



专题:智慧城市

基于政府数据开放的智慧城市构建^{*}

顾磊,王艺

(中国电信股份有限公司上海研究院 上海 200122)

摘要:政府数据开放是一个新兴的领域。作为智慧城市的重要组成部分,开放数据使得更多参与者能对数据进行加工和挖掘,从而创造新的价值。政府数据开放技术涵盖了元数据管理、关联数据、数据可视化和社交网络等各类新兴信息技术,与此同时,也对城市的各种基础如数据中心、网络通信提出了新的需求,并且通过集中的使用和配置有效减少了冗余设施,提高了使用效率。通过分析现状和目标,提出了智慧城市中数据开放的需求,进一步阐释了政府数据开放的方式和核心技术,最后给出智慧城市政府数据开放的架构方案。

关键词:政府数据开放;智慧城市;元数据管理;关联数据;数据可视化;社交网络

doi: 10.3969/j.issn.1000-0801.2014.11.007

Construction of Smart City Based on Open Government Data

Gu Lei, Wang Yi

(Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China)

Abstract: Open government data(OGD) is a new evolving research field in smart city. As an important part of smart city, open government data allows new participation and collaboration in data reuse and utilization to create added value service. The key technologies of open government data include new information technologies such as metadata management, linked data, data visualization and social networks. Meanwhile, open government data propose additional infrastructure demands in smart city such as data center, network facilities and so on. In the open government data scheme, resources are utilized in a more intensive and efficient way. The state-of-the art of open government data in smart city was proposed. Moreover, the demand and technologies of open government data were introduced. Last but not least, the architecture design of open government data in smart city was illustrated.

Key words: open government data, smart city, metadata management, linked data, data visualization, social network

1 政府数据开放的现状

政府数据开放(open government data, OGD)^[1-7]是指按照用户特定的需求和一定的互联网协议、规则、框架,对Web数据进行存储和组织,所利用的数据或来自不同的数

据源,或属于不同的数据类型,最终目标是实现信息在网络空间的开放、共享与重用,以寻求信息数据最大可能的无限获取与重用。政府数据开放来源于一系列开放运动^[8],如开放源码、开放硬件、开放内容和开放通路。随着信息化发展进入移动互联网和大数据时代,数据的价值愈发凸

^{*} 国家社会科学基金重大项目“基于语义的馆藏资源深度聚合与可视化展示研究”(No.11&ZD152),国家社会科学基金一般项目“关联数据的理论与应用研究”(No.11BTQ041),国家社会科学基金青年项目“W3C的RDB2RDF标准规范在关联数据服务构建中的应用”(No.13CTQ008),上海市社会科学基金资助项目“关联数据的复用与整合在图书馆知识服务体系中的应用模型构建”(No.Y159I21031)

显,开放数据运动在全球迅速兴起。

政府数据开放的历史最早可以追溯到欧洲议会的 2003 公共部门信息开放令^[9]。2009 年,以美国率先开通政府数据服务网 data.gov 为标志,国家政府开放数据正式大规模推广,截至 2014 年全球已有 63 个国家和地区推进政府数据开放,并开放了超过 700 000 个政府数据集。通过开放政府数据促进社会转型,带动大数据产业的发展,已经成为各国的普遍共识。政府数据开放一般是门户+应用的形态,以美国 data.gov 为例,通过门户可以浏览、下载、搜索各个部门和地区的数据。数据开放的维度体现在 3 个方面:数据的数量、数据的质量和基于开放数据的应用数量。数据开放的质量根据关联数据的五星级别进行定义^[10]:基本的信息罗列属于零颗星;有数据呈现是 1 颗星;有 xls、txt 格式是 2 颗星;有 csv 数据呈现是 3 颗星;利用 URL 进行定位是 4 颗星;关联数据呈现则是 5 颗星。

图 1 挑选了具有代表性的 15 个国家,包括美洲、欧洲、亚洲、非洲的国家,统计时间为 2014 年 5 月 10-12 日。横坐标代表该国家所公布的数据集数量,纵坐标代表该国家利用数据集所开发的应用数量。

在开放政府数据计划之前,各国政府机构大多设有各自的网站,使用者为了找到所需信息,首先要了解政府机构的职能,然后再逐个寻找,要找到所需信息非常困难,而且信息支离破碎,内容不完整。此外,不同机构、网站发布的数据格式各异,使用极其不便。“一站式”政府数据下载网站则可以解决这些问题,以美国的 data.gov 为例,只要是不涉及隐私和国家安全的相关数据,均需在該网站公开发

布,并且开放标准一致的可机读数据格式和 API。

推进政府数据开放需要自上而下的顶层设计,以美国为例,开放的法律、方案和管理都由相应的主管部门负责。立法部门通过《信息自由法》、《数字化问责制和透明度法案》和《开放政府令》等保障了数据公开。美国联邦政府机关管理与预算办公室与各联邦政府机关的代表合作拟定关于开放数据、开放内容与网络应用程序编程接口的实施政策。美国预算管理办公室(OMB)负责政策施行的协调和领导,并制定遵行规范、标准的实际操作方案。data.gov 由美国联邦 CIO 委员会开发、总务管理局主管,保障了数据的安全和隐私。

我国的政府数据开放正处于起步阶段,目前北京和上海已建成了智慧城市政府数据开放平台 www.bjdata.gov.cn 和 www.datashanghai.gov.cn,并从 2012 年开始试运营。青岛、武汉、厦门也正在积极开发类似的数据平台。

2 政府数据开放的目标

数据开放将成为下一个 10 年拉动经济增长的经济引擎,研究表明^[11],综合 G20 各国的经济,开放数据将在未来 5 年中实现总额为 13 万亿美元的增长。因此,开放数据将能在未来 5 年中为 G20 国家贡献大约 1.1 个百分点的 GDP 增长,这将占到 2%GDP 增长目标的 55%。麦肯锡全球研究院^[12]分析报告则指出,开放数据在如教育、运输等领域加以应用,每年可释放约 3 万亿美元的经济潜力。因此政府数据开放作为数据开放的重要部分,蕴藏着巨大的潜力。

我国也将政府数据开放作为重要的工作规划。《国务

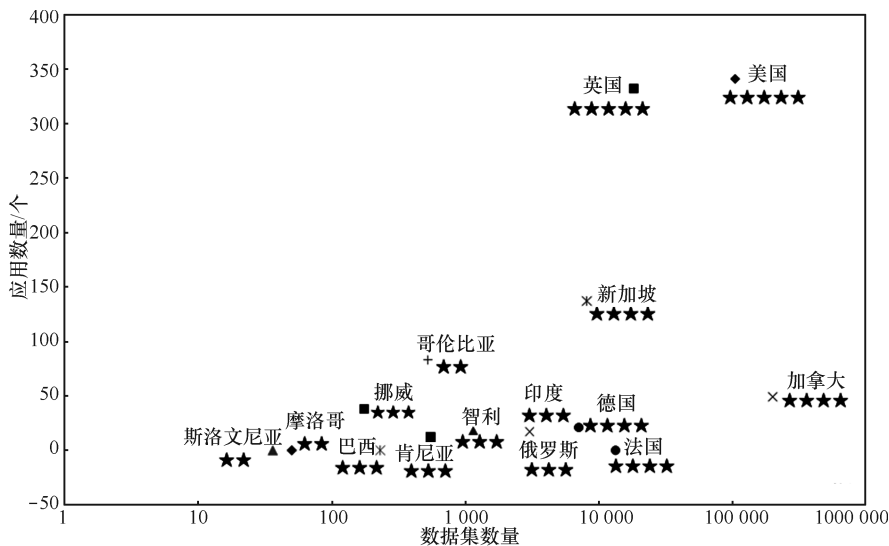


图 1 世界各国政府数据开放对比



院关于加快促进信息消费扩大内需的若干意见》指出,加快智慧城市建设,在有条件的城市开展智慧城市试点示范建设,各试点城市要出台鼓励市场化融资、信息系统服务外包、信息资源社会化开发利用等政策。《国家新型城镇化规划(2014-2020年)》也提出推进智慧城市建设,促进跨部门、跨行业、跨地区的政务信息共享和业务协同,强化信息资源社会化开发利用,推广智慧化信息应用和新型信息服务。对于政府来说,政府数据开放的主要目的体现在如下3个方面。

(1) 加强数据再利用

政府数据开放的首要目标是提供易于发现、访问和理解的数据,并使用高价值、可以机读的数据库,改善数据的利用率,进而推动政府透明化,提升行政效率。“一站式”门户需要提供各种数据及标准数据接口,方便用户下载,参与政府事务,提供反馈意见和建议,消除政府和公众之间的信息不对称问题。同时,政府数据开放门户建有数据目录中心,便于用户根据自己的需求使用数据,从而提高数据检索能力。

(2) 整合现有数据中心

政府数据开放的另一个目标是对现有数据中心进行整合。以美国为例,美国联邦政府拥有超过1100个数据中心^[2],数据中心的电力消耗也随着服务需求的提高而持续增加,因此美国联邦政府提出了数据中心整合行动,计划将政府数据中心削减到100~200个。

(3) 打造完整的数据产业链

在现代信息社会中,对信息的增值利用将成为推动财富积累和文明发展的重要途径。政府信息是一种特殊类型的信息资源,它的数据开放使政府信息得到更为广泛和有效的利用。政府拥有超过80%^[3]的社会数据,蕴含着惊人的经济潜力^[3]。通过政府数据开放,将推动城市范围内的数据沉淀、数据交易,带动信息消费,并构成一个完整的信息生态系统,第三方的信息技术企业可以对于原始开放的数据进行加工和再开放,创造新的价值并带动信息消费的蓬勃发展。

3 政府数据开放与智慧城市的关系

智慧城市是一种通过综合运用现代科学技术、整合数据资源、统筹业务应用系统,增强城市规划、建设和管理的新模式,是一种新的城市管理生态系统。智慧城市综合运用以物联网、云计算、大数据和移动互联网为代表的现代信息技术和手段,通过对城市数据资源的全面感知、全面整合、全面挖掘、全面分析、全面共享和全面协同,提高城

市管理和服务水平。城市“智慧化”的每一个领域无不是建立在数据的充分开发和利用基础之上的。智慧城市的建设和健康发展,都需要基于涉及该城市各方面的大量数据的采集、整合、组织、挖掘、分析和利用,这种对大量数据的开发和利用需求与政府数据开放的趋势是契合的。政府数据开放的基本要求是可以为社会自由获取、广大公众与政府机构可以互动、API对外开放。现有的政府数据开放平台(如data.gov)的数据采集基本都是采取分工协作、多点聚合的方式完成,对于跨部门、跨领域、跨地域的情况是可行的。如果将政府数据开放与智慧城市建设有机结合起来,既可以实现政府信息的高效公开和深层开发,又可以实现数据资源的便捷获取和对公众需求的有效引导,不仅能提高政府运转的效率和城市发展的智能化水平,而且将大大提升对市民的服务水平。

4 政府数据开放的形式

政府数据开放的主要形态是“一站式”的政府数据开放门户,该门户将各个部门数据整合到单一的入口,并提供浏览、检索、下载、用户定制等功能^{[1][3]},其主要优点如下。

通过“一站式”门户,将地理、环境、经济、民生、医疗、教育等各方面的数据以扁平化目录架构和搜索引擎的方式提供浏览和检索功能,并将分散的数据整合,从而有效减少管理和运维成本。此外,数据的发布采用可机读的数据,支持基于机器的大规模的实时数据处理,同时有利于防止欺诈与滥用。数据的浏览、检索和下载是“一站式”数据开放门户的主要功能,以北京政务数据资源网bjdata.gov.cn为例,网站提供两种方式来获取数据资源:下载数据、在线调用API。从数据类型上看,bjdata.gov.cn网站在上线初期主要提供3种类型的数据:一是带有地理坐标信息的空间数据;二是地理空间数据所对应的属性信息数据(无坐标信息),以csv格式文件提供;三是国内外、北京市出台的软件信息服务业相关政策法规文件,以WPS格式文件呈现。在bjdata.gov.cn网站中,数据按照一般性数据和地理性数据进行分类,一般性数据再根据主题和机构进行分类:主题分类包括17个一级目录、50个二级目录 and 88个三级目录;机构分类则包括市教委、市科委、市民委、市公安局等36个机构。

“一站式”门户将提供更加便捷的入口完成对政府数据的获取,打造开放、透明的政府,并提高政府的美誉度。同时,门户提供了在社会网络环境下网站信息的分享和定

制机制 RSS,以改善用户体验,用户可将网络链接共享到微信、微博、论坛等相应的网络社区,从而不仅扩展数据资源利用的用户群,也扩大了数据门户的知名度和影响力。此外,以上海市政府信息服务网 datashanghai.gov.cn 为例,门户还提供调查问卷、需求交流等内容,实现信息交流和资源打分评级机制以及用户个性化参与机制。

“一站式”门户中统一的格式和标准的可机读数据,将为网络应用服务的开发者提供更加便捷的方式,通过对数据的加工和二次开发,创造新的商机和就业机会,并满足和丰富用户的需求与体验。政府数据开放门户提供标准的 API,方便第三方机构开发专业化且有特色的应用。由于大多数政府数据与地理信息有关,因此,鼓励各企业根据自身的需要,开发大量地理信息服务的可视化应用,可以提升用户体验度。以第三方工具制作网站 http://logd.tw.rpi.edu/demo/clean_air_status_and_trends_-_ozone 为例,利用 data.gov 中的数据,开发了美国各地区臭氧浓度的可视化工具。在该应用中,企业用户点击任意地点即可查看该地点的臭氧情况,为企业碳排放及相应的收费提供参考,以做出更为准确的选址策略。

应用是政府数据开放的重要组成部分,通过开放专门的移动互联网应用调用数据 API,移动用户可以便捷地获取所需的服务,如路径导航和基于位置的商品推荐等。

5 政府数据开放的核心技术

政府数据开放依托于“一站式”门户,政府数据网站主要提供数据浏览、检索、下载,数据接口开发、应用下载、数据目录以及用户参与功能。底层的支撑技术则涵盖了元数据管理、大数据存储技术、关联数据技术^[14-16]、数据可视化技术、搜索引擎技术、社交网络等各个方面,这些技术极大地拓展了政府数据开放的边界,并引发很多革命性的应用。以英国第三方机构南安普顿大学所开发的基于位置应用^[17]为例,通过对不同地区的犯罪记录进行关联,可以展示所在城市各个街区的犯罪数量,警察可以根据历史数据进行推演,从而合理地分布警力。政府数据开放的技术中又以元数据管理^[18]、关联数据技术^[19]、数据可视化技术^[20,21]和社交网络技术^[8,21]为核心。

(1) 元数据管理

元数据是描述数据的数据,是对数据及信息资源的描述性信息。元数据被定义为:提供关于信息资源或数据的

一种结构化的数据,是对信息资源的结构化描述。其作用为:描述信息资源或数据本身的特征和属性,规定数字化信息的组织,具有定位、发现、证明、评估、选择等功能。在政府开放数据中,对于数据的定义和描述是非常关键的,现有的元数据规范包括都柏林核心 (Dublin core, DC) 元数据标准、FOAF (friend of a friend) 标准等。都柏林核心元数据是一个致力于规范 Web 资源体系结构的国际性元数据解决方案,定义了一个所有 Web 资源都应遵循的通用的核心标准,其内容较少,得到了其他相关标准的广泛支持。面向其他类型资源的元数据标准,基本上都兼容 DC 标准,并对它进行了扩展。

(2) 关联数据技术

关联数据是第一种可行的语义网表达形式,采用 RDF (resource description framework, 资源描述框架) 数据模型,利用 URI (统一资源标识符) 命名数据实体,发布和部署实例数据和类数据,从而可以通过 HTTP 揭示并获取这些数据,同时强调数据的相互关联、相互联系以及有益于人机理解的语境信息。RDF 是万维网联盟 (W3C) 提出的一组标记语言的技术标准,以便更为丰富地描述和表达网络资源的内容与结构。RDF 模型是由“主体—谓词—客体”3 部分 (即三元组) 组成的,核心就是用 URI 将 RDF 和 HTTP 链接起来,即 <http://> 是唯一的 URI 模式,也是 Web 中处理资源信息的途径。关联数据技术是把所有可公开的相关数据关联起来,并发布到统一的数据平台上供大家使用,可以采用 W3C 标准中的关联数据技术和方法实现。为了实现这个目标,需要处理的问题有:信息的来源问题、数据的格式问题、数据的相互关联问题以及数据的使用问题等。关联数据构建的基本流程是:语义描述→语义发布→语义发现→语义调节。

(3) 数据可视化技术

数据可视化技术是另一个政府数据开放的关键技术,是将数据库中每一个数据项作为单个图元元素表示,大量的数据集构成数据图像,同时将数据的各个属性值以多维数据的形式表示,可以从不同的维度观察数据。通过数据可视化技术和关联数据技术的结合,用户可以对开放数据进行数据挖掘和分析,发现蕴藏在数据之间的关联和知识。

(4) 社交网络技术

社交网络技术即通过微信、微博、交友、博客、网络社区转发和分享数据。现代的政府数据开放利用 RSS 技术



实现数据资源内容的聚合、共享、推送、订阅和发布,主要应用于各类新闻报道、服务推送、书目或其他数据的查询定制等。通过社交网络技术,这些数据可以通过 API 直接转发和分享到社交网络平台。现代的复杂网络理论表明,通过社交媒体的病毒式传播,是数据传播最有效的方式,因此社交网络技术是保障数据有效传播的重要技术。

6 智慧城市政府数据开放架构方案

城市中存在着大量数据,其运行的各个方面如公共设施、交通、教育、医疗、环境保护、经济运行等构成了信息的来源。这些数据处于割裂和沉睡的状态并分散在各类型的部门和机构单位。智慧城市作为城市发展的高级阶段,需要活化碎片化数据并充分挖掘潜在价值,建设能实现数据共享、交换和开放以及系统协同的平台。

正因为数据共享和活化的重要性,中国电信认为智慧城市应当搭建数据开放平台,将政府开放数据和其他涉及公共信息的数据进行统一的开发和运营。中国电信智慧城市政府数据开放平台总体框架如图 2 所示。平台的顶层是

公共信息服务门户,主要提供数据目录、开放浏览检索和下载服务,进一步开放统一格式的数据接口 API,基于数据的智能应用和个性化的用户交流和定制功能等。在数据层面,则提供数据申请、数据发布、数据交换和数据分析服务,这些数据服务通过底层统一的数据中心和通信服务实现,其核心的四大技术分别是元数据管理、关联数据、数据可视化和社交网络技术。

数据共享和活化的基础是研究数据开放的原则和边界,包括开放的内容、方式和对象。根据开放的不同对象,内容和方式也可能不同,而其核心是在个人隐私、数据安全、社会效率之间找到合适的均衡点。政府数据开放需要完善的数据开放安全体系和标准规范体系来保障数据开放的安全性和规范性。根据开放对象和方式的不同,政府数据开放应当分为两种模式和两种类型(2+2)。两种模式分别是针对一般公众开放和针对特定对象开放,而针对特定对象开放又包括针对本人信息开放和针对其他特定对象开放;两种类型为专题型和分类型,专题型指的是根据专用的目的构建的数据,如个人征信系统数据库数据和个人血

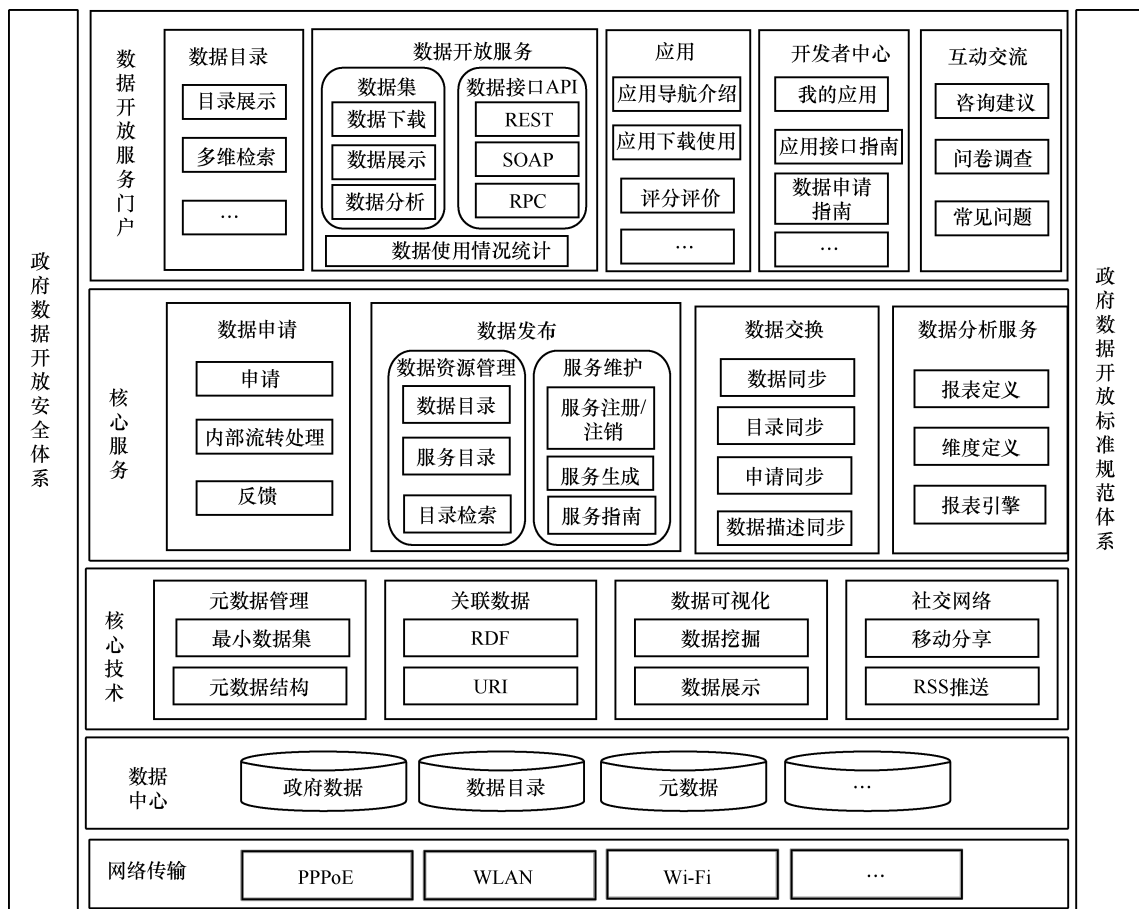


图 2 中国电信智慧城市政府数据开放平台总体框架

型匹配数据,分类型则是根据数据来源和结构的不同,按照各个职能部门和组织单位进行分类和汇总得到的数据。

数据共享和活化不仅是智慧城市建设的趋势,也是使智慧城市智慧化的关键因素,特别是智慧城市需要建立完善的信息生态系统,包含信息采集、信息汇聚、信息集成、信息加工和信息消费等各个方面,其重点是推动城市范围内的数据沉淀、数据交易,带动信息消费,并构成一个完整的信息生态系统。生态系统的底部是传统经济及其参与者,包括个人、家庭、企业等,他们是城市数据的创造者。这些来自经济个体的数据通过两类机构被搜集和汇总:传统的数据收集者,如政府相关部门、电信运营商、金融机构;新兴的数据收集者,如互联网企业。再上层则是智慧城市的核心力量——企业级别的数据分析者、新兴创业者、个人开发者和科研机构,这里是智慧闪现的地方,包含大量的创新开发团队和中小型科技公司。政府数据开放将带动整个生态体系,产业链将带来巨大的商业机会并有效促进信息消费。

参考文献

- 1 刘增明,贾一苇. 美国政府 Data.gov 和 Apps.gov 的经验与启示. 电子政务, 2011(4):90~95
- 2 付熙雯,郑磊. 政府数据开放国内研究综述. 电子政务, 2013(6):8~15
- 3 陆健英,郑磊. 美国的政府数据开放:历史、进展与启示. 电子政务, 2013(6): 26~32
- 4 徐慧娜,郑磊. 国外政府数据开放研究综述:公共管理的视角. 电子政务, 2013(6):2~7
- 5 吴曼. 开放数据在英、美政府中的应用及启示. 图书与情报, 2012(1):127~130
- 6 邓珊妮. 美国政府信息增值利用:经验、模式及启示. 图书馆理论与实践, 2011(3):42~44
- 7 周志峰,黄如花. 国外政府开放数据门户服务功能探析. 情报杂志, 2014, 32(3):144~148
- 8 Ding L, Lebo T, Erickson J S, *et al.* TWC LOGD: a portal for linked open government data ecosystems. Web Semantics: Science, Services and Agents on the World Wide Web, 2011(9): 325~333
- 9 Shadbolt N, O'Hara K, Berners-Lee T, *et al.* Linked open government data: lessons from data.gov.uk. IEEE Intelligent Systems, 2012(3): 16~24
- 10 Artigas F, Chun S A. Visual analytics for open government data. Proceedings of the 14th Annual International Conference on Digital Government Research, New York, NY, USA, 2013: 298~299
- 11 Unsworth K, Townes A. Transparency, participation, cooperation: a case study-evaluating twitter as a social media interaction tool in the US open government initiative. Proceedings of the 14th Annual International Conference on Digital Government Research, New York, NY, USA, 2013: 90~96
- 12 吴明,李占羽. 基于关联数据开放政府数据. 电脑知识与技术, 2010, 6(31): 8688~8691
- 13 袁远明,吴产乐,艾浩军. 关联开放政府数据的研究与应用进展. 电信科学, 2012, 28(9):69~73
- 14 Marienfeld F, Schieferdecker I, Lapi E, *et al.* Metadata aggregation at GovData.de - an experience report. Proceedings of WikiSym'13, San Diego, California, US, 2013
- 15 Ding L, Peristeras V, Hausenblas M. Linked open government data. IEEE Intelligent Systems, 2012, 27(3): 11~15
- 16 Graves A, Hendler J. Visualization tools for open government data. Proceedings of the 14th Annual International Conference on Digital Government Research, New York, NY, USA, 2013:136~145
- 17 侯人华,徐少同. 美国政府开放数据的管理和利用分析. 图书情报工作, 2011,55(4):119~122
- 18 Mutuku L N, Colaco J. Increasing kenyan open data consumption: a design thinking approach. Proceedings of ICEGOV'12, Albany, NY, USA, 2012
- 19 Prieto L M, Rodríguez A C, Pimiento J. Implementation framework for open data in Colombia. Proceedings of ICEGOV'12, Albany, NY, USA, 2012
- 20 Yang T M. From inter-agency information sharing to OpenData: a case study of Taiwan e-Government. Proceedings of ICEGOV'12, Albany, NY, USA, 2012
- 21 Braunschweig K, Eberius J, Thiele M, *et al.* The state of open data. Proceedings of WWW2012, Lyon, France, 2012
- 22 贾伟. 开放数据:提高绩效,释放创新潜能. 科技参考, 2013(62)

[作者简介]



顾磊,男,博士,复旦大学和台湾大学博士后出站,中国电信股份有限公司上海研究院高级工程师,主要从事物联网及智慧城市的研究工作,已主持、参与国家科研项目 5 项,发表论文 10 篇。



王艺,男,中国电信股份有限公司上海研究院高级工程师,主要从事面向政企客户的产品开发和技术研究工作,目前主要研究方向为 M2M 和物联网业务、网络适配和通用平台技术、智慧城市顶层设计、系统架构和数据活化技术等。

(收稿日期:2014-10-10)