



联合国
粮食及
农业组织

Food and Agriculture
Organization of the
United Nations

Organisation des Nations
Unies pour l'alimentation
et l'agriculture

Продовольственная и
сельскохозяйственная организация
Объединенных Наций

Organización de las
Naciones Unidas para la
Alimentación y la Agricultura

منظمة
الأمم المتحدة
للإغذية والزراعة

C

理事会

第一五五届会议

2016 年 12 月 5 - 9 日，罗马

对联合国系统评价职能的分析（JIU/REP/2014/6）

1. 联检组报告附有总干事的简明意见及联合国系统行政首长协调理事会（首协会）共同提出的更为宽泛的意见（A/70/686/Add.1）。

粮农组织总干事提出的意见

2. 粮农组织欢迎联检组报告，赞同其关于评价职能分析的总体方法及首协会意见。

3. 该报告发现粮农组织的评价职能是联合国系统中最先进评价职能之一，建议增强与其他监督职能合署办公的中央评价职能（建议 1）、采用涉及问责和学习目的的双重评价方法（建议 2）、制定列入预算的评价计划（建议 3）、确定评价事务负责人的任期限制（建议 4）、建立质量保证制度（建议 5）。

4. 除不适用于粮农组织的第 1 条建议外，上述建议要么已为粮农组织《评价办公室章程》所涵盖，要么评价办公室目前正在采用，尤其是 2014 年评价办公室为加强评价职能和质量而采取改革措施之后。该报告还提出了另外 4 条建议，为落实这些建议而有必要提供额外资源。

5. 粮农组织注意到，不久前结束的“对粮农组织评价职能的独立评价”¹对联检组报告所述事项做了更加深入的阐述。此次独立评价报告已提交计划委员会第一二〇届会议讨论。

¹ PC 120/5

本文件可通过此页快速响应二维码读取；粮农组织采用此二维码旨在尽量减轻环境影响并倡导以更为环保的方式开展交流。

其他文件可访问：www.fao.org。



mr629



第七十届会议

议程项目 140

联合检查组

对联合国系统评价职能的分析

秘书长的说明

增编

秘书长谨向大会成员转递他本人和联合国系统行政首长协调理事会对联合检查组题为“对联合国系统评价职能的分析”报告(见 A/70/686)的评论意见。



摘要

联合检查组在题为“对联合国系统评价职能的分析”报告(见 A/70/686)中,审查了联合国系统评价职能的演变、发展和改进,以评估该职能支助联合国系统各组织的进步、发展水平和能力,以及可能存在的有效评价职能的任何替代办法。

本说明反映了联合国系统各组织对该报告所提建议的看法。这些意见是在联合国系统行政首长协调理事会成员组织提供的意见基础上进行的汇总,各成员组织对该报告表示欢迎并支持其中的一些结论。

一. 引言

1. 联合检查组(联检组)在题为“对联合国系统评价职能的分析”报告(见A/70/686)中,审查了联合国系统评价职能的演变、发展和改进,以评估该职能支助联合国系统各组织的进步、发展水平和能力,以及可能存在的有效评价职能的任何替代办法。在报告中,联检组分析了联合国系统28个组织的评价职能,并通过其向行政首长、立法机构和作为联合国系统行政首长协调理事会主席的秘书长提出的9项建议,力求改善评价职能的效力并加强协调和协作。

二. 一般性评论

2. 认识到当前的国际形势要求对发展和人道主义援助采取新办法,并提高效力,联合国系统各组织欢迎这份报告,指出该报告既全面又深入。

3. 各组织承认联检组在其分析中所强调的评价职能的重要性,并支持报告中的许多建议。它们指出,这些建议旨在巩固和加强这一重要职能。此外,这份报告对于整个联合国所需的战略方向非常重要,有助于在联合国建立评价和循证的文化,加强联合国贡献的影响力和可持续性。

4. 总的来说,各组织认为这些建议明确而且注重成果,并赞赏联检组努力为联合国系统评价职能制定循证和战略上协调一致的道路,以满足日益增加的要求以及问责和学习的机会。虽然一些建议难以实现,各组织认为,在高级别的领导和支助下,它们能够实现。一些组织已经看到执行联检组建议的效果。

5. 但是,各组织还指出,一些建议需要相关组织和全系统具体的供资,并进一步指出,在财政紧缩的情况下,这些资源可能很难找到,会减少执行各项建议的效力。此外,各组织指出,一些涉及分散评价工作的建议可能会对具有复杂任务规定的大型外地组织带来挑战。

6. 此外,各组织指出,一些涉及评价机构结构的建议可能对联合国系统较小实体构成挑战。这些组织指出,这些较小实体已建立了能满足自身需要的评价职能,但可能不具备联合国系统较大实体更完善的评价职能。将大小组织的评价职能相提并论,可能会造成所介绍的情况并非完全准确。此外,对联合国系统规范性工作的评价不同对其发展工作的评价,而且,审查工作原本可以受益于对这一区别的更多考虑,以及这种区别对具有规范和技术领域任务机构的评价职能的结构、作用和责任的影响。

7. 最后,一些组织表示,尽管它们可以看出联合国系统各实体复杂用词之间的联系,仍然难以理解许多描述各种评价模式的用词之间的差异,如“中央评价职能”、“分散评价职能”、“中央评价单位”、“评价职能”、“内置的评价职能”,以及“联合国秘书处的中央评价职能”等。

三. 对各项建议提出的具体评论

建议 1

有合署办公和与其他监督职能整合在一起或与行政管理职能整合在一起的中央评价职能的联合国系统各组织行政首长应加强该职能并确保其质量、完整性、可见度和增加值。

8. 尽管有些组织希望对“中央评价单位”有一个更明确的定义，各组织支持建议 1。

建议 2

联合国系统各组织行政首长应采用平衡方法来解决对问责制的评价目的问题，并促进发展一个具有激励革新、敢担当风险和采用多学科观点的适当机制的学习型组织。

9. 联合国系统各组织支持建议 2，并一致认为，问责制不应是进行评价的唯一动力，评价结果的经验教训也是促进改善的有用工具。但是，它们也指出，若要实现上述平衡，则需要为建立学习文化做出一些努力，并指出，该建议可通过纳入行政首长将其实体变成学习型组织所需采取的行动或步骤而得到加强。

建议 3

立法机构应要求联合国系统各组织行政首长根据为维持一个能为本组织增加价值的有效和可持续评价职能所需的成本为各自的评价职能制定综合预算框架和资源分配计划。应在现行预算和报告机制和流程范围内向立法机关提交这些计划供审议。

10. 各组织指出建议 3 是针对立法机构，同意需要为评价职能制定综合预算框架，并强调这些框架有助于为评价活动的可持续筹资确定供资目标和计划。各组织指出，如果该建议也适用于分散评价职能，其效益将产生更大的影响。

建议 4

立法机构应指示联合国系统各组织行政首长酌情审查和修订用于任命评价办公室负责人的现行政策，以加强独立性、廉政、道德操守、公信力和包容性，同时要适当考虑以下标准：

- 应为单一、不可延长的任期设立一个 5 年至 7 年的期限，而且任职者没有再次被该组织雇用的可能性；

- 评价负责人应具有评价方面的资格和实质性经验，并有在战略规划、基础和业务研究及知识管理等相关领域方面的经验作为辅助，以及应有出色的管理和领导能力。

11. 各组织指出，建议 4 是针对立法机构，但有一项谅解，即该建议应提到中央评价职能的负责人。因此，支持这一建议，不过，有一些组织对任期限制问题表示保留意见。各组织指出，报告提供的证据表明，任期限制方面的做法和重新进入组织的可能性差异很大。似乎没有任何分析具体地将评价职能在任期限制和轮换政策方面的独立性(感觉上或实际的独立性)与公信力联系起来。虽然对重新进入组织的限制可能有赖于避免利益冲突的理由，但似乎没有任何明确的证据可以证明关于任期限制政策的建议比“没有任期限制”的政策更有优势。如果与建议 2 一道考虑，若要在问责制和学习之间采取一种平衡办法，可以说具有较长工作经验的评价职能独立负责人，作为方案管理人员没有机会重新进入组织，可以通过留在现职而更好地满足该组织在评价问责和学习双重目的方面的平衡。此外，联合国系统较小组建立一个专门的评价办公室可能没有成本效益，因此一个专门技术组织的任期限制会导致经验和技术的丧失。

建议 5

联合国秘书长应以联合国系统行政首长协调理事会(首协会)主席的身份要求联合国评价小组进行协作，以拟定一个可用于整个联合国系统评价职能的健全和统一的质量保证制度。

12. 各组织指出，联合国评价小组一直在致力于制定整个联合国系统各组织改进评价职能质量保证程序的方法，因此，欢迎并支持建议 5，特别是如果该请求所涉工作在联合国评价小组今后的会议上作为其年度工作方案提出。然而，各组织也指出，除非质量保证制度能够在联合国评价小组控制和影响之外独立运作，其价值和公信力可能是有限的。它们还指出，需要有资源来支持制定质量保证制度的努力，并指出目前的现状是，大型组织已能够采取质量保证程序，而较小实体发现很难这样做，特别是因为所涉费用。

建议 6

联合国系统各组织行政首长应根据对评价职能的明确界定的愿景、战略和成果框架把利用评价作为一个组织优先事项，并就利用评价的程度、性质和影响力情况向其立法机关提交报告。

13. 各组织支持建议 6，它们指出，该建议促进学习和透明度，以及将评价纳入整个工作的努力，包括通过制定其分散评价职能。然而，一些组织指出，该建议将重点放在进程，这可能因组织的要求和规模而各不相同。

建议 7

联合国系统各组织行政首长应要求评价办公室重新评估涉及评价职能的政策、战略和优先事项。他们应把评价职能置于各自组织的战略位置，从而加强其相关性，以增强联合国系统在应对当前变化和挑战及在实现影响力和可持续性方面的能力。

14. 各组织支持建议 7，确认必须根据全球变化和可持续发展的要求对评价职能进行战略定位。

建议 8

秘书长应以首协会主席的身份采取步骤并支持创新，以促进联合国系统各组织之间及与其他伙伴之间的协作，加强国家在涉及问责制、学习及具有国家和全球价值的知识发展方面的评价能力。

15. 各组织支持建议 8。

建议 9

联合国系统各组织行政首长应制定体制框架和必要的支持系统，以提高分散评价的质量和增加值，提高其在支持联合国系统应对新出现的挑战(包括 2015 年后发展议程的那些挑战)方面所能发挥的作用，以及加强联合国系统各组织内部和全系统范围及与国家机构的一致性和协调。

16. 许多组织表示支持建议 9，但有几个组织指出，应由一个中央评价职能负责执行该建议，因为相关办公室提供方法上的支助，并确保分散评价机构的评价能力得到充分发展。各组织还指出，权力下放有可能是一项复杂的工作，需要在改善外地办事处的评价能力和工作人员方面进行大量投资，并将需要时间、领导能力、投资和持续的承诺。

对联合国系统内的评价职能的分析

编写者

苏凯·普罗姆-杰克逊
乔治·巴尔齐奥塔斯

联合检查组



联合国，2014年日内瓦

内容提要

对联合国系统内的评价职能的分析

JIU/REP/2014/6

在目前资源受制约及各国和全球在发展和人道主义援助方面面临需要采用新对策的各项挑战的情况下，联合国系统各组织越来越需要说明资源使用情况并展示成果和其工作的增加值。这些组织的业务环境要求它们就各自政策、方案和活动的相关性、增加值、实效、效率、影响力和可持续性提供公正、有效和可信的证据。在这种环境下，评价职能为各组织提供了一个平台，以应对整个系统内的这类要求。

评价是用于在以下方面支持联合国系统的主要工具之一：对成果和增加值及对学习和知识发展进行问责、加强其在全球治理方面的领导作用，以及实行能对世界各地人民的生活产生影响的改革。因此，持续不断地发展评价职能对于保持联合国系统在以下方面的能力至关重要：实现其目标、说明成功的原因，以及实现必要的改革以改善国际发展和治理状况。

本报告载有联合检查组(联检组)在 2013 年对联合国系统评价职能的演变、发展和改善情况进行的全系统范围的研究成果。这项研究讨论了下列问题：联合国系统的评价职能在成长和发展方面是否有进展？评价职能目前处于何种发展水平？其用于支持联合国系统各组织的能力如何？用于应对各种要求、变革和挑战的对策怎样？以何种方式增加价值？在当前情况下，作为给联合国系统提供服务的一个有效评价职能可采用什么备选方法？

本研究报告试图促进当前在全系统范围内为加强评价职能的能力所作的努力，以使评价职能能符合专业标准、应对新出现的挑战并在加强联合国系统的价值体系方面发挥作用。研究重点是中央评价职能，因为该职能一般用于支持组织一级的整体政策和战略决策。研究报告还快速评审了分散评价职能；这种评价职能在中央评价部门之外运作，一般嵌入联合国全系统范围内的方案和业务部门，以支持业务领导作出决策。

研究报告涵盖了参与世界各地发展、人道主义和规范性工作的联合国系统 28 个组织的评价职能：12 个基金或方案、12 个专门机构、3 个联合国其他实体和联合国秘书处的中央评价职能。但是，研究报告不包括联合国秘书处各部厅或区域委员会的嵌入式评价职能，因为联合国秘书处最近进行的一次评价研究已经涉及了这个问题。

本报告提出了 9 项建议：向联合国系统各组织行政首长提出 7 项，向联合国系统立法机构提出 2 项。

下文概述了主要结论及相关调查结果和本报告所讨论的关键问题。

A. 中央评价职能

1. 这些年来，中央评价职能有所成长，力求提高质量和效率，但整个联合国系统对评价的承诺水平仍与对该职能不断增长的要求和重要性不相符合。

经过评估后发现，在 36 年(1977-2013 年)的时间跨度里，联合国系统各组织的中央评价职能在作用、结构、制度和标准方面都发生了变化。这一职能已从主要以发挥监督和确保分散评价的质量的作用转向注重支持广泛和战略性的组织一级的决策。联合国系统面临的新出现的全球挑战要求中央评价职能更多地参与交叉和多部门类型的评价，以向联合国系统提供更综合和整体的解决办法。

各组织在发展自己的评价职能方面始终是务实和有成本意识的。这一点既反映在该职能在组织结构中的定位，也反映在为提高评价报告的效率和质量而拟定的业务模式上。但很明显，这些年来分配给中央评价职能的资源水平并没有以符合需求的方式发生显著改变。

中央评价职能的最初创建和发展曾受到联检组关于组织管理、行政和治理问题的几份报告的影响。大会的一些决议强调了评价职能的重要性并扩大了该职能作为一项问责制和学习工具所能发挥的作用的范围。联合国评价小组在过去 8 年里在加强该职能的专业发展和在统一评价方法方面同样发挥了重要和非常突出的作用，从而为提高该职能的地位提供了一个有效平台。

2. 在联合国系统里，评价系统的质量、机制、流程和产出各不相同。根据各自评价职能的发展水平，可将各组织分为 4 个群组。在大多数情况下，影响发展水平的是相关组织的规模、分配给评价工作的资源，以及该职能的结构所在地。

联检组这次在全系统范围内对 28 个组织评价职能发展状况的评估是通过分析能确定评价职能是否有效的 5 个核心构成部分进行的：(a) 适当的有利环境；(b) 独立性，同时具有包容性和公正性以增强公信力；(c) 质量，以增强公信力；(d) 对该职能的使用，重点在于利用评价结论和建议；以及(e) 相关性和适应性，以满足组织要求，并随时准备应对变化和挑战。

在分析中央评价职能的成熟度后，将所有组织分成 4 个群组：9 个组织优良、6 个组织一般、9 个组织低于一般水平，还有 4 个组织处于初级水平，连明确界定的评价职能都没有。这些成熟度是按照系统性和标准化的评估方式确定的，采用了基准框架(即，联检组的中央评价职能成熟度矩阵)。该框架针对输入的各种数据有 66 个业绩指标，包括联合国评价小组的规范和标准、基于联检组前几次报告的参数，以及其他国际发展行为体在评估组织实效时得出的数据。

与小型组织相比，大多数大型和中型组织被评为具有较高的评价职能发展水平。具有独立评价职能的组织的排名高于其中央职能是合署办公的组织，即评价职能或设在负责评价、审计、检查和调查的监督办公室内，或设在负责政策、战

略规划和研究的高级管理层办公室内。主要履行规范性任务的组织做得不如主要以发展为重点的组织那么好。但有例外，即一个机构只要有发展一个高质量评价职能的意愿和决心就能够克服组织障碍。

有 4 个组织没有中央评价职能，而就此提出的理由也不充分。其中 2 个组织已着手拟订评价政策或已开始磋商以制定一个正式的中央评价职能。

在对评价职能的 5 个核心构成部分进行分析后，从联合国系统 24 个有中央评价职能的组织中得出了以下结论。在考虑交叉结论时必须铭记各组织之间存在差异。

有利环境

3. 各组织在制定评价职能时侧重于对问责制方面的要求作出回应，而没有充分顾及其他重要要素，例如发展一种评价文化和把评价作为本组织的一种学习工具。这就限制了该职能的可持续性和评价的增加值。体制框架中需要一个以本组织任务、共同目标和改革议程为基础的对评价的总体愿景和战略，从而能用平衡的方法来发展评价职能。

会员国和立法机构在设立评价职能、确保在执行政策方面的合规和要求更新政策方面发挥了重大作用。会员国的主要重点一直是资源和成果问责制。各组织一直在回应这些要求，但没有同等重视发展学习职能和评价文化。它们没有使评价成为本组织结构的组成部分或承认评价的战略作用不仅仅是报告成果或业绩。问责制与评价的学习目标之间的不平衡不会提高评价职能的增加值和可持续性，特别是在当前这种快速变化的全球环境下；因为在当前的环境下，知识流的跨界速度很快，因此需要严谨和具有多维思考能力的灵活的专职知识工作者。

大多数组织的评价政策已到位。这些政策在很大程度上采用了联合国评价小组颁布的规范和标准。但其中大部分政策机械地采用了联合国评价小组的规范和标准，而没有根据机构的具体情况进行调整。同样，大多数组织对中央评价职能没有形成能顾及战略调整的总体愿景和战略，例如与其他监督和学习体系、与分散评价职能、与其他联合国组织或与国家体系的战略调整。缺乏一个广泛的战略愿景会限制适应性和决策，而在发展这一职能时采用临时性办法则会带来风险。

4. 联合国系统评价职能可用的资源不足，捉襟见肘。

经评估，专业评价工作人员对整个组织工作人员的百分比为 0.2%，对评价职能的供资额在组织支出中大约占 0.3%，这不足以应付对覆盖范围、质量和体制支持的不不断增长的需求。这种有限的很难满足对该职能广泛并日益增长的要求，也很难加强其作用以支持整个联合国系统。

大多数组织都需要向评价职能提供更多资源并制定适当的成本计算规范和框架，以指导对评价工作的预算拨款。还有一些组织需要更高效地利用工作人员

来集中精力履行核心评价职能，而不是把精力放在其他监督或管理职能上。在提高中央职能效率方面同样重要的是，要发展体制支助职能，例如改善监督制度和数据质量、目标明确和设计完善的方案、成熟的成果框架和业绩指标，这些都有助于提高方案的可评价性。

在提高效率和增加值方面具有同等重要性的是，要探讨方案交付的替代方式，以便在发展评价系统时有更多的相互依存性和协作，并增加联合评价。目前正在进行的关于 2015 年后发展议程的辩论表明，能使联合国系统存活和发展壮大的最佳途径就是打破其各自为政的“筒仓”式格局。这就意味着，应重组联合国系统的评价职能，扭转那种每一个组织“单打独斗”的普遍做法，尤其在小型组织里，并让更多的伙伴和机构参与进来。这种评价职能在提供多部门交叉信息方面能发挥更大的作用，有助于联合国系统做出整体战略性决策。

评价职能的公信力：独立性和质量

5. 通过增强独立性和发布质量更高的评价报告，在提高评价职能的公信力方面取得了进展，但还需要做更多的工作，特别是要解决结构独立性、评价职能的外部质量保证和报告质量等问题。

评价职能的独立性是在用于确定高质量评价职能的 5 个构成部分中发展水平最高的构成部分。对照独立性 5 项标准(结构、职能、专业、技术、包容性和行为独立性)所作的评估结果表明，2 个组织具有高度业务独立性、16 个组织高于一般水平、6 个组织处于一般水平、6 个组织远低于可接受的独立性水平。在增强独立性方面取得了很大进展，因为评价管理人员和评价工作人员在维护高标准的对独立性的专业和个人忠诚方面发挥了作用。高级管理层在保障独立性方面同样有所进步。具有包容性的独立性是联合国评价系统一个重要而又非常受重视的特征。对于一个像联合国系统这样的多文化、多层面的政府间机构而言，独立性能增强信任和透明度。联合国系统各组织的大多数评价部门在提高利益攸关方对评价工作的参与度、同时又保障独立性方面做得很出色。在加强结构独立性方面仍然存在挑战，主要是要解决评价职能主管的独立性与上下级关系问题。在加强独立性及提高评价职能的技术质量和公信力方面需要解决的其他全系统范围的问题包括，要解决评价部门主管的区域和性别不平衡问题，并确保部门主管具有专业评价知识。

在评价职能的专业和技术质量方面，主要优势就是工作人员的能力和行为，但需要做更多的工作来支持工作人员的专业发展并使他们出类拔萃。很多组织的一个常见做法是让利益攸关方参与，以加强评价报告的内容、价值和意义。重大挑战包括：(a) 在对报告的质量评估或元评价方面，使用独立的外部质量保证机制的程度有限；(b) 缺乏关于评价报告质量的可靠信息；(c) 目前的报告质量不高。于 2006 年启动、负责评估评价职能和报告质量的联合国评价小组同行审议机制只在 9 个组织中得到落实。对报告质量的外部评估有一些共同要素，但使用了不

同的衡量制度。最近几年来，双边组织对评价职能使用外部评估的次数有所增加。需要协调用于对评价职能及其产品作质量评估的所有现行机制，以避免重复、加强可靠性、限制交易成本及增强统一性和效率。

效用：利用评价和利用的实效

6. 各组织不倾向于高度利用评价来支持循证政策和决策，以制定战略方向、改善活动方案和创新。

研究发现，大多数组织对评价报告按既定目的加以利用的比率一直很低。甚至在那些评价职能被认为运作良好的组织里，目标受众利用报告的水平也仅达到一般水平。利用率不高与问责制驱动的重点及上文提及的评价职能在发展学习型组织方面的作用受到限制有关。需要改进用于评估评价利用情况的现行制度。同样，需要改善用于评估评价对组织实效的影响的制度。目前，用于评估这种影响的制度是不成熟和临时性的。由于关于评价报告的利用率和影响力的信息有限，因此很难进行分析，以充分了解评价职能的总体价值。但是，在利用评价影响决策并将学习化为行动方面业绩不善的并不单单是联合国系统。现有证据表明，其他发展伙伴有同样的问题；这就表明，需要作出一致努力来解决这一问题，并加强了解评价职能的价值。

鉴于利用评价职能是有益的，因而该职能也是有价值的，因此研究报告以探索性方式研讨了评价职能的全面性价比情况。分析显示，评价职能最发达的 9 个组织在联检组 28 个参与组织中占组织支出总额的将近 70%。虽然这个数据只是指示性的，但有证据表明，联合国的评价职能具有良好的性价比，而只要继续发展该职能，包括提高评价报告的利用率，性价比就会更高。

是否做好了支持变革的准备

7. 联合国各组织的任务和业务模式不同、对“筒仓效应”（即不以广泛的战略方式思考或不考虑联合国各机构内和之间的更为广泛的联盟和协调）的趋向性，以及一些组织的评价能力普遍过低等因素阻碍了评价职能在驱动联合国系统内进行变革方面发挥关键作用。与此相关的是联合国系统的分类职能结构。

评价职能用于解决联合国系统变革和新出现挑战问题的准备状态在大多数情况下是不到位的。在以联合国评价小组大量规范产品为基础形成的方法和标准中存在着统一和一致性。但是，还有待各组织将这种统一和一致性转化为广泛的联合工作。联合评价的内容有所增加，主要是在性别平等、粮食安全和人道主义领域，但因系统性制约因素而受到限制。大多数组织的评价职能尚未适应不断变化的全球动态、对相互依存和相互联系的现实需求和如何把它们当作机会加以利用以重组评价职能。许多组织仍在传统做事方式中寻找最佳做法，而现在对联合国系统最有用的可能是要思考采用以各组织之间的联系和分系统为重点的替代性综合协作系统，以使业绩达到最优化。

包容性(即多样性、性别平等和人权)及评价的国家自主权和能力议题是发展领域里的两项重要挑战。妇女署的持续参与一直在促进增强在管理、设计、进行和报告评价结果方面的性别平等。在接受评审的组织中有 9 个组织显示在将性别平等观点纳入评价工作方面业绩优良。

许多组织一般不把发展国家的评价能力(有几项决议强调了这一点)视作为一项交叉任务。除了对国家治理有价值外,联合国系统大多数组织还承认,国家评价能力的发展会对各组织自身评价职能的可用性和可持续性产生影响,不管其组织规模、结构、活动或实地存在的情况如何。但是,评价职能对国家评价能力发展所提供的支持是不到位的。各组织显示,为增强评价能力或评价结果的所有权在与国家评价机构和专家的联系方面采用了不同的形式。只有 3 个组织根据战略方针落实了支持发展国家评价能力的活动。

尽管有人认为,加强国家评价能力是方案责任,但要求拥有自有评价专家的中央评价职能提供支持的呼声从未减弱过。这些要求及其他若干对中央评价职能越来越强烈的要求导致对联合国评价系统交付预期成果的能力产生疑问。在改善评价质量和增加值方面,评价部门主管和工作人员的活力和决心一直是主要驱动力,但仅凭他们自己不足以确保评价职能的可持续性和相关性。评价部门主管现已着手实施联合国评价小组用于应对其中一些挑战的一项新战略,但需要充分思考非传统的替代方法,并需要包括管理层、会员国和私营部门伙伴在内的所有利益攸关方提供支持。

B. 分散评价职能

8. 缺乏一个以评价政策或成果管理制为基础的全面和明确界定的体制框架使这一分散职能变得很脆弱。

分散评价是在中央评价部门之外规划、管理和开展的。它们嵌入在负责规划和实施项目和方案的方案部门和管理部门内。在结构上,它们不独立于业务主管部门。这适用于由项目管理人员自行酌定规划和管理的评估和委托或外包给外部咨询人的那些评价。今天,大多数组织没有坚实的体制框架来支持这一分散职能。只有 11 个组织在评估时被确认具有明确界定的系统或系统要素可用于分散评价。

研究报告分析了联合国系统内分散评价的不同典型和模式,并确认了分散评价系统的两个主要模式。

第一种模式是没有明确体制框架的临时性系统,延续了总部和外地办事处项目工作人员进行的自行酌定的项目自我评价传统。这是 28 个组织中除 1 个组织外都采用的做法,涉及大量项目管理人员;他们作为一种知识力量在不停地生产大量项目评价。他们的评价工作的数量和质量不为人知,因此减少了这类评价的战略效用。

对临时性分散评价系统缺乏关注一事表明，联合国系统正在失去更好地利用已产出的评价报告所产生的资产的机会，正在失去与参与整个系统正在实施的这类评价的工作人员进行联系的机会。在目前情况下这是个重要问题，因为这类工作人员在满足对及时提供实地证据的需求和在实现最重要的即时变革方面发挥着关键作用。关于 2015 年后议程的辩论强调了这个问题的重要性，因为如果要对快速变化和复杂的挑战作出反应的话就必须进行持续不断的形成性评价以在执行过程中随时作出矫正。

第二种模式是一种包含明确计划、质量保证体系和系统性报告制度的分散评价系统。如上所述，经评估，在联检组 24 个参与组织中只有 11 个组织有这类系统，但程度不同，制度化的水平也不一样。经评估，这些组织中的 10 个组织目前依靠在总部及在区域和 150 多个国家办事处的方案和技术部门的 923 个点(即评价部门或指派负责监督的评价任务及评价专家或联络人)开展运作，每年产出 640 多份报告。平均而言，这些报告中的 64%(这 10 个组织为一组)是产出一级的评价，10% 是成果评价。影响力评价有 12 个，其余的评价类型不明。

研究报告收集了关于这 10 个组织的进一步证据，下文的意见适用于这 10 个组织。

9. 政策、规范和标准，以及用于分散评价的资源是不够的。

被抽样评审的这 10 个组织的分散职能缺乏一个“与目的相符”、与分散评价的作用及比较价值和增加值相符的适当的规范性框架。联合国评价小组的规范和标准框架不能完全满足分散职能的需求。但正在建立专业网络，以讨论分散评价、其价值体系和游戏规则、其关键问题和挑战，以及为各组织和整个系统的利益应发挥何种作用等问题。这些网络有可能会大大促进评价职能的发展，因此需要扩大范围、得到授权和支持。需要联合国各机构作出一致努力，以促进发展这类网络。

现有的组织评价政策在某种程度上为发展分散职能提供了指导。但这些政策未涉及两个重要领域：(一) 一致性；(二) 与国家评价的协调。就其发展水平而言，这两个领域属于分散职能中最薄弱的环节。

在一些组织中，用于分散评价的资金来自于项目和方案预算。与依赖预算外供资的情况相比，这种做法提供了某种稳定性。但是，相对于其他国际发展实体用于分散评价的资金，为开展评价提供的款额是非常低的。

10. 分散职能的公信力有所改善，但仍有很多工作要做。

为限制偏见和增强评价的公正性的许多措施已经到位，因而提高了公信力。分散评价报告的质量有所改善，但需要做更多的工作来增强各个制度，还需要安置更多的人员以进一步改善质量。同样，希望加强国家一级的一致性，特别是要

统一通过联合国发展援助框架(联发援框架)提出的框架。采用联发援框架已有 10 年了,但在许多情况下,这类框架尚未得到统一。分散评价中的协调或协作程度有限。在国家一级评价中,分散与中央评价职能之间的协调几乎是没的。

11. 对分散评价的利用并从中学习的程度有限,因为组织文化的重点是问责制和对捐助方的回应。

像中央职能一样,分散职能对问责制要求的回应程度很高。分散职能还侧重于项目和方案的形成性改善,因为这被认为是分散职能的一个主要价值。但没有想到要利用分散职能来加强对工作人员的赋权和促进从评价中学习。即使是大量项目管理人员正在实施的专案评价,事实表明,大多数是因捐助方的要求驱动的,因为这是与预算外供资挂钩的议定书的组成内容。

12. 分散职能在解决当前及新出现的变化和联合国系统的改革问题方面可发挥重要作用,但若要使分散职能作出贡献,就必须使其成为就发展评价职能所采用的整体战略方法的组成部分。

发展的速度很快,这就要求持续提供及时的证据,以指导变革和发展进程,取得预期的成果和影响力。分散评价是作为项目实施工作的组成部分运作的,因此在提供经常性证据以进行矫正或在采用替代方法方面发挥着重要作用。联合国系统的分散评价职能有各种模式,未得到明确界定的体制框架的充分支持,其质量和增加值尚不清楚。需要发展战略重点,为提高分散评价的作用和相对价值做出规划,以在组织一级和全系统一级支持联合国系统。

建议

建议 1: 有合署办公和与其他监督职能整合在一起或与行政管理职能整合在一起的中央评价职能的联合国系统各组织行政首长应加强该职能并确保其质量、完整性、可见度和增加值。

建议 2: 联合国系统各组织行政首长应采用平衡方法来解决对问责制的评价目的问题,并促进发展一个具有激励革新、敢担当风险和采用多学科观点的适当机制的学习型组织。

建议 3: 立法机构应要求联合国系统各组织行政首长根据为维持一个能为本组织增加价值的有效和可持续评价职能所需的成本为各自的评价职能制定综合预算框架和资源分配计划。应在现行预算和报告机制和流程范围内向立法机关提交这些计划供审议。

建议 4: 立法机构应指示联合国系统各组织行政首长酌情审查和修订用于任命评价办公室负责人的现行政策,以加强独立性、廉政、道德操守和包容性,同时要适当考虑以下标准:

- 应为单一、不可延长的任期设立一个 5 年至 7 年的期限，而且任职者没有再次被该组织雇用的可能性；
- 评价负责人应具有评价方面的资格和实质性经验，并有在战略规划、基础和业务研究及知识管理等相关领域方面的经验作为辅助，以及应有出色的管理和领导能力。

建议 5：联合国秘书长应以联合国系统行政首长协调理事会(首协会)主席的身份要求联合国评价小组进行协作，以拟定一个可用于整个联合国系统评价职能的健全和统一的质量保证制度。

建议 6：联合国系统各组织行政首长应根据对评价职能的明确界定的愿景、战略和成果框架把利用评价作为一个组织优先事项，并就利用评价的程度、性质和影响力情况向其立法机关提交报告。

建议 7：联合国系统各组织行政首长应要求评价办公室重新评估涉及评价职能的政策、战略和优先事项。他们应把评价职能置于各自组织的战略位置，从而加强其相关性，以增强联合国系统在应对当前变化和挑战及在实现影响力和可持续性方面的能力。

建议 8：秘书长应以首协会主席的身份采取步骤并支持创新，以促进联合国系统各组织之间及与其他伙伴之间的协作，加强国家在涉及问责制、学习及具有国家和全球价值的知识发展方面的评价能力。

建议 9：联合国系统各组织行政首长应制定体制框架和必要的支持系统，以提高分散评价的质量和增加值及提高其在支持联合国系统应对新出现的挑战(包括 2015 年后发展议程的那些挑战)方面所能发挥的作用，以及加强联合国系统各组织内部和全系统范围及与国家机构的一致性和协调。

目录

章次	页次
内容提要	3
简称	14
一. 导言	1-35 16
A. 概述	1-5 16
B. 背景：评价职能的概念框架——要求和发展高质量的评价职能.....	6-19 17
C. 方法	20-34 21
D. 鸣谢	35 25
二. 中央评价职能：演变、成长和发展水平.....	36-139 26
A. 发展中的变化：目的、结构安排、覆盖范围和管理.....	36-50 26
B. 组织一级的业绩：各组织的模式和各组织之间的差异.....	51-60 31
C. 职能的构成部分：优势、挑战和系统性制约因素.....	61-139 36
构成部分 1：有利环境、体制框架和支持系统	62-85 37
构成部分 2：具有包容性的独立性，以提高公信力.....	86-98 44
构成部分 3：质量-技术和管理严格性，以提高公信力	99-105 50
构成部分 4：效用-利用和利用所产生的影响.....	106-114 51
构成部分 5：相关性和为支持联合国各组织和全系统的改革及 为应对新出现的变化和挑战所作准备的状态	115-139 55
三. 分散评价职能	140-189 62
A. 导言	140 62
B. 体制和规范性框架	141-152 63
C. 政策、执行和制度化	153-161 66
D. 资源	162-164 68
E. 公信力(公正和质量)和效用.....	165-175 69
F. 为应对变化和挑战所作准备的状态：不断变化的要求、能力和 准备状态	176-189 72

附件

一. 联合国系统内的评价职能成果框架.....	75
二. 联合国系统内的评价职能构成部分和次级构成部分.....	77
三. 来自各基金和方案、专门机构、联合国秘书处和其他实体的联检组参与组织.....	79
四. (a) 联合国系统内评价职能的架构.....	80
四. (b) 联合国系统内评价职能的架构(关于联合国秘书处结构安排的详细情况).....	81
五. (a) 联合国系统评价职能一些术语的定义：中央和分散评价.....	83
五. (b) 联合国系统内的分散评价职能和原型.....	86
六. 联合国系统 23 个组织所使用的业务模式.....	87
七. 对评价职能超出实施和提交评价报告之外的要求.....	88
八. 1977-2013 年联检组参与组织评价办公室的演变.....	89
九. 学习型组织：促进成果、问责制、双圈学习、共享和利用评价的文化.....	91
十. 领导力属性和职能.....	92
十一. 非核心供资、对有效和独立利用的保障与评价职能发展水平之间的关系.....	93
十二. 为应对全球变化和挑战包括 2015 年后发展议程规定的任务以促进可持续发展所作准备的状态.....	94
十三. 各参与组织就联合检查组的建议将要采取的行动概述.....	95

简称

首协会	联合国系统行政首长协调理事会
粮农组织	联合国粮食及农业组织
原子能机构	国际原子能机构
国际民航组织	国际民用航空组织
劳工组织	国际劳工组织
海事组织	国际海事组织
国贸中心	国际贸易中心
电信联盟	国际电信联盟
联合检查组	联检组
人权高专办	联合国人权事务高级专员办事处
监督厅	内部监督事务厅
艾滋病署	联合国艾滋病病毒/艾滋病联合规划署
贸发会议	联合国贸易和发展会议
联发援框架	联合国发展援助框架
开发署	联合国开发计划署
环境署	联合国环境规划署
教科文组织	联合国教育、科学及文化组织
人口基金	联合国人口基金
人居署	联合国人类住区规划署
难民署	联合国难民事务高级专员公署
儿基会	联合国儿童基金会
工发组织	联合国工业发展组织
毒品和犯罪问题办公室	联合国毒品和犯罪问题办公室
项目厅	联合国项目事务厅
近东救济工程处	联合国近东巴勒斯坦难民救济和工程处
妇女署	联合国促进性别平等和增强妇女权能署
旅游组织	联合国世界旅游组织

万国邮联	万国邮政联盟
粮食署	世界粮食计划署
世卫组织	世界卫生组织
知识产权组织	世界知识产权组织
世贸组织	世界贸易组织

一. 引言

A. 概述

1. 本报告阐述并分析联合国系统内评价职能的演变和发展。编写本报告的目的是要帮助加强评价职能的能力，以达到专业标准并能应对全系统新出现的全面挑战。

2. 报告沿用了联检组在 1977 年至 1998 年之间就联合国各组织评价职能发表的前几份全面报告。¹ 报告是对联检组在 2005 年至 2011 年之间就监督和问责制、² 成果管理制³ 和知识管理⁴ 等问题进行的全系统研究的补充。此外，联检组目前业务战略的一部分就是要进一步加强工作，为发展评价职能作出贡献。通过本报告，联检组力求促进三大发展：支持执行新制定的《2013 年联合国评价小组战略》、为 2015 国际评价年的庆祝计划提供资料，以及在 2015 年后发展议程中加强联合国系统评价职能的准备状态、战略定位和优先次序。

3. 本报告有大量辅助材料作为补充；可在联检组网站上查阅这类材料并应结合本报告加以阅读。这些材料包括各种附件，以表格和图表说明形式提供了详细的数据分析摘要。相关数据为本报告得出的结论提供了详实证据，并包含各组织业绩的详细情况，从而为制定基准提供了丰富的信息基础。辅助材料还包括用于收集数据的评估工具，其中包括：关于中央和分散系统的调查表、针对无明确职能的组织分发的调查表，以及用于评估中央评价职能发展水平的联检组成熟度矩阵和方法。联检组完成了联合国系统 24 个组织中每一个组织的成熟度矩阵，并在 2014 年第一季度向各组织提供了这一矩阵。⁵ 一些组织报告说，已利用联检组的矩阵评估结果来解决确认需要改善的领域，还有一些组织利用该矩阵作为正在进行的联合国评价小组同行审评的组成部分，另有一些组织利用矩阵来更新其评价政策。

4. 根据《联检组章程》第 11.2 条，检查专员们在本报告定稿前进行了协商，目的是用联检组的集体智慧来检验报告的结论和建议。

¹ 联检组在 1977 年至 1998 年期间分发的相关报告有：JIU/REP/98/2、JIU/REP/88/6、JIU/REP/85/10、JIU/REP/83/5、JIU/REP/82/12、JIU/REP/82/10、JIU/REP/81/6 和 JIU/REP/77/1。

² 可作为本报告补充材料的联检组报告有：联合国系统的调查职能 (JIU/REP//2011/7)、联合国系统的问责制框架(JIU/REP/2011/5)、联合国系统的审计职能(JIU /REP/2010/5)、联合国系统的知识管理(JIU/REP/2007/5)和联合国系统的监督缺陷(JIU/REP/2006/2)。

³ 关于联合国系统成果管理问题系列报告的概述(JIU/REP/2004/5)。

⁴ JIU/REP/2011/7、JIU/REP/2011/5、JIU/REP/2010/5 和 JIU/REP/2006/2。

⁵ 联检组中央评价职能成熟度矩阵：一个用于评估联合国系统各组织评价职能发展水平的客观和标准化的方法，联检组，2013 年。

5. 为便于落实本报告所载各项建议并加以监测，附件八载有一个表格，列明是否已将报告提交给相关组织以采取行动或供其参考。该表格列明了与每个组织相关的建议，具体说明是否需要由各组织立法或理事机构做出决定，或由该组织的行政首长采取行动。⁶

B. 背景：评价职能的概念框架——要求和发展高质量的评价职能

评价的目的：问责制、改善和应对日益增长的要求，以支持改革和全球变化

6. 人们普遍认识到，联合国系统的改革能力和持续相关性在很大程度上依赖于它是否有能力实现自己的目标、获得成功和实现必要的改善。还依赖于它是否有能力学习和应用替代办法、大力积累知识和良好做法并加以利用，以充分了解国家、区域和全球环境，在国际发展和治理方面实现必要的变革。

7. 在使联合国有能力满足对成果问责制和对其活动产生增加值的要求方面，评价发挥着关键作用。评价就是要判断联合国系统活动的价值或作用。它提出下列问题：我们在做正确的事(即，这些事是否切合实际、能够因应需求、适当、具有创新性和不是重复性的)吗？我们在以正确的方式做事(即，我们是以高效、协调一致和包容性的方式做事)吗？我们正在取得成果、发挥作用和增加价值(即，我们在有效地取得成果并且我们工作的贡献和意义有多大)吗？我们是否能确保我们的贡献能产生影响、我们的工作是可持续的？因此，评价是要“根据一项活动的目标和效益尽量系统性和客观地确认其相关性、效率、实效和影响力”。⁷

8. 评价不仅要说明“什么”，而且要说明“为什么和如何”及“在什么情况下”。在这样做时，评价就提供了一个良好的基础，以支持管理层作出必要决定和采取必要行动，改善政策、战略、项目、方案及业务活动和规范性活动。评价的主要优势在于其方法的严格性，可通过系统性的关键调查来提供强有力、有效、可靠和可信并可放心地用于决策的证据。

9. 正是评价专业的那种反思型离散思维模式使评价成为一个可用于建设一种所谓学习型组织的文化的重要工具。学习型组织需要知识，并以足以在一个快速变化的环境中存活和壮大的速度进行创新和作必要的转型变革。学习型组织：
(a) 创造一种鼓励和支持雇员不断学习、通过评价开展批判性思维和勇于提出新

⁶ 必须指出，这些建议并不一定是具体、可衡量、可实现、相关和及时的，因为那样做的话就无异于是确定项目参数，从而限制以后独立评估的能力。鉴于各组织之间存在差异，这些建议允许每个组织在拟定行动方向时采用专业方法。本报告提供了实质性信息，可用于指导根据各种组织的特点拟定对每一项建议的适当对策，同时做到尊重评价职能的标准、目标和完整性。

⁷ 《方案规划、预算内方案部分、执行情况监测和评价方法条例和细则》(ST/SGB/2000/8)，附件，第 15 页。

观念的文化；(b) 允许犯错误并重视雇员的贡献；(c) 从经验和实验中学习；以及(d) 在整个组织里传播新知识以纳入日常活动。在一个学习型组织里，众人参与“能增强创造力的再生性学习，他们不断学习以了解全貌”。⁸

对评价职能的日益增长的要求和影响

10. 这些年来，对联合国系统评价职能的要求增多了，并且在关于联合国系统的相关性问题的辩论中，评价被置于中心地位。评价被认为是必不可少的，不仅可帮助联合国满足对成果问责制和改善做法的日益增多的要求，而且还可帮助满足各种其他要求。其中包括：(a) 支持发展一种学习和评价文化；(b) 监测和报告评价政策的执行和遵守情况；以及(c) 根据改革议程促进一致性、相互成果问责制、国家评价能力发展、性别平等和人权，以及多样性和包容性。这些要求在大会的一些关于评价问题的决议⁹ (有些可追溯到 1980 年代)及在一些关于援助实效、相互问责和成果管理制的文书里得到了反映。¹⁰

11. 还有最近在全球背景下发生的变化，包括 2015 年后发展议程规定的任务，要求转型、重组及为快速行动和创新持续提供证据。要做到这些就必须采取跨部门的综合方法，分析联合及全系统工作和全球参与情况。这会在三个层级上产生影响。

12. 第一，由于需要进行改革和必须制定战略方向，大量中央评价部门被越来越多地要求将注意力转向评价联合国各组织的战略性重新定位和相对增加值问题，从而作为变革的推动者发挥着更大的作用。这种影响导致所进行的评价类型发生了变化，现在变得更为复杂了。在资源有限的情况下，一些中央职能现在不那么关注业绩评估或具体的项目评价了，这些工作正转由嵌入在管理部门的分散评价职能处理。现在的问题是，分散职能是否有能力生产高质量的评价报告。第三章讨论了中央与分散评价系统之间的结构关系和中央职能在提高分散评价的能力和质量和/或控制评价报告质量方面所正在发挥的作用。

13. 第二，在国家和全球平台上发生的快速变化突显出在困难环境下新出现的复杂发展进程。这就需要采用新的评价模式，以应对下列情况：(a) 在评价时不可把发展视作是一种有明确界定和确定的假设，而应视作是一种内源性转型进程，因为发展是在自然环境中发生的，因此不是总能控制得了的；(b) 在转让发展干预技术中使用最多的最佳做法模式在发展环境中通常是行不通的，这类受环境限制的干预技术是杂乱无章、难以评价的；以及(c) 需要就什么东西行之有效、为

⁸ BusinessDictionary.com。Peter Senge, 《第五种学科》，麻省理工学院，1990 年。

⁹ 例如，大会第 59/250 号、第 62/208 号和第 67/226 号决议。

¹⁰ 《援助实效问题巴黎宣言》、《阿克拉行动议程》、《釜山有效发展合作伙伴关系》和《国际发展筹资会议蒙特雷共识》。

什么和如何等问题提供及时或实时信息，以支持及时、相关和快速的决策流程和中期矫正，而所有这些对于最大限度地提高干预技术的相关性、实效和价值都至关重要。

14. 第三，2015年后发展议程对联合国系统提出了以相互依存和相互联系的方式一起工作的要求。这就要求在跨组织方案编制和进行评价方面不仅仅需要协调，而且还要加强协作，包括审查联合国系统对跨部门的全球、区域和国家各级需求和优先事项所作回应的总价值和总影响力。最近对各组织评价职能提出的参与联检组全系统评价的要求就反映了这一点，例如通过秘书长的说明(A/68/658-E/2014/7)所载的联合国系统发展方面业务活动独立全系统评价政策。这同样反映在增强营养倡议里，因为该倡议要求联合国在制定方案和进行评价时与多个行为体以伙伴关系方式开展工作。

15. 这些要求对评价职能产生了重大影响，要求该职能重组和解决互补和协调问题并使之合理化。应该在联合国系统的内部评价职能之间完成这项工作。为加强伙伴关系，还应与各种国家伙伴和发展伙伴及私营部门和其他新出现的、参与评价工作的全球行为体进行协调。

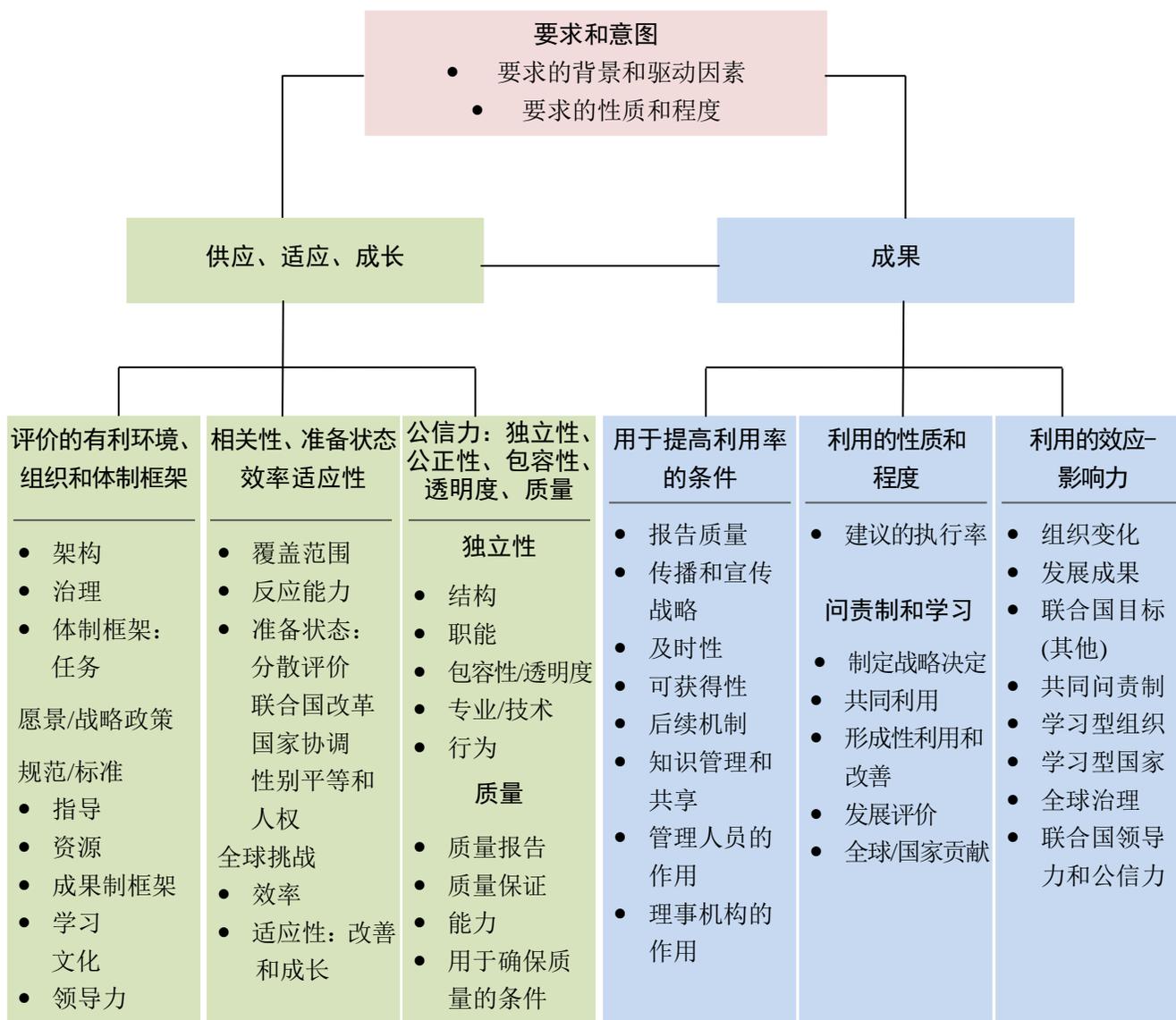
联合国系统为满足对高质量评价职能和发展这一职能及支持系统的日益增多的要求所采取的对策

16. 为应对不断增多的要求，这些年来联合国系统的一些组织一直在寻求发展高质量、高实效和反应灵敏并能支持问责制、学习和改革的评价职能。它们为发展评价职能的核心构成部分制定了业务模式、结构、政策、制度和机制，本报告述及其中一些部分：

- 构成部分 1. 一个能支持评价职能的适当有利环境和体制框架
- 构成部分 2. 独立和/或公正的评价，以促进公信力
- 构成部分 3. 高质量的评价流程和产品，以促成公信力
- 构成部分 4. 评价职能的效用，包括对评价的利用和影响力
- 构成部分 5. 系统在应对组织对方案的覆盖要求方面的相关性和在应对要求方面的适应性；以及与是否已准备好满足在改革及新出现的全球变革和挑战方面的要求相关的层面

17. 下文图一列出了评价职能的概念框架，突出了要求与对应要求的关键构成部分之间的联系。

图一
评价职能的构成部分——联检组模式



18. 当上文第 16 段概述的各个构成部分的成果得到实现，预计评价职能将：(a) 根据可信的证据加强联合国系统的问责制和监督；(b) 提高计划、战略、政策、方案和项目的质量和加强体制改革；以及(c) 加强联合国系统在与国家伙伴及在区域和全球一级共享从评价中得到的教训方面的领导作用和贡献。附件一阐述了评价职能在支持这些目标上的逻辑模式或成果框架及对人民生活的最终实效或影响力。

19. 本研究报告注重的是与这些构成部分有关的结构、制度和机制。附件二为评价职能 5 个构成部分的每一个部分提供了一整套指标，界定了对加强评价职能而

言至关重要的重要制度、业务和产出。已确认有 66 个指标¹¹ 对评价职能至关重要。它们可与双边和多边指标作比较。可在联检组网站的辅助材料里查阅关于这些指标和业绩标准及评估方法的详细情况。

C. 方法

主要问题

20. 就评价职能提出的主要问题如下：

- 在联合国系统评价职能的成长和发展中取得了哪些进展或变化？
- 参照用于加强其相关性、公信力和效用的构成部分，当前评价职能的发展或成熟度达到了什么水平？用于支持变革和应对挑战的准备状态达到了什么水平？
- 在不同的组织之间存在着何种差异？有哪些良好做法？有哪些主要优势、系统性制约因素和挑战？有哪些可用于提高效率 and 实效的最佳替代模式？
- 用什么来加强评价职能的持续发展和进步？

范围

21. 本次研究是在全系统范围内进行的，涵盖了联检组 28 个参与组织，包括各基金、方案、专门机构和其他组织。附件三提供了本次研究涵盖、按类型和任务分列的组织名单。应当指出，就联合国秘书处而言，只包括了内部监督事务厅(监督厅)和中央评价司。名单里未列入秘书处内有嵌入式评价职能的部厅。监督厅最近的一份报告载有关于这些部厅的信息。¹²

22. 研究报告的首要重点是中央职能，其次是联合国系统的分散评价职能。附件四(a) 概述了联合国系统内评价职能的架构及这两个职能如何支持许多组织作出不同层级的决策。可在附件四(b) 里看到该架构的衍生部分，看到关于联合国秘书处里在监督厅之下有嵌入式评价职能的部厅的信息。

23. 鉴于联合国系统各组织在阐述评价系统各要素时所使用的术语有很大差异，因此必须进一步明确区分联合国系统两个层级的评价架构，即中央职能和分散职能，这也是本研究报告的重点。附件五提供了关于这两个层级的详细情况。联合国系统内的中央级职能运作时独立于业务主管部门的控制，在某些情况下，在规划、管理、实施和报告评价工作时独立于高级管理层或组织负责人。虽然中央职

¹¹ 联检组成熟度矩阵有 70 个指标，但本研究报告只评估了 66 个指标。2 个指标未得出有效评估结果。2 个指标未以全面方式用于评估所有组织。因此，本研究报告放弃了 4 个指标。

¹² 2012 年 1 月 1 日至 12 月 31 日期间内部监督事务厅与维持和平行动有关的活动(A/67/297(Part II))。

能在许多情况下开展不同类型的项目评价，但总体上用于支持组织的广泛政策、方案和战略决策。

24. 分散评价是在中央评价部门之外规划、管理和开展的。它们嵌入在负责规划和实施项目和方案的方案部门和管理部门内。这包括在技术、区域和国家办事处一级的项目和方案或部厅。因此，分散评价职能一般是作为方案管理的组成部分运作的，负责处理评价、监测和其他相关分析活动以支持作出与项目和方案的各个实施阶段相关的决策。

25. 因此，分散评价在结构上是不独立的，因为它们评价的规划、管理、开展、核准和报告方面受业务主管部门的控制。由方案工作人员自行酌定管理和开展的评价(在一些组织里将这类评价称作自我评价)和委托或外包给外部咨询师的评价(这类评价报告的最后核准权在方案管理人员手上)就属于这种情况。这同样包括联合国发展援助框架(联发援框架)的评价。本研究报告未述及联发援框架评价系统，因为该系统可能成为对发展方面业务活动独立全系统评价拟议进行的试验研究工作的对象。¹³ 在其他可信的独立机构控制下在分散层级上规划的评价，并且相关的最后核准和报告的发布是独立的，就可评估为独立评价。但是，其公信力依赖于所有关键利益攸关方的参与是否透明。附件五(b)载有分散系统原型的详细情况，并在第三章里进行了讨论。

26. 附件四(a)显示评价职能的架构包括一个第三层级，即联合国机构间和全系统评价。本研究报告不包括处理全系统评价的职能，因为秘书长最近已委托进行侧重于这个方面的研究。¹⁴ 在联检组 28 个参与组织中，4 个组织没有中央职能，但开展专项评价。研究报告简要介绍了这 4 个组织的评价职能。因此，本研究报告始终把重点放在联检组其余 24 个参与组织的中央职能上。

27. 本研究报告没有评估评价职能的影响力。附件一内评价职能成果框架的前面部分载有对这一问题的说明。但是，研究报告确实评估了影响力评估系统。研究报告还评估了附件一所列成果或代用成果指标中一些特选指标的实现情况。

28. 本研究报告的目标是：

目标 1. 就评价职能的演变情况提供可追溯到 1977 年的信息。这些信息为突显评价职能的进步及其为应对要求而转换作用的过程提供了一个背景。这些信息具有历史重要性，对于追溯评价职能的变化很有用。

目标 2. 阐明各组织中央评价职能的状况，分析在何种程度上达到了评价职能构成部分的既定专业标准，并指出这些构成部分的交叉优势和弱点。这就

¹³ 联合国系统发展方面业务活动独立全系统评价政策(A/68/658-E/2014/7)。

¹⁴ Angela Bester 和 Charles Lusthaus，全系统独立评价机制：根据大会第 64/289 号决议的任务规定全面审查联合国系统发展方面业务活动全系统评价的现有体制框架(最后报告，2012 年 4 月)。

为各组织之间的差异提供了信息，同时又显示了共同和不同的组织挑战及示范性做法。

目标 3. 分析评价职能为适应全球变化及应对全球挑战和 2015 年后发展议程规定的任务所作准备的状态，借鉴联检组和评价办公室负责人确认的属性，因为这些属性对于应对联合国系统的变化和要求至关重要。

目标 4. 通过快速评审阐明分散评价职能的状况并分析其发展情况。研究报告增加这一目标，因为要就联合国评价系统提供一个更全面的视角，从而提供一个基础，以平衡的战略性方法支持中央和分散职能。

目标 5. 确认应予消除的结构和系统制约因素，以加强用于交付可信的评价证据的能力并提高整个联合国系统对这类证据的使用率。

目标 6. 以所得出的结论和经确认的示范性做法为基础提出建议，以说明如何加强联合国系统评价职能的持续发展和加强与涉及该职能的所有关键角色(立法机构、管理和评价部门)的相关性。

方法

29. 研究报告审视了 2013 年的评价情况，同时还展望未来的需要，以使评价在支持联合国系统实现各项目标方面发挥重大作用。严格地应用了定量和定性方法，以确保为达成结论所使用的证据的准确性、有效性和可靠性。对 24 个组织中央级业绩所作分析的基础是：对 500 多份文件的二次分析(平均每个组织 20 份文件)、交叉验证数据、对来自联合国评价和管理部门的 150 多位工作人员和管理人员的访谈和咨询，以及对会员国和外部伙伴的访谈。

30. 还验证了概念和结构、关键利益攸关方的参与情况、对证据的交叉验证和三角划分；广泛使用显着性统计检验法，以评估对所提出的调查结果的信心度；¹⁵ 以及利用定性方法评估对立假设。按照大会授权，研究报告既涉及了评价职能的硬系统和机制，也涉及了评价和学习文化。评价职能有多重角色，担负着互补性责任——立法机构、行政首长和工作人员，以及评价管理人员和工作人员。因此，本研究报告对评价职能发展水平所作的评估反映了所有这些角色的共同努力。

31. 虽然中央评价职能和分散评价职能都具有用于界定评价职能的 5 个构成部分，但它们在目标上有不同的优先事项，组合、发展水平和标准也不同，因此需要采用不同的评估方法。因此，参照有一整套 66 个指标的矩阵评估了 24 个中央职能在构成部分方面的业绩，而这些指标的基准对照物则是经联合国评价小组、联检

¹⁵ 为加强严谨性，研究报告采用了适用于参数和非参数统计的多重统计分析，包括方差分析、Kruskal 及 Wallis 式单向差异分析和一些“发生于其后者必然是其结果”的检验。应承认，析因、多重回归法和其他高级统计本可增强本研究报告的解释性权威。本研究报告如在工作 and 资源方面有更详尽的范围则本可采用这类方法。

组和发展伙伴认可的既定标准。根据这些指标对中央层级所作的评估结果分成 5 个业绩挡(较低、低于一般、一般、高于一般和较高)。下文第 53 段详细说明了评定这些业绩挡所使用的标准尺度。

32. 对于分散评价职能来说,由于缺乏一个组织为支持建立一个高效率和高实效的评价系统所需到位的关于政策和系统的商定标准,评价工作受到了质疑。因此,在开展研究工作时首先进行了磋商,然后确立了适用于分散评价的概念框架。接着,利用这个框架对整个联合国系统分散评价的状况作了一次快速评估。向表明自己有明确分散职能的各个组织和被确认对分散职能承担监督责任的官员分发了调查表。¹⁶ 作为补充,还与工作人员和利益攸关方进行了广泛磋商,并作了文件分析。

33. 对于没有中央职能的 4 个组织,分发了一个关于关切问题的调查表,以评估在发展整体评价职能方面的当前状况、问题、系统性制约因素和打算。3 个组织完成了该调查表。在联检组网站可看到本研究报告所采用的关于所有调查表的信息和成熟度矩阵所使用的方法。

在评估和在全系统分析方面的挑战

34. 由于各组织的任务规定和结构不同、数据无可比性,以及各组织在提及评价及其职能时采用了意义不同的概念,因此,研究工作在全系统衡量和分析时遇到了挑战。为了应对这些挑战,采用了不同方法。为概念或观念确定了一个定义,并在整个研究报告中始终采用该定义。在评估中使用了标准化和参照标准的方法。指出了不同类型组织之间的差异并提供了可能的解释。采用了权重方法并对缺失的信息作了数据校准。在出现巨大的数据空白或完全不可比的情况时,在作交叉比较时放弃使用推论性统计方法。在这类情况下,提供了说明性分析。本研究报告特别关注对小型组织评估结果的有效性,而作出使用成熟度矩阵的决定的基础是磋商和对其在评估小型组织评价职能方面的有效性和价值的分析。对小型组织的数据校准未对调查结果产生明显作用。

限制

- 就中央职能而言,成熟度矩阵的发展和使用时是同时进行的,但最后 2013 年的所有数据集都经过订正和更新以求得一致性。¹⁷
- 当时计划访谈各种利益攸关方,但由于时间和预算限制及在及时获得所需要的准入方面的困难,并没有完成所有访谈。

¹⁶ 被确认具有这类监督责任的官员各不相同。在一些组织里,监督由中央评价办公室提供。在另外一些组织里,监督由一个中央支持部门或主管成果管理制的部门提供。

¹⁷ 对联合国人口基金(人口基金)的最后评估是个例外,因为这次评估是在 2014 年 2 月进行的。评价职能在 2013 年期间正在进行重组,因此需要在 2014 年初最后确定评估结果。此外,人居署的数据未以访谈和信息更新的方式经过验证。

- 在评估和分析中未列入以下基准：因无法找到一个以上可靠来源而被认为证据不够健全的基准和被认为不适用于各组织不同任务规定的基准。
- 为迅速评审分散评价，相关分析的基础是取自由承担监督责任的官员填写的调查表的数据及来自访谈和文件分析的数据。¹⁸ 本来应该在 940 多个负责管理和开展分散评价的部门/办事处及从进行所谓自我评价的众多工作人员(如项目管理人员)那里作更为深入的数据收集工作。可是没有足够的时间或资源来做这种严格的数据收集工作。但是，来自监督官员填写的调查表、访谈和文件分析的信息为快速评审提供了一个适当基础。联检组计划于 2015 年在联合国系统里对成果管理制进行的系列评审可能可以更全面地审查分散评价系统。
- 本研究报告对评价职能在 36 年期间的演变所作的分析综合了来自各种报告的不同数据集。因此，相关信息不具绝对性，仅可用于说明问题。这项分析以各种磋商和在过去至少 15 年里参与评价职能的各个利益攸关方的验证作为补充。
- 鉴于在能力方面的限制，研究报告未使用本可加强其解释权威性的先进统计方法。但在一些特选的分析中确实应用了足够数量的描述性和推论性统计方法。¹⁹
- 研究报告确认了在评价职能和质量标准一些选定层面上的业绩最佳者。对何种力量驱动导致业绩优良作了有限度的解释。要作出更为广泛的结论则必须有更多的数据和分析。但是，研究报告还是列入了对假设性关系的许多分析，以支持所作的推测。

D. 鸣谢

35. 联检组检查专员谨感谢所有提供指导、实质性意见和有用投入的人。研究报告的范围、方法、工具和分析受益于与联合国评价小组的磋商，他们还提供了大量的信息以供评估。研究报告在概念化、验证和咨询方面还受益于一位国际专家和一个非正式的公益性咨询机构，其成员包括：以评价专家身份提供帮助的联检组前执行秘书、联合国各组织的管理人员和工作人员、来自大不列颠及北爱尔兰联合王国国际发展部的一位评价主管、美利坚合众国霍华德大学机构评估和评价办公室的一位代表，以及一位区域发展问题专家。在项目早期阶段与下列部门进行的磋商对于制定基准非常重要：监督厅检查和评价司、世界

¹⁸ 这些数据仅限于各基金、方案、专门机构和联合国其他实体，不涉及联合国秘书处有嵌入职能的那些部厅。

¹⁹ 所采用的分析方法包括斯皮尔曼的 rho 和皮尔森的相关性检验、t 检验、方差分析法、Kruskal 及 Wallis 式分析法和许多重大差异的“发生于其后者必然是其结果”的检验。具有描述性数据的交叉表提供了具体的信息。

银行、发展和援助委员会及经济合作与发展组织发展合作局(经合组织发展局)。最后但并非最不重要的是，全体检查专员希望感谢联检组各位协调人提供了广泛的协调支助。

二. 中央评价职能：演变、成长和发展水平

A. 发展中的变化：目的、结构安排、覆盖范围和管理

结论 1：自联检组发布 1977 年至 1990 年的报告以来，中央评价职能在系统运作方面有了实质性演变和成熟，其作用、结构和标准也发展了。作用多样化了：中央职能一直在继续支持分散评价职能，同时还支持广泛和战略性的组织/中央级决策。为应对不同要求，其作用也有所增加。各组织在发展评价系统时一直比较务实并具有成本意识，在作结构安排和在设计用于管理和进行评价的业务模式方面也都是这样做的。但很明显，以实际价值计算，提供给中央职能的资源没有什么变化。联合国系统必须反思：是否真的想知道自己正在产生何种影响力？如何才能加强自己在国家、区域和全球一级的相关性？

不断变化的作用

36. 在 1977 年时，联合国系统有一个以学习和改善的必要性为重点的自我评价系统，²⁰ 而中央评价办公室的首要作用是“中央监督和协调，以确保自我评价这一主要系统的质量和业绩”；²¹ 自那时以来，联合国系统内的评价职能已有实质性发展。²² 到展开本次研究时，在经本研究报告评审的联检组 24 个参与组织中的中央评价办公室的作用已多样化了，其重点是支持作出战略和组织一级的决策。许多组织继续以平行方式开展自我评价或分散评价及其他各种与任务相关的分析和评价以应对各种要求。下文表 1 说明了自 1977 年以来的这种进展情况。

结构安排

37. 评价部门在各组织结构中的地位已变得更加稳定。在最初几年，它们附属于：……“各组织内的最高首长；被纳入规划和方案编制活动；或与行政财务部门合为一处”。²³ 中央评价部门被称为“是以一种相当不稳定和无保障的方式嫁接到现有结构上，尚未真正成为组织管理流程的一个既定组成部分”。²⁴ 突出的

²⁰ 这反映在当时对评价所下的定义内，即作为一项学习和以行动为导向的管理工具(A/34/286, 1979 年 6 月 6 日)。

²¹ “关于联合国系统内评价工作的第二次报告”，JIU/REP/81/6，第 4 页。

²² 见联检组 1977 年至 1985 年期间关于评价职能的系统性全面研究报告。

²³ 同脚注 21。

²⁴ 同脚注 21。

问题是工作人员人数太少：在此期间有 60 个专业人员在评价系统工作，在专业人员总数(23 个组织共 18 500 人)中所占比例不到 1%(0.7%)，而可利用的资源又极其有限，与各组织项目和方案成果的数目和范围不相称。²⁵

38. 在作本次研究时，在联检组 28 个参与组织中有 24 个组织有明确的政策、明确规定的治理和结构安排，并有用于提高评价效率和实效的明确的制度和业务模式。它们还有用于提高评价职能独立性、覆盖面、公信力和效用的广泛机制。这些机制的发展水平是本研究报告的重点。

39. 检查专员们认为，各组织在发展评价系统时一直比较务实并具有成本意识，各组织在作结构安排和在设计能在进行评价时有效利用工作人员和咨询人的相对优势的业务模式方面也都是这样做的。但检查专员们不清楚，对评价的承诺和供货水平是否发生了很大程度的变化。下文各段落和表格述及这些问题。

表 1
评价职能各阶段和中央评价职能不断变化的作用

阶段	第一阶段 (1977-1980 年)	第二阶段 (1980 年代)	第三阶段 (2000-2005 年)	第四阶段 (2006-2013 年)	第五阶段 (2013 年至今)
目标和重点	学习 供在项目中使用的评价	学习 在项目中使用时，开始由中央部门提高质量	问责制 重视独立性并改善方案和项目以实现目标	平衡 问责制 独立性 质量 学习 一致性 国家评价和其他要求	新出现的趋势 平衡不同的要求及新出现的变化和挑战
评价系统的主要特点	嵌入项目执行工作的广泛项目评价(后来自我评价被称为分散评价)	嵌入项目执行工作的广泛项目评价	确立整体中央评价：重视制定用于中央和分散职能的制度	在中央一级的系统发展以促进独立性、公信力和效用(以联合国评价小组的规范和标准为基础)	在明确界定的战略框架范围内的中央和分散评价系统的战略、日常化和互补作用，以支持组织改革和转型
		中央评价部门协调和提高分散评价的质量	有两个主要模式的分散评价：(一) 具有体制框架的分散系统；(二) 由工作人员和管理人员自行酌定的临时系统	有两个主要模式的分散评价：(一) 具有体制框架的分散系统；(二) 由工作人员和管理人员自行酌定的临时系统	协调和协作以评价全系统价值 协调以促进国家能力发展和全球伙伴关系

²⁵ 同脚注 21，第 18 段。

40. 因此，对于设有用于支持中央级决策的中央评价职能的 24 个组织而言，对中央职能所处组织地位作出的选择是不同的：10 个组织有非常明确的独立评价职能；14 个组织的评价职能属于合署办公型，其中 9 个组织的评价职能在监督办公室负责人的领导下与范围更为广泛的监督职能合署办公(或者作为一个有自己主管的单独单位或者与其他监督职能完全整合在一起)；另外 5 个组织的评价职能与行政管理部门合署办公(以及与政策、规划、管理或研究部门整合在一起)(见下文表 2)。

41. 据称，地位的选择受许多因素的影响，以不同方式与在独立性、效率、相关性和效用方面的假设收益联系在一起。在 7 个大型组织(以预算款额来定，年度组织支出超过 10 亿美元)中的 4 个组织里，就此作出的选择倾向于创立独立的单独常设部门，这部分是因为要满足联合国评价小组的规范和标准及执行局的要求。与此相对照，在小型组织(年度组织支出低于 3 亿美元)里就此作出的选择主要是与其他职能合署办公，反映出对成本和效率或对以下优先次序的考虑：(a) 建立独立性(通过使评价职能与其他独立监督职能合署办公)，或(b) 管理层所认为的相关性和赋予评价职能对组织文化的直接影响力(通过使评价职能与行政管理部门合署办公)。研究报告就不同结构地位对评价职能的发展所产生的影响提出了调查结果。

42. 在 28 个组织中，有 4 个组织没有中央评价职能：国际电信联盟(国际电联)、世界旅游组织(旅游组织)、万国邮政联盟(万国邮联)和联合国项目事务厅(项目厅)。它们为这种情况提出了各种理由，其中突出的理由是：(a) 评价不是高优先事项，因为理事机构对评价没有要求，为此相关资金有限；(b) 有一个更倾向于交付活动而不是考虑所取得的成果的组织文化；以及(c) 本组织的总体预算有限。²⁶ 据报告还存在一些顾虑：鉴于组织具有的特定类型，包括其能力和文化背景，评价是不可行或不会增加价值的。

43. 检查专员们认为，这些解释无法充分证明没有评价职能是有道理的。这些组织需要重新审视评价职能在加强其工作的实效、效率和可持续性方面的作用。在完成本报告时，国际电联和旅游组织告知，它们现在已开始制定发展评价职能的计划。检查专员们建议，鉴于目前的情况，当前在反思小型组织评价职能的机构框架和政策的发展问题之前应有一个深思熟虑的愿景和战略，包括其增加值、优先事项、协同作用和与组织内现有的其他监督和知识发展系统的协调，以及与联合国其他机构和外部伙伴的协调。旅游组织利用项目厅道德操守干事的服务而不是建立自己的办公室的做法就是一个高效率交付服务的实际例子。

²⁶ 在这 4 个组织中，3 个组织的年度组织支出总额为每年大约 1.96 亿美元。28 个参与组织的预算范围在每年 2 700 万美元至 53 亿美元之间。

表 2
中央评价职能的结构：联检组 24 个参与组织的规模和结构安排

结构安排	按年度预算分类的组织规模(美元)			
	大型 10 亿至 55 亿美元	中型 3 亿至 10 亿美元	小型 少于 3 亿美元	
A. 单立 N = 11	A.1: 职能作为一个完全单独的评价办公室运作	开发署 粮食署 粮农组织	劳工组织 人口基金	工发组织
	A.2: 职能作为一个单独办公室但在行政首长办公室内运作	儿基会	环境署	人居署 妇女署 毒品和犯罪问题办公室
	B.1: 职能设在独立的内部监督办公室内 职能在以外处长或司长领导的单独部门里运作，与其他监督职能平行运作 3 个组织	联合国秘书处/监督厅	教科文组织 近东救济工程处	
B. 合署办公 N = 13	职能与其他监督职能整合在一起 6 个组织	世卫组织 ²⁷	原子能机构 知识产权组织	国际民航组织 海事组织 气象组织
	B.2: 与行政管理(政策、规划、研究)部门合署办公 职能完全与政策、规划、管理或研究部门整合在一起： 4 个组织	难民署	艾滋病署	贸发会议 国贸中心

核心职能——编写报告以支持决策

44. 中央评价职能的重点作用是开展评价和编写评价报告，其核心职能就是支持就政策、方案、项目和业务作出决策，同时执行与评价相关的各种各样的任务以支持该职能的制度化。在 2012-2013 年期间，24 个组织的中央评价办公室编写了 584 份不同类型的评价报告，其中：26% 内容复杂，涉及专题、国家、区域和全球方案；4% 是联合评价；64% 是对小型和大型项目的评价。后面的项目评价主要是小型组织和那些以执行规范性任务为主的组织做的。一般来说，就 7 个大型组织和那些具有明确的分散结构安排的组织而言，²⁸ 项目评价主要是在分散评价

²⁷ 这里用于分析的数据源自 2013 年。最近的一些结构变化包括在世卫组织里的变化；该组织在 2014 年 8 月不再让内部监督事务办公室负责评价职能，使该职能独立出来作为一个新职能——“评价和组织学习”——的组成部分。

²⁸ 教科文组织、粮食署、妇女署、世卫组织、人口基金、儿基会和开发署。

一级完成的，每年至少要编写 636 份评价报告。²⁹ 鉴于报告数目众多，在中央和分散一级编写的报告构成了联合国系统的一个大型知识库。³⁰

45. 但是，联合国系统的这一知识资产分散在联合国系统各组织的办公室和网站上。经过多年辩论，联合国评价小组于 2014 年 9 月为其成员和中央办公室的所有评价报告设立了一个公共数据存放库。存放库保存了可追溯至 1977 年的 4 800 多份报告，到目前为止是由联合国评价小组的 11 个成员提供的。³¹ 这一数据库可大大加强知识共享，是为促进在全系统范围利用交叉分析结果和吸取经验教训所作出的一项重大努力。最近在对联合国系统在发展方面工作所面临的变化和挑战作了一次调查后建议，鉴于“联合国的因特网内容对于局外人来说极难查找，索引作得也很差”，³² 因此应为联合国的知识资产和出版物建立一个单一的网关。联合国评价小组的这项努力就是应这一要求所采取的一个步骤。

46. 在大多数情况下，对评价对象所作选择的基础是一系列明确界定的战略标准。据报告，评价的覆盖面不足以应对来自组织内部的要求。³³ 据说，造成覆盖面不足的主要制约因素是资源不足。在目前的情况下，有必要优化对现有资源的利用，以提高评价质量，限制公信力受损的风险。

商业模式

47. 在开展这些评价时，24 个中央部门采用了各种不同的业务模式。在早些年间，中央职能委托外部独立咨询人进行了大量评价，³⁴ 但在作本次评审时，中央职能全局掌控，并对评价报告负全部责任。不过有少数例外。采用的一个主要模式是，工作人员在管理和确保遵守质量标准及评价的机构准确性和价值方面发挥领导作用，³⁵ 同时与专家咨询人合作开展评价工作。各组织在管理评价工作时所

²⁹ 这个证据来自就这 7 个组织的分散评价所提供的数据库。

³⁰ 如果加上联合国秘书处有嵌入职能的部厅的报告，这个数字就会大幅度增加。

³¹ 见 www.uneval.org/evaluation/reports。用户可按机构、评价类型、国家、区域或关键词查询。

³² Barbara Crossette, “联合国系统还有可能改善吗？全球专家对一项独立调查作出回答”，2014 年 5 月 26 日。可查阅 <http://passblue.com/2014/05/26/is-a-better-un-system-possible-global-experts-respond-to-an-independent-survey/>。

³³ 各组织在选择将要开展的评价时采用了各种标准，其中主要的标准是：(a) 与组织战略的关联(20 个组织)；(b) 管理层的内部要求(20 个组织)；(c) 捐助方的外部要求，通常有协议规定(17 个组织)；(d) 供资数额(14 个组织)；以及(e) 相关活动的可评价性(12 个组织)。采用的其他标准包括：风险评估(10)、对新出现的全球趋势的考虑(9)、与联合国系统其他组织计划的一致性(8)，以及根据评价结果可能会扩大活动规模(4)。

³⁴ 委托进行的评价在订约负责开展和报告评价结果的外部机构的控制下。以往这种形式盛行，因为一些评价职能缺乏独立性。

³⁵ 这包括拟定职权范围、与咨询人签订合同、管理合同、在适当顾及组织和体制结构及评价部门和联合国评价小组的标准和准则的情况下管理评价的实施，以及质量保证。

采用的业务模式不同，因为必须提高资源的利用效率和改善质量，同时还要确保公正性(见附件六)。

工作人员配置和咨询人

48. 中央职能在 2012-2013 年期间展开的 584 次评价是由在中央部门的 160 位全职专业工作人员和管理人员管理和进行的，他们占相关组织工作人员总数的大约 0.18%。³⁶ 在进行这些评价时利用了 989 位专家咨询人的服务，估计他们提供了相当于 32 个全时职位。因此，所有工作人员和咨询人(大约相当于 192 个全时职位)³⁷ 加在一起在组织工作人员总数中所占比例仍不到 0.3%，比 1980 年代所报告的 0.7% 低很多。³⁸ 在整个联合国系统，平均而言，每个工作人员(加上咨询人)每两年编写大约 3 份不同类型的评价报告。这些数字是根据现有信息估算得出的，这类信息并不总是稳定不变的，因此在解读这些信息时需谨慎。鉴于零星的证据表明，评价工作人员实际上每天工作 8 个多小时，因此在解读这些数字时必须谨慎。

49. 工作人员平均把他们 24% 的时间用于其他与评价相关或其他监督、研究和战略管理活动及向分散评价提供评价工作以外的支助。附件七提供了关于工作人员在应对要求方面所承担的各种任务的信息。各组织在关于非评价报告活动方面的工作类型和范围也是不同的。该附件中的信息是很重要的，因为显示出对中央评价职能的影响和效率及对该职能供资情况的分析必须考虑到这类除了编写评价报告以外用于加强各组织内评价文化发展的活动。

50. 虽然在一些领域里取得了进展，但在另一些领域里的进展则不均衡，需要做更多的工作来发展和持续推动评价职能并提高其地位，同时要适当考虑评价职能在加强联合国系统实效和可持续性方面可发挥的作用。下文各节着重说明评价职能的发展水平和改进建议。在这方面至关重要是要采取更综合性的方法，在运作中要采用深思熟虑的战略方法，平衡优先事项和为评价职能定位，以提高其在支持目前处于“发展十字路口”的联合国系统方面的增加值和作用。³⁹

B. 组织一级的业绩：各组织的模式和各组织之间的差异

结论 2：联合国系统内所有组织的中央评价职能在达到为高影响力评价职能核心构成部分设立的标准方面的总体发展水平上取得了相当大的进展(即，体制框架、对于新出现的挑战的相关性和准备状态、独立性、质量和效用)。但是，各

³⁶ 分析是以 17 个参与组织现有的全部数据为基础的。2012-2013 年用于 160 名工作人员的人事费总数为大约 534 711 127 美元。用于 989 名专家咨询人的支出总额为 21 856 019 美元。

³⁷ 相当于全时职位的工作人员和咨询人人数在整个组织工作人员中所占比例不到 0.3%。现在没有可用于评估这个比例的基准。但正如上文第 36 段所示，这一比例比 1981 年的 0.7% 的比例要低。

³⁸ 这是根据现有数字得出的估计数。计算的基础是 2012 年 52 915 名工作人员这一数据(不包括秘书处各部厅，因为本研究报告未包括这些部厅)。

³⁹ Bruce Jenks 和 Bruce Jones，“联合国发展处于十字路口”(2013 年，纽约大学国际合作中心)。

组织之间在这些构成部分的发展水平和性质上存在显著差异。根据不同构成部分的发展水平和速度，可将这些组织分成 4 个不同成熟度或发展水平的群组。

51. 对联合国系统内的中央评价职能作追溯到 1977 年的历史分析后可以看到在系统业务发展成熟度方面的进展：(a) 评价职能系统是否得到发展但仍然初级水平；(b) 得到发展但属于临时性质或被整合了；或(c) 它们是否被完善整合并提升到更高的系统业务水平，以及是否在加强评价职能的质量、相关性和影响力方面有一个范围更为广泛的战略性外向视野。

52. 如附件八和来自参与发展中央职能的联合国系统官员的传闻证据所示，数据趋势表明，发展的速度和规模有过起伏，该职能的发展也有停滞不前的时候。中央职能的发展似乎在 2000 年代中期加快了，原因是：(a) 大力推动成果问责制和联合国系统提出的价值观或影响力；(b) 联检组评审了治理和监督情况并提出了设立评价职能的建议，不管组织规模小到什么程度；以及(c) 由联合国评价小组牵头在联合国系统内发展了评价专业并在 2005 年颁布了规范和标准。⁴⁰ 这些规范和标准为评价职能的发展和对该职能的监督提供了一个系统性体制框架。它们还为用于指导在本次研究中全面分析评价职能所采用的标准提供了投入。⁴¹

53. 就当前的发展水平而言，对各构成部分的评估是以下文表 3 提供的进展分级表为基础的，按在执行该职能时在 5 个构成部分方面的 5 个可能的成熟度进行评估。联检组网站辅助材料里的联检组成熟度矩阵载有关于评估方法的详细资料。

表 3

中央评价职能的发展水平(成熟度)

成熟度或发展水平	得分范围	在达到既定基准方面的模式说明*
1 级 初级 很低	1-2	表明系统各构成部分不存在或处于初级状态：相关组织至多正在寻找设立一个职能的方法和预期成果、制度、机制和所需要的管理结构。
2 级 临时性的 低	3-4	评价系统是临时性的：存在该职能各构成部分的某些要素，但这些要素不完全一致或得到一个明确界定的体制框架的支持。
3 级 系统化和常规化 一般*	5-6	该职能得到较明确界定：各构成部分的关键措施和机制已到位，运作不再是临时性的，而已常规化并有一定程度的稳定性。重点是加强各要素的整合、质量和制度化。因此，该职能的定向主要着眼于内部。

⁴⁰ 见 www.uneval.org。

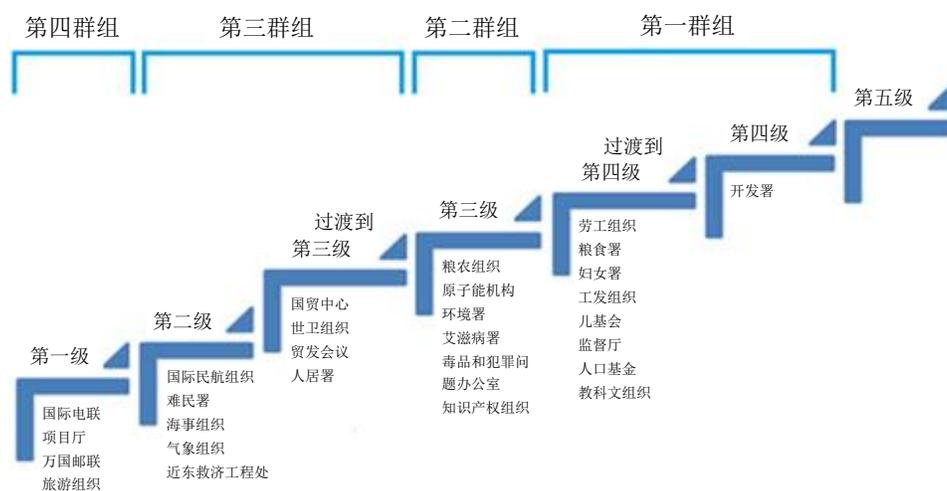
⁴¹ 采用的其他标准包括在评估联合国系统各组织时为多边组织制定基准框架所用的那些标准。

成熟度或发展水平	得分范围	在达到既定基准方面的模式说明*
4 级 完善和加强使用 高	7-8	该职能构成部分各要素的整合和制度化大部分已经实现，改善的重点已转向为相关组织和整个联合国系统加强增加值效应。更加重视伙伴关系和用于加强实效、效率、可持续性和影响力的替代办法。
5 级 更新 很高	9-10	这一阶段的重点是使该职能成为促进组织、联合国系统、国家、区域或全球一级变革的主要驱动力。还有一个重点就是促进催生尖端的评价方法和以综合方式进一步提高各级评价工作的价值(在组织、联合国系统、全球和国家各级)。

54. 下文图二显示联检组 28 个参与组织如何根据各自在评价职能所有构成部分的评价职能“总体发展”分成 4 个不同群组的情况。

图二

联检组 28 个参与组织中央评价职能的总体成熟度



55. 在总结每一个群组在 2013 年的组织状况时发现，其主要要点和特点如下：

- 第一个群组有 9 个组织，在这些组织里用于支持中央评价职能的大多数预期制度和政策已经到位并高度制度化了。经评估，这些组织在达到为评价职能构成部分设定的既定基准方面平均有 65% 的成功度。它们面临的主要挑战是如何加强评价产品的使用和影响力及如何平衡不断增加的要求，包括通过在解决新出现的组织和全球挑战方面发挥关键作用。

该群组的组织主要是基金和方案(5个),有单独的中央职能(7个),以及是大型或中型组织(6个)。值得指出的是,这些基金和方案的发展水平符合预期,因为它们长期与捐助方合作从事发展工作,而捐助方会要求对获得资助的项目进行评价。但必须指出,统计分析表明,基金和方案与专门机构之间在发展水平方面无显著差别。本群组各组织把大量资源用于评价,有许多较高职等的专业工作人员。

- 第二个群组有6个组织,评价职能的发展水平一般。用于支持中央评价职能的大量预期制度和政策已经到位,但不是所有制度和政策都已全面运作了。因此,中央职能的主要重点仍然是使它们投入运作以保持稳定发展,而不是注重其利用或影响力、或确保有一个外向型重点以提高评价的增加值。总的来说,本群组中各组织在本次对所有5个构成部分的研究中都达到了56%预期基准的一般业绩水平。本群组同样向评价工作投入了大量资源。应当指出,与第1个群组相比,本群组利用了大量咨询人来执行评价工作。
- 第三群组有9个组织,预期制度和政策尚未全部到位,那些已到位的也未充分或以综合方式运作起来。本群组各组织的业绩低于一般成熟度。本群组的主流是:(a)与审计部门、监督部门或与涉及政策和规划的管理部门合署办公的小型组织(组织预算低于3亿美元)和(b)与其他部门合署办公的大型或中型组织,面临同小型组织一样的挑战。本群组平均达到40%或低于一般的业绩水平,在发展评价职能方面还有很长的路要走。相比较而言,本群组中各组织没有大量高职等(P4-D2)专业人员来执行评价任务。
- 第四群组包括万国邮联、旅游组织、国际电联和项目厅。尚未建立组织级中央评价职能,虽然存在某些初级评价形式。如上文第43段所述,国际电联和旅游组织正在制定程序,以发展一个明确界定的职能。

欲知各组织之间差异的更多详情,请参阅联检组网站上的辅助材料。

结论 3: 证据表明,与监督部门合署办公或整合进行政管理部门的评价职能在其发展水平上通常是滞后的。在小型组织中这种情况尤其普遍,因为相关组织中的资源制约是使评价职能合署办公的一个决定因素。但在合署办公和规模制约方面都有例外。成功标准包括是否有一个能理解评价相对于监督职能的战略性和增加值的治理和管理结构、是否提供了足够的资源及评价质量是否得到保障。

56. 对24个具有中央评价职能的组织所作的统计分析表明,与合署办公的职能相比,单独的职能发展水平更高。大型和中型组织的发展水平比小型组织要高。因此,在观察业绩水平一般或偏低的第三群组的9个组织时可以看到,所有9个组织的评价职能都是合署办公的:5个与管理部門、4个与其他监督部门合署办公。6个是小型组织(年度预算低于每年3亿美元)。下文表4显示了这种相互作用的效应。

表 4

按各组织年度预算总额规模和中央评价职能所处地位排列的评价职能发展水平

	小型		中型		大型	
	组织	得分	组织	得分	组织	得分
单独	毒品和犯罪问题办公室	5.2	劳工组织	6.7	粮食署	6.7
	工发组织	6.4	环境署	5.8	开发署	7.1
	妇女署	6.4	人口基金	6	儿基会	6.3
	人居署	4.2			粮农组织	5.9
与管理部門合署办公	国际电联	4.9	艾滋病署	5.3	难民署	3.8
	贸发会议	4.4				
与监督/审计部門合署办公	海事组织	3.7	教科文组织	6.1	监督厅	6.2
	气象组织	3.5	近东救济工程处	3.4	世卫组织	4.7
	国际民航组织	3.9	知识产权组织	5.2		
			原子能机构	5.9		

3.0-4.9 = 低于一般和较低	5.0-6.0 = 一般	6.1-8.0 = 高于一般和较高
-------------------	--------------	-------------------

57. 但必须指出，有一些优秀业绩实例表明，这两个因素，不管是规模还是合署办公，都没能阻止高级别职能的发展，这两种例外情况都有。当小型组织向评价职能投资(即，有良好的预算支持、人数相当多的高职等专业人员级评价工作人员及对评价职能的作用、增加值和对相关组织的支持作用有清楚认识)时，如第一和第二群组里的联合国促进性别平等和增强妇女权能署(妇女署)、联合国工业发展组织(工发组织)和联合国毒品和犯罪问题办公室(毒品和犯罪问题办公室)那样，其发展就会超过一般水平。同样，像监督厅和联合国教育、科学及文化组织(教科文组织)这样的大型和中型组织或实体，虽然评价职能是合署办公的，但运作时却独立于与其合署办公的监督职能；这些组织有数量合理的工作人员或资源，业绩高于一般水平；这再次说明，单是合署办公并不是造成业绩较低的决定因素。

58. 对解释性因素所作的分析表明，问题在于，一些组织内在监督办公室合署办公的评价职能被认为是丧失了评价职能的特性。相关分析结果表明，这种看法起到了限制评价职能增加值的作用，包括其在应对变化及加强评价文化和学习型组织方面的复原力和灵活性。同样，在某些小型和/或合署办公职能的情况下，质量保证和质量提高机制(例如对报告质量的同行评审和评估)较弱，这可能反映出对评价职能的投资力度较低。评价职能及其增加值的可见性同样受到限制，或可能未得到充分保障，因为这些组织一般不向其立法机构提交、也不讨论关于评价工作的全面年度报告。统计分析表明，与立法机构的这种讨论是在确保评价职能的可见性和发展方面最为重要的一项活动。在大多数情况下，关于评价工作的信息被嵌入在关于其他监督活动信息的年度报告里。在评价职能以合署办公方式运作的组织里要加强评价职能的能力和可见性时应考虑到这些因素。

59. 与预期相反，分析结果表明，与其他管理职能(例如，战略规划、研究和一般管理)整合在一起的评价职能并没有提高评价工作的利用率或建立成更好的学习型组织。在假设中往往会觉得这两个特征能给这类合署办公的职能提供增加值。但是，这种合署办公的组织在发展评价文化或学习型组织方面不一定就能在组织内实现明显更高的吸引力或重大收益。对这类职能的优势和弱点作分析后并没有看出任何一致性或优先次序模式，表明对评价职能采用了临时性方法。

60. 执行以下建议将会提高评价职能的实效。

建议 1

中央评价职能是合署办公和与其他监督职能整合或与行政管理职能整合在一起的联合国系统各组织的行政首长应加强该职能并确保其质量、完整性、可见性和增加值。

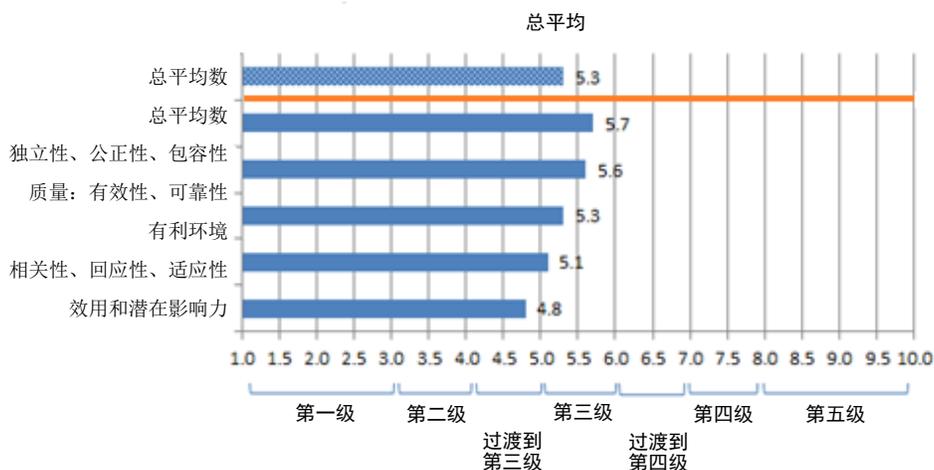
C. 职能的构成部分：优势、挑战和系统性制约因素

结论 4：24 个组织在评价职能 4 个构成部分的发展水平上相对处于同样水平，在成熟度评估表上都被评为“一般”。但在第 5 个构成部分(评价职能以利用评价为重点的效用)上情况则不同。对该构成部分而言，发展水平明显落后，在评估表上低于一般水平。总体而言，在每一个构成部分内都看到了优势和弱点。优势和弱点模式往往反映出各组织可能共同存在的系统内的不平衡、系统性制约及在价值体系或风险因素方面的差异。

下文图三反映了上述结论，并显示出评价职能 5 个构成部分的发展水平。

图三

5 个评价构成部分每一个部分的发展水平和平均得分



61. 根据对 5 个构成部分中每一个构成部分的优势和弱点所作调查分析的结果，对评价职能每个构成部分提出了如下意见和结论。

构成部分 1：有利环境、体制框架和支助系统

62. 联检组早先的一份报告已经指出，“……没有来自理事和立法机构及来自整个组织管理层和工作人员的要求和坚定支持，即使用意再好、设计得再精心的评价系统也无法成功。而没有一个明确界定的政策、规范、游戏规则和标准体制框架，评价系统也无法存活”。⁴²

63. 表 5 显示根据对本构成部分各项指标评估后得出的优势和弱点。下文概述了对本构成部分一些主要成果模式的分析和结论。

表 5

有利环境——优势和挑战

已取得进展的领域	进展不足的领域
<ul style="list-style-type: none"> • 理事和立法机构关于评价工作的任务规定/决议 • 有明确界定的问责制和责任的治理结构 • 成员国/立法机构的作用(推动问责制) • 涵盖联合国评价小组规范和标准的评价政策 • 定期审查政策执行情况和修订政策 • 用非核心资源来支持交付方案 • 联合国评价小组的规范和标准及其应用和适应 	<ul style="list-style-type: none"> • 资源充足性 • 明确界定的规范性框架以指导资源分配 • 高级管理层的领导作用和评价的价值及在组织里的促进情况 • 促进成果、问责制、评价和学习的组织文化在组织里充分扎根 • 基于组织的愿景和战略 • 与其他知识、监督/审计和检查，以及决策支助系统相协调——战略协调至关重要

成员国的作用、任务规定和政策

结论 5：持续不断的强烈要求似乎是使评价系统和政策制度化的组织越来越多的主要原因。这种持续的要求主要来自理事机构。但是，这导致评价职能主要注重满足问责制要求。根据联合国评价小组的规范和标准确立评价政策的做法同样加强了发展。这些规范和标准也有一个主要由问责制驱动的结构。

⁴² 内部评价在联合国系统各组织内的地位，JIU/REP/85/10。

64. 大会决议及组织任务规定和政策在发展评价的体制框架方面发挥了关键作用。从 1980 年代以来已经通过了大量决议。在 2000 年至 2013 年期间，就评价职能通过了 3 项决议。⁴³ 成员国在指导拟订评价政策和执行政策方面发挥了重要作用。分析结果表明，成员国不那么积极参与的那些组织的评价职能发展水平较低。但与评价职能的其他目标相比，成员国关注得更多的是成果问责制和资源的利用情况。这种对问责制的关注对于满足在援助实效和最大限度地提高援助的效果方面的要求而言非常重要，但需要作更好地平衡，以加深了解能加强发展实效的评价职能的关键特征(另见下节——结论 6)。

65. 在 2005 年开始采用的联合国评价小组规范和标准明显影响了评价职能制度化的进程，并使大多数组织对其政策作了相应调整。联合国系统各组织中有三分之二的组织有一个力求符合联合国评价小组规范和标准的评价政策，其余组织则在一个通用的组织问责制框架下运作。评价政策在界定评价职能的框架(做什么、作用、责任和问责制、评价类型和评价结构)方面发挥了重大作用。应当指出，联合国评价小组规范和标准同样有一个主要由问责制驱动的结构。

66. 评价政策的确立也在评价职能的发展方面发挥了关键作用。对现行评价政策的内容作分析后发现，各组织在这类政策中广泛采纳了联合国评价小组的规范和标准。这并不奇怪，因为所有 24 个有中央评价部门的组织都是联合国评价小组的成员。尽管如此，但发现，这些组织中只有 46% 的组织有一个详细说明如何采用联合国评价小组规范和标准的政策并采纳其他适合本组织情况的规范。但是，没有充分立足于组织优先事项和相关组织目标所要解决的那类挑战的政策就很可能缺乏相关性。除 3 个组织(国际劳工组织(劳工组织)、教科文组织和妇女署)外，绝大多数组织都缺乏关于评价职能的组织战略，也未利用这种战略来指导政策制定工作。建议联合国系统各组织的行政首长确保制定一个立足于本组织的明确界定、一致和全面的战略及在当前情况下进行改革的计划。这包括小型组织，因为对小型组织而言，在各组织内和组织之间的协调对于在提高效率的同时还能保持评价职能的专业完整性及其对本组织的增加值而言至关重要。

学习文化和领导力

结论 6: 为回应理事机构的要求，联合国系统大多数组织对评价工作的重视主要体现在把评价工作作为支持组织问责制的一个有效机制。相比之下，没有同等重视确保把评价职能纳入本组织的结构，确保让评价职能在组织发展中发挥作用，以及确保利用评价来给本组织作战略定位以应对不断变化的环境。解决评价的这方面问题对于评价职能的可持续性至关重要，特别是鉴于变化的速度很快、要求层出不穷、挑战异常复杂。如果评价职能的发展更倾向于问责制、偏离学习型组织的发展，那么就会限制评价职能的价值和可持续性。

⁴³ A/58/587、A/RES/59/250、A/59/488Add.1、A/62/424/Add.2、A/RES/62/208/19、A/RES/67/226、A/RES/67/226(27of30)、A/RES/67/226 (28of30)。

67. 虽然由问责制驱动的文化仍然是重要的，但是不断变化的全球动态发展和全球格局新出现的需求，包括正在举行的关于 2015 年后发展议程的辩论所提出的新要求，都呼吁要加强对学习及发展学习和评价文化(即学习型组织)的重视。

68. 上文第 9 段对学习型组织的定义及在磋商中提出的意见为拟订成熟度矩阵中的规则(即，标准和发展水平)提供了指导。附件九列出了学习型组织和评价文化的指标，并汇集了相关信息。联检组网站辅助材料里联检组中央职能成熟度矩阵载有在评估中所使用的规则或业绩标准资料。

69. 对接受评审的学习型组织的评估：(a) 成果、学习和问责制文化在本组织的深入扎根情况；(b) 可能从评价工作中受益最多的那些人对评价的利用情况(双圈学习)；(c) 共享信息和围绕基于证据的信息发展实践社区；(d) 评价职能的适应性及持续变革和自我更新，在产生知识和为促进发展评价作贡献方面具有创造性 and 生成性；以及(e) 与其他监督和知识生成职能建立战略联系和协调。

70. 附件九所载分析结果表明，参照用于界定一个学习型组织的标准，联合国系统的业绩表现“低于一般”水平，包括那些已将成果和评价文化嵌入本组织结构的那些组织。分析结果还表明，联合国系统在发展学习型组织方面的最大优势是：(a) 共享知识和发展实践社区；(b) 不断评估和调适评价职能；以及(c) 为促进发展评价作出贡献所作的努力。最弱的业绩领域是：(a) 建立同相关职能(其他监督和知识管理系统)的战略联系；(b) 外联和对分散评价职能或对国家或其他伙伴系统的支持；以及(c) 利用评价作出决定和采取行动。下文涉及评价职能对变革和挑战产生影响力所必需的条件和关于分散职能的章节将讨论这些领域。

71. 必须指出，近年来，一些组织已重视发展学习型组织；其中包括联合国环境规划署(环境署)、世界卫生组织(世卫组织)和劳工组织。世卫组织对其评价职能作了重大改组并在 2014 年 8 月设立了一个单独的评价部门作为“评价和组织学习”职能的组成部分。

72. 对于发展学习型组织至关重要的是组织的领导力和要制定一个重视创新和风险管理的激励机制，以预防采用那种小心翼翼、害怕失败、不足以应对挑战的做法。评价职能要取得成功就必须在确定方向和管理来自创新的风险方面有坚强的领导力。有证据表明，各组织在关键领导力属性和做法方面的业绩不是十分强劲。平均而言，只有 4 个组织在用于评估领导力的所有标准中表现出高水平的业绩。对各组织领导力的评估内容包括联检组的 5 个领导力属性⁴⁴ (见附件十)。对优势和弱点作分析后表明，领导作用在涉及问责制和独立性系统时高于一般水平。参照这个标准，大约有 11 个组织表现出高水平业绩。在评价的制度化和为评价职能制定愿景和战略方面则要差劲一些。如果执行以下建议将会加强战略方法的发展和评价职能的实效。

⁴⁴ 联检组领导力属性包括价值观和榜样、宣传和交流、指导和纪律，以及在利用评价方面的参与。

建议 2

联合国系统各组织行政首长应采用平衡方法来解决如何将评价用于促进问责制的问题，解决如何将评价用于发展具有适当激励机制以促进创新、敢于承担风险和采用多学科视野的学习型组织的问题。

资源、资金和预算分配的适当规范

结论 7：是否能维持中央评价职能的成熟度并确保在与适当评价质量和影响力作平衡的情况下适当覆盖将要评价的各种问题依赖于是否能提供适当、可预测和妥善管理的资源。在许多组织中用于评价职能的资源不足，尤其是小型组织，供资情况不稳定，也不可预测。普遍缺乏可用于指导和保障预算分配的规范性框架。只有 3 个组织在其评价政策中规定了确立的预算规范，但规范的基础似乎是临时性的，没有以坚实的费用和健全的财务分析作为基础。对这种分析框架的实质作出界定尚未成为优先事项，部分原因在于存在一些困难，包括评价职能的预算结构未涉及该职能的全部范围，也无法在各组织之间进行比较或设立基准。

资源不足又难以预测

73. 根据评价办公室负责人报告的信息和对为回应要求而要采取的措施所作的分析结果，可以得出如下结论：联合国系统大多数组织的中央评价职能在很多程度上资源不足。其平均运行费用占组织支出的 0.3%。分析结果表明，0.3%的投资比例对于各组织要运行一个高质量的职能是不够的，而这个职能的目的是要使联合国系统了解自己在世界上发挥了什么作用，对自己受委托去帮助的人民的生活产生了什么作用。

74. 除 2 个组织(2013 年，妇女署中央和分散职能支出所占比例为 1.3%；工发组织为中央级职能的拨款占总支出的 0.5%)外，其他大多数组织的供资水平在 0.05% 至 0.39% 之间。本研究报告所作统计分析强调，必须要有适当供资，因为分析结果表明，所评估的供资适当度与所评估的评价职能成熟度之间有高度关联性。同样，第一个群组中业绩优良的组织 and 单独组织对评价工作投入的资源相对更多。

75. 对有足够数据集供分析的 13 个组织而言，非核心资源占该组资源的 24%，虽然各组织之间的差异很大。据报告，非核心资源极大地帮助了评价方案的交付。然而，这类资源同样不可预测，产生高昂交易成本，并影响评价职能的稳定性、覆盖面的相关性、效率和独立性。由于数据不够无法得出明确结论，但分析结果仍然显示，提供高数额核心资金的各组织往往有较高的发展水平，虽然这种情况并不具有统计意义。这些组织往往还采取了强有力的措施以保障独立性、限制交易成本和有效利用非核心资源以支持交付工作。同样，与提供较高数额非核心资源的组织相比，它们的独立性也较高(见附件十一)。

预算规范和用于建立规范的基础

76. 在目前削减预算和组织改革的情况下，保护用于一般评价的预算拨款是关键利益攸关方的关切问题。目前缺乏规范和规范性或以价值为基础的框架来指导预算拨款。人们普遍认为，提供预算规范将有助于显示评价职能的价值并保障其可见性。4个组织(妇女署、联合国人口基金(人口基金)、儿基会和世卫组织)确定了用于预算拨款的规范。它们在制定规范时采用不同的参数；但是，用于界定这些规范的实际财务基础不明显。⁴⁵

77. 评价部门负责人指出，拟订预算规范时采用了不同方式，这是整个联合国系统在制订基准和在标准化过程中存在的一个问题。第一，针对方案制订规范的做法是假设，评价职能只生产对方案的评价结果，但联合国系统内的评价职能除此之外还要开展一系列与评价相关、对本组织十分重要的其他活动。第二，今日开展的评价工作比过去作的典型业绩评估工作更为复杂。因此，财务基础必须顾及不同类型评价的复杂程度和质量。第三，组织和任务规定的规模不同会对所开展的各种类型评价产生不同的影响，因此“一刀切”的做法行不通，与此相关的是，还必须考虑到规模经济问题。⁴⁶ 这一原则造成小型组织相对大型组织会支出更多。在这方面，一般的看法是，根据组织的任务规定、组织规模、评价类型及该职能在制度化方面的作用和在支持加强分散评价、国家评价能力和评价中的伙伴关系情况，占组织支出 0.5%至 3%之间的供资范围是值得考虑的。

78. 除了这些因素之外，在核算中央评价职能的成本时还必须作战略思维，同时考虑到与为组织决策提供证据基础的其他职能的互补性和协同效应。因此，必须考虑到分散评价职能、监督办公室就业绩评估所作的工作，以及知识管理在集中所有现有知识集以为本组织找出经验教训方面所起的作用。所有这些职能都为组织决策提供了各种证据基础。还必须考虑到每一个业务机构的性质：单一地点对多个地点、业务情况和服务交付方面的不同性对行政和管理。

79. 应在制订用于资源分配的规范性框架时考虑这类因素。建议对联合国开展的工作进行分类，以使具体的评价方法与相关资源能相匹配。联合国评价小组的一些成员认为有必要进行分析，制订参数和基准框架，以指导不同类型的组

⁴⁵ 妇女署的目标定在为中央和分散职能提供方案预算的 3%；人口基金定在为中央和分散评价职能提供方案预算总额的至多 3%。世卫组织定在为中央和分散职能提供方案预算的 3%至 5%。对儿基会而言，相关政策呼吁将方案资金的 1%拨给评价工作。世卫组织、人口基金和妇女署这三个规模不同的组织有一个相同的成本规范，这一点导致产生对用于界定相关规范的基础的疑问。但它们表明，为界定评价职能在组织内的价值和为保障其可见性作出了努力。

⁴⁶ 规模经济的定义：“成本优势随着一个产品产出的增加而加强。规模经济的产生在于数量与每单位固定成本之间的反比关系，即所生产的一种物品的数量越多，每单位固定成本就越低，因为这些费用被大量物品分摊了。规模经济还可能因为提高了运行效率和协同效应而降低每单位的可变成本。规模经济可分成两个大类：内部——从公司内部产生；外部——由外部因素产生，例如产业规模”。(摘自 Investopedia.com)

织制订用于资源分配的预算规范。这将是一个需要采取的重要步骤，不仅能使联合国系统受益，而且还能使包括双边和多边组织在内的大量其他发展伙伴受益。关于预算拨款的证据和可适用的规范表明，这些发展伙伴同样面临着上文概述的那些挑战。⁴⁷

成本核算和预算结构

80. 鉴于目前在界定用于评价职能的预算规范方面存在挑战，因此有必要制订并在较长时间内监测一个经过全面成本核算的工作方案，作为制定规范的一个基础；正如世卫组织和妇女署(这两个组织分配了方案预算的 3%)最近在政策中所要求的，必须作这方面的考虑。但是，加强预算框架的可比性同样有帮助。目前，各组织的预算结构不同，采用同样措辞的预算项目却意味着不同的东西。就各组织的评价职能而言，必须有一个用于评估其工作性质的标准方法，以及相关的评价方法及随之而来的结构和资源需求。

用于加强资金效用的选项

81. 虽然要满足不断增加的评价要求就需要更多的资金，但评价工作负责人已提出了各组织之间进行协作和合作以提高效率的各种方法。这包括汇集资金和利用一个服务提供方来开展某些共同活动，一个恰当的实例就是对评价报告的质量评估。联合王国国际发展部的所有评价职能就使用了一个服务供应方，由此得出的经验教训可能是有用的。此外，从联合国系统各组织在过去 5 年里所开展的 44 个联合评价中得到的经验教训同样为评估制约因素及界定跨组织协作和在全系统如何提高效率提供了一个良好基础。还可以从人道主义行动问责和绩效动态学习网络倡议中吸取经验教训，例如，对其成员评价报告的质量评估是由该网络作的。在这方面，可以预见联合国评价小组将在质量评估方面发挥更强有力的作用。因此，强烈建议评价部门利用各种经验教训，增强各组织之间的协调、合作和协作，以提高资源的使用效率和加强评价方面的伙伴关系。

小型和资金不足的组织的资源筹措和有效管理资源

82. 研究结果显示，评价方面的大多数资源增长都发生在早已完善确立的职能部门，虽然仍与该职能的要求不相匹配。必须重视加强成熟度较低、资金严重不足和工作人员的专业职等不高的那些办公室。

83. 从数据分析中得出的模式显示，第三个群组中业绩评分较低的组织显示出在对评价职能的供资水平上存在很大差异。两个供资模式显示 4 个组织的组织支出平均值为 0.07，另 4 个组织则是 0.6。第三群组中的这一组从比例上看雇用了更多较低职等(P2-P3)的工作人员(请参见辅助材料里的各个表格)。该组还往往利用这批工作人员来开展各种非评价活动(见下文表 8)。

⁴⁷ 见经合组织，“改善援助工作：发展机构的评价工作”(2010 年)。

84. 各组还有相当多的小职能和合署办公职能。有证据强力显示，具有合署办公职能的组织拨付的专用资源比有单独职能的组织要少。有证据表明，需要认真考虑增加对合署办公组织中评价职能的供资。证据还表明，这类职能，特别是当评价职能与其他监督或管理职能完全整合在一起时，应加强对专业评价工作人员时间的有效利用，应把这种时间专用于评价而不是从事其他任务。

表 6

不同类型组织的支出和专用于编写报告这一核心职能的时间与用于其他任务的时间的比较

组织类型	在评价方面的组织支出所占百分比	在编写评价报告和提高质量方面所用时间百分比(相对于在该部门其他活动上所用时间)
A. 业绩水平	0.33	69
第一群组 发展水平较高	0.38	84
第二群组 发展水平一般	0.30	51
第三群组 发展水平低于一般		
B. 组织规模⁴⁸		
大型组织(在 10 亿至 50 亿美元之间)	0.18	68
中型组织(在 10 亿美元至 3 亿美元)	0.17	73
小型组织(少于 3 亿美元)	0.33	53
C. 结构地点		
单独	0.33	73
与监督部门合署办公	0.14	55
与管理部門合署办公	0.28	64

85. 考虑到评价对组织和发展实效的价值及不开展评价会产生的风险，特别是在目前的财政状况下，因此执行下列建议将增强评价职能的实效、可持续性、增加价值和效率。

建议 3

立法机构应要求联合国系统各组织行政首长根据为维持一个可为本组织增加价值的有效和可持续的评价职能所需要的成本，为各自的评价职能制订综合预算框架和资源分配计划。应在现行预算和报告机制及流程范围内向立法机关提交这些计划供审议。

⁴⁸ 请注意，小型组织资源百分比较大的原因在很大程度上与上文提到的规模经济概念有关。

构成部分 2：具有包容性的独立性，以提高公信力

86. 独立性有助于确保评价的公正性和客观性，从而能产生公信力。研究结果表明，这是评价职能所有构成部分中发展水平最高的构成部分，虽然根据本次研究确立的标准它仍在一般成熟度的范围内运作。下文表 7 显示了这个构成部分的优势和弱点领域。根据对调查结果的分析，本研究报告作出以下结论：

表 7

独立性构成部分的优势和弱点

已取得进展的领域	进展不足的领域
<p>专业/技术独立性：</p> <ul style="list-style-type: none"> 评价员和评价管理人员的专业忠诚和技术质量 	<p>结构独立性</p> <ul style="list-style-type: none"> 任命负责人 负责人的任期和在组织内的轮调
<p>结构独立性：</p> <ul style="list-style-type: none"> 评价办公室设在行政首长办公室和管理部门之外，并有独立决策能力 	<p>职能独立性</p> <ul style="list-style-type: none"> 预算流程独立性
<p>职能独立性：</p> <ul style="list-style-type: none"> 负责人对工作方案及在无需通过组织负责人编写和向成员国及公众分发报告方面有充分酌处权 组织和管理部门允许接触所有信息 	
<p>建立促进公正性和透明度的机制：</p> <ul style="list-style-type: none"> 利益攸关方的广泛参与以促进平衡视角和透明度及利用独立外部专家 	

结论 8：尽管在制度发展以加强评价职能的独立性方面取得了进展，但理事机构和各成员国仍然认为公信力是个问题。特别是，该制度需要处理与结构独立性有关的限制因素，特别是对评价部门负责人的任命及其任期、多样性和专业教育背景。

87. 所有利益攸关方都关心职能独立性问题。对职能独立性的重视是被成员国的要求驱动的，因为它们来说，有独立性的问责制对于公信力是必不可少的。评价部门同样从专业角度和为了使双边组织更多地利用其评价成果、从而限制双边组织的平行评价而非常重视这个领域。⁴⁹ 管理部门也在保障独立性方面发挥了

⁴⁹ Ian C. Davies 和 Julia Brümmer, “联合国评价小组评价职能同行评审的经验教训研究。最后报告”, 2013 年 5 月。

表 8
24 个组织评价职能的报告关系

报告线的独立性	开发署	人口基金	监督厅	粮食署	原子能机构*	劳工组织	教科文组织	工发组织	儿基会	世卫组织	知识产权组织	气象组织	贸发会议	妇女署	粮农组织	国际民航组织	环境署	毒品和犯罪问题办公室	近东救济工程处	人居署	难民署	国贸中心	海事组织	艾滋病署	
1. 部门设在行政首长办公室之外，完全独立于所有/几乎所有管理职能																									
或																									
2. 部门设在行政首长办公室之内																									
和																									
3. 在管理部门不参与的情况下，评价报告直接分发给成员国(有时通过执行局的一个监督和咨询委员会)																									
4. 在管理部门不干预的情况下，关于评价工作的年度报告直接分发给成员国，报告是全面性的																									
5. 工作方案得到确定并直接分发给成员国(以得到成员国的批准)											X														
6. 在管理部门不干预的情况下，向成员国直接提出预算要求																									
7. 行政独立性(包括完全控制对财务和人力资源的管理、对工作人员和咨询人的甄选，以及经常性行政职能)																									
机制是适用的																									
其他机制是适用的																									
无需成员国批准																									X

* 原子能机构具有该组织特有的差异。

89. 一个主要关切问题是，就独立性作出的重点努力是否会导致中央部门被完全孤立并脱离本应由其服务的组织。人们越来越认为，在利益攸关方参与下或有包容性而不是孤立的独立性对于联合国系统评价职能的公信力、信任和透明度至关重要。相当多的组织在这个指标上取得了很好的业绩。用于使利益攸关方参与评价流程各个阶段的机制提供了各种重要的替代办法，以把评价职能与行政首长联系在一起，如果这样做的目的是加强与组织的联系。

90. 从本研究报告收集的证据中看出的对评价职能的重大挑战是结构独立性：需要关注的主要问题与评价部门的负责人和报告关系有关。问题在于：(a) 谁选择或决定任命评价负责人；(b) 负责人的任期、他/她在当过评价职能负责人或仍在任时是否可以轮调回方案管理部门；(c) 负责人在评价工作方面的背景和教育；以及(d) 区域和性别的多样性，因为目前是不平衡的。下文表 9 根据各组织的报告说明了现状。

91. 就评价职能领导力及其在加强各种利益攸关方中的公信力和信任的直接作用所提出问题对于大多数组织来说都是富有挑战性的。就甄选和任命负责人来说，无论由行政首长或是由成员国来甄选都无法做到没有政治偏见。为支持监督职能的独立性，联检组曾提出建议，即在任命内部监督负责人时，每一个组织的立法机构应决定：(a) 根据广为宣传的空缺公告确认合格的候选人；(b) 任命前应与管理机构协商并获得其事先同意；(c) 终止任期应有正当理由，并应受理事机构的审查和征得其同意；(d) 应设立一个 5 年至 7 年的不得连任任期，并在任期结束后不存在被同一个联合国组织作进一步雇用的期望。⁵⁰

92. 研究报告仅审视了最后一点和其他能保障负责人的独立性的因素，例如报告关系。分析结果显示，在如何对待这个问题方面采用了各种不同的模式，反映出所有组织都必须重新审视整套独立性指标并拟订一个合理和平衡的方法来加强和保障独立性。

93. 有 13 个组织的评价职能在结构上未独立于高级管理层并在组织负责人的办公室内运作(要不作为一个单独的评价职能，要不与该办公室的管理职能全面整合在一起)。其中 6 个组织(环境署、毒品和犯罪问题办公室、联合国近东巴勒斯坦难民救济和工程处(近东救济工程处)、联合国人类住区规划署(人居署)、国际贸易中心(国贸中心)和国际海事组织(海事组织))没有报告方面的独立性。5 个组织(国际民用航空组织(国际民航组织)、妇女署、联合国贸易和发展会议(贸发会议)、联合国儿童基金会(儿基会)和工发组织)在报告方面有有限的独立性，但如上文表 8 所指出的，在程度上各不相同。在 13 位评价负责人中有 11 位是定期任用，在离开评价职能后会在本组织内部轮调。有两个任期限制：儿基会的评价负责人可连任两届，但在第二个任期后不能再延长；国际民航组织的评价负责人有一个 5 年至 7 年的任期；这两个组织的评价负责人在报告关系方面都只有有限的独立性。这些组织的立法机构应重新审议其关于评价职能结构独立性的政策，并决定如何以最佳方式加强和保障结构独立性，包括在任命评价部门的负责人方面。

94. 有 9 个组织(联合国内部监督事务厅、开发署、人口基金、粮食署和教科文组织、开发署、世界知识产权组织(知识产权组织)和监督厅检查和评价司)在以下的结构独立性方面做得较好：(一) 评价职能设在行政首长办公室之外；(二) 任命能尊重独立性的负责人；以及(三) 在报告方面具有独立性。但在这个群组中，监督厅、知识产权组织和教科文组织评价负责人的独立性最强，因为他们只有一个分别为 5 年和 6 年的任期，任期不得延长，也不得再次被本组织雇用。只有一个任期、任期不得延长的政策被认为比任期可延长政策赋予了更强的独立性。在大多数情况下延长任期决定由各组织的负责人作出，如果任职者寻求留在本组织，这种做法就有限制独立性的作用。5 年至 7 年的时间提供了一个长期运作的机会，从而能充分了解组织情况并能将评价职能置于能支持变革和发展的最佳位置。从

⁵⁰ 联合国系统监督制度的缺陷，JIU/REP/2006/2。

人力资源管理角度看，并为了采用适当激励机制以吸引外部的优秀人才，需要至少 5 年的服务时间才能受益于联合国养恤基金。建议开发署、人口基金和粮食署的立法机构重新审议两个不可延长的任期政策，并决定采用一个 5 年至 7 年的不可延长任期。

95. 未列入前面两个段落的组织包括：劳工组织、世卫组织、世界气象组织(气象组织)、粮农组织、联合国难民事务高级专员公署(难民署)、联合国艾滋病毒/艾滋病联合规划署(艾滋病署)，其中大多数是具有不同程度结构独立性的专门机构，这使人不得不问：作出这类选择的基础是理性考虑或是在某种程度上与任务规定或规模有关。劳工组织有一个在行政首长办公室之外的办公室，但评价部门的负责人通常是长期任职的工作人员，并可在组织内部轮调。粮农组织的评价负责人可做完两个不可延长的任期，但在报告方面实际上没有独立性。世卫组织的评价负责人在行政首长办公室之外运作，但相关任命允许在组织内部进行轮调。气象组织的评价负责人在该组织负责人的办公室之外运作，但在报告关系方面独立性受到严重限制。难民署、艾滋病署、国贸中心和贸发会议的评价负责人完全与管理部门整合在一起了，在报告关系上没有独立性，并是以定期方式雇用的。建议这 8 个组织的立法机构重新审议结构独立性、雇用评价负责人的方式和组织内的轮调做法，以加强和保障评价职能的独立性，并适当考虑可作比较的那些组织的组织条件和经验。

表 9

结构独立性：按任命制度和组织内轮调情况分列的各个组织

	由理事/立法机构任命	由行政首长任命，同时听取理事机构和/或其他独立的人权或甄选机制的意见	由行政首长任命，不听取理事机构或其他方面的意见
任期 负责人有一个 5 年至 7 年的任期并不得再次被本组织雇用		联合国秘书处(5) 知识产权组织(5) 教科文组织(6) 国际民航组织(5)	
负责人的任期不超过两届并不得再次被本组织雇用		开发署(4+4) 儿基会(4+4) 粮食署(4+4) 粮农组织(5+5) 人口基金(5+5)	
定期任用并与组织内其他工作人员一样适用在本组织内部轮调的规则		毒品和犯罪问题办公室 气象组织妇女署 世卫组织**	工发组织、贸发会议、环境署、人居署、难民署、近东救济工程处、劳工组织、海

			事组织、艾滋病署、 原子能机构*
--	--	--	---------------------

* 原子能机构对所有工作人员的政策是一个 7 年的任期，然后离开本组织，但允许在组织内部轮调，从而对独立性产生影响。

** 世卫组织在 2014 年调整其评价职能，现正在制订新政策。

96. 鉴于评价工作的复杂性越来越高，人们对评价作为一个在应用复杂方法方面有明确标准和严格性的专业和学科的了解也越来越清楚，因此就提出了关于评价负责人的教育和背景的问题。D-1、D-2 和 P-5 一级的负责人具有不同的背景。在监督部门的评价职能负责人是审计员，⁵¹ 他们同时兼任监督负责人，在联检组 24 个参与组织中占这类负责人的 38%。大多数，但不是全部，其他评价负责人是根据评价背景任命的。在某些情况下，具有扎实的机构知识被认为比具有评价知识更为重要。有人指出，负责人的评价背景不如使负责人得到评价员的支持那样重要。与此同时，由于缺乏批判性和战略性调查方式，领导力和公信力可能受到了损害。

97. 在联合国目前严格的等级制度下，当一个 P-5 级工作人员被指派为评价负责人时，评价职能的公信力和地位就受到了限制。为改变这种情况，现在在大型组织里出现了一种日益增强的趋势，即任用 D-1 及以上级别的评价负责人，他们不仅要满足领导能力方面的要求，而且要有评价方面的教育背景或扎实的评价经验。最后，分析结果表明，目前关于负责人的人口信息让人质疑：在选择负责人时在何种程度上考虑到了联合国的包容性和多样性原则？因为 85% 的负责人来自发达国家，75% 是男性。

98. 执行以下建议将会提高评价职能的公信力和实效。如上文所述，鉴于各组织之间在独立性组合方面的差异很大，本建议虽然是通用的，但在执行时必须适当顾及组织的不同类型。

建议 4

立法机构应指示联合国系统各组织行政首长审查和酌情修订用于任命评价办公室负责人的现行政策，以加强独立性、忠诚、道德操守、公信力和包容性，并要适当考虑到下列标准：

- 应为单一、不可延长的任期设立一个 5 年至 7 年的期限，而且任职者没有再次被该组织雇用的可能性；
- 评价负责人应具有评价方面的资格和实质性经验，并有在战略规划、基础和业务研究及知识管理等相关领域方面的经验作为辅助，以及应有出色的管理和领导能力。

⁵¹ 2014 年出现了一个例外：新任命的教科文组织监督办公室负责人是一位专业评价员。

构成部分 3：质量-技术和管理严格，以提高公信力

99. 下文表 10 汇总了在成熟度矩阵中界定的本构成部分各指标经评估后得出的优势和弱点。

表 10
评价流程和产品的质量和公信力——优势和挑战

已取得进展的领域	进展不足的领域
<ul style="list-style-type: none"> 评估的内部质量保证(内部同行组、来自方案部门的参考小组的参与、咨询小组的参与，以扩大视野和看到全面价值) 工作人员的能力：技术、专业、管理和机构知识 	<ul style="list-style-type: none"> 对报告和评价职能的外部独立、质量评估(例如，联合国评价小组/经合组织发展援助委员会的同行评审和理事会要求进行的其他独立审查) 评价报告的质量 对评价职能的外部评估：联合国评价小组对小型和合署办公组织的同行评审 工作人员的专业发展 在评价方法和类型方面的创新 技术和管理导则和工具

结论 9：各组织力求通过采取一系列措施来确保评价职能的有效性和严格性，包括内部部门质量保证、利用外部专家、利用关键利益攸关方的参考小组和加强工作人员的能力。虽然各组织使用多种内部质量保证机制，但没有采取两种外部措施，而这两种措施对于评价职能的高度发展具有重大意义。它们是：(a) 对评价职能的外部评估，包括通过联合国评价小组/经合组织发展援助委员会对评价职能的同行评审；和(b) 对中央评价报告质量的独立评估。关于联合国系统中央评价职能所产出评价的质量的现有独立证据很有限。这就限制了评价职能的公信力。

100. 研究发现，许多组织的评价职能采用了大量措施来加强有效性和可靠性。评价办公室内的内部部门质量保证机制包括：让本部门或其他组织的同行担任评价质量保证小组的成员、利用外部咨询机构、利用关键利益攸关方参考小组，以及在编写报告的各个阶段举行公众和利益攸关方会议以验证调查结果和结论。所有这些都助于提高评价报告的质量、可信度和效用，

101. 在加强评价职能公信力方面的另一个要素就是工作人员的能力。据报告，工作人员的技术、专业、管理、机构和方案知识及在专业和个人忠诚方面的高标准均有助于确保独立性和加强作出平衡和客观的评价。评价部门的核心工作

人员通常都会在确保应用评价原则以维持技术严格性和在确保评价对本组织是有意义的方面发挥领导作用，同时还要管理评价的有效流程。工作人员与订约外部专家/咨询人一起开展评价工作；当评价部门要处理一系列广泛主题事项而无法适当地由本部门长期承担时，雇用外部专家/咨询人能加强评价内容的有效性。

102. 虽然上文第 100 段阐述的内部(部门)质量保证机制相当强健，但该系统没有采用明确界定、客观和系统性方法对评价职能和评价报告作可信、透明、独立的外部质量评估。有 9 个组织利用联合国评价小组的同行评审机制和/或由执行局或管理部门委托进行的外部评价或利用双边小组对其职能作外部评估。

103. 就对评价报告的质量评估而言，研究报告未列入对这类报告的实际元评价。就本研究报告内其他大多数类目而言，检查专员们根据对评价报告质量的现有外部评估分析了质量情况。有 13 个组织对其报告质量作了外部评估。这类评估表明，在这些组织中只有 7 个组织(劳工组织、监督厅、工发组织、儿基会、开发署、环境署和粮食署)的评价报告的质量在符合专业标准方面达到了较高水平。各组织确实借鉴了联合国评价小组为优质报告制定的标准，但是它们还应用了其他标准并采用了不同的衡量体系。因此，很难进行解读和作比较分析。

104. 由于系统性外部评估的数量和关于评价报告质量的了解有限，并且评价报告的质量不令人满意，因此对整个系统评价工作的产出和价值就产生了一些重大关切。但如同其他许多评估领域一样，各组织之间存在差异。现在双边组织越来越多地对评价职能及其产品作系统性外部评估。虽然这很重要，但不同的外部评估系统的扩散(因为不是所有系统都采用同一标准)可能会导致重复工作、高昂的交易成本和工作上的相互矛盾。

105. 执行以下建议将会提高评价职能的公信力和实效。

建议 5

联合国秘书长应以联合国系统行政首长协调理事会(首协会)主席的身份要求联合国评价小组进行协作，以拟定一个可用于整个联合国系统评价职能的健全和统一的质量保证制度。

构成部分 4：效用-利用和利用所产生的影响

106. 研究报告评估了以下一些从理论上来说对加强评价职能的成果和影响力至关重要的领域：(a) 用于提高利用率的各项条件已经到位，包括工具、技巧和流程；(b) 优质报告；(c) 利益攸关方的参与；(d) 组织文化和领导力；(e) 对评价产品的利用程度和性质；(f) 评价员、管理人员和理事机构在加强利用方面的互

补作用和责任；以及(g) 用于评估利用评价后产生的影响力的系统已经到位。前面一个章节已经述及对评价的利用具有极其重要意义的组织文化和劳动力，因此本节就不作论述了。

107. 评估结果表明，评价职能的效用构成部分是该职能全部 5 个构成部分中最为薄弱的，总体业绩低于一般水平(下文表 11 左栏所列的那些指标是例外；表 11 提供的是能观察到进展情况的指标)。

表 11
与评价职能效用相关的优势和弱点

已取得进展的指标	进展不足的指标
用于提高利用率的措施 <ul style="list-style-type: none"> • 要求和利用意图及与要求相关和回应要求的规划标准 • 利益攸关方的参与 	对评价的利用(成果指标) <ul style="list-style-type: none"> • 通过建议执行率看利用水平
用于产生直接效应的定向工具、系统和措施已到位 <ul style="list-style-type: none"> • 及时性和向决策进程提供信息 • 在网站和其他平台可取得报告以促进利用和透明度 • 内部交流 	主流文化的影响 <ul style="list-style-type: none"> • 评价文化(文化组织、领导力；在知识系统方面的调整)
	用于提高利用率(直接效应)的定向工具、系统和措施已到位 <ul style="list-style-type: none"> • 传播和宣传战略 • 建议追踪系统(用于利用评价后的后续执行) • 对外交流
	影响评估系统 <ul style="list-style-type: none"> • 组织愿景和战略及评价以何种方式对此加以支持 • 评价职能的影响指标 • 影响的后续行动制度和报告

108. 各组织在效用构成部分上的得分显示，没有任何组织处于得分表的高端位置。2 个组织处于“高于一般”水平，15 个组织处于“一般”水平，其余 11 个组织处于“低于一般”水平。

图五
联检组 28 个参与组织在效用方面的业绩



结论：在决策时系统地利用评价证据的水平较低。这种情况与为确立评价机制和系统以加强利用中央评价部门的评价结果所作努力的水平不相符。对评价结果的有限利用无法加强作出基于证据的决策和改进做法。从而限制了评价职能的价值。各组织没有重视利用评价结果所能产生的影响。用于记录和评估影响力的制度如果存在的话也是不成熟的。要使对评价结果的利用所产生的影响力概念化就必须为评价职能、其预期成果和影响力指标及为这一影响力如何与本组织内其他监督和知识生成系统的成果加以协调制订一个全面和综合的愿景和战略框架。

利用水平和用于评估评价结果的利用情况的后续追踪系统

109. 评估评价结果利用情况的一个具体方法是审查利用从评价中得出的各项建议的情况。对评价中得出的各项建议的利用水平远远低于预期业绩的一般水平。只有 21% 的组织的利用水平较高。这意味着在 3 年内得到执行的建议超过 85%。这一结论的基础是基于后续追踪系统得出的执行率所报告的信息。用于监测建议执行情况的追踪系统为评估对评价结果和建议或评价中所提出的问题的利用水平提供了一个具体基础。40% 的组织(11 个组织)⁵² 建立了完善的追踪系统，就评价结果的利用情况作了较好的报告。

利用的性质

110. 本研究报告确认了对评价结果的 5 类利用方式：利用目的：(a) 战略决策；(b) 整体(在方案或机构方面的)累积目的；(c) 发展的目的——方案和项目设计和

⁵² 监督厅、开发署、妇女署、儿基会、原子能机构、国际民航组织、教科文组织、工发组织、劳工组织、知识产权组织和气象组织。

调整；(d) 外部伙伴(国家伙伴、发展伙伴或专业团体)要求；以及(e) 用于加强增加值(例如，通过元分析和经验说明)，以为评价专业作出贡献，并促进提高评价方法和办法。调查结果显示，评价结果最频繁的用途是用于改进本组织的方案和成果的实现，其次是用于战略决策(见辅助材料，附件 21)。⁵³

已到位的用于加强利用评价结果的条件

111. 本次研究审查了各组织为确保有一个高影响力评价系统而提供的条件和要素，包括：

- 主流文化和领导力的影响；
- 使用该系统的要求和意图及其覆盖范围；
- 一整套明确界定和直接的关键工具和流程以加强利用和影响力。

112. 分析结果显示，评价系统用于加强利用的最发达的机制是：努力使覆盖范围有意义并满足要求、利益攸关方的参与、通过网站和评价数据库提高可获取性和透明度，以及评价报告的及时性。总的来说，为加强利用评价结果来设立的制度和条件处于发展的“一般”水平，正在实现制度化。但是，特别是管理层需要做更多的工作，以加强评价和学习文化。劳工组织启动了一个用于加强利用的重点、系统性进程；需要记录这个进程并在更广泛的范围内共享来自该进程的经验教训。⁵⁴

用于评估评价影响力而已到位的系统处于初级水平

113. 在本次研究过程中人们发现，虽然有传闻证据，但各组织几乎不评估利用评价及其产品所产生的影响力。因此，研究报告评估了用于评估影响力而且已到位的制度。有证据显示，在发展相关制度时缺少以下内容：(a) 关于评价职能的愿景和战略，包括成果框架和关于评价职能的一整套全面的指标，以及将评价职能的成果框架与组织成果相协调；(b) 强制性和行之有效地执行建议并有一个后续追踪系统；(c) 根据既定成果框架报告执行情况和影响；以及(d) 报告同样有助于加强利用效应和影响力的其他因素。各组织必须利用评价来影响问责制、学习和决策，并使利用评价所产生的影响力作为战略优先事项。与此相关的是关于必须发展一个学习型组织的建议，利用评价信息促进本组织的变革和发展及加强其交付更好成果的能力。各组织还应在向立法机构提交的关于评价工作的年度报告中报告利用评价所产生的影响力情况。

114. 执行以下建议将会提高评价职能的实效。

⁵³ 见 www.unjiu.org/en/Pages/default.aspx。

⁵⁴ Janet Neubecker、Matthew Ripley 和 Craig Russon，“提高利用率：国际劳工组织的案例”，载于 Marlène Läubli Loud 和 John Mayne 编辑的书籍，“加强利用评价：内部评价部门的见解(2014 年，伦敦，Sage 出版社)。

建议 6

联合国系统各组织行政首长应根据对评价职能的明确界定的愿景、战略和成果框架把利用评价作为组织的优先事项，并就利用评价的程度、性质和影响力情况向其立法机关提交报告。

构成部分 5：相关性和为支持联合国各组织和全系统的改革及为应对新出现的变化和准备状态

115. 这一构成部分涉及(a) 相关性和通过联合评价提高效率，和(b) 评价职能在帮助联合国系统在各组织内和在全系统范围一级应对变化、挑战和新出现的转型变革议程和可持续发展议程方面的能力和战略定位。

116. 在当前世界的变革和挑战格局中，有人质疑评价职能是否能支持处于发展十字路口的联合国系统。⁵⁵ 提出的一个问题是联合国系统分门别类和功能性结构的问题，因为这种结构限制了能向新出现的议程(包括 2015 年后促进可持续发展的议程)提供支持的全系统视角的转变。这种结构同样限制形成一个关于能联系并影响各组织所开展的评价工作的评价职能的全系统视角。换言之，各组织的中央评价职能至少将是“全联合国系统”的一个“分系统”。但实际情况远非如此。在这种情况下，人们期望看到对其工作方案产生影响的各中央评价职能之间的联系，并提出如下问题：从整个联合国而不是从单个组织的角度来看，哪种类型的评价将带来最大的增加值？而这种增加值反过来又影响了不同中央评价职能之间的资源分配决定。

117. 但迄今为止，关于 2015 年后发展议程的辩论对能满足这类标准的评价的必要性显然保持了沉默。对作为问责制报告框架组成部分的数据和数据报告十分重视，但却没有深入反思可持续发展议程对评价作用的影响，包括分散评价在支持形成性评价和中期纠正方面的关键作用。它同样回避了国家系统在评价和评价能力发展方面的重要性和联合国系统在加强这类能力方面的作用。

118. 研究报告审查并评估了各组织在解决与组织对覆盖范围要求的相关性问题时所采用的方式，包括通过联合评价，同时还涉及对支持现代改革和变革的回应和准备状态具有重要性的一些因素。

结论 11：所审查的 24 个组织的评价职能一直力求协调和统一工作，主要是通过联合国评价小组，并采用主要在各自组织内集体制定的规范和导则。但尚未在各机构之间进行协作和开展联合工作。联合国系统实行单独的治理、业务程序和预算编制流程，这种“分门别类和功能性”的性质阻碍了为进行联合评价所作出的

⁵⁵ Bruce Jenks 和 Bruce Jones，“处于十字路口的联合国发展”。

努力。有证据表明，联合方案编制为联合评价提供了机会，通常是在交叉主题，如粮食安全、性别平等、环境和劳工问题。

119. 检查专员们发现，联合国系统大约 50%的组织未参与联合评价；大多数组织并不认为这是一个优先事项，而更在意如何把评价职能放在正确的轨道上。认为联合或机构间工作有困难的想法同样限制了开展联合工作的动机。在 24 个组织中有 11 个组织开展了联合评价：至今总共开展了 44 次联合评价，主要是过去 5 年中在这些联合国机构里开展的。劳工组织作为一个规范性机构为其他机构持续进行的评价工作履行了咨询、协调和支持职能，反映出主要是规范性的组织就评价开展的一种不同形式的机构间工作。劳工组织在过去 5 年中开展了 150 多项这种类型的活动。

120. 对参与联合评价的一些组织团体的访谈显示出要继续致力于开展联合评价的决心，但也存在大量会影响联合和机构间评价的结构和系统性问题。这些问题包括：单独的治理结构、要启动联合评价就需要支付先期投资费用(包括资金和时间，因为交易成本很大)、需要提供管理和协调方面的支持以使各机构能够把重点放在联合工作的实质性内容上、在统一方案和预算周期方面存在困难、不同组织采用不同的方法、受理和讨论联合报告的立法机构不同、不确定应在组织还是在机构间机构一级开展联合评价，以及管理层的接受度不高，缺乏后续行动。

121. 能促进中央评价办公室之间开展联合评价的一个重要因素是就交叉问题进行联合方案编制。联合方案是载于一个共同工作计划和相关预算里的一整套活动，涉及联合国两个或更多组织和国家以下一级的伙伴。⁵⁶ 性别平等、粮食安全和保护生计，以及体面工作就是开展联合评价的例子(见辅助材料，附件 26)。⁵⁷ 从组织的中央一级联合评价得到的经验教训可以在推动评价一致性议程方面发挥重要作用。

122. 除了联合方案，联发援框架本身作为协调机制同样应该增加对国家方案的联合评价，包括由中央职能开展的评价。有证据表明这一点尚未实现。国家一级的分散评价一般不会纳入联发援框架。同样，在(a) 分散评价、(b) 联发援框架评价和(c) 由中央评价部门开展的战略性国家方案评价之间很少有或根本就没有协调。检查专员们仅发现一个能说明开发署作出努力要试行这类协调评价的例子。据报告，管理这类协调是非常困难的，要使这种联合评价获得成功政府就必须发挥领导作用。建议联合国系统各组织行政首长充分支持其评价办公室积极参与联合评价。

结论 12: 调查结果表明，评价职能并没有达到本应达到的强健程度，而这种强健程度对于将评价职能置于有效战略位置以支持联合国系统而言必不可少。经评估，6 个最成熟的中央职能在满足不断变化的要求方面达到了“适当”水平。对于其余 18 个组织而言，它们仍然没有能力抓住和利用不断变化的全球情况带来的机会，以加强各组织之间的协作、创新和工作，从而加强其相关性。

⁵⁶ 联合国发展集团，《关于联合方案编制的指导说明》，2003 年 12 月。

⁵⁷ 见 www.unjiu.org/Pages/default.aspx。

123. 以下是评价部门负责人就关于组织反思和今后向前推进的计划的问题所作的开放式评论：“……在思考和前景方面要有战略性；注重战略方向的制定；在应对变化方面要有弹性；在加强联合国的相关性方面要发挥促进作用，而不是被动反应；要有创新精神，不断调整方法、方式和优先事项，针对快速变化要不断完善和更新；参与全球平台并加强伙伴关系，包括与目前正在推动知识议程的分析中心和全球智库的伙伴关系；以及以灵活方式开展合作”。

124. 全系统评价负责人和工作人员所述的这些特征对于评价职能在支持联合国系统方面所要发挥的作用至关重要，因为评价职能力求在全球变化和对可持续发展提出各种要求的情况下找到一种比较优势和一条前进的道路。研究报告通过成熟度矩阵的各种类目审查了这些特征。其中一些类目与在业务上界定学习型组织所用的那些类目是一样的，因为两者之间存在密切关系。下文表 12 显示了用于评估评价职能的相关性、回应性和准备状态的成熟度矩阵里的关键层面和指标。表内的数字显示各组织的平均分数，表明在支持联合国系统方面的准备水平。总体而言，在应对变化和挑战方面的准备水平或能力低于一般水平。平均而言，经评估，只有 6 个组织的评价职能已在应对挑战方面准备就绪。它们是劳工组织、儿基会、开发署、工发组织、妇女署和粮食署。经评估，它们被认为具有必要的前瞻性重点，并明确打算要应对全球变化和新出现的任务。但是，还有许多组织处于一般准备水平，尽管它们正开始重视加强评价的覆盖范围和质量。

表 12

用于评估相关性和准备状态的次级构成部分和指标

构成部分 5：在加强对变化和挑战的相关性、回应性、适应性和准备状态方面的能力(4.8 分)(最高平均分是 10 分)

次级构成部分

在思考、方法和业务方式方面的战略性(平均得分 4.9 分)

在覆盖本组织各方案和项目方面的相关性(评价规划和覆盖)：有一个立足于本组织和整个联合国系统的评价职能愿景和战略、领导力、成员国的作用及对评价的增加值的了解和欣赏、对评价的战略性利用和利用所产生的影响力

创新、适应性和更新/灵活性和复原力(平均得分 5.3 分)

适应性及不断改善和成长——动态职能：平衡各种要求：持续评估和适应；努力促进发展评价和发展对于联合国系统所采用的那类干预措施有效的适当方法

一致性、包容性、自主权和协调(平均得分 4.4 分)

对支持一致性、包容性、自主权和可持续性的回应性，重点是：(a) 分散评价；(b) 联合国改革以促进一致性，包括开展联合评价；(c) 性别平等与人权；(d) 国家评价能力发展和与国家系统的协调及加强自主权

应对全球变化和挑战(平均得分 4.7 分)

应对在评价方面的全球变化和挑战，参与全球平台和全球治理问题，包括共享知识和指导变革：发展学习型组织和不同思维文化、创新和承担在组织转型方面的风险；积极参与联合国评价小组的工作，发挥主动作用以促进一致性，并作为不仅是各组织而且是联合国系统和全球性的促进变革的推动力

125. 尽管目前的准备状态不尽人意，但评价和方案管理人员表示，他们非常明确打算让评价职能以替代和更具活力的模式开展工作：⁵⁸

“作为一个重要的变革推动者：参与全部门和其他更具战略性的评价；加强分散评价的制度化，以使其为增加价值发挥补充作用；有共同共享的系统(共同的外部咨询人信息数据库、用于评估报告质量的共同系统、共同风险框架)；一起工作，限制各自为政而使各组织各干各的做法；在评价员之间进行直接同行交流；执行全系统评价政策、协调和联合规划”。⁵⁹

126. 这些想法表明，一个动态进程已经开始了。它们回应了关于 2015 年后发展议程的辩论所产生的那些原则：⁶⁰ 独立性和相互联系；跨越界线的合作；做符合目的的事情、增加价值和加强联系，包括国家能力。它们反映了一种新的运作模式；这种模式更灵活，但需要所有各方的协作和行动。

127. 但是，为确保评价职能具有这些特征而作出的回应不能是临时性的。它要求(a) 有一个明确界定的优先事项战略；(b) 在各个方面进行合作，这就需要打破限制，同时在各组织之间开展工作以推动在转型和相互依存方面的变革；以及(c) 制定新的激励制度。评价负责人和管理人员指出，这种回应不限于在联合国评价小组领导下开展的协调和统一活动。必须有充足的人力和财政资源，还必须为不同思维和替代方法的探索提供智力空间、承诺和时间。

128. 上文第 14 段提及的对发展方面业务活动的全系统独立评价为联合国系统各组织与联检组协作提供了一个机会，因为联检组承担了促进全系统独立评价的

⁵⁸ 见对评价负责人和其他管理人员就成熟度矩阵相关开放问题所作答复作的内容分析结果。

⁵⁹ 评价部门负责人和工作人员对关于评价在迈向未来时的方向的思考所提出的开放问题的答复。

⁶⁰ 2015 年后发展议程高级别小组，“一种新的全球伙伴关系：通过可持续发展消除贫穷并使经济转型”(联合国，2013 年)。

任务。以现有系统、机制和能力为基础制订的全系统独立评价政策为对发展方面业务活动的全系统评价提供了一个伙伴关系框架。该政策的评价原则——辅助性、经济和效率——为全系统各层级之间的协调和协作提供了明确基础，包括有效利用现有评价库。评价模式包括(a) 利用在中央和分散一级形成的现有知识体系进行全系统一级的元合成；和(b) 不同的中央部门之间进行协调和协作，在一个共同的框架下开展工作，围绕在组织和全联合国系统一级均可加以利用的专题开展评价工作。

129. 全系统独立评价政策一旦得到执行就能在以下方面发挥作用：加强协作和效率，以及更全面地说明联合国系统的一致性情况和发展方面业务活动的实效。但由于缺乏资金，为启动两个试点评价所作的努力失败了，尽管大会通过了要求提供额外预算支持的决议。⁶¹ 增强营养倡议成功地使包括私营部门伙伴在内的广泛伙伴参与其中，也许可以借鉴这方面的经验来为全系统独立评价筹资。

130. 执行以下建议将会加强一致性和统一性。

建议 7

联合国系统各组织行政首长应要求评价办公室重新评估评价职能的政策、战略和优先事项。它们应把评价职能置于各自组织的战略位置，从而加强其相关性，以增强联合国系统在应对当前变化和挑战及在实现影响力和可持续性方面的能力。

国家评价能力促进联合国系统评价职能的可持续性

结论 13：大会关于国家评价能力发展的各项决议可追溯到 1989 年。联合国系统各组织认识到这类能力发展的价值，并认为与国家评价系统的协调，包括在数据和统计方面的协调，不仅对各国而且对联合国系统评价职能的可持续发展至关重要。目前在支持国家评价能力发展包括加强评价结论自主权方面的参与和努力水平是初级的。许多组织并不认为自己承担着促进国家评价能力发展的任务。另有一些组织不认为这是评价职能的一种责任，而认为是一种方案责任。将国家评价能力发展列入其评价政策的组织很少。各评价部门所采用的战略重视在进行中央评价时利用国家咨询服务。有 3 个评价部门(开发署、儿基会和妇女署的评价部门)是例外：它们在其评价政策中规定了这样的任务，提出了愿景和战略或计划，并开展计划好的活动。⁶² 总之，目前整个联合国系统的做法并不是根据一个综合框架运作的，并且尚需就能力发展和国家自主权作为一个内部进程达成共识。

⁶¹ A/RES/68/229。

⁶² 妇女署在 2014 年也制定了一个专门针对国家评价能力及全球外联和伙伴关系问题的全球评价战略。

131. 加强国家评价能力不仅对各国具有价值，而且也是联合国系统评价职能的一个关键成功因素。在审查评价职能如何看待协调其为加强国家评价能力所作的努力时，这是一个业务前提。检查专员们充分了解关于以下问题的辩论：国家评价能力发展在某种程度上是一种方案努力，应与项目周期和管理及与旨在加强治理和公共部门管理的干预措施挂钩。联检组 1982 年的报告⁶³ 审查了联合国 23 个机构及国际复兴开发银行、经合组织和欧洲联盟委员会所作的各种方案努力，重点指出了已采取的行动和面临的挑战。联检组今后在研究中值得作出努力，评估各组织在这方面所取得的进展。

132. 方案部门承认自己的能力有限，希望评价界支持发展国家评价能力。联合国系统的评价员承认国家评价能力在促进评价职能的影响力和可持续性方面所具有的战略重要性。根据许多评价部门负责人的看法，联合国评价职能与国家系统和能力发展的协调有多方面的重要性：(a) 对成果的相互问责；(b) 纳入各种知识体系，以提高有效性；(c) 参与规划和管理，以提高利用率和影响；(d) 参与以加强统一和相互能力发展；以及(e) 与实地的现有系统和机构联系，以加强联合国和各国评价职能的效率和可持续性，在目前情况下这是至关重要的。那些确认自己承担着促进国家评价能力发展任务的人和那些表示自己没有这项任务但认为国家评价能力发展对于联合国系统评价职能的效用、影响力和可持续性具有重要性的人都表示了这类看法。因此，没有人反对评估其为支持国家评价能力发展所作的努力。

133. 分析报告上来的信息后发现，中央评价职能与国家系统和国家评价能力发展之间的联系加强了。目前对 8 个组织而言，与国家系统的协调不在考虑之内。其他 16 个组织则不同程度地在接触国家专家和机构。对一些组织来说，所开展的活动和各组织之间的活动协调通常都缺乏战略指导。参与的主要方式是，让国家咨询人或机构参加所进行的评价。在大多数情况下，这样做的目的是要加强在利用评价效应方面的某种自主权，但同时也为了通过学习、联系和参与加强能力。这种努力对于相互能力发展的影响并不是微不足道的。但需要以更系统性和较少临时性的方式做更多的工作。

134. 如上文所述，有 2 个组织(开发署和儿基会)把国家评价能力发展作为中央评价职能业务的一个组成部分。虽然这些组织在促进国家能力发展方面所开展的活动一般是值得赞扬的，但在拟订这类活动时未参照一个能加强协作和利用彼此在支持国家评价能力发展方面的比较价值和增加值的框架。中央职能的工作与其方案部门的工作的协调同样是现在才出现的一个战略组成部分。最近在 2014 年，妇女署将国家评价能力发展作为其全球评价战略的一个战略领域。全球战略在扩大与联合国各机构、国际金融机构和私人基金会的伙伴关系方面的做法是值得赞扬的。

⁶³ 联合国系统在各政府发展评价工作方面的合作，JIU/REP/82/12。

135. 从国家评价能力发展方面的成功中能吸取哪些经验教训呢？一个最重要的成功条件是把对国家评价能力发展的支持置于国家现实之中，以及利用本土知识和问责制。⁶⁴ 旨在为采用联合国评价小组提供的关于国家评价能力发展问题的最新导则而作出的努力中必须采取这种方法。⁶⁵ 开发署的模式⁶⁶ 是以国家的内部能力和领导作用作为基础的，被公认为是一个用于支持国家在加强能力方面的自主权的模范方法。在评价这类方法时遭遇很多困难，因为这种方法见效慢、形式多变，并需要对发展进程、倒退类型和所取得的重大进展作认真的人种学研究。

136. 在加强国家能力方面的其他重要考虑包括：利用非传统和私营部门机制，例如利用风险资本机制与国家 and 区域高级别专家合作以建立能力。同样重要的是，与现有的国家和地区知识发展机构合作，例如大学、分析中心和智库，包括那些通过世界银行关于促进评价和成果学习的区域中心项目发展起来的机构。这些建议没有一个是新建议，因为联检组自 1982 年以来一直在提出这类建议。

137. 人们还认识到，要使国家评价能力发展取得成功就必须采取更整体的方法，从方案设计阶段就开始直到执行、监测和最后的评价。在这方面，虽然联合国各组织的评价职能确实可以在加强能力方面发挥关键作用，但方案部门在评价能力发展方面的作用同样重要。评价是各国治理和公共部门管理的组成部分，因此应成为这些领域里一个重要的支助构成部分。⁶⁷

138. 自 1989 年以来，联合国通过了好几项呼吁促进国家评价能力发展的决议。⁶⁸ 一直在作出努力，但都是零散的努力，相关结果也不为人所知。现阶段需要联合国系统各组织采取一系列深思熟虑的干预措施，以促进在加强国家评价能力方面的协作和伙伴关系，并为这项工作提供必要的框架、资源和激励机制。在采取这个步骤之前必须就联合国系统为加强国家评价能力所作的整体努力的现状作一次经过适当设计的研究。可以在联检组的领导下开展这项研究，作为对以往所作工作的延续。将要作出的努力应与 2015 年的评价工作挂钩，因为其重点是国家评价能力。

⁶⁴ Sukai Prom-Jackson, “非洲评价职能的重生、恢复、改造和责任”，载于《评价问题：我们如何加强国家评价系统——来自全球各地的看法》，《非洲开发银行知识季刊》，第 2 卷，第 3 号（2013 年 9 月），第 76-81 页。

⁶⁵ 联合国评价小组，国家评价能力发展：关于如何加强国家评价系统的实用技巧(2012 年)。

⁶⁶ 开发署关于国家评价能力的国际会议(2009 年和 2013 年)。

⁶⁷ 又见，联合国系统在各政府发展评价工作方面的合作，JIU/REP/82/12；和 Sukai Prom-Jackson, “非洲评价职能的重生、恢复、改造和责任”，载于《评价问题：我们如何加强国家评价系统——来自全球各地的看法》，《非洲开发银行知识季刊》，第 2 卷，第 3 号(2013 年 9 月)，第 76-81 页。

⁶⁸ 大会关于联合国系统发展方面的业务活动三年期全面政策审查的第 44/211、第 59/250 和第 62/208 号决议，以及关于四年期全面政策审查的第 67/226 号决议。

139. 执行以下建议，同时考虑到上文提到的经验教训，将加强评价职能的可持续性和与该职能的协作。

建议 8

秘书长应以首协会主席的身份采取步骤并支持创新，以促进联合国系统各组织之间及与其他伙伴之间的协作，加强国家在涉及问责制、学习及具有国家和全球价值的知识发展方面的评价能力。

三. 分散评价职能

A. 引言

140. 引言部分第 24 和第 25 段对分散评价所作定义适用于本研究报告的目的。附件五对此有详细阐述。辅助材料中的附件 27 同样概述了切合目的的概念框架，⁶⁹ 其重点是分散评价所产生的价值和其作为嵌入管理的一项职能所具有的某些特点。本研究报告利用这一框架来回答就分散评价职能提出的下列主要问题。

1. 要求

谁是要求的主要驱动方？对分散评价职能的日益关注和要求有哪些特点？如何使分散评价职能处于能应对任何日益关注/要求的地位？

2. 目的、比较价值和规范性框架

联合国系统分散评价职能的性质是什么？该职能的目的是什么？在联合国系统架构中发挥着什么作用？该职能有什么显著特点、规范性框架和标准？该职能如何在联合国评价系统中增加价值？

3. 结构安排

分散评价如何在结构安排上支持以下层级的决策：(a) 项目一级和(b) 机构一级，作为明确界定的部级业务计划的组成部分。评价设在何处：总部、区域办公室、国家办公室还是技术部门？

4. 体制框架和支助系统

分散评价职能的体制框架和支助系统的性质是什么：(a) 评价政策、战略、政策的执行；(b) 政策执行支助系统，以及管理、实施和质量保证；(c) 资源和资金；以及(d) 类似于联合国评价小组的专业网络？

⁶⁹ 见 www.unjiu.org/en/Pages/default.aspx。

5. 业绩

分散评价职能的业绩有哪些特点：(a) 重点和覆盖范围；(b) 一致性和协调；(c) 为加强公正性而设立的机制；(d) 报告的质量及质量保证制度和能力；以及(e) 效用-为加强利用、利用水平、利用的性质和利用所产生的效应而设立的条件？

6. 相关性和前进战略定位

在应对新出现的变化和要求方面具有何种准备或能力水平？

B. 体制和规范性框架

结论 14：目前，联合国全系统一级的分散评价未得到明确界定的体制框架、制度和政策的充分支持。联检组 28 个参与组织中的大多数组织，如果不是全部的话，在进行分散评价，而这类评价可分成两个不同的类型。对联检组 11 个参与组织而言，分散评价在不同程度上受到本组织的评价制度、政策和程序的影响。对其余组织而言，则不是这种情况。

141. 在说明分散职能的这两种类型或模式之前，先简要说明一下分散职能的演变情况。早些年间，联合国评价系统的特点是，各组织主要进行自我评价以向项目和方案提供支持。这在很大程度上是因为需要改善发展方面合作活动中的具体项目。值得注意的是，当时对评价所作的定义同样反映了该职能的这一目标。⁷⁰ 当时评价的主要目标是：覆盖各组织的全部活动；就新出现的成果快速和持续地提供信息反馈，并利用评价来迅速调整目标和战略、改善今后的活动；将评价作为管理周期的一个正常组成部分和一个参与性进程，以提高工作人员的坚持和参与度，促进发展成为一个学习型组织。⁷¹

142. 同时还假定，如果由负责项目的管理人员来制订，评价结果就更可能被接受并就此采取行动。此外，在纳入正常的管理流程后，与更详细和复杂的评价系统相比，就减少了“附加费用”。

143. 人们在此期间认识到的缺点是，由无评价专业背景的工作人员所作的评价的质量和工作人员是否能客观地评价自己的工作的问题。因此，随着在 2000 年代对具有独立性和专业质量的问责制的要求日益增多和要评价对组织一级决策提供支持的要求日益增多，内在自我评价(现称作分散评价以与中央评价作区分)在评价职能总体架构中的角色就不那么明显了。重点转向了中央评价职能。本研究报告得出的证据表明，越来越需要平衡作为一个综合系统组成部分的这两个职能，以应对发展格局中的要求和新出现的优先事项。

⁷⁰ 如欲查看评价术语词汇表，请见 JIU/REP/78/5。又见，JIU/REP/91/6。

⁷¹ 关于联合国系统内评价工作的第二次报告，JIU/REP/81/6。

联合国系统内分散评价职能的模式和原型

144. 对分散评价各种原型的分析结果显示有 6 个不同类型。并非所有类型都是相互排斥的，有些类型则不清楚谁来负责在中央评价职能控制下产出的报告质量。可以看到，各组织在同一时间在执行这些原型中的任何两个原型。附件五(b)提供了这 6 个原型的详细情况。还有第七个原型，即由为项目供资的捐助方管理和实施的评价。该原型未列入图表。总体而言，这 6 个类型可分为本研究报告所阐述的两种模式：(a) 临时和自行酌定的分散评价；和(b) 分散的，但有一个体制框架，根据明确界定的覆盖范围标准规划评价并委托给外方。下文将阐明这两种类型。

145. 临时和自行酌定的分散评价。在联合国系统 27 个组织内仍然存在着由担任项目管理工作人员的工作人员在其业务主管批准下以自行酌定方式开展、其结果得到业务主管利用并作为方案管理职能组成部分运作的自我评价。⁷² 这类评价充其量只能说是临时和自行酌定的分散评价。它们通常没有战略选择方面的指导，是在工作人员和管理人员酌定的情况下启动的。一些组织的评价政策承认它们是自我评价，但在大多数情况下，这类评价在运作过程中得不到任何具体体制框架的指导。在某些情况下，它们被视作是成果管理制的组成部分，但并不是以制度的方式得到这种承认的。很难评估以这种临时分散评价为基础的证据，因为各组织评价和监督系统不是以一致方式报告或获取这类评价的结果的。这类评价结果是联合国系统内一个巨大的知识库，有助于本组织“创造一种把评价视作是学习工具的文化”。⁷³ 这种评价和学习文化对于联合国组织的发展和实效而言至关重要。但是，联合国系统正在错失调动和利用这类知识资产并将权能赋予参与这类评价的工作人员的机会。

146. 从报告中获得的信息表明，为满足向特别项目和方案提供预算外资金的捐助方的要求，目前正越来越多地在开展这类评价。因此，首要重点已变成是支持问责制，从而有别于上文第 141 段概述的形成性评价最初的学习和促进改善项目的目的。鉴于其重要性，加强分散评价系统必须基于：扎实地了解其目的、作用和对联合国系统的比较增加值；能加强其形成性作用的标准；能加强而不是“削弱”工作人员广泛参与的制度(鉴于工作人员可能不具备《联合国评价小组联合国系统评价员的核心能力》所界定的专业技能)；在整个联合国系统实施关于评价的适当的培训和学习方案；以及发展或编纂形成性评价的方法以满足在持续改善项目方面的需求。

147. 在支持广泛制度化、包括发展一种评价文化和统一评价方法方面，联合国评价小组曾为联合国系统的工作人员和伙伴启动了一个全系统面对面培训方案。由于联合国评价小组的能力有限，这一方案已经中断了。但是，联合国各组织内

⁷² 项目厅不开展评价工作。

⁷³ 联合国系统内的问责制框架，JIU/REP/2011/5。

正在进行广泛的评价培训。联合国系统内需要整合关于各种培训的信息和关于评价的指导材料，共享在良好做法方面的经验教训。个别和全系统机构(联合国职员学院)提供的这类培训必须得到从联合国系统评价经验中发展出来的专家知识的支持，以使这类培训具有相关性。

148. 各组织关于评价的年度报告必须述及并报告所有分散评价，以强调联合国的知识资产和参与建设学习和评价文化的广大工作人员。应发展一个联合国各组织之间的协作制度，以加强工作人员在管理和开展各类评价方面的能力。

149. 有体制框架的分散评价。另一个类型就是受到有明确规则的明确界定的体制框架影响的分散评价，有一套激励和宣传制度，以支持对这类评价的规划、执行和报告采用一种一致和战略性的方法。用于这类分散职能的框架有如下特点，但需适当考虑到各组织之间在范围方面的差异：(a) 现行评价政策界定了分散评价职能的作用；(b) 由各部厅和国家办事处负责管理评价工作；(c) 这类评价是有规划的；(d) 它们根据一整套标准进行运作；(e) 有制度性的质量保证或控制和质量提高机制；以及(f) 会就遵守政策和该职能发展的情况提交报告。中央评价办公室参与这类职能的程度不一，但在所有情况下，这类职能是在业务管理部门领导下运作的。中央职能提供了各种类型的质量提高和/或质量控制服务。关于分散评价职能模式的附件五(b)对此作了概述。

150. 在联检组 28 个参与组织中有 11 个组织(劳工组织、艾滋病署、开发署、教科文组织、人口基金、儿基会、近东救济工程处、妇女署、世卫组织、粮食署和知识产权组织)有这类框架，但程度不同。⁷⁴ 在这些组织中有 10 个组织(除艾滋病署外)对调查表作了答复并提供了可供分析的数据。对来自这 10 个组织的数据所作分析的结果产生了以下结果、结论和建议。

151. 这 10 个组织的各方案办公室，包括技术或方案部、专题组和在总部的区域办公室，把分散评价委托出去，在区域中心和 150 多个国家办事处或技术中心加以实施。在这些组织中的某些组织里，明显看到最近对分散评价职能有所加强，各管理部门和评价部门新近采取的一系列举措表明，现在越来越重视分散评价和思考该职能能为各组织增加何种价值(见联检组网站上的辅助材料)。⁷⁵

结论 15: 整个分散评价系统缺失一个切合目的并能加强有效管理和决策的明确界定和适当的规范性框架和标准。

152. 诚然，联合国评价小组相当数量的规范和标准是适用的，但它们未得到发展，同时应铭记分散评价作为业务管理工作和形成性评价系统的组成部分所具有的目的和性质及其作为各组织评价系统组成部分所具有的比较增加值。联合国评

⁷⁴ 知识产权组织提供的数据有限，但仍然被列入分析。

⁷⁵ 见 www.unjiu.org/en/Pages/default.aspx。

价专家区域评价网络就分散评价问题开展的工作可包括发展切合分散评价目的规范、适当估值和标准，作为对联合国评价小组规范和标准的补充。联合国亚洲及太平洋评价发展工作组是成立时间最长的网络，过去 8 年来一直在运作。加强对这类评价团体的支持和联系对于发展切合分散职能目的的规范、标准和方法而言十分重要。行政首长应支持为分散评价职能和切合目的的方法发展适当的规范性框架和标准，并提供适当的激励机制以将权能赋予广大参与评价和学习的工作人员。

C. 政策、执行和制度化

结论 16：在作为本项目研究对象的 10 个组织中，分散评价政策的发展在指导分散职能方面一直起着重要作用。这些组织已采取了大量措施来支持执行政策。各组织之间执行政策和制度化的程度各不相同。在制订政策方面缺失的两个关键环节是如何加强联合国各组织之间的一致性和如何与国家系统协调。因此，各组织的分散职能在这两个领域里的发展水平很低。如前几章所述，这两个领域在中央评价一级反映出类似的缺陷。

153. 在所审查的支持分散评价的 10 个组织中有 8 个组织的评价政策详细阐述了分散评价的作用和对分散评价的支持，但对联合国系统内的一致性及与国家系统和能力的协调问题的叙述则很有限。在所审查的 10 个组织里，所有层级的分散评价都是委托进行的，虽然各组织之间的程度不同，这取决于相关组织的结构和分散的程度。对这些组织评价政策的审查结果显示，相关政策全面而详细地阐述了分散评价的作用和对分散评价的支持。涵盖的问题包括：(a) 分散评价在各组织内的重要性、增加值、作用和目标；(b) 为分散评价的规划、管理和筹资所作的安排；以及(c) 一些被认为对效率和可持续性而言至关重要的协调和协同作用。

154. 未得到中央评价政策重视的领域是：(a) 联合国系统各组织之间在分散评价方面的协调和一致性，以及组织委托的分散评价与联发援框架的评价之间的关联；和(b) 分散评价在支持发展国家评价能力方面的作用。检查专员们认为，这些领域里的现有制度是最弱的；这证明，评价政策在界定完成何种任务方面具有重要性。行政首长应确保各自组织的评价政策充分涉及联合国的一致性改革和各组织之间的协作模式，并充分阐述评价职能在支持国家评价能力发展方面的作用或其他能促进该职能实效和影响力的重要优先事项领域。

155. 在加强执行政策方面，在作为研究对象的 10 个有分散系统的组织中，有 8 个组织投资于支持分散评价，主要通过制订导则和提供技术支持。此外，它们还开展了一系列广泛的支持活动，其中有如下突出的活动：提高认识、举办会议和讲习班、为发展共同体制基础而采取特别举措、通过手册促进能力发展、质量评估、培训及发展实践社区或学习团体和网络。据报告，两个最有效的执行政策的形式是联系本组织的领导层和向执行局报告分散职能的进展情况，从而使分散评价发挥作用，并使其提出的问题在决策时受到关注和得到机构认可。

156. 在支持执行分散职能政策方面最有价值的是中央评价职能，特别是开发署、儿基会、人口基金和妇女署的中央评价职能。它们为分散评价制定了相关标准、导则和业务程序。检查专员们还了解到，联合国区域评价网络的作用已变得越来越重要。一个堪称典范的做法就是联合国亚洲及太平洋评价发展工作组组办的联合培训方案；该方案在发展执行政策的重要能力方面发挥了作用。

结论 17：相关评价政策的制度化程度各不相同，几乎没有任何中央评价职能就如何支持分散评价或就期望从分散评价系统的业绩中得到何种成果制订了明确的战略。

157. 在作为研究对象的 10 个组织中，那些在与分散评价相关政策的制度化方面领先的组织取得了不同程度的成功。根据对来自关于分散评价的调查表的数据所作的分析，5 个组织在政策的制度化和在用于执行政策的制度到位方面达到了较高水平；它们是劳工组织、开发署、人口活动基金、儿基会和妇女署。

158. 用于支持制度化的一系列机制已经到位，它们包括：指导方针、手册和指南、咨询人专家名册、评价数据库或资源平台、后续回应系统和评价网站。各组织的这类机制在支持制度化方面的程度不同。现有的专家咨询人名册的用处似乎有限，但据报告很有必要，理由是：委托开展评价的基础是利用具有不同背景的独立专家咨询人和语言群体来实施评价。因此，现有用于支持分散评价的名册上的评价员和专家的人数和类型有限，这对分散职能的公正性和公信力构成了重大问题。对该职能的供资有限也意味着无法负担某些咨询人。但据报告说，总的来说，在咨询市场并没有太多合格的评价专家。

159. 各组织最近为支持制度化推出了新的举措。妇女署推出了一个在线管理人员评价指南并在全球各地对数百名评价员展开了关于顾及性别平等的培训。开发署为工作人员和管理人员开发了网上培训课程，并重点发展与分散职能和与国家系统的战略联盟。儿基会一直在评价培训方面发挥领头作用。儿基会目前有两个在线培训课程，同时还支持开发了一个在线培训课程，现在已通过 EvalPartner 网站(MYM&E)推出了这个课程。劳工组织有一个管理人员认证方案。开发署、教科文组织、人口基金、儿基会、妇女署和世卫组织在发展质量保证制度和执行情况追踪系统方面采用了各种创新做法。国际民航组织、教科文组织和近东救济工程处已开始展开编制所有现有自我评价情况和/或元分析的进程。世卫组织作为组织改革组成部分在 2013 年开始的一个全组织范围最大的举措是发展了全球评价网络。

160. 粮食署中央评价职能在 2014 年作出了一个有趣的努力：在经过多年与分散评价职能的有限关联后，中央评价职能正在发展一种模式，通过对管理和质量保证采用创新性方法来支持分散评价，将分散评价临时置于中央评价部门的管控之下，直到发展出一个有利于分散评价的全面框架和能力发展。这一规划中的由中央评价管控的方法具有一些与妇女署所采用的模式及与粮农组织、环境署和工发

组织⁷⁶ 现行做法和新出现的做法相同的要素。在这些组织中，在质量控制的程度和类型及在中央或分散一级对最后评价报告的利用方面存在着差异。需要作进一步分析，以说明在中央和分散一级与组织结构、发展和能力水平相关的新出现的一些分散评价的模式。

161. 各组织还指出了对制度化的各种关键挑战，包括：资金不足、高级管理层的支持不足、评价文化薄弱，以及监测、数据系统和方案的可评价性都差强人意。如上文所述，中央职能同样面临其中一些挑战。下一个章节将说明在分散一级得到的调查结果。

D. 资源

结论 18: 对大多数组织来说，关于预算分配或分配的规范性基础的信息未得到充分界定或透明度不够。支出被认为是不够的，与双边机构分散评价的供资额相比，支出额非常低。为提高管理效率提出的建议包括：跨越组织界限的新工作方法和以相互关联的方式开展工作，包括在联发援框架范围内界定评价计划。

162. 在回答关于在加强分散评价职能方面的挑战问题时，10 个组织中有 9 个组织强调，资金不足是最大的障碍之一。7 个组织的评价政策确定了用于分散评价的预算分配性质。对 4 个组织而言，资金来源一般是项目或方案预算的一个百分比。很难获得关于向分散评价提供的实际数额的信息：在 10 个组织中有 4 个组织提供了关于专用于分散评价的具体数额的信息。其他组织的预算与监测和评估相关的其他职能混在一起。各组织为每次评价工作平均支付 35 000 美元(可变费用不包括工作人员的时间)，支付范围在 20 000 至 55 000 美元之间。虽然所支付的数额一般比用于联发援框架评价的数额要高(每次联发援框架评价的费用是 30 000 美元)，并且联发援框架评价的范围要广泛得多，但与双边捐助机构专用于分散评价的支出相比则不算高。⁷⁷ 这就引起了对评价质量的质疑。

163. 虽然据报告说资金不足和需要更多资金，但一些组织作为优先事项提出了管理费用的新方法，包括一个利用工作人员资源和各组织就评价系统和机制开展合作的综合战略。

- 开发署指出，尚有提高效率和整合资源的余地，特别是在国家一级或在处理共同专题时，创造更大的成果及改进评价规划和一致性，以及一起建设专家能力。联发援框架提供了一个可以这样做的机会，只要在“一个方案”的框架和联发援框架原则的范围内作适当构想和规划。

⁷⁶ 这 3 个组织都表示，它们没有分散评价职能：评价工作由项目管理人员启动，在外地展开，但由中央部门的一个评价单位管理或作质量保证，报告由该中央部门分发，因此被认为是中央评价。需要就各种模式作进一步澄清。

⁷⁷ 见联合国国际发展部和联合国援助机构，“对联合国国际发展部嵌入式评价的快速审查：最后报告”（2014 年 2 月）。

- 国际民航组织和近东救济工程处提出，必须利用每一个组织编写的许多准则并编制一份工具概要供各组织使用。
- 其他组织建议，有必要审查所有现行质量保证和追踪系统，以找出共同和各组织特有的经验教训，并制订一个共同框架。
- 毒品和犯罪问题办公室建议，分散的和小型组织应利用共同机构来执行对评价报告的所有质量评估，这种做法与联合王国国际发展部的经验相似。
- 其他组织建议加强和利用国家系统，同时要加强它们的能力。

164. 各组织就这类问题建立一个论坛，以促进分散评价职能之间的协调和协作并使评价部门的代表及分散评价工作的管理人员和行政人员参与进来，这将为启动分散评价一级的协作提供一个重要平台。联合国各组织的行政首长应在发展能尊重联合国一致性改革议程、限制交易成本和加强国家自主权的分散评价职能方面加强协作和提高效率。

E. 公信力(公正性和质量)和效用

公正

结论 19：在加强分散职能的公正性方面取得了进展。在 10 个组织中有 8 个组织已经建立了一系列制度和机制。这些组织利用了 9 个机制，以限制偏见和管理层的控制、提高透明度和加强主要利益攸关方的信任。要满足对该职能问责制要求的期望就必须这样做。

165. 分散评价的公信力一直受人关注，因为评价是在管理部门内部规划、管理和实施并由管理人员批准或通过管理人员提交报告，因此有人说缺乏独立性。考虑到分散评价结构的性质，自然不符合结构或职能独立性标准。但是，各组织为限制管理层对证据的偏见已采取了各种为加强分散评价的公正性、公开性和透明度的措施和机制。以下是优先重点：(a) 利用外部咨询人；(b) 利益攸关方的参与；(c) 利用评价管理小组或评价委员会；以及(d) 实施道德操守行为守则和行为独立性准则。这些机制在加强公信力、透明度和自主权方面发挥了关键作用。

166. 对许多组织而言，在分散评价中利用了外部咨询人就表明，评价是独立的。在大量已知案例中，这一机制没有确保独立性，因为最后的报告是要委托编写该报告的管理人员批准的。在许多情况下，一些双边组织力求利用自己的独立手段来评价其供资的项目。一些组织现在由中央办公室做加强和控制质量的工作并确保评价工作符合独立性质量标准。妇女署采用的就是这个模式。毫无疑问，这是因为对问责制的重视才推动在分散一级作出广泛努力以保持独立性。要使分散评价切合目的就必须开始重新考虑该职能首先在支持开展形成性评价以改善并作必要纠正从而最终实现成果方面的重要性。联合国系统在制订战略和排定评价职能作用的优先次序时需要考虑这个问题。

质量

结论 20:在联检组 10 个参与组织中有 8 个组织对加强评价技能能力进行了投资,以支持提高分散评价的质量。但还需要做更多的工作,采用能在合理程度上顾及技术标准和有效管理要求的更客观和更系统性的方法来加强评价报告的透明度和质量。

167. 在加强分散评价质量方面的一个关键重点领域一直是,要在评价工作中雇用专家作为工作人员以提高质量。就 10 个接受审查的组织而言,在总部及在区域和国家一级都有监测和评价专家。近年来他们的人数一直在增加(开发署、儿基会、人口基金、妇女署和粮食署),虽然有人说人数仍然不够。但在许多情况下,这些专家在评价工作上所花费的时间不到其总工作时间的 25%。他们的大部分时间反而用于其他项目周期管理工作:规划、监测、审查和其他评估。虽然让这类专家来支持分散和中央职能的系统发展及加强成果管理制系统和可评价性是有道理的,但是在这些领域里的过度使用就会影响对他们作为评价专家在加强分散评价职能及其发展方面的有效利用。鉴于新出现的变化和对分散职能的要求,必须重新考虑监测和评价专家的人数和对他们的有效利用,以支持核心分散评价。

168. 另一方面,妇女署采用了不同的模式:支持分散评价的区域评价专家根据共同问责制框架的规定同时向评价办公室主任和区域办公室主任报告工作并要发挥提高分散评价质量这一关键作用。这种模式的好处在于把专家的重点放在评价工作上。这种联合问责制制度的实效尚需在妇女署得到验证。

结论 21:平均而言,根据作为研究对象的 10 个组织所提供的信息,这些组织的绝大多数分散评价已得到改善,⁷⁸但仍低于一般质量水平,需要在制度发展方面做更多的工作,以在全系统范围内加强发展关于分散评价质量的扎实信息。

169. 中央评价部门开展了一系列旨在支持改善分散评价的活动。总体而言,对这些活动最宝贵的支持就是评估评价报告的质量、就需要改善之处提供的反馈和作为关于评价工作年度报告组成部分报告组织一级的质量。

170. 接受审查的 10 个组织每年产出大约 640 份评价报告,涉及产出、成果、影响力、复杂性和其他评价。在联检组 10 个参与组织中有 7 个组织在过去 3 年里对分散评价报告的质量进行了深入的系统性评估。⁷⁹ 已到位的制度有一整套与联合国评价小组评价报告质量标准挂钩的共同标准,但在许多地方有差别,包括质量指标的数量和类型、评估程序和所使用的评分制度。

⁷⁸ 对改善的分析是基于过去 4 年的年度报告和一些特定访谈。

⁷⁹ 劳工组织、妇女署、开发署、教科文组织、人口基金、儿基会和世卫组织。

171. 但对于这些系统为各自组织所服务的目的而言，对 6 个组织所报告的数据作分析后可看出，总体而言，在所评估的评价中，50%至 74%属于高质量。⁸⁰ 这与从其他 3 个没有评估系统的组织来的受访者的看法不同，因为他们认为，分散评价的质量通常低于一般或较差。因此，一个用于对联合国系统分散评价的质量作判断的稳定而全面的基础尚未在所有组织中形成。开发署、儿基会和人口基金发展的现有系统提供了示范性基础，可加以利用促进发展。上文提及的其他替代方法包括利用一个中央来源或机构来支持一些组织开展评估工作。

利用和问责制

结论 22: 同中央职能一样，分散职能对于成果问责制和对于利用评价证据改善项目和方案方面的要求的回应度很高。但尚未在对问责制的要求与发展评价和学习文化及在联合国各机构内部和在国家、区域和全球一级机构之间加强共享评价成果的必要性之间作充分平衡。

172. 本结论与为中央一级作出的结论一模一样。联检组 10 个参与组织发展了一系列制度来提高利用率；这些制度可分为 6 大类，涉及：(a) 提高工作人员和管理人员的参与度；(b) 加强在流程和用于提高利用率的激励机制方面的包容性；(c) 伙伴的参与；(d) 后续追踪系统；(e) 网站；以及(f) 反馈机制。

173. 被认为在提高利用率方面最有用的制度是：有一个可查阅所有评价报告的网站、用于执行评价建议的行动计划、用于追踪后续行动和管理层对建议执行情况的回应的系统、将来自分散评价的证据列入将提交给高级管理层和/或理事机构的年度报告里以提高可见性；使国家一级的利益攸关方参与评价的规划、实施和/或管理以加强包容性和自主权，以及确保与工作人员和管理层进行有意义的协商和他们的参与，同时保证客观性。

174. 在 10 个组织中有 5 个组织已经建立了用于追踪建议执行情况的系统，这被认为对于提高这些组织的利用率至关重要。这些系统揭示，这 5 个组织在执行建议方面的程度是不一样的。据报告取得了不同的利用实效，最重要的是：加强了在资源利用和成果方面的追责能力、方案决策方面的质量、改善了方案和项目、提高了透明度从而加深了信任、加强了捐助方的支持，以及促进了资源调动。

175. 同中央评价职能的调查结果相似，所报告的关于分散职能的信息一致显示，对绝大多数组织来说，在问责制及改善具体方案和项目方面取得的利用实效较高，在发展学习和评价文化方面则要低得多。这一分析结果显示，在国家一级、区域一级或全球各区域的联合国机构之间共享评价结果方面的实效非常低。对问责制方面的要求较大，之后的重点是改善正在实施的方案和项目。对增强工作人员权

⁸⁰ 劳工组织、开发署、教科文组织、妇女署和世卫组织——来自过去 4 年年度报告的证据显示，在此期间质量有所改善。

能使他们成为知识工人方面的要求较少。按重要性递减顺序排列，分散评价的主要用户是捐助国政府，之后是项目管理人员和执行者。

F. 为应对变化和挑战所作准备的状态：不断变化的要求、能力和准备状态

结论 23：接受审查的联检组 10 个参与组织为分散评价设立的现有制度为便利应对全球变化产生的要求和特别是来自 2015 年后发展议程的新出现的需要提供了一个机会。当分散评价根据明确界定和一致的愿景和战略开始应对在方法、优先事项和一致性方面的挑战并避免采用临时方法和干预措施时，该制度就能发挥重大作用。

一个有效的对策必须具有战略性和一致性，并以增加值作为基础

176. 检查专员们注意到，大多数组织对分散评价的重要性越来越重视。这是由若干因素驱动的。首先是日益认识到，分散评价的质量普遍不足限制了其作为主要证据来源或中央组织评价基石的有效价值。与评价质量低下关联的是用于支持评价工作的数据和信息管理系统的质量低下。这就要求更加重视分散评价系统，以加强数据收集、分析和监测能力。

177. 其次是缺乏最佳做法模式和快速的全球发展速度，这就要求：(a) 需要进行持续、形成性和流程评价，其重点是测试关于什么东西行之有效、为什么和如何等问题的假设。(b) 工作人员必须全面参与评价工作，以扩大组织内评价思维的广泛范围和加强发展学习型组织；(c) 在实地的联合国各机构和各级分散职能需要实现评价职能的一致性和协调；以及(d) 认识到分散评价需要有足够程度的质量和公信力，以限制在决策中因评价造成的失误并提高分散职能各利益攸关方(捐助方、管理人员、政府、项目管理人员和执行者)的信心。

178. 一些组织在 2012 年和 2013 年采取了一些举措，以应对这些新出现的要求。一些组织展开了调查(国际民航组织和粮食署)或进行了元研究(劳工组织和教科文组织)，以便更多地了解各种组织内分散评价的现况和结果。世卫组织有一个用于在整个组织里发展分散评价的最全面的战略，并且与该组织正在实施的关于组织权力下放的改革方案挂钩。

179. 提出的一个意见是：各组织必须抵制使用临时性策略，必须根据一个明确界定的关于发展分散评价职能的战略框架来开展业务。因此，在作出关于增加分散职能工作量的决定时必须考虑到现有的能力。在为发展学习型组织和文化制订优先事项时必须同时建立适当的激励制度。为避免出现不一致和不平衡发展的风险，必须评估会影响分散职能发展的所有因素，并制订一个一致性的推进计划以限制风险。

评价职能在应对联合国系统所面临的变化和挑战方面的准备状态处于初级水平

180. 检查专员们根据一整套重要标准对分散职能的准备状态作了评估。分散评价具有下列优势和弱点。

主要优势

181. 全组织范围的工作人员作了广泛的项目评价，但各组织之间有差异。这就为发展学习型组织提供了基础，在加强作出努力时要采用适当的方法和重视承担风险和创新的激励制度。分散评价与其他分析系统(监测、审查、评估和流程评价)之间有平均协调，以便为方案一级正在进行的决策进程迅速提供证据。有一个日益增多的参与分散评价领域工作的高质量评价专家和顾问人才库，这个人才库需要得到充分扩大和支持。

主要挑战

182. 缺乏一个明确界定的愿景和战略来应对上文所概述的新出现的趋势，因此所开展的活动往往是临时性的，并且通常不会充分考虑到对整个系统的影响和风险因素。

183. 要在新兴的转型系统中采用快速学习方法就必须重新引进更多形成性评价和其他方法，例如实时评价。同样，系统运作的性质和发展的非结构性会对非传统评价模式和方法的采用产生影响。缺乏规范的统计数据库和完善的信息系统是极其重要的问题。这就要求加强联合国组织和国家统计能力。

184. 对联合国各组织之间的一致性要求越来越多。正如已指出的，过去 10 年至今，一直在运作的联合评价很有限，即使在联发援框架的情况下。必须全面处理和管理已确认因各组织单独治理结构造成的制约联合评价的整套体制因素。这就需要以采用高度积极的合作做事的方式来破除官僚主义界线和加强相互联系。一位受访者的话描述了这样做的决心：有志者事竟成。

185. 在国家一级，对所有评价结果的整合程度要低一些；记录显示，如果将这类评价结果汇集成一个大型资料库的话就能形成一个广泛的知识基础，可用于指导各国的发展。举例而言，对国内有联合国业务的至少 150 个国家所作的国家一级的评价包括：联合国各组织就其各自方案及其区域办公室所作的分散评价和中期审查、在驻地协调员和联合国国家工作队管理下所作的联发援框架评价、中央评价办公室所作的国家一级评价、各个伙伴就相关项目所作的评价，以及其他评价，如全球环境基金(全环基金)和其他发展伙伴所作的评价。因此，任何一个国家现有的知识库都很庞大，尽管质量水平不一，但联合国无法集中获取这类信息，政府和相关伙伴就更无法获取了。拟议的联发援框架 2015 年全系统独立元评价和元合成就是在最大限度地利用分散一级评价结果方面作出的必要努力。

186. 在国家一级的这类评价库里几乎不作任何协调。检查专员们得知一个在开发署在蒙古国所作的一个国家一级评价与联发援框架评价之间进行协调的案例；这个案例被认为是在协调方面的一次困难经历。所进行的一些访谈表明，是否能成功进行协调在很大程度上取决于国家政府是否能在确保联合国系统协调和整合其工作方面发挥领导作用。在为支持国家能力发展作出努力时必须考虑到这个

非常重要的要点，包括政府和国家伙伴在评价联合国系统所作工作的长期影响力方面的作用。

187. 高级管理层必须应对在学习型组织、评价文化和信息共享方面不足这一挑战；高级管理层需要在评价方面发挥更加积极主动的作用，而不是被动地发挥作用。

188. 与国家系统的协调有限，这是个需要解决的重要问题；可以以各组织之间的相互联系为基础加以改善，同时让国家政府和机构发挥领导作用。大多数国家缺乏外部独立办公室。支持各国发展这类独立系统将大大加强在利用国家评价系统方面的公信力。

189. 执行以下建议将会提高分散职能的一致性和完整性。

建议 9

联合国系统各组织行政首长应发展体制框架和必要的支持系统，以提高分散评价的质量和增加值及提高其在支持联合国系统应对新出现的挑战(包括2015年后发展议程的那些挑战)方面所能发挥的作用，以及加强联合国系统各组织内部和全系统范围及与国家机构的一致性和协调。

附件一

联合国系统内评价职能的成果框架

(所列数字系指中央评价职能成熟度矩阵中的项目)

* 2014年研究报告未作评估的领域。但是评估了对影响力评估的准备状态和该职能的总体价值

成果等级	指标
对联合国系统的影响*	实现了联合国系统的目标：发展、人道主义、规范、和平与安全
中期影响：	<ul style="list-style-type: none"> • 加强联合国系统的公信力
利用评价产品所产生的影响力*	<ul style="list-style-type: none"> • 理事机构和关键利益攸关方(国家和发展伙伴)的有效监督和治理 • 提高战略规划、政策、方案和项目的质量及加强体制改革 • 加强联合国系统对全球格局的领导作用(知识、决策和治理)
成果：	
高度利用评价	<ul style="list-style-type: none"> • 加强成果和资源问责制 • 加强学习和改善：高度利用评价建议和获得的经验教训(79)，以为决策和改善战略、方案和活动提供信息(81 82 83 84) • 加强在促进发展评价及全球知识和影响力方面的贡献(43) • 加强联合国系统评价工作的协作、效率和协调
产出：	
优质产出和机构能力	<ul style="list-style-type: none"> • 覆盖：全面、相关和足够，以支持战略决策及制订政策、方案和项目(项目 30)(核心) • 高质量且可信的评价报告(公正、有效、可靠、具有包容性的观点)(72)(核心) • 采用高质量方法以促进发展评价(65) • 加强知识管理：对采集工作的评价、合成、经验教训及在内部和外部进行共享(77、78) • 强大的学习型组织和评价能力：加强能力和成果文化、学习和改善及知识发展和利用 (23、21、26、34、41、43、54、65、66、76、77、78、51、79、81、82、83、84) • 应对全球挑战的准备状态和参与(37、9、22、7、24、81、86、28、41、43、65、77、78、34、36、37)
投入和活动：	高质量的系统、流程、程序、工具、机制

核心职能-活动和流程	有利环境-组织和体制框架	
一. 相关性、响应性、适应性、应对变化和挑战的准备状态	基础	体制制度
<ul style="list-style-type: none"> • 规划、管理和实施评价 • 向理事机构或管理层报告 • 经验教训和合成 • 回应改革、任务和挑战 <ul style="list-style-type: none"> - 回应联合国的改革、全球挑战、国家能力发展、人权/性别平等和促进发展评价 • 复兴和进步 <ul style="list-style-type: none"> - 自我评估 - 外部评估和同行评审 - 政策审查和更新 	<ul style="list-style-type: none"> • 宪章 • 决议 • 任务 • 问责制 • 框架 	<ul style="list-style-type: none"> 政策 规范和标准(评价小组等) 准则(评价小组) 成果管理制和问责制系统 知识管理和学习系统
	评价部门编写的准则和手册	
二. 独立性、公正性、包容性	资源	实践社区
<ul style="list-style-type: none"> • 体制机制 • 职能独立性 • 专业技术独立性 • 行为独立性 	<ul style="list-style-type: none"> • 经常预算 • 预算外 • 人力资本 • 社会资本 	<ul style="list-style-type: none"> 联合国评价小组 联合国区域评价小组
三. 质量-技术和管理	促进问责制和学习的组织文化	
<ul style="list-style-type: none"> • 工作人员的能力 • 评价报告的标准 • 咨商小组 • 外部咨询小组 • 工作人员的培训和职业发展 	<ul style="list-style-type: none"> • 领导力 • 激励机制 • 网络 	
四. 效用	治理安排-作用和责任	
<ul style="list-style-type: none"> • 用于提高利用率的条件和措施 <ul style="list-style-type: none"> - 管理部门的对策和后续行动，以及追踪系统 • 传播、宣传和知识管理 <ul style="list-style-type: none"> - 分析工作的发展和出版物 	<ul style="list-style-type: none"> • 立法 • 行政 • 评价 <p>评价架构</p> <ul style="list-style-type: none"> • 与监督/研究/知识管理系统的协调 	

附件二

联合国系统内评价职能的构成部分和次级构成部分

A: 对评价的要求

要求和要求的用意

组织情况

- 要求的性质和程度(开放式问题)(又见构成部分三和六下的内容)

B: 供应、适应、成长

构成部分：加强有利环境——组织和体制框架及对评价的支持和是否适度回应了要求

- 评价架构：(在管理和进行评价及与其他监督、知识发展和学习部门协调时的组织情况如何)
- 治理：(立法、行政和评价部门的作用)
- 任务、愿景、战略和评价政策
- 支持执行评价政策：(颁布、指导、政策监测、反馈机制)
- 资源(规范和足够度)和筹资
- 促进成果、问责制、评价和学习的组织文化：(成果制框架、领导力和发展学习型组织，即支持雇员不断学习；利用评价并根据评价结果采取行动、思考、利用严格调查和在运作中采用不同思维、在所有层级进行培训以了解情况、采用激励机制以鼓励在测试不同思路时承担风险；获取信息、共享知识和建立实践社区)

构成部分：加强相关性(覆盖范围)、效率和适应性以及在应对变化和挑战及迈向未来方面的准备状态

- 在覆盖本组织方案和项目方面的相关性(评价规划和覆盖范围)
- 适应性和不断改善和成长
- 随时准备提供支持：(一) 分散评价；(二) 联合国改革；(三) 国家评价能力发展；(四) 全球挑战；以及(五) 性别平等及人权和多样性
- 就评价职能如何迈向未来制定方向和进行思考

构成部分：加强公信力：独立性、公正性、包容性和利益攸关方参与

- 结构独立和对评价部门负责人的政策
- 职能独立——规划、管理和交付产出

- 建立公正性机制：控制和利益攸关方参与以维持平衡视角和公正性
- 专业/技术独立和忠诚
- 行为独立

构成部分：加强可信性：有效性和可靠性(技术质量)

- 评价员和评价小组：工作人员和咨询人的质量
- 在应用标准方面用于确保质量和一致性的各种条件已到位
- 评价报告的质量(产出)

C：成果

构成部分：提高效用和潜在影响力

1. 用于提高利用率的条件已到位
2. 产出水平：利用的性质和水平(成果)
3. 影响力水平：利用的实效(影响力)

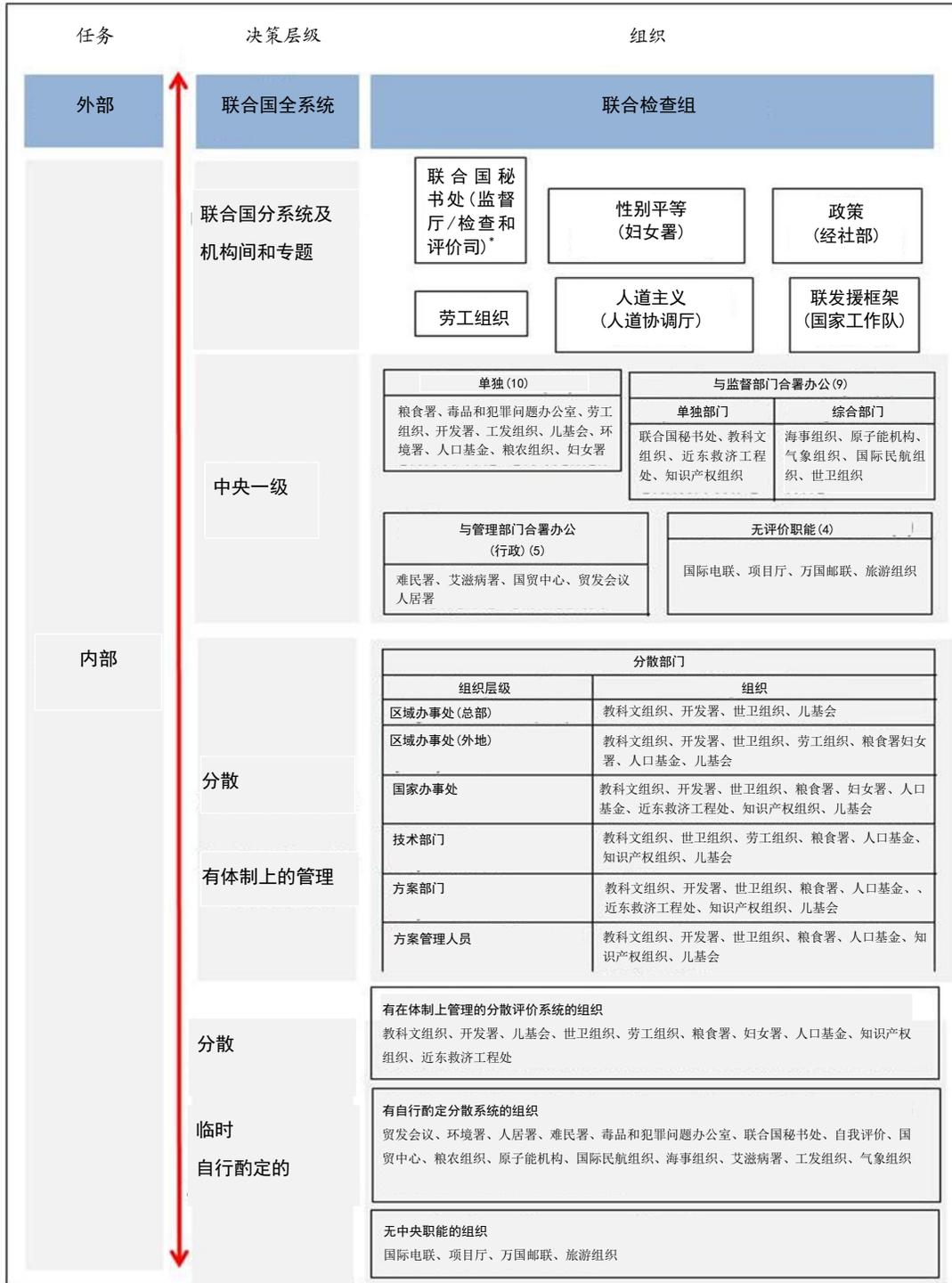
附件三

来自各基金和方案、专门机构、联合国秘书处和其他实体的
联检组参与组织

		任务和主要(P)任务			
		发展	人道主义	规范性	
类型	联检组参与组织				
	联合国贸易和发展会议(贸发会议)	X		X (P)	
各基金和方案	国际贸易中心(国贸中心)	X (P)			
	联合国开发计划署(开发署)	X (P)			
	联合国环境规划署(环境署)	X	X	X (P)	
	联合国人口基金(人口基金)	X (P)	X	X	
	联合国人类住区规划署(人居署)	X (P)	X	X (P)	
	联合国难民事务高级专员公署(难民署)	X	X (P)	X	
	儿童权利和紧急救济组织(儿基会)	X(P)	X		
	联合国毒品和犯罪问题办公室(毒品和犯罪问题办公室)	X (P)		X	
	联合国近东巴勒斯坦难民救济和工程处(近东救济工程处)	X	X (P)		
	妇女署	X (P)	X	X	
	世界粮食计划署(粮食署)	X(P)	X (P)		
	国际劳工组织(劳工组织)	X	X	X (P)	
	联合国粮食及农业组织(粮农组织)	X	X	X (P)	
	联合国教育、科学及文化组织(教科文组织)	X	X	X (P)	
	世界卫生组织(世卫组织)	X	X	X (P)	
	国际民用航空组织(国际民航组织)	X	X	X (P)	
	国际海事组织(海事组织)			X (P)	
	专门机构	国际电信联盟(国际电联)	TC		X (P)
		万国邮政联盟(万国邮联)	TC		X (P)
世界气象组织(气象组织)				X(P)	
世界知识产权组织(知识产权组织)		X		X (P)	
联合国工业发展组织(工发组织)		X (P)		X	
联合国世界旅游组织(旅游组织)		X (P)		X	
联合国内部监督事务厅(监督厅)		X	X (P)	X	
联合国秘书处和其他实体	联合国项目事务厅(项目厅)	执行机构			
	国际原子能机构(原子能机构)	TC		X (P)	
	联合国艾滋病病毒/艾滋病联合规划署(艾滋病署)	X	X	X	

附件四

(a)：联合国系统内评价职能的架构



附件四

(b)：联合国系统内评价职能的架构(关于联合国秘书处结构安排的详细情况)

任务	组织和实体	决策层级
外部	联合国系统联合检查组 监督厅/中央评价职能 (独立评价司) 嵌入联合国秘书处各部 厅、办事处和方案的评 价职能	联合国全系统 全分系统 <ul style="list-style-type: none"> • 和平与安全部门：维和部、外勤支助部、政治部、地区事务办公室、外空厅 • 区域委员会：非洲经委会、亚太经社会、欧洲经委会、拉加经委会、西亚经社会 • 总部以外的联合国办事处：日内瓦办事处、维也纳办事处、内罗毕办事处 • 人道主义和人权：人道协调厅、人权高专办 • 管理支助：秘书长办公厅、大会部、管理部、安保部、新闻部、法律厅 • 其他：最不发达等国家高代办和非洲问题特别顾问办公室
内部	在监督厅监督之下有自 有中央和分散评价职能 的各基金和方案及专门 机构	妇女署、环境署、 贸发会议、劳工组织 毒品和犯罪问题办公室 国贸中心 人居署、近东救济工程处

本研究报告未列入联合国秘书处各部厅/办事处/方案：和平与安全部门(维和部、外勤支助、政治部、地区事务办公室、外空厅)；区域委员会(非洲经委会、亚太经社会、欧洲经委会、拉加经委会、西亚经社会)；总部以外的联合国办事处(日内瓦办事处、维也纳办事处、内罗毕办事处)；人道主义和人权(人道协调厅、人权高专办)；管理支助(秘书长办公厅、大会部、管理部、安保部、新闻部、法律厅)及最不发达等国家高代办和非洲问题特别顾问办公室。因此，研究报告未覆盖和平与安全及人道协调厅和人权高专办的人道主义和人权机构间协调机制。

这些部门的中央评价职能是由监督厅/检查和评价司履行的。本研究报告列入了监督厅/检查和评价司，作为分析联合国各组织中央评价职能的内容。

各部厅、办事处和方案有嵌入式评价职能，即该职能是各部厅、办事处或方案内部自有的。这类评价职能有一种混合结构。有嵌入方案、作自我评价的评价职能，由自己酌定如何规划和实施。有单独设立的评价部门，雇用专门开展评价工作的专业评价员(如新闻部、维和部)，负责回应管理部门的评价要求以指导部厅的业务。还有专门负责评价和其他职能的单独设立的评价部门(如大会部)。另有一些单位没有专门的职能或能力(如经社部、管理部和政治部)。对这些单位作了访谈和磋商，并利用相关信息来指导决定本研究报告的范围。

2010年监督厅关于联合国秘书处评价能力的双年度报告列入了这些部厅。联检组力求避免作重复工作，避免开始另一次收集数据进程，因为在最初决定研究范围的阶段时就提出了这个关切问题。联检组注意到对本研究报告对联合国这些部厅、办事处和方案评价职能的总体目的、结构、效率和切合目的的情况所作分析的兴趣。联检组欢迎有机会与监督厅/检查和评价司就此进行协作。

有8个组织的部分资源来自秘书长办公厅，因此它们也是联合国秘书处的组成部分。这些组织是：妇女署、环境署、贸发会议、劳工组织、毒品和犯罪问题办公室、国贸中心、近东救济工程处、人居署。它们同样处于监督厅的监督之下。它们是以各基金和方案或专门机构的身份列入本研究报告的。它们的结构安排与各基金和方案及专门机构的典型评价职能不同，有一个明确界定的中央评价职能和一个分散职能。

附件五

(a)：联合国系统评价职能一些术语的定义：中央和分散评价

联合国系统的评价职能已发生演变。随着这种演变发展出各种概念和结构，并随之产生术语方面的混乱。所用的术语或指不同的事物或在意思上有细微差别。这是本研究报告遇到的挑战之一。下文对一些关键概念和结构作了定义；本研究报告在提及中央和分散评价包括自我评价概念时是按这种定义使用这些术语的，由此支持了对所用术语的标准化。欢迎评价界为本清单增添新内容，以支持评价职能及其新出现的特征的演变。

中央评价职能

联合国系统的中央评价职能得到评价办公室的支持；评价办公室在开展评价的规划、管理、实施和报告业务时独立于业务管理部门和方案部门的控制。

在一些组织里，中央评价办公室作为单独的办公室运作，有两种类型：(一) 在行政首长办公室之外运作；和(二) 在行政首长办公室之内作为一个单独的单位运作。在其他情况下，中央评价办公室在行政首长办公室内作为研究、政策和战略规划的重要组成部分运作(不是单独的)。在其他组织，中央评价办公室在本组织监督办公室内与其他监督办公室一起运作，这里也有两种类型：(一) 在监督办公室作为一个单独的单位运作；和(二) 作为其他监督职能的综合组成部分运作。

中央评价职能的显著特点是产出评价结果，以支持本组织广泛的中央级组织政策、方案和战略决策，虽然存在许多以不同程度进行项目评价的中央职能。

报告主要提交给政府间机构、高级别咨询机构和高级管理层，用于促进全组织范围的变革和在政策、范围广泛的方案(国家、区域、全球)和交叉专题及体制改革方面的改善。

参照结构、职能、包容性、行为及专业技术独立性的全部标准，中央评价显示出不同程度的独立性。这种独立性不是自动就有的，而需要设立用于促进结构、职能、技术、透明度和行为独立性的机制。

一些组织把中央职能称之为“共同”职能，意思是，评价具有范围更广泛的共同实效。但是，“共同”一词同样应用于分散评价。

分散职能评价

分散评价是在中央评价部门之外规划、管理和实施的。从历史上看，它最初被命名为自我评价，但这个术语现在被弃用了，因为甚至在某些中央职能里也包含这个术语(见下文自我评价这一条目)。

分散评价嵌入在方案部门内，项目和方案的规划和执行则由业务管理职能负责。这包括在技术、区域和国家办事处一级或部厅一级的项目和方案。

因此，分散评价职能通常作为涉及评价、监测和其他相关分析活动的方案管理工作的组成部分运作，以支持作出与执行项目和方案各个阶段有关联的决策。

这种类型的分散评价在结构上没有独立于方案管理人员在评价的规划、管理、实施、批准和报告方面的控制。这适用于方案工作人员自己酌定管理和实施的评价及委托或承包给外部咨询人的评价，因为方案管理人员对报告有最后批准权。这同样包括联发援框架评价。

非独立分散评价由工作人员实施或委托给外部咨询人，但在业务管理部门的控制之下，因为规划、实施和报告都不是独立的。

独立分散评价在分散一级和其他层级规划，在其他可信的独立机构的控制之下，对报告的最后批准和分发都是独立作出的，因此被认为是独立的。但在这类情况下的可信性依赖在关键利益攸关方参与方面的透明度。

自我评价

这一术语引起的混乱最严重。联检组在本研究报告中决定，不用这个术语作职能的分类。这个术语主要用作形容词。

总的立场是，自我评价不是独立的，因为：

- (一) 它不是由中央部门实施的。但有人说，有些中央部门在位于行政首长之下时就不是独立的，或如果考虑到联检组成熟度矩阵确认的所有独立性因素，则它们会显示出不同程度的独立性
- (二) 它不是由外部咨询人实施的。但有人说，只要报告仍然在方案管理人员的控制之下，仅凭使用外部咨询人并不能保证独立性。外部咨询人也并不总是以独立的视野来运作的。

项目管理人员进行自我评价的目的是自己用来改善项目和方案的。

- (三) 有人说，驱动管理人员的这种利益关切意味着，管理人员将以忠诚的态度运作，而不会篡改证据，因为需要知道相关证据以进行改善。因此，独立性问题没有意义，可用管理层的忠诚和信任来取代。

严格为方案所用的自我评价(如某些定义所描述的特征)是很罕见的，因为在不同情况下自我评价还用于支持满足在资源使用的问责制方面的要求，特别是在吸引更多预算外支持的情况下。有人说，最好是使用形成性评价这个词来描述用于设计和改善项目和方案的评价。

下文是联检组过去和监督厅最近对自我评价所下的定义。在本研究报告里这两个定义现在都归类在分散职能之下。

自我评价——从前的定义(摘录自 JIU/REP/81/6)：“覆盖组织全部活动范围的评价，……一个就新出现的成果及利用成果的情况作持续不断迅速信息反馈的进

程，以快速调整目标和战略及改善今后的活动，……纳入评价使其成为管理周期的一个正常组成部分，……作为一个参与性进程以加强工作人员的承诺和参与并促进建设一种评价文化和促进发展学习型组织”。所有组织在 1980 年代都是这样做的。

自我评价(监督厅的关键监督术语清单，版本 1.0，2013 年 4 月)

项目和方案管理人员进行的主要为自己所用而进行的评价。自我评价的重点是次级方案，可应用于所有部门，包括政治、法律、人道主义、经济和社会事务、新闻和共同事务。作为管理流程的一个组成部分，自我评价程序的设计和规划在规划和方案编制阶段就结合相关次级方案的设计作了明确规定。方案管理人员利用调查结果对执行工作作必要的调整或以对相关次级方案或项目的设计和/或方向提出改动的方式把调查结果反馈回规划和方案编制进程。虽然通常不在政府间一级报告自我评价的结果，但来自为一个方案内各次级方案和项目产生的分析信息的结论可用于评估整个方案。

本研究报告中的分散评价的类型

如下文所概述的，本研究报告把分散职能分成两个模式。在两个模式的每个模式下都有不一定相互排斥的各种原型。下一页附件五(b)提供了这些原型。

临时、自行酌定的分散评价

工作人员和管理人员自行酌定实施的评价，并不符合明确界定的体制要求。它们不与用于提交报告和作出其他方案一级战略决定的体制框架挂钩。它们遵循联检组和监督厅的定义，虽然也可以用于问责制目的，关于最近趋势的传闻证据也证明了这一点。

用有体制框架的评价替代分散评价

鉴于临时模式的特性，发展此模式的目的是为了给分散评价职能一个更有规划和战略性的价值。

这类分散评价受明确界定的体制框架的影响，有明确的规则及激励和宣传制度，以支持对这类评价的规划、执行和报告采用一致和战略性的方法。用于这类分散职能的框架有如下特点：(一) 现行的评价政策界定了分散评价职能的作用；(二) 评价由各部厅和国家办事处管理；(三) 这类评价是有规划的；(四) 它们按照一整套标准运作；(五) 有系统性的质量保证或控制和提高质量机制；以及(六) 对遵守政策的情况和该职能的发展情况提出报告。

中央评价办公室在这类职能上的参与程度各不相同，但总之，它们由所属的业务管理部门管理。中央职能提供了各种类型的质量提高和/或质量控制服务。各组织内确定存在着数量不同的制度化因素。

附件五

(b)：联合国系统内的分散评价职能和原型

不同的原型不会相互穷尽。各组织可能同时适用一个以上原型。

模式 1：临时和自行酌定——几乎联合国所有组织			模式 2：提供了明确界定的体制框架——联合国一些组织		
原型 1:	原型 2:	原型 3:	原型 4:	原型 5:	原型 6:
由方案部门作规划、管理。	由方案部门作规划、管理。	由方案部门作规划和管理。	由方案部门作规划和管理(根据政策)。	由方案部门作规划(根据政策)。	由管理部门或根据管理部门的要求作规划(根据政策)。
没有界定体制框架或由成果管理制度提供这个框架。	没有界定体制框架或由成果管理制度提供该框架。	没有界定体制框架或由成果管理制度提供这个框架。	不同程度地界定了用于政策、规划、管理和实施的体制框架。	关于评价对象的决定要经行政管理部批准。	关于评价对象的决定要经行政管理部门批准。作的是战略性选择。为选择评价对象制定了标准。
由项目或方案工作人员作评价。	评价委托给外方并根据技术标准甄选出来的外部咨询人实施。咨询人独立于被评价的方案或项目。	由工作人员或外部受委托的咨询人实施评价。	由工作人员或外部受委托的咨询人实施评价。	由外部咨询人实施评价。但由业务管理和方案部门管理。	由中央评价部门管理和实施评价。
质量的提高和保证措施是临时性的。	质量的提高和保证措施是临时性的。	由中央评价办公室负责质量的提高、支持和指导(通过手册和指导材料)。	由中央评价办公室负责质量的密集提高、支持和指导(通过直接的服务台帮助和/或通过手册和指导材料及在评价和项目设计质量方面的培训，即可评价性评估)。	由中央评价办公室负责质量的提高、支持和指导，大多数限于提供指导材料。	由中央评价办公室负责质量的提高、支持和指导。
由业务管理人员作最后批准和许可。	由业务管理部门对报告作最后批准和许可。	由业务管理部门负责最后批准和分发报告。	由中央评价办公室在对每一份报告作验证或核实或其他质量控制机制后进行批准、许可和分发。	由业务管理部门负责最后批准和分发报告。由中央评价部门评估报告质量后向组织提交报告并检查合规情况，以向决策工作提供支持。	由中央评价办公室负责最后批准和分发报告。
报告为自己所用——改善和纠正。	报告为自己所用——改善和纠正。	报告为自己所用——改善和纠正。	报告的用途较广。	报告的用途较广。	报告的用途较广并是中央职能报告库的组成部分。
报告未被知识管理系统收集。	报告未被知识管理系统收集。	报告未被知识管理系统收集。	报告被知识管理系统或评价资源平台收集。	报告被知识管理系统或评价资源平台收集。	报告被知识管理系统或评价资源平台收集。

附件六

联合国系统 23 个组织所使用的业务模式

模式类型	使用该模式的 组织数量	2012-2013 年 采用该模式完成 的评价数量
<p>模式 1: 评价委托给外面的专家或公司。该外部来源对质量负全责。它对评价工作提出报告。</p> <p>评价部门的责任是支持评价的实施。评价的质量保证可由外部评价管理小组或外部读者负责。</p>	2	5
<p>模式 2: 评价部门对评价负全责。它领导报告的整体管理、质量保证和完成工作。它要确保评价对本组织是有意义的(机构有效性)。</p> <p>实施评价的团队领导工作承包给同时具有评价或广泛分析背景和战略前瞻眼光的主题事项专家。同样，还与能为评价工作提供支持的专家小组订约。评价部门还可以参与评价的试点或实施工作，在某些情况下，可作为小组共同负责人。</p>	16	417
<p>模式 3: 评价部门对评价负全责。它领导报告的管理、质量保证和完成工作，并负有对评价的机构价值的责任。</p> <p>评价部门担任小组领导。与外部专家共同担任小组领导同样是一个选项。评价部门可视需要利用订约承包专家提供支持。</p>	10	72
<p>模式 4: 评价在内部实施，不与外面的来源订约。</p>	3	9
<p>模式 5: 其他模式：包括在分散一级进行的项目，但在评价部门的质量控制之下并由该部门作最后批准。</p>	3	35

附件七

对评价职能超出实施和提交评价报告之外的要求

对组织的内部支持, 包括分散评价	为组织的利益发展评价职能	外联(全球、区域和国家)
成果管制和支持项目和方案设计的质量	发展政策 审查和更新政策(由外部小组负责) 颁布政策	全球平台的参与和全球知识: 会议讲习班 出版物
发展关于评价的指南和工具——全组织范围	关于评价职能和评价的年度报告	规范性工作、一致性、统一和知识共享: 联合国评价小组和联合国改革: 联合国评价小组年度大会: 联合国评价小组评价做法交流
培训工作人员(网络研讨会、面对面、远程学习) 培训高级管理人员 开发基于网络的培训工具	吸取来自各种评价的经验教训(合成)以采取管理行动	方法方面的全球伙伴关系: 作为影响力评价网络的成员和EvalPartners的成员参与
分散评价: 提高质量和控制质量	宣传和传播	国家评价能力发展
后续追踪系统的发展和报告 数据系统: 各组织所有评价和知识管理平台	中央评价职能的方法指南	与独立的全系统评价的关联

附件八

1977-2013 年联检组参与组织评价办公室的演变

在此期间在发展水平和工作人员人数增加方面的进展情况一览表

(分析的基础是从多个数据来源获得的数据。数据来源未就同样的组织全部报告数据)。日期表明设立该职能的年份

		中央评价职能									
		发展水平					在评价部门的工作人员和咨询人人数				
		1980 年		1980-2005 年 (无数据)			2011 年		2013 年		2012-2013 年
		1977 年	20 个组织	工作人员 人数	2006 年	2013 年	工作人员 人数 (21 个组织)	专业工作 人员人数 (24 个组织)	一般事务 人员人数	工作人员 总数	咨询人人数 (17 个组织)
成熟度*							132	160	39.5	199.2	988.5(2 年)
5											
4											
过渡到第 4 级						开发署	15	18	6	24	160
						劳工组织	7	5.5	1	6.5	85
						粮食署	8	16	3	19	188
						妇女署 (2011 年)	8	12	2	14	X
						工发组织	4	4	3	7	141
						儿基会	7	14	3	17	X
						联合国监督厅	19	22	4	26	X
						人口基金	5	6	1	7	46
						教科文组织	X	6.5	1	7.5	X
						粮农组织	15	17	4	21	318
3			粮农组织	8		开发署					
(职能的主要要素已到位, 日常业务在持续进行)			人口基金	5		粮食署 国贸中心					
			世卫组织	1		工发组织					
			粮食署	10		儿基会					
			开发署	1		教科文组织					
			国贸中心	2		粮农组织					
			环境署	2		毒品和犯罪问题办公室					
						人居署					
过渡到第 3 级											
(系统的一些要素已到位, 机制性运作)		粮食署(1963 年)	劳工组织(1977 年)	3		劳工组织		2	1	3	7
		开发署(1967 年)	教科文组织(1987 年)	2		妇女署		4	0.5	4.8	X
		粮农组织(1968 年)	联合国人居中心(1979 年)	1		联合国监督厅(1994 年)		2	0.5	2.5	11
		气象组织(1968 年)	气象组织	2		人口基金		2	2	4	1
		人口基金(1972 年)	工发组织	1		原子能机构					
		环境署(1975 年)	难民署(1980 年)	4		环境署					
		国贸中心(1975 年)	儿基会	2		贸发会议					
		儿基会(1976 年)	原子能机构(1978/80 年)	10							
		工发组织(1976 年)	联合国经社部(1980 年)	4							
		世卫组织(1976 年)									
		联合国管理部(1974 年)									

		中央评价职能									
		发展水平				在评价部门的工作人员和咨询人人数					
		1980年		1980-2005年 (无数据)	2006年	2013年	2011年	2013年		2012-2013年	
成熟度*	1977年	20个组织	工作人员 人数	2005年 (无数据)	2006年	2013年	工作人员 人数 (21个组织)	专业工作 人员人数 (24个组织)	一般事务 人员人数	工作人员 总数	咨询人人数 (17个组织)
							132	160	39.5	199.2	988.5(2年)
2 (正在得到发展, 临时性的)	教科文组织 劳工组织					国际民航组织 难民署 海事组织 气象组织 近东救济工程处	2 5 2 2 2	2 7 1 1 2	0.2 1 X 0.25 X	2.45 8 0.5 0.75 2	1 13 3 X 2
1 (不存在/朝发展 评价职能的方向 前进)	原子能机构 国际电联 海事组织 贸发会议	海事组织 贸发会议 国际电联 万国邮联				项目厅 万国邮联 旅游组织 国际电联					
分散职能评价											
有自行酌定的评价的组织						2013年: 所报告的中央评价部门向分散职能提供的支持质量的水平					
开发署 工发组织 劳工组织 教科文组织 万国邮联 国际电联 组织 国际民航组织 教科文组织 世卫组织 劳工组织 工发组织 国际电联 W/开发署						广泛支持			有限支持		
						毒品和犯罪问题办公室 联合国 监督厅 人口基金 儿基会 劳工 组织 工发组织 世卫组织			贸发会议 人居署 难民署 近东救济工程 处 粮食署 粮农组织 国际民航组织 海 事组织 教科文组织 知识产权组织		
						良好支持			无支持		
						国贸中心 开发署 妇女署 艾滋 病署			环境署 原子能机构 气象组织		

说明: 成熟度或发展水平的概念化基础是联检组关于成熟度矩阵的概念计划。对 1977 年至 1980 年所作分析的基础是联检组以该职能为重点的各次报告。对 2006 年所作的分析有许多来源: (一) 联合国评价小组质量合格标志工作队: 联合国评价小组成员核对表答案的基线概要-以联合国 23 个组织的自我评价为基础, 其中 16 个是联检组参与组织; 这是参照联合国评价小组的规范和标准进行的; (二) 2006 年和 2007 年之间各组织进行的同行评审(开发署、儿基会); 联检组关于 2005/6 年监督和问责制框架的报告。对妇女署(2011 年)所作分析的基础是妇发基金评价职能记录和在联合国监督厅监督之下运作的其他相关机构关于性别平等和增强妇女权能的记录。

附件九

学习型组织：促进成果、问责制、双圈学习、共享和利用评价的文化

跨组织分析：平均得分和参照学习型组织指标处于不同发展水平上的组织数量

	平均得分	处于成熟度矩阵不同水平的组织数量			
		第1级 1-2	第2级 3-4	第3级 5-6	第4级 7-8
促进成果、问责制、学习的组织文化 (23)					
(核心项目——在成熟度矩阵里的总体汇总评分)	4.3	1	11	12	-
一. 战略协调、外联、生成性					
与其他支持系统的协调(21)	4.7	-	10	11	3
对分散评价的支持(26)	4.8	1	11	3	8
对国家评价能力发展的支持(34)	3.5	8	9	5	2
共计	4.35				
二. 发展、贡献、转型、创造					
职能的持续不断评估(41)	5.5	1	7	7	10
对促进评价知识的贡献(43)	4.8	4	8	4	8
控制和利益攸关方参与及包容各种视角(54)	6.0	-	3	11	10
方法——发展和创新(65)	5.3	-	9	11	4
工作人员的专业发展(66)	5.1	-	12	5	7
共计	5.34				
三. 共享和实践社区					
增强可获取性和透明度(76)	6.4	2	3	3	16
内部共享(77)	5.4	1	5	10	8
外部共享(78)	4.7	6	5	10	3
提供获取信息的渠道(51)	6.0	1	2	11	10
共计	5.6				
四. 双圈学习——利用和行动					
执行建议(79)	4.8	7	1	11	5
用于战略决策(81)	5.0	1	7	12	4
集体/共同利益(82)	5.2	1	7	8	8
形成性利用(83)	4.9	1	9	9	5
组织以外的利用(联合国其他机构、国家、发展伙伴)(84)	4.3	3	9	11	1
共计	4.84				
总平均得分	4.8				
处于不同水平的组织平均数		2.1	7.3	8.8	6.2

说明：指标边上的数字是成熟度矩阵里的类目数。

附件十

领导力属性和职能

跨组织分析：平均得分和处于四种发展水平的组织数量

领导职能	平均得分	处于成熟度矩阵不同水平的组织数量			
		第1级 1-2	第2级 3-4	第3级 5-6	第4级 7-8
高级管理层的整体领导力*					
核心项目(22) 主要责任(核心)	4.4	0	11	13	0
一. 前瞻性、战略、关系					
前瞻性和战略(9) 共同责任	5.2	1	7	12	4
治理结构(6) 连带责任	5.8	0	7	6	11
二. 职能的制度化					
颁布政策(12) 共同责任	5.1	1	4	16	3
学习型组织和评价文化(23) 主要责任	4.3	2	10	12	0
相关支持系统和协调及用于加强评价的激励机制(21) 主要责任	4.7	0	10	11	3
有专用、稳定和足够的资源(15、19、50) 主要责任	4.8	2	10	9	3
确保利用评价(79、81、82、83、84、85、86) 共同责任	4.5	2	8	14	0
三. 问责制和独立性					
保障方法的独立性和完整性(61) 主要责任	5.9	1	1	12	9
确保可获取用于评价的信息(51)(独立性和透明度) 但关键问题更在于数据的质量和在获取数据和信息方面的效率及已到位的制度。 主要责任	5.9	1	2	11	10
确保工作方案的独立性(49) 主要责任	5.7	1	4	9	10
可获取性和透明度(76) 共同责任	6.4	2	3	6	16
平均得分	5.0				
处于不同水平的组织平均数		1	6.7	11.4	4.8

* 高级管理层充分了解评价的作用和增加值，在本组织内积极支持和促进评价职能并发展适当的激励制度。联检组关于领导力及发展学习和问责制文化的原则是存在的，包括领导评价职能：按价值和实例、按信息和宣传、按动机、按指导和纪律、按参与度。

说明：指标边上的数字是成熟度矩阵里的类目数。

附件十一

非核心供资、对有效和独立利用的保障与评价职能发展水平之间的关系

13 个组织非核心资源百分比 (采用完整的数据集)	发展水 平的平 均得分	对评价职能独立性、可持 续性、有效利用和限制交 易成本的保障的平均得分 (指标 18)	平均得分独立性 (结构、职能、行 为、专业、技术) (指标 45 至 62)
非核心供资水平较低			
1-10%: 开发署 教科文组织 人口基金	6.3	6.3	6.7
11-20%: 粮食署 妇女署 原子能机构 贸发会议	5.8	6.6	6.4
共计	6.1	6.5	6.5
非核心供资水平一般			
21-30%: 工发组织 国贸中心	5.7	5.5	5.4
31-40%: 儿基会 粮农组织	6.1	4.5	6.4
共计	5.9	5.0	5.9
非核心供资水平较高			
40%以上: 环境署(46%) 人居署(82%)	5	4.2*	5.1

* 人居署的是估计数据。

附件十二

为应对全球变化和挑战包括 2015 年后发展议程规定的任务以
以促进可持续发展所作准备的状态

	平均得分	处于成熟度矩阵不同水平的组织数量			
		第 1 级 1-2	第 2 级 3-4	第 3 级 5-6	第 4 级 7-8
应对全球变化和挑战					
(核心因素)(37)*	5.0	-	8	5	2
一. 在思考和展望方面的战略性					
前瞻性和战略(9)	5.2	1	7	12	4
领导力(22)	4.4	-	11	13	0
成员国的赞赏(7)	5.5	-	7	11	6
对战略性和优化覆盖作出规划(24)	5.8	-	6	7	11
利用战略(81)	5.0	1	7	12	4
利用的实效——影响力(86)	3.9	4	10	10	0
共计	4.9				
二. 创新、适应和更新					
平衡活动(28)	5.6	-	6	9	9
持续评估(41)	5.5	1	6	7	10
为促进发展评价作出贡献(43)	4.8	4	8	4	8
方法和创新(65)	5.3	-	9	11	4
共计	5.3				
三. 参与，包括全球平台					
内部交流(77)	5.4	1	5	10	8
对外交流(78)	4.7	5	5	10	3
共计	5.0				
四. 一致性、包容性、自主权、协调					
支持国家评价能力(34)	3.5	8	10	10	0
性别平等、人权和包容性(36)	5.2	1	9	5	9
支持分散评价(37)	4.7	3	10	3	8
共计	4.4				
发展学习型组织					
(总分——清单项目)	4.8	-	8	16	0
跨组织平均得分	4.9				
处于不同水平的组织平均数		1.5	7.7	9.1	4.5

* 充分认识到全球趋势和挑战。寻找新的方法和伙伴关系(在规划、覆盖、共同工作、方法和考虑因素方面)及为认识多样性和新任务改善旧的伙伴关系和方法。

说明：指标边上的数字是成熟度矩阵里的类目数。

