will limit the possibility, at least initially, to use robust quantitative evidence against which to measure progress (See Section 6.3);
- c. the Demand for evaluation of the SDGs within the UN governance system, and on which aspects of the process and progress; this will likely change significantly across organizations, countries and over time (see Section 6.1). In this respect, it is important to note that, differently from the MDGs context, the strive for accountability included in the Agenda 2030 firmly demands that evaluations be conducted systematically and comprehensively, of the performance, progress and results of the initiatives that will be implemented for the achievement of the SDGs.

195. The extent to which UNEG members can act, or interact with, on the three Ds above will differ. With respect to Design of strategies, policies, plans and programmes, Sections 3.2 and 7.5 do provide some evidence and suggestions, respectively. With respect to Data, the scope for action will reside in the use of targets and indicators as benchmarks and the added value of the UN evaluation system will reside in its ability in validating, or challenging as required, the information available from the Statistical Systems at the various levels. With respect to Demand, the use of evidence originated through independent evaluations, as a contribution to the design of national SDG-related policies and programmes should be strongly advocated for by parliaments, Civil Society and national evaluation associations, and the UN where appropriate.

196. Another key question that was raised during the Review was “what type of evaluations will have to be carried out to evaluate the SDGs”. An attempt at breaking down the ‘big picture’ of the SDG evaluation, on the basis of evaluation purpose, led to the following groupings:

- a. Evaluations of the contribution of the various stakeholders, namely Governments, UN and other multi- and bilateral organizations, Multi-Stakeholder Partnerships, Civil Society and the private sector, to the achievement of each/multiple SDGs at the various levels. Initially, these will likely focus on processes of alignment, partnerships, extent of integration of the Agenda 2030 principles into strategic plans, policies, programmes and implementation approaches, to progressively focus on the analysis of effects, sustainability of the initiatives and initial impacts. The scope for these evaluations would be at the sub/national-, regional and global level, depending on the evaluation subjects and stakeholders; meta-analysis of evaluations conducted by different stakeholders with similar scopes and comparable subjects, e.g. projects with similar objectives in the same country, or similar policies in different countries, may allow consolidating the findings at different levels of aggregation;
- b. One or more evaluations, inspired by the Paris Declaration evaluation, on the initial impacts of the political decisions embedded in Agenda 2030, upon the direction and effectiveness of the collective development efforts towards the SDGs. These evaluations would probably take place around mid-point of Agenda 2030 and could include in the scope, the efficiency and effectiveness of the monitoring and review process. The main purpose would be the identification of major gaps in the political commitment/s, either for all SDGs or for sub-sets of these at the same time, and in the monitoring and review mechanisms.
- c. Evaluations of the progress made, with a specific focus on the long-term impacts, towards the achievement of the SDGs, taken individually or clustered, at sub/national, regional and global levels, including constraints and enabling factors; the national, regional, thematic and global targets for each of the SDGs, would be the benchmark against which impacts should be measured.

197. In this set-up, the role for UNEG members, typically in partnerships, appears to mostly reside in:

- taking the lead for evaluations under group a. on the work of their respective organizations, at the regional, thematic and global level; at the national level, this would also include leading the UNSDF evaluations (see Section 5.1);
- supporting Member States and the Regional Fora for Sustainable Development (See Section 6.2) for country- and region-led reviews, upon request;
- contribute to evaluations under group c., either through system-wide or other coordination and leadership mechanisms to be developed (see below);
- UNEG in its own right, could be involved in the evaluation/s in group b and in the coordination of its members’ support to the country- and region-led reviews.

198. Another key issue will also be the decision on who will do what, and finding ways for these evaluations to ‘talk to each other’ and feed into the global follow-up and review system. As of April 2016, there is no agreement yet on the shape and modalities of the global, regional and thematic follow-

up and review mechanism for Agenda 2030. This will have a bearing on the country level as well. Further, the whole structure of the UN development system is under the magnifying glass of Member States and up for discussion, while the UN is readying itself to work in the new context.¹²⁷ These important political and institutional processes represent crucial opportunities for UNEG to firmly position evaluation in the Agenda 2030 framework.

199. As initially suggested by OIOS in its MDG evaluation, this Review also identifies the need for a UN evaluation system strategy that addresses the contribution of the system to the Agenda 2030. UNEG is the only and legitimate platform that can issue such a strategy on behalf of its membership, thanks to its significant political capital. UNEG members should urgently develop and agree on the strategy they want, for this to be taken forward by UNEG leadership: to the political level, by advocating and lobbying for a strong role of the UN evaluation system in the review mechanism; and to the institutional level, to contribute to the definition in the UNSDF guidance, of the role of the UN evaluation system in the new UN set-up at country level.

200. This Review, on the basis of the evidence gathered and its analysis, suggests a number of pillars on which the strategy for the UN evaluation system should be based. These are briefly synthesized here and further detailed in the subsequent sub-sections, as required:

- a. UNEG and its members should embrace in their future evaluations, the vision that underpins and informs Agenda 2030 and the SDGs, and in particular, the principles established in paragraph 74 of Agenda 2030 on the follow-up and review process at all levels; UNEG and its members should also stand ready to support Member States in the conduct of the country-led reviews, should there be a demand in this sense;¹²⁸
- b. The UN evaluation system can contribute to learning and accountability at the national, regional, thematic and global levels, by bridging across monitoring, statistics and evaluations, through both country-level evaluations feeding into the country-led reviews, and System-Wide Evaluations harmonized with the focus, frequency and timing decided by the High-Level Political Forum for the review of the SDGs;¹²⁹
- c. Conducting evaluations through partnerships should become the ‘default’ mode of work for UNEG and its members, as feasible and appropriate at all levels; partnerships should be established with Member States and their National Evaluation Systems, other multi-lateral evaluation networks, EvalPartners and its sub-networks, VOPEs, Major Groups and their members, multi-stakeholder partnerships, academia and foundations, and possibly others;¹³⁰
- d. Through these same partnerships, UNEG should also encourage evaluation demand, advocate for and enable wide public access to data and information on the progress made by Member States towards the SDGs, and contribute to meta-analysis and meta-evaluations conducted by diverse stakeholders;¹³¹

¹²⁷ See Sections 6.1 and 5.1 respectively.

¹²⁸ See Section 5.1.

¹²⁹ See Sections 5.1 and 6.1

¹³⁰ See Sections 5.7 and 7.7

¹³¹ See Section 7.2

- e. UNEG should contribute, together with the international development community, to develop new, and/or fine-tune the existing evaluation criteria, approaches, methods and tools that will be required to adequately assess the challenges inherent in the Agenda 2030 principles and the complexity of the SDGs and their indivisibility;¹³²
- f. UNEG members should actively engage and take the leadership of the UNSDF evaluations, possibly in substitution of their country-level evaluations or in addition to these, which should however be closely coordinated and harmonized through common evaluation frameworks; UNSDF evaluations should also be conducted in close collaboration with National Evaluation Systems or proxy governmental institutions, to increase their credibility and utilization as building-blocks for country-led reviews;¹³³
- g. Collectively, the UN evaluation function, through the ISWE mechanism, has a unique comparative advantage in conducting regional, thematic and global evaluations, as well as in the evaluation of the SDGs at the impact level;¹³⁴
- h. UNEG and its members, should continue supporting National Evaluation Capacity Development through advocacy and normative work at the global level, within the UNEG agreed boundaries of action, and respective resources and mandates;¹³⁵
- i. In-country technical assistance and support for the development of effective National Evaluation Systems at the institutional, and individual level, and related advocacy for an enabling environment, should be coordinated and led by UNCTs as part of the UN mandate on Capacity Development, in collaboration with VOPEs, IFIs and bilateral partners; UNEG and its members should facilitate these processes through normative guidance and networking;¹³⁶
- j. UNEG and its members should closely collaborate with the Statistics system, at global, regional and national level, on validation, review, fine-tuning and development of indicators, use and reliability of different sources of data.¹³⁷

201. To achieve most of the above, the UN evaluation system in the Agenda 2030 framework would benefit from the existence of a higher-level mechanism, with both a normative and a coordination mandate. The model of the Inter-Agency and Expert Group, already used by the UN Statistical Commission, appears to be particularly appropriate.¹³⁸

202. On the basis of some exploratory thinking, a future IAEG-Eval should draw its membership from National Evaluation Systems, JIU, OIOS, UNEG and possibly some of its members on a voluntary rotational basis, EvalPartners, Major Groups as representatives of Civil Society, and possibly academia. Its purpose would primarily be normative, to provide guidance and facilitate experience sharing and

¹³² See Sections 5.3, 5.4, 5.5, 5.6 and 7.4

¹³³ See Section 5.1. and 7.2

¹³⁴ See Sections 5.7

¹³⁵ See Section 5.2 and 7.3

¹³⁶ See Sections 5.2. and 7.3

¹³⁷ See Section 6.3

¹³⁸ See Section 6.3

lessons learning on evaluation in the context of the SDGs and Agenda 2030 for the global evaluation community. The mandate of the body should include:

- a. the oversight and programming function for System-Wide Evaluations, which would enrich and ground its function into a more tangible and results-oriented role; this would also strengthen the institutional legitimacy of the ISWE mechanism, and possibly facilitate the grounding of the IAEG in the JIU itself, which is one of the ‘natural’ host organizations, together with UN-DESA and OIOS;
- b. collaboration with Member States for the programming of their country-led evaluations, and high-level coordination of support and efforts for leveraging resources;
- c. facilitate meta-analysis of evaluations on key SDG-related themes, to enable access by the wide public to evaluative information and data originated by diverse stakeholders, including among others, Member States, the UN, the International Financial Institutions, Multi-Stakeholder Partnerships, Major Groups.

203. An additional key institutional element in this architecture would be the regional level, to enhance the utility and use of evaluation evidence there. Ideally, the evaluation units of the Regional Commissions, would have a coordinating function for UN evaluative efforts in the respective regions through the UNEG regional networks, identifying gaps of evaluative evidence, extracting and consolidating findings, conclusions and recommendations from regional evaluations, and making these available to the RFSD for their regional review processes. An additional role for the UNEG regional networks, would be facilitating peer-learning based on both evaluability of policies and programmes, and evaluative evidence, across countries in the same region. UNEG and its members should therefore support, including through advocacy and/or human resources as appropriate, a strengthened evaluation function at the regional level.¹³⁹

204. A major and legitimate question in all of this, concerns the resources available for the expected UN evaluation effort. The OIOS MDG evaluation and the JIU 2014 analysis of UN evaluation functions, both stressed the need for additional resources for the SDGs evaluation framework. This was not explicitly mentioned among the issues at stake in the Member States’ informal consultation on the global follow-up and review mechanism currently on-going in New York. At the time of writing, a zero-growth scenario is the only possible default projection. Considering that the work of the UN should all converge under the Agenda 2030 framework and contribute to the achievement of the SDGs; progressively, also all UNEG members will have to re-allocate a major portion of their current resources to evaluations of SDG-related work. This should happen, from the national level with the joint UNSDF evaluations, to the regional, thematic and global, conducted insofar as possible, in a joint modality. The lessons learned through the ISWE will again, be very relevant on this issue as well.

205. The above are proposals, for UNEG members to agree or disagree, add new ones and/or take some out. The key point here is that the preparation of a UNEG strategy proposal for the UN evaluation system in the framework of Agenda 2030, should be taken as a matter of the utmost importance and urgency. If the opportunity offered by the currently on-going political and institutional processes is missed, the international recognition and visibility achieved by UNEG through the major advocacy efforts that led to the 2014 UNGA resolution on National Evaluation Capacity and the International Year of Evaluation, might be seriously undermined. Suggestion 1 addresses this key issue.

¹³⁹ See section 6.2

Suggestion 1. To UNEG and its members

UNEG and its members should urgently develop and agree on a strategy proposal for the role and responsibility of the UN evaluation function in the Agenda 2030 framework. The window of opportunity for making a contribution to both currently on-going intergovernmental and institutional processes is very short and requires an immediate action. Advocacy on key issues should start soon after the AGM 2016.

7.2 On country-level evaluations and country-led reviews

206. Agenda 2030 emphasis on country ownership is leading to significant changes in the way the UN will work at country level, including the key role of the UNCTs in Delivering as One. New guidelines for the UNSDF are being prepared at the time of writing this report, which will have a strong bearing on how UN evaluations at country level will have to be conducted. All of this, in addition to the evidence available on UNCTs' capacity to manage quality evaluations and the complexity inherent in the UNSDF evaluations, point to the need for UNEG members to take the lead in conducting UNSDF evaluations and closely coordinating all evaluations at country-level. In this scenario, the structure of UN evaluations at country level would be the following:

- a. UNEG members should directly engage and lead the UNSDF evaluations, in close collaboration with the national governments and their evaluation systems, to make these exercises credible and accepted building blocks for country-led reviews; internal arrangements will have to be developed, at least initially, through a case-by-case approach; external partners should also be involved as appropriate;
- b. whenever separate country-level evaluations will have to be conducted by individual organizations for internal accountability purposes, these will have to be planned in close collaboration with the UNCT, with the concerned line-ministries and with the UNEG members involved in the UNSDF evaluation; these evaluations will have to be harmonized with the UNSDF evaluation framework;
- c. whenever UN work focused on one or more Goals is conducted outside the UNSDF, or more in-depth evaluations pitched at SDG level will be required, respective evaluations should be carried out as joint exercises by concerned UNEG members; these evaluations will have to be harmonized with the UNSDF evaluation framework and involve external partners as appropriate;
- d. UNCT Monitoring and Evaluation teams, under the responsibility of the Resident Coordinator, would contribute to plan, coordinate and support both the UNSDF evaluation and other evaluation efforts by UNEG members, including liaising and facilitating collaboration and partnerships with the respective National Evaluation Systems.

207. Additional benefits from this approach would be that the quality of the UNSDF evaluations would dramatically improve, the demand on UNCTs and national governments for resources would decrease, including staff time to follow-up individual organizations' evaluations, and the evaluations would be an additional trigger for the UN organizations to jointly deliver.

208. An alternative option, taking into account the complexity and the long-term nature of the work to be assessed, as well as the potential existence of evaluative evidence originated by a variety of stakeholders at national level, could be for the UNSDF evaluations to take the form of meta-analysis

drawing on separate and joint evaluations conducted by UN organizations, and by other stakeholders, all informed insofar as possible by similar analytical frameworks.

209. These models are only two, among possible others. Also, as repeatedly stressed in the Agenda 2030 and related discourse, there is no such thing as ‘one-size fits all’. Various approaches and options will have to be developed to meet the specific circumstances of the various countries, or sub-sets of countries. Similarly, key principles for the UNSDF and other country-level evaluations by the UN may have to be defined, if possible also at the level of UNEG Standards.

210. Furthermore, it appears of the utmost importance that UNEG engages with UN/DOCO and presents the internally agreed position of UNEG members, for this to be integrated in the next version of the UNSDF guidance document. UNEG should also not miss the opportunity to participate in the piloting of the new guidance, in two to three countries, as this will be a building block on roles and responsibilities for a number of years to come.

211. With respect to the role of the UN evaluation system in country-led reviews, there is too limited information as of April 2016 about these processes, how they will be conducted and what will be their contents. Different groups of countries, moreover, will have different requests for support from the UN evaluation system. Initially, a case-by-case approach will likely be the only option, with UNEG Secretariat in a clearing-house role, coordinating the available support from UNEG members as feasible. A significant step forward in the efficiency and effectiveness of this support would be the possibility to plan the country-led reviews, at least two years in advance. The IAEG-Eval proposed in this Review, should have a role in this.

212. In the light of all the above, the Review proposes Suggestion 2 below.

Suggestion 2. To UNEG and its members, on country-level evaluations and support to country-led reviews

UNEG and its members should:

- i) Urgently discuss the proposal in this Review about their role in the UNSDF evaluations and on the basis of the outcome of the consultation, contribute to the preparation of the UNSDF guidelines; participation in the pilot testing in a limited number of countries would also be advisable;
- ii) Follow-up closely the presentation of the upcoming 21 country-led reviews, the ensuing debate and recommendations and debate on their respective roles and responsibilities should requests come to them for supporting the future reviews.

7.3 On National Evaluation Capacity Development

213. Expectations from Member States vis-à-vis the UN are high on National Evaluation Capacity Development. Only few UNEG members are actively involved in N/ECD, and the very diverse activities so far, have led to useful normative products of global relevance. Most members do not contemplate this type of work in their mandates and among those who do, resources are often a constraint. Alternative no-cost approaches are thus followed, including engaging national governments and other stakeholders in evaluation processes at country level and ‘in-service’ development of skills and competences in evaluation of consultants from the Global South.

214. UNEG members are divided on their role and responsibilities on NECD. However, the demand in Agenda 2030 for the UN to take up the responsibility, requires that clarity be urgently made. In this

context, the debate within UNEG should adequately take into account the core principles underpinning the UN evaluation system, which are:

- a. the long fought-for and distinctive requirement of independence for the UN evaluation system, including the requirement for its total separation from management and programme operations;
- b. the strong and indivisible link between independence, credibility and utilization of evaluation;
- c. the need for all UN work to be independently evaluated, irrespective of who designs, manages and implements it;
- d. the added value and comparative advantage of the UN evaluation system when compared to other functions, including training, research, policy analysis and support.

215. Moreover, it is important to take into account that there is an increasingly active and populated network of actors in evaluation at different levels, who could operate where UNEG members cannot, lest stepping over the boundaries of their mandates. Partnerships will be the key in this approach. Furthermore, the same mechanism proposed below with respect to country-level outcome and impact monitoring (see 7.4) could apply in the case of support by the UN to UNCD.

216. Suggestion 3 below addresses this issue.

Suggestion 3. To UNEG members, for a definition of their role in NECD

UNEG members, taking into account the views of all its members and the considerations raised in this Review about independence of the evaluation function and its comparative advantage, as well as the need for all the UN work to be independently evaluated, should define the boundaries of their mandates in National Evaluation Capacity Development. The final decision should be transformed into a UNEG Norm.

217. Independently from the decision on UNEG members' role in NECD, UNEG should address a key global need on NECD. The UNDP recently published "Baseline on National Evaluation Systems" is a useful initial effort in mapping the 'who is who' in evaluation in 43 countries. The publication itself recognizes that more work would be required, and this would mean extending the mapping of NES to all UN Member States and analyzing more in depth with each of them at least two key issues:

- a. the assessment of the demand for NECD in each Member State and identifying with national stakeholders and external partners including the UN system and IFIs, on who could do what and with what resources;
- b. the potential pathways for those country-led reviews that want to take into account stakeholders' contribution, including the UN evaluation system, VOPEs, Major Groups, bilateral and multi-lateral partners;

218. This would undoubtedly be a major effort for UNEG, also in financial terms. It would however be fully in line with its mandate and would represent an important contribution to the international evaluation community. UNEG Secretariat could have a leadership and coordination role in such an assessment, that would have to be commissioned to a group of consultants. If this will not be feasible UNEG could still facilitate the work at country level of those UNEG members who wish to engage

more systematically with National Evaluation Systems in conducting their evaluations. Suggestion 4 addresses both proposals.

Suggestion 4. To UNEG, for the mapping and analysis of National Evaluation Systems

UNEG should take the leadership for commissioning a world-wide mapping and analysis of national evaluation systems, with a view to assess the need for National Evaluation Capacity Development and potential pathways of collaboration on country-led reviews.
As a first step in this, UNEG Secretariat should compile and keep updated a repository of contacts for National Evaluation Systems in Member States. This could entail an on-line system that NES themselves could keep updated.

7.4 On evaluation criteria, methods and approaches

219. The Review has identified on several occasions, the need for capturing the SDGs principles through new, and a more systematic use of existing evaluation criteria, as well as for developing and fine-tuning approaches, methods and tools. This appears necessary to more adequately address the complexity inherent in the work that will be conducted for the achievement of the SDGs, as well as to make evaluation responsive to the principles of Agenda 2030 itself. This sub-section consolidates all suggestions in this domain, with the exception of proposals on evaluability assessments that are discussed in Section 7.5 below.

220. The main purpose of including additional evaluation criteria to those currently in use among UNEG members, is to contribute to building large data-sets of comparable evaluative evidence, on the basis of which analysis of performance and achievements can be conducted at various levels of aggregation. An additional benefit is linked to the empirical experience showing that the introduction and systematic application of evaluation criteria on critical issues, for example progress in mainstreaming gender equality, contributes to positive behavioral and institutional change.¹⁴⁰

221. This suggestion also stems from the observation that the request made to the UN and international evaluation community to provide evaluative evidence, and related learning and accountability, at the national, regional, thematic and global levels of the progress made towards the SDGs, is of a tall order. In a context of hugely diverse sets of initiatives across the globe implemented by highly heterogeneous organizations and actors, meeting this challenge will require, among others, frameworks for data-gathering as harmonized and comprehensive as possible, to enable making any sense of the information and evidence available, at the moment of consolidating and analyzing it. And expanding the list of common criteria and evaluation questions from which to draw information and data, would help in this sense.¹⁴¹

222. Furthermore, most of the proposed additional evaluation criteria are, albeit to different extent, already “common currency” for several UNEG members, for example gender equality, partnerships, participation. Not all the new proposed criteria will be useful for all UNEG members, given the diversity among UN organizations, nor all will be useful in all evaluations by any single UNEG member. Also, for evaluations pitched at different levels of analysis, different criteria will be appropriate. The main

¹⁴⁰ Experience proves that this is effective if other incentives are also in place of course.

¹⁴¹ This already happens in reality, for example the criterion ‘impact’ is often not assessed for a number of legitimate reasons. Nevertheless, having it in the list of standard criteria, helps both as a standard and in broadening the potential scope of the evaluation functions.

rationale of the proposal is to expand the range of agreed evaluation criteria and related questions that can help in contributing to data aggregation and analysis at higher levels, from which each evaluation will draw, as useful, relevant and appropriate.

223. The “missing” criteria that have been identified as most closely relevant to Agenda 2030 and the SDGs, and that are also highly relevant for the majority of UNEG members seem to be the following (alphabetic order): country-ownership, equity, gender equality, human rights, inclusiveness, participation, partnerships. An argument could be easily made for adding Environmental Sustainability and Sustainable Development as standard criteria: both are quite complex and not easily captured through one criterion or indicator. At the same time, both are partly included in the international standard criterion of sustainability, which incidentally often tackles the environmental dimension of Sustainable Development. Other criteria of high relevance to Agenda 2030, e.g. resilience, sustainable natural resources management or natural resources status just to mention a few, hold valid for sub-sets of UNEG members, who could agree among themselves about their adoption as well.

224. The integration of the existing set of criteria with the new criteria should be complemented by the development of one or two key standard evaluation questions for each criterion, possibly supported by detailed rubrics for reducing the influence of subjective assessments, to be included in all evaluations, as appropriate.

225. Data generated on each criterion through UNEG members’ evaluations, would be consolidated by each UNEG member, made available to both the respective organization and UNEG, to be included into a system to be developed, publicly available, on the UN performance in supporting Member States in the implementation of Agenda 2030. This would represent a very powerful contribution to the accountability framework for the UN system as a whole. Suggestion 5 synthesizes this proposal.

Suggestion 5. To UNEG and its members, on expanded evaluation criteria and their use

- i) UNEG and its members should agree on expanding the list of evaluation criteria to be potentially used in the evaluations of initiatives implemented in the Agenda 2030 framework. Proposed criteria common for all are: country-ownership, equity, gender equality, human rights, inclusiveness, participation, partnerships.
- ii) Groups of UNEG members whose focus of work coincides, e.g. the environment-focused or the humanitarian-focused organizations, should also develop additional sets of common criteria for their evaluations;
- iii) UNEG should develop a data-base where it would be possible to consolidate the information generated on each criterion by all UNEG members, and make its contents publicly available.

226. An additional step has been suggested, to develop “networked evaluation frameworks” to facilitate analysis of evaluative evidence from various sources, for UNEG members that share mandates for the same SDGs and will be involved in related joint evaluations at national, regional, thematic and global level. This might be an interesting approach to be tested, for example, by the Rome-Based Agencies, which share the common goal of Food Security, to which each organization contributes from a different perspective.

227. Mainstreaming new criteria in evaluations will be a necessary, but not sufficient step to adequately capture, and comply at the same time, with the principles underpinning Agenda 2030. For some of these, namely Human Rights, Gender Equality, Universality and possibly Partnerships, approaches, methods and tools are already included to some extent in the common “tool-box” for UN evaluations, even if they are by no means fully acquired and mastered by all. Other principles

nevertheless represent quite new and complex ground. For example, the discussion on the evaluation of indivisibility of, and inter-linkages across the SDGs raised the following observations:

- evaluations will have to be adaptive, and expand their scope beyond the boundaries of individual organization's work to address the work of the UN as One and how this will interact with 'all the rest'; this will likely require unpacking complex initiatives into more easily evaluable components, based on common frameworks and on syntheses of findings and results that look at interactions also at the level of the whole;
- innovation will be required in the conceptual framing of evaluations, including new ways of developing or addressing theories of change, articulation among these, as well evolution and changes in programmes themselves (emergence);
- Contribution Analysis and Qualitative Comparative Analysis offer valuable insights and may be very helpful, but may also not be sufficient;
- evaluability assessments will have to establish boundaries defining what components or parts thereof, will be possible to evaluate with an acceptable level of rigour and credibility, and those that will only be assessed at a lower level of accuracy or confidence.

228. The existing UNEG guidance documents are all solid building blocks offering some direction in this respect. Also, some work for their upgrading to the new challenges has already started, as for example is the case with the on-going work on Equity-focused and Gender-responsive evaluation. The existing experience accumulated over long years of participatory approaches also applied to evaluations, should prove useful and relevant, including at the level of impact analysis. However, this seems to have been partly forgotten, or simply unknown to the new generations of evaluators.¹⁴²

229. What is still missing in the UNEG guidance portfolio is an analysis about how UNEG members could integrate the concept of Sustainable Development in their evaluations. The concept of SD is highly challenging, which is also why there is not one single indicator for it. It is likely that complex discussions around it would be required to make it somewhat amenable, if at all, to any measurement. In addition, the long time-frames required to assess sustainability and resilience, and often impacts on social and institutional structures or natural resources at levels beyond the household or community, will require different conceptual frameworks, and organizational arrangements as well, including subsequent cycles of evaluations or meta-evaluations over long time-spans.

230. Any suggestion on the issues at stake here could only be formulated at a very generic level, and indeed, the purpose and usefulness of a suggestion is in itself questionable. The needs are broadly laid out here. Now, decisions should be made by UNEG and its members, on priorities of action. A possible choice could be to tackle the low-hanging fruit first, i.e. to improve what is already there; or exactly the contrary, decide to break ground on those topics which are more nebulous and complex, because some clarity is better than none.

¹⁴² This incidentally raises issues of internal knowledge management and institutional memory. This author found partly exciting and partly worrying, in the recent New York event, listening to proposals that for the audience were really innovative, though they referred to approaches that had been part of the bread and butter of the development community, evaluators included, until not so long ago.

231. A final consideration has a place here, on the long-term and complex nature of the Goals. These will require establishing assessment systems to track results and changes, both expected and unexpected, as well as emergence and inter-linkages, in real time. This, to enable more timely adjustments and steering actions than could ever be possible through evaluations, even if new arrangements will be developed for these. It is therefore suggested for UNEG to promote that outcome and impact-level monitoring be included among the tools that the UN system will put in place in support of country-led reviews, possibly in collaboration with National Evaluation Systems.

232. The responsibility for developing and implementing the monitoring systems should primarily rest with the UNCT Monitoring, Evaluation and Data teams foreseen in each UNCT, to provide support to all UN organizations working in the country. The regional UN Evaluation Group Inter-agency Networks could provide methodological support to the UNCT M&E teams for establishing the systems, and could in turn be supported through a UNEG-level Task Team in this activity. As UNEG members would be among the primary users of the monitoring data produced, some caution would be required in ensuring absence of conflict of interest at the time of conducting UNSDF evaluations. Suggestion 6 synthesizes the proposal.

Suggestion 6. To UNEG and its members, on support to country-level outcome and impact monitoring

UNEG and its members should propose and advocate with UNCTs for the establishment of outcome and impact-level monitoring systems, to enable real-time tracking of results, expected and unexpected, from the initiatives implemented by the UN in support of Agenda 2030. The responsibility for developing and running the systems would rest with the UNCT M&E teams; the regional UNEG Inter-agency Networks could provide methodological support, and be in turn supported in this by a UNEG-level task team.

7.5 On Evaluability assessments

233. The Review has found that the definition of Evaluability by the OECD/DAC ‘*The extent to which an activity or project can be evaluated in a reliable and credible fashion*’ is broadly accepted, although the term in itself is somewhat awkward to grasp and has been misinterpreted at times. In turn, the useful model for Evaluability Assessment focused on the assessment of Design, Data and Demand for evaluation, has the advantage, and disadvantage, of being very broad, and to be applicable and applied to very different exercises. This points, in the opinion of this Review, to the need for more clarity about terminology, purpose, roles and responsibilities for the exercises currently conducted under the term Evaluability Assessment.

234. The Review also identified a “dearth of evaluability” within the UN system. A few UNEG members have conducted Evaluability Assessments, intended as ex-ante/design reviews of Strategic Plans and UN system-wide planning approaches, which have proved to be useful to enhance internal coherence of plans and the robustness of the evidence-base generated by evaluations, about UN relevance and effectiveness. Their methodological knowledge and experience in this domain, complemented with the broader experience of the international evaluation community, should be harnessed in one single UNEG guidance document, that appears to be particularly useful in the context of the SDG evaluation challenges. This should also include suggestions for roles and responsibilities in the management of these reviews, including the caveat for segregation of roles within evaluation units. Suggestion 7 and Box 5 address both issues.

Suggestion 7. To UNEG, on a guidance document on Evaluability Assessments and ex-ante and early reviews

UNEG, building on the experience of its members and of the international evaluation community on Evaluability Assessments, ex-ante/early reviews and related tools, should develop a Guidance note that:

- i) Clarifies the terminology and arrangements for Evaluability Assessments and ex-ante/early reviews, along the lines proposed in Box 6;
- ii) Consolidates best practices for evaluability assessments in preparation for an evaluation;
- ii) Defines the procedures and standards for ex-ante/early reviews of Strategic Plans/Frameworks and programmes;
- iii) Includes standards tools and templates for both.

Box 5. Proposed terminology and arrangements for Evaluability Assessments and ex-ante/early reviews

UNEG and its members, together with other stakeholders in the evaluation community, should contribute to define what an Evaluability Assessment is, who should conduct it and how. In line with Rick Davies' identified purposes of EAs, and UNEG Norms 7.1 and 7.2, it is suggested that:

i) As per UNEG Norm 7.2, the term evaluability assessment should exclusively indicate preparatory steps of evaluations, that focus on the availability of Data, the Design and the Demand for the evaluation;

ii) EAs should not be typically used to cancel evaluations because they are 'too complicated or sensitive'; however, an EA may show that a planned evaluation is poorly timed, poorly framed, significantly under-resourced, or even useless because it would not add anything to what already known;

iii) Ex-ante or early assessments of strategies, plans, policies, programmes and projects, that focus on Design, rather than results, should be called 'ex-ante/ design review' depending on the moment in time they are conducted; these design reviews should be conducted in such a way as to avoid any conflict of interest with the planned subsequent evaluations.

235. In addition, UNEG members should advocate within their organizations, and provide support in an advisory role, for improving the evaluability of the work done by the UN in support of the implementation of Agenda 2030. This through ex-ante or early analysis of strategies, plans and programmes, leading to better design and integration of Theories of Change and of measurable indicators, aligned with the national, regional and global indicator frameworks. These analyses could also be used to detail the scale of the challenges that the evaluation of a given programme is likely to face, and for comparison across programmes, either by using a composite index and /or through qualitative descriptions.

236. The overall purpose of the advocacy work would be to enable a more rigorous and credible assessment of the UN contribution, or lack thereof, to the established targets and should be directly done by the UNEG members within the UN system. Similarly, the UNCTs and VOPEs should engage in the same advocacy processes with Member States. Suggestion 8 articulates this proposal.

Suggestion 8. To UNEG members, on ex-ante and early reviews of Strategic Plans and Frameworks in the context of the SDGs

Within their organizations, UNEG members should:

- i) In an advisory role, advocate with Management, and indirectly with Member States, for the systematic application of Theory of Changes in the development or revision of their SDGs-related strategic plans and frameworks;
- ii) In an advisory role, advocate with Management, and indirectly with Member States, for conducting ex-ante or early Design reviews of their SDG-related revised strategic plans and frameworks, to ensure the internal clarity, consistency, alignment of indicators with SDGs and country-level indicators;
- iii) Following the example of a few UNEG members, the evaluation policies of UN organizations should clarify that ex-ante and early reviews exclusively focused on Design of strategic plans and frameworks or programmes and projects, should be the responsibility of Management, with Evaluation units exclusively involved in an advisory role, if any at all.

7.6 On Lessons learned from past evaluations

237. The UN system has been remarkably inactive in conducting evaluations of the MDGs, with few notable exceptions. UNEG members may consider identifying why this was the case and share their views. This evaluation gap should be avoided at all costs in the SDG context and this Review is indeed one of the building blocks for UNEG and its members to address the challenge.

238. Despite the limited number of evaluations that met the criteria established for this Review, these do contain a number of key lessons on evaluation-specific issues, listed below:

- a. Any complex evaluation tackling multiple countries and diversity of adoption, adaptation and implementation, requires developing at inception, possibly in a participatory and inclusive manner, a very clear central framework for analysis and synthesis, with a clear results logic, that allows consistency in inquiry and findings while allowing for some flexibility to address diversity and the unforeseen at the local level;
- b. Independence is critical, particularly where potentially contentious findings are likely to emerge; this should include strong procedures and clear governance systems, articulated from an early stage, reinforced throughout the process, and checked and verified as part of ongoing evaluation management;
- c. Theory-of Change, contribution analysis and focus on gathering qualitative evidence have proved to be appropriate and useful approaches in these evaluations; this should not exclude other potential approaches, in particular taking into account that linear ToCs are not fully appropriate to unpack complex interlinked systems, as the SDG-related interventions will be.

239. The evaluations also point to a number of key systemic challenges, which are still relevant in the post-2015 context. The meta-evaluation of the UNDAFs prepared by the ISWE should further contribute to a list of lessons learned; experience also tells that additional useful methodological evidence, findings and recommendations should be available from UNEG members' country-level evaluations. All of this represents a wealth of learning opportunities for the UN system, and possibly for Member States as well. Extracting and consolidating this information in a suitable format for decision makers in Governments, Senior Management, Operations and UNCTs, would mean extending

and increasing the utility of past evaluations and increasing the relevance of evaluation units. Suggestion 9 articulates this proposal.

Suggestion 9. To UNEG and its members, on the use of past evaluation findings and recommendations

- i) The few evaluations available of the MDGs and similar complex initiatives, such as of the Paris declaration and the DaO, contain key lessons for UNEG members and the international evaluation community to take into full consideration in the context of the Agenda 2030 evaluations;
- ii) UNEG may consider conducting a membership survey on the reasons why only very few evaluations of the MDGs were conducted, oriented to learn lessons for the future;
- iii) evaluations of the UNDAFs and UNEG members' country-level evaluations contain a wealth of evaluative evidence highly relevant to the mandate of the UN in the SDG era. This should be made available in consolidated form and easily digestible format, for the UN system and Member States. UNEG could help in coordinating, consolidating, publishing and making available this information for the wider set of stakeholders.

240. Last, despite the UNEG evaluation reports data-base has been in function for several years now, it is still very short from including all the evaluations issued by UNEG members; some members have no report uploaded, and others have only a few. UNEG Secretariat might also consider improving the user-friendliness of the download function from the database. Suggestion 10 addresses this issue.

Suggestion 10. To UNEG and its members, on a common repository of UN evaluation reports

- With the aim of making UN evaluation reports more easily and publicly accessible:
- i) All UNEG members should commit to upload all their evaluation reports in UNEG repository;
 - ii) UNEG Secretariat should improve the user-friendliness of the repository, by creating additional search categories and facilitate downloads of large numbers of evaluation reports.

7.7 On Partnerships

241. Partnerships play a key role in the SDGs and Agenda 2030, that are built on complex partnerships and propose partnerships as a goal in itself. The UN evaluation system will also have to systematically embrace partnerships, as a modality of work in evaluation to tackle the challenges ahead, as both a goal and criterion of performance. UNEG members have a solid base of experience in joint evaluations, including the work done within the IAHE and the lessons that will be learned from the ISWE pilots. More lessons reside in past evaluations and further guidance and standards may still be required, expanding what already exists in UNEG with more recent experiences.

242. For UNEG and its members, developing partnerships will be of particular importance to meet many of their responsibility within the Agenda 2030 framework, as follows:

- a. to develop, update, fine-tune and test new approaches, methods and tools, UNEG should closely collaborate with the international evaluation community, including EvalPartners and its sub-networks, multilateral evaluation networks, VOPEs at all level, academia and evaluation think-tanks, to join forces to issue relevant and high quality products and enhance their utilization;

- b. to contribute to country-led reviews, UNEG members should develop partnerships with National Evaluation Systems, Parliamentarians, VOPES, Major Groups, multilateral evaluation networks;
- c. to contribute to National Evaluation Capacity Development, UNEG should facilitate partnerships between UNCTs, EvalPartners, VOPEs and Major Groups, with the latter to expand the outreach of NECD work, multilateral evaluation networks, as well as with National Statistics Systems;
- d. to comply with the Universality principle, UNEG members should develop partnerships with evaluation units in the respective line ministries in all Member States, aiming at collaborative and joint efforts for the evaluation of normative work, including Global Public Goods, at national and sub-national level; OECD/DAC Evalnet members could be key facilitators in this process;
- e. to contribute to tracking progress towards the SDGs, accountability and lessons learning at all levels, UNEG and its members should closely collaborate with the Statistics systems, on the issues of common interest.

243. In all the above, UNEG association with EvalPartners and its sub-networks, can expand the reach-out potential, as well as demand and use for evaluation at national, regional and global level, for all concerned. This partnership should be perceived as a fundamental piece of the future evaluation architecture and should be maintained and nurtured, building on the respective comparative advantages, interest, capacities and resources. However, this Review identified a gap among UNEG members, and possibly within EvalPartners as well, in understanding what and how this partnership can offer and function, for the benefit of all parties. Suggestion 11 addresses this issue.

Suggestion 11. To UNEG leadership

UNEG leadership should engage in a debate with UNEG members on the opportunities offered by the association with EvalPartners to tackle the challenges embedded in the SDGs and in Agenda 2030. Joint pilot initiatives should be identified and carried out, to develop a shared experience on what is feasible. Areas of collaboration that appear particularly promising are:

- i) networking with UNCTs to meet the demand for national evaluation capacity development;
- ii) raising awareness about evaluation among Parliamentarians;
- iii) production of joint guidance documents; and
- iv) collaboration with Civil Society through Major Groups members to enhance the use of evaluation for learning, accountability and ‘transformational change’.

244. The scenario for the implementation of Agenda 2030 will become increasingly complex. Among the various players, Multi-Stakeholder Partnerships represent a particular group of very diverse initiatives, linked to the UN but accountable through different frameworks. It is important that UNEG and its members advocate with UN Senior Management and with the MSPs themselves, for transparency and inclusiveness of their evaluation processes, fully including the evaluation units of the UN organizations concerned. Suggestion 12 addresses this challenge.

Suggestion 12. To UNEG and its members, with respect to Multi-Stakeholder Partnerships.

UNEG members whose organizations are involved in Multi-Stakeholder Partnerships with separate accountability frameworks, should advocate with their Senior Management, if necessary with UNEG support, for their full and systematic participation in the Evaluation Management Groups for the MSPs’ evaluations.

245. Another set of important players are the Major Groups, that represent multi-tier organizations at national, regional and global level and whose vast memberships are grounded in the civil societies of Member States. Major Groups are an ideal entry-point to reach out to associations that need support in capturing the available opportunity to report on their contribution to the implementation of Agenda 2030, and to participate in the country-led reviews. Suggestion 13 contains this proposal.

Suggestion 13. To UNEG, EvalPartners and their members, on partnering with Major Groups

UNEG and EvalPartners should develop partnerships with the Major Groups and their members, at the appropriate levels, to enable:

- i) capacity development of MG members on the use of evaluation as an accountability and lessons learning tool;
- ii) collaboration in identifying priorities for evaluation agendas at national, regional and global levels;
- iii) participation of MG members in evaluations as stakeholders; and
- iv) improved use of evaluation findings and recommendations through enhanced ownership and relevance for MG members.