美国政府NIST大数据互操作性框架的特点研究及启示

张 斌 王露露 张 臻

【摘 要】[目的/意义]对美国政府大数据互操作性框架提出的背景、具体内容和主要特点进行分析与总结,以期为我国制定大数据参考框架、促进跨界合作提供有益的参考。[方法/过程]以内容分析法和文本分析法为主要研究方法,以从美国NIST官网获得的公开政策、研究报告等作为主要数据来源,从数据层、框架层、角色层和应用层等方面分析总结美国大数据参考框架的特点。[结果/结论]分析发现:NIST构建了一个具有较强参考性与适用性的大数据概念框架,着重体现了大数据范式的前后变化并鼓励挖掘大数据应用的可能性。启示我国政府在制定大数据参考框架时,应当在理论层面达成共识的前提下,关注可参考价值与利益相关者的开发需求,同时在需求与价值之间构建起映射关系。

【关 键 词】大数据;政府;参考框架;概念模型;利益相关者

【作者简介】张斌(1967-),男,中国人民大学信息资源管理学院教授,博士生导师,研究方向:信息资源管理、知识管理(北京 100872);王露露(通讯作者)(1993-),女,中国人民大学信息资源管理学院博士研究生,研究方向:信息资源管理(北京 100872);张臻(1989-),男,中国人民大学信息资源管理学院讲师(北京 100872),北京电子科技学院管理系(北京 100070),研究方向:信息管理与网络安全治理。

【原文出处】《现代情报》(长春),2019.11.3~12

【基金项目】国家社会科学基金重大项目"大数据环境下政务信息资源归档和管理研究"(项目编号:17ZDA293)。

大数据已成为推动经济发展、完善社会治理、提升政府服务和监管能力的新动力和新途径。各国在积极制定和实施大数据发展战略的过程中,面临一个重要挑战就是如何处理好跨部门、跨领域的大数据管理问题从而发挥大数据的基础性和战略性价值。2016年5月,美国国家标准与技术研究院(National Institute of Standards and Technology,简称NIST)发布了大数据互操作性框架(NIST Big Data Interoperability Framework)并于2018年3月进行了更新[1],以适应新阶段的发展要求。美国的NIST大数据互操作性框架针对的是跨部门大数据管理与应用问题。本文通过分析与研究该框架,对面临同样发展困境的我国大数据发展具有一定的参考价值。

以"大数据+互操作/参考框架/参考架构/标准/概念模型"为检索关键词,笔者在中国知网检索到了87篇相关文献,在Springer、Science Direct和EBSCO检索到了323篇相关文献。通过中外对比,发现在关键词分布上国内外呈现出较为明显的区别。国内文献重点关注的是大数据指导标准的建立,譬如,肖筱华等[2]和张群[3]对当前国内大数据标准体系及标准研制情况的研究。相较而言,大数据参考架构和概念模型的研究成果不如标准多,但是也占据了较高的比例,譬如,郑大庆等综合了大数据治理的内部要素和外部应用特征构建了一个大数据治理参考框架[4]。国外文献相较于标准制定,更偏重于对参考架构的研究,Nadal S等遵循软件工程原则细化了大数据系

统的参考模型，并用它创建支持Semantic-aware大数据系统的软件参考体系架构[5]。Pääkkönen P等认为将Twitter、LinkedIn和Facebook等大数据开发案例的方法抽取到统一概念模型上尚且存在研究空白，因此，对已公布大数据用例实现架构进行了分析，由此提出了大数据系统的技术独立参考架构[6]。笔者认为，国家标准和行业标准提供的是相对具体的指导，在大数据范式尚处于探索阶段时，宏观概念层次的参考架构可以为大数据领域的创新提供更多的空间，抽象化的体系也更加有利于不同技术、组织和资源的融合与交流，然而，国内对该主题的研究尚显得较为薄弱，这为本文提供了研究空间。另外，笔者未发现以NIST大数据参考性框架为研究对象的文章，因此，本文以该框架作为介绍与分析的对象，具有一定的研究意义。

本文选择美国NIST大数据互操作性框架作为研究对象的主要原因如下：第一，该框架旨在促进政府各部门、学界与企业之间开展有效合作，所针对的问题是当前大数据发展过程中所有国家政府都需要面临的问题，大数据的概念之所以成立，在于数据通过有机、大规模集合可达成量变引起质变，该特性决定了必须进行跨部门、跨界合作，而在合作过程中的优劣互补、利益协调等问题同样困扰着我国政府部门。第二，2016年10月，习近平在主持中央政治局第三十六次集体学习时指出："以数据集中和共享为途径，建设全国一体化国家大数据中心，推进技术融合、业务融合、数据融合，实现跨层级、跨地域、跨系统、跨部门、跨业务的协同管理和服务[7]"。该指导理念与美国政府"大数据研究和发展计划"的核心原则有共通之处，都强调了对国家大数据开展工作进行集中指导与统一规划。NIST大数据互操作性框架是美国"大数据研究和发展计划"的政策产物，与我国自上而下的工作部署方向相一致，因此，可为我国的大数据战略开展提供一定的参考。第三，该计划于2016年形成，截至目前已实施了两年多的时间，在这期间并未废止且在向第二阶段推进，可见该框架具有较强的可行性；同时，该框架还对第三阶段的工作重点提前进行了规划，对于未来大数据的趋势形成了一定的洞见，因此，也具有一定的前瞻性。

1 提出背景

1.1 大数据的潜在价值催生合作需求

早在2002年，为了对大容量的流数据进行实时数据分析，美国政府就开发大规模可拓展的集群基础设施与IBM公司展开合作[8]。由此带动IBM后续开发的IBM InfoSphere Stream和IBM Big Data等大数据产品受到了美国政府和企业的广泛欢迎。2009年，美国政府Data.gov网站开始运行，大大推动了美国的政府信息公开和数据开放。所建设的数据仓库整合了涵盖交通、经济、卫生保健、教育和人类服务等领域的数据以及多个应用的数据源[9]。2010年，总统科学技术顾问委员会在其《设计数字化未来：联邦资助的网络和信息技术研究与开发》(*Designing a Digital Future：Federally Funded Research and Development in Networking and Information Technology*)报告中明确阐述了美国即将实施大数据战略。2012年，奥巴马政府启动"大数据研究和发展计划"(Big Data Research and Development Initiative)，总投资为2亿美元，计划涉及80多个合作项目，要求多个联邦部门共同参与，包括白宫科技政策办公室，国家科学基金会，国家卫生研究院，国防部，国防高级研究项目局，能源、健康和人类服务部以及美国地质调查局。该计划明确要求产业界、研究型大学和非营利组织与联邦政府合作，最大限度地利用大数据带来的机遇[10]。

由上述发展趋势及其政策要求可见，当前美国无论是政府部门、商业界，还是学术界，都已经充分认识到大数据在推动经济社会发展和增进人类福祉等方面的潜在价值。美国已从总统层面开始推动各个部门之间积极开展合作，同时，美国政府也与IBM、Aamazon、Google等公司展开合作，从技术研发、产业应用等方面共同推动大数据的发展。因此，可以说，大数据的潜在价值已促使利益相关者之间广泛构建和发展合作关系。

1.2 大数据技术应用带来挑战和问题

尽管跨部门和跨界合作的政策环境已经基本具备，但是在具体的实施过程中却面临着诸多问题与挑战，主要表现为两个方面：一是在大数据的几大关键问题上尚未达成共识。NIST大数据公共工作小组(Big Data Public Working Group，NBD-PWG)认为，未

达成共识的问题包括:1)哪些属性可以用来界定大数据解决方案;2)大数据与传统数据环境的应用流程有何区别;3)大数据环境的基本特征是什么;4)新环境如何与当前部署的体系结构进行集成;5)为加速部署强大的大数据解决方案,需要解决哪些核心科学、技术和标准化问题带来的挑战。二是尚未形成足够的大数据应用能力[11]。美国白宫科技政策办公室前主任霍尔德伦(John P Holdren)认为:美国拥有大量善于生成数据的机构,但作为一个国家,还没有充分发挥我们的能力来共享潜在竞争资源、协作分析与分享经验[12]。不同于其他物质型的国家资产,他们所对应的实现场景和所具备的价值是清晰可见的,大数据属于信息导向型资产,需要多元化的利益主体共同参与,通过持续的试验与探索才可以发现其潜在的应用价值,因此,需要足够的协作经验与顶层指导为大数据战略的开展保驾护航。

根据2012年"大数据研究和发展计划"要求,NIST开始着手制定大数据互操作性框架,以促进大数据有关专业力量间的合作,进一步确保大数据的安全和有效应用。2013年1月15~17日,NIST举办了"云与大数据论坛",专门成立了大数据公共工作组负责开发大数据互操作性框架。2016年5月11日,NIST正式发布了大数据互操作性框架1.0版本,将美国的大数据发展分为3个阶段,不同阶段的工作任务对应参考框架的特定环节。2018年3月23日,NIST又对大数据互操作性框架进行了更新,明确指出当前美国大数据的发展已步入第二阶段[13]。

2 核心概念界定

要在大数据关键领域达成共识,确保利益相关者合作项目的顺利开展,必然要进行核心概念的界定。因此,该框架的目标之一是形成基于共识的理论范式,为实际操作的交流消除误区,同时也促进对大数据技术有更深刻的理解与认知,扩大其影响力。

尽管大数据具有很多特征,但是大体量(Volume)、多样性(Variety)、时效性(Velocity)和可变性(Variability)的"4V"特征真正推动了新型数据密集型并行架构的产生,并且决定了对大数据系统的整体设计和大数据生命周期模型的构建。基于大数据的"4V"特征,NIST将大数据界定为:"大数据由大量数据集组成,主要集中在数量、种类、速度和/或可变性等特征上,这些数据集通过建设可扩展架构可实现高效的存储、操作和分析。"值得注意的是,NIST在概念界定中强调了各个特征之间的相互作用关系,同时重点关注了为了满足所需性能和成本效率需求可以使系统架构变得可扩展。"系统架构可拓展"通常被描述为垂直或水平拓展两种思路,垂直拓展意味着增加处理速度、存储和内存的系统参数,以获得更高的性能。这种方法受到物理能力的限制,其改进需要引入更复杂的元素(例如,硬件和软件),无疑会增加现实过程中的时间和经济成本。另一种方法是使用水平扩展,即利用集成的分布式单个资源作为单个系统,而这种横向扩展才是大数据革命的核心。同时,NIST也将与大数据系统设计相关的子概念进行了界定,譬如,大数据范例(Big Data Paradigm)包括跨水平耦合的独立资源分布数据系统,旨在提供有效处理大量数据集所需的可扩展性[14]。

3 美国的NIST大数据互操作性框架及其特征

NIST大数据互操作性框架的开展以NIST大数据参考架构(NIST Big Data Reference Architecture,NBDRA)的构建过程为主线,分为以下3个阶段:第一阶段,确定高级别大数据参考架构关键组件,这些组件是技术、基础架构和供应商当前所不可知的。第二阶段,定义NBDRA组件之间的通用接口。第三阶段,通过通用接口构建大数据通用应用程序来验证NBDRA。不同的发展阶段对应不同的框架版本,指导相应阶段大数据公共工作小组目标的实现。

NIST大数据互操作性框架主要由概念、分类、应用案例和一般要求、安全和隐私、架构白皮书调查、参考架构和标准路线图七大主题组成,这些主题并非随意选择,是由大数据公共工作小组通过调查与研究所得。本文在进行介绍性分析时,并未按照该框架的主题顺序展开,原因在于各个主题之间前后逻辑顺序与相互关联性较弱,不便于在文章中进行系统性分析与特征总结。因此,笔者对各个主题进行了整合与概括,将其分为数据层、框架层、角色层和应用层。

3.1 数据层:关注新旧数据范式的变化

理解大数据工程首先需要理解数据本身的特征。通过检查不同颗粒度的数据在数据资源中所占

的比例情况，可以更好地看到数据是怎样改变了大数据范式以及不同数据层级需要重点解决的问题。因此，NIST提供了基于不同数据粒度的数据特征分析，见图1。数据特征层级模型(Data Characteristic Hierarchy)将大数据的数据状态分为数据、文件、数据集和多个数据集4个层次。

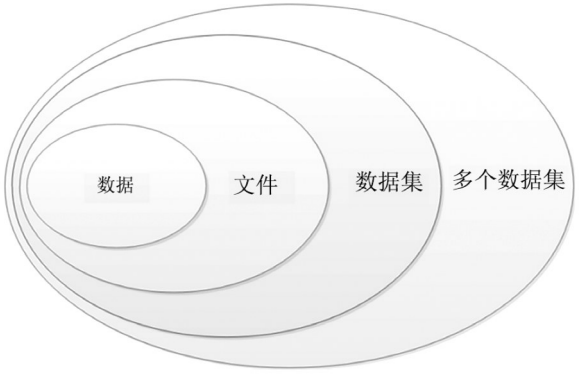

图1 数据特征层级模型[15]

数据层在新的大数据范式中没有发生大的变化，还是通过自身的数据类型和其他上下文数据(或元数据)，元数据提供关于数据的历史记录等进行理解。数据层关注的是数据格式、数据价值和词汇表、元数据和语义、质量和真实性。数据会被分配到描述具体实体、事件或者事物的文件中，即文件层。在文件层，体现出了大数据带来的变化，譬如，在非结构化的文本中，1个数据文件可以指的是1个短语或者句子。该层次关注的是文件格式、复杂性、容量、元数据和语义。文件分组后即形成了数据集，数据集层次也体现了大数据带来的变化。例如，在非结构化文本中，数据集指的是完整的文档。该层次关注的是质量与一致性。对多个数据集的关注即形成了集成或融合多个数据集的需求，该层次体现的是大数据的多样性。大量的数据集不能总是转换成一个集成结构，例如，大量的天气数据无法都转换在同一个时空网格上。由于无法将大容量数据集简单复制到规范化结构中，因此正在开发新的技术来根据业务需求来集成数据。例如，在非结构化文本中，多个数据集可以同时引用一个文档集合。该层次关注的是个体数据集的标识。

3.2 框架层：提高系统框架的可参考性

制定统一的参考框架(Reference Architectures)可以通过权威的信息来源，为某个主题领域存在的多样化的系统架构和解决方案提供指导并给予一定的约束[16]。鉴于大数据领域的复杂性，NIST专门推出了大数据参考框架(NIST Big Data Reference Architecture，NBDRA)。为此，专职工作小组调查了目前支持大数据框架的领先企业或个人发布的大数据平台，并对收集到的资料进行了分析，从中提炼出了当前普遍的大数据开发架构之间的一致性，并将调查结果形成了白皮书，即架构白皮书调查(Architectures White Paper Survey)[15]。

NIST大数据架构主要由2个坐标轴、5个角色与2个底层结构组成。首先，该框架围绕信息价值(Information Value)横轴和信息技术价值(Information Value Technology)纵轴展开。沿着横轴，通过数据收集、保管、分析和可视化等价值链后续流程来创建价值。沿着纵轴，通过提供网络平台、基础设施、应用工具和其他IT服务来创建价值，这些服务用于承载和操作大数据，以支持所需的数据应用程序。其次，5个角色指的是系统协调员(System Orchestrator)、数据提供者(Data Provider)、数据用户(Data Consumer)、大数据框架提供者(Big Data Framework Provider)和大数据应用提供者(Big Data Application Provider)。在这些角色中，需要注意的是，大数据应用程序提供者和大数据框架提供者使用“Provider”一词表示这些组件在系统中提供或实现特定的技术功能，并非普遍意义上的“提供”。总体上看，这5个角色是在任何大数据系统中都必然存在的技术角色。其中，系统协调员负责定义所需的数据应用程序活动，并将其集成到一个可操作的垂直系统中；数据提供者负责将新的数据或信息导入大数据系统；大数据应用提供商负责执行数据生命周期，满足安全、隐私需求和系统编配定义需求；大数据框架提供商负责建立计算框架，在转换特定应用的同时，保护彼此数据的隐私和完整性；数据用户指的是终端用户及其使用大数据应用程序成果的其他外部系统。最后，容纳5个角色的2个底层结构分别是隐私与安全(Security and Privacy)、管理(Management)，这两个底层结构是所有大数据系统都必不可少的，负责为系统的所有组件提供保护隐私与安全的职能和管理的

服务。此外,图2中的服务应用代表软件的可编程接口,“DATA”表示数据在组件之间通过引用或者直接的物理流动,“SW”表示在处理流程中大数据软件工具发生转移。

3.3 角色层:增强参考架构的适用性

在传统的数据项目中,数据系统一般是由一个组织进行主持、开发、部署和资源承载,而在大数据时代,系统的开发布局则是转变成为分布式的。由2.3的大数据参考框架可见,在系统中会出现多种技术角色,这些角色可以是个人、组织、硬件或者软件,某个角色可以固定在某个业务实体中,也可以由不同的业务实体共同实现,无需指定具体的参与角色或在合作情况下划分清晰的业务边界。因此,大数据系统需要适用于各种不同的业务环境,既要满足紧密集成的企业系统,又要适应松散耦合、依赖不同利益相关者合作的垂直行业。

NIST构建的大数据互操作性框架就是为了达成上述需求,他们认为在大数据系统开发项目中,“角色”(Roles)与“演员”(Actors)之间的关系与电影角色类似,某个角色可以由不同的演员来承担,而不同的演员也可以重复扮演同一个角色,同样地,某个活动可以由不同的行动者来承担,而不同的行动者也可以承担多种活动。为此,NIST提出了基于NBDRA系统的“角色”与“演员”样本分类体系(Roles and a Sampling of Actors in the NBDRA Taxonomy),具体参见下页图3。NBDRA“角色”与“演员”分类体系中的7个角色是由2.3的大数据参考框架中的5个角色与2个底层结构组成,即系统协调员、数据提供者、数据用户、大数据框架提供者、大数据应用提供者、隐私与安全和管理角色组成。

7个角色的含义与职能如下:1)系统协调员负责提供并确保系统必须满足的总体需求,包括策略、治理、体系结构、资源、业务需求、监视或审计等。虽然该角色的出现早于大数据系统,但在大数据范式中,一些与之相关的设计活动实则已经发生了变化,应当进行相应的调整与更新。2)数据提供者负责为自己或者其他角色提供数据。NIST提出的这一概念本身并不新鲜,但是大数据带来了强大的数据收集和

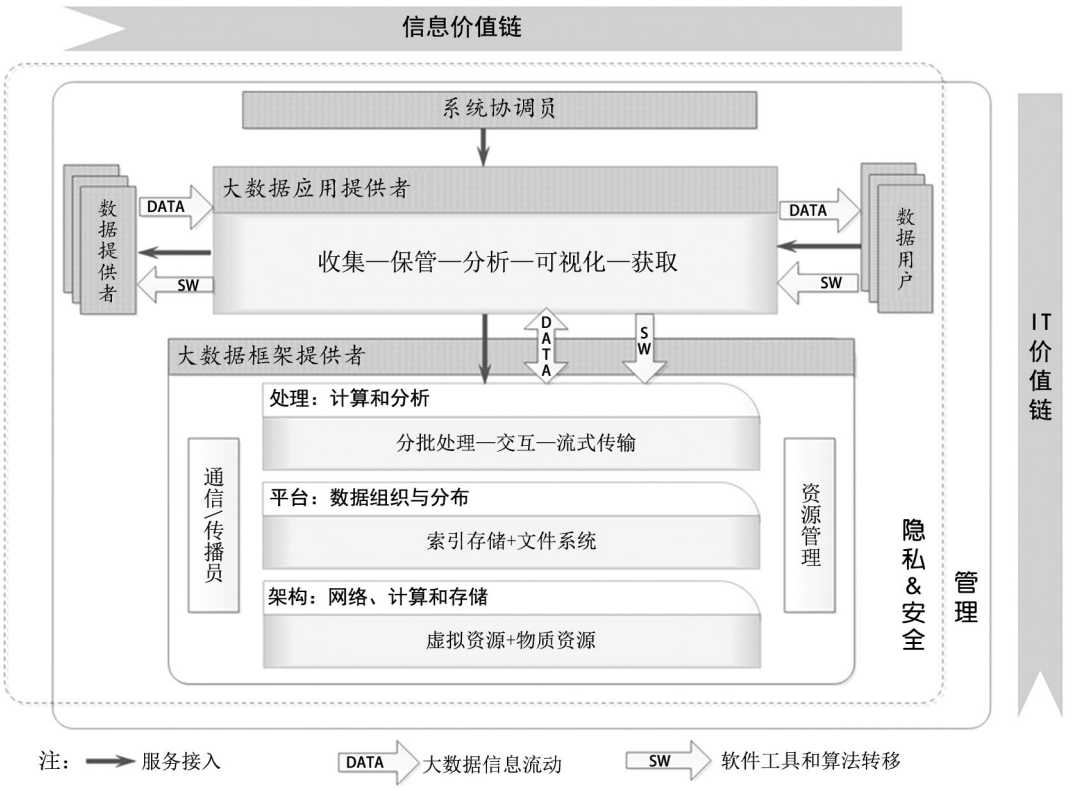

图2 NIST大数据参考概念模型[17]

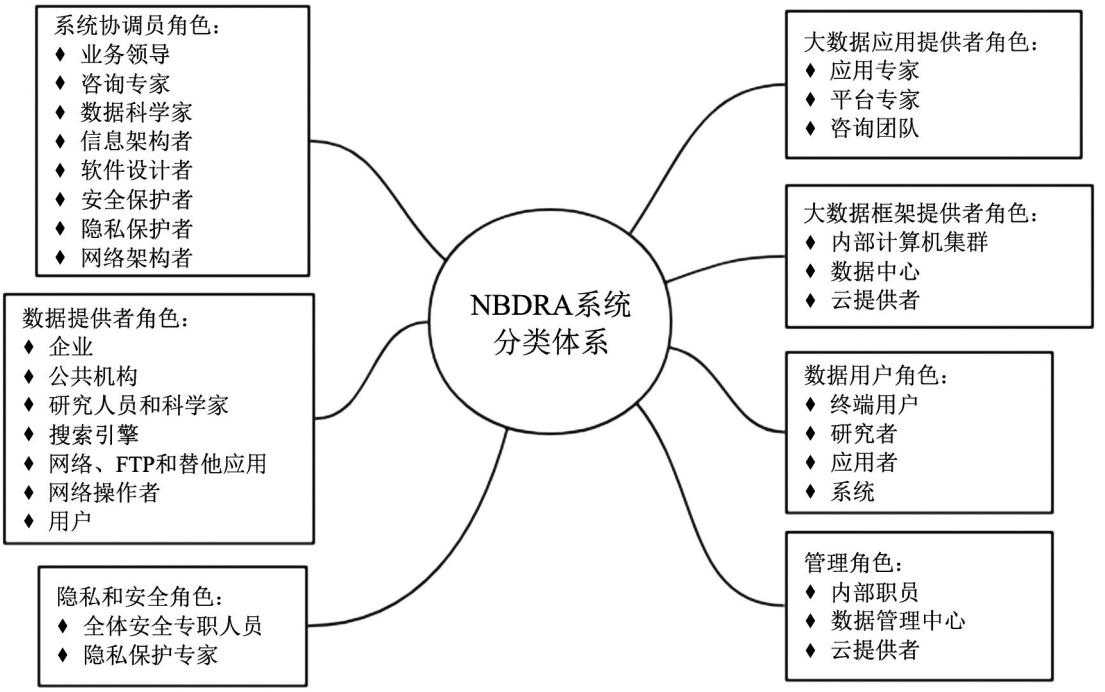

图3 基于NBDRA系统的“角色”与“演员”样本分类体系[17]

分析功能，为该角色创建数据价值开辟了新思路。此外，政策环境也为该角色提供了助力，美国政府积极倡议开放数据，作为公共数据管理者的联邦机构也应积极承担数据提供者的角色。3)数据用户是大数据系统的价值输出所在。正常情况下，数据用户收获的价值应当与大数据提供者提供的服务相对接。该角色受大数据系统的影响较小，较为明显的相互作用主要在于反馈市场需求。4)大数据框架提供者拥有大数据应用程序提供者在创建特定应用程序时所需要的一般资源或服务，包括基础架构框架、数据平台框架和处理框架等。大部分情况下，该角色提供的是多种技术的混合实现，这是大数据所带来的新变化，也是未来需要关注的新领域。5)大数据应用程序提供者按照数据生命周期执行具体的操作，满足系统协调者提出的需求，同时满足安全和隐私需求。该角色是将大数据框架内的一般功能结合起来产生特定数据系统的地方。6)隐私与安全角色主要开展的活动时隐私与安全的政策制定与监控，需要与系统协调员在政策、需求和审计等方面进行合作，还需要在系统开发、部署和操作方面，与大数据应用程序提供者和大数据框架提供者进行交互。7)管理角色是为了顺应大数据的4V特征而建立的多样性、复杂性和多功能平台，主要用于存储、处理和管理复杂数据。该角色既涉及处理大数据环境下的系统相关，也要处理大数据环境下的数据相关。

3.4 应用层：挖掘大数据应用的可能性

如果将大数据系统视为一个黑箱，想要尽可能地输出更多的数据价值，就必须从源头掌握大数据最新的技术、面临的挑战、市场发展等相关情况，从而向黑箱输入更加合理的需求，因此，应当重视挖掘大数据应用的更多可能性，应用问题也是大数据发展过程中较为突出的一大挑战。NIST为此专门进行了相关调研，形成了一份跨利益相关者的需求清单，该清单列举了九大基础应用领域，分别是：政府运行、商业活动、国防、医疗与生命科学、深度学习和社交媒体、研究生态系统、天文学和物理学、地球、环境与基地科学和能源九大领域。基于上述九大领域，NIST又从中归纳出了七大数据需求，使之更有延伸空间和概括性。具体的需求是：数据资源(如，数据大小、文件格式、增长率、静态或动态)；数据转换(如，数据融合、分析)；功能(如，软件工具、平台工具、硬件条件)；数据使用(如，以文本、表格、可视化和其他格

式处理结果);安全性和隐私;生命周期管理(如,策划、转换、质量检查、分析前处理)和其他需求。针对七大需求,可以与2.3的系统架构的7个角色相对应,形成数据需求—NBDRA组件映射表,具体参见表1。

值得注意的是,在挖掘大数据应用可能性的同时,也要注意隐私和安全的保护。NIST将大数据系统中的隐私与安全分为以下5种情况,分别是数据保密性、数据来源、系统状况、公共政策、社会和跨组织主题。前3种情况大致符合传统的数据机密性、完整性和可用性要求,在大数据范式下,又被重新定位为需要并行考虑的大数据隐私与安全问题。

4 对我国的启示

4.1 以可控性为前提,在概念层面达成局部共识

我国国务院2015年8月印发的《促进大数据发展行动纲要》将大数据界定为:"以容量大、类型多、存取速度快、价值密度低为主要特征的数据集合,正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析,从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态"[19]。无论国内还是国外,当前对于大数据尚未有一个公认的定义,不同的定义基本是从大数据的3v或者4V等特征出发,3V与4V都是当前在大数据领域较为普遍地达成共识的大数据特征描述,除了之前提到的4V,此处的3V指的是大体量(Volume)、多样性(Variety)和时效性(Velocity)。总之,各界学者试图通过这些特征的阐述和归纳试图给出其定义[20],包括本文研究的NIST大数据框架。尽管如此,在概念上对大数据形成局部共识还是十分必要的,一方面,在实践过程中,利益相关者需要致力于构建同一个解决方案,相互之间需要共同语言进行交流与理解,方可使各方作用于同一着力点,尽量减少沟通问题和扩大参与范围;另一方面,大数据解决方案与大数据的特征是密不可分的,大数据的特征决定了大数据应用方案的创新,大数据思维带来了不同粒度层次的数据价值的变化,这些是在具体实施过程中无法绕开的理解性问题,必须在特定工作范围内达成一定的共识。

笔者认为,在实践过程中,对于概念的界定,应以可控性为重点,以适用人群为导向,达成局部的共识,不必过度拘泥于具体的表述。任何范式在最初的时候概念都是纲要性的,探索性的概念界定不一定是完美的。因此,从大数据的3V或者4V特征出发,确保利益相关者在讨论数据和数据系统时是可控的即可。譬如,NIST的概念界定是从4V特征出发,重点服务于以下人群:一是面向管理者,为他们理解这一变化领域所需的整体规划提供支持;二是面向组织者,有利于理解组织需求并区分不同的解决方案;三是面向应用者而言,可以促进大数据解决方案和应用的创新与协作;四是面向技术人员,将提供一种通用语言,方便他们更好地区分大数据的特定技术产品。在概念界定的结构上,可以借鉴NIST的方案,围绕核心概念,扩散出子概念,形成一个逻辑清晰、逐步深入的"概念树";在形式上,最终可以形成一份专有名词词汇表,方便社会各界进行查阅。另外,笔者认为,在进行概念界定时,需要体现出变化以及不同范式之间的差异性。Kuhn T在《科

表1 数据需求—NBDRA组件映射表[18]

数据需求		系统架构组件
数据资源	→	数据提供者
数据转换	→	大数据应用程序提供者
功能	→	大数据框架提供者
数据使用	→	数据用户
安全和隐私	→	安全和隐私底层架构
生命周期管理	→	系统协调者、管理底层架构
其他需求	→	所有组件和底层架构

学革命的结构》中提出，不同科学范式之间具有不可通约性，即在革命和范式转换过程中，是世界观的转变，即便是同样的用词，他们的真实含义也已改变[21]。Jim Gray将大数据视为“第四范式”，大数据在概念上带来的混淆与不确定性需要通过新旧范式对比来进行解释。

4.2 以指导性为目标，设计统一的系统构建框架

从国家层面规划大数据发展是一个系统性、互操作的工程，需要融合多种技术与方法来共同解决问题。因此，构建一个无关技术与基础设施的、统一的、中立的概念性结构模型，一方面，可以提高利益相关者对各种大数据组件、流程和系统的理解，鼓励他们遵守国家推荐的标准、规范和模式；另一方面，又可以为政府部门、机构和其他用户提供技术参考，以便共同理解、讨论、分类和比较大数据解决方案。我国政府在制定统一架构时，可以适当借鉴NIST制定的大数据参考架构，在此基础上，笔者认为，以下4点可供进一步思考与分析。

一是在架构构成元素方面。大部分参考架构在设计参考元素时都需要基于已有的开发案例数据进行归纳推理，而数据流、数据存储和功能组件这三大基本角色是必须具备的[6]。

二是在大数据应用者环节。该环节分为收集、保管、分析、可视化和获取等活动，但是针对不同的垂直领域，需要制定用于子组件之间定义与交换的元数据策略，不便于进行标准化。另外，尽管这些活动还是传统数据管理活动的基本流程，但是大数据实则从本质上改变了他们的含义、价值和实现方式，因此需要对算法、机制和应用程序进行调整与优化，使之具有较强的可扩展性和较高的响应能力。

三是在大数据框架提供者环节，大数据领域对处理容量、多样性、速度和可扩展性等的新要求大部分都发生在该环节，促使大数据框架的相关技术研发逐渐成为迭代更新的热点，因而当前该环节具有相对充足的参考信息可供选择，可以进行进一步的细化与标准统一。

四是在整体上，需要关注对互操作性、可移植性、可重用性和可扩展性等方面的分析与介绍[22]。此处，值得注意的是美国大数据互操作性框架的制定者是美国国家标准与技术研究院，该机构主要为美国发展提供标准、标准参考数据及有关服务[23]，可见大数据框架所承担的作用相当于标准类知识产品，制定机构本身的职能也使之可以起到较好的统筹规划、提供指导的作用。

4.3 以适用性为要求，考虑各利益相关方的需求

NIST制定的“角色”与“演员”样本分类体系实则是对NBDRA框架的补充，但是因为该体系的形成来自NIST对已有大数据用例的市场调研，且这样的列举可以从某种程度上提高不同参与者对大数据标准的响应度与参与度，因此该体系的制定也值得我国进行借鉴与参考。

笔者认为，在制定类似的参与者框架时，需要注意以下几点：一是对业务环境适应性的关注，大数据分布式的开发布局不仅促进了跨界合作，还使得大数据利益相关者边界日益模糊化，对此，NIST采取的解决思路是对利益相关者的需求进行抽象化，转化为对不同业务环境的适应性。二是对抽象与具体的把握。NIST大数据参考架构并不是特定大数据系统的架构，而是一个基于公共参考框架用于描述、讨论和开发的工具。该模型不绑定到任何特定的供应商产品、服务或参考方案，也不定义限制创新的说明性解决方案。尤其是在“角色”与“演员”关系的构建方面，始终重点阐述的是角色的含义及目标，并未过多限定“演员”的可能性与职责。三是对数据所有权和数据治理的关注。NIST在第3阶段的发展目标中明确提出了对这两个问题的关注。这两个问题属于社会影响问题，虽然并非是当前亟须考虑的，但是鉴于顶层架构的宏观指导性与社会影响力，都应当将它们纳入讨论范围内。四是NIST在设计大数据架构时灵活运用了分类法。无论是3.1数据层的数据层级还是3.3角色层的角色分类，都可以更加清晰地展现大数据的概念结构，不断收集新的数据并逐渐完善分类也是归纳推理法的体现，是NIST构建技术架构、数据层级、应用领域和角色体系的重要方法。

4.4 以应用性为重点，构建需求与价值的映射关系

NIST的应用层构建主要特色在于：一方面，鼓励相关机构、学界和普通公众自下而上提交市场案例；

另一方面，将基于市场调研提炼出来的数据需求与大数据架构的功能组件相对应，形成一个相互映射的需求功能表。我国政府发布的《促进大数据发展行动纲要》提出了："促进大数据应用市场化服务为重点，引导鼓励企业和社会机构开展创新应用研究，深入发掘公共服务数据，在城乡建设、人居环境、健康医疗、社会救助、养老服务、劳动就业、社会保障、质量安全、文化教育、交通旅游、消费维权、城乡服务等领域开展大数据应用示范[19]"。为了实现上述目标，可以适当借鉴NIST分析大数据应用案例时所形成的需求功能表，在政府重点提到的应用示范领域构建大数据应用范例，以便于分析利益相关者的需求和构建指导性框架所需的功能组件。

具体可以从以下3个方面参考：一是重视规范化市场数据的收集，可以制定标准化的大数据应用案例信息提交表，引导大数据实践工作者详细收集数据的生命周期流程、数据资源的4V特征、数据用户、应用软件、分析工具、安全与隐私保护措施等相关信息。二是实现需求输入—价值输出的对接，这是构建需求功能表的关键。其具体步骤为：1)政府机构发布大数据应用案例信息提交表，利益相关者提交大数据应用需求，收集需求数据；2)将具体的应用需求进行提炼与归纳，完成需求输入；3)基于已构建的大数据系统架构中的职能组件，完成价值输出；4)将可能影响大数据应用布局的挑战与变量列举出来，总结过程参数；5)综合对接需求输入与价值输出，形成需求价值清单，可参见图4。三是对安全与隐私的关注。NIST发布了大数据安全与隐私架构和分类清单，作为NBDRA框架的补充，同时还主张探索安全与隐私分类清单与NBDRA框架之间的映射关系，这一关系的建立实则是将安全与隐私的需求落实到了具体的职能组件集合，更有利于指导解决方案的形成。

大数据既是新一代信息技术，也是一种服务业态，还改变了新的经济、社会乃至政治环境。在这一背景下，美国NIST的大数据互操作性框架为美国政府部门、商业界、学术界和用户等各个利益相关方提供了一个开展沟通和合作的重要框架，与大数据系统、资源和解决方案等有关的探讨都可以在这个框架下展开。作为在美国推动大数据发展的重要举措，NIST制定大数据互操作性框架这一行动本身，值得我国政、产、学、研等各界关注和重视。从框架的具体内容上看，美国NIST在设计过程中所采取的设计思路、所构建的概念映射关系等方面，都可以为构建具有中国特色的大数据发展指导框架提供一定参考。

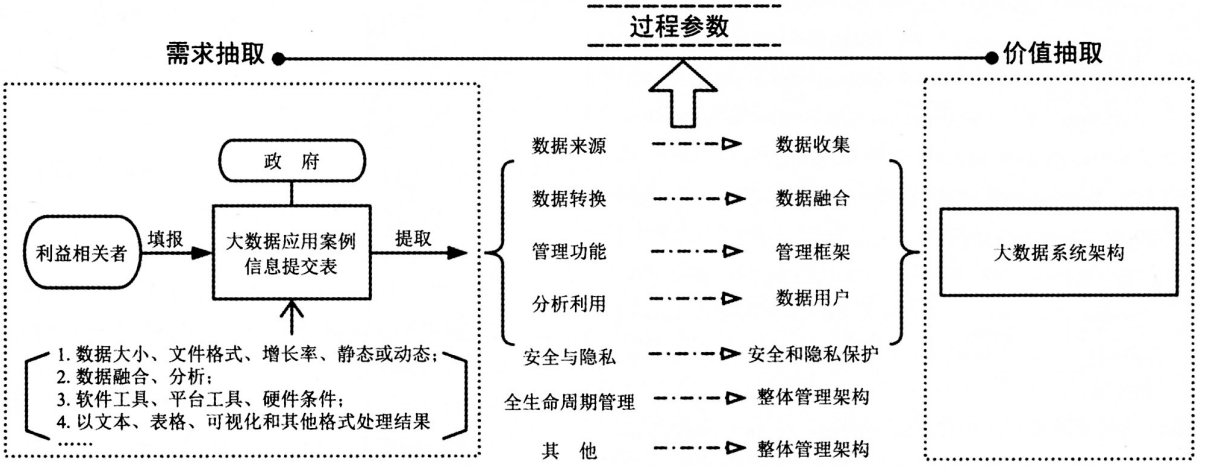

图4 大数据应用需求与价值的功能映射流程

参考文献：

[1]NIST. Big Data Information[EB/OL].https://www.nist.gov/el/cyber-physical-systems/big-data-pwg, 2019-03-05.

[2]肖筱华，周栋.大数据技术及标准发展研究[J].信息技术与标准化，2014，(4)：34-38.

[3]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.

[4]郑大庆，黄丽华，张成洪，等.大数据治理的概念及其参考架构[J].研究与发展管理，2017，29(4)：65-72.

[5]Nadal S, Herrero V, Romero O, et al. A Software Reference Architecture for Semantic-aware Big Data Systems[J].Information and Software Technology, 2017, 90.

[6]Pääkkönen P, Pakkala D. Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems[J].Big Data Research, 2015, 2(4).

[7]吴韬.习近平国家治理现代化思想的大数据观及其现实意义[J].云南行政学院学报,2018,20(5):104-109.

[8]President s Council of Advisors on Science and Technology. Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology. Washington [EB/OL].http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd-report-2010.pdf, 2019-03-14.

[9]U.S. Government[EB/OL].Data. gov: http://www.data.gov, 2019-03-14.

[10]Office of Science and Technology Policy, Executive Office of the President. Fact Sheet: Big Data Across the Federal Government[EB/OL].http://www.whitehouse.gov/administration/eop/ostp, 2019-03-14.

[11]NIST. NIST Special Publication 1500-6[EB/OL].https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500- 6.pdf, 2019-03-14.

[12]OSTP."Data to Knowledge to Action" Event Highlights Innovative Collaborations to Benefit Americans[EB/OL].https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/Data2Action%20Press%20Release.pdf, 2019-03-06.

[13]NIST. NIST Big Data Interoperability Framework: Volume 5, Big Data Architecture White Paper Survey[EB/OL].https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500- 5.pdf, 2019-03-06.

[14]NIST. NIST Big Data Interoperability Framework: Volume 1, Definitions[EB/OL].https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf, 2019-03-06.

[15]NIST. NIST Big Data Interoperability Framework: Volume 2, Big Data Taxonomies[EB/OL].https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-2.pdf, 2019-03-06.

[16]Office of the Assistant Secretary of Defense, Reference Architecture Description[EB/OL].http://dodcio.defense.gov/Portals/0/Documents/DIEA/Ref_Archi_Description_Final_v1_18Jun10.pdf, 2019-03-06.

[17]NIST. NIST Big Data Interoperability Framework: Volume 6, Reference Architecture[EB/OL].https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-6.pdf, 2019-03-06.

[18]NIST. NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements[EB/OL].https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf, 2019-03-06.

[19]国务院.促进大数据发展行动纲要[EB/OL].http://www.gov.cn/zhengce/content/2015- 09/05/content_10137.htm, 2019-03-06.

[20]孟小峰,慈祥.大数据管理:概念、技术与挑战[J].计算机研究与发展,2013,50(1):146-169.

[21]Kuhn T.科学革命的结构[M].金吾伦,胡新和,译,北京:北京大学出版社,2012:5.

[22]Chang W. NIST Big Data. Reference Architecture for. Analytics and Beyond[EB/OL].https://bigdatawg.nist.gov/Day2_15_NBDRA_Anaytics_and_Beyond_WoChang.pdf, 2019-03-06.

[23]百度百科.NIST[EB/OL].https://baike.baidu.com/item/%E7%BE%8E%E5%9B%BD%E5%9B%BD%E5%AE%B6%E6%A0%87%E5%87%86%E4%B8%8E%E6%8A%80%E6%9C%AF%E7%A0%94%E7%A9%B6%E9%99%A2/3931459? fr=aladdin&fromid=6274256&fromtitle=NIST, 2019-03-06.

Characteristics of NIST Big Data Interoperability Framework and Its Enlightenment

Zhang Bin Wang Lulu Zhang Zhen

Abstract: [Purpose/Significance] Through the analysis on the background and content of NIST Big Data Interoperability Framework, the main characteristics were summarized to provide beneficial advice for the big data development in China.[Method/Process] With the methods of content analysis and text analysis, from the aspects of data layer, frame layer, role layer, and application layer, the analysis and summary on NIST Big Data Interoperability Framework were made based on the data of public policy and research report mainly collected from the NIST official websites. [Result/Conclusion] It is found that NIST established a comparatively perfect reference framework, and emphasis on the changes in the big data paradigm and encourage the possibility of big data applications. When formulating the big data reference framework, China can pay attention to the reference value and the development needs of stakeholders under the premise of reaching a consensus on the theoretical level, and build a mapping relationship between demand and value.

Key words: Big data; government; Reference framework; Conceptual model; Stakeholder

