



Final Inception Report

Joint Evaluation of Joint Gender Programmes in the UN System

Prepared for //
Evaluation Management Group

Date //30th August 2012

By// Julia Betts, Cathy Gaynor,
Hope Kabuchu, Angelica
Arbulu, Hatty Dinsmore, Judith
Friedman

IOD PARC is the trading name of International
Organisation Development Ltd//

Omega Court
362 Cemetery Road
Sheffield
S11 8FT
United Kingdom

Tel: +44 (0) 114 267 3620
www.iodparc.com

Contents

Acronyms	ii
1. Introduction	1
1.1 Context of the evaluation	1
1.2 Structure of the Evaluation	2
1.3 Purpose, objectives and scope of evaluation	2
2. Theoretical and Conceptual basis of the Evaluation	8
2.1 Theoretical basis of JGPs	8
2.2 Conceptual approach to the evaluation	11
3. Evaluability Assessment	15
3.1 Completeness of documentation set	15
3.2 Specific data gaps	15
3.3 Counterfactual	17
3.4 Revising the Evaluation questions	18
3.5 Evaluation criteria and how they will be applied	23
3.6 Summary of feasibility and limitations	24
4. Methodology	27
4.1 Evaluation process and stages	27
4.2 Overall methodological and analytical framework	27
4.3 Operationalising the methodological approach	29
4.4 Detailed breakdown of Evaluation Phases	35
5 Final outputs and quality assurance	43

Acronyms

BPA	Beijing Platform for Action
CCA	Common Country Assessment
CEDAW	Convention for the Elimination of All Forms of Discrimination Against Women
CSO	Civil Society Organisation
DAO	Delivering As One
EMG	Evaluation Management Group
ERG	Evaluation Reference Group
GEWE	Gender equality and women's empowerment
HR	Human Rights Based Action Plan
JGP	Joint Gender Programme
MDG	Millennium Development Goals
MDG-F	Millennium Development Goals Achievement Fund
NGO	Non-governmental organisation
OECD DAC	Organisation for Economic Co-operation and Development Development Assistance Committee
QCPR	Quadrennial Comprehensive Policy Review
RBM	Results Based Management
ToR	Terms of Reference
UN	United Nations
UNCT	United Nations Country Team
UNDAF	United Nations Development Assistance Committee

UNDOCO	United Nations Development Coordination Operations Office
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
UNFPA	United Nations Population Fund
UNICEF	United Nations Development Fund for Children
UNIFEM	United Nations Development Fund for Women
UNRC	United Nations Resident Co-ordinator

1. Introduction

1.1 Context of the evaluation

This Joint Evaluation of Joint Gender Programmes in the UN System is taking place at time of change and reform. The launch in 1997 of the UN reform agenda by the then-Secretary-General, aimed at creating a more effective and efficient UN system¹, placed joint programmes at the centre of a change process that sought to harness the comparative advantages of agencies in a more coordinated way.

Various statements and declarations since then have reinforced the importance of joint programmes for supporting UN reform.² The ‘Delivering As One process’, which started with eight pilot countries, now expanded to include ‘self-starter’ nations, has led the ‘simplification and harmonisation’ process on the ground.³ Initiatives such as the Government of Spain’s Millennium Development Goals Achievement Fund (MDG-F) also provided support to improve UN effectiveness at country level through funding and a specific programmatic window on gender equality and women’s empowerment (GEWE). The endorsement of the Paris Declaration on Aid Effectiveness in 2006 further linked joint programming inextricably to the aid effectiveness agenda.

With these drivers, the number of joint programmes within the UN system has gradually increased.⁴ Yet at the same time, UN contributions to supporting governments to achieve gender equality and women’s empowerment were, in the late 2000s, acknowledged to be still lacking.⁵ The new entity of UN Women was created in 2011, with the mandate to lead and promote coherence in the UN system on gender equality, and to coordinate the efforts of the UN system to support the full realization of women’s rights and opportunities.⁶

While the number of joint programmes within the UN system is increasing, including those on gender equality and women’s empowerment (GEWE), there have been limited assessments of joint programmes (JPs) in general and of joint gender programmes (JGPs) specifically.⁷ Consequently, seven partners - the United Nations Development Fund for Children (UNICEF), the United Nations Development Programme (UNDP), the United Nations Entity for Gender Equality and Women’s Empowerment (UN Women), the United Nations Population Fund (UNFPA), the Millennium Development Achievement Fund

¹ Secretary General Report *Renewing the United Nations: A Program for Reform: Measures and Proposals* 1997.

² For example, the 2006 High Level Panel on System-Wide Coherence in the Areas of Development, Humanitarian Assistance and the Environment report, which launched the Delivering as One (DaO) initiative to further UN reform at the country level.

³ See UNDG (2003) Guidance Note on Joint Programming

⁴ Terms of Reference, available at Annex 1

⁵ General Assembly resolutions in 2009⁵ and 2011 identified the pressing need to enhance the accountability of the UN system on gender equality and women’s empowerment. The Delivering as One, Report of the Secretary-General’s High Level Panel on System-wide Coherence in the Areas of Development, Humanitarian Assistance and the Environment (A/61/583) found that despite the UN system’s key role in it, its contributions in this area had been incoherent, under-resourced and fragmented.

⁶ General Assembly Resolution 64/289

⁷ See attached Glossary for the terms applied for the Evaluation, including Joint Programme, Joint Gender Programme and joint programming

(MDG-F); and the Governments of Norway and Spain - have come together to undertake this first major joint evaluation of JGPs in the UN system.

1.2 Structure of the Evaluation

This Evaluation's management and governance structures are as follows:

- The **Evaluation Management Group (EMG)** are the main decision-making body for the evaluation. The EMG is composed of designated representatives from the evaluation offices of the key joint evaluation partners, and includes a Secretariat, managed by UN Women Evaluation Office, a core group within the EMG tasked to oversee the day to day business of the evaluation and communication with the Evaluation Team. The EMG is responsible for the dissemination of the Evaluation and for developing the management response.
- The **Evaluation Reference Group (RG)** is composed of Gender Focal Points, Joint Programme Focal Points, UNDG Gender Team members, representatives from donor countries, UN Resident Coordinators, UNCT members and UN Gender Theme Group members. This group acts as a source of knowledge for the evaluation, coordinates feedback from HQ and from the field; supports the identification of external stakeholders to be consulted; and plays a key role in disseminating the findings of the evaluation and implementation of the management response
- The **Evaluation Steering Committee (SC)** is composed of the Executive Directors/Directors of the joint evaluation partners and donor countries. This group are tasked with providing political and institutional support for the evaluation at the highest level; endorsing a robust and credible evaluation; and ensuring the use of the findings and recommendations developed through formalized management responses and associated action plans.

Terms of Reference for each of these groups are available on request.⁸

1.3 Purpose, objectives and scope of evaluation

Purpose

The **purpose** of the Evaluation, as articulated in its Terms of Reference (ToR), is 'to provide credible and useful evaluative information on the added value of JGPs in enhancing achievement of results on gender equality and women's empowerment through improved UN system coherence and efficiency by using joint design and implementation processes'.⁹

⁸ Available from the UN Women Evaluation Office

⁹ Terms of Reference

The **objectives** of the Evaluation are similarly presented within the ToR: these are, ‘to assess, taking into account local and national circumstances:¹⁰

- the overall contribution of JGPs to national development results on GEWE, including intended and unintended results and the efficiency of JGPs in achieving their objectives
- the extent to which JGP focus and support are relevant to UN and national development goals and policies
- the overall sustainability of JGP results, including the level of national ownership, national capacity development, partnerships between the UN system and national partners, as well as sustainability aspects in programme design and programme exit strategies
- the extent to which JGPs have created synergies that contribute to gender equality and women’s empowerment in UN efforts at the national level
- the overall level of integration of a human rights based approach in JGPs

These broad objectives are interpreted and operationalised for the study within the Evaluation Questions and design, below.

Scope

The study has both summative and formative aspects. Its two key **perspectives** are:

- **Jointness** – arising from the post-1997 (and particularly post-2006) reform agenda to create a more effective and efficient UN, and the Paris principles of aid effectiveness. The practical implications of jointness, including design, modalities, processes, management, performance assessment and results, will be assessed from the perspective of participating UN agencies and national partners. The presumed added value of working through joint programmes – the core thesis of the Evaluation - will also be explored and assessed
- **Results at the national level**, including the contribution of joint programming to national development results on gender equality and women’s empowerment (GEWE), sustainability (including national ownership), national capacity development, partnerships between the UN system and national partners and improved co-ordination in UN efforts for GEWE nationally.

The Evaluation examines JGPs operating at national level established between 2006-2010, across a range of geographical and pre-defined thematic areas (Eliminating Violence Against Women: Health including HIV and AIDS; Education: Economic Empowerment; Governance; Human Trafficking; and Integrated programmes).¹¹ Aspects not covered are:

¹⁰ Some changes have been made to wording, based on input from the Evaluation Management Group

¹¹ The thematic areas were set during the preparatory stage of the study. See Portfolio Review (UN Women 2011)

- Programmes implemented before 2006, since a Portfolio Review , conducted by UN Women as part of the preparatory stage of the Evaluation,¹² established that JGPs prior to 2006 lack sufficient documentary record
- Programmes designed post-2010, since these will be insufficiently established to assess progress through the programme cycle and to enable assessment of results
- Programmes which operate at regional level – this was a decision taken in the preparatory stage of the Evaluation by the Secretariat and Evaluation Management Group, who wish to concentrate efforts on results at national level
- One programme in portfolio which operates at a global level
- Programmes which mainstream gender, rather than those with an explicit objective of GEWE and/or which place women and girls as the main beneficiaries/programme partners. This was again a decision of the preparatory phase, since in theory all JPs mainstream gender, and would therefore be eligible for assessment.

Preparatory and Inception Phases

This Inception Report is a culmination of two phases of the Evaluation; firstly, a Preparatory phase undertaken internally by partner agencies; and secondly, an intensive Inception Phase undertaken by the Evaluation Team. Methods applied and products developed (at the request of the Evaluation Management Group) during the latter are as follows:

Preparatory Phase

Scan of JGPs and development of a database (produced in 2010 by the Evaluation Unit of the former United Nations Development Fund for Women (UNIFEM))

Production of a more comprehensive Portfolio Analysis of JGPs, based on the initial UNIFEM scan (conducted in 2011 by the UN Women Evaluation Office)

Developing the set of Evaluation Questions, through a fully consultative process

Reaching out to UN agencies and donor countries to partner in the Evaluation in light of the 2002 Secretary General's Report and General Assembly resolution 62/208 encouraging UN agencies to conduct joint evaluations, and given the collaborative nature of JGPs. ¹³ (undertaken by UN Women in 2011)

Design and development of the Terms of Reference for the Evaluation

¹² UN Women (2011) Analytical Overview of Joint UN Gender Portfolio: Scoping for the JGP Evaluation Final Report

¹³ See the 2002 Secretary General's report "Strengthening the United Nations: an Agenda for Further Change" (A/57/387)

Inception Phase

Initial inception discussions with the Evaluation Management Group and Secretariat around study design, approach and workplan

Screening of an initial sample of 21 JGP and other documents for data availability/completeness/indicative substantive issues

Inception Mission to New York to meet with the Evaluation Management and Reference Group, including understanding of their perspectives on the study, and how utilisation can be maximised

Over 70 semi-structured phone interviews in person and by phone to key stakeholders, including their perspectives on how to maximise the use of the study

Development of sampling criteria and main sample base for analysis 24 JGPs to be included for in-depth desk review

Refinement of evaluation questions (as presented in the ToR)

A full 2-day Team Meeting in London, to orient the full team on the approach, methodology and Evaluation Matrix for the study

Draft and revised Inception Reports

A combination of document review; discussions with the EMG; and over 70 detailed interviews conducted both in person during the Inception mission and by phone during the follow-up period, have yielded valuable perspectives and information.

During the Inception period, initial screening of documentation identified a number of key substantive issues and lines of enquiry arising.

- It is important to unpack the assumptions and definitions (e.g. ownership) contained within the design of JGPs
- JGPs often fail to articulate an explicit theory of change or provide clarity on choice and rationale for the selected intervention model
- There has been limited analysis around the conduciveness of the environment for JGPs e.g. aid coordination, policy coherence within countries UN system
- JGPs generally present a high level of ambition based on time and resources available;
- Documents are rarely explicit about if and how human rights based approaches were used in planning and programming (beyond acknowledgment of human rights instruments);
- There are apparent shortcomings in strength of the framework for accountability and the coordination mechanisms for JGPs
- The role of and relationship with Gender Theme Groups is highly varied
- Many JGPs have experienced delays and bottlenecks in implementation
- Monitoring and reporting against results is often limited in documentation

- Little explicit attention to sustainability is present within the JGP documentation

These substantive issues have contributed significantly to the Evaluation design, particularly the revision of the Evaluation Questions, the Evaluation Matrix, interview guides and analytical tools described in Sections 3 and 4 below.

Outputs from the Inception Phase were:

- A workplan approved by the Evaluation Management Group
- Protocols for semi-structured interviews with key stakeholders
- Set of proposed revised Evaluation Questions
- Proposed sampling criteria and main sampling base for detailed desk review
- Set of proposed Analytical Categories for detailed Desk Review of JGPs
- An Inception Mission Note, which captured sampling approaches, substantive and data issues arising from initial document review,¹⁴ and which are addressed below
- Proposal for field study sampling of JGPs

Subsequent stages

The Evaluation is organised into the following subsequent stages, described below:

- Detailed desk review (August 2012)
- Field study of JGPs (September-November 2012)
- Synthesis (December 2012-March 2013)

Structure of the Inception Report

This Inception report builds on the multiple products generated by the Inception Phase by:

- Taking into account the perspectives of the EMG and other stakeholders expressed through interviews and review of documents
- Clarifying the Evaluation object (JGPs in the UN System) and Evaluation Questions
- Operationalizing the scope, focus, questions, methodology and tasks within the ToR
- Presenting a methodology for sampling

¹⁴ Available from the Evaluation Secretariat

- Identifying constraints on evaluability and limitations identified so far;
- Setting out the organisation, conduct and deliverables of the study, including the timing for detailed desk review
- Making proposals for content and length of subsequent reports

The report is organised as follows:

1. **Introduction** including context, purpose, objectives, scope and process
2. **Theoretical and Conceptual Basis** of the Evaluation design
3. **Evaluability Assessment** including completeness of the dataset; evaluation criteria; evidence gaps and proposed mitigation strategies; revised evaluation questions; feasibility and limitations
4. **Evaluation Methodology and Approach** including overall analytical framework; methods to be applied in each stage of study; sampling criteria; detailed description of phases of study
5. **Final outputs** of the study and quality assurance
6. **Annexes**

A **Glossary** of the terms used within the Evaluation is available at Annex 2.

2. Theoretical and Conceptual basis of the Evaluation

This section of the Inception Report sets out the theoretical basis of JGPs, and the conceptual approach to the Evaluation. It starts with a discussion of the Evaluation object (JGPs); and then sets out the consequent conceptual approach adopted, situating this within recent approaches to development evaluation.

2.1 Theoretical basis of JGPs

The definition of Joint Programmes applied for the Evaluation is the current one for the UN System, provided by UNDG.¹⁵ ‘A joint programme is one that is undertaken within the framework of a joint programme document signed by all partners, governed by a joint committee and that adopts an agreed fund management modality’.¹⁶ Within this, the parameters of a JGP, set out above, are applied.

The associated UNDG Guidance on Joint Programming does not prescribe a single ‘model’ for Joint Programmes.¹⁷ Data analysis during the Inception Phase confirmed the diversity of JGPs; their varied aims, objectives and intended results; diverse designs and implementing strategies; different operational modalities and partnerships. JGPs respond to very different contexts – from fragile situations to countries with mature and well-harmonised aid architectures and/ or those which are Delivering As One or self-starter pilots. As evaluation objects, therefore, they are diverse and highly complex.

Yet as the EMG has made clear,¹⁸ setting out a common theoretical basis for Joint Programmes is a crucial step. In the absence of any explicit overarching programme theory, therefore, the Evaluation Team has applied the source UNDG Guidance, supplemented with data gathered and analysed during the Inception Phase, including the substantive issues arising, above. This has resulted in the production of a very preliminary programme theory,¹⁹ below. It is structured around the features of JGPs (which could be considered as strategies); the process-level and interim results articulated in the Guidance and within programme documents surveyed; and the overall objectives conveyed in terms of better development results for GEWE.

¹⁵ Terms of Reference

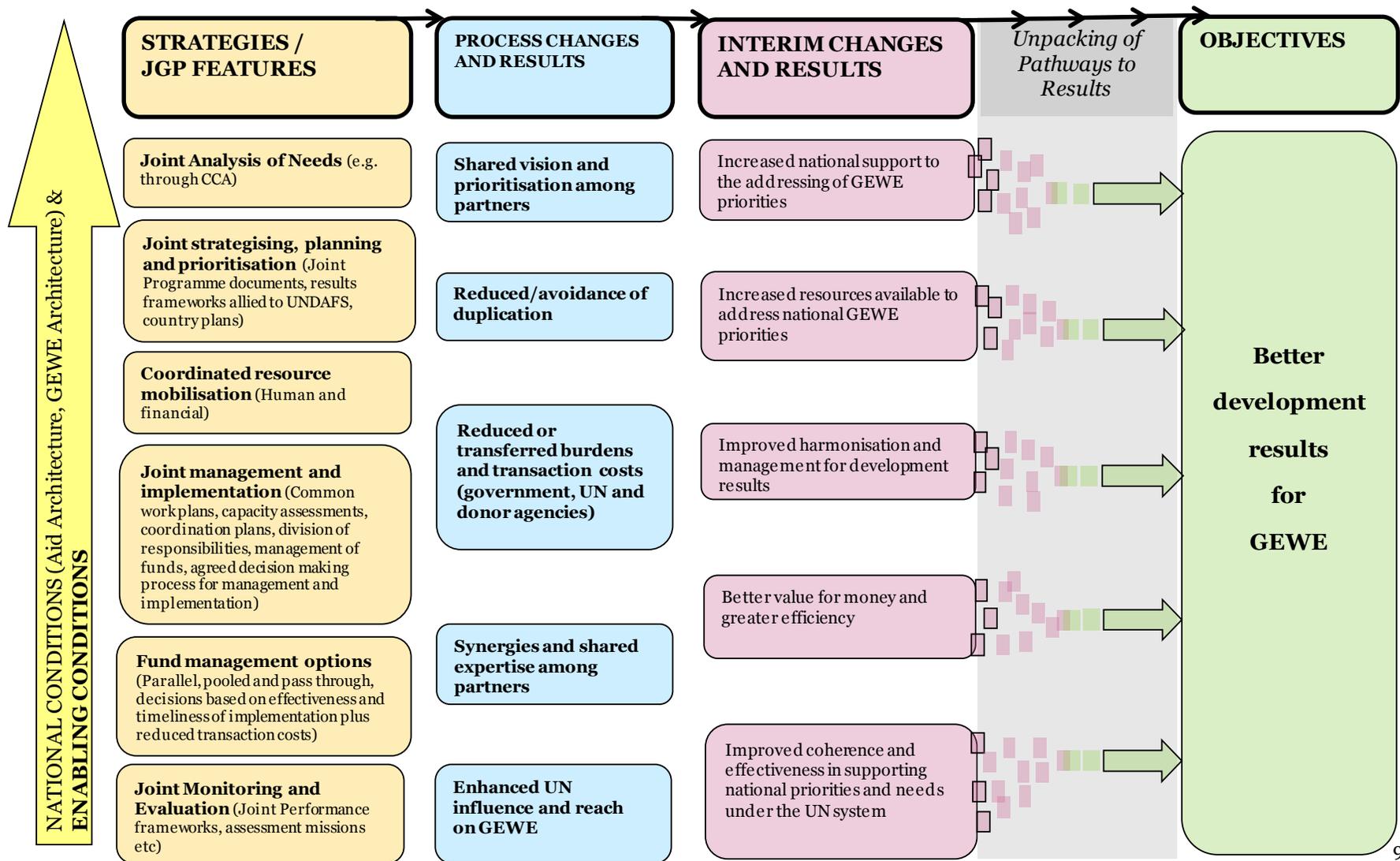
¹⁶ UNDG (2003) Guidance Note on Joint Programming

¹⁷ As applies to Joint Evaluations for example - ‘There are various degrees of “jointness” depending on the extent to which individual partners cooperate in the evaluation process, merge their evaluation resources and combine their evaluation reporting (OECD DAC 2010).

¹⁸ Inception Mission meeting plus feedback on first draft of Inception Report

¹⁹ The classic definition of a programme theory is “a specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated’ Chen (1990).

Figure 1: Preliminary Programme Theory



Some of the main weaknesses in the preliminary programme theory are: the need for intended results to be more clearly articulated, particularly at process and interim levels; the lack of explicit interconnections between the stages of different results; and the absence of clearly articulated assumptions at different levels. For example, reducing duplication and transaction costs (process level) may not automatically lead to improved harmonisation and management for development results (interim change / result). Even more significant is the gap – often linked to assumptions – between such interim results and the intended objectives i.e. whether and how improving UN effectiveness and efficiency in supporting national priorities and needs will lead to development results for GEWE.

A major task of the Evaluation will be to unpack these interconnections and pathways, and to clarify and confirm the often flawed assumptions that underlie them. Some initial work here has been undertaken during the Inception Phase; this will be developed during the study, as explained below.

Box 1: Assumptions identified to date in UNDG / JGP documentation:

- The national context and aid architecture is conducive to joint programming, including capacity and political will for design and implementation
- Efforts to provide an enabling environment will be made by headquarters, regional and field offices management and staff
- A common vision, intended results for GEWE and the modalities for supporting programme implementation exists or can be developed
- Sufficient incentives and willingness exist for partner agencies to work operate jointly, even where incentives are external; and insufficient institutional or organisational barriers exist
- Partner UN and national agencies have the capacity and will to strategize and devise implementation methods to support common results
- The capacity and resource (human, financial and time) requirements of joint planning and coordination are considered and addressed
- Differences among agencies in methodology and approach can be identified and resolved at an early stage
- Adequate dedicated expertise to address GEWE is available within the UN and partner agencies involved in JGPs
- Monitoring and Evaluation planning can take place within the framework of the UNDAF, and information and progress updates will be shared
- Fund management options will be selected on the basis of how to achieve the most effective, efficient and timely implementation, and to reduce transaction costs for national partners, donors and the UN, rather than on the basis of expediency.
- Joint operation will lead to reduced or transferred transaction costs and burdens
- Those involved in JGPs understand the requirements and ways of working appropriate to joint programmes
- Joint accountability (upwards, downwards, across) can be achieved with relatively light touch measures to support this

This preliminary programme theory, including the assumptions above, forms the central plank of Evaluation design. It is of course only indicative at this stage; the process for its interrogation,

refining and development is described below. A refined programme theory, which is relevant to the future planning and implementation of JGPs, will be one of the key outputs of the Evaluation.

2.2 Conceptual approach to the evaluation

The conceptual approach to study design has four main pillars; the application of a theory-based approach to a collective case study; the use of contribution analysis; a utilisation focus; and the application of a human rights based approach. These are explained below.

Collective case study and theory based approaches

The Evaluation is essentially a collective case study across multiple cases.²⁰ The unit of analysis is the individual and collective set of JGPs. Modern approaches to case study regard the case as a complex entity, operating in a ‘real world’ setting, in which multiple ‘causes’ and influences interact not always predictably.²¹ JGPs, as stated, are highly complex objects, which operate in complex and diverse environments.

A helpful way forward is to apply theory-based approaches to JGP evaluation.²² These emphasise context, presume complexity, and focus on seeking out ‘causative pathways’ – or the explanations for why what happened, happened, to test the theories implicit in an intervention’s design. Theory-based approaches are well-suited to evaluations which seek to assess highly complex and multivariate aid processes which are operating in multiple and diverse settings.²³

A theory based approach will be applied in the Evaluation by a) emphasising the importance of context in affecting JGP design and results b) searching out the explanations and reasons for events c) testing, validating and elucidating the indicative programme theory above and d) generating ‘predictive statements’ (i.e.. a broader programme theory, or theory of change) at final synthesis level, that can inform the subsequent design and implementation of JGPs.

Contribution analysis

Contribution analysis²⁴ aims to construct a ‘contribution story’ by building up the evidence to show the contribution made by an intervention whilst also establishing the relative importance of other influences on outcomes. For JGPs, for example, a contribution analysis approach would assume that many reasons contribute to the achievement of development results on GEWE, with JGPs playing a contributory role. The challenge for the Evaluation is to describe the nature – or ‘pathways’ - of this contributory role.²⁵

Contribution analysis is particularly appropriate for an evaluation of JGPs because it recognises that attributing development results on GEWE directly to JGPs is generally complex and often unfeasible. It also enables the Evaluation to explore the very complex interplay between JGP activities, contexts and outcomes – part of the complexity paradigm above.

²⁰ See e.g. Byrne and Ragin (2009)

²¹ Stern et al (2012)

²² Stern *2009) (e.g. theory of change (Connell et al 1995) and realist approaches (Pawson and Tilley 1997; Koenig 2009.

²³ Theory based approaches have been applied in several major cross-national studies in which the evaluators have been involved such as the global Evaluation of the Paris Declaration Phase II across 21 countries and 7 donor agencies

²⁴ Developed by Mayne (2001) A recent issue of the journal ‘Evaluation’ summarises current thinking on Contribution Analysis

²⁵ White and Phillips (2012)

Contribution analysis involves a structured approach of a) establishing the results of an initiative (here the results at different levels within the programme theory) b) establishing the contributions of the intervention (JGP) to those results as well as any other possible explanations and c) establishing the pathways of contribution that have occurred. Section 4.2 describes how this has been applied within the Evaluation Matrix.

Utilisation-focused evaluation

Utilization-focused evaluation also part of the new wave of thinking in development evaluation.²⁶ This begins with the premise that evaluations should be judged by their utility and actual use.²⁷ Evaluations therefore need to be designed and implemented with careful consideration of how everything that is done, from beginning to end, will affect use.

There is no single ‘right’ way of doing any given evaluation to maximise use. The approach emphasises the importance of evaluators and stakeholders working closely together, recognising the responsibility of stakeholders (here represented in the first instance by the Evaluation’s Management and Reference Group (EMG and ERG) members) for applying evaluation findings and implementing recommendations. Utilization focused approaches have much in common with the UNEG approach to participation and inclusion (See Table 5 below and Glossary in Annex 2).

Groups of stakeholders, separate from the partners to this Evaluation, identified to date²⁸ are:

- Government and other national duty-bearers including Ministries of Planning and Finance, Gender or similar, National Women’s Machineries, and local governments
- Civil society, including National Women’s Movements, networks and federations of women’s groups; groups representing rights holders involved in JGPs
- Donor agencies, including bilateral and multilateral agencies
- UN agencies including those involved in JGPs and other partner agencies, including their governing bodies, senior management, joint programme managers/focal points and gender advisers; the UN High Level Panel for System-Wide Coherence; the United Nations Development Coordination Operations Office (UNDOCO); the UN Multi-Party Trust Fund the preparation of 2012 Quadrennial Comprehensive Policy Review (QCPR) of UN system operational activities by the General Assembly; the UNDG; and the wider community of UN agencies engaging in joint programmes
- International and regional evaluation networks

Utilization-focused evaluation also emphasises however that participation should be relevant and appropriate rather than aiming for breadth for the sake of breadth. ‘High *quality* participation is the goal, not high quantity participation.’²⁹

The Evaluation adopts a utilization (including participatory and inclusive) focus in a range of ways:

- Firstly through ongoing engagement with the EMG and ERG, who are themselves representatives of a far wider group of stakeholders³⁰

²⁶ Quinn Patton (2000)

²⁷ Use’ ‘concerns how real people in the real world apply evaluation findings and experience the evaluation process’ ibid.

²⁸ Terms of Reference

²⁹ Quinn Patton (2000)

- Secondly through an extensive round of interviews (over 70 to date within the Inception phase alone) with stakeholders identified by the Secretariat, whose aim was not merely to extract information, but to listen to needs and concerns around the Evaluation, to assess the focus and relevance, and therefore help refine, the pre-set questions, and to inform design. All interviewees were asked the question ‘How can this Evaluation be made most useful for you?’.
- Thirdly by outreach to stakeholders through an Electronic Survey, described below, in order that those whose views and opinions cannot be accessed through interview have the opportunity to contribute and participate
- Fourthly through a further range of interviews – over 100 in total - to be conducted during detailed desk review and field study of JGPs which will adopt the same principles as above
- Fifthly through being committed to the production of a range of appropriate evaluation products, and producing these in as timely way as feasible to inform other aspects of the development agenda, such as the Quadrennial Comprehensive Policy Review.³¹

Perspectives on Evaluation relevance and use voiced by interviewees during and subsequent to the Inception Mission have been integrated into design. Primarily, these orient around requests in reporting for: clear demonstration of results; the identification and citing of good practices; the production of recommendations which are sufficiently clear and focused to inform programming at national level. Stakeholders also expressed a wish for a communications strategy around the Evaluation findings (below) and the associated development of products to inform policy processes going forward. Those elements which fall within the realm of the evaluators have been integrated into the Evaluation design, below.

As the main architect of the utilization approach, Michael Patton, however, explains: ‘*Evaluation use is too important to be left to evaluators*’.³² During the Inception Mission, the Evaluation Team asked the EMG to propose ways in which use within their individual agencies could be maximized. A communications and dissemination strategy is under development by the EMG, and the idea of further engagement with Knowledge Management functions within agencies has been proposed. The responsibility for this rests with the EMG, but the Evaluation Team have voiced their commitment to contributing as far as they can, for example through supporting the production of more user-friendly products than a necessarily dense and detailed Inception Report.

Of importance is maximising the use of the report to inform the wider development agenda going forward, and to maximise utility beyond the partner agencies involved in this study. This needs to be decided and owned, as stated, by the EMG, but ideas might include: the QCPR process itself: regional meetings of Resident Co-ordinators; meetings of the UN High Level Panel for System-Wide Coherence and UNDOCO; and regional donor / evaluation network meetings. The joint management response to the Evaluation by the partner agencies involved will also be a critical aspect of communications.

³⁰ See the list of EMG and ERG members at Annex 14

³¹ The quadrennial comprehensive policy review (QCPR) is the primary policy instrument of the General Assembly to define the way the UN development system operates to support programme countries in their development efforts. The QCPR deals with issues of funding UN operational activities for development, the functioning of the UN development system and the development effectiveness of the work of the UN system for development. <http://www.un.org/en/ecosoc/about/qcpr.shtml> accessed 3.8.12

³² Ibid.

Application of a human rights based and gender equality approach to the evaluation

UNEG guidance states that ‘An evaluation that is Human Rights & Gender Equality (HR/GE) responsive addresses the programming principles required by a human rights based approach and gender mainstreaming strategy. It contributes to the social and economic change process that is at the heart of most development programming by identifying and analyzing the inequalities, discriminatory practices and unjust power relations that are central to development problems. HR & GE responsive evaluation, can lead to more effective interventions and better, more sustainable results’.³³

The Evaluation Team are committed to the application of human rights and gender equality-sensitive processes during the Evaluation, which has two aspects: firstly, the substantive aspects related to HR/GE issues within JGPs; and secondly, the application of a HR/GE-sensitive approach to the Evaluation.

Approaches to the former are described in Sections 3 and 4 below. For the latter, the principles of the UNEG Guidance will be fully applied throughout. These include: framing the evaluation questions to ensure the recognition of HR and GE issues; the use of disaggregated data where feasible; ensuring stakeholder participation to the maximum extent possible (see Utilization, above); using mixed method approaches; and ensuring that individual methods, analytical tools and reporting incorporate HR/GE issues. Particular attention will be paid to human rights dimensions during identification of stakeholders and their interests, during selection of interviewees and in the selection of methods, timeframe etc.

The evaluation will be conducted in full coherence with the UNEG Ethical Guidelines and Code of Conduct. Key features of the ethical code to be applied are:

- Respecting gender and human rights principles throughout the Evaluation process, including; the protection of confidentiality; the protection of rights; the protection of dignity and welfare of people; and ensuring informed consent. Feedback will be provided to participants wherever possible, and data validation will take place at all levels with participant consent
- Maximising the degree of participation of stakeholders in the Evaluation itself wherever feasible and a commitment to using participatory approaches in field studies in particular
- Ensuring that the Evaluation matrix (above) integrates CEDAW and human rights commitments
- Disaggregating data by gender and social group where feasible
- Ensuring that outputs use human-rights and gender-sensitive language
- Ensuring that, where possible, Evaluation findings are disseminated back via partner agencies and through the dissemination strategy to stakeholders including rights holders and duty-bearers where appropriate and feasible.

³³ UNEG (2011) Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance

3. Evaluability Assessment

This section of the report considers the key evaluability issues for the Evaluation identified during the Inception phase and subsequently. It covers: data availability based on review to date; information gaps arising; options for providing a counterfactual; and how comparability across cases can be assured. In the light of this, revisions to the original set of Evaluation Questions are described and Evaluation Criteria proposed. Finally, a summary of feasibility, including limitations, is presented.

3.1 Completeness of documentation set

A key task undertaken during the Inception Phase was an initial screening of the documentation available on JGPs, plus other key texts. This aimed to assess the completeness of the documentation available and the quality, content and coverage of data within it. A stratified random sample of JGP information was screened against specified analytical parameters.³⁴ The extent of substantive information in key areas (theories of change, results information, design processes, and others) was assessed; and any information that might inform the sampling process and analytical categories for the next phase of detailed desk review captured.

There were some initial concerns about data paucity. The initial sample of 21 JGPs contained nine programmes for which only the basic design document (Prodoc) was available. This provided useful information on programme context, intended content and proposed governance structures for JGPs, as well as forming a useful reference point for comparing the intentions of JGPs with the actuality, but minimal data on design processes; theories of change; or results. However, document search during detailed desk review (currently ongoing) found sufficient evidence available to allay concerns around evaluability, including data in the three areas mentioned above. Only two JGPs had to be substituted in the sample for lack of information. Two were removed because the programme was non-operational.³⁵

During the Inception Mission of June 2012, the EMG requested a strategy for addressing data paucity at the next main stage of data analysis: detailed desk review. This essentially comprised the need for at least 30% of sample JGPs satisfying minimum data requirements. The bar was set at a relatively low level: a design document plus some form of results or evaluative reporting. This strategy has ensured that programmes with lesser amounts of evaluative information are not screened out, whilst also reducing the possibility of a biased sample, where well-managed programmes, which have consequently greater data availability, are unduly represented.

3.2 Specific data gaps

Initial data screening highlighted some potential gaps in the evidence base.³⁶ Ongoing data search and analysis has addressed some of these, but potential weaknesses in information relating to

³⁴ For the initial screening the following parameters were used: Region, Budget, Theme, UN-Lead (including unspecified), Funding Source, DAO (original pilot countries) and Conflict/Post-Conflict (World Bank 2011 harmonised list). See Annex 4

³⁵ See Section 4.4. Those substituted were: Equatorial Guinea and Ecuador (programmes non-operational / not official); Sierra Leone and Moldova (lack of evaluative information within the required timeframe)

³⁶ As well as the four main areas cited below, these included: lack of evidence on outreach to meso and micro levels; limited analysis of the operating context; limited information on design processes; lack of information on sustainability and resource-raising strategies. These concerns appear to be mitigated to some extent by detailed desk review.

theories of change; results; baselines and efficiency/ transaction costs remain. Potential strategies to address these are set out below.

a) Theory of Change / programme theory: With some exceptions,³⁷ few of the JGP documents so far have revealed clear theories of change or logics of intervention. Addressing this will imply making implicit theories explicit, something that is feasible for field study of JGPs, where programme-specific theories can be developed and validated, incorporating views of stakeholders (UN and donor agency, government and civil society) at national level. The analytical tools for desk review, below, are also designed to extract the elements of a theory of change in desk study of JGPs, though full theories of change cannot be developed. The description of the Synthesis process below specifies how the overarching programme theory will be developed.

b) Results: While almost all JGPs have a Results Matrix, initial screening found limitations to the technical merits of these and, perhaps more significantly, a paucity of reporting in some cases against results. This caused concern during the Inception Phase of the evaluation. Yet the potential for achieving and tracking development results on GEWE should in theory be greater in a context where multiple-partners contribute

Documentation search for detailed desk review has however found that evaluative information such as Annual Reports, monitoring returns and Mid Term Reviews and evaluation reports are available, although the quality of information within them is varied. A fuller picture will be available at the end of detailed desk review stage. Alternate sources such as UNDAF Annual Reviews, Resident Co-ordinator annual reports, national reports (including Poverty Reduction Strategy or Gender Policy reports) and DAO and donor reports will be sought out in mitigation. Field study is likely to provide the strongest source of data here, and results chains will be constructed to the highest level at which robust data is available, as per the programme theory.

Successfully mapping the pathways, or ‘stories’ of plausible contribution will depend on the levels of results information available, which is likely to be different for each programme. This will be tackled by clearly delineating the different *levels* of results available as they relate to the programme theory. It should be feasible to identify, through triangulated evidence, pathways of contribution to interim level at least, even where impact level information is not available, as well as any unintended results.

Where baselines are not available, careful triangulation will be needed to firmly establish results available (particularly important for the application of contribution analysis) using combined sources such as programmatic information, national datasets and records, and qualitative information such as interview and survey. Where results cannot be robustly established, at any of the different levels within the programme theory, analysis will report frankly on the measures of progress available, plus any apparent contributions to results, but be clear on the limitations of a systematic application of contribution analysis.

c) Efficiency and transaction costs: The UNDG definition of transaction costs focuses narrowly on costs related to programming that do not directly affect beneficiaries, such as monitoring and evaluation.³⁸ No internationally agreed definition for the term exists,³⁹ but even applying the broader concept of efficiency,⁴⁰ initial and subsequent data analysis so far shows only

³⁷ For example JGPs in Vietnam and Namibia

³⁸ See UNDG Guidance (2003)

³⁹ See e.g. Lawson (2009)

⁴⁰ See Glossary (Annex 2)

limited information available beyond operations and funding/disbursement concerns.⁴¹ UNEG guidance points out that results on gender mainly happen in the long term and are complex, so may not lend themselves to traditional efficiency analysis.⁴²

Other studies,⁴³ grappling with the same dilemma, have applied the concept of aid burdens to represent the administrative, management and legal obligations incurred by partners (donors, partner governments, UN and other agencies) which occur within the aid relationship.⁴⁴ This study will also apply the term, understood as whether working jointly has maximised the use of resources; allowed for cost-sharing; reduced time and resource requirements for partners; streamlined management and administrative burdens; and affected the pace of implementation for national partners and participating UN organizations. This information is likely to be more readily available from documentation and interviews which comprise detailed desk review, as well as from field study. It will be sought out through specific targeted questions. Where information is available to support efficiency, such as costing models, this will be analysed.

3.3 Counterfactual

‘Counterfactual logic seeks to answer the question: ‘what would have happened without the intervention?’ by comparing an observable world with a theoretical one, where the latter is intended to be identical to the former except for the presence of the cause and effect. The latter is described as ‘counterfactual’ because it cannot be observed empirically.’⁴⁵

This Evaluation does not meet the conditions required for a true counterfactual, which in any event is not methodologically appropriate for the nature and object of the enquiry.⁴⁶ Reflecting these concerns, a counterfactual was not integrated into the design of the Evaluation from the outset.

That said, the central thesis of this Evaluation and programme theory – that operating jointly has advantages over a operating in alternate ways, such as via single-agency programmes – implies the need for a comparator. If not full counterfactual then *comparative* aspects need to be addressed.

Detailed desk review of programmes is still ongoing, but examples of possible single-agency comparator programmes, (even using minimal parameters of donor- of UN agency-led, GEWE initiatives in approximately the same timeframe) operating alongside the JGPs under study are few and lack features that might make them suitable comparators. This may change but cannot be guaranteed. Of note is the fact that the UNFPA evaluation of JPs in Reproductive Health found that such comparisons were not possible based on the information available to the evaluation team.⁴⁷ Feasibility issues are also key, given time and resource limitations of the study. However, examples of possible comparator JGPs are available; indeed, one issue arising is that in some contexts, multiple JGPs are operating concurrently – raising the wider, and very significant, issue of whether the UN system as a whole is programming effectively and efficiently for GEWE in a given context.

⁴¹ This is not unique to JGPs - the UNDG Iraq Trust Fund Lessons Learned Exercise (2011) found little insight provided in evaluation reports to enable in-depth analysis of cost efficiency.

⁴² See UNEG (2011) Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance

⁴³ See for example the Paris Declaration Evaluation Phase 2 (Wood et al, 2010)

⁴⁴ See the Glossary (Annex 2)

⁴⁵ Stern et al (2012)

⁴⁶ ‘Counterfactuals answer contingent, setting-specific causal questions ‘did it work there and then’ and cannot be used for generalization to other settings and timeframes, unless they are accompanied by more fine-grained knowledge on the causal mechanisms actually operating within the process leading from potential cause to effect. Stern et al (2012)

⁴⁷ John Snow Inc (2008)

These challenges are common to multi-case studies. Recent work proposes alternate means of addressing ‘counterfactual inference’ through a) a focus on comparison, even where ‘perfect’ comparators are not available, and b) the identification/confirmation of causal processes or ‘chains’, as well as identifying the supporting factors and mechanisms at work in the context.⁴⁸ This can be applied to internal comparison across JGPs at least; and potentially to external comparison with other programmes which share common features or attributes, even where these are not perfect comparators. This approach will also allow the wider issue of UN programming for GEWE at country level as a whole to be integrated.

Presuming a body of evaluative information exists on these programmes already,⁴⁹ the following strategies will be applied within the different phases of the Evaluation:

- a) Across the 24 different cases for desk review, identify attributes that can be internally compared, and conduct analysis across these
- b) For field study selection, apply the criterion of more than one JGP / a single-agency gender initiative operating in a country, for which evaluative information exists, as a selection criterion for at least three of the five projected field studies. Although parallel evaluations cannot be conducted, carry out limited interviews to support comparative analysis across the attributes of a broader body of JGPs within the same operating environment
- c) Within field study, assess in the round the range of other UN JGPs in operation in a country, exploring assessing the role (with particular attention to efficiency and effectiveness) of the specific JGP under study within this broader programming environment
- d) At synthesis stage, include analysis of a body of comparator programme evaluations (single agency or joint gender-focused initiatives) within the evidence base, applying similar sampling parameters to those for desk study selection. Conduct comparative analysis across findings; and clarify some of the explanatory factors – why events may plausibly have happened, or not happened, in JGPs as opposed to in single-agency programmes.

The alternative strategy, of redesigning the study to be fully comparative (e.g.: detailed desk study of 12 JGPs and 12 single-agency programmes; field study of JGPs in three countries, and single-agency interventions in three others), has been rejected by the EMG as unfeasible⁵⁰ and to have disadvantages in terms of a reduced body of evidence from field study of JGPs; requiring a highly purposive sample of JGPs; and a significant extension of the study’s timeline.

3.4 Revising the Evaluation questions

Given the findings from the Inception Phase above, the Evaluation Team embarked on a process of revision of the original Evaluation Questions. These had been developed prior to commissioning the Team, by a consultative process across partner agencies to the Evaluation, co-ordinated by the Evaluation Secretariat.

The original set of Evaluation Questions comprised a list of some 27 sub-questions, grouped under 4 Strategic Priority questions, as follows:

⁴⁸ See Stern et al (2012)

⁴⁹ This is a critical factor, since the Evaluation cannot undertake parallel evaluations

⁵⁰ Feedback on Revised Draft Inception Report, August 2012.

Box 2: Strategic Priority Questions

Strategic Priority 1: Design of JGPs To what extent have JGPs been conceptualized, planned and designed jointly to respond to international, regional and national commitments on GE/WE?

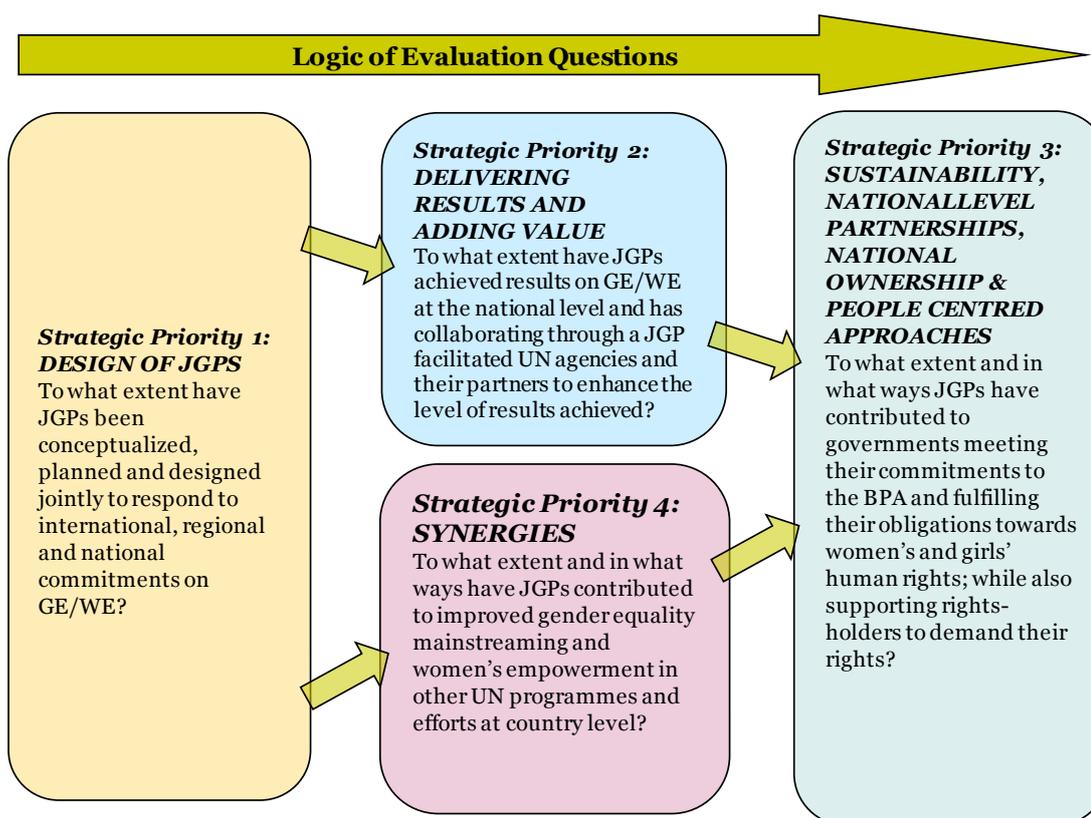
Strategic Priority 2: Delivering Results and Added Value To what extent have JGPs achieved results on GE/WE at the national level and has collaborating through a JGP facilitated UN agencies and their partners to enhance the level of results?

Strategic Priority 3: Sustainability, National Level Partnerships, National Ownership & People Centred Approaches To what extent and in what ways JGPs have contributed to governments meeting their commitments to the BPA and fulfilling their obligations towards women’s and girls’ human rights; while also supporting rights-holders to demand their rights?

Strategic Priority 4: Synergies To what extent and in what ways have JGPs contributed to improved gender equality mainstreaming and women’s empowerment in other UN programmes and efforts at country level?

The diagram below sets out the logic and flow across these questions:

Table 2: Logic and flow of Strategic Priority Questions



Sub-questions

The original list of sub-questions was excessively broad for feasible evaluation. Some of the sub-questions themselves were very large in scope, being broken down into 5 or more sub-components.⁵¹ The Evaluation Team therefore embarked on a process of narrowing down and focusing these questions in the light of the substantive data gathered during the Inception phase and the evaluability assessment above. This resulted in a refined set of 18 evaluation sub-questions under the main Strategic Priority Questions, which the Evaluation Team understands to be fixed.

These sub-questions are conceptualised in two ways:

- a) Firstly, they form a set of dimensions or components of the Strategic Priority Questions. Combined, and subject to data availability, the sub-questions should enable a comprehensive response to the overall Strategic Priority Question.
- b) Secondly, the sub-questions integrate the set of assumptions extracted so far, and set out above, which underlie joint programmes and their contributions to results on GEWE. Integrating these will allow them to be tested / interrogated / validated/expanded through the evaluative process, as part of developing the overarching programme theory.

The first refined set of questions were presented to the EMG during the Inception Mission, informed by the substantive issues arising from initial screening of documents. Feedback on the first draft of the Inception Report from UNFPA in particular requested an in-depth presentation of the Evaluation Questions, including their logic, understanding, the logic applied to “break down” / “deconstruct” and operationalize each question, as well as the overall logic and approach to be adopted to answer them. A useful template for this purpose was provided, although the Evaluation Team found this to be most appropriate for use at the level of the Strategic Priority questions.

The final proposed list of Evaluation Strategic Priority Questions and their associated sub-questions; their rationales;⁵² the theory of the upwards logic from sub-questions to Strategic Priority Questions (chain of reasoning); and the Evaluation Criteria assigned against them can be found at Annex 5. The table below presents, based on this and the evaluability analysis above, an assessment of their feasibility.

This assessment is of course only indicative at this stage; only the evaluative process itself, and field study in particular, will reveal whether questions are truly evaluable. The Evaluation should not at this stage reject an Evaluation Question, simply because the Inception Phase has not shown whether data is available (as may arise, for example, within field study). The important thing for the Evaluation is to ask the questions that matter, and to report where information is not available.

⁵¹ For example Question 1 under Strategic Priority 1: ‘When and why is a joint programme design approach considered a viable and relevant execution instrument to attain development results on GE/WE? To what extent did the level of complexity, the funding, and the expected results factor into the decision-making process?’

⁵² bearing in mind that the Strategic Priority Questions were largely pre-set by the preparatory phase

Table 3: Summary breakdown and feasibility of the Evaluation Questions

<p>Strategic Priority Question 1: DESIGN <i>To what extent have JGPs been conceptualized, planned and designed jointly to respond to international, regional and national commitments on GE/WE?</i></p>	
Sub-questions	<p>1.1. To what extent did the design process include a collaborative process, shared vision for delivering results, strategies for delivery and sharing of risks among UN partner agencies? What incentives / barriers were in place?</p> <p>1.2. To what extent were key national partners involved in conceptualization and design process? To what extent were the capacities of government and national implementing partners, as well as the capacity and conduciveness of the aid architecture, assessed during the design process?</p> <p>1.3. How has the programme prioritized the GEWE and HR needs of the country (those articulated by the national women’s movement or similar), including in fragile situations? Are JGP designs based on quality analysis?</p> <p>1.4. To what extent do JGP designs capitalise on the comparative advantage and added value (e.g. strengths and expertise) of each participating agency and establish coherence in regards to their roles, modalities and strategies?</p> <p>1.5. To what extent did JGP design processes integrate human rights principles and strategies of the HRBAP?</p>
Feasibility	<p>High, based upon the results of the inception stage. Detail:</p> <p>a) Interviews and initial document analysis showed that Questions 1.1 and 1.2 should be fully answerable through interviews during detailed desk review stage (plus some documentary evidence may be available), and through field study (providing that relevant stakeholders are available)</p> <p>b) Initial document analysis showed that Questions 1.3 and 1.5 can be fully answered through documentary review.</p> <p>b) Interviews and initial data analysis show that Question 1.4 can be answered from a range of sources: through documentary analysis; through interviews with stakeholders in country during detailed desk review stage; and through field study</p>
<p>Strategic Priority Question 2: DELIVERING RESULTS AND ADDED VALUE <i>To what extent have JGPs achieved results on GE/WE at the national level, and has collaborating through a JGP facilitated UN agencies and their partners to enhance the level of results achieved?</i></p>	
Sub-questions	<p>2.1 What evidence exists that JGPs have delivered short, medium and longer term results, from processes through to benefits? Have any unintended results been delivered?</p> <p>2.2 To what extent do UN agencies act with coherence (shared delivery strategies, division of labour, shared monitoring and measurement etc) in their implementation and performance assessment of JGPs?</p> <p>2.3 How has the joint nature of JGPs affected efficiency of delivery (reduced duplication and increased cost-sharing, reduced / transferred burdens and transaction costs)?</p> <p>2.4 Has there been effective leadership and management of JGPs at country level, including the structuring of management and administration roles to maximise results? Where does accountability lie?</p> <p>2.5 What, if any, types of innovative / good practices have been introduced in joint gender programmes for the achievement of results in GE and WE? In what contexts have these innovative practices worked better?</p>

Feasibility	<p>Medium, based on findings from the Inception Phase. Detail:</p> <p>a) Initial and subsequent document review has found limitations in the reporting of higher level results; however, data at process and interim results level appears available. The sample for detailed document review has some level of results reporting as a minimum threshold for evaluability; the strategies outlined in Section 3.2 above and 3.5 below should help mitigate the difficulty if it arises.</p> <p>b) The main data sources for Q2.2 question are phone interview and field study; the greatest source of evidence is likely to be field study, though interviews to date have already supplied some data.</p> <p>c) Initial documentation review showed little evidence available on efficiency. Question 2.3 is likely to be at least partially answerable from phone interviews and field study, but may not be fully answerable.</p> <p>d) For Question 2.4, interviews and initial documentation review have already indicated some available data on management and leadership, as well as on accountability; this has been supplemented (and triangulated) by in-person and phone interviews. Detailed desk review and field study should enable this question to be fully answered.</p> <p>e) Question 2.5 will be answered largely through field study, where forms of innovation and good practice can be identified in detail and validated / triangulated fully.</p>
<p>Strategic Priority Question 3: SUSTAINABILITY, NATIONAL LEVEL PARTNERSHIPS, NATIONAL OWNERSHIP & PEOPLE CENTRED APPROACHES <i>To what extent and in what ways JGPs have contributed to governments meeting national and international commitments to GEWE and fulfilling their obligations towards women’s and girls’ human rights; while also supporting rights-holders to demand their rights?</i></p>	
Sub-questions	<p>3.1. To what extent are JGPs integrated into national development plans and budgets, and national machinery (governmental and / or that of civil society and the women’s movement?)</p> <p>3.2. What voice and influence do key national partners including women’s movements etc, within JGP decision-making structures and hierarchy? Is consultation carried through into partnership / resource allocation?</p> <p>3.3. What steps were taken to develop and/or reinforce the operating capacities of national partners during implementation?</p> <p>3. 4. To what extent, and with what timing, have robust sustainability strategies been considered / developed / implemented? Are these feasible and appropriate?</p>
Feasibility	<p>Medium to high, based on findings from the Inception Phase. Detail:</p> <p>a) Initial and subsequent document review has found some – though limited - information on the integration of JGPs into national structures and machinery (Question 3.1); this can however be explored through field study, particularly for JGPs that have completed / are near completion. One issue highlighted by the initial document review is a lack of clarity on what is meant by ‘ownership’. The evaluation will explore the different meanings implicit or explicit in JGPs as well as the strategies to build this and their effectiveness</p> <p>b) Question 3.2 is potentially evaluable through interview at detailed desk review stage, as well as through the online survey and (mostly) through field study. Some data is already emerging through detailed desk review</p> <p>c) Some evidence has already emerged through initial analysis for Question 3.3 on capacity development intentions and results; detailed desk review and field study are likely to provide the evidence required.</p> <p>d) Initial documentation review and interviews have provided some evidence on sustainability strategies (Question 3.4) ; detailed desk review and field study are likely to provide the evidence needed.</p>
<p>Strategic Priority Question 4: SYNERGIES <i>To what extent and in what ways have JGPs contributed to improved gender equality mainstreaming and women’s empowerment in other UN programmes and efforts at country level?</i></p>	

Sub-questions	<p>4.1. To what extent have JGPs enhanced communication, planning, co-ordination and collaboration between the UN and governments/ other development partners?</p> <p>4.2. To what extent have JGPs promoted or led to improved communication, planning, coordination and collaboration among national stakeholders, e.g. between different line ministries and among government and civil society?</p> <p>4.3. To what extent have JGPs promoted or led to improved communication, coordination and information exchange within the ‘UN family’ in relation to GE/WE including between HQ and field offices?</p> <p>4.4. Are JGPs able to attract increased and new resources (such as those from non-traditional donors or philanthropic sources)?</p>
Feasibility	<p>Medium to High, based on findings from the Inception Phase. Detail:</p> <p>a) Initial document review has found some limited information, mainly in evaluations, around improved synergies between the UN and its national partners (Question 4.1). This question is most likely to be evaluable through interviews as part of detailed desk review, as well as through field study.</p> <p>b) As above, the information emerging on improved synergies among national partners (Question 4.2) is mainly contained in evaluations – though detailed desk review and field study are likely to reveal more.</p> <p>c) Some evidence has already emerged through interviews in the inception phase on co-ordination between UN agencies on GEWE (Question 4.3); there is very likely to be more available through field study and detailed desk review.</p> <p>d) Early desk review indicates some availability of funding information for JGPs, including whether new resources have been attracted (Question 4.4). This information is likely to be available through desk review and field study.</p>

3.5 Evaluation criteria and how they will be applied

The Evaluation criteria of the study were pre-assigned to the original list of Evaluation Questions. These are mainly the OECD DAC criteria, with the addition of ‘participation and inclusion’, which is drawn from UNEG guidance.⁵³ The definitions of the criteria are listed in the Glossary, attached.

The table below sets out how the Evaluation Criteria will be applied within the study. They are also reflected in the Evaluation Matrix.

Table 5: Application of Evaluation Criteria

Relevance	The planning, design and implementation processes of JGPs in relation to responsiveness and alignment with national priorities and needs, as well as national, international and UN commitments, policies and priorities, UN mandates and UNDAFs, and individual agency policies, mandates and comparative advantages. The modality of the JGP in relation to the operating context.
Effectiveness	The success or otherwise of JGPs in achieving their stated objectives on GEWE, and any intended or unintended long-term effects (particularly whether and how the joint programme has enhanced ownership, including within the UN system, and contributed to developing national capacity). Also includes evidence of innovation.

⁵³ See <http://www.oecd.org/dac/evaluationofdevelopmentprogrammes/daccriteriaforevaluatingdevelopmentassistance.html>. Also: UNEG Integrating Human Rights and Gender Equality Perspectives in Evaluation: Towards UNEG Guidance(2011). UNEG/G(2011)2

Participation and inclusion	The extent to which a development intervention is designed, implemented and monitored to promote the meaningful participation of a range of stakeholders (both rights holders and duty bearers) and to minimize negative effects of social exclusion. ⁵⁴
Efficiency	Whether the JGP has affected (in terms of reducing or transferring) transaction costs or burdens in terms of joint working; that is, whether working jointly has maximised the use of resources; allowed for cost-sharing; and reduced time and resource requirements; streamlined management and administrative burdens; and affected the pace of implementation for national partners and participating UN organizations and how.
Impact⁵⁵	The positive and negative, primary and secondary long-term effects produced by a JGP, directly or indirectly, intended or unintended.
Sustainability	The extent to which the results of JGPs given the level of national ownership has generated effective partnerships and national capacity strengthened through JGP processes

Additional criteria may be applied depending on the final selection of JGPs for field study e.g. those operating in fragile situations (OECD DAC principles for operating in fragile states).

3.6 Summary of feasibility and limitations

Feasibility

The Evaluation is ambitious in its breadth and scope, as reflected in the original list of evaluation questions. Revising the Evaluation Questions in the light of documentation availability and an assessment of their feasibility indicates, however, presents reasonable confidence in evaluability, as far as can be assessed at this stage. It is not, of course, possible to confirm full evaluability against any particular Evaluation Question, since desk review is only partially underway, and the status of evidence at field study level will only become apparent once the team are deployed to the field. As stated above, however, this should not deter the Evaluation from asking the questions which matter.

In summary: data availability appears sufficient at this stage to provide confidence that most of the Evaluation questions, tailored as above, can be robustly undertaken. A Summary Analysis of detailed desk review, to be provided at the end of the next stage of the Evaluation, will comment further on feasibility.

Limitations

Overall, the Evaluation has a number of limitations to its design, as follows:

- 1. The complex nature of the object*, namely JGPs which operate at policy level, involving multiple stakeholders, within complex and fluid environments, and which have changed over

⁵⁴ See UNEG (2011) Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance

⁵⁵ As described below, the Evaluation design is realistic on the likely limitations of assessing impact.

time. This limits the possibilities for applying standard evaluation methodologies which imply more linear causality.

Mitigation: the application of a theory-based model, as discussed, should enable complexity and context to be explored and understood, and causative pathways to be developed – leading ultimately to an overarching programme theory for JGPs.

2. Data paucity concerns- for some JGPs, is possible that inadequate information will be present to robustly assess results, particularly at higher levels of intent (though this is unlikely to be the case across the full sample base). This may limit the full application of contribution analysis at the upper levels of results.

Mitigation strategies include: the opting for a more qualitative appraisal applying contribution analysis at different levels, rather than an audit of inputs and outputs; the use of proxy indicators and alternative datasets where initial research indicates that data is simply unavailable; the adoption of a mixed-method approach (described below) to ensure triangulation through the use of multiple sources (qualitative and quantitative) and cross-checking these on an ongoing basis. At analysis stage, assumptions of pathways to results based on the theory of change implicit (or explicit) in programming will be assessed.

3. Impact evaluation in the strict OECD/DAC sense of the term cannot be conducted: firstly because time-lag issues mean that JGP activities may not yet have delivered demonstrable results at this level; secondly because the Evaluation questions focus on broader issues than impact including the design process itself such as unintended outcomes, ownership and sustainability; and thirdly since impact level results are the most challenging to obtain. Desk review should shed light on what is feasible.

Mitigation: the Evaluation is not an impact evaluation in the classic sense. As above, the difficulty of attribution and the multiplicity of actors within JGPs make a contribution analysis the most logical approach.

4. Timing: Practical issues of timing and resources are also significant influences on feasibility. The gap between the confirmation of field study sites and the anticipated field visits is short and may constrain both the stakeholders available for consultation and those JGPs which can readily accept a mission in the timeframe; these concerns have been discussed with the EMG.⁵⁶

Mitigation: An indicative set of sample JGPs is therefore presented within this Inception Report (see section Table 8 below) for consideration by the EMG. The Secretariat has already begun the process of consultation with the sample JGPs to assess their readiness to accept a mission during the timeframe.

5. Comparability – The diversity of contexts within which JGPs operate, and the diverse nature of JGPs themselves, mean that data generated may suffer from comparability constraints. Ensuring a common and robust core analytical approach is essential if the evidence base generated is to permit comparison at synthesis level.

⁵⁶ See Proposals for Field Study Sampling of JGPs, prepared for the EMG on 17th July 2012 (available from Evaluation Secretariat)

Mitigation: The Evaluation design places a strong focus on comparing across contexts, applying a common framework for analysis across diverse JGPs in very varied operating circumstances. Common enquiry tools will be applied around a central Evaluation Matrix (see section 4.4 below) all geared to the preliminary programme theory, above. The application of common core questions, methods for data collection, and analytical frameworks; as well as common formats for reporting, e.g. for field study reports, will support this. Comparison will take place across features, or attributes, rather than using a whole-case approach; this should maximise the options for analysis and support the construction of internal validity.

4. Methodology

4.1 Evaluation process and stages

The four phases of the Evaluation as outlined are: Detailed Desk Review (August 2012); Field Study of JGPs (September-November 2012); E-survey; and Synthesis (December –March 2013). These are explained in detail below.

Methodologically, each phase acts as a building block in generating the composite evidence base. Detailed desk review, for example, will be informed by the data gathered from initial document screening and interviews during the Inception phase. Field study of JGPs will build on the evidence created during detailed desk review of a wider sample of JGPs. The Synthesis stage will draw the composite body of evidence together for full analysis.

All four phases will operate within a common methodological framework, whose function is to operationalise the conceptual approach above. Each phase will however apply different *methods*, in line with the kinds of evidence needed to be generated. The evaluation criteria above will be applied across the phases. This section of the Inception Report describes how this will be achieved, taking into account the evaluability issues and concerns identified above.

4.2 Overall methodological and analytical framework

The methodological framework forms the ‘lens’ through which evidence generation and analysis will take place, and within which different methods will be applied. This section sets out how this has been constructed.

The Evaluation design is oriented around an extrapolation, testing, possible validation and refinement of the preliminary programme theory, above. As section 2.1 notes, however, this is preliminary in nature. A key element of the evaluation will be to determine the aims and intentions of JGPs at different levels; to reveal the interconnections between the strategies applied to realise these; and to explain and explore the range of assumptions inherent within design and implementation.

A summary of how this will be operationalised against the Strategic Priority Questions is set out in Table 4 below:

Table 4: Operationalisation against the Strategic Priority Questions

Extraction and analysis of contextual features and their influence on the design, implementation and results of JGPs, how and why.	Utilisation focused approach	<p><i>Field and desk study stage</i></p> <p><i>Q1 Design</i></p> <p>Exploring and testing the assumptions and theories implicit within a ‘joint’ process surrounding JGP design</p> <p>Identifying the features of the operating environment within which design took place, and their influence upon the design process</p>	<p><i>Synthesis stage</i></p> <p>Comparing across different design processes within the sample of JGPs to extract common attributes of successful and less successful design as they relate to the operating context</p>
		<p><i>Q2 Delivering Results and Sustainability</i></p> <p><i>[contribution analysis]</i></p> <p>Identifying the results of JGPs, included unintended (positive or negative) results at different levels including those related to the substantive issues of GEWE</p> <p>Identifying the pathways of contribution of JGPs to results, plus alternative explanations</p> <p>Identifying the explanatory pathways from contribution to results, including the influence of context</p>	<p>Comparing the achievement of results, pathways of contribution, explanatory pathways, as well as alternative explanations, across contexts to identify commonalities and differences.</p>
		<p><i>Q3: National Level Partnerships, National Ownership & People Centred Approaches</i></p> <p>Exploring and testing the assumptions and theories implicit within efforts to build national capacity and ownership, and to design and implement sustainability plans</p> <p>Identifying the features of, and exploring and testing the assumptions implicit within, accountability strategies, as well as the implementation of these</p> <p>Identifying the features of the operating environment within which partnerships, accountability and sustainability strategies (including efforts at capacity development) took place, and their influence upon this process</p>	<p>Comparing across features of different experiences, to extract common attributes of successful and less successful partnerships, ownership and sustainability strategies occurred, as they relate to the operating environment</p>
		<p><i>Q4: Synergies</i></p> <p>Exploring and testing the assumptions and theories implicit within efforts to create synergies, and where the balance of synergies lies (within the UN system, between the UN and government, within national systems)</p>	<p>Comparing the achievement of synergies, pathways of contribution, explanatory pathways, as well as alternative explanations, across contexts to identify commonalities</p>

	<p>Identifying the features of the operating environment within which synergies have been developed, and their influence upon this process</p> <p><i>[Applying contribution analysis]</i> Assessing the extent to which synergies have changed over the duration of the JGP including those related to improved gender capacity and awareness within the UN and national partners</p> <p>Identifying the pathways of contribution of JGPs to improved synergies, plus alternative explanations</p> <p>Identifying the explanatory pathways from the contribution of JGPs to improved synergies including the influence of context</p>	<p>and differences.</p>
--	---	-------------------------

4.3 Operationalising the methodological approach

The theoretical and conceptual basis for the Evaluation set the parameters within which it will be conducted, and define the framework of the methodological approach. The framework above operationalises this against the Strategic Priority Questions. The next step is the development of the framework into a feasible mechanism for analysis; the Evaluation Matrix.

The Matrix forms the main ‘spine’ of the Evaluation. Its function is to guide analysis and enable robust and evidence-based assessments to be made against the Evaluation Questions. As such, it includes relevant indicators (along a continuum of positive direction) around which evidence can be collated, enabling progress to be assessed. It also includes detail on anticipated methods and forms of evidence (as far as can be assessed at this stage), and the evaluation criteria to be applied. The Matrix can be seen at Annex 6.

The Matrix applies the preliminary programme theory in the following ways:

- The strategies / features of JGPs are all reflected in the Matrix ‘indicators of positive progress’ (joint analysis of needs, joint strategising, planning and prioritisation; coordinated resource mobilisation; joint management and implementation; fund management options; joint monitoring and evaluation)
- The process changes or results (shared vision and prioritisation among partners; reduced/avoidance of duplication; reduced or transferred burdens and transaction costs; improved synergies and shared expertise among partners; enhanced UN influence and outreach on GEWE) are all reflected within the Evaluation sub-questions / Strategic Priority Questions 1-4, around which the Matrix is constructed
- The interim changes and results (increased national support to the addressing of GEWE priorities; increased resources available to address national GEWE priorities; improved harmonisation and management for development results; better value for money and greater efficiency; improved coherence and effectiveness in supporting national priorities and needs under the UN system) are all reflected within Strategic Priority Questions 2,3 and 4 (Strategic Priority Question 1 being concerned with the precursor issue of design) within the Matrix.

The Matrix currently has broader scope than the preliminary programme theory, since the latter is derived mainly from the source UNDG Guidance, supplemented with data gathered and analysed during the Inception Phase. Applying the Matrix should therefore support the process of interrogating and elaborating the programme theory, by applying its wider range of indicators against the Evaluation Questions and Sub-Questions.

The Matrix arises from the conceptual approach above, but is an operational, not a conceptual, tool. It will be applied in the following ways:

- Within desk study – the analytical tool for desk study is drawn from the Matrix, including the Strategic Priority Questions, sub-questions and indicators
- Within the E-survey – questions and responses geared to different stakeholder groups are geared to the sub-questions and indicators of the Matrix
- Within field study – the Matrix will function as the analytical template for field study, with field teams plotting in data, supporting evidence and analysis as the basis for drafting reports. This will enable clear and transparent chains of evidence for report-writing.
- At synthesis level – the Matrix will form the main analytical framework for analysis of the composite evidence base.

Methods and forms of analysis applied for all these stages are described below.

The Matrix applies contribution analysis, as described above. Its design, with each Strategic Priority Question being extrapolated through the supporting Sub-Questions and explored through the indicators assigned, will allow findings to emerge robustly and analysis to follow a logical path. Particular emphasis has been placed on making the tool feasible to apply at field study level, given that over-ambitious evaluation matrices often result in thinner data coverage against evaluation questions.

Methods to be applied

The Evaluation applies a mixed-method approach, advocated by some current thinkers⁵⁷ as a compensatory principle and as a means of meeting criteria for causal inference. ‘Combining methods is a way to overcome limitations and enhance strengths’, recognising that ‘different techniques meet specific purpose, from measurement and description of events and states to understanding of a situation or a process, bringing their own strengths and limitations.’⁵⁸ It also facilitates responsiveness to gender and human rights considerations.

Methods will be combined in the Evaluation in several different ways:

- **Triangulation** – to confirm and corroborate results reached by different methods – e.g. confirming the articulation of a joint design process which is present in design documents.
- **Complementarity** - to explain and understand findings obtained by one method by applying a second. E.g. explaining and understanding the nuances around the results of JGPs stated in reports

⁵⁷ Stern et al (2012)

⁵⁸ Ibid.

- **Interrogation** - where diverging results emerge from the application of different methods – these will need to be interrogated to either reconcile, or explain, the differences apparent.

A summary of the intended methods to be applied, relevant to the Evaluation Questions outlined above, is presented below. For each stage in the Evaluation process, different methods and forms of analysis will apply, as well as different sampling techniques. Methods are further set out per individual sub-question, within the Evaluation Matrix. For field study of JGPs in particular, these remain statements of intent. As always with evaluations conducted at field level, the sort of data actually available, plus new or intensified forms of analysis needed, or key interlocutors to speak to, often only emerge clearly during actual field visit. Every effort will be made to implement the methodology as described below, but to remain open to other critical data sources / forms of analysis needed if these emerge.

The following table (Table 5, which is followed by a detailed breakdown of the purpose, methods to be applied and approaches to analysis and reporting within each phase) sets out how methods will be applied within the phases of the Evaluation. As described above, the different phases should be seen as building blocks in generating the composite body of evidence. Each phase, and the methods ascribed within it, informs and refines the next, so that each individual phase is not taking place in isolation, but building on - and validating - the evidence generated in the previous stage.

The Evaluation Matrix at Annex 6 assigns specific methods against the individual Evaluation Questions and Sub-questions. Further detail is therefore also available at Annex 6. Contribution analysis, as indicated above, is applied across all phases, being integrated within the Evaluation Matrix design.

Table 5: Methods to be applied per phase of the Evaluation

Evaluation Phase	Specific methods to be applied	Output to be generated
Detailed desk review of 24 JGPs	<p>Systematic analysis of JGP documentation, ⁵⁹according to a pre-devised analytical tool geared to the Evaluation Matrix</p> <p>Semi-structured interviews, applying a format also geared to the Evaluation Matrix</p>	<p>Summary output of detailed analysis, including interview data</p> <p>Proposals for field study sampling</p> <p>Further elaboration of initial programme theory</p>
E-survey (concurrent to desk review / field study)	<p>Systematic survey of three key stakeholder groups at national level:</p> <ul style="list-style-type: none"> • Key informant from UN intensively involved in the JGP at country level, e.g. Programme Coordinator • Key national stakeholder informant e.g. representative from Ministry of Women’s Affairs, Ministry of Planning, Civil Society • A stakeholder with significant insight/oversight and engagement with the JGP e.g. a representative of a funding/donor agency or an evaluator <p>Application of a systematic survey tool, diversified for different stakeholder groups, which uses the Likert scale, to collect perception and qualitative data on the core indicators against the Evaluation Questions</p>	<p>No distinct product: summary analysis will be generated and fed into synthesis analysis. Subject to timing, may also inform field study.</p>

⁵⁹ To include a range of: Documentary analysis of UNDAFs, CCAs, UNRC Annual Reports, UNCT reports, MDG reports, DAO reports and evaluations, Resident Co-ordinator reports and other relevant data; programme documents including Prodoc; Memo of Understanding; annual progress reports; mid-term and end of programme evaluations and any studies or reviews generated by the JGP or by donors which include the JGP

	and Sub-Questions (as reflected in the Evaluation Matrix)	
Field study of JGPs	<p><i>Note: methods will be selected dependent on locations selected, and outcomes of desk study, but are likely to include:</i></p> <ul style="list-style-type: none"> • Stakeholder and perceptual analysis; interviews using a semi-structured interview format and standard stakeholder analysis tool;⁶⁰ • Financial and budgetary analysis, to the extent supported by available documentation, of JGP performance against targets • Analysis of national datasets e.g. poverty profile, gender equality profile, (e.g. any data on GBV prevalence, political representation, women’s share of national income etc); other datasets relevant to the programme e.g. education, health, HIV and AIDS • Systematic documentary review of data unavailable centrally⁶¹ applying the common analytical tool geared to the Matrix, above • Interviews, using semi- structured interview guides (to be developed once detailed desk review has been conducted) • Participatory tools may be used where appropriate for context: these may include: the use of focus groups; process tracing; and others • Comparison with the body of evidence from available comparator JGPs • Analysis / elaboration of specific theory of change 	Field study reports, written to a standardised structure (geared to the Matrix) and length, including a specific theory of change for the programme

⁶⁰ See Annex 10 for draft tools

⁶¹ To include: internal JGP and comparator documentation; National partner reports and analyses, including meeting minutes, progress reports; internal UN and donor agency reports; Resident Co-ordinator reports and other relevant data

Synthesis	<p>Analytical methods to include:</p> <ul style="list-style-type: none"> • Application of a standardised analytical filter across different data categories, geared to the Evaluation Matrix • Identification of common trends, contradictions and differences • Identification of common assumptions, and how these can / have been identified and managed • Identification of explanatory factors (including alternative explanations) related to the operating environment, the internal design, implementation and synergies • Identification of different pathways of contribution and causality at different levels of results • Analysis of evidence arising from comparator joint programmes, applying the same parameters as for JGP analysis 	<p>Emerging findings note, integrating the findings from the</p> <p>Synthesis report, to include an elaborated and validated programme theory.</p>
-----------	--	--

4.4 Detailed breakdown of Evaluation Phases

1. Detailed desk study

Purpose: A major component of the Evaluation design is to conduct a detailed systematic analysis of a sample base of 24 JGPs. The purpose of this phase is to:

- Identify key contextual parameters within which JGPs are operating and which may have influenced the achievement of results
- Allow for initial findings to be developed against the Evaluation Questions (thereby testing the theories and assumptions explained above), which can then be further explored / triangulated / validated during fieldwork
- Generate a composite evidence base against the Evaluation Questions, which can then be systematically analysed across all 24 JGPs
- Interrogate and refine the preliminary programme theory, above

Sampling: The Evaluation Team were directed⁶² to develop a ‘fully representative’ sample of 24 JGPs, derived from the ‘universe’ of 80 JGPs identified by the Portfolio Review as eligible for study.

Sampling criteria and content were developed accordingly, though the evaluation team felt that geography (region) and thematic area were the only feasible areas in which the sample could be made truly ‘representative’. Other parameters (country income/Human Development Index/Gender Inequality Index status/programme budget/lead agency/range of partners/fragility/Delivering As One status etc), agreed with the EMG during the Inception Mission, were included on the principle of maximising spread, on the basis that exploring a wider distribution of programmes across varied contexts will give greater depth and breadth of data. Full details of the process of applying these can be found in Annex 7.

The EMG were presented with, and agreed, the sampling criteria for selection during the Inception Mission. The sample base has been constructed accordingly, with just three JGP substitutions to the sample made to date.⁶³ The full sample base can be found at Annex 8.

Methods: detailed desk analysis of JGPs is currently ongoing. The document set includes some or all of: a concept note, memorandum of understanding, Programme document (Prodoc), work plan, budget, annual progress reports, medium-term reviews and evaluations plus Common Country Assessment, Country Programme Action Plan, UNDAF, donor reviews, and external evaluations plus relevant wider evaluations that comment on JGPs e.g. DAO evaluations.

The key to systematic data analysis of documents is the application of a core analytical tool which applies analytical categories geared to the Evaluation Matrix. This has been pre-developed, applying the indicators of the Evaluation Matrix, and has been separately submitted. As analysis progresses, it may be appropriate or useful to add or amend categories; revision will take place on an iterative basis and following full agreement by all members of the Evaluation Team.

⁶² by EMG members

⁶³ Those screened out for various reasons were: Equatorial Guinea and Ecuador (programmes non-operational / not official); Sierra Leone and Moldova (lack of evaluative information within the required timeframe). These were substituted by JGPs from Albania; Colombia; Eritrea and Namibia respectively..

A cross cutting issue for analysis of JGP documents is evidence of innovation (in design, programme focus, ownership, results monitoring and reporting etc). This will be identified by the team in documentary analysis, using relatively open parameters, and confirmed through interview with key stakeholders (below). The integration of a human rights based approach to JGPs has been included within the analytical tool through specific indicators.

Desk review will be supplemented by semi-structured interviews with (where feasible) key stakeholders of JGPs, classified as:

- Key informant from UN intensively involved in the JGP at country level, ideally in design and implementation e.g. Programme Coordinator or someone from the Lead Agency
- Key national stakeholder informant with an interest in and good knowledge of the JGP e.g. representative from Ministry of Women’s Affairs, Ministry of Planning or a Women’s Organisation⁶⁴
- A stakeholder with significant insight/oversight and engagement with the JGP e.g. a representative of a funding/donor agency or an evaluator.

Ideally three, but a minimum of two interviews per JGP analysed will be conducted, applying a semi-structured format which has been seen and commented on by the EMG (see Annex 10).

Analysis and Reporting: analysis will take place across the composite body of evidence generated by combining individual analyses (on Excel spreadsheets, separately attached), within a composite database. Attributes of context and JGPs will be compared, to identify and interpret concentrations of evidence within the analytical categories; and differences and divergence interrogated. Output from this stage will be in the form of a Summary of Desk Analysis, which presents the preliminary findings from desk study (available mid-September 2012, proposed structure in Annex 11). This will include any extension or refinement to the initial programme theory, based on its interrogation and refinement through the desk review process.

2. E-Survey

An electronic survey will be conducted as part of efforts to maximise participation and inclusion, to reach out to stakeholders whose opinions and views cannot be accessed through interview. The survey will also support the triangulation and validation of data.

During the Inception Mission, discussion took place with the Secretariat on how to maximise the utility and outreach of the proposed survey. In particular, the importance of outreach beyond HQ was emphasised.

Accordingly, an E-Questionnaire has been designed, using [FluidSurveys](#). This includes a targeted questionnaire (drawn from the Evaluation Matrix) for different groups of respondents. For the Evaluation: these will initially comprise programme and partner agency staff; representatives of national governments; and representatives of civil society or the women’s movement, though other groups may well be identified by the EMG.

Subsequently, invitations and guidance are sent to proposed respondents, using the Likert scale to respond to questions. Respondents have a 2-3 week period in which to complete the questionnaire;

⁶⁴ The need to include balanced representation of rights-holders within the sample of interviews has been emphasised by UN Women; the Evaluation Team will make efforts to this effect, pending these interviews being facilitated by country programme staff

data is then collated and analysed centrally. Lead agencies in each country will be asked to identify key stakeholders at whom the survey should be targeted including government, civil society / the women’s movement, partner agencies in country, relevant regional bodies etc.

The advantages of this method are that data is clearly labelled and can be quickly synthesised using in-built reporting tools. All information provided is confidential. It also allows for standardisation of data to allow for analysis across countries and regions. The survey includes as much space for free-flowing information as possible, in order that rich qualitative data can be obtained

The survey tool has been prepared and can be implemented at any point. ⁶⁵We would propose that it is issued either prior to the field study period if possible, to allow for the analysed data to inform the country studies themselves.

Once analysis is complete, and if appropriate, a set of ‘further enquiry’ questions could be developed from the data and an e-discussion held at a later date – or indeed, the EMG may wish to consider this as part of the dissemination strategy for the Evaluation.

Analysis and reporting: analysis will take place using a standard computerised data synthesis tool. Reports can be generated across different categories and for different groups of stakeholders. The findings of the survey will not be a separate output of the Evaluation, but will be integrated into the composite evidence base for Synthesis stage, below.

3. Field study

Purpose: Field study of JGPs is a major element of the Evaluation. It has been agreed with the EMG that up to five field studies will be conducted. Aims are as follows:

- To complement and deepen the evidence base generated by detailed Desk Review, above, of 24 JGPs, while retaining the systematic approach
- to validate and enrich the desk review and to generate new information that will confirm or refute the conclusions of the desk review
- To deepen enquiry in areas where desk review alone, even where supplemented by interviews, is insufficient. This applies to all the Evaluation Questions; e.g. a detailed assessment of design cannot take place through desk analysis alone
- To enable a range of contexts and JGPs to be studied in-depth, in order that causative relations and pathways of contribution can be assessed in detail and in a grounded understanding of the operating conditions in the country
- To further elaborate the initial programme theory

Sampling The sampling for field study of JGPs has been debated between the Evaluation Team, the Secretariat and the EMG during Inception Phase. The original directive⁶⁶ to the Evaluation Team was that sampling criteria, and the consequent group of countries, should arise from the findings of desk study; that is, be fully evidence-based, and representative. The Evaluation Team recognize the rigour of this approach, but have highlighted the need to balance it a) with feasibility

⁶⁵ A draft version has been reviewed by UN Women.

⁶⁶ This point was emphasised by the UN Women in particular

issues, including the requirement for sufficient lead-time between finalizing criteria, selecting countries, and organizing visits and b) the methodological emphasis on illustration, below.

A Proposal for Field Study Sampling was developed and sent to the Secretariat on 17th July 2012.⁶⁷ The Evaluation Team were requested by the EMG to resubmit their proposals for sampling at field study within the revised Inception Report, herewith.

We propose that the selection of countries for field study of JGPs is:

1. Drawn from the sample base of 24 JGPs, since this is a) broadly representative of the universe of the Portfolio Analysis and b) covers the range of regions, operating environments, size, scale and thematic area of JGPs and c) will enable field study to be built on an already solid base of evidence generated from desk study, enabling enquiry to be more readily deepened and broadened
2. Conceived as illustrative rather than generalisable; focused on exploring the theories and assumptions surrounding JGP design and implementation, and causative connections to results
3. Focused, as for the sampling criteria for desk review, on the principle of maximising spread within the framework for comparability, on the basis that exploring a wider distribution of programmes across varied contexts will give greater depth and breadth of data

The following set of criteria can be applied, according to the reasoning below:

Table 6: Criteria for field study sampling

Criterion	Rationale
Regional diversity plus contextual features: maturity of aid architecture ⁶⁸ ; DAO context; fragile situation	Exploring a range of operating environments will enable the assumptions underlying the role of the context in influencing the design, implementation and results of JGPs to be assessed. <i>A decision will be needed on whether 1 JGP per region is required, or since the bulk of JGPs (55% of the ‘universe’) are in Africa, whether concentrating study in Africa (e.g. 2 JGPs in the region) is appropriate</i>
Income status and Gender Inequality Index rating	The extent of aid dependency and poverty incidence among women, plus the gender inequality status of the country, are likely to be key determinants in affecting design, implementation and results of JGPs.
Thematic area	Assessing a range of JGPs across diverse thematic areas will enable the ‘conduciveness’ of different thematic areas to joint programming for GEWE to be assessed, as well as whether the interconnections and assumptions within different design and implementation processes are common or different for particular thematic areas
Scale / budget	Combined with other features, this will allow assessment of whether and how budget and scale affects the design and implementation of JGPs and the different challenges and opportunities arising
Number of partners	As for budget and scale; combined with other features, this will allow the challenges and opportunities of designing and implementing JGPs with both large and small numbers of partners to be explored

⁶⁷ This set out some proposed categories for stratification by context such as middle and low income status, DAO and self-starter, fragile and non-fragile, Gender Inequality Index ratings etc; to confirm whether the sample would include a range of thematic areas of JGPs and cover all five geographical regions and a spread of budget size / range of lead agencies.

⁶⁸ As defined by e.g. Paris Declaration Phase II country evaluations and other reports

Recent / current evaluation	There would be little value to be gained from reviewing programmes which have been recently evaluated, particularly given the burdens of missions on country offices. Programmes with recent evaluations will have maximum value retained as part of the sample for desk study. Some MDG-F programmes are also being evaluated during the period for field study.
Comparator JGPs and / or more than one JGP in the country exists	Reasons of comparability, described above – this is currently being explored through desk analysis
Whether the JGP is funded through the MDG-F	These programmes are generally well-documented which may possibly imply strong management and a potential link to improved results. It will be important to avoid bias by ensuring proportionate representation.
Strengths and weaknesses identified	Studying a range of JGPs which are identified through desk study as having different strengths and weaknesses will enable better exploration of causative connections and explanatory factors

Applying these parameters (bar the last one which will arise from detailed desk review) in composite, and aiming for a cross-section of contexts which satisfy the range of criteria, gives rise to, for example, the following possible sample set (see Annex 8 for JGP details):

Table 7: Possible set of JGPs satisfying sampling criteria and alternatives

POSSIBLE JGP SELECTION								
Location	Context features	Income status	GII	Thematic area	Value (\$)	Lead agency	Partners	Other
OPT	Fragile	Low middle income	Unknown	Governance	\$9m	UNDP	6 partner agencies	MDG-F More than one JGP
Uganda	Mature aid environment	Low income	Below 0.5	Integrated	\$24m	UNIFEM	11 partner agencies	Bilateral donor; possible single agency comparator
Nepal	Fragility aspects	Low income	Above 0.5	Health	Unknown	UNFPA	3 partner agencies	More than 1 JGP
Mexico		Upper middle income	Below 0.5	EVAW	\$994,000	ECLAC	6 partner agencies	
Albania (OR an Africa JGP)	DAO	Upper middle income	Below 0.5	Governance	\$1m	UNIFEM	4 partner agencies	

Possible alternates								
Kenya	Mature aid environment	Low-income	Above 0.5	Integrated	\$28m	UNIFEM	13 partner agencies	
Iraq	Fragile	Low-middle income	Above 0.5	Education	\$12m	UNICEF – led	7 partner agencies	More than 1 JGP
Thailand		Upper middle income	Below 0.5	EVAW	\$1m	UNIFEM lead	3 partner agencies	
Liberia	Self-starter / post-conflict	Low income	Above 0.5	EVAW	\$2.8	UNFPA	6 partner agencies	More than 1 JGP
Lesotho		Low income	Above 0.5	Health	\$663,000	unspecified	3 partner agencies	More than 1 JGP
Uruguay	DAO	Upper middle income	Below 0.5	Governance	\$1.3m	UNFPA and UNIFEM	9 partner agencies	

Detailed desk study, plus checking with country programmes on feasibility, will enable the inclusion of these JGPs to be assessed, and/or propose any suitable revisions to the sample.

Methods: The Evaluation, as stated is a collective case study involving multiple cases. To ensure that findings are comparable for synthesis purposes, whilst remaining illustrative, individual studies will apply the Evaluation Matrix described above. Data and analysis will be plotted in to the matrix, to provide clear and transparent chains of evidence for report writing, and as for desk review, to ensure a fully systematic approach. The identification of innovative practice will form part of the lines of enquiry.

The particular set of methods for field study will be selected by Evaluation Team members once the locations of field studies are known, and the evidence base available clear. However, methods to be applied will combine qualitative and quantitative approaches, and include those generally widely justified for in-depth fieldwork. These include: stakeholder and perceptual analysis; interviews using a semi-structured interview format;⁶⁹ financial and budgetary analysis; analysis of national datasets; and documentary review of data unavailable centrally applying systematic tools. Interview checklists will be applied to guide semi-structured interviews; these will be developed once detailed desk review has been conducted, in order that they are based on the findings arising from this stage. Participatory tools will be used where the field teams consider that their use will enhance the quality and accessibility of information – these may include for example the use of focus groups; process tracing; and others. The integration of a human rights based approach to JGPs will be assessed through the use of the specific indicators included within the Evaluation Matrix.

Theory of change models will be developed to map the logic of the selected JGP case studies. This will be done by the Evaluation Team, working with and facilitating programme staff and key partners to retrospectively model how the JGP has contributed to expected GEWE outcomes; what

⁶⁹ See Annex 10 for draft tools

interconnections arise between JGPs and the different levels of results observed; and what assumptions are evident, as well as whether and how these have been managed.

To support triangulation / complementarity / interrogation; findings from desk review will be plotted onto the relevant Evaluation Matrix template in advance of field study, and areas where enquiry needs to be deepened / validated and tested / interrogated identified. All pieces of data arising from desk review will be triangulated during field study, to ensure that internal validity is maximised, for example by applying any independent data from civil society which reflects on the JGP performance, the partnerships and synergies it has supported or otherwise, etc.

In the discussions surrounding Team appointment, the Secretariat proposed that the three-person teams proposed for field study include a member of an Evaluation Unit of Office of a partner agency to the Evaluation. The Evaluation Team agreed to this, providing that issues of independence were addressed (e.g. that the member of the UN agency participating in the team was not also a member of the lead agency for that JGP). It was further agreed that collaboration and support would be provided from one of the UN offices involved in programming, for example to help set up interviews, with the possibility of the Resident Co-ordinator Office support also. Local consultants, of which IOD PARC has an extensive global network, will also form team members, partly for the valuable local knowledge and experience which they apply, and partly as part of IOD PARC's professional commitment to capacity building in development evaluation.

Analysis and Reporting: Analysis for field study reports will apply the Evaluation Matrix as the main analytical tool across data streams, grouping evidence around the indicators within it, including those on human rights and gender equality, and proving summary evidenced progress assessments. Reporting will take place to an agreed structure and length, to ensure comparability of findings and maximum contribution to the final report. Reports will be written in clear and concise language, without the use of jargon or acronyms. Content will focus on analysis and progress assessments, rather than description. The report structure will be that reflected in the Evaluation Matrix (i.e. oriented around the Evaluation Strategic Priority Questions). The proposed format can be found at Annex 11.

Synthesis stage

Purpose: The purpose of the Synthesis stage is essentially analytical. It will bring together the composite body of evidence arising from the different streams of data: interviews, initial and detailed desk review; and field study in a coherent form, to present findings, lessons learned and conclusions that are logically interconnected, fully respond to the evaluation questions pending the availability of evidence, and with a clear chain of evidence behind them.

The synthesis report needs to be grounded in the desk review, country studies and other evaluation components, so that it has full substantive validity, but also provides coherent aggregate-level findings.

Analytical Methods: The core requirement of a robust synthesis process is the application of a standardised analytical tool across different data categories, as well as an assessment of data quality. As noted, all components and tools of the evaluation design are geared towards the Evaluation Matrix, which itself forms the basis for the synthesis report. This allows data to be systematically comparable. During the field study phase, the Evaluation Team will concurrently develop a core analytical tool, geared to the Evaluation Matrix, to allow for the robust aggregation

of data at synthesis level. This will then filter the composite body of evidence (findings arising from detailed Desk Review plus field study of JGPs), in the following ways:

- Across the analytical fields developed, which are drawn from the preliminary programme theory and applied across the Evaluation Questions and all enquiry tools, common trends, contradictions and difference will be sought out and explored.
- The different pathways of contribution and causality at different levels of results will be tracked and identified
- Explanatory factors related to the operating environment, the internal design, implementation and synergies of the programme will be assessed
- Common assumptions, and how these can / have been identified and managed, will be drawn out
- Findings at the different levels of the preliminary programme theory will be identified and reported
- Gaps in information available will also be reported

Programme theory: The analytical tools for the Evaluation, as stated, are geared towards the extraction and refinement of the preliminary programme theory. At synthesis stage, the composite body of evidence, providing evaluability requirements are satisfied, should provide information on the results generated at different levels, the interconnections and assumptions between them, and should also allow for an overall testing and validation of the logic inherent within it.

Drawing on the series of programme-specific theories developed, a composite programme theory can be developed by applying analysis to extract commonalities and interrogate differences. In keeping with the utilisation and participatory approach outlined, this should incorporate the views of the Evaluation Reference Group, Management Group and key agency staff in particular.⁷⁰ It is possible – but not definite – that suggestions may be made towards a revised definition of joint gender programmes.

Reporting: We also propose early agreement of an agreed content outline and length at an early stage, to serve as a continuing reference point. The proposed format is at Annex 11. It is a core principle of the utilisation approach that evaluation reports be accessible and readable; we therefore propose to write the report without jargon or use of acronyms other than essential ones (UN, JGP, GEWE etc).

⁷⁰ This iterative approach has been used in other evaluations, where there are a variety of interventions and no explicit theory of change e.g. Evaluation of AusAid’s Engagement with Civil Society: Evaluation Plan and Annex on Theory of Change

5 Final outputs and quality assurance

5.1 Outputs of the Evaluation

Key outputs and timing, presuming that the Workplan at Annex 12 can be implemented as planned, will include:

- Summary of desk analysis and sampling criteria for field study (September 14th 2012)
- Selected Joint Gender Programme case study reports (15th December 2012)
- Preliminary Findings paper (15th January 2013)
- Draft Synthesis Evaluation Report (February 15th 2013)
- Final Evaluation Report, evaluation and methodology briefs and PowerPoint presentation (March 2013)

5.2 Quality assurance

IOD PARC is committed to delivering credible and high quality evaluations based on a clear set of organisational processes. These include a Quality Assurance process that is transparent to all stakeholders. Quality Assurance in this context refers to both the reliability, traceability and efficacy of evaluation management processes and also the professional and intellectual rigour and standard of the resulting outputs from those processes.

IOD PARC has built its approach to quality control systems on its foundation as a centre of excellence in the provision of both innovative and pragmatic M&E assistance to organisations such as UN agencies and bilateral donors. A named Director of IOD PARC will also take responsibility for ensuring the quality of outputs. Key elements of the system will include:

- A **process monitoring documentation protocol** (QM protocol) to outline the key requirements from the tools that will be used, how data will be validated and presented.
- The use of the **ethical framework** described above, in accordance with UNEG standards, to ensure that ethical standards are fully maintained throughout the evaluation process
- The **systematic time management** of staff and consultants through the use of time sheets to enable clear attribution of time to deliverables.
- The use of quality assured Associates, whose work we know and who are familiar with, and able to deliver on, the high standards IOD PARC expects
- Use of **‘after action review’** learning processes at key evaluation milestones. These semi-formal, internal review processes involves a short (less than 1-hour) semi-structured interview with the consultant / team that delivered the project, which in turn results in a 1-page report. This report captures the key learning points from the project, including any actions for improving our service to clients, our own internal systems, and the methodologies and processes that we apply.

- The use of **formal quality assurance reviews** between a named Director of IOD PARC and the Evaluation Team on an appropriate basis, with action plans agreed to ensure effective process documentation and the implementation and monitoring of any required changes.
- The use of ongoing **client assessment of our performance**, which will take place via regular review meetings with the Secretariat, focussing specifically on performance and processes. These meetings are used to explore performance in detail.