

元数据在城市可持续发展能力评估数据的应用分析

The use of metadata in the evaluation of urban sustainable development capacities

姜天祺 李思远 张娜娜 董仁才（中国科学院生态环境研究中心/城市与区域国家重点实验室/系统分析与可持续发展研究组，北京 100085）

摘要 可持续发展能力评估是可持续发展战略实施中的重要环节，然而传统的评估数据获取方式已经难以满足城市可持续发展能力评估的数据需求。本文通过对城市可持续发展能力评估中数据问题进行深入分析，得出影响评估的数据层面的关键因素。并应用元数据理论，提出一套针对城市可持续发展能力评估的数据获取与管理方法，包括元数据标准与框架，以及针对每一项数据所需填报的元数据项。旨在为解决评估中的数据问题提供思路，同时对元数据在科学评价工作中的应用展开探索。

关键词 元数据; 可持续发展能力; 科学评价; 指标体系

免责声明 本文仅代表作者个人观点，与中国 21 世纪议程管理中心、中国可持续发展研究会及联合国均无关。

Disclaimer: This brief was submitted through the Administrative Centre for China's Agenda 21 (ACCA21), Ministry of Science and Technology, China, and the Chinese Society for Sustainable Development (CSSD). The views and opinions expressed are those of the author(s) and do not necessarily represent the views of, and should not be attributed to, the Secretariat of the United Nations, the ACCA21 or the CSSD. Online publication or dissemination does not imply endorsement by the United Nations. For further information, please contact Mr. Sun Xinzhang (sunxzh@acca21.org.cn)

1 引言

随着可持续发展理念的推行，我国在可持续城市建设方面取得了显著进展。但如何全面评估城市的可持续发展能力一直是一个棘手问题，除了要有一个科学有效的评价指标体系外，评估数据的获取与管理也是一个关键因素。从目前来看，传统的仅通过向被评估方派发评估数据统计表的数据获取方式已经难满足对评估数据使用需求。在现实中，评估数据本身的不准确和难以追溯，是造成评估结果不准确的普遍原因。特别是在历史数据中，例如某个指标在不同地区的统计方法相异，而该统计指标又出现在评估指标当中，评估人员无从知晓，造成评估漏洞。类似的情况时有发生，严重影响了评估的科学性与准确性。针对以上情况，城市可持续发展能力评估需要一套科学严谨的数据获取与管理方式，以保证评估的科学有效。而元数据的应用，可以为解决上述问题提供思路。

元数据是关于数据的信息，是具有描述、解释、定位信息资源功能的结构化信息，是说明数据内容、质量、状况及其他有关特征的描述信息^[1]。在中国可持续发展工作中，一向非常重视元数据的规范化工作。早在本世纪初，为了加强对国家层面可持续发展信息中的上交集中共享数据的监督与管理，规范上交集中共享数据的数据格式，保证数据质量，保障集中共享数据的集成、发布和综合分析工作的顺利进行，就已制定中国可持续发展信息共享元数据标准，该标准的主题内容是“提供中国可持续发展共享信息元数据的内容，包括可持续发展数据标识、内容、质量、状况及其他有关特征”，该标准可用于国家层面的数据集的状态描述、数据集编目及信息交换网络服务^[2,3,4]。

长期以来，元数据在众多领域扮演着极其重要的角色。元数据最初起源于图书管理领域，用以描述图书与期刊的基础信息。进入新世纪，随着对网络信息资源进行有效管理和使用的需求不断增长，元数据的理论得到

了极大丰富，同时元数据正在被应用到更多的领域。尤其在空间信息领域，空间数据的多尺度、多精度、多地域、多时序等特点使得元数据对空间数据集的描述尤为重要。然而在统计领域，元数据的应用还处在起步和探索阶段，在科学评估中的元数据应用更为鲜有^[5, 6]。

本文运用元数据理论，针对城市可持续发展能力评估，对元数据在科学评估中的实践应用展开探索。试图通过对元数据的应用，解决评估中的数据问题，保证评估结果科学有效，为评估过程中的数据获取和管理提供思路。

2 城市可持续发展能力评估中的数据问题

2.1 评估中的数据问题分析

城市可持续发展能力由经济、社会、资源和环境等要素共同决定，各要素之间相互作用相互制约的关系反映了城市可持续发展的整体状态和变化趋势^[7, 8]。城市可持续发展能力评估是综合以上要素，依据城市的历史发展情况，对城市未来可持续发展潜力的评估。评估对象覆盖全国 156 家可持续发展实验区，其涉及领域之广、涵盖指标之多，都超出一般科学评估。

城市可持续发展能力评估数据受制于评估本身的特殊性和复杂性，使得数据的获取方式、生产方式等均存在着不同程度上的差异。由于评估中多数指标是由不同的机构生产和管理，地方可持续发展实验区管理小组需要通过大量不同的渠道收集数据再上报给评估者，以日照市可持续发展实验区为例，实验区管理小组需要从十余个政府机关、事业单位获取评估数据。另一方面，同一指标在不同地域、不同部门的计算方式也有差异，例如高新技术产业产值占工业总产值比重这一指标，不同地区对高新技术产业的认定标准不同，造成统计差异。加之一些其他因素，如获取数据的设备种类和数据格式不同，指标的时间分辨率、空间分辨率和处理技术不同，都可以导致同一指标在不同地区有着不同的生产方式。

以上这些数据背后的信息，都与评估结果有着直接或间接的联系，而以往使用的数据获取机制中并没有对这些信息进行有效的记录和管理，这种对数据基础信息无记录的数据收集方式是导致评估中数据问题的根本原因。导致的问题主要体现在以下三个方面：

(1) 评估数据的真实性。城市可持续发展能力评估数据来源复杂多样，数据生产涉及国土、规划、发改、经信、能源、水务、环保、市政、教育等多个政府机构的统计部门。而且在评估数据中，并不是所有的数据都是公开数据。事实上，在评估指标中只有部分数据可以获取自国家或地方统计年鉴、公报。另一部分数据为非公开数据，这些非公开数据，实验区管理小组需要与相关单位直接联系才能取得。以日照市可持续发展实验区为例，非公开数据往往占到以往实验区上报数据的三分之一以上，有些地区甚至更多。面对以上如此复杂多样的数据获取渠道，如果不对数据的获取方式加以记录和管理，就会难以避免的发生一些数据谎报、错报的情况，使得数据的真实性难以保障。而面对这样的情况，当数据出现问题，评估者无从知晓其数据来源，无法对问题数据进行追踪和纠正，对评估的进展造成阻碍。

(2) 评估结果的可比性。当评估指标在不同地区的生产方式不同时，评估的可比性将受到影响。针对城市可持续发展能力评估，其评估对象为国家可持续发展实验区，而国家可持续发展实验区包含多种不同类型，如大城市城区型、中小城市型、县域型、城镇型等^[9]。在不同尺度、不同地域、不同行政级别的实验区之间，对于同一指标的理解和计算时常存在差异，这些差异体现为计算方法上的差异，空间、时间范围上的差异，统计频率上的差异等等。这些数据生产方式的差异将导致评估在不同地区间缺乏可比性。对于目前的评估，由于对收集上来的评估数据的生产方式缺乏记录，评估人员无从知晓同一指标在不同地区间生产方式的差别，导致很多缺乏可比性的数据被应用在评估当中。

(3) 评估的可持续性。城市可持续发展能力评估是一项长期的评估工作，评估由历史延伸到未来，需要逐年进行对比，得出一个地区在一段时间内的可持续发展能力变化情况，因此要求评估数据的获取必须具有可持续性。但由于目前的评估中并未记录数据的获取来源以及其他一些必要的基础信息，使得数据收集工作无法很好的衔接，数据收集人员无法通过查看历史数据的收集方式来指导新的数据收集工作，同时也会有对数据的理解偏差和使用不当的情况发生。还有一些获取来源以外的数据基础信息同样重要，如数据的公开性，对一个数据是否为公开数据的了解对于数据的获取和使用都有着明确的指导意义。

可持续性的另一方面体现在对于数据信息的完整记录和保存上。随着评估的数据使用需求不断增加、数据量不断增大，一项完整有效的科学评估需要建立在数据信息完备的基础之上。只有当一个数据拥有完整的信息记录，该数据才具备长久的使用价值。

2.2 元数据在评估中的应用意义

如上所述，如何有效记录和管理数据背后的关键信息，保证数据的真实性、评估结果的可比性和评估的可持续性，是城市可持续发展能力评估中关于数据的核心问题。

元数据是有着广泛应用的用以记录数据信息的重要工具，面对以上评估中的数据问题，元数据的应用可以起到很好的应对作用。针对数据的真实性，元数据可以通过记录数据来源信息，使数据的收集过程规范化、透明化，降低人为导致数据错误的几率，最大限度的保证数据真实可靠；针对评估的可比性，元数据可以通过记录数据的生产方式信息，向评估人员展示同一指标在不同地区的生产方式的差异情况，使其可以以此为依据，对数据进行修正或对评估方式进行调整，保证不同地区的评估结果可比；针对评估的可持续性，元数据可以通过记录数据的获取方式及公开性等基础信息，指导数据

收集人员对数据进行逐年获取，确保评估可以持续稳定的进行。并使评估数据具有完备的信息记录，使其可以被可持续利用。

综上所述，元数据可以将评估数据背后的重要信息加以记录和归纳整理，使评估者可以更加全面的认识和了解收集来的数据，利于对数据的处理和使用，在避免数据问题的同时，也对评估工作的持续开展起到促进作用。

3 元数据在城市可持续发展能力评估中的应用

3.1 元数据标准的构建

元数据的应用需要通过构建元数据标准来实现，合理的元数据标准设计对提升元数据库的利用效率至关重要。我国现有元数据国家标准 20 余项，其中对于元数据标准的构建规范已初步形成。针对城市可持续发展能力评估，由于评估本身涉及指标众多，且每一年每一项指标的数据都具有元数据信息，因此元数据的设计应在保证信息完整度的前提下，尽量精简元数据库中的元素，减少数据冗余。此外，与其他领域的元数据标准不同，城市可持续发展能力评估的元数据意义不仅在于数据记录和管理，也在于为数据获取提供规范。一些原本在数据收集过程中不被记录的信息将被收集整理，用以满足评估的使用需求。

同时结合以上问题分析，将需要描述的元数据分为三大类，分别为数据来源信息、统计方式信息和数据基础信息。

数据来源信息包括数据发布和数据统计单位信息。发布信息用来描述评估数据的发布日期及发布形式。统计单位信息用来描述数据的生产方信息，包括数据统计单位名称、负责人、联系方式等。以上信息对数据的来源起到记录和监督作用，明确数据来源，为数据的真实性提供保障。

统计方式信息包括计算方法、统计频率、时间范围、空间范围四个方面，其可以准确描述一个数据的生产方式。其中计算方法以公式的形式展现，统计频率用以描述数据的更新周期，时间范围和空间范围可以有效区分指标的时空分辨率，这些都为评估结果在不同地区间的可比性提供了信息支撑。

数据基础信息包括限制信息、数据质量信息等。限制信息是描述评估数据安全等级的信息，级别分为公开和非公开。限制信息用以对数据的安全等级加以识别，为数据的处理、共享和保护提供参照。数据质量信息是对评估数据的辅助说明，在评估中，由于一些因素的限制或特殊情况的发生，一些数据的获取来源、统计方式不同于常见情况，或以上的信息不足以对数据进行描述，这时需要对数据做进一步的解释说明和补充。例如涉及“人均”的评估指标，有必要对该数据采用哪一年的人口统计进行说明。这些说明为评估数据的质量又提供了一层保障，同时提高了数据可持续获取的效率。结合以上，将涉及到的元数据实体进行整合分类，设计针对城市可持续发展能力评估的元数据框架如下。该元数据框架包含7个元数据子集，分别为标识信息、覆盖范围信息、内容信息、维护信息、限制信息、数据质量信息和分发信息。

(1) 标识信息：包含唯一标识数据集的信息，内容包括城市可持续发展能力评估数据名称、发布日期、数据提供方、数据分类和数据量等信息。标识信息实体是必选的。

(2) 覆盖范围信息：提供数据覆盖范围的描述信息，内容包括城市可持续发展能力评估数据的空间范围和时间范围。覆盖范围信息实体是必选的。

(3) 内容信息：包含描述城市可持续发展能力评估数据的统计内容信息、统计方式信息、数据统计单位信息、数据审核方信息及数据汇总方信息。内容信息是必选的。

(4) 维护信息：包含有关数据的更新频率的信息。维护信息实体是可选的。

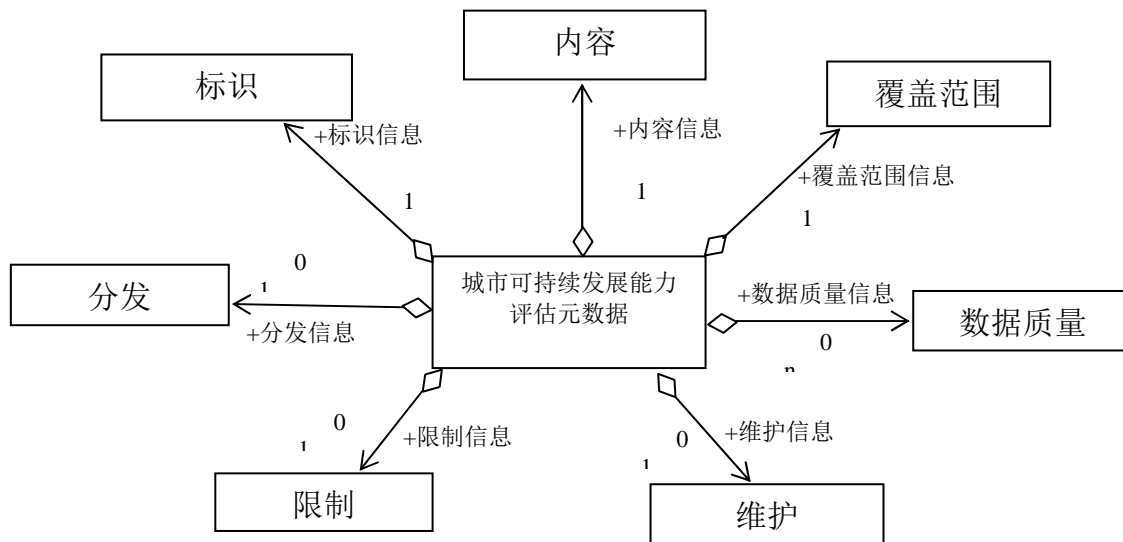
(5) 限制信息：包含访问和使用数据的限制信息。限制信息实体是可选的。

(6) 数据质量信息：包含数据的数据志说明信息。数据质量信息实体是可选的。

(7) 分发信息：包含获取数据的途径分发格式信息，如在线资源链接地址。分发信息实体是可选的。

3.2 需填报的元数据项

结合以上元数据标准，对元数据项进行整合、筛选



和补充，最终得到 16 个需要数据提供方提供的核心元数据项。即数据发布日期、数据发布形式、指标解释、计算方法、空间范围、时间范围（统计起始时间、统计结束时间）、统计频率、数据安全限制分级、数据志说明、在线资源链接地址、数据统计单位信息（单位名称、联络人、联系电话、单位地址、邮件地址）。选择其中最为重要的 8 项为必填项，包括数据发布日期、数据发布形式、指标解释、计算方法、空间范围、统计频率、数据安全限制分级、数据统计单位名称。

通过这 16 个元数据项，可以清晰完整地描述一个评估数据从生产到获取的全方位信息，为评估提供有力支撑。

数据发布日期*	
数据发布形式*	
指标解释*	
计算方法*	
空间范围*	
时间范围	统计起始时间
	统计结束时间
统计频率*	
数据安全限制分级*	
数据志说明	
在线资源链接地址	
数据统计单位信息	单位名称*
	联络人
	联系电话
	单位地址
	邮件地址

(*表示为必填项)

4 结论

针对城市可持续发展能力评估，以上元数据标准与 16 个需填报的元数据项可以为其提供完善的数据保障，支撑评估开展。

在当前的可持续发展制度环境下，一套完善的数据获取与管理机制是保证各项评估能够顺利开展的前提条

件，也是评估得以具备科学性的重要保障。在以往的评估工作中，评估数据的基础信息往往得不到重视，但这也恰是造成一些评估问题的根源所在。

相较于传统的评估数据获取方式，元数据对于评估数据的全面记录，可以完整的描述数据基础信息，提高数据的真实性，保障评估的可比性，维持评估的可持续性，减少评估中对于数据本身问题的争论。

综上所述，元数据的应用可以有效支撑科学评估的开展。随着评估对象的多样化，评估问题的复杂程度不断升高，评估人员应更多地应用元数据理论来收集和管理数据，用以提升评估的效率和意义。

参考文献

- [1]龙军,陈舜子,刘高嵩.科技评价系统中的元数据表示与管理[J].计算机技术与发展,2010,20(10):12-15+20.
- [2]吴显义.我国元数据研究现状分析[J].情报科学,2004,22(1):55-58+62.
- [3]刘若梅,等.可持续发展信息共享标准化研究和相关标准制订[J].资源科学,2001,23(1):23-28.
- [4]孙成永,等.中国可持续发展信息共享的理论与实践[J].资源科学,2001,23(1):3-11.
- [5]杜阳.规范统计数据存储模式 提高统计数据服务水平:统计元数据库体系的建立与研究[A].北京市第十五次统计科学讨论会获奖论文集[C].2009.
- [6]陈海云.以元数据建设推动统计业务标准化和管理规范化[J].数据,2007,(9):38.
- [7]郑小晴,许慧.城市可持续发展能力研究[J].西安建筑科技大学学报:自然科学版,2012,44(1):82-89.
- [8]郑伯红,廖荣华.资源型城市可持续发展能力的演变与调控[J].中国人口·资源与环境,2003,13(2):92-95

作者简介:姜天祺,硕士研究生,专业领域:系统分析与可持续,联系地址:北京市海淀区双清路18号,邮编100085, 邮箱:fido89@163.com