

共治共创视角下的开放数据发展：趋势、挑战和反思

高丰
开放数据中国, 上海 200438

摘要
开放数据发展至今, 已经从原始政府透明化推手的定位逐渐演变为激发社会创新和重塑经济结构的要素, 而伴之成长的公民科技在近年来的迅猛发展, 更是突出了数据开放作为社会共治共创基础机制的地位。因此, 数据开放本身定位的演变、机制的进化和形式的转变都会影响社会各界共同参与并贡献于社会治理与创新的方式。从共治共创的视角, 反思开放数据发展的历程和趋势, 讨论其中的挑战与机遇。

关键词
共治 ; 共创 ; 开放数据

中图分类号 : C19 文献标识码 : A doi: 10.11959/j.issn.2096-0271.2016017

Open data development in the era of co-governance and co-creation: trends, challenges and reflections

GAO Feng
Open Data China, Shanghai 200438, China

Abstract
Beyond transparency, open data has been considered as new type of tool to stimulate innovation and change economic structure. The rise of civic technology also reflects that open data is fundamental to today’s co-governance and co-creation. Building upon these, it is important to recognize that the evolution of the concept, mechanism and form of open data has significant impact on how multiple-stakeholders are involved in governance and innovation. From this perspective, the path of open data development was reflected and relevant challenges and new opportunities were discussed.

Key words
co-governance, co-creation, open data

1 引言

开放数据运动一般来说,被认为起始于2009年data.gov的启动。作为政府透明化工作的重要一步,奥巴马政府大力推进政府数据的开放,以此来推动政府内外人士对政府运作过程,特别是政府财政预算与开支的监管。也因此,开放数据在发展伊始,常常被认为是政府透明化工作的一环,是政府信息公开工作的自然延伸,通过提供可实际获取并易于使用的数据而非纸面或pdf文本信息,加强公众对政府公信力的监管^[1]。

而在各国不断接纳开放数据思维,推动本地政府数据开放的同时,也不难观察到,开放数据所扮演的角色正在悄悄发生变化。2013年麦肯锡研究院的报告^[2]揭示,开放数据能够带来3~5万亿美元的经济价值,首次让人们把视线从政府透明化转向了开放数据的经济价值。透过美国纽约的BigApps大赛、英国开放数据研究院与Nesta的开放数据竞赛(open data challenge)等政府牵头的共创平台,大量基于开放数据的创新应用得以开发,并进一步得到资助、孵化,从而使“政府即平台”^[3]的理念得以真正落地,为社会提供基于数据的创新解决方案。而在中国,上海市经济和信息化委员会和交通委员会联合主办的上海市开放数据创新应用大赛也秉承着这一思路,通过开放1 TB的交通数据,激发社会创新力量,最终诞生了横跨“新能源车”、“交通保险”、“定制公交”等不同应用主题的优秀作品。

在这一系列以开放数据作为基础的共创活动中,开放数据所扮演的角色不仅仅是为共创提供了必要的数据资源,其本身也重塑了政府与公众间的交互关系,以数

据为纽带,形成了共同协作的机制,一同解决社会问题。而在这个过程中,也观察到共治共创的实际需求,正在重新定义开放数据本身的运作机制和形式。政府在思考数据开放时,从原本“我有什么就开放什么”逐步过渡到“公众需要什么我先去开放什么”,并进一步朝着“我和公众期望共同解决什么问题,我就先开放什么”的开放思路所演变。

而实际上,这场由政府率先主导推动的开放数据运动,也慢慢影响着社会各界接纳开放数据作为跨界协商合作的渠道。共享经济的典型代表企业Airbnb,就在2015年底主动对外开展有限的数据开放^[4],以期换来社会各界特别是政府对其共享经济模式的理解,并协商讨论可能的创新监管模式。无独有偶,专车模式的鼻祖Uber也采取相同的策略,主要面向政府做数据开放,以此来换取政府接纳专车模式,不再限制非经营性车辆的介入,并寻求协同治理专车的新模式。

面对这一系列的新变革,开放数据已然不再简单作为透明化的工具,更不仅仅是政府一端的事务,其正逐步成为21世纪社会协同治理和协同创新的基石,需要社会各界多元的参与及贡献。而在这个过程中,作为开放数据的实践者和研究者,也需要重新反思开放数据的定位、机制与形式,并探讨开放数据如何更好地服务于社会的共治共创。

2 政府即平台:可读可写的开放数据接口

“政府即平台”并不是一个新鲜的概念, Tim O'Reilly提出政府即平台的概念是基于其对互联网的深入理解,从互联网角度来讲,即政府应当成为社会运作的基

础技术设施,对外暴露其读写接口,允许第三方基于接口获取政府基本服务,从而加值利用,建立起更为用户导向的对外服务。

而从这一点来看,实际上可以认为政府运作开放数据本身就是一个对外提供“读”接口的过程,虽然这个接口可能有的时候仅是一个数据下载而非真正的API(应用程序编程接口)。而基于开放数据平台提供的数据,无数优秀的应用已经被开发,并为市民提供服务。最为人们熟悉的当然莫过于交通类的应用,伦敦交通局在开放其交通数据后,有超过5 000名开发者注册获取数据,有超过300多个交通应用帮助人们了解车辆的实时到站情况,超过400万人每天在上下班时使用这些应用,规划自己的出行,无形中为人们节省的时间可以换算为2 000万~6 000万英镑^[5]。

然而,对于任何一个所谓的“平台”而言,绝不会仅仅让人“读”,而不能“写”。事实上,政府在现实治理过程中有大量面向公众的服务,这些服务过程从数据角度来看即由公众向政府提供数据的过程。比如市民投诉热线12345服务,是政府接受公众对非紧急事务投诉的服务,公众投诉内容即是一种由公众向政府提供的数据,其包括了投诉内容、投诉针对地点、投诉种类等,而这样的服务过程实际也可以接口化,面向开发者提供一个“写”API,使得数据可以通过API由公众端流向政府端。这一类型的政府服务数据接口化中,最为著名的就是美国Open311接口,它是一个由民间组织OpenPlans制定并实现的接口标准,起初是为华盛顿311热线专门定制的一个API,帮助其拓展数据流入的渠道,并标准化数据流入标准。之后,这个接口得到了不同州、市政府的欢迎,甚至开始有欧洲城市也加入使用该接口标准,实现自己政府的311接口。而311接口的实现,使得政府在接受市民投诉的过程中不再局限

于电话热线,而是可以通过由市民自己选定的应用、社交媒体来获取详细的数据,不但包括更为精准的经纬度坐标,还能获得额外的多媒体信息。这些都是原本电话热线时代所不具备的。

如果说,政府将数据通过接口方式提供给公众使用是开启了数据多元化呈现的大门,那么政府提供一个“写”接口,则是大大丰富了其数据获取的渠道,双向的数据接口开放,将政府真正地变成了可读可写的平台。

3 加速“数据”到“服务”的转化:开放数据的标准化和互通性

对于开发者而言,开放数据虽然帮助他们获取了原先封闭的数据,并且确保了数据是提供在可机读格式下而非pdf或word格式下,但在面对不同地方释放的同类数据采用完全不同的数据范式时,也是非常头疼。

针对此种情况,于2015年9月正式启动的国际开放数据宪章^[6]就提出了开放数据必须保证互通性的倡议,其提出对于拥抱国际开放数据宪章的国家、地区、城市,都将努力“制定并实现统一、开放的数据标准,从而确保数据的格式、结构和通用标识符都具备互通性”^[6]。而这一倡议,实际上是对政府开展开放数据提出了更高的要求,不再是简单要求确保可机读格式,而是要求特定的数据能够满足特定的标准,从而保证其能够和其他数据发布者所发布的同类数据互通。

事实上,对于数据标准的探索,美国政府和民间机构已经尝试了很久。其中最为经典的一个案例就是美国点评网站Yelp和旧金山市政府共同合作的LIVES(local inspector value-entry specification)标

准。这一标准的诞生起源于2013年初Yelp和旧金山市开展的一项合作：由旧金山市政府向Yelp直接提供餐厅卫生检查结果数据，从而Yelp可以在其网站上展现餐厅的卫生情况，帮助消费者找到卫生健康的餐厅就餐。

这可能看上去仅是个简单的数据共享合作，但实际Yelp和旧金山市想得更远：如何让更多城市的卫生检查结果得以展示在Yelp以及其他的应用中。虽然2013年不少美国的城市已经开展了开放数据的工作，餐厅卫生检查的数据大多也已经开放在本地的开放数据门户之上，但若仔细检视各地的数据，就会发现数据的利用者（如Yelp），想要直接使用各地发布的形式各异的数据是十分困难的，因为不同地方的数据中包含的信息并不一致，例如波士顿提供的数据中包含了对餐厅违例的具体说明、状态、等级，而其他地方并不提供这一信息。而即使多个地方都提供同一信息，其在数据库中的表示也会

不同，比如同样是提供餐厅地理位置，有的地方仅提供其文字地址，有的地方却是以经纬度表示的。

为此，旧金山市的科技部门联合Yelp和纽约市的科技部门一起提出了LIVES数据标准，该标准对餐厅卫生检查结果数据应当包含的字段内容和字段要求标准化，其规定将原本复杂的餐厅卫生检查结果拆分成数张表格，包括餐厅信息的business表、检查结果的inspections表、检查违例详情的violation表。同时，也要求提供该信息的各个城市同时提供数据源说明表feed以及数值范围表legend。对于上述任意一个表内的信息，比如检查结果inspections，LIVES都基于不同州、市当前的数据采集标准和记录方式设定了标准化的字段，如图1所示。

同时，为了使得各个城市能够以最低的成本（无论从时间上还是从财政上）完成工作并快速接纳LIVES，数据的提供方式并未要求使用API，而是简单要求

Inspections

The inspections.csv file contains information about inspectors' visits to businesses. This file is required. (view sample file)

Name	Type	Required	Description
business_id	string	Yes	Unique identifier of the business for which this inspection was done
date	date	Yes	Date of the inspection in YYYYMMDD format
score	number	No*	Inspection score on a 0-100 scale. 100 is the highest score * This column must always be present in inspections.csv. However, it can be safely left blank for inspection rows that don't have an associated score (For example, some municipalities don't associate a follow-up inspection with a score) * If scores are available for a business, there must be at least one inspection with a score provided
result	string	No**	For cities that do not capture a score, this string represents the non-numeric result of the inspection. The string must be 4 characters or fewer Examples: 'Pass' or 'Fail' ** This field is required if scores are unavailable
description	string	No	Single line description containing details on the outcome of an inspection. Use of this field is only encouraged if no violations are provided
type	string	No	String representing the type of inspection Must be one of: initial, routine, followup, complaint

图1 LIVES 标准示例——inspection 表

各城市能够按照标准要求将所有数据表打包成一个zip文件,以最简单的文件替换更新形式来完成数据更新。这样做的一个好处是,操作方式简单,各地方仅需将数据按照要求制作成对应的数据表,而无需担心自己是否有能力去开发一个API,或者是否有足够人力去维护后续的接口更新。目前,已经有多达11个地方按照LIVES提供了数据,使得Yelp可以快速将卫生评价引入自身的平台。而在开放数据平台解决方案提供商Socrata的帮助下,LIVES接下来将会在一个州内的50个郡进一步推广,同期也将在英国食品标准署推广,从而在英国多达350个地方被使用。

LIVES无疑是一个成功的数据标准,而在它的发展过程中,不难看出政府作为数据发布者、企业作为数据消费者都扮演着极为重要的角色,在标准的制定、推广、采纳、使用的生命周期中充分参与和积极反馈,才最终使得LIVES能够被迅速接纳。而LIVES之外,也有其他不同类型的数据标准(见表1)通过类似的方式被制定、推广、接纳、消费,通过标准的制定,数据提供方和数据消费者之间终于说上了一样的语言。

4 清晰开放数据的目标: 为应用成效而开放

2015年底,普华永道会计师事务所受雇于开放数据研究院和Nesta公司,针对开放数据竞赛进行了影响力的评估。结果显示,在2013-2015年3年间,由7个不同主题组成的竞赛成功孵化了21家游戏的创业企业,而大赛投入的每一英镑都将收获5~10倍的回报^[7]。同时,较之经济受益更令人振奋的是,竞赛完全达到了原有对应用成效的预期,通过竞赛成功吸引了社会力量,利用开放数据,为社会紧要问题提供了可能的解决思路。

回顾英国开放数据竞赛的组织,不得不提到的是其对应用成效的注重(见表2),所谓应用成效即期望开放数据能够被运用于解决何种问题,并能够取得怎样的对应成效,这个成效应当是一个可以量化评估的具体目标。而事实上,这一思路不仅适用于竞赛的组织,更应当被实际运用于政府规划数据开放优先级的策略之中。

表1 各类开放数据标准示例

标准名	主题	说明	提出者
GTFS (general transit feed specification)	交通	标准化的公共交通时刻表及相关信息。GTFS-realtime版本则标准化交通实时信息	Google公司与波特兰交通部门提出
Open311	投诉举报	标准化的居民投诉举报性数据及标准接口	OpenPlans提出
LIVES	餐饮	标准化的餐厅卫生检查结果数据	Yelp联同旧金山和纽约市技术部门
House Facts Standard	住房	标准化的居住性用房检查结果数据	旧金山市、Accela、Code for America 等多元群体志愿性参与
Open Contracting	合约	标准化合约数据,特别是公共招标结果合约、基金会资助项目合约的数据	Open Contracting Partnership,由基金会资助形成的非官方标准联盟
BLDS (building & land development standard)	建设项目	标准化建筑项目许可和土地开发项目许可数据	Zillow、Socrata、Accela 等多家企业、机构、政府单位合作提出

表 2 英国开放数据竞赛系列主题及对应挑战问题

主题	挑战问题
治安和司法	开放数据如何能够提升群众在社会治安司法系统建设中的参与度，如何进一步为服刑人员有效改造教育提供指导以及如何降低个人犯罪率
教育	开放数据如何帮助家长为孩子的教育需要作出高效决策，包括：如何高效择校、如何选择合适专业、如何参与孩子学习过程
能源和环境	开放数据如何帮助人们制定团购能源的策略从而节省开支、如何指导面向社区的节能项目以及如何帮助人们自行生产能源
住房	开放数据如何帮助人们租借到最理想的房屋，这不仅针对在市场上租借市场房屋的人们，也包括向政府租借廉租屋的人群
食品	如何利用开放数据帮助人们吃得更健康、更为可持续化以及获知更安全的食品供应链
文遗与文化产业	开放数据如何帮助吸引更多的不同背景的人来关注文化遗产和文化产业
就业	开放数据如何帮助人们找到更好的工作或者帮助创造更多就业机会，包括：帮助人们找到合适岗位或开始自己创业、协助职业晋升或改善工作环境以及使人才市场运作更有效率

美国新墨西哥州阿尔布开克市的开放数据领导者Mark Leech近日也撰文^[8]提倡在思考该开放什么数据时，除了应当思考人们需要什么，同时也必须清晰政府期望数据释放后能够带来怎样的可能成效。以应用成效为导向思考开放数据的释放规划有以下3点好处。

- 作为数据的拥有者，政府部门可以更容易建立起数据开放和自身部门的利益联系。事实上，一直以来开放数据推进过程中如何获取相关政府部门支持是一件很令人头疼的事情。拥有数据的政府部门通常并不认为开放数据是他们需要做的事情，至少听上去这更多是某一个主管开放数据部门的主要任务，而非他们的任务。但如果能够优先思考数据的应用成效，帮助部门从自身角度思考哪些问题期望得到解决，释放数据可能找到合适的解决方案，那么这将不失为一个极大的推动力，去推动相关部门释放数据。
- 使得数据将更具备故事背景，从而和不同数据更自然地聚类在一起。现有的开放数据门户通常通过数据发布机构或者内容主题来分类数据，但事实上这样的分类方式往往会将解决同一问题的

数据再次打散，增加了数据利用者寻找数据的难度，并且也同时丢失了一个启发开发者如何将一些数据组合起来去解决一个具体问题的绝佳机会。如果能够提前思考数据的可能应用成效，然后根据应用成效将相关的数据组合归类，并提供相应的问题背景，那么数据的利用率将会大幅增加。

- 将帮助政府更好地规划数据从释放到应用再到真正落地发挥成效的路径和所需支持。数据从释放到被开发者增值利用并非是故事的结束，事实上，很多开发者并没有足够的能力去推广应用，或者使得应用真正地被使用，特别是当一个应用专门针对社会公共问题或者服务于政府部门。为了使应用能够真正落地，政府部门还需要采用诸如PPP（公私合作）模式等方法协助应用落地，发挥真正的作用，才能达成预期的成效。若相关的部门从未在数据释放前对此有足够的认识和预期，那么自然不会去思考如何帮助开发者落地应用，这也正是问题的症结所在。也因此，应当鼓励数据发布的机构，去真正思考数据的应用成效，对其有所预期，并规划可能的支持，实现相应的效果。

5 以开放为核心的基础架构：展望多元数据开放与汇流

在引言中回顾到，开放数据最初的主要定位是作为政府透明化工作的一环，作为政府信息公开工作的自然延伸而存在。而如今，开放数据则更多被视作政府大数据战略的一部分，其主要目的是释放政府数据潜力，供公众特别是专业机构和企业加以利用创造价值。在这一过程中，开放数据的定位则更多是对当前大数据市场的数据资源补充。然而，开放数据不应当被狭隘地理解为一种数据供应的手段，而更应当被视作一种社会协作治理和创新的基础体系，其提供的不仅是数据资源，更是一种新的协作模式。

实际上，若思考当前社会治理范式的革新，不难发现，当前的社会治理，因为大量新技术的应用和模式的创新，已经迈入了算法管控 (algorithmic regulation)^[9] 逐步代替人工管控，多权力中心逐步代替单权力中心的新形态。共享经济的治理问题即是一个典型的代表，无论是Airbnb还是Uber，实质都已经摒弃了原本先准入后监管的传统治理模式，共享经济平台以数据为基础，构建信誉度算法，来实时监管评价并且奖惩进入平台的经营者。而这一过程，实际上也就演变为了由共享经济平台取代政府作为监管者，对经营者做到了更为细致和实时的监管。这一根本性的变化实际反映了当前社会治理过程中，政府无法再垄断所有社会治理相关数据，而改由不同具备技术能力的企业和机构来分化这样的权力，使得原本的单权力中心演变成了多权力中心。

面对这样的变革，见证到的是在过去的几年中，不同国家的政府都试图把共享经

济平台重新纳入现有的治理体系，试图通过法令将非经营性的车辆、房屋重新排除在市场之外，从而使得监管问题重新回到既有的规则体系里。但实际这是错误锁定了问题点，问题的根本并不在非经营性车辆、房屋是否可以进入市场，而是政府如何能够像共享经济平台一样，监管好这些在既有政府监管体系之外的经营主体，而这也直指问题的症结：如何确保不同利益相关体（政府、企业、第三部门等）对已经分散的经营主体监管数据拥有访问和利用的权限，从而可以基于数据去完成必要的监管。

为了解决这一问题，政府需要做的不是试图要求完全掌握企业数据或取代企业去采集相应数据，而是应当作为协调者去搭建一个用于社会治理的数据基础服务架构，使得各方的数据得以汇流，而在这个过程中，为了保障各方的均等参与和信任，开放应当是最为核心的理念，即各不同利益相关体应当将数据开放，使得任何第三方能够访问和利用数据。

在这一过程中，一个重要的问题是如何扩展和界定哪类数据应当予以开放汇流。现有的开放数据定义往往将开放的范围限定在公共数据，即由公共资金投入所产生的数据。而实际上，当前的社会运作过程中已不仅仅是公共机构掌握着社会治理的关键数据，应当拓展和延伸原有的定义，将事关公共利益和社会基本运作的的数据纳入开放的范畴，界定哪些数据是社会运作过程中最为基础的、关键的数据，对其予以开放，并纳入数据基础服务架构的体系中。特别的，当前政府积极推进PPP模式吸引社会资本代替原有公共机构提供公共服务时，要确保开放的理念予以贯彻，确保相应的数据能够得到及时的开放，从而确保政府本身能够有相应的数据做进一步的决策监管，同时也确保公众、第三

部门等利益相关体能够在过程中对掌握的关键数据进行监督。

6 结束语

随着社会多元主体积极投身于社会治理和创新中,数据开放的进化发展将是一个迫在眉睫的问题。本文中提及的开放数据的双向接口化、数据标准化、应用成效导向规划数据开放以及数据开放作为社会治理基础体系,都是如今各国正在摸索的新趋势。作为一个正在追赶西方各国的发展中国家,中国如何在这一过程中摸索出自己的路,则是一个需要政府与企业、第三部门、公众共同努力的事情。

参考文献:

- [1] 高丰. 开放数据: 概念、现状与机遇[J]. 大数据, 2015014.
- [2] MANYIKA J, CHUI M, FARRELL D, et al. Open Data: Unlocking Innovation and Performance with Liquid Information[R].[S.l.]: Mckinsey Global Institute, 2013.
- [3] TIM O. Government as a platform[J]. Innovations Technology Governance Globalization, 2011, 6(1):13-40.
- [4] ISSAC M. Airbnb releases trove of New York City home-sharing data[N]. New York Times, 2015-12-01.
- [5] HOGGE B. Open Data: Six Stories About Impact in the UK[R]. London: Omidyar Network, 2015.
- [6] Government of Canada. International open data charter[EB/OL]. (2015-03) [2016-01-20]. <http://opendatacharter.net>.
- [7] PwC. Open Data Challenge Series Final Report[R]. London: PwC, 2015.
- [8] LEECH M. Why you should consider ROI when releasing open data[EB/OL]. [2016-01-20]. <https://sunlightfoundation.com/blog/2015/11/30/opengov-voices-why-you-should-consider-roi-when-releasing-open-data/>.
- [9] TIM O. Open data and algorithmic regulation[EB/OL]. [2016-01-20]. <http://beyondtransparency.org/chapters/part-5/open-data-and-algorithmic-regulation/>.
- [10] GAO F. Open data: concepts, status and opportunities[J]. Big Data Research, 2015014.



高丰(1986-), 英国南安普敦大学计算机博士, 现为开放数据与创新独立咨询顾问, 兼复旦大学数字与移动治理实验室特邀研究员。自2013年起, 担任英国开放知识(原英国开放知识基金会)大使, 在中国大陆地区倡导、推广、支持开放数据运动。2014年2月, 联合发起开放数据中国, 致力于打造中国开放数据生态圈。先后作为受邀专家参与开放数据指数(open data index)、开放数据晴雨表(open data barometer)、英国城市级别开放数据普查等研究和咨询项目。已与英国开放知识、开放数据合作伙伴计划、世界银行、瑞士Lift创新会议等在开放数据事务上开展过合作。

收稿日期: 2016-01-20