

ISO/IEC JTC1/WG9大数据国际标准研究及对中国大数据标准化的影响

光亮¹, 张群²

1. 华为技术有限公司, 北京 100085
2. 中国电子技术标准化研究院, 北京 100007

摘要

ISO/IEC JTC1/WG9是制定大数据基础性国际标准的官方组织, WG9在研的大数据术语和参考架构等国际标准和相关技术报告对中国的大数据标准有重要的指导和借鉴意义。简要介绍了WG9工作组的构成、在研的国际标准和技术报告项目, 阐述了WG9标准在大数据国际标准体系中的基础性位置。对比了WG9国际标准和我国大数据标准体系的关系, 论述了WG9国际标准对我国大数据标准工作的意义和价值。给出了参与WG9标准工作组的建议。

关键词

大数据标准; 国际标准; 大数据参考架构

中图分类号: TP399

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2017038

ISO/IEC JTC1/WG9 big data international standards and the impact on Chinese domestic standards

GUANG Liang¹, ZHANG Qun²

1. Huawei Technologies Co., Ltd., Beijing 100085, China
2. China Electronics Standardization Institute, Beijing 100007, China

Abstract

ISO/IEC JTC1/WG9 is dedicated to the development of foundational big data standards. The on-going standardization works on vocabulary and big data reference architecture have great value and impact on the Chinese domestic standards. The structure and current projects of WG9, and the foundational status of its standards in international big data standards system were presented. In addition, the standards work in the WG9 and the Chinese domestic standards system were compared, and the values of participating in and contributing to WG9 standards were discussed. At last, suggestions on how to get involve in and contribute to WG9 were given.

Key words

big data standard, international standard, big data reference architecture

1 引言

ISO/IEC JTC1/WG9大数据工作组(以下简称WG9)是由国际标准化组织/国际电工委员会的第一联合技术委员会(ISO/IEC JTC1,以下简称JTC1)直接领导、致力于大数据基础性国际标准的专门工作组,不隶属于任何一个分委会(sub committee, SC)。WG9编制的基础性国际标准对于各国大数据标准和行业标准都有重要参考和借鉴意义。随着中国大数据产业的高速发展以及中国大数据标准化工作的全面展开,向ISO国际标准学习和借鉴,有助于快速吸收国际标准的成功经验,并在术语和参考架构等基础性标准上和国际标准、规范保持一致。同时,将中国大数据产业的优势贡献到WG9国际标准,将提升中国大数据标准的国际影响力。

2 WG9简介

WG9成立于2014年11月,属于JTC1比较年轻的工作组(与物联网工作组同时成立)。WG9的主要任务是制定大数据基础性(foundational)的国际标准,作为整个JTC1大数据标准工作的基础和参照,具体任务包括:

- 制定大数据基本性的标准,例如参考架构和术语;
- 在其他工作组不能覆盖的情况下,制定相关大数据国际标准;
- 识别大数据标准化中的缺失;
- 和JTC1内外的标准组织合作,分析正在进行的标准项目或探索新的标准项目,推动大数据标准化进程。

2.1 WG9构成

如图1所示,WG9由一名召集人(来自美国国家标准与技术研究院的Wo Chang)、24个国家代表团的157名专家组成(截至2016年底)。除了在研2项国际标准和3项技术报告外,还成立了3个特设组对特定标准问题进行深入研究。为了加强与其他标准组织的标准合作,WG9与多个国际标准和产业组织建立了联络函机制。

2.2 国际标准和技術报告项目

WG9目前在研的有两项国际标准和3项技术报告,见表1。

在研的两项国际标准如下。

- ISO/IEC 20546: 大数据—概述和术语^①。作为大数据官方标准,WG9制定的ISO/IEC 20546《大数据—概述和术语》为目前蓬勃发展的大数据技术和标准提供了权威的概念说明和术语解释。这对于统一产业和用户对大数据的认识、一致化技术和标准词汇都有重要意义。目前标准草案正在充分地研讨和修改中,除了对大数据的4个核心V——大量(volume)、高速(velocity)、多样(variety)、变化(variability)给出权威的标准解释,也对大数据技术和系统(如数据存储、计算模型等)以及相关技术(如云计算、集群计算等)涉及的常见术语给出标准定义。

- ISO/IEC 20547-3: 大数据参考架构—第三部分: 参考架构^②。ISO/IEC 20547-3《大数据参考架构—第三部分: 参考架构》是WG9正在编制的重要标准。由于大数据架构的纷繁多样,各种商业、开源版本共存,由WG9定义一个通用、标准并作为各种实现共同的基础架构,对提高各版本之间的兼容性(compatibility)、功

① <http://isotc.iso.org/livelink/livelink?func=ll&objId=17270859&objAction=browse&viewType=1>

② <http://isotc.iso.org/livelink/livelink?func=ll&objId=17287288&objAction=browse&viewType=1>

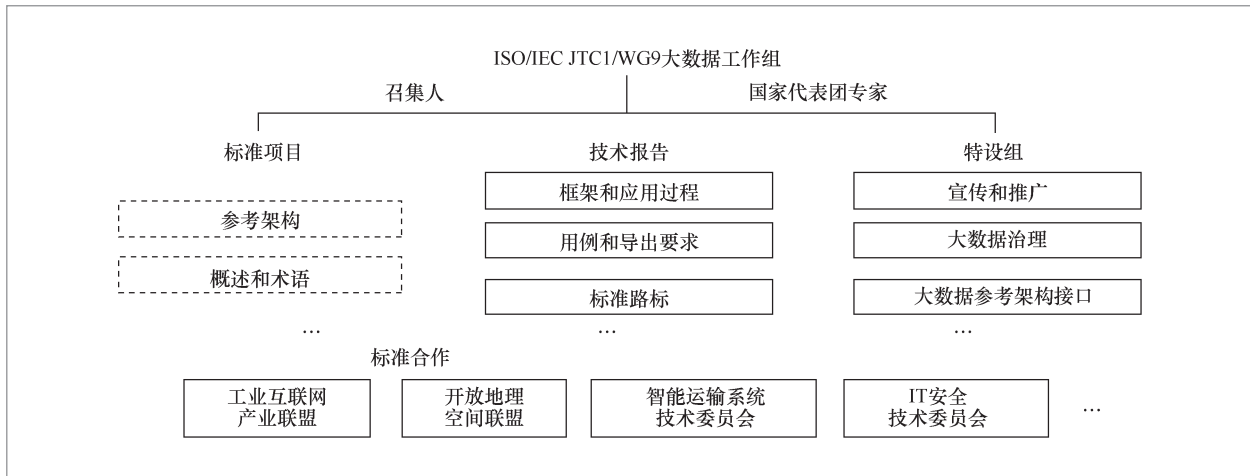


图1 WG9 构成与主要工作内容

表1 ISO/IEC JTC1/WG9 正在制定的标准 / 技术报告项目^③

类别	标准/技术报告项目	主要内容	价值和贡献
国际 标准	ISO/IEC 20546: 大数据—概述和术语	大数据的基本概念和重要术语	为大数据提供统一的基本术语和理解
	ISO/IEC 20547-3: 大数据参考架构—第三部分: 参考架构	大数据通用架构, 包括用户视图和功能视图	为大数据系统定义通用、标准的基础架构
技术 报告	ISO/IEC 20547-1: 大数据参考架构—第一部分: 框架和应用过程	构建大数据参考架构的框架以及利用参考架构构建大数据系统的过程	为构建大数据通用架构提供方法论依据; 为构建大数据系统提供标准化过程
	ISO/IEC 20547-2: 大数据参考架构—第二部分: 用例和导出要求	大数据典型用例及导出的对大数据系统的要求	为用户构建大数据系统提供参考借鉴, 并从典型应用中提炼出大数据系统的通用要求
	ISO/IEC 20547-5: 大数据参考架构—第五部分: 标准路标	大数据当前的国际标准、标准化存在的缺失以及今后的发展路标	综述大数据标准化情况, 识别标准缺失, 提出标准发展路标

^③ ISO/IEC 20547系列的第四部分《大数据参考架构安全与隐私》由ISO/IEC JTC 1/SC27 IT security techniques 制定

^④ <http://isotc.iso.org/livelink/livelink?func=ll&objId=17287288&objAction=browse&viewType=1>

能模块的互操作性 (interoperability) 以及上层应用的移植 (portability) 有重要意义。目前这个标准正在充分研讨和制定中, 主要是从用户视图 (不同角色、子角色和各种活动之间的关系) 和功能视图 (分层架构、功能模块以及模块之间的接口) 两个角度来定义大数据参考架构。

在研的3项技术报告如下。

- ISO/IEC 20547-1: 大数据参考架构—第一部分: 框架和应用过程。ISO/IEC 20547-1《大数据参考架构—第一部分: 框架和应用过程》主要介绍大数据参考架构的整体框架以及基于大数据参考架构进行系统设计的应用过程。

- ISO/IEC 20547-2: 大数据参考架构—第二部分: 用例和导出要求。ISO/IEC 20547-2《大数据参考架构—第二部分: 用例和导出要求》^④从政务、商务、防御、医疗与生命科学、深度学习与社交媒体、科研生态圈、天文和物理学、地球环境及极地科学、能源九大领域选取众多大数据的典型用例, 描述每个用例的引用场景、当前的解决方案、下一代的解决方案要具备什么特性及突破什么瓶颈。从这些用例中分析总结出下一代大数据系统和架构的共性要求, 作为制定大数据标准的依据。

- ISO/IEC 20547-5: 大数据参考架构—第五部分: 标准路标。ISO/IEC

20547-5《大数据参考架构—第五部分：标准路标》主要调研了JTC1和ISO内外的相关大数据国际标准以及这些标准和大数据参考架构之间的对应关系，从中识别出大数据标准化的主要缺失。

2.3 特设组

为了对特定议题进行深入的分析 and 讨论，WG9设立了多个特设组，当前有3个特设组：宣传和推广(awareness and engagement)特设组、大数据治理(big data governance)特设组、大数据参考架构接口(big data reference architecture interface)特设组。

其中，大数据参考架构接口特设组研究的是大数据功能模块相互之间接口的标准。例如大数据平台和应用之间的接口、数据分析层和计算及存储资源之间的接口以及功能模块的管理接口等。大数据功能模块众多，不同厂商的功能模块由于架构及接口的异构，容易造成系统不兼容、模块间不能互操作的情况。因此，在参考架构中定义统一、标准的接口能力，对功能模块的接口方式进行规范，能提高互操作性和兼容性，避免厂商锁定(vendor lock-in)。大数据参考架构接口已经成为ISO/IEC 20547-3《大数据参考架构—第三部分：参考架构》的重要章节，今后也可作为单独的技术报告，对模块间的接口能力和实现方式进一步规范、细化。中国代表团专家目前是特设组的主席。美国、印度、韩国等代表团的专家也在积极参与。

2.4 和其他标准组织的合作

为了推进大数据国际标准体系的发展，WG9与数据管理与交换(JTC1/SC32)、IT安全技术(JTC1/SC27)、云计算和分布式平台(JTC1/SC38)、统计方法的应用(ISO/TC69)、运动图像和音频编码(SC29/WG11)、智能型运输系统(TC204)、国际电联电信标准化部门第13研究组(ITU-T SG13)、工业互联网联盟(IIC)、开放地理空间协会(OGC)等多个标准及产业组织建立了联络函机制，互派联络员跟踪和对齐标准的进展。

与这些组织的合作(例如共同制定相关的大数据国际标准)有助于大数据标准体系的整体发展，推动构建一个统一、端到端的大数据系统架构，支撑越来越丰富的应用场景(例如工业互联网、下一代智能交通系统)等。

2.5 WG9主要职位情况

由于WG9标准在大数据国际标准中的重要位置，美国、德国、日本等发达国家在WG9中都争取了重要职位，包括2个标准项目和3个技术报告的首席编辑。随着中国代表团在WG9中的持续贡献，来自华为技术有限公司、中电长城网际系统应用有限公司等单位的专家获得了两个标准项目(ISO/IEC 20546、ISO/IEC 20547-3)的联合编辑职位、特设组的主席以及与其他标准组织的联络代表等职位(见表2)。

表2 中国代表团在WG9获得的主要职位

类别	项目名称	中国代表团职位	备注
编辑	ISO/IEC 20546: 大数据—概述和术语	联合编辑	首席编辑为美国人
	ISO/IEC 20547-3: 大数据参考架构—第三部分：参考架构	联合编辑	首席编辑为爱尔兰人
特设组主席	大数据参考架构接口特设组	主席	
联络代表	与JTC 1/SC 27的联络代表	联络代表	

2.6 评价

作为专门针对大数据这个新兴领域的工作组, WG9侧重编制对统一大数据的认知、规范大数据平台架构有重要价值的基础性标准, 符合大数据产业对国际标准的诉求。WG9工作组汇聚了美国、德国、中国、日本、印度等技术领先或是市场广大的国家和地区, 吸收了他们在大数据标准上的经验和成果。同时, 近几次WG9会议均同时召开大数据标准论坛, 吸引当地的厂商、本地标准机构, 特别是用户参与标准的讨论和制定。

除了组内的工作, WG9积极和JTC1内、外的标准和产业组织合作, 将大数据标准和上层应用标准结合, 这对于推动标准落地以及大数据生态的发展意义重大, 例如与工业互联网联盟和智能型运输系统的标准合作。

作为JTC1较新成立的工作组, WG9还有很大的发展空间。目前WG9的标准和技术报告大多是理论性的方法描述, 或是参考框架。大数据用户需要更加详细地了解大数据平台应该具备的能力、特性, 甚至

适合的接口方式, 才能支撑特定的大数据应用。WG9可以在这一方面扩展和深入, 例如以优秀实践、参考案例的方式为用户提供更有可操作性、更具参考价值标准或技术报告。

3 WG9和其他国际大数据标准

除了WG9之外, 还有不少国际标准组织、产业联盟在制定大数据相关的国际标准(见表3), 包括JTC1内的其他分委员会(例如IT安全技术、数据管理和交换等)、行业标准组织(例如国际电联电信标准化部门第13研究组)、其他标准和产业联盟等(例如W3C万维网联盟、开放地理空间信息联盟等)。这些标准组织侧重定义具体场景(例如地理信息系统、万维网)、特定接口(例如数据访问的SQL接口)或者具体特性(例如大数据安全)的国际标准。

WG9的国际标准在大数据国际标准体系中起到地基性的作用, 是其他大数据相关标准和规范的基础和参照(如图2所示)。

- 为国家标准和行业标准针对本地市场、行业诉求的定制及细化提供统一术语

表3 相关标准组织及其大数据标准工作

组织	委员会/工作组	主要标准/技术报告(举例)	侧重标准内容
ISO	数据管理和交换(JTC1 SC32)	结构化查询语言(SQL)对多维数组的支持 SQL对流数据的支持	数据库访问的接口、流数据访问的接口等
	IT安全技术(JTC1 SC27)	大数据参考架构—第四部分: 安全和隐私	大数据安全和隐私
	云计算及分布式平台(JTC1 SC38)	云计算—参考架构	云计算参考架构
ITU-T ^⑤	SG13下一代网络(包括云计算、移动和下一代网络)	云计算—参考架构	电信行业大数据标准
W3C ^⑥	资源描述框架工作组	资源描述框架, SPARQL(资源描述框架的一种查询语言)	互联网各种类型数据的统一语义层
OGC ^⑦	开放地理空间信息联盟	尚未发布	地理信息系统相关的大数据标准
TPC ^⑧	资源描述框架工作组	TPC-HS(用TeraSort做负载的大数据系统测评)、TPC-BB(测评基于Hadoop的大数据分析系统)等	大数据平台、大数据分析系统测评标准

⑤ <http://www.itu.int/zh/ITU-T/Pages/default.aspx>

⑥ <https://www.w3.org/>

⑦ <http://www.opengeospatial.org/>

⑧ <http://www.tpc.org/information/benchmarks.asp>

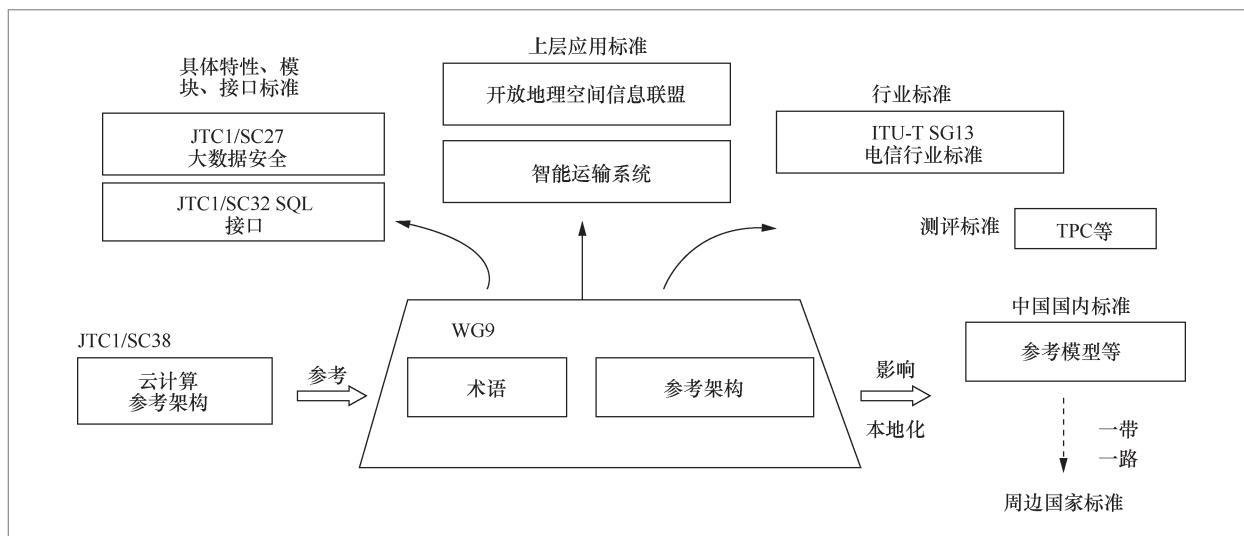


图2 WG9 标准在大数据国际标准体系中的地基性位置

解释和参考架构。

- 为各标准组织制定的具体特性（如安全）、特定模块（如管理模块）、接口标准（如SQL接口）提供统一的参考框架和基本能力要求，推动这些扩展的模块适配、系统兼容。

- 为上层应用（工业物联网、地理信息系统、智能交通系统等）提供统一的大数据基础平台，推动上层应用向大数据平台移植和部署，形成端到端的标准架构（体系）。

4 WG9标准对我国大数据标准的价值

随着我国大数据产业的高速发展，国内的大数据标准也在持续发展和制定中，特别是由全国信息安全标准化技术委员会大数据标准工作组牵头研制的大数据标准，包括《信息技术大数据技术参考模型》《信息技术大数据术语》《信息技术通用数据导入接口规范》在内的10多项国家标准^⑨。国内标准如何吸收和借鉴国际标准，特别是WG9大数据标准，并在此基础上扩

展，满足国内产业诉求，值得深入分析和讨论。

4.1 WG9标准和中国大数据标准的对比

中国的大数据标准相比WG9在研的标准，是一个更加广泛和深入的体系（如图3所示），包含7个方面。

- 基础：主要包括术语和参考架构，基于WG9的术语和参考架构标准。

- 数据：重点是数据资源的开放共享（例如开放数据的格式和开放的层级）和数据交换。

- 技术：包含数据生命周期管理、互操作技术等。

- 平台和工具：包括平台基础设施、平台管理工具、数据库产品等方面的标准。

- 管理：包括数据本身的管理及平台的运维管理和评估等。

- 安全和隐私：这一部分主要参考ISO/IEC JTC1/SC27。WG9不再做安全相关的大数据标准。

- 行业应用：针对具体行业（例如工业大数据、健康大数据等）定制和细化。

⑨

<http://www.sic.gov.cn/News/260/4166.htm>

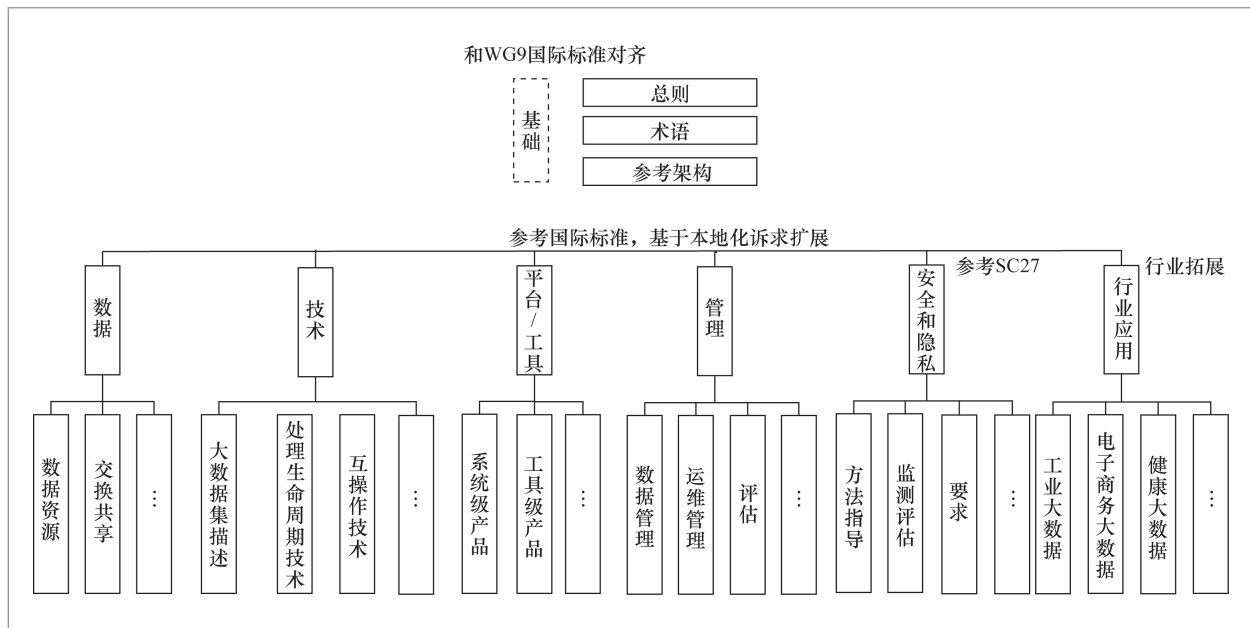


图3 中国大数据标准体系框架

对比WG9的标准项目和中国大数据标准体系，除了基础部分基于WG9的两项国际标准，其余国内标准都是针对中国产业诉求做的标准细化和扩展。WG9的3项技术报告——框架与应用过程、用例和导出要求、标准路标可以作为标准研究的参考。国内的大数据标准侧重数据开放共享、大数据平台各项能力要求及行业标准等。这些标准体现了中国本地市场对大数据平台的诉求（如庞大的用户数和丰富的业务类型对多租户能力、可靠性的要求）、特定行业（比如公共安全、工业制造等）对大数据技术的迫切诉求以及国家对数据开放、安全隐私的法规和要求。这些内容都超出了国际标准的一般范畴，却对中国大数据产业发展有重要意义。

4.2 WG9标准对中国大数据标准的价值

WG9的工作对应中国大数据标准体系的基础部分。这些基础性的标准对国内

的大数据标准建设有重要的指导和借鉴意义。

第一，基础性国际标准可以供国内标准引用和参照。例如在制定中的国家标准《信息技术 大数据 技术参考模型》^⑩就借鉴了ISO/IEC 20547-3《参考架构》中的分层架构和角色、活动、功能模块定义等。借鉴国际标准的定义、描述和要求是为了基于本地诉求进行扩展和细化。例如大数据参考架构中提及的功能组件和能力，可以作为国内大数据系统的基础要求，再基于国内用户诉求进行扩展，比如集群管理的要求、可靠性的要求等。

第二，通过参与、贡献WG9国际标准可以提升中国大数据产业、大数据标准的国际影响力。WG9工作组内，中国代表团已经获得了多个重要职位，包括两个标准项目（ISO/IEC 20546、ISO/IEC 20547-3）的联合编辑。这对于提升中国大数据标准的国际影响力以及中国标准在JTC1的整体影响力有着积极的推动意义。

第三，国内大数据厂商可以通过贡献

^⑩ <http://www.nits.org.cn/getIndex.req?action=query&req=modulenvpromote&id=3151&type=0&moduleId=83&sid=5>

WG9标准,推动我国大数据产品特性、产业诉求埋入国际标准,并通过国际标准的影响力提升在国际市场的竞争力。尤其是通过ISO/IEC 20547-3《大数据参考架构—第三部分:参考架构》将国内大数据厂商的优势和能力埋入大数据系统通用的参考架构中,在国际标准中体现我国的技术优势,提高市场竞争力。

第四,通过WG9与其他标准组织的合作,可以融入大数据产业的国际生态圈,推动我国大数据产业的整体发展和相关行业领域的信息化水平。WG9和其他国际标准组织的合作有助于行业应用向大数据时代转型。应该借助WG9和这些标准组织的合作,将中国的大数据标准融入国际大数据标准和产业的生态圈,推动国内的行业应用符合国际标准,与大数据平台无缝兼容,充分发挥大数据系统在提升工业生产力方面的潜能。

5 参与WG9标准的建议

由于对WG9国际标准的持续贡献,中国的大数据标准工作已经在WG9获得了一定的影响力。参与、贡献WG9的方式主要有提交修改意见、担任标准职务、申请新立项3种。

- 提交修改意见。针对当前标准或技术报告草案提出修改意见,包括对文字、图例进行修正,删减草案内容,增加新章节等。

- 担任标准职务。包括标准或技术报告的联合编辑甚至首席编辑、与其他标准工作组的联络代表、特设组主席等。尤其是大数据参考架构接口特设组,目前接口相关标准已经成为WG9的重要讨论内容和下一步细化工作的主要方向之一,也是国内接口标准的重要依据。我国参与和贡献这一特设组,可以主导大数据参考架构接

口标准的制定,推动国内产品的接口特性埋入国际标准,获得竞争优势。

- 新的立项申请(new work item proposal, NWIP)。如果所提内容符合WG9的工作范围,又尚未包含在当前标准或技术报告项目内,可以通过中国代表团提交新立项(标准或是技术报告)申请。

在全国信息安全标准化技术委员会大数据标准工作组,特别是国际专题组的组织下,国内专家向WG9提交了多份高质量的提案,涉及术语、参考架构、接口规范等各方面内容,获得了WG9的高度肯定,并担任编辑、特设组主席等多项职位。但是相比美国、德国、日本等发达国家,我国参与大数据国际标准的时间还不长,工作上也存在不足。特别是当前我国参与WG9的工作侧重提升影响力,真正的技术贡献不足。WG9还没有中国立项的标准或技术报告项目。当前项目的首席编辑也是由外国专家担任的。虽然我国有一定数量的提案,但技术层面的贡献和亮点有限,不能充分体现我国大数据产品的特性和优势。建议在争取话语权和领导力的基础上,提交更有广度和深度的技术提案,更加体现出中国大数据产品的优势和竞争力。

6 结束语

本文介绍了ISO/IEC JTC1/WG9大数据工作组以及在研的大数据国际标准和技術报告。通过调研相关大数据国际标准的内容和侧重,分析了WG9标准在大数据国际标准体系中的基础性作用。WG9的大数据国际标准为各种上层应用、特定行业标准、不同功能模块/接口标准等提供了统一、通用的参考架构和基础平台,推动行业应用向大数据平台移植和部署,提升系统兼容性。

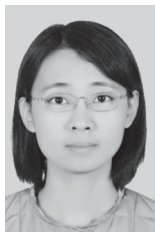
同时,本文对比了WG9国际标准和
中国大数据标准体系,讨论了参与、贡献
WG9国际标准对中国大数据标准化的价
值。中国大数据标准的基础部分(术语和
参考架构)引用和吸收了WG9的相关国际
标准,并基于本国的产业和市场诉求进行

多方面的扩展和细化。这些进展和成果可
以贡献、反哺到WG9国际标准中,埋入中
国的产业诉求、产品优势,并借助WG9和其
他标准组织的合作融入国际大数据标准生
态圈。最后也给出了参与WG9国际标准的
多种方式和建议。

作者简介



光亮(1982-),男,博士,华为技术有限公司高级标准经理,主要研究方向为云计算和大数据相关标准、国际标准。ISO/IEC 20547-3《大数据参考架构—第三部分:参考架构》标准的联合编辑,大数据参考架构接口特设组主席,ISO/IEC 29075《SQL对流数据的支持》技术报告的联合编辑。



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。

收稿日期:2017-02-07