

基于兽药电子追溯的兽药大数据平台建设研究

郝毫刚,高录军,张积慧,刘玲,唐军,李晓平,赵丽丹*

(中国兽医药品监察所,北京 100081)

[收稿日期] 2017-01-05 [文献标识码] A [文章编号] 1002-1280 (2017) 03-0004-07 [中图分类号] S851.66

[摘要] 从国家实施兽药产品电子追溯着手,分析了兽药大数据平台建设的必要性,介绍了大数据的关键技术和分析应用,重点阐述了兽药大数据平台的建设原则和技术架构,对兽药大数据平台建设的因素进行了思考,以期为兽药行业发展提供数据支撑。

[关键词] 大数据;兽药电子追溯;数据分析

Research on Construction of Veterinary Drugs Big Data Platform Based on Veterinary Drugs Electronic Trace

HAO Hao-gang, GAO Lu-jun, ZHANG Ji-hui, LIU Ling, TANG Jun, LI Xiao-ping, ZHAO Li-dan*

(China Institute of Veterinary Drug Control, Beijing 100081, China)

Abstract: In this paper the necessity of construction on veterinary drugs big data platform was analyzed, the key technology and analytical applications of big data were introduced from the current situation of state implementing veterinary drugs product electronic trace. Meanwhile, emphasis was on the specification of the construction principle and technology framework of veterinary drugs big data platform. In addition, reflection was made on the key factors in the construction of veterinary drugs big data platform in order to provide data support for the development of veterinary drugs industry.

Key words: big data; veterinary drugs electronic trace; data analysis

兽药作为预防、治疗、诊断畜禽等动物疾病的物质,是一种特殊商品,既要保障动物疾病得到有效的治疗,又要保障动物和人的安全。为强化兽药安全监管,保障动物产品质量安全,对兽药产品实施追溯管理,国家实施了兽药产品电子追溯码(二维码)标识制度,为政府、企业和广大消费者提供相关的公共信息服务。两年的实践证明,国家兽药产品追溯系统(以下简称追溯系统)的实施,规范了生

产企业质量管理行为,提高了监管部门工作效率,为经营者、养殖者辨别非法产品提供了快捷可靠的手段,有效遏制了造假售假行为。随着信息技术发展,全国乃至地方兽药管理和研究机构均不同程度地吸纳了国内外先进的信息技术,建立了不同的系统平台,兽药行业的信息化程度得到很大提高。

“大数据”被看作是继云计算、物联网之后,人类历史上又一次深刻的信息技术革命,是生活、工

作者简介: 郝毫刚,硕士,工程师,从事兽药信息化建设和大数据研究工作。

通讯作者: 赵丽丹。E-mail: zhaolidan@ivdc.org.cn

作与思维的大变革^[1]。2015年,《国务院关于印发促进大数据发展行动纲要的通知》^[2]正式发布,建立“用数据说话、用数据决策、用数据管理、用数据创新”的管理机制,实现基于数据的科学决策,将推动政府管理理念和社会治理模式进步。兽药追溯系统建设及各级管理部门和企业信息管理系统建设,为兽药大数据平台的建设积累了宝贵的经验和基础数据。由此,认识大数据、理清兽药大数据平台的建设思路,及早进行兽药大数据平台建设并应用,是兽药企业、监管部门、检验机构的当务之急。

1 兽药大数据平台建设的必要性

1.1 国家兽药电子追溯历程 2015年1月,农业部颁布中华人民共和国农业部公告第2210号^[3],开始利用国家兽药产品追溯系统实施兽药产品电子追溯码(二维码)标识制度。国家兽药产品追溯系统是为行业提供服务的公益性系统,主要在兽药产品包装上印制电子追溯码(二维码),生产者、经营者、监管者和消费者可扫描电子追溯码(二维码),查询兽药产品信息,实现兽药产品可追溯。目前,追溯系统主要包含:国家兽药产品追溯信息系统、国家兽药基础信息查询系统、国家兽药进销存软件、数据采集设备(内含数据采集系统)、国家兽药查询手机APP等。

2016年4月,农业部办公厅发布《关于在全国范围内组织开展兽药经营追溯试点工作的通知》^[4],在全国范围内组织开展兽药经营追溯试点工作。2016年6月,农业部发布《关于加快推进农产品质量安全追溯体系建设的意见》^[5],应用现代信息技术加快推进全国农产品质量安全追溯体系建设。2016年10月,农业部办公厅印发《农业农村大数据试点方案》的通知^[6],推动农业农村大数据应用,在农产品质量安全监管领域研发支撑智慧农业建设的大数据应用系统,争取2019年底实现数据共享取得突破和大数据应用取得突破。

1.2 我国兽药行业信息共享现状 近年来,我国兽药信息体系发展迅速,已经积累了可观的数据量。兽药追溯系统体系框架初步建立。但兽药生

产、经营和使用主体呈现“多、小、散”特点,兽药基础信息仍处于多单位和多部门分散管理状态,政企间信息资料得不到充分利用,无法实现信息资源共享。

1.2.1 国家兽药信息数据库不断完善 中国兽药信息网自2001年开通以来,宣传兽药产业政策方针、法律法规,普及推广科学知识,成为兽药行业信息化建设基础。依托中国兽药信息网建设的国家兽药基础信息查询系统2011年11月正式上线运行,目前已有12个数据库,28.5万余条数据,包括兽药生产企业(许可证和GMP证书)、兽药产品批准文号、兽用生物制品批签发(国产、进口)、兽药标签说明书等信息。

1.2.2 兽药电子追溯不断汇聚兽药产业数据资产 国家兽药电子追溯将连接生产、经营、使用、监管四个环节,使我国兽药产品首次实现流向可追溯、来源可查询。目前,已有5000余家兽药生产、经营企业和监管单位在使用追溯系统,兽药生产、经营、使用、监管等多个环节中所产生的海量数据,逐渐形成兽药产业数据资产。截止2016年12月底,追溯系统中的数据总量已近百亿条。

1.2.3 兽药不同环节信息化水平不均衡 兽药使用环节的信息化应用水平远落后于生产、经营环节。长期以来,兽药行业信息化基本呈金字塔模式:以部分大型兽药生产企业和监管单位为金字塔尖,大部分生产企业和部分兽药经营企业组成金字塔身,多数经营企业和养殖场构成金字塔底。而兽药使用信息恰恰是兽药大数据的重要部分。

1.2.4 重系统建设,轻数据分析利用 近几年来,兽药行业对硬件基础设施及应用系统开发的重视程度普遍较高,建成了各种管理系统,投入也越来越大,甚至用到了小型机和大型网络设备,认为信息化建设与系统搭建实施可以直接划等号,忽视了信息资源的开发和利用,出现了“重系统轻数据”的问题,对数据的分析利用、深度挖掘重视不足,难以满足政府决策需求,信息化总体效益和效果不尽如人意。

1.2.5 兽药信息共享不足,存在“单兵作战”和“信

息孤岛”现象 我国在兽药信息化建设尤其是兽药监管和企业管理方面发展迅速,建立了国家兽药产品追溯系统、企业防伪防串货系统、实验室信息管理系统(LIMS)、广西动物卫生监督信息管理平台、河南省饲料兽药追溯系统等,但各省、市兽药系统多数独立运行,尚未实现业务数据的纵向交换和横向共享。因数据采集的渠道不同、标准各异、重复冗余,系统之间难以兼容,信息资源难以共享,形成了“信息孤岛”,由于缺乏沟通,没有充分挖掘利用数据的潜在价值。

1.2.6 数据质量较差 不同的系统因多种原因数据采集和录入准确性不高,数据标准出现前后不一致的现象。不同结构与属性的数据来源分属于很多部门,甚至个人所有,数据采集质量较差。

总的来看,目前我国兽药行业整体尚未能实现信息共享。随着信息化水平不断提高,兽药行业单位和企业积累了大量基础数据和业务数据,这些“原始数据”是信息化建设和管理的基础。随着大数据等技术的成熟,信息化建设的重心将逐步从信息技术向数据技术转化,未来信息化建设的重心将是如何对系统内的数据进行深入的挖掘和分析。

1.3 兽药大数据平台建设的必要性 有效、适时的大数据管理能够从我国兽药生产、流通、使用和监管领域的数据中分析出很多深入的和有价值的信息,为建立完善的兽药追溯体系和风险预测机制提供契机和依据,也有助于解决兽药安全信息不对称的问题。

对于监管部门来说,建设兽药大数据平台有助于探索建立新的兽药安全风险防控机制。兽药大数据平台建成之后,利用大数据,可以分析不同地区的用药情况,找出控制用药的重点地区;分析不同畜禽品种的用药情况,找出重点的控制用药的畜禽品种;分析不同的季节与养殖规模和畜禽用药的关系,给出有效应对措施;分析抗菌药的使用特点,采取针对性的措施;分析假劣兽药的生产、流通和使用规律,提高假劣兽药的打击力度和兽药残留控制的效率;实现大视角、全方位、多角度的观察和分析,找出重点、找准关键点,建立新的兽药安全风险

防控和风险预测机制。

对于生产、经营企业来说,建设兽药大数据平台有助于提高企业管理水平。国家兽药电子追溯码的推行加快了兽药企业信息化建设,基础设施及应用系统开发的重视程度逐渐增高,但却出现了“重系统轻数据”问题,对数据分析利用、深度挖掘重视不足。行业普遍认为,信息化建设等同于系统开发应用,忽视信息资源的开发和利用。目前兽药企业信息系统并未按照统一标准开发,多数只能满足业务流程的电子化和功能性需求,无法有效提取数据中含有的巨大价值。

对于终端兽药用户来说,建设兽药大数据平台可提供广泛的查询及用药服务。兽药作为一种特殊商品,兽药综合信息查询可以提高用药安全性。目前,国内的兽药综合信息查询平台仅有国家兽药基础信息查询系统,功能仅限于基础兽药相关信息查询,尚不能满足终端兽药用户用药服务。兽药大数据平台未来可通过定制服务为包括政府机构、监管单位、企业、普通大众等不同层级提供服务,从基础应用到深度应用,从查询服务到数据服务,实现兽药基础信息查询、流通信息追溯、监管信息共享,整合兽药生产环节、经营环节、监管环节、养殖场、终端用户的数据信息,逐步建立兽药产业数据资产。

2 大数据概述

2.1 大数据的定义 目前对大数据尚未有统一公认的定义。维基百科给出的大数据定义为:数据量规模巨大到无法通过人工,在合理时间内达到截取、管理、处理并整理成为人类所能解读的信息;IBM 则用4个特征相结合来定义大数据:数量(volume)、种类(variety)、速度(velocity)和真实(veracity)^[7]。目前人们对大数据具有数量巨大(volume)、种类多样(variety)和处理时效紧(velocity)的特征存在共识。大数据可以在大规模数据的基础上,挖掘分析获得新的认知、创造新的价值并以此改变市场、组织机构以及政府与公民的关系^[8]。同时,也是政府做出科学、准确决策的“数据技术支撑”。通俗的讲,大数据指的是经过长期收集所得的资料,因规模巨

大而复杂,无法通过目前主流的数据库管理工具和数据处理软件来处理^[9]。大数据与过去海量数据的区别不仅仅在于数据量更大,更在于数据类型繁多^[10],不单纯包括“数字”,还包括文本、图片、音频、视频等多种格式,涵盖内容十分丰富^[11]。

根据来源的不同,大数据大致可分为如下三类^[12]:(1)来自于人。人们在互联网活动以及使用移动互联网过程中所产生的各类数据,包括文字、图片、视频等信息。(2)来自于机。各类计算机信息系统产生的数据,以文件、数据库、多媒体等形式存在,也包括审计、日志等自动生成的信息。(3)来自于物。各类数字设备所采集的数据。如摄像头产生的数字信号、医疗物联网中产生的人的各项特征值、天文望远镜所产生的大量数据等。

兽药大数据属于生物大数据范畴。在我国目前兽药监管体系下,兽药大数据有动态数据和静态数据,主要含有:兽药监管机构历年兽药质量抽查检验数据;历年兽药基础数据;兽药注册审批资料;企业 GMP 及过程控制信息;兽药电子追溯信息;来源于公共数据库和文献的与药物分析、安全性评价等有关的数据。这些信息既包括数字型数据,如兽药检验结果;也有文本类信息,如抽验报告、注册资料等。上述信息或数据具有明显实效性,随着时间的推移,使用价值递减。因此,如何在合理的时间内对上述信息的提取、处理并整理成可服务于兽药行业的有用信息则成为关键。

2.2 大数据的关键技术 大数据处理涉及数据的采集、管理、分析与展示等。相关技术示意图如图 1 所示。

2.2.1 数据采集与预处理 大数据处理的第一步是从数据源采集数据并进行预处理操作,为后继流程提供统一的高质量的数据集。数据采集与预处理主要含有数据清洗、数据集成、数据存储三个方面。

在大数据环境下,大数据的维度包含了数量、多样性、速度、精确性等,在如此大维度中不可避免地存在冗余、错误、粗糙的数据,如何将这良莠不齐的非清洁数据有效转化成高质量的干净数据,需

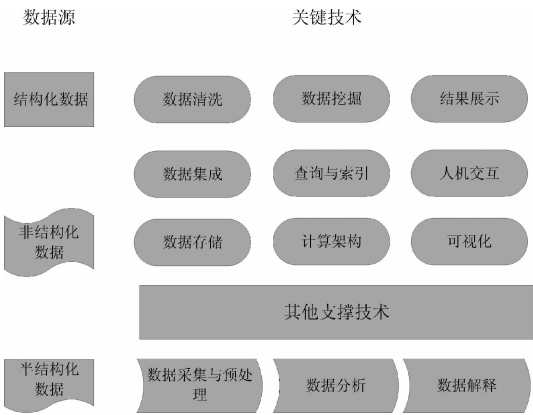


图 1 大数据相关技术示意图

要数据清洗。数据的质量体现出数据的价值,数据清洗是数据质量研究的起点,其本质是发现并解决问题的过程,最终目的是提高数据的质量^[13]。数据清洗主要是将不正确、不完整和不一致的数据进行过滤和规范,从而提高决策系统的准确性;并根据待分析数据的特点对规范后的数据进行归约,从而提高决策系统的效率。

由于大数据的来源不一,可能存在不同模式的描述,甚至存在矛盾。因此,在数据集成过程中对数据进行清洗,以消除相似、重复或不一致的数据是非常必要的。

数据存储与大数据应用密切相关。某些实时性要求较高的应用,如状态监控,更适合采用流处理模式,直接在清洗和集成后的数据源上进行分析。而大多数其它应用则需要存储,以支持后继更深度的数据分析流程。为了提高数据吞吐量,降低存储成本,通常采用分布式架构来存储大数据。

2.2.2 数据分析 数据分析是大数据应用的核心流程。根据不同层次大致可分为三类:计算架构、查询与索引以及数据挖掘。在计算架构方面,MapReduce是当前广泛采用的大数据集计算模型和框架。在查询与索引方面,由于大数据中包含了大量的非结构化或半结构化数据,传统关系型数据库的查询和索引技术受到限制,而 NoSQL 类数据库技术得到更多关注。由于大数据环境下数据呈现多样化特点,所以对数据进行挖掘时,就较难统一术语进而挖掘信息。

2.2.3 数据解释 数据解释旨在更好地支持用户对数据分析结果的使用,涉及的主要技术为可视化、人机交互和结果展示。目前已经有了一些针对大规模数据的可视化研究,通过数据投影、维度降解或显示墙等方法来解决大规模数据的显示问题。由于人类的视觉敏感度限制了更大屏幕显示的有效性,以人为中心的人机交互设计也将是解决大数据分析结果展示的一种重要技术。

2.2.4 其他支撑技术 虽然大数据应用强调以数据为中心,将计算推送到数据上执行,但是在整个处理过程中,数据的传输仍然是必不可少的。由于虚拟集群具有成本低、搭建灵活、便于管理等优点,人们在大数据分析时可以选择更加方便的虚拟集群来完成各项处理任务。

2.3 大数据的分析应用 目前,大数据的分析应用主要用于科学、医药、商业、金融、互联网等各个领域,用途差异巨大,但是其分析应用主要有以下三个方面^[14]:(1)推测趋势:由于大数据中包含大量原始、真实的信息,大数据分析能更准确把握事物背后的规律,有效消除个体特征差异,对自然或社会现象进行预测,如可以根据 Twitter 信息预测股票行情。(2)分析个体特征:个体特征在满足群体特征时,具有鲜明的个性化特征。大数据中积累了长时间、多维度的数据,可以分析用户行为规律,更准确描述个体特征,为用户提供个性化产品和服务。如 Google 通过其大数据产品对用户的习惯和爱好进行分析,帮助广告商评估广告活动效率,预估在未来可能存在高达数千亿美元的市场规模。(3)分辨真相:大数据来源广泛而多样,在一定程度上它可以帮助实现信息的去伪存真。如 Yahoo 利用大数据分析技术来过滤垃圾邮件。

3 兽药大数据平台建设思路

3.1 建设原则 按照“整体规划、统一标准、分步实施、逐步完善”原则,从兽药行业需要出发,统筹整体规划、强化顶层设计,统一数据定义、制定数据标准,逐步健全平台、分步实施应用,建设数据中心、完善数据模块,实现对兽药数据的深度利用和信息共享,辅助行政管理部门进行合理、有效地决

策。为了保证兽药大数据平台发挥其应有的作用和功能,建设兽药大数据平台时应遵循以下原则:

经济适用原则:按照经济实用、需求主导的要求,确定平台建设的规模和软硬件配置,要求系统具有很强的环境适应性、较好的开放性;功能模块间轻耦合,互相通信,交换数据。**统一建设原则:**平台建设要统一数据标准(数据结构标准),同时兼顾特殊性,形成纵向和横向的一体化数据中心。**安全可靠原则:**平台建设要符合信息安全的一般特征,保证存储安全、逻辑安全、访问安全和传输安全,结合平台自身安全体系,充分确保数据、网络安全。**可扩展性原则:**平台建设应具有较大的可升级空间,以保证业务流程的连续性和系统信息的一致性,立足现有数据资源整合建设,同时为未来不同数据融合利用留有余地。**分步实施原则:**信息化平台建设不会一蹴而就,是一个长期、持续的过程,既要考虑短期内成效,又要保护持续发展。因此,平台架构设计要先进合理,符合主流趋势,才能确保之后运行稳定、安全可靠、后续建设。

3.2 平台架构 兽药大数据的开发建设采用智能化数据处理工具,对国家兽药产品追溯系统、国家兽药产品基础信息查询系统、LIMS 等系统产生的海量数据按照数据规范进行清洗、整合后,根据业务需求对数据进行二次计算,并以可视化交互方式进行展示,通过简单、灵活、可扩展的操作方式获取分析统计结果,为兽药监督管理、企业运行等提供服务。兽药大数据平台架构如图 2 所示。

3.3 平台建设关键因素 兽药电子追溯码标识制度是兽药行业发展的需要,兽药电子追溯体系收集的信息资源对于前期大数据平台建设十分重要。兽药大数据平台建设关键因素如下:

3.3.1 建立制度体系,加强数据共享,打破“信息孤岛模式”,形成数据资源共享的体制与机制 在这方面,兽医行政管理机构可充分发挥统筹和引领作用,基于对核心单位和企业调研,制定规划目标,建立农业部门内部不同司局、单位、部门之间的沟通协作机制,梳理各部门之间的数据共享交换标准,同时部级、省级、市级、县级各主管部门之间建

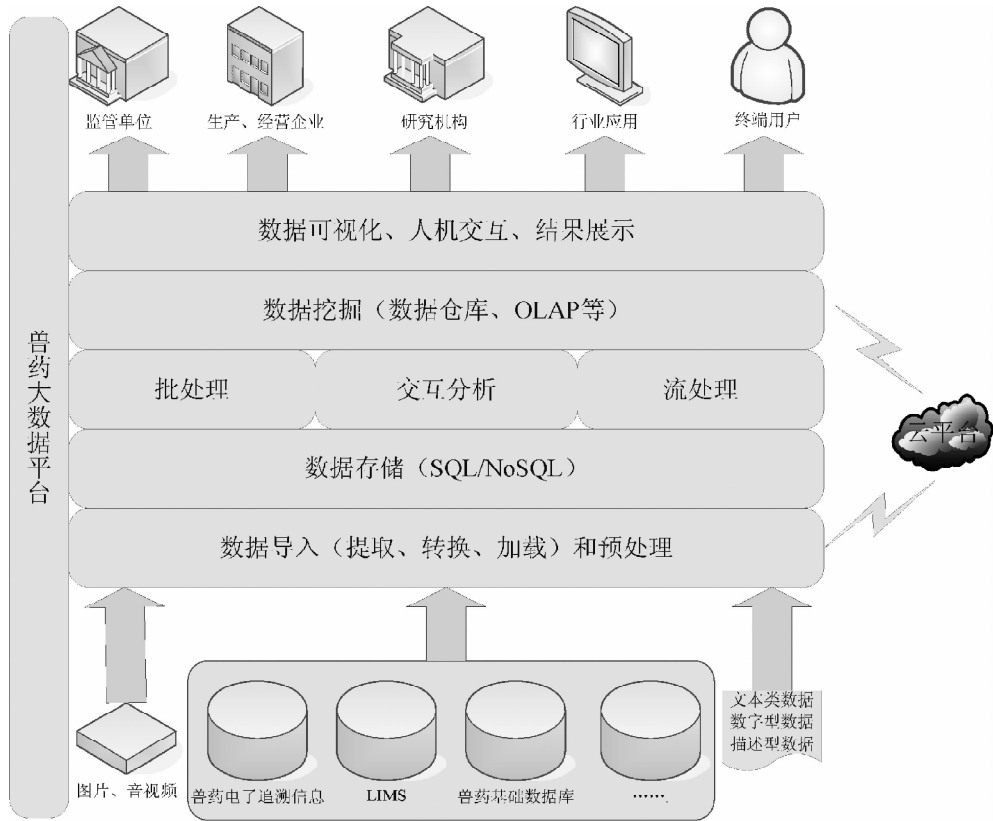


图2 兽药大数据平台架构

立数据共享汇聚渠道,建立我国兽药大数据中心,形成我国兽药数据资产。

3.3.2 完善数据标准,保证业务数据准确性、唯一性,确保原始业务数据质量 对已有数据需彻底清洗和规范,使之能符合数据分析的要求,此过程需要和业务部门商讨,总结出数据的规范和标准,并通过一些专业的数据处理工具对原始数据进行自动清洗。还需引入业务专家对清洗后的数据进行人工审核,继续找出数据中存在的问题,并给予修正,直到满足后期数据分析的要求为止。

3.3.3 整合数据资源,强化顶层设计,建设数据中心 在兽药大数据平台建设过程中,应统筹规划数据中心与之前系统,加强统一指导,分级建设,按照统一的信息化标准规范,加快整合各类业务应用系统,逐步建设成国家、省(市)两级的兽药大数据中心,实现资源整合和数据共享,为科学决策、智能监管以及行业服务提供数据支撑。

3.3.4 加强人才培养,建设复合型人才队伍 大数据时代的到来对兽药行业人才的技能和素质提出了新的要求,急需擅于处理庞大规模和复杂结构数据的专业人才,包括数据分析师、数据可视化人员和数据管理人员等,以适应深度分析数据的需要。除要具备较强的数理统计知识和技能,还要拥有一定管理、兽医兽药、信息技术等专业知识,才能具备跨学科、跨领域的数据综合分析思想,驾驭大数据时代的数据挖掘与分析应用。

4 结 语

大数据是一种从海量数据中快速获取信息的能力,为提高我国兽药行业发展水平提供了新的思路和切入点。兽药大数据平台的建设是一项技术、业务、管理紧密结合的复杂工程,目前还处在探索和研究阶段。在平台建设的设计过程中,应充分考虑平台的标准化和开放性,以及完善的配套法规制度,如平台运行保障制度、数据质量保障制度、数据

安全保障制度等。

在兽药电子追溯制度实施和大数据时代的背景下,期待进行大数据和兽药行业的大融合,进一步推动兽药行业的信息化建设,为提升兽药监管效能,提供科学、有效的数据支撑。

参考文献:

- [1] 维克托·迈尔-舍恩伯格,肯尼思·库克耶. 大数据时代[M]. 杭州:浙江人民出版社,2013.
- [2] 国务院. 国发[2015]50号国务院关于印发促进大数据发展行动纲要的通知[S]. 2015.
- [3] 农业部. 中华人民共和国农业部公告第2210号[EB/OL]. (2015-1-27). http://www.moa.gov.cn/govpublic/SYJ/201501/t20150127_4352158.htm
- [4] 农业部办公厅.《关于在全国范围内组织开展兽药经营追溯试点工作的通知》农办医[2016]22号[EB/OL]. (2016-4-15). http://www.moa.gov.cn/govpublic/SYJ/201604/t20160419_5100452.htm
- [5] 农业部.《关于加快推进农产品质量安全追溯体系建设的意见》农质发[2016]8号[EB/OL]. (2016-6-23). http://www.moa.gov.cn/govpublic/ncpzlaq/201606/t20160623_5184476.htm
- [6] 农业部办公厅.《关于印发<农业农村大数据试点方案>的通知》农办市[2016]30号[EB/OL]. (2016-10-14). http://www.moa.gov.cn/zwllm/tzgg/tfw/201610/t20161018_5308511.htm
- [7] 陈刚. 科学研究大数据挑战[J]. 科学通报, 2015, 60(5/6):439-444.
- [8] 孟宪民,李磊. 大数据应用及相关财税思考[J]. 地方财政研究, 2013, (10):70-72.
- [9] Wikipedia. Big data[EB/OL]. http://en.wikipedia.org/wiki/Big_data,2014-05-16.
- [10] Grobelnik M. Big data tutorial [EB/OL]. (2012-10-22). <http://viedeolctures.net/eswc2012-grobelnik-big-data/>
- [11] 陈冠如. 善用“大数据”成就智慧监管[N]. 中国医药报, 2014-04-08(8).
- [12] 李国杰,程学旗. 大数据研究:未来科技及经济社会发展的重大战略领域[J]. 中国科学院院刊, 2012, (6):647-657.
- [13] 蒋勋,刘喜文. 大数据环境下面向知识服务的数据清洗研究[J]. 图书与情报, 2013, (5):16-21.
- [14] 冯登国,张敏,李昊. 大数据安全与隐私保护[J]. 计算机学报, 2014, (1):246-258.

(编辑:李文平)