

# 大数据技术在环境监测中的应用

解辉

(天津市环境监测中心,天津 300191)

**摘要:**简述了环境监测信息化建设应用现状及问题;分析了大数据技术的特点及其在环境监测信息化中的优势。提出,基于“大数据”技术,可以建设环境监测大数据中心,突破传统关系型数据库局限性,运用大数据的核心技术,实现环境质量综合评价技术的新突破,提升环境状况综合预警能力;应用大数据可视化技术,提高科学决策水平,提高人体健康风险评价能力;建立环保舆情分析云平台,提升公众服务能力。

**关键词:**环境监测;大数据;环境质量综合评价;可视化;环保舆情

中图分类号:X821

文献标志码:A

文章编号:1674-6732(2016)04-0062-05

## Application of Big Data Technology in Environmental Monitoring

XIE Hui

(Tianjin Environmental Monitoring Center, Tianjin 300191, China)

**Abstract:** This paper described the current status and issues of information construction of environmental monitoring and analyzed the characteristics and advantages using big data technology in environmental monitoring. It was proposed that based on the big data technology, environmental monitoring big data center could be constructed to break through limits of the traditional data bank. Application of the big data core technology might find new breakthroughs in the comprehensive evaluation technology of environmental quality. The first was to improve comprehensive early warning ability of environmental conditions; the second was to use the big data visualization technology for improving in the level of scientific decision-making; the third was to improve the ability of human health risk assessment; the fourth was to establish a “cloud” platform for public opinions and analysis of environmental protection in order to improve the public service ability.

**Key words:** Environmental monitoring; Big data; Environmental quality comprehensive evaluation; Visualization; Environmental protection public opinion

### 1 环境监测信息化建设应用现状及问题剖析

#### 1.1 现状

从信息化角度而言,对环境监测过程合理而简要的描述就是获取、分析、呈现环境监测数据的过程,每一环节皆与信息技术的发展密不可分。伴随着中国环境监测事业40余年的发展历程,环境监测信息化工作也开展了近30年,已成为环境监测现代化的重要组成部分。在环境监测工作的每一次跨越发展中,环境监测信息化建设也同样经历了重大突破与挑战,大体归纳为3个阶段:

第一阶段,20世纪70年代中后期至90年代中期,环境质量评价与污染动态分析报告的初级阶段。环境监测工作经历了20世纪70年代的起步阶段、80年代的调整巩固阶段和90年代初期的充实提高深化阶段,这期间对信息化的需求逐步加

大,建立了一些小规模信息处理系统,工具化统计软件得到应用,分析报告水平逐步提高,对提高民众环境意识、为各级政府下决心加大投入进行污染综合防治起了很大的推动作用<sup>[1]</sup>。

第二阶段,“九五”至“十五”期间,以单要素为主的环境质量综合分析与表征技术相结合的发展阶段。巩固和完善了国家环境监测网,环境监测能力迅速提高,环境监测数据及信息的收集、处理、传输已实现计算机化,并应用多媒体技术编制环境监测报告,初步建立了全国、省、流域环境监测地理信息基础数据库和数字地图。这期间现代信息技术得到普遍应用,在运用先进、简明、实用的综合评价

收稿日期:2015-12-04;修订日期:2016-02-29

作者简介:解辉(1967—),女,研究员级高级工程师,本科,从事环境监测与信息化工作。

方法和表征技术说清楚环境质量现状及其变化规律方面,取得了重大突破。

第三阶段,“十一五”至“十二五”期间,基于多环境要素环境质量综合评价的探索阶段。环境污染呈现更加复杂的态势,环境保护的一项核心任务就是探索一条代价小、效益好、排放低、可持续的中国特色环境保护新道路。国家适时提出了构建先进环境监测预警体系的大思路,这一时期的环境监测信息化技术主要应用在综合指数法、模糊综合评判法、灰色聚类法、人工神经网络法、投影寻踪模型法、基于层次分析法的区域环境质量综合评价体系等,多方尝试突破按空气、地表水、噪声、海洋等单一环境要素评价模式,全面展开了向基于多环境要素的综合信息评价技术的探索。

## 1.2 面临的问题与机遇

进入“十三五”以来,环境监测事业面临着前所未有的压力与挑战,要实现全面设点、全国联网、自动预警,尤其对生态环境监测信息的集成、共享、开发与应用提出了更高要求,因此,环境监测信息化面临的问题愈加突出。

(1)宏观层面上,目前环境监测信息呈现工作虽然有一些技术方法尝试将社会、人口、经济等指标纳入环境质量综合评价体系,但总体仍是一元应对模式,没有整合社会各方资源,建立起环境问题高效治理的多元协同模式;

(2)微观层面上,目前的信息利用技术方法均存在一定缺陷,如:指数评价法难以反映环境质量的综合状况;模糊综合评判法对每个参评指标的权重值需要人为给定,且参评指标数量较多时,往往低估了主要指标的贡献<sup>[2]</sup>。这些方法大多限于环境质量评价,并未真正体现环境和社会系统的关联和可持续发展思想;

(3)数据库管理仍然沿用传统模式,各系统独立部署,数据分散,无法实现跨系统的数据分析;大量外部数据未能有效管理,如经济社会发展、气象、地理国情、人口等,尤其是大量非结构化信息,如互联网、健康、金融、经济、交通等活动过程所产生的各类数据,包括微博、文字、图形、视频等尚未开发利用;系统间接口效率低下,对于数据存储、数据共享等造成较大压力,维护成本高昂。

## 2 大数据技术特点及在环境监测信息化中的优势

### 2.1 大数据技术及其特点

大数据(Big Data)是从多种来源中搜集得到

的海量数据信息的总称,是数据量大、类型复杂、需要即时处理和价值提纯的各类数据。大数据技术是一代全新的数据科学领域的技术架构或模式,是对预测分析、数据挖掘、统计分析、人工智能、自然语言处理、并行计算、数据存储等技术的综合运用<sup>[3]</sup>。大数据有4个基本特征:一是数据种类多,处理和分析方式区别大。二是数据量巨大,据统计,每天产生2.5艾字节(quintillion bytes)的数据信息,全球将近90%的数据是过去2年创造出来的。三是处理速度快,多数为高速实时数据流。四是价值密度低,需要价值提纯<sup>[4]</sup>。实质上,容量大是大数据的整体特点,增速快是大数据的过程性特点,多样性是大数据的具象性特点,价值性是大数据的时代性特点<sup>[5]</sup>。

### 2.2 大数据技术应用于环境监测信息化的优势

#### 2.2.1 提升生态环境综合预警能力

应用大数据从海量、分散、实时变化的环保数据中挖掘出有价值的信息,通过分析获取当前状况的总体描述,进行态势推演,综合研判和决策,及时提供给政府和公众。第一,大数据能够驱动预警预报的精确度。在大气污染预警中,气象局可以对星及风云系列气象卫星反馈的数据进行监测和分析,准确地预警大气污染过程,作为启动预警的技术支撑。第二,大数据能够驱动预警信息的时效性。大数据的时效性,在于对海量信息的快速正确研判。理论上讲,污染事件发生后,大数据的功能可以加倍放大,并以一种没有提前规划却及时有效的方式监测和推动事件进程,如发布环境污染预警、发布避难场所信息、请求救援等,海量涌现的数据信息在网络平台上下呼应,相互关联,这都是传统的污染监测、信息公开手段望尘莫及的<sup>[6-7]</sup>。

#### 2.2.2 提升科学决策水平

大数据可视化技术应用在环境监测中有2个优点:首先是基于可视化技术的环境监测数据分析结果可以提高科学决策水平,让数据自由流动,不断提高其附加值,并且将分析结果完整地呈现给公众,将各种环境数据信息转换为有资讯价值和行动含义的表达,为政府决策管理层、企业生产管理者、环境组织和社会公众提供决策、管理、生产、行动和生活信息;其次,通过虚拟的数据及治理模型的立体化展现是一种极具创意的环境治理方式,可以测试环境保护治理方案是否有效。

### 2.2.3 提升健康风险评价能力

当前环保的核心已逐步转移到保护民生健康方面,有毒有害污染、重金属污染以及环境中病原体等的监测与评价必将逐步列为重点监测内容,这不仅需要先进的仪器设备和技术方法,更需要科学的健康风险评价技术。大数据下的人体健康风险评价,就是通过有害因子对人体不良影响发生概率的估算,评价暴露于该有害因子的个体健康受到影响的的风险。其主要特征是以风险度为评价指标,将环境变量的大数据与出行、疾病预测、城市资源配置相关联,将环境污染程度与人体健康联系起来,通过模型分析,定量描述污染对人体健康产生危害的风险。

### 2.2.4 提升公众服务能力

应用大数据技术可以建立面向环保的舆情分析云平台,警示人们对于环境保护、生态文明建设等问题的重视,推进环境保护的公共治理问题。例如:对于污染源企业的部分监管工作,可以借鉴“数据众包”思路,通过平台自助式地把各类数据采集任务发布给公众人群,公众利用手机参与应用,即可直接完成各类数据采集任务,整个数据采集过程无须人工干预。同时利用多种数据挖掘及自然语言处理技术对网页内容进行分析,获取新闻事件主题及关键信息,定位环境事件发生地点,发现热点事件并跟踪发展趋势,一方面使得环境管理者可以更好地了解社会热点事件、政策实施效果等;另一方面,可以将公众交互行为产生的最新信息及时记录下来进行分析,进而面向社会开展精细化服务,实现个性化的推荐功能,为公众提供更多便利,产生更大价值。

## 3 环境监测大数据采集与集成策略

### 3.1 环境监测大数据的采集类别

#### 3.1.1 根据内容分类

环境监测大数据来源主要有两个方面,一是监测部门自身产生的数据,包括生态环境监测数据和污染源排放数据,以结构化数据为主,主要数据的时间跨度将近40年;另一种是外部产生的数据,主要是人群活动产生的与环境相关的数据信息,非结构化数据所占比例较高,时间跨度不等,典型的如气象数据在百年以上,而现代人与环境相关数据往往年代较近。

(1)生态环境综合监测数据。可细分为大气监测、水质监测、土壤监测、噪声监测、降水监测、气

象监测、电磁辐射监测、森林植被防护监测数据等,涵盖环境质量监测点、环境功能区、水源保护区、自然生态保护区等环境敏感区域。由全国环境质量监测网产生,用于描述整体生态环境质量状况。

(2)污染源排放数据。可细分为废气污染源、废水污染源、固体废物在线监管数据等。包括污染源基本情况、污染源监测、治理设施运行、总量控制、污染防治、排污费征收、监察执法、行政处罚、环境应急等环境监管信息。

(3)人群活动产生的与环境相关的数据信息。可细分为经济社会发展规划、基础气象、地理国情、人口经济统计、用水量、用电量、生活垃圾、交通等。主要来源于各类官方统计年鉴、网络、媒体等,以数据库、文件、图形、视频等方式存在。

#### 3.1.2 根据数据的生命周期分类

(1)冷数据,历史归档数据及未经处理的原始数据。在生态环境监测数据及污染源排放数据中,有相当比例是该类数据。特点是极少被访问,计算需求较低。存储需要大容量硬盘,低端CPU和较低内存即可;

(2)温数据,供历史查询、周期较长的各种分析报表数据等。如:环境质量日报、月报、季报、年报,统计年鉴信息等。特点是访问频率较低,运算要求中等。存储需要大容量硬盘,低端CPU和相对较低内存;

(3)热数据,指当期使用数据,包括处理后的入库及部分原始数据。如水、气、污染源、气象等需要实时发布的在线监测数据及模型数据等。特点是频繁被访问、高运算、高并发、高实时。存储需要大容量高速硬盘,中端CPU、高内存和高速网络。

### 3.2 数据集成策略

大数据集成的主要目的是保证数据质量,这是大数据应用的前提。为保证大数据分析结果的准确性,需要将不真实的数据剔除掉,挑选出最正确有效的数据。具体步骤是将来自底层数据源系统的不同类型数据,通过数据抽取、分发、清洗、转换和装载等过程,将源数据存储到数据共享平台中。根据环境监测大数据类型及特点,在集成过程中应侧重考虑以下环节。

(1)数据转换模块。对类型或格式不一致的数据进行转换。由于历史原因,环境监测数据始终存在编码不统一的问题,所以应将不同数据源或网元的数据转换成统一格式编码;

(2)数据补采模块。为应对数据中断、数据不完整等问题,需要根据不同类型数据设定不同的补采规则,对于冷、温数据,可以人工进行补采,而对于热数据要实现补采任务的自动发起;

(3)数据清洗模块。负责对“脏数据”进行剔除,消除数据的不一致。“脏数据”包括不规则数据、不符合事实数据。要设置尽量详细的规则,如:取值范围、完整性规则、拼写检查等;

(4)采集监控模块。整个数据采集与集成过程需要发起大量的数据采集、转换、加载等任务,所以需要具有完备的任务调度管理及监控能力,通过界面对各类任务进行配置、启动、跟踪、管理和监控,包括加载任务的状态,对加载任务进行启停和优先级设置等操作。

## 4 环境监测大数据平台架构及其主要功能

### 4.1 大数据管理与传统数据库的区别

传统数据库一般为关系型数据库,有很大局限性,很难满足大数据高效率存储和访问的需求,以及高扩展性和高可用性的要求。大数据中心管理与传统数据库管理存在本质区别。

(1)数据规模不同。传统数据库以 MB 为基本单位;大数据则以 GB、TB、PB 为基本单位;

(2)数据类型不同。传统数据库以结构化数据为主,种类单一;大数据包括结构化、半结构化和非结构化数据,种类繁多;

(3)产生模式不同。传统数据库先设定模式,后产生数据;大数据很难预先设定模式,通常是数据出现之后才有模式并且不断演化;

(4)处理对象不同。传统数据单纯是处理对象,大数据已经成为一种资源;

(5)处理工具不同。传统数据仅需一种或少数几种;大数据则需要多种不同的处理工具。

传统模式下的环境监测数据库一般为独立部署,数据分散,形成数据孤岛,无法实现跨部门数据分析;系统间接口效率低下,数据共享会对原有系统造成较大压力;数据存储、维护成本高昂,不适用大数据容量的急速膨胀。

### 4.2 环境监测大数据平台架构

“大数据”平台与一般的业务信息应用系统有本质区别。大数据平台不是一般的“生产型”数据平台(水质、空气、污染源在线、例行监测、lims 等是生产业务性数据平台),而是多方汇聚、吸纳、整

合、统筹,再次组织社会各类、甚至各行业信息,这些均参与或涉及生态环境监测信息的主动产生、专业产生、间接产生,是大数据平台的上游、源头、供给侧。这些多元、异构、结构化和非结构化信息在大数据平台大量交织,信息价值密度差异很大。大数据平台的一个重要功能就是利用信息挖掘手段,根据环保专业的规则、策略,进行更高层次的改造、转换、有序组织,从单一因子、单一维度向多维关联转化,从而认知和洞察传统的、面向某一具体领域的环境监测数据库难以准确呈现的规律、知识、趋势特征等。

环境监测大数据平台主要架构示意如图 1。由数据采集、存储、处理、共享、应用等系统组成。整体技术架构是基于物理资源及网络资源,采集整合所有环保相关的数据汇聚于大数据平台;对数据进行分析挖掘,提供基于可视化的数据分析结果应用,建立面向对象的环保业务应用系统和信息服务门户,为第三方环保应用提供商提供统一的应用展示平台,为公众、企业、政府等受众提供环保信息服务和交互服务<sup>[8]</sup>。

#### 4.2.1 数据采集与存储

为保证环境监测大数据分析结果的准确性,数据中心需要将不同类型的环境数据,通过数据抽取、分发、清洗、转换和装载等过程,将大数据中不真实的数据剔除掉,保留最准确的数据,再将源数据存储到数据共享平台中。存储的数据按照生命周期配置为历史归档数据和当期使用数据,提供查询接口并对外开放。数据仓库可以满足大量的环境监测实时数据分析处理的要求。

#### 4.2.2 数据处理与共享

数据处理是按预先定义的计算处理需求进行批量计算处理,实现数据建模、数据计算功能。如:针对大气、水体、土壤的环境质量连续监测,引入预测模型,预测环境质量变化趋势,从而为环境容量核定、产业结构布局、城市规划建设、资源开发利用等提出更加合理的生态环境保护建议。同时,在污染企业监控预警方面,引入生产排污工况监控,深度监控排污企业生产、排放、存储、运输各个环节,为监察部门提供可靠的执法依据,并结合大数据挖掘技术来计算企业排污对当地环境的影响。

数据共享层支持建立数据服务的标准化接口,促进联防联控,将监察执法处理情况、环境监测情

况、污染源在线监控情况进行统一汇总分析,实现 跨部门协作,切实加大环境监管力度。

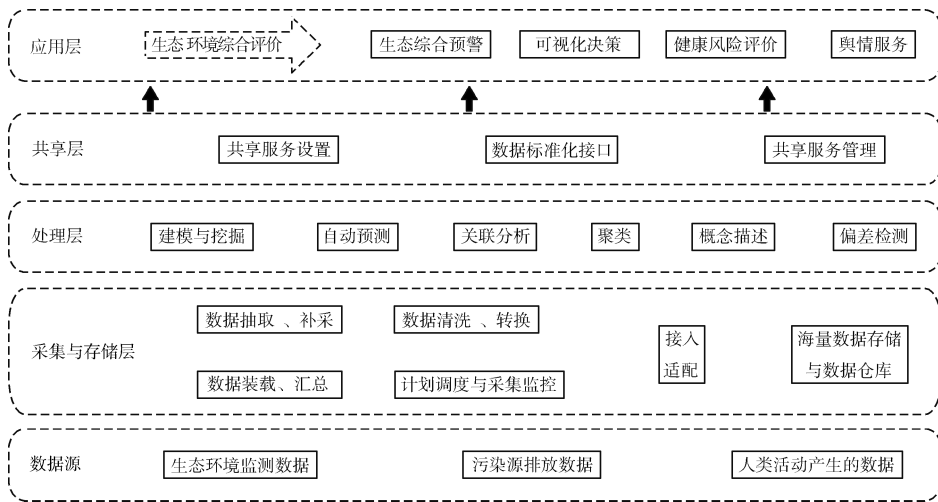


图 1 环境监测大数据平台功能架构

#### 4.2.3 数据挖掘应用

应用环境监测大数据平台的最大优势,就是把大量的、以往没有得到重视的水质、空气、污染源监控视频信息的“大数据”进行集成、挖掘,把来自社交网络的大量很发散、缺乏有效组织、价值密度低,但仍具有生态环境信息意义的数据资源进行嗅探、抓取、整合、处理,赋予数据系统新的深度,从而达到提高环境质量综合评价技术水平的目的。

#### [参考文献]

[1] 魏复盛. 我国环境监测的回顾与展望[J]. 环境监测管理与技术, 1999, 11(1): 1-4.  
 [2] 商博, 于光金, 王桂勋, 等. 基于 PCA 的区域环境质量综合评价及

应用实例研究[J]. 中国环境监测, 2013, 29(5): 12-15.  
 [3] 张锋军. 大数据技术研究综述[J]. 通信技术, 2014, 47(11): 1240-1242.  
 [4] 陶安. “大数据”时代下的信息资源建设[J]. 软件导刊·教育技术, 2014(10): 59-60.  
 [5] 任志锋, 陶立业. 论大数据背景下的政府“循数”治理[J]. 理论探索, 2014(6): 82-83.  
 [6] 黎如昊, 黄云生. 广东省地表水自动监测系统数据传输协议的设计及应用[J]. 环境监控与预警, 2016, 8(2): 59-62.  
 [7] 董艳平, 喻义勇, 金鑫, 等. 基于在线监测数据的青奥会废气重点污染源临时管控效果评估[J]. 环境监控与预警, 2015, 7(4): 9-11.  
 [8] 戴秀丽, 朱培瑜, 王蕾, 等. 环境监测数据审核体系的回顾与展望[J]. 环境监控与预警, 2015, 7(3): 52-55.

栏目编辑 周立平

(上接第 19 页)

[2] ZHANG W H, WU Y X, SIMONNOT M O. Soil Contamination due to E-Waste Disposal and Recycling Activities: A Review with Special Focus on China[J]. Pedosphere, 2012, 22(4): 434-455.  
 [3] ZHAO G F, ZHOU H D, WANG D H, et al. PBBs, PBDEs, and PCBs in foods collected from e-waste disassembly sites and daily intake by local residents[J]. Science of the Total Environment, 2009, 407(8): 2565-2575.  
 [4] SONG Q B, LI J H. A systematic review of the human body burden of e-waste exposure in China[J]. Environment International, 2014, 68: 82-93.  
 [5] ZHAO G F, WANG Z J, DONG M H, et al. PBBs, PBDEs, and PCBs levels in hair of residents around e-waste disassembly sites in Zhejiang Province, China, and their potential sources[J]. Science

of the Total Environment, 2008, 397(1-3): 46-57.  
 [6] 董亮, 张秀蓝, 史双昕, 等. 新型持久性有机污染物分析方法研究进展[J]. 中国科学: 化学, 2013, 43(3): 336-350.  
 [7] 董黎静, 戴玄吏, 滕加泉, 等. SBSE-TDU-GC/MS 法同步测定饮用水源水中 16 种多溴联苯化合物[J]. 环境监控与预警, 2015, 7(5): 37-41.  
 [8] 林静佳. 应用 DPX 方法快速测定蔬菜中有机磷残留量[J]. 广州医药, 2012, 43(3): 57-59  
 [9] 李凯, 刘永强, 张金玲, 等. DPX-Q 萃取吸管结合气相色谱/串联质谱法同时检测水果蔬菜中 102 种农药残留[J]. 分析试验室, 2015, 34(2): 236-242.  
 [10] 李桂香, 张大为, 李文. PTV-GC/MS 检测环境样品中的痕量有机氯农药[J]. 实验室研究与探索, 2010, 29(11): 241-242, 375.