

大数据能否破解数据造假难题?



林宣雄,1962年9月生,江苏宜兴人,西安交通大学环保大数据研究中心主任。1998年获陕西省优秀青年科技工作者称号,1999年主持研发的“国家环境监理信息系统”被确定为国家级火炬计划项目,同年获省优秀软件人才称号;2003年获首届省优秀青年科技创新奖;2008年承担环境保护部节能减排核心重大软件项目。长期从事环保物联网技术研究及应用。

对话人:西安交通大学环保大数据研究中心主任 林宣雄
采访人:本报记者李莹

大数据可否解决数据造假问题?

■可以通过环比、同比、类比,发现数据中的异动

中国环境报:在线监测数据造假已经成为当前环境管理工作中的大难题。一些专家认为,以大数据为代表的新技术可以破解这一难题。您怎么看?

林宣雄:我也这样认为,用大数据的分析方法可以发现在线监测数据造假。其实,这也是我的研究重点领域之一。

在解释大数据如何破解在线监测数据造假问题之前,必须先了解大数据的基本特征。

大数据是近年来的一个热词。什么是大数据分析?通俗地讲,就是运用一些数据分析软件工具,对海量的、混杂的数据进行分析,在融合丰富的实践基础上,运用创造性思维,得出突破性的结论。大数据包括3个特征:一是具有海量的、混杂的基础数据;二是熟练运用Hadoop和Spark等分析软件工具;三是具有开放的、有创造性的思维方式。只有具备了这3项,才能真正做好大数据分析。

大数据和以前的数据分析有3个明显的区别:一是原来的数据分析针

对部分样本,大数据是所有的数据都要参与计算;二是大数据中,相关关系重于因果关系;三是大数据允许混杂数据甚至错误数据。

我们得到海量数据后,首先,要对这些数据进行分析,即对其本身进行分析,如污染源数据、环境质量数据等。其次,要做扩展分析,如分析清楚污染源数据和环境质量有什么关系。第三,要做延伸分析,即将污染源数据、环境质量数据和人口数据、产业结构的数据结合起来进行分析。如果不做这些分析,就不算是大数据分析,或者说做不好大数据分析。

中国环境报:您刚刚谈到,大数据允许混杂数据甚至错误数据,这是为什么?

林宣雄:大数据允许混杂数据甚至错误数据。这是因为,大数据能够通过造假数据的特征将其辨识出来。造假的数据和平常的数据不一样,可以通过环比、同比、类比,发现数据中的

基础数据获得方面急需哪些突破?

■首先要形成拉直、拉真、拉准的机制

中国环境报:要进行大数据分析,前提是有海量数据。现在我们具备大数据分析的基础吗?

林宣雄:近年来,环保物联网的建设已经为环境大数据分析提供了一定的基础。什么是环保物联网?通俗地讲,就是将应用在环保领域的传感网连接到互联网上就构成了环保物联网。例如,各种环境监测设备收集了大量的环境相关数据,如COD排放量、SO₂排放量、空气质量等,将这些数据传到互联网上,就形成了环保物联网。

目前,我国的国控污染源已经全部联网,仅污染源一类每年全国就有近50亿条的基础数据产生,这些数据为我们进行大数据分析提供了一定基础。

中国环境报:据了解,目前我国地市级、县级污染源联网推进工作并不十分顺利,这是为什么?

林宣雄:目前的数据造假不仅是

大数据可否使数据分析更为科学?

■没有大数据的创新思维,即使躺在大量的数据上,也不会有突破性的结果

中国环境报:环境保护部近日表示,要大力提高数据采集、合成和综合分析能力,提高环境管理的精细化水平。现实工作中发现,如果没有运用科学的分析方法,即使有大量基础数据,也不能得出科学的结论。对此,您有何看法?

林宣雄:确实如此。现在很多人在谈大数据的概念,但其实谈的人多,做的人少,真正运用互联网思维解决实际问题的例子目前还不多。还有人认为,只要运用大数据就可以解决环境问题,这也是一个思维误区。

◆李合敏

基层环保部门在向司法机关移送危废案件时,因证据不足被退回补充侦查的现象经常会发生。笔者认为,在移交过程中易出现以下一些问题:

一是部门之间协调配合不紧密。因危废污染构成刑事犯罪的,需环境监测大队、环境监测站等多部门密切配合。但往往因协调不够,难以形成执法合力,使证据收集不规范。

二是检验检测体系不完善,取证方法不科学。相关设备投入不足,致使在危废的检测上大多依靠pH值,对重金属或者危险化学品没有更科学的检测手段,取证时缺少必要的环境监测设备。

危废案件移交证据不足怎么办?

一方面,大数据分析需要在大量基础数据、大量实践的基础上进行。没有数据、没有物联网的高效运转,大数据分析无从谈起。

另一方面,在各种数据、各种技能都已掌握的基础上,还要有大数据的思维。根据多年的实践和对环保问题的深层理解,将混杂的数据融会贯通进行思考,才能产生思维上的突破。如果没有大数据的创新思维,即使躺在大量的数据上,也不会有突破性的结果。

现在很多地方都在做生态文明指标体系、绿色创建指标体系。但是,他们对指标的运用仍停留在传统思维层

环境监测数据的真实性直接关系到环境管理和综合决策。然而,目前的在线监测数据造假已经成为困扰环保部门的难题之一。近年来,大数据如同浪潮一般席卷全世界,直接带来了科研、商业、政府运作方式乃至人类思维方式的变革。很多人期待,环保大数据的应用和发展能够从根本上扭转监测数据造假局面。环保大数据可否发现、纠正数据造假?要推动环保大数据的发展和应用还需突破哪些障碍?

目前,我们正通过分析数据异常波动为环保部门精准执法提供线索。例如,我们每周都要为浙江省嘉兴市环保局提供一份在线数据出现异动的企业名单,环保局可以根据这份名单,有针对性

地执法检查,执法效率可大大提高。

中国环境报:据您了解国外有没有通过大数据研究,发现环保数据造假的案例?

林宣雄:目前,国外大数据在环保领域还没有典型案例。这是因为,一些发达国家的环境问题已经得到了较好解决,而大数据的概念是这几年才提出的。但大数据应用在国外有一个经典案例,值得借鉴。2009年,甲型H1N1流感暴发的几周前,谷歌公司通过对运用谷歌软件搜索流感相关信息的人群进行分析,成功地预测了流感在美国境内的传播,其分析结果甚至具体到特定的地区和州,并且非常及时,令公共卫生官员倍感震惊。因为通常来说,美国疾病控制中心要在流感暴发一两周之后才可以做到这些。大数据拥有如此大的威力,对于环境问题十分严峻的我国来说,其应用意义更加巨大。将大数据应用于环保领域,也将成为我国的一大创新。

中国环境报:对污染底数不清,也是环保数据利用的一个重大难题。您认为,当前有哪些数据急需收集?

林宣雄:污染源数据方面,目前只有末端的数据,也就是排放口的数据,缺乏过程的数据。如果有了生产过程的数据,如来料数据、用水、用电数据等,数据分析就能产生更大威力。此外,土壤方面相关数据也比较少。

中国环境报:很多企业认为,来料数据、用水、用电数据涉及商业机密,如何平衡数据公开与保护商业机密的关系?

林宣雄:生产过程的数据对于民营企业来说,并不涉及商业机密。只是少数的军工企业需要保密。在做相关分析时,并不需要工艺、发明等涉及商业秘密的相关数据。所谓商业机密,只是企业不愿意公开数据的借口。

中国环境报:很多企业认为,来料数据、用水、用电数据涉及商业机密,如何平衡数据公开与保护商业机密的关系?

林宣雄:生产过程的数据对于民营企业来说,并不涉及商业机密。只是少数的军工企业需要保密。在做相关分析时,并不需要工艺、发明等涉及商业秘密的相关数据。所谓商业机密,只是企业不愿意公开数据的借口。

否则,数据分析只是一种形式,并不能真正为决策提供科学依据。

李莹

废运输的个人或单位进行登记,并随时监控其运输情况。此外,可对司法机关办理危废案件的处理情况、衔接机制的运行情况进行自查。

加大设备投入,提高危废检测水平。设备先进与否关系到对危废的检测水平,还牵涉到对危废污染的治理情况。只有明确污染物性质,才能有针对性地消除影响。因此,要尽量配置先进的危废检测设备,使取证方法更科学。

定期进行执法规范化培训,增强及时固定收集证据的意识。在执法过程中要注意收集物证、书证、视听资料、危废电子检测数据等证据材料,在刑事诉讼中可以作为证据使用。

作者单位:山东省茌平县环保局

探索与思考

监测工作如何华丽转身?

◆罗岳平 田耘 郭卉

“十一五”以来,我国各级环境监测站历经了近十年的黄金发展期,装备能力、人员力量、技术水平等取得长足进步,一些长期制约环境监测工作深入发展的瓶颈问题逐步突破。陈吉宁部长在环境保护部组织的“环评和监测工作创新”大讨论中希望改革激发创新活力,为环境监测新发展凝聚力量。结合工作实际,笔者认为,各级环境监测站应主动适应环保工作新形势,在加强环境质量管理、监督、巡视性监测和监测质量监督等方面有所作为。

加强对环境质量的监测与评价

环境质量逐步改善是环境保护工作的根本使命。新环保法规定各级人民政府对本行政区域的环境质量负责。但在目前,普遍采用环境质量各级人民政府自评、自评、自评的工作模式,还是一种绩效的自我评估和审查,不仅在科学性上不符合接受上级或第三方机构独立评估的原则,而且真实性也难以完全保证。

为此,对环境质量要增加和建立上一级环境监测机构不定期核查制度。也就是说,本级人民政府下属环保部门直属的环境监测机构要全面开展辖区内的环境质量监测,并将监测结果按要求报送上一级环境监测机构。上一级环境监测机构根据既定的现场核查计划,以及日常审核上报的监测数据,不定期赴现场开展核查。

这种模式,将环境质量自测与上级监督性质的现场核查结合起来,只要严肃使用现场核查结果,就能使环境质量绩效考核体系正常运转起来。当务之急,各级环境监测机构应牢牢把握环境质量监测这条主线不动摇,守土有责,科学布设各种环境监测点的监测点位,心无旁骛地做好本辖区环境质量监测,主动接受上级环境监测机构的日常监督和不定期的检查,排除来自本级领导的行政干预,如实反映环境质量现状。

加强污染源监督性和巡视性监测

新环保法明确了企事业单位达标排放的主体责任和环保部门的监督责任。企事业单位对自身的排污行为负责,并开展自行监测;环保部门

◆张旭

近年来,甘肃省积极推进全省环境监管网格化工作,目前已取得显著成效。以兰州市创新网格化管理模式为例,据环境保护部统计数据,兰州市已成为全国重点监测城市中综合污染指数下降最快的城市,这充分证明了网格化监管的优势。

甘肃省之所以能在环境监管网格化管理上取得成效,主要做法如下:

顶层设计、配套细则,有序推进工作。甘肃省在全国第一个以政府令的形式颁布施行了《甘肃省环境保护监督管理责任规定》(甘肃省人民政府令第101号),为落实各级政府、各部门的监管责任和各企事业单位的环保主体责任提供了制度保障。先后印发了《甘肃省环境保护厅关于推行环境监察网格化管理等四项环境监管模式的通知》和《甘肃省环境保护厅关于进一步完善环境监管网格化管理工作的通知》等文件,对网格化管理的思想、基本原则、工作内容、工作目标进行了详细安排。

分级负责、扎实推进,做好管理工作。一是做实一级网格。指导全省开展环境监管网格化工作,制定全省环境监管网格化管理工作规划、制度,并监督实施。二是做强二级网格。市(州)级环保部门在网格化管理工作中起着承上启下的关键作用。要组织开展本辖区环境监管网格化管理工作,对国家和省级确定的重点污染源和建设项目进行监管,对辖区县(区)级环境监管网格化工作进行指导和检查,对网格划分范围内的环境违法企业和案件进行调查和后督察。三是做细三级网格。县(区)级环保部门作为网格化管理的根本落脚点,要切实履行环境监管网格化责任,按照辖区内重点污染源、环境风险源、建设项目、环境敏感区的分布状况和相关的本

信息,合理配置环境监管人员,具体实施责任区域内的环境监管工作。

落实责任,基本实现环境监管无盲区。甘肃各级环保部门通过实施“分块管理、网格划分、责任到人”的网格化监管模式,逐步形成了一个覆盖全省,以各县区基层执法力量为主体的环境监管网络,基本实现了“三清三到位”的总体目标(区域清、职责清、底数清和监管到位、服务到位、互通到位)。

重心下移,稳步提升执法效能。环境监管网格化工作实施以来,环境监察执法重心全面下移,基层环保部门熟悉情况的优势得到了充分发挥。目前,基层网络的监管人员基本做到“三能、三清”,即能初步分清自己分管网格内的国控、省控、环境风险源企业情况;能基本摸清各类排污企业污染物的排放数量及去向;能点清环境敏感区域、主要环境风险隐患点及群众关注的重点环境问题,网格员能第一时间对管辖网格内的环境违法行为进行举报处理,有效维护人民群众的环境权益。
作者单位:甘肃省环境监察局

监管网格化怎么做实?

甘肃省平凉市崆峒区、陇南成县等地还成立了乡镇环保所,实现了覆盖省一市一县一乡(镇)的四级环境监管。

举一反三、创新方式,推行“6+1”监管模式。在实施网格化管理的同时,为进一步落实环境监管责任,实现执法工作全覆盖、无盲区,甘肃省积极探索建立了环境监管分类化、痕迹化、流程化、模板化、智能化和编制执法计划等强化环境执法的“6+1”监管新模式。甘肃省环保厅每年年初编制全省环境执法计划,分别确定省级重点监管企业和市、州和县、区重点监管企业名单,每季度对全省环境执法检查情况进行调度,督促各市州按计划开展环境执法工作,年底进行考核,做到“事前有计划、事中有方案、事后有考核”。对1619多家企业实施分类环境监管,下发污水处理厂、燃煤电厂等5个主要行业模板全部运用于现场执法过程。