

大数据标准体系

张群, 吴东亚, 赵菁华

中国电子技术标准化研究院, 北京 100007

摘要

随着大数据的发展, 标准化的内容越来越广, 标准化的对象也越来越复杂。系统地分析了国内外大数据标准化工作现状, 结合《促进大数据发展行动纲要》、中华人民共和国国民经济和社会发展第十三个五年规划纲要等国家战略以及大数据产业发展对标准化工作的需求, 分析我国大数据标准化工作面临的问题, 描述了大数据参考架构, 提出大数据标准体系框架, 给出了未来的工作建议。

关键词

大数据; 标准化; 参考架构; 标准体系

中图分类号: TP399

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2017037

Big data standards system

ZHANG Qun, WU Dongya, ZHAO Jinghua

China Electronics Standardization Institute, Beijing 100007, China

Abstract

With the development of big data, standardization involves more and more content, and the objects are becoming more and more complex. The status of national and international big data standardization work was systematically analyzed. In combination with the national strategy of "Platform for the Development of Big Data" and the Thirteenth Five-Year Plan of National Economic and Social Development of the People's Republic of China, as well as the demand for big data standardization, the problems of national big data standardization were pointed out, big data reference architecture and standards system was proposed, and suggestions for future work were given.

Key words

big data, standardization, reference architecture, standards system

1 引言

大数据的初步应用已经开始对生产、流通、分配与消费模式产生重要影响,但目前也存在核心技术差距较大、政府数据开放共享不足、应用水平低、产业基础薄弱、法规标准滞后、信息安全问题日益突出等问题。由于大数据领域相关的标准缺失,导致大数据应用建设没有标准可依据,形成数据信息孤岛,影响数据的复用性和互联互通性,阻碍产业化发展。

2015年9月,国务院印发《促进大数据发展行动纲要》(以下简称《行动纲要》),明确提出“建立标准规范体系。推进大数据产业标准体系建设,加快建立政府部门、事业单位等公共机构的数据标准和统计标准体系,推进数据采集、政府数据开放、指标口径、分类目录、交换接口、访问接口、数据质量、数据交易、技术产品、安全保密等关键共性标准的制定和实施。加快建立大数据市场交易标准体系。开展标准验证和应用试点示范,建立标准符合性评估体系,充分发挥标准在培育服务市场、提升服务能力、支撑行业管理等方面的作用。积极参与相关国际标准制定工作”^[1]。2016年3月,中华人民共和国国民经济和社会发展第十三个五年规划纲要,简称“十三五”规划,指出要“完善大数据产业公共服务支撑体系和生态体系,加强标准体系和质量技术基础设施建设”^[2]。

本文通过总结国内外大数据标准化工作现状,紧紧围绕《行动纲要》和“十三五”规划的要求,结合大数据产业发展对标准化工作的需求,充分分析我国大数据标准化工作面临的问题,提出未来工作的几点建议。

2 大数据参考架构

2015年,美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)发布了《大数据互操作框架第6卷:参考架构(big data interoperability framework volume 6 reference architecture)》,描述了大数据参考架构的总体框架^[3]。2016年,全国信息技术标准化技术委员会大数据标准工作组结合NIST的《大数据互操作框架第6卷:参考架构》,提出了我国大数据参考架构,如图1所示。该参考架构方便大家对大数据复杂性操作的认识,中立于供应商,并在技术和基础设施方面独立,为大数据标准化提供基本参考点,为大数据系统的基本概念和原理提供了一个总体框架,为各种利益相关者提供一种交流大数据技术的通用语言,鼓励大数据实践者遵守通用标准、规范和模式。

大数据参考架构围绕代表大数据价值链的信息价值链(水平轴)和IT价值链(垂直轴)两个维度组织展开。信息价值链表示大数据的应用理论作为一种数据科学方法,从数据到知识的处理过程中所实现的信息价值,其核心价值通过数据收集、预处理、分析、可视化和访问等活动实现。IT价值链表示大数据作为一种新兴的数据应用范式为IT技术产生的新需求带来的价值,其核心价值通过为大数据应用提供存储和运行大数据的网络、基础设施、平台、应用工具以及其他IT服务实现。

大数据参考架构提供了一个构件层级分类体系,用于描述架构中的逻辑构件以及定义逻辑构件的分类。逻辑构件被划分为3个层级,从高到低依次为角色、活动和组件。最顶层级的逻辑构件代表大数据

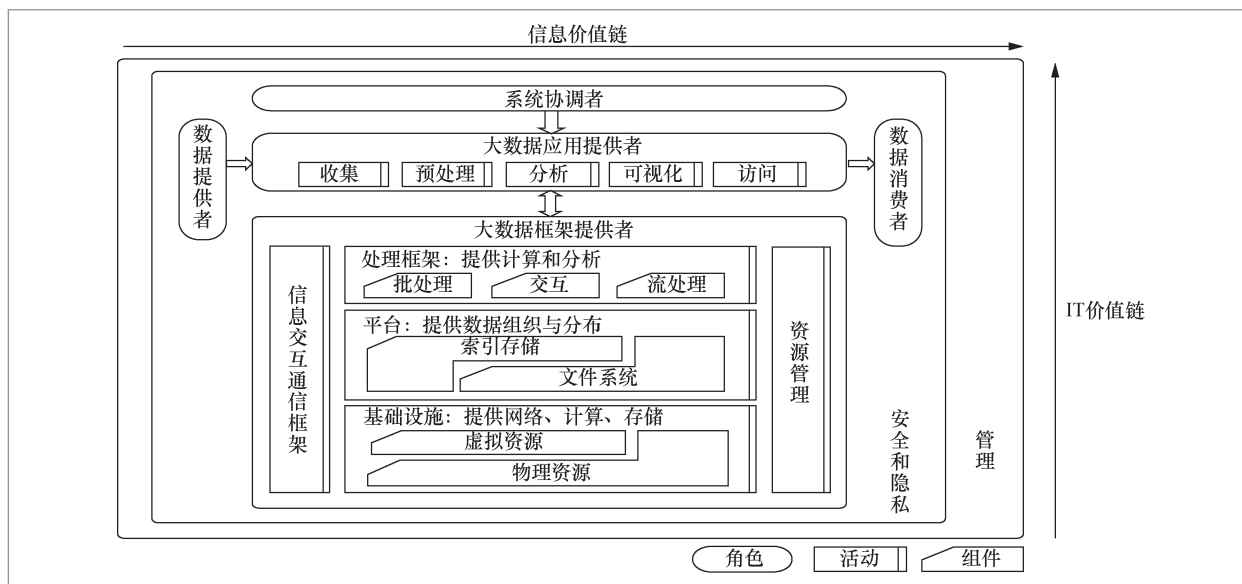


图1 大数据参考架构

参考架构中存在的5个角色，包括系统协调者、数据提供者、大数据应用提供者、大数据框架提供者、数据消费者。另外两个非常重要的逻辑构件是安全和隐私、管理，它们为大数据的5个角色提供服务和功能。第二层级的逻辑构件是每个角色执行的活动。第三层级的逻辑构件是执行每个活动需要的功能组件。

该架构可以用于表示由多个大数据系统组成的堆叠式或链式系统，其中一个系统的数据消费者可以作为后面一个系统的数据提供者。该架构支持各种商业环境，包括紧密集成的企业系统和松散耦合的垂直行业，有助于理解大数据系统如何补充并有别于已有的分析、商业智能、数据库等传统的数据应用系统。

3 国内外大数据标准化工作部署

当前，许多国家的政府和国际组织纷纷将开发利用大数据作为夺取新一轮竞争制高点的重要抓手，实施大数据战略。随着大数据技术的发展与应用，大数据标准

研制已成为国际各标准化组织共同关注的热点，然而尚处于初期发展阶段。

3.1 国际标准化现状

国际大数据标准化工作主要集中在ISO/IEC JTC1/WG9大数据工作组（以下简称WG9）。除此之外，ISO/IEC JTC1/SC32数据管理和交换分技术委员会（以下简称SC32）和国际电信联盟电信标准分局（ITU-T）也在从事大数据标准化相关的工作^[4]。

(1) WG9大数据标准化情况

WG9于2014年11月正式成立。工作重点包括：聚焦和支持JTC1的大数据标准计划；编制大数据基础标准，以指导JTC1中其他大数据标准的编制；编制建立在基础标准上的其他大数据标准（当JTC1下属相关组不存在或不能编制这些标准时）；识别大数据标准化中的差距；建立和维护与JTC1中那些将来可能提出大数据相关工作的所有相关实体以及任何下属组的联络；识别那些正在编制有关大数据的标准和相关资料的JTC1（和其他组织）实体，并在适当时候调查有关大数据的正在进行

中和潜在的新工作；与JTC1之外的相关社区共同提升意识，并鼓励其参与JTC1的大数据标准化工作，根据需要建立联络。

目前，WG9的国家成员有22个，各国代表超过190名。正在研制《信息技术大数据概述和术语(information technology-big data-overview and vocabulary)》《信息技术大数据参考架构第1部分：框架和应用(information technology-big data reference architecture-part 1: framework and application process)》《信息技术大数据参考架构第2部分：用例和需求(information technology-big data reference architecture-part 2: use cases and derived requirements)》《信息技术大数据参考架构第3部分：参考架构(information technology-big data reference architecture-part 3: reference architecture)》《信息技术大数据参考架构第5部分：标准路线图(information technology-big data reference architecture-part 5: standards roadmap)》国际标准。

(2) SC32大数据标准化相关情况

SC32是与大数据关系最为密切的标准化组织，持续致力于研制信息系统环境内及之间的数据管理和交换标准，为跨行业领域协调数据管理能力提供技术性支持。其工作内容包括研制开发和维护有利于规范和管理的数据、元模型和本体的标准，此类标准有助于理解和共享数据、信息和过程，支持互操作性、电子商务以及基于模型和基于服务的开发。

2012年SC32成立了下一代分析技术与大数据研究组。2014年6月，启动4项为大数据提供标准化支持的新工作项目，包括结构化查询语言(structured query language, SQL)对多维数组的支持、SQL对JS对象标记(JavaScript object notation, JSON)的

支持、数据集注册元模型、数据源注册元模型。SC32现有的标准制定和研究工作为大数据的发展提供了良好基础。

(3) ITU-T大数据标准化相关情况

根据ITU-T现有工作基础开展的标准工作包括：高吞吐量、低延迟、安全、灵活和规模化的网络基础设施；汇聚数据机和匿名；网络数据分析；垂直行业平台的互操作；多媒体分析；开放数据标准。

目前，ITU-T大数据标准化工作主要是在第13研究组(SG13)开展，并由第17课题组(Q17)牵头开展ITU-T大数据标准化工作。2015年8月，发布“基于云计算的大数据需求和能力”，正在研究的课题包括“针对大数据的物联网具体需求和能力要求”“大数据交换需求和框架”“大数据即业务的功能架构”。

3.2 主要国家大数据标准化现状

目前，较系统地开展大数据标准化工作的国家主要包括美国和中国。美国国家标准与技术研究院在2013年6月建立了大数据公共工作组(NBD-PWG)，致力于开发大数据互操作性框架。中国在2014年12月2日由工业和信息化部信息化和软件服务业司指导成立了全国信息技术标准化技术委员会大数据标准工作组(以下简称“大数据标准工作组”)，全面开展我国大数据标准化工作。

3.2.1 美国NBD-PWG标准化情况

NBD-PWG工作范围是建立来自于产业界、学术界和政府的公共环境，形成共识的定义、术语、参考架构、安全与隐私和技术路线图，提出数据分析技术应满足的互操作、可移植性、可用性和扩展性需求，安全有效地支持大数据应用的技术基础设施，为大数据相关方选择最佳方案提供支持。

NBD-PWG是一个开放工作组，欢迎来自于产业界、学术界和政府的各方面力量参与并贡献力量。其发布《大数据互操作框架第1卷：定义(big data interoperability framework volume 1 definitions)》《大数据互操作框架第2卷：大数据分类(big data interoperability framework volume 2 big data taxonomies)》《大数据互操作框架第3卷：用例和一般需求(big data interoperability framework volume 3 use cases and general requirements)》《大数据互操作框架第4卷：安全和隐私(big data interoperability framework volume 4 security and privacy)》《大数据互操作框架第5卷：架构调研白皮书(big data interoperability framework volume 5 architectures white paper survey)》《大数据互操作框架第6卷：参考架构(big data interoperability framework volume 6 reference architecture)》《大数据互操作框架：第7卷：标准路线图(big data interoperability framework volume 7 standards roadmap)》等研究报告。

3.2.2 中国大数据标准工作组标准化情况

中国大数据标准化工作主要集中在全国信息技术标准化技术委员会大数据标准工作组，主要负责制定和完善我国大数据领域标准体系，组织开展大数据相关技术和标准的研究，申报国家、行业标准，承担国家、行业标准制订和修订计划任务，宣传、推广标准实施，组织推动国际标准化活动，对口WG9大数据工作组。

其下设7个专题组：总体专题组、国际专题组、技术专题组、产品和平台专题组、安全专题组、工业大数据专题组、电子商务大数据专题组，负责大数据领域不同方

向的标准化工作。目前，正在研制的国家标准有12项，详见表1。

大数据技术更新快速，为了适应新形势下标准化工作的新需求，大数据标准工作组不断加强标准的试验验证，快速迭代标准化验证和制定工作，积极推动标准化工作的快速成熟与落地。

4 大数据标准体系研究现状

大数据标准体系是为实现大数据领域的标准化而形成的体系。标准体系的建立应具有先进性，在应用系统科学理论和方法的基础上，运用标准化的工作原理，着眼于寻找整套的标准内容，基于这些内容，在标准体系的内在联系上进行统一、简化、协调和优化等处理，力求体现出系统内标准的最佳秩序，防止在标准之间存在不配套、不协调、互相矛盾及组成不合理等问题。随着大数据的发展，标准化的内容越来越广，标准化的对象也越来越复杂，大数据领域标准之间都存在着相互依存、相互衔接、相互补充、相互制约的内在联

表1 在研12项大数据领域标准

序号	计划号	标准名称
1	20141191-T-469	信息技术大数据术语
2	20141190-T-469	信息技术大数据技术参考模型
3	20141184-T-469	数据能力成熟度评价模型
4	20141172-T-469	多媒体数据语义描述要求
5	20141194-T-469	信息技术科学数据引用
6	20141200-T-469	信息技术数据交易平台交易数据描述
7	20141201-T-469	信息技术数据交易平台通用功能要求
8	20141202-T-469	信息技术数据溯源描述模型
9	20141203-T-469	信息技术数据质量评价指标
10	20141204-T-469	信息技术通用数据导入接口规范
11	20160597-T-469	信息技术大数据分析系统基本功能要求
12	20160598-T-469	信息技术大数据存储与处理平台技术要求

系,最终形成科学的有机整体。

大数据涉及各方面的内容越来越多,标准化工作的广泛性、复杂性主要体现在以下5个方面。

4.1 数据开放共享标准化缺乏顶层设计

政府开放数据不是政府信息公开,开放数据要把底层的、原始的数据进行开放,更多是要保障公众对政府数据的利用。真正的开放数据要满足完整性、可机读、一手、非歧视、及时、非私有、可获取、面授权等标准。数据开放、共享是数据运用的前提,许多地方政府不知道怎样开放、开放什么、开放程序是什么、管理方式是什么、考核评价标准是什么等,地方政府在推进数据开放工作上无所适从。我国政府数据开放共享时存在数据量少、价值低、可机读比例低、开放的数据多为静态数据等问题。因此,数据开放共享标准化缺乏顶层设计以及至上而下的执行标准、开放标准等。

4.2 大数据交易缺少标准

在大数据上升为国家战略的背景下,数据交易发展更是拥有了市场和政策的双重机遇。但数据交易发展的机遇与困难同在,数据交易、交换和服务发展面临一些问题:数据商品化需要先解决标准化问题,缺乏经过实践检验的有效的数据交易市场机制和运营模型,数据商品定价和数据资产评估困难,数据隐私保护和数据安全仍需加强,政府与企业的数据开放与商品化动机不强。

在数据标准化方面,交易所产品的重要特点就是交易产品的标准化。而大数据由于数据种类繁多,格式多样,难以形成一种普适的标准化方法,直接影响到其成为一种集中化、大规模交易的产品。由于

数据的应用场景和价值不容易标准化,数据应用水平和程度有限,数据标准化程度很低,无法按照传统的商品销售模式进行销售。

4.3 数据质量缺少规范

数据质量是影响大数据产业健康有序发展的重要因素之一。如何从海量数据中快速分析出有价值的信息,很大程度上取决于分析处理的数据能否真实地反映实际情况、分析的数据是否按一定要求在相同条件下收集、不同数据之间是否具有同质性、最终获得的数据是否具备合并统计分析的基础。然而,大数据时代下的数据质量应满足什么样的规范、是否达到规范的要求、大数据时代的数据质量与普通的数据质量之间的区别是什么、大数据时代的数据质量评估维度是什么,这些都是需要从标准的角度去解决的问题。

4.4 大数据系统评估标准缺乏

面向大数据需求的新硬件、软件和服务将形成巨大的市场空间。目前,开源软件平台为大数据存储管理和处理提供了基础,国内外主流解决方案提供商纷纷基于这些开源软件推出商用解决方案。在国家层面建立统一的测试方法,对大数据平台产品与服务的功能进行评价,是引导技术研发、系统建设、系统调优、采购选型等工作,促进大数据产品成熟的关键。为此,需要建立一套评价大数据系统产品的指标体系和评价方法。需要广泛吸取学术界和开源测试软件的成果,联合国内外厂商和用户,共同建立一套评价大数据系统和服务的测试标准,在确保测试结果能够充分反映系统特性的同时,简化测试配置,降低测试成本。

4.5 工业大数据问题突出

纵观大数据产业生态体系,我国工业大数据正面临一系列问题,阻碍产业化进程。“重硬件轻软件”变为“重软件轻数据”,工业大数据意识淡薄;工业大数据基础设施薄弱,企业数据安全问题突出;工业大数据标准尚未建立,数据获取效率低下;工业大数据技术创新与应用能力滞后,难以满足转型升级需求。

对于国内制造业企业来说,虽然很多企业已开始意识到将物联网和大数据技术应用到产品和服务中去,并积极地进行了初步的应用实践,但因为在建设过程中只强调数据获取的途径、性能、量级,没有考虑到数据的具体分析和利用以及相应的功能与目标,造成许多数据采集回来后没有可用之处或使用不充分,甚至一些关键数据反而没有采集。针对各领域的大数据标准化工作将为大数据相关技术在领域中的应用和发展提供重要的规范,因此十分有必要开展各领域的大数据标准化工作。

5 大数据标准体系框架

结合大数据参考架构、国内外大数据标准化工作部署、大数据标准体系研究现状及标准化需求,根据数据自身标准化特点、数据生命周期管理、当前各领域推动大数据应用的初步实践以及未来大数据发展的趋势,初步构建了大数据标准体系框架,如图2所示。

大数据标准体系框架由7个类别的标准组成,分别为:基础标准、数据标准、技术标准、平台/工具标准、管理标准、安全标准和行业应用标准。

基础标准为整个标准体系提供包括总则、术语、参考模型等基础性标准。数据标准主要针对底层数据相关要素进行规范以及数据交易、数据开放共享等方面的标准。技术标准主要对应大数据参考架构中大数据应用提供者的相关活动,针对大数据集描述、大数据处理生命周期和互操作等大数据相关技术进行规范。平台/工具标准主要对应大数据参考架构中大数据框架

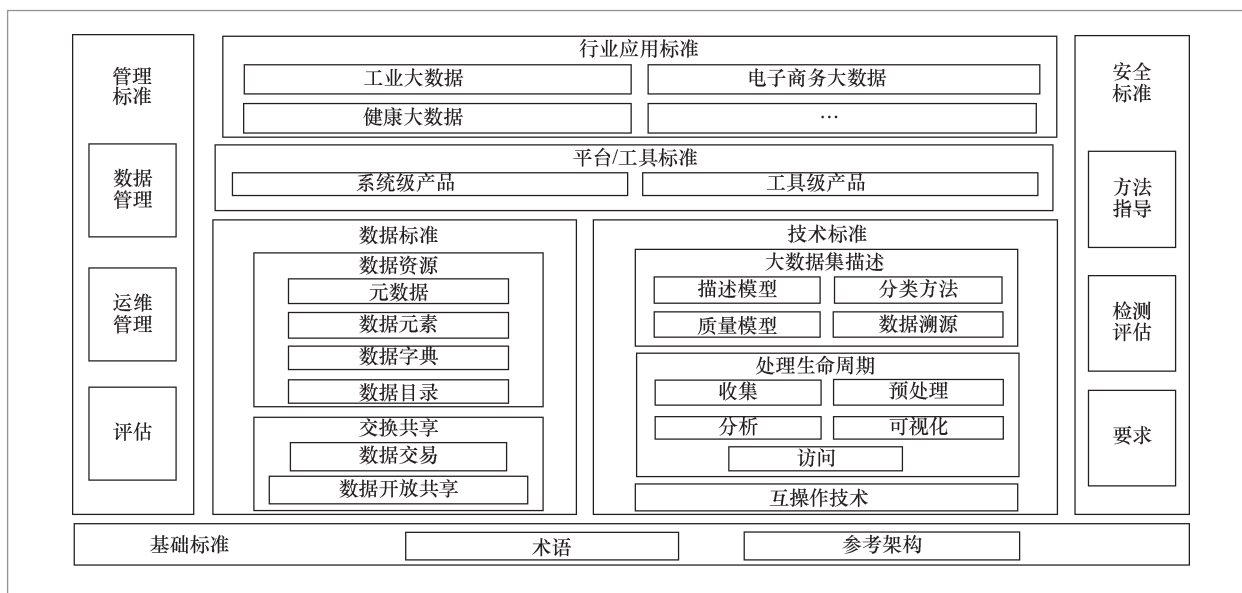


图2 大数据标准体系框架

提供者的相关活动,针对系统级产品和工具级产品等大数据相关平台和工具以及相应的测试方法和要求进行规范。管理标准以及安全标准作为数据标准的支撑体系,贯穿于数据整个生命周期的各个阶段,主要对应用大数据参考架构中安全与隐私、管理等相关活动进行管理规范。行业应用标准主要是从大数据为各个行业提供的服务角度出发制定的规范。

通过对现有各类标准情况进行分析可以看出^①:

①

http://www.cesi.cn/cesi/xxzx/bi_aozhunhuayanjiu/2014/1205/11689.html

- 在数据资源方面,我国已经研制的一些相关标准同样适用于大数据环境,目前急需加强这类标准的推广应用。

- 在交换共享方面,加快数据开放共享是国家重要任务,然而尚缺乏数据开放共享方面的标准,尤其是适用于政府数据开放共享的标准。虽然在研2项交易类的国家标准,但是尚缺乏交易流程、交易数据管理等方面的标准。

- 从技术标准上来看,在数据访问方面,目前已经发布和在研的数据导入和数据库相关标准适用于大数据底层数据接口,但是尚缺乏分析、可视化类标准;数据质量是大数据应用和发展的基础,目前有多项在研标准,但是均尚未发布,较为缺乏。大数据安全方面,部分现有标准适用,但是尚缺乏针对大数据的安全框架、隐私、访问控制类标准。

- 针对大数据平台和工具,目前发布和在研多项数据库、非结构化数据管理产品类标准,缺乏大数据系统级相关产品的标准;在大数据环境下,数据也已成为产品,而针对系统级和工具级产品等新兴产品,尚缺乏相应的标准。

综上所述,针对大数据,我国在数据管理、信息安全等方面已经发布和在研一些标准,适用于大数据环境,提供了一定的基础,但是缺乏标准化整体规划;数据开

放共享、数据交易、数据安全、系统级产品等方面的标准以及管理和评估类的标准较为缺乏,急需研制。

6 大数据标准化工作建议

(1) 完善大数据标准化工作平台建设

完善和维护大数据标准化工作平台,力争做到及时、准确、高效、有序,使得工作平台更具人性化和便利性。以我国大数据应用需求为基础,充分凝聚相关“产学研用”的力量,积极组织标准相关的应用调研,研究我国的大数据标准体系,加强大数据标准化顶层设计,积极出台相关建设指南,指导大数据标准化后续工作。加强研制大数据产业及应用急需的总体性标准草案,同时开展关键支撑技术标准、工程实施标准以及重点行业/领域应用标准的预研工作。

(2) 加强重点标准研制和验证推广

结合大数据产业发展需求,建立并不断完善涵盖基础、数据、技术、平台/工具、管理、安全和应用的大数据标准体系。加快基础通用国家标准和重点应用领域行业标准的研制。选择典型企业、重点行业、重点地区开展标准试验验证和试点示范,加强标准的宣贯、实施和应用推广。建立标准符合性评估体系,强化标准对市场培育、服务能力提升和行业管理的支撑作用。加强国家标准、行业标准和团体标准等各类标准之间的衔接配套。

(3) 推进国际标准化

我国拥有丰富的数据资源和应用市场优势,应尽量深入地进入国际标准化工作,大力推动标准化工作的国际化开放程度,加强我国标准化组织与相关国际组织的交流合作。组织我国产学研用资源,加快国际标准提案的推进工作,积极提交

国际贡献物。支持相关单位参与国际标准化工作并承担相关职务, 承办国际标准化活动, 扩大影响并争取国际关键标准的主导权。

参考文献:

- [1] 中华人民共和国国务院. 促进大数据发展行动纲要[R]. 北京: 中华人民共和国国务院, 2015. The State Council of the People's Republic of China. Action platform for greater data development[R]. Beijing: The State Council of the People's Republic of China, 2015.
- [2] 中国共产党中央委员会. 中华人民共和国国民经济和社会发展第十三个五年规划纲要[N]. 北京: 人民日报, 2016-03-17. Central Committee of the Communist Party of China. The 13th five-year plan for economic and social development of the People's Republic of China[N]. Beijing: People's Daily, 2016-03-17.
- [3] National Institute of Standards and Technology. NIST big data interoperability framework volume 6 reference architecture [S]. [S.l.:s.n.], 2015.
- [4] 张群. 大数据标准化现状及标准研制[J]. 信息技术与标准化, 2015(7): 23-26. ZHANG Q. Big data standardization current situation and standards development[J]. Information Technology & Standardization, 2015(7): 23-26.

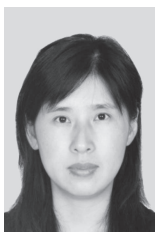
作者简介



张群 (1988-), 女, 博士, 中国电子技术标准化研究院设备与数据研究室副主任, 从事信息技术相关标准的制订、修订及研究工作, ISO/IEC JTC1/WG9大数据工作组专家, 主要研究方向为大数据领域相关技术及标准。



吴东亚 (1972-), 女, 中国电子技术标准化研究院信息技术研究中心高级工程师、副主任, 国家OID注册中心副主任。负责我国信息技术标准化工作, 先后主持承担数据、网络、信息化等领域国家标准、行业标准, 主导制定多项家庭网络、物联网等领域国际标准, 承担工业和信息化部、国家发展和改革委员会、科学技术部、国家质量监督检验检疫总局等多项科研项目, 获国防科技进步奖两次、上海市科技进步奖一次, 主要研究方向为信息技术标准化。



赵菁华 (1977-), 女, 中国电子技术标准化研究院高级工程师, 负责电子政务、数据库、办公软件、游戏和信息无障碍等相关技术领域国家标准和行业标准的制订、修订工作, 主要研究方向为信息技术标准化。

收稿日期: 2017-01-18

基金项目: 国家社会科学基金特别委托基金资助项目 (No.15@ZH012)

Foundation Item: Special Project Commissioned by the National Social Science Fund (No.15@ZH012)