

环境因素识别中的数据质量初探

李爱红¹ 林宣雄¹ 弓仰军²

(1. 西安交通大学管理学院 陕西 西安 710049 ;
2. 深圳柯赛电子有限公司 广东 深圳 518054)

摘要 环境监测、环境评价等过程中的数据质量是正确识别环境因素的关键。参考国外对环境管理中的数据质量方面的研究,简述了数据产品、数据质量的概念,指出现有的环境因素识别中存在的数据库质量方面的问题,提出在环境因素识别中借鉴生命周期评价(LCA)应用全面数据质量管理体系解决问题的思路。

关键词 环境管理系统 环境因素识别 数据质量管理 LCA

中图分类号:X820.1 文献标识码:A 文章编号:1004-693X(2004)04-0019-03

建立环境管理体系的基础是对环境因素的识别和确认,识别环境因素才能确定环境目标和指标及环境管理方案。虽然 ISO14000 中规定“组织应建立并保持一个或多个程序,用来确定其活动、产品、或服务中它能够控制,或可望对其施加影响的环境因素,从中判定哪些对环境具有重大影响,或可能具有重大影响的因素”^[1],但具体如何实施并没有规定。企业可以选择多种途径从自己的角度阐述该标准。所获得的数据带有很大的个性与主观性,不利于汇总或企业间的比较,数据质量低。

1 数据质量概述

1.1 数据产品

数据产品是数据生产系统通过对原始数据的加工获得的结果,与一般意义上的实物产品类似,有生产成本、效益、利润等属性^[2]。环境因素的识别就是 1 种数据产品,通过规定的程序如环境评价系统,根据相关的评价标准分析监测数据,比较监测结果,找出主要污染源、主要污染物及其变化规律。

1.2 数据质量

数据质量包括多方面,如表 1 所示^[2]。

表 1 数据质量属性

数据质量	属 性
数据内在质量	准确性、可信度、再现性、客观性
数据的访问性	访问性、安全性
数据的关联性	关联度、时间度、完整性、值的附加度、数据量
数据的代表性	可解释性、代表性、易理解性、简洁、一致性

数据内在质量是数据本身的性质,和数据的来源有关,数据的访问性从用户使用数据的难易出发,

如权限的设置,一方面使数据的安全性提高,另一方面又使数据的访问受到限制;数据的关联性从相关数据角度出发满足数据在量和值上的使用要求;数据的代表性从类的概念出发,找出相关数据的共性。

1.3 数据质量管理体系

如图 1 所示,首先,定义数据的质量属性,找出重要的属性维;其次,测量这些属性,即找出它们的度量属性,根据度量属性建立测量标准;然后,分析实际的监测数据,根据测量标准得出结论;最后,提出改善数据质量的方案。

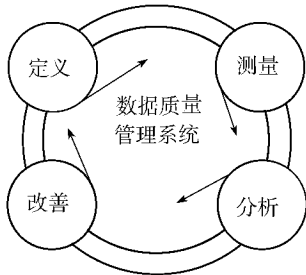


图 1 全面数据质量管理体系

2 识别环境因素中的数据质量存在的问题

环境因素的识别过程实际就是数据的加工过程,确定问题的范围,然后收集数据,最后分析数据,结果就是应关注的环境因素。组织一般在 ISO14000 的指导下根据本行业的特点和经验确定环境管理体系的范围、监测设备、人员等,监测结果的再现性、客观性差,具体如下:

a. 界限模糊。企业往往根据经验、法规、预算、市场要求、环境影响等确定环境因素涉及的范围,企

作者简介:李爱红(1976—),女,山西翼城人,硕士研究生,研究方向为数据仓库与环境管理。

业内部有关部门的职责不够明确,文件化程度低,对环境管理系统的理解不准确,只重视生产部门,如产品加工中的能源消耗、废水、废气、固体废弃物等的排放,忽视与其他管理部门的联系,对于原料供应、产品运输、销售和最终的处理等辅助生产环节中的环境因素的识别做的较少,执行人员的理解不一,影响数据的准确度和客观性。界限模糊,数据源杂,数据量少,原始数据的低可靠性降低了分析结果的数据质量。

b. 数据收集不统一。数据的收集有多种方法,如工艺流程图、专家咨询、现场监测、问卷调查、查阅文件记录等,但如何选取并具体应用它们,无章可循,不同结果的兼容性差,同一企业内部不同部门的数据差异较大,数据汇总不便,比较性差。另外,缺乏数据质量标志,如可信度、完整性等。数据的访问性差。数据的类型不统一,完整性差。这些都给数据的进一步分析带来困难。

c. 数据分析过程不清。重要环境因素的评定一般通过环境影响评价即数据的分析完成。环境因素有多种,包括污染物、能源消耗、产品本身等各个方面。按 ISO14000 规定,据相应的标准做出评价。但选用的标准和评价方法的随机性大。即使评价指标统一,评价模型的建立过程如权重的选取往往通过专家评审,带有很大的主观性。分析过程的非程序化降低了数据可信度、代表性。

总之,数据质量存在的主要问题是有关环境因素的识别和评价的数据收集、分析过程的文件程序化差,人为因素太多,致使数据的可信度、准确度、客观性、访问性、值的附加度、所获得的环境因素的再现性降低,代表性差。

3 提高环境因素识别的数据质量

借鉴生命周期评价(LCA),把全面数据质量管理体系应用到识别和评价环境因素中,从定义、测量、分析、改善各个环节控制数据的质量。

3.1 定义

确定环境评价系统的范围,按 ISO14000 统一有关术语。如所涉及的部门、流程、人员等。定义术语时要辨认数据质量属性并明确标示。可以借用数据库中的数据表示: $\langle qA, R, qV \rangle$,其中 qA 是质量属性, R 是属性关系, qV 是质量属性值,把数据质量的属性量化。

概述企业的活动,如原料供应、生产加工、产品、产量、运输、销售、废品等,从产品生命周期的各个阶段分析所涉及的范围。对内部的生产系统,可以采用过程分析法分析生产、物料中排放的废弃物,也可以采用水平对比法比较同行业的资源或能源消耗

等,其中,过程或工艺分析法对明确边界十分有用,可以找到不同数据源的联系。产品生命周期法适用性广,可以参照它的环境问题明细表识别环境因素的范围。外部系统可以采用运输链评价、管理行为评价等,通过评价供应商的管理能力、环保的实施及运输中的影响识别环境因素。定义的关键是明确边界,做到文件化、程序化,减少人为的干扰。

3.2 监测

类比生命周期分析法中的清单分析,包括元数据、数据收集机制、数据的收集、数据的量化等。

元数据包括监测单元的名称、范围、测量方法和频率、环境影响的程度及持续时间、环境法律法规的要求及评价标准、可能的环境因素、质量属性等,详细陈说监测单元的各种正常与反常情况,各种情况下的排放物的类型、可回用性,监测方法、装置、监测指标,可能的失误,数据质量指标等。元数据就是把数据的监测、来源程序化、文件化,提高数据的再现性、透明性、代表性。

数据收集机制说明监测单元包括的部门,各部门的人员编制(包括数据质量管理人员如工程师、数据质量评价员工,他们应熟悉全面质量管理体系)及负责监测的环境因素、指标、监测标准,物流及能量流,原料消耗,测量的时间及人员纪录等。把数据的收集程式化,提高数据的客观性。

数据的量化可以提高它的可比性、再现性,易于评价数据的质量。如清单分析中的嵌套数据存储格式,利用 LCA 中的物料平衡法、能量平衡法得出不同监测单元的资源 and 能源消耗指标,根据监测标准,决定监测物的类型、数量、监测频率;根据环境评价标准和评价模型转化监测结果,统一单位,利用等标污染负荷法得出不同监测点的评价指数,实行 LCA 的国家大多根据 ISO14042 进行^[3],国际环境毒物学和化学学会 SETAC 对此也有规定,如温室气体的排放可以转化为 GWP(Global Warming Potential)。关键是确定了评价指标后,结合地域特点和经验值统一评价模型中的参数标准,减少人为因素的干扰。

3.3 分析

根据定义的边界和监测的数据把监测数值和环境影响指标联系起来识别环境因素,分析数据质量。如找出某水域主要污染源,从污染程度上确定主要污染因素,找出重要环境影响因素。类似于清单分析中的影响评价。根据不同的影响类型对监测数值分类,如资源消耗型或生态影响型;再对数据归一化,与有关的监测标准对比或利用评价模型,把不同的监测指标统一;最后,根据影响大小给各指标打分,得出重要的环境影响因素。根据评价标准选定

某项环境影响指标的计算方法后 , 一般不要随意重选 , 这样便于对比不同时期的环境因素。

数据质量的分析可以采用可信度评价系统 CAS (Confidence Assessment System)^[4] , 从环境因素识别的各个阶段为数据的质量属性打分。根据数据的来源 (实测值、文献记录、专家评审等)、时效性、聚合度、环境影响程度等给予分值得出数据质量分数 ; 然后再从质量属性指标方面打分 , 如是否达到完整性、准确度、再现性、比较性等方面的要求 , 达到加分 , 否则不加分 , 得出数据质量指标分数 , 最后把数据质量分数和数据质量指标分数相乘得出总的的数据质量。具体请参阅文献 [4]。核查表法也是采用比较多的一种方法 , 把数据和同行业的其他监测值比较或和理论值比较。监测单元内部或之间相似监测指标可以用加权平均法。

3.4 改善

识别出重大环境因素后 , 根据有关的环境法规、企业特点及其环境方针、目标和指标制定环保方案 , 改善环境质量。同时 , 根据分析阶段的结果找出主要数据质量因素改善数据质量。分清是受哪方面的影响 , 如数据源、监测方法、监测模型等。关键是借鉴 LCA 实施全面质量控制 , 把环境影响因素识别过程程序化、文件化。

PHASETS^[5] 是一种描述生产工艺过程的概念模型 , 它把数据质量融入环境管理系统中 , 如图 2 所示 , 把数据处理过程分为 6 个阶段 , 指出在各个阶段如何保证数据质量 , 可以借鉴它在生产过程中改善数据质量。具体请参阅文献 [5]。

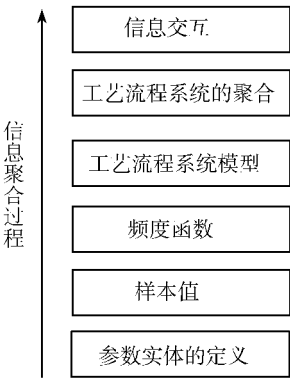


图 2 PHASETS 模型

4 举 例

4.1 数据质量量化

根据表 2 , 对数据质量属性量化。共考虑了 8 项 , 总分为 80 分 (部分取值参考了文献 [4])。

根据表 2 给数据打分后 , 再根据式 (1) 得出数据质量得分 (Data Quality Score) :

表 2 数据质量属性量化

属 性			属 性		
分 数			分 数		
可 信 度	政府检测机构	10	代 表 性 (测 量 范 围)	企业内部或乡镇	10
	企 业	9		全省范围	8
	数据源未知	2		地域范围	6
再 现 性	印刷文档	10		全 国	4
	数据库	9		全国以上	2
	非印刷文档	5		其 他	0
客 观 性 / 准 确 性	实际检测数据	10	关 联 度	单个检测值	10
	样本数据	7		月平均值	8
	模型计算值	7		年平均值	6
	专家意见	6		一般取值	4
	类型未知	2		其 他	2
时 间 度	1 ~ 2 a	10	标 准 度 / 可 比 度	国家标准	10
	3 ~ 4 a	9		地方标准	8
	5 ~ 6 a	7		企业标准	6
	7 ~ 8 a	5		其 他	2
	9 ~ 10 a	4	访 问 性	统一的监测网络	10
	11 a 以上	3		规范的文档资料	9
	时间未知	2		内部资料	7
				其 他	4

$$S_{DQ} = (\sum_{i=1}^8 S_i) \times 100 / 80 \tag{1}$$

式中 : S_{DQ} 为数据质量得分 ; S_i 为单个属性得分。

4.2 评价实例

以 1998 年西安市大气污染调查情况为例 , 监测由西安市环境监测站承担 , 监测范围主要是西安市县及县以上企业 , 监测频率据有关的国家标准进行 , 监测结果以报告书的形式存放。根据表 2 , 各项属性得分如表 3 所示。

表 3 数据质量得分

属 性	得 分	属 性	得 分
可信度	10	关联度	6
再现性	10	标准度/可比度	10
客观性/准确性	10	访问性	9
时间度	10	总分	75
代表性	10	S _{DQ}	93.75

由上可知 , 大气监测报告书的数据质量比较高 , 采用等标污染负荷法确定大气污染物的主要来源是燃煤 , 主要污染物是二氧化硫 , 由此确定了主要的排放企业 , 完成环境因素的识别。

5 结 语

数据质量在识别环境因素中很重要 , 如何建立有效的分析模型提高数据的客观性、再现性、准确性是研究的难点。提高数据收集和分析过程的程序化 , 借鉴 LCA 的清单分析的经验 , 实行全面数据质量管理是对该问题的一个试探 , 具体如何实施有待进一步研究。

(下转第 27 页)

理地控制下泄流量,既能充分利用淮河干流环境容量,又能避免污染水体过量下泄造成水质污染。因此,做好水质预测预报等基础工作非常重要。要建立水质模型,对水质进行科学预测,才能为调度提供决策依据。

5 关于进一步发挥水利工程调度改善水质的建议

新修改的《中华人民共和国水法》一个显著特征就是对过去水资源开发、利用中重开源、轻保护,重经济效益、轻生态与环境效益的不足进行了修改,规定了在水资源开发、利用过程中,要维持河流合理流量,保证生态用水。维持河流合理流量,维持河流自然净化能力,是《中华人民共和国水法》对水利工程调度提出的新的要求。

淮河流域水闸防污调度缓解了水污染。为进一步发挥水利工程调度改善水质的作用,以下 3 个方面的工作需要完善和努力。

a. 防止河道断流,提高水体自净能力。淮河流域很多水污染事故是由于污染水体集中下泄诱发的。减少水污染事故发生,除了减少排污,同时也要避免水闸长时间关闭,保证河流一定的生态流量。各级水行政主管部门要树立可持续发展观,合理开发利用水资源,统筹兼顾生产、生活和生态用水。

b. 对洪水实施风险调度,充分利用汛期洪水,实现洪水资源化。淮河流域多年平均年径流量为 581.83 亿 m³,目前全流域污废水排放量约 35 亿 m³,污径比 1:17。但由于淮河流域降雨分布时空严重

不均,汛期大量水资源得不到有效利用,导致枯水期水污染十分严重。如果能有效利用洪水资源,淮河流域的水污染将会得到缓解。

合理开发利用洪水资源,除了建设必要的水利工程之外,通过对现有工程科学调度可以发挥一定的作用。比如对汛期洪水实施风险调度,确保汛后水利工程能蓄积到充足的水源,对淮河干流行、蓄洪区等湖泊洼地进行合理利用,提高水资源利用率。

c. 合理调度,实现清污分流。通过水利工程调度,改变污染水体走向,对确保重要供水水源地水质安全有着非常重要的作用。位于沂沭河流域的石梁河水库是江苏连云港市重要供水水源地,库区养殖业也十分发达。20 世纪 80 年代以来,由于受新沭河下泄污染水体的影响,多次发生死鱼事故,引起江苏、山东两省水污染边界纠纷。2003 年汛期沭河发生第一场洪水时,由于洪水前峰水质污染严重,为了避免石梁河水库再次受到水污染危害,沂沭泗水利管理局在防汛调度时,将进入新沭河污染水全部调入老沭河,借道江苏新沂河内的排污走廊入海,保证了石梁河水库水质安全。

参考文献:

[1] 程绪水,贾利,沈哲松. 2001~2002 年度淮河沙颍河水污染联防调度与经验[J]. 治淮, 2003(12):3.
[2] 沈哲松,杨智,刘耀宾. 加强污染联防 减轻水污染危害[J]. 水资源保护, 2001(3):14.
[3] 沈哲松. 淮河治污 呼唤联防[N]. 中国水利报, 2001-05-26(3).

(收稿日期 2004-02-24 编辑:傅伟群)

(上接第 18 页)紧缺、水污染严重、城市生活用水多样化及低质杂用水需求量日益增长的今天,我们必须改进现有的给排水系统,开源、节流,加强管理,保证对水资源的利用“既能满足当代人的需要,又不对后代人满足其需要的能力构成危害”,给我们的子孙留下一片青山绿水。

参考文献:

[1] 杭州年鉴编辑部. 杭州年鉴(1987~2001)[M]. 北京:中华书局,1987~2001.
[2] Sieker F. On-site stormwater management as an alternative to conventional sewer systems: a new concept spreading in Germany[J]. Wat Sci Tech, 1998, 38(10):65~71.
[3] Pratt C J. Use of permeable, reservoir pavement constructions for stormwater treatment and storage for re-use[J]. Wat Sci Tech, 1999, 39(5):145~151.
[4] 郭勇伟,王伟. 城市生活用水的多样化趋势雨水资源合理利用途径研究[D]. 杭州:浙江大学, 2001.

(收稿日期 2003-05-03 编辑:高渭文)

(上接第 21 页)

参考文献:

[1] 徐敏. 快速精通 ISO14000:1996[M]. 广州:广东经济出版社, 2002.
[2] Richard Y W, Mostapha Z, Yang W L. Data Quality[M]. Kluwer Academic Publishers, 2001.
[3] Zobel T, Almroth C, Bresky J, et. al. Identification and assessment of environmental aspects in an EMS context: an approach to a new reproducible method based on LCA methodology[J]. Journal of Cleaner Production, 2003(10):381~396.
[4] Tzivilakis J, Broom C, Lewis K A, et al. A strategic environmental assessment method for agricultural policy in the UK[J]. Land Use Policy, 1999(16):223~234.
[5] Raul C, Ann-Christin P. Industrial environmental information management for technical systems[J]. Journal of Cleaner Production, 2001(9):424~435.

(收稿日期 2003-07-06 编辑:傅伟群)