

# 中国开放政府农业数据分析与评价

姜恩波<sup>1,2</sup>, 李娜<sup>1,2</sup>

(1. 中国科学院 成都文献情报中心, 成都 610041;

2. 中国科学院大学 经济与管理学院图书情报与档案管理系, 北京 100190)

**摘要:** [目的/意义] 农业农村部 and 各省的农业行政主管部门是推动农业农村大数据发展的主要力量。本文通过调研全国 31 个省、自治区、直辖市的农业开放政府数据现状, 了解、分析并评价全国开放政府数据中农业数据的发布、共享、标准化和应用情况, 据此提出可行的建议。[方法/过程] 通过调研、统计全国 31 个省市、自治区、直辖市的开放政府数据平台, 对平台总体情况以及各省、自治区、直辖市农业行政主管部门所发布涉农数据的内容进行整理、规范和总结, 从数据更新时间、数据应用许可、数据下载等几个方面进行分析。同时, 结合 FAIR 规范, 从数据的可发现性、可活动性、可互操作性以及可重用性 4 个方面进行评价。[结果/结论] 目前中国农业数据的整理、发布和共享都还处于初级阶段。虽然各级政府制定了相应的法规、政策, 但是从政策落地的效果来看并不理想。农业农村部应该与各省级农业行政主管部门联合起来共同推动政府农业数据的开放与应用。同时, 积极与国际开放政府数据的建设模式接轨, 提高数据发布、管理和服务的能能力。

**关键词:** 开放政府数据; 农业大数据; 智慧农业

**中图分类号:** G250

**文献标识码:** A

**文章编号:** 1002-1248 (2020) 10-0004-12

**引用本文:** 姜恩波, 李娜. 中国开放政府农业数据分析与评价[J]. 农业图书情报学报, 2020, 32(10): 04-15.

## Analysis and Evaluation of Chinese Open Government Agricultural Data

JIANG Enbo<sup>1,2</sup>, LI Na<sup>1,2</sup>

(1. Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041; 2. University of Chinese Academy of Sciences, School of Economics and Management, Department of Library Information and Archives Management, Beijing 100049)

**Abstract:** [Purpose/Significance] The Ministry of Agriculture and Rural Affairs of China and the agricultural administrative departments of various provinces play a major role in promoting agricultural and rural big data develop-

收稿日期: 2020-06-30

基金项目: 四川省科技厅国际合作项目“面向知识服务的农业科学数据与科技文献关联融汇服务系统研究与示范”(2017HH0094)

作者简介: 姜恩波 (ORCID: 0000-0001-7890-9917), 硕士, 研究馆员, 中国科学院成都文献情报中心。李娜, 硕士, 研究方向为政策计量学

ment. This paper aims to learn, analyze and evaluate the data publishing, sharing, standardization and applications by investigating the open governmental agricultural data in 31 provinces, autonomous regions and municipalities in the mainland of China, and propose suggestions for improvement accordingly. [Method/Process] We investigated the open government platforms built by 31 provinces (autonomous regions and municipalities), sorted out the content published in these platforms, and standardized and summarized the investigation data. We made an analysis based on the update time of data, license of data applications, and download of data. Meanwhile, combined with FAIR data principles, we evaluated the data from the findability, accessibility, interoperability and reuse of the data. [Results/Conclusions] At present, the management, publishing and sharing of agricultural data in China is at an elementary stage. Although various governments formulated regulations and policies, the implementation of these regulations and policies is unsatisfactory. The Ministry of Agriculture and Rural Affairs of China should join hands with the agricultural administrative departments of various provinces to promote the opening and use of government agricultural data. Meanwhile, they should be in line with the construction mode of the international open government data and improve their capabilities in data publishing, management and services.

**Keywords:** open government data; agriculture big data; intelligent agriculture; scientific data

## 1 引言

近年来,世界主要发达国家都将数字农业作为战略重点和优先发展方向,相继出台了“大数据研究和发 展计划”“农业技术战略”和“农业发展 4.0 框架”等战略,构筑新一轮产业革命新优势<sup>[1]</sup>。中国有悠久的农业生产历史,同时也有着丰富的农业生产经验。农业是立国之本,也是社会经济稳定的重要保证力量。自 2016 年以来,中国相继出台了《农业部关于推进农业农村大数据发展的实施意见》《农业部办公厅关于公布农业农村大数据实践案例的通知》《数字乡村发展战略纲要》和《数字农业农村发展规划 2019—2025》等规划,提倡“大力发展数字农业,实施数字乡村战略,推动农业数字化转型”<sup>[2]</sup>。数字化农业,其核心是数据驱动的精准农业和智能农业建设。

农业数据的最大拥有者是中国各级政府部门。长期以来,政府部门作为公共资源的管理者,在履行职责过程中生成、采集和保存了海量的数据,是数据的最大生产者和保有者。政府数据占据整个社会数据的 80% 以上。政府数据是指产生于政府内部或虽然产生

于政府外部,但对政府活动、公共事务和普通民众有重要影响的数据资源<sup>[3]</sup>。推动政府涉农数据的开放与应用具有极大的经济价值、社会价值和政治价值。

2017 年 11 月,农业农村部发布了《农业部关于推进农业农村大数据发展的实施意见》(以下简称《意见》)。《意见》中提到“未来 5~10 年内,实现农业数据的有序共享开放,初步完成农业数据化改造”。另外,《意见》分别从农业农村部以及省级农业行政主管部门两个方面中给出了不同时间阶段的建设目标:①到 2017 年底前,农业农村部及省级农业行政主管部门数据共享的范围边界和使用方式基本明确,跨部门、跨区域数据资源共享共用格局基本形成。②到 2020 年底前,逐步实现农业农村部和省级农业行政主管部门数据集向社会开放,实现农业农村历史资料的数据化、数据采集的自动化、数据使用的智能化、数据共享的便捷化。

由此看出,中国农业大数据的汇聚、整理、发布和应用,主要是由农业农村部和各省农业主管部门来实施和推动的。二者互相融合,缺一不可。2018 年,中华人民共和国国务院发布《促进大数据发展行动纲要》(以下简称《纲要》)。《纲要》提出“大力推动

政府部门数据共享”，优先推动农业等 20 个领域的政府数据集向社会开放。在公共服务大数据工程中，《纲要》强调，“整合构建国家涉农大数据中心，推进各地区、各行业、各领域涉农数据资源的共享开放，鼓励各类市场主体充分发掘平台数据”。因此，各省的政务公开平台或者开放政府数据平台是重要的政府对外发布公共数据的窗口。

近些年，对于中国开放政府数据的研究逐渐增多。其研究对象主要以国内外开放政府政策<sup>[4-7]</sup>、开放政府数据平台建设<sup>[8]</sup>以及平台数据质量<sup>[9,10]</sup>、开放政府数据元数据<sup>[11]</sup>、隐私保护和服务质量<sup>[12]</sup>的整体情况为研究目标。本文从农业这一特定领域的开放政府数据进行调研和评价，希望能够进一步完善开放政府数据领域的研究内容。

## 2 开放政府农业数据调研及解读

为了了解各省、自治区、直辖市农业行政主管部门对于农业数据开放与应用现状，并对调研结果进行解读和分析。对中国 31 个省、自治区、直辖市的开放政府数据平台进行了调研。农业数据涉及广泛，以农业产业链来划分包括农业生产数据、农业市场数据、农业管理数据以及与农业生产密切相关的农业资源与环境数据。而这也是非常宏观的划分。仅以农业生产数据来说，就可以再划分为农（种植）、林、牧、副、渔等 5 个领域。种植业数据又包括了农机数据、农肥数据、农药数据、农膜数据等多种类型。因此，涉农数据类型繁多，体系庞大，数据来自多个行政主管部门。如果要将所有涉农数据的发布和利用情况完整调研，非政府部门很难做到。另外，虽然农业数据分类体系完整，但从目前中国开放政府数据的实际发布情况来看，类型数据存在一定困难。综上考虑，笔者选择了将调研的对象定位在各个开放政府数据平台上的省农业行政主管部门发布的数据。这种数据遴选规则虽然会使得数据范围缩小，数据量减少。但是农业行政主管部门是农业农村数据的主要出口，也是《农业部关于推进农业农村大数据发展的实施意见》中提到

的农业农村数据的建设主体之一，界限更为明晰，避免了工作内容情况不可控。

至于分析和评价模型，笔者拟通过 FAIR 规范来对数据的部分情况开展评价。2014 年，为了更好地推动科学数据的管理，学术界、产业界、科研资助机构和学术出版机构在荷兰莱顿大学共同提出了 FAIR 规范。2016 年，FAIR 规范通过 *Nature* 的 *Scientific Data* 期刊进行发布。FAIR 规范从 4 个方面概括了优质数据管理应该达到的目标，即数据的可发现（Findable）、可获取（Accessible）、可交互（Interoperable）和可重用（Reusable），如图 1 所示。4 个方面内容独立，又有所关联。FAIR 强调机器用户对于数据的获取、处理和理解。FAIR 规范一经发布，就引起了各界的广泛关注。GO FAIR 发布了“FAIR 原则阐释”，欧盟委员会发布了“将 FAIR 变为现实”的报告，研究数据联盟发布了 FAIR 数据成熟度模型。还有一些团体也形成了数据管理 FAIR 化的实施方案、实施注意事项等内容<sup>[13]</sup>。笔者对本次调研数据基本情况的整理后，通过 FAIR 规范中的一些描述性指标来解读和评价部分调研数据。

The FAIR Guiding Principles	
<b>To be Findable:</b>	
F1.	(meta)data are assigned a globally unique and persistent identifier
F2.	data are described with rich metadata (defined by R1 below)
F3.	metadata clearly and explicitly include the identifier of the data it describes
F4.	(meta)data are registered or indexed in a searchable resource
<b>To be Accessible:</b>	
A1.	(meta)data are retrievable by their identifier using a standardized communications protocol
A1.1	the protocol is open, free, and universally implementable
A1.2	the protocol allows for an authentication and authorization procedure, where necessary
A2.	metadata are accessible, even when the data are no longer available
<b>To be Interoperable:</b>	
I1.	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2.	(meta)data use vocabularies that follow FAIR principles
I3.	(meta)data include qualified references to other (meta)data
<b>To be Reusable:</b>	
R1.	meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1.	(meta)data are released with a clear and accessible data usage license
R1.2.	(meta)data are associated with detailed provenance
R1.3.	(meta)data meet domain-relevant community standards

图 1 FAIR 规范及 15 个二级指标

### 2.1 平台名称与平台管理机构

通过 3 个途径查找到各省、自治区、直辖市的开

放政府数据平台：①各省、自治区、直辖市人民政府官网；②近年来发布的开放政府数据报告、专著；③互联网上的相关指引。共找到了17个省、自治区、直辖市的19个开放政府数据平台（表1）。

17个网站的名称可以分为4类。第一类叫做“XX省（市）公共数据开放平台”。这一类包括了上海、河南、江西、山东、山西、四川、湖北、贵州8个省区市，也是最多的一类。第二类网站的名称简单明了，叫做“开放XX”。这一类有广东、福建、宁夏和浙江。第三类叫做“XX省（市）政府（政务）数据资源（服务）网”。这一类有北京、上海、江西、湖南、青海。第四类叫做“XX省（市）信息资源统一开放平台”。这一类包括福建、天津和海南省。目前上海市和海南省都是新老两个平台同时运行服务。上海市采用了两个不同的名称，而海南省则保持了一致。

另外，从网站的管理机构来看，有6个省自治区、直辖市的平台是由省办公厅主办。值得指出的是，有4个省（贵州、四川、山东、海南）的平台是由本省大数据管理机构主办的。2017年，继贵州省率先设立大数据发展管理局后，山东、海南、湖南和福建等省先

后设立大数据局。这些大数据局的职责主要是统筹规划省内大数据基础设施建设，推动“互联网+电子政务”以及政府数据开放共享利用<sup>[14]</sup>。因此，通过大数据行政管理机构管理开放政府数据平台，有利于海量政府数据的汇聚，打破数据孤岛，推动各省的大数据战略、智慧城市战略、人工智能战略的顺利实施<sup>[15]</sup>。

## 2.2 平台数据集数量和发布机构

表2列出了17个平台上的农业数据集的数量、数据发布机构以及数据最早和最新的更新时间。

17个开放政府数据平台中，有3个省的农业主管部门没有将数据在网站发布和共享。其他14个省市平台发布的农业数据中，最早的更新时间是2012年7月，最新的更新时间是2020年3月，时间跨度达到8年。

2009年，美国、英国先后推出开放政府数据网站，开启开放政府数据的先河。在中国，北京、上海两地于2012年率先建立了开放政府数据平台，开始对政府数据的开放和应用进行探索。从调研数据来看，这两个平台也是数据发布时间跨度最大的。当然，我们也

表1 平台名称以及主管（主办）机构

省、自治区、直辖市	网站名称	主管（主办）机构
上海市	上海市数据服务网、上海市公共数据开放平台	上海市人民政府办公厅、省经信委
天津市	天津市信息资源统一开放平台	天津市人民政府办公厅
广东省	开放广东	广东省人民政府办公厅
河南省	河南省公共数据开放平台	河南省人民政府办公厅
浙江省	浙江. 数据开放	浙江省人民政府办公厅
湖北省	湖北省公共数据开放平台	湖北省人民政府办公厅
山东省	山东公共数据开放网	山东省大数据局
海南省	海南省政府数据统一开放平台	海南省大数据管理局
贵州省	贵州省政府数据开放平台	贵州省大数据发展管理局
四川省	四川省公共数据开放网	四川省大数据中心
陕西省	陕西省公共数据开放平台	陕西省互联网信息办公室
湖南省	湖南政务大数据公众门户	湖南省人民政府发展研究中心
江西省	江西省政府数据开放网站	江西省发展和改革委员会
青海省	青海省人民政府政务公开	青海省人民政府信息与政务公开办公室、省电子政务办公室
宁夏回族自治区	开放宁夏	宁夏回族自治区网络安全和信息化委员会
北京市	北京市政务数据资源网	北京市经济和信息化局
福建省	开放福建 或 福建省公共信息资源统一开放平台	福建省经济信息中心

表 2 平台农业数据的基本信息

省、自治区、直辖市	最早更新时间	最新更新时间	数据集/个	数据发布机构
北京市	2012-07	2020-01	34	北京市发展和改革委员会、北京市农业农村局、北京市妇女联合会、北京市农村工作委员会
上海市	2015-10	2020-02	49	上海市农业农村委员会
天津市	2019-02	2020-01	27	市农委、市粮食和物资局、市商务局、市民族宗教委
广东省	2016-09	2020-01	23	省农业农村厅
海南省	2018-10	2019-09	7	省交通厅和统计局
河南省	2018-06	2019-11	23	省粮食局、省扶贫办、省农机局、省农业厅、省统计局
江西省	2017-09	2019-01	3	
宁夏回族自治区	2016-09	2017-06	10	水利厅、农垦局、统计局
山东省	2017-05	2020-03	1 646	由下属 16 个市提供数据
陕西省	2018-01	2018-01	8	省农业农村厅
浙江省	2019-06	2020-03	53	省农业农村厅、省统计局、省林业局、杭州海关、省粮食物资局、省生态环境厅
贵州省	2018-04	2019-12	112	整合了下属政府机构数据
四川省	2019-07	2019-12	98	整合了下属县市政府机构数据
湖北省				
湖南省				
青海省	2018-01	2019-05	6	省林业厅、省科学技术厅、省农业农村厅
福建省	2018-10	2020-01	71	省农业农村厅、省政府办公厅、海洋渔业局、省市场监管局、省统计局、省发改委、省水利厅、省粮储局

必须要看到，大多数省份在时间上跨度并不大，除去北上广之外，最大的时间跨度就是山东省。山东省可以说是后来而居上的典型。平台整合了下属 16 个市的数据，是开放政府数据平台中较为优秀的案例。

从发布的农业数据集数量上看，总体而言并不乐观。首先，各省之间的数据量相差很大。有的省份农业主管部门没有通过开放政府数据平台发布农业数据，而有的省份发布的数据集已经上千。其次，相较于客观存在的农业数据总量来说，现在各省、自治区、直辖市公开发布的数据数量几乎可以忽略不计。另外，从数据发布机构来看，大多数平台的数据是由省级农业相关主管部门发布的数据。只有山东、贵州和四川 3 个省不仅包括了省级机构发布的数据，而且还展示了下属市县的涉农数据。在理想情况下，省级开放政府数据平台应该通过数据的发布带动全省上下涉农数据的整理，形成本省农业本底数据，展现全省在农业农村建设方面的成果。

分析数据发布机构的情况，可以看出各省市区的农业农村厅（局）是开放农业数据发布的主体。除了没有发布农业数据的 3 个省之外，只有海南省的数据是通过其他厅局发布的。另外，由于涉农数据内容丰富，既包括农林牧副渔等生产数据，也包括农业市场数据、农业管理数据。因此，从各个机构发布的数据数量来看，各农业行政主管部门并不占很大优势。相当一部分数据是由其他厅局发布，其中以各省统计局参与发布最多。从数据来看，发布机构最多的福建省共有 8 个政府机构发布了涉农数据。这说明除了行政主管部门之外，还有大量农业数据还分布在其他相关的政府机构中。

### 2.3 平台农业数据的组织模式

表 3 呈现的是各个平台对农业数据的组织模式，主要包括农业信息所属的主题以及该主题在整个数据体系中的定位。在目前还没有全国性的数据组织标准

表 3 平台农业数据栏目设置和栏目名称

省、自治区、直辖市	主题数量/个	栏目级别	栏目名称
北京市	20	一级主题	农业农村
上海市	12	二级主题	“经济建设”主题下
天津市	21	一级主题	工业农业
广东省	12	二级主题	“经济建设”主题下
海南省	12	一级主题	农田水利
河南省	22	一级类别	农业
江西省	12	无农业主题	未见农业数据
宁夏回族自治区	20	一级主题	农业水利
山东省	20	一级主题	农业
陕西省	12	无独立主题	分布于公共服务、生态环保等多个主题之下
浙江省	22	一级主题	工业农业
贵州省	20	一级主题	工业农业
四川省	22	一级主题	工业农业
湖北省	12	无农业主题	省农业农村厅无数据
湖南省	15	无农业主题	无省农业主管部门数据
青海省	13	一级主题	农业水利
福建省	23	一级主题	工业农业

的情况下,从这两点也能看出农业信息在该省数据体系中的重要程度。

17个开放政府数据平台中,有3个省的平台上没有农业信息主题。另外14个省、自治区、直辖市中有11个省将农业数据放在一级主题里,使其首页可见。其次,从农业数据所属的类别名称来看,各省趋同但又有一些差异:“农业”“农业水利”“农业农村”“农田水利”“工业农业”。另外,从网站的类别设置上,各省也都有差异。总体来说分为两类:12个左右和20个左右。

数据的类别名称、数量以及整个网站的组织体系是数据开放共享标准方面的建设内容。《促进大数据发展行动纲要》中提出“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。”目前,中国国家政府数据统一开放平台还没有完全建成,各种标准还未出台。各省之间相互参照,同时根据本省的情况

对各类数据进行了规划和组织,总体来说差异不大。统一标准化的分类体系和主题名称能够为将来大批量数据的集中融合打下很好的基础,同时也为第三方机构汇聚分析数据提供便利。

## 3 数据的 FAIR 化分析

### 3.1 数据可发现性 (Findability)

FAIR 规范的第一条就是数据的可发现性。因为数据唯有被用户发现之后才会产生后续的各种可能:使用、分析和挖掘等等,也才能充分发挥出数据本身的价值。因此,起到描述和揭示数据信息的元数据也就成为可发现性的核心内容,同时也是整个 FAIR 体系的重要内容。FAIR 规范的可发现性内容包括:①为数据分配一个永久、唯一的标识符 (PUID);②用丰富的元数据对数据进行描述;③数据在可检索资源中进行注册;④元数据描述要包括其所描述数据的标识符。

通过调研,17个开放政府数据平台都发布了数据集的元数据信息。经统计,元数据字段最丰富的是广

东省，一共使用了 18 个字段来描述数据集。元数据最少的平台也使用了 9 个字段。各个平台使用较多的字段有 12 个：数据名称、数据发布机构、数据分类、数据摘要、数据发布日期、数据更新日期、数据类型、数据关键字、开放类型、更新频次、下载次数和浏览次数。

从各平台数据集元数据信息页面来看，只有为数不多的省份显示了数据集的唯一标识符。但是可以看出，这些标识符并非 FAIR 所提到的永久和全局唯一 (PUID)。例如“河南省知名农业品牌目录公示信息”的标示符为 41/100001043/00006。虽然这个标识符已经通过层次来尽量减少重复，提高唯一性。但是永久和全局唯一是指在整个互联网范围内都不会重复，同时也不会其他系统中被使用。

全局唯一且永久的对象标识符通常是由一些国际性的组织进行管理和分配。不同的领域都有相关的机构。数据的唯一标示符可以看作是数据资源的身份证，能够实现数字资源统一编号、规范管理、精确定位和透明利用，是数字时代重要的信息基础设施（数字资源唯一标识符体系系统分析）。以数字对象标示符 DOI 为例，它是一套识别数字资源的机制，包括了资源命名和识别号解析的机制。目前，全世界已有上千万个已经分配并解析的 DOI 号码，应用领域也扩展到政府部门<sup>[6]</sup>。目前中国也有 DOI 的分配与管理机构。但是其应用目前还限制在期刊论文领域。从国外的发展来看，DOI 已经不仅仅限制于期刊论文领域。Dryad、Figshare 等数据管理平台都已经为其数据集分配了 DOI 作为唯一标识符。

### 3.2 数据的可获取性 (Accessibility)

FAIR 的可获取性规范是指数据提供方应该保证在技术层面上让用户，特别是机器用户无障碍地获取数据以及元数据的内容<sup>[3]</sup>。通常来说，开放数据集需要以人可读和机器可读两种方式进行发布，并逐步以机器能够解析和处理为主。因此，FAIR 规范中，这个无障碍的可获取性包括了以下几项内容：① 获取数据的访问协议是标准的、通用的、开放和免费的；② 数据可

以是部分开放或者认证开放，但管理模式不应该成为数据获取的技术障碍；③ 数据和元数据有其不同的作用，即使数据无法访问，也能通过元数据了解到相关的信息。

对照 FAIR 指标来看，17 家开放政府数据平台都采用了 HTTP(s) 作为发布数据的协议。HTTP 协议是典型的开放、免费的互联网标准协议。基于 HTTP 协议发布数据，全互联网可见，无疑降低了访问数字资源的成本。

17 个省、自治区、直辖市的开放政府数据平台，14 个平台对用户下载数据都设置了注册登录的环节 (表 4)。平台把用户分为个人用户和机构用户两类。个人用户注册时需要填写个人的真实信息，例如手机号信息、身份证信息、电子邮箱等。机构用户需要填写机构名称、统一社会信用代码、法人代表姓名、法定代表人身份证号码等信息 (图 2)。部分省市平台把数据内容分为数据产品和数据接口。数据产品是完整的数据内容，可以认证下载。而数据接口，则是通过检索条件获取数据集的数据片段。

表 4 平台涉农数据的下载管理模式和数据接口情况

省、自治区、直辖市	数据下载管理
北京市	需要注册登录
上海市	数据产品不需要，数据接口需要实名认证
天津市	需要注册登录
广东省	需要注册登录
海南省	需要注册登录
河南省	需要注册登录
江西省	需要注册登录
宁夏回族自治区	需要注册登录
山东省	需要注册登录
陕西省	需要注册登录
浙江省	部分数据下载不需要注册
贵州省	不需要注册
四川省	需要注册登录
湖北省	需要注册登录
湖南省	需要注册登录
青海省	不需要注册登录
福建省	需要注册登录

笔者访问了美国、英国、法国、加拿大、印度和

图2 用户注册界面

澳大利亚的国家开放政府数据平台。这些平台虽然也有用户注册、登录的要求。但是对于所发布数据的下载都不需要用户注册和登录。注册用户通常是需要进行数据的提交、编辑和修改。美国开放政府数据平台使用的是 MAX.gov 的用户认证授权功能。只有联邦政府雇员和承包商通过联邦政府电子邮件进行注册。印度开放政府数据平台在用户下载数据时, 只需要告知用户类型和用途即可 (图 3)。用户姓名和电子邮件则不是必填项。

图3 印度开放政府网站数据下载界面

另外, 从调研情况来看, 17 个开放政府数据平台中有 15 个都提供了数据接口服务。北京和上海的平台还提供了数据接口的使用说明文档, 图文并茂, 操作方便。

各平台提供数据接口服务的方式比较一致, 即 API

基地址 + 数据集唯一 Token。以北京平台的数据访问接口为例 ( [http://data.beijing.gov.cn:80/cms/web/API-Interface/userApply.jsp?id=文件编号 &key=个人唯一标识码](http://data.beijing.gov.cn:80/cms/web/API-Interface/userApply.jsp?id=文件编号&key=个人唯一标识码))。文献编号就是数据集的 ID。“北京市农业产业化重点龙头企业名单”这个数据集的 XLS 版本 ID 就是 1694771829。另外, 大多数平台需要用户注册之后, 分发一个个人唯一标识码才允许构建完整的 API 链接。在 API 返回结果上, 大部分平台返回的是对应编号的完整文件。而上海市的数据 API 访问返回信息不是完整的文件, 而是由 XML 编码的数据片段。以“基因资源种质发芽数据”为例, 返回的内容就包括了种质名称、发芽率、发芽势等 20 多个字段。

总体来说, 实名制认证的数据获取模式为后续基于数据的应用开发和推广增加了障碍, 这一点从各个平台数据的浏览量与下载量的差异上就能看出。

### 3.3 数据的互操作性 (Interoperability)

FAIR 规范的核心内容之一是机器用户对数据的可理解、可操作。互操作原则是指在多个机构进行同一类型数据融合时, 机器能够“理解”不同机构对于不同数据的描述和定义, 从而能够自动完成数据集成。数据的互操作性侧重于 (元) 数据信息的“知识表示”, 也就是数据的语义描述, 旨在通过通用的机器“语言”实现对数字资源的“共同理解”。目前为止, 万维网联盟 (W3C) 提议的资源描述框架 (Resource Description Framework, RDF) <sup>[3]</sup> 应该是接受程度和应用面最为广泛的选择。

关于开放政府数据的语义描述、知识表示, 目前很少能够看到公开的资料。这些应该还属于部分政府机构内部的工作规范或操作手册。另外, 我们也基本没有看到基于各省市间的政府开放数据集成融合的服务案例。因此, 我们暂时无法对各个平台数据的互操作性进行具体评价。

### 3.4 数据的可重用性 (Reusability)

FAIR 的可重用性同样也包括了多项指标。但是从本次调研获取到的数据来看, 只有平台数据使用许可



信息在这个范畴之内。数据使用许可是用户使用数据的依据。许可方式直接影响到数据能否被再利用以及数据可利用的范围和方式。一个数据没有使用许可并不意味着“开放”，反而让数据的应用处于不确定状态。政府机构需要通过数据使用许可在法律层面上真正落实开放数据原则，使公众能够对政府数据在最大范围内进行多元化应用而不受法律的潜在阻碍和限制。目前开放数据所采用的许可类型主要包括知识共享协议（Creative Common License，简称 CC）、开放政府数据协议（Open Government License，简称 OGL）、开放数据库许可（Open Database License，简称 ODbL）以及直接进入公共领域方面的协议，如 CC0、PDDL 等。

17 家数据平台里，有 5 家平台（海南省、陕西省、湖北省、青海省、宁夏回族自治区）没有发布自己的数据使用许可。11 家平台使用不同的名称来描述数据的使用许可，如“使用条款”“法律声明”“服务协议（条款）”“平台声明”等。另有 1 家虽然有服务声明，但是缺少关于数据使用许可的内容。

(1) 基本情况。通常来说，数据使用许可包含但不限于这些内容：①对数据使用许可中所使用到的重要概念进行说明和定义，力求其含义清晰、准确；②说明用户针对数据的可能行为（如对数据的复制、署名、发布、传播、关联、包含等）及其组合的权利和义务；③说明用户对数据在商业以及非商业模式下使用的责任与义务，包括声明信息来源、后续使用数据者需要遵循的规范等；④免责条款和保留权利：特定情况下，数据提供方的责任免除以及回收数据或取消（部分）服务的权利。

11 家平台的数据许可大都包括“网站服务内容”“用户权利与义务”“隐私保护”和“免责声明”等几个部分。上面提到的内容大都能够包含进来。从各个部分的内容来看，平台之间内容互相借鉴的情况较多。这说明各省之间对开放政府数据的使用管理是趋于一致的。不足之处在于，各家平台的数据使用许可都并非法律意义上的数据使用许可，所规范的内容比较宏观。

(2) 信息获取。目前大部分平台都遵循“不受歧

视、免费获取数据”的原则。用户可以检索、浏览平台发布数据的元数据信息。但用户如果需要下载数据信息或者通过数据 API 访问数据，则需要注册和认证。而平台注册需要提供用户的真实信息。

(3) 数据传播。开放数据概念的核心是数据可以被任何人自由地使用、再利用和重复发布<sup>[17]</sup>。北京、上海、贵州等大部分省、自治区、直辖市明确在不违反国家相关法律的前提下，用户享有“自由传播与分享”的权利。只有广东省在数据使用许可中明确说明不得将从平台获取的公共信息资源进行有偿转让。

(4) 数据利用。大多数平台指出用户可以利用数据进行展现、分析、融合和挖掘，形成自己的成果。但是这些成果不能“侵犯任何第三方的合法权益”。另外，北京、广东、山东、四川、福建等省市要求用户在成果中注明数据信息的来源，并（建议）将成果上传到数据平台。贵州省则提出用户可以“不受限制的进行商业、和非商业性利用”<sup>[18]</sup>。部分省市平台列出了收费条款，即目前允许用户免费获取数据，但是平台保留对（部分）数据收取费用的权利。另外，部分平台还指出，可以根据特定的需要，为用户提供数据加工服务。

(5) 许可内容的呈现方式。目前，各平台的数据使用许可都没有以机器可读形式进行呈现，也没有与不同的数据集进行单独绑定授权。这对于机器用户使用数据形成了比较明显的障碍。

## 4 开放政府农业数据呈现的特点与建议

### 4.1 开放政府数据的政策体系还需要进一步完善

虽然近年来，国家和农业农村部相继出台了若干关于（农业）大数据、智慧农业方面的政策和规划，各省市也相继出台了不少相关的政策。2019 年，《南方都市报》大数据研究院在其《政府数据开放准备度》报告中提到，31 个省份 2014 年至今出台了 231 份提及“政府数据开放”的政策文件。其中，2019 年以来新出

台 20 余份。但是大部分省市的政策仍然是以宏观规划为主, 例如大数据产业、智慧城市、城市大脑。而政府数据开放是一件实践性、持续性非常强的政务工作, 需要多个部门纵横联合, 按照特定的标准规范去整理数据、整合数据、发布数据, 并且按照“开放为常态、不开放为例外”的原则, 这个工作将成为政府部门的日常工作。因此, 在中国顶层政策体系设计逐步完善的情况下, 特别需要各级政府机构出台一些具体的规章制度、技术标准和考核指标, 用以实实在在地将政府管理的农业大数据发布出来<sup>[9]</sup>。没有具体的制度和标准, 开放政府数据必将沦为一纸空谈, 也就无法对中国快速发展的智慧农业、精准农业提供有力的支撑和保障。

2017 年, 《贵阳市政府数据资源管理办法》由贵阳市第 54 号人民政府令发布, 于 2018 年 1 月 1 日起施行。这是中国第一部针对地方政府数据资源的管理办法。2019 年 10 月 1 日, 《上海市公共数据开放暂行办法》开始施行。《上海市公共数据开放暂行办法》在内容上比《贵阳市政府数据资源管理办法》更为完整和详细, 包括了市公共数据开放工作的职责分工、开放机制、协调机制、安全保护机制、数据质量管理与数据服务等, 较为细致地对开放政府数据的管理与应用进行了规定和规范。可以想见, 将来会有越来越多的“办法”出台, 中国开放政府数据的发展的框架也将越来越具体化、规范化和标准化。

## 4.2 农业农村部与省农业行政主管部门的纵横联合

《农业部关于推进农业农村大数据发展的实施意见》中规定, 政府农业数据的发布与服务是由农业农村部和省级农业行政主管部门共同完成。从数据归属来说, 农业数据是每个省公共数据的一种。因此, 农业行政主管部门的建设规划也需要遵循本省开放政府数据建设的整体框架。因此, 从目前情况来看, 它们基本上是在独立建设和运转。但是, 每个省的农业数据在所属部门、管理模式以及数据类型上又具有极大的同质性。另外从规划来看, 中国会通过汇聚不同领

域大数据, 形成开放政府数据的国家平台。而农业是这其中一个重要的建设领域。因此, 农业农村部 and 各省农业行政主管部门可以在开放政府农业数据的建设方面进行纵横联合, 可以在某些方面进行一体化、协同化的建设, 包括但不限于: ① 农业农村部充分发挥其行业规划与业务指导的优势和能力, 从数据建设与服务的各项标准制定入手, 例如农业数据资源目录标准、农业数据分类体系、农业数据描述标准以及农业数据交换标准等。因为这些标准的作用范围限制在农业科学与行业之内, 不会和各省的建设相冲突。因此, 各省农业数据管理部门可以结合本地开放政府数据的建设情况参照使用。② 各省农业主管部门依据统一的业务标准来建设农业大数据, 形成本省的农业大数据服务。同时, 各省可以逐步建立与农业农村部的数据通道, 为后续数据汇聚和整合打下基础。③ 在统一的环境下, 农业农村部建立的大数据和各省的农业数据可以根据业务联系建立关联和补充, 完善彼此的数据资源体系和服务能力。

## 4.3 农业数据发布数量少, 基于农业数据的应用少

各省农业数据的发布数量以及基于已发布数据形成的各种应用数量都非常少。特别是后者, 在调研过程中几乎没有看到。另外, 从数据发布机构来看, 除了极少数的省、自治区、直辖市外, 都是以省级机构为主在发布数据。2020 年 4 月, 《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》第一次把数据作为一种新型生产要素和土地、劳动力、资本、技术并列, 并提出在政务信息公开、共享的基础上扩大数据开放, 加强数字基础设施建设。开放政府数据平台无疑是重要的数字基础设施。而数据要在国民经济与社会发展过程中发挥更重要的作用, 首先就要在数据的数量上满足需求, 达到一定数量之后才能在“质”上有所改变, 真正起到支撑和服务的作用。

因此建议: ① 省级行政管理部门应该积极带动各级农业数据管理部门, 建立自己的数据目录清单, 摸清家底, 逐步将数据发布出来, 释放数据的能量; ②

农业行政管理部门应该和相关的数据管理部门进行沟通,建立协作、合作机制,让与农业相关的配套数据,如气象数据、农业进出口贸易数据等能够尽量多的予以发布出来;③积极举办面向农业领域的开放数据大赛,既宣传了政府开放数据,同时能够形成多样化的app应用。

#### 4.4 基于开放政府农业数据的服务需要提高

目前开放政府数据平台提供的服务比较单一。大部分省、自治区、直辖市除了数据的浏览、检索和API服务之外,基本就没有其他服务了。极少数平台向用户提供了数据定制的申请页面。数据的开放就是为了应用。如果没有应用,数据的开放就失去了其基本的意义。农业和农村数据其主要的应用和服务对象应该是政府决策机构、智慧农业生产与服务机构、面向农业农村的基层服务机构、相关的研究机构以及农业生产者。因此,从需求来看还是很多样和旺盛的。但是,需求方往往需要的是有一定时间跨度,并且能够系统性、持久性地提供服务。这对政府的数据发布和持续更新能力也会是极大的考验。因此建议:各级政府首先应该面向重点用户,搜集和了解它们所需要的数据内容、数据类型,积极优先发布相关数据,或者面向这些用户开展数据定制服务,使之形成优势产品和服务。另外,政府部门应该面向基层服务机构和农村生产力开展持续的培训服务,切实提高他们了解和利用开放农业数据的意识和能力,帮助他们利用数据解决农业服务和生产过程中的困难、政府部门应该树立良好的服务意识。开放政府数据平台就是政府面向各类机构和普通民众的一个重要服务场所。政府的服务应该是可以通过SERVQUAL等服务质量评估模型来考核的。服务评估可以包括数据维度(数据数量、质量等)、平台维度(可用性、易用性等)、宣传推广维度等等。而且服务的评估可以通过第三方机构来开展,使得结果更加客观、准确。

农业大数据是中国大数据建设的重要领域。开放政府数据是农业大数据的重要组成部分,也是推动中国智慧农业、数字农业的重要手段。随着近些年中国

持续推动数据在经济社会建设中发挥更大的作用,可以预见,未来开放政府农业数据的数量、质量和服务水平都将会会有一个较大的提升,也将会对中国农业发展起到越来越重要的作用。

#### 参考文献:

- [1] 农业农村部. 中央网络安全和信息化委员会办公室关于印发《数字农业农村发展规划(2019—2025年)》的通知[EB/OL].[2020-03-03]. [http://www.gov.cn/zhengce/zhengceku/2020-01/20/content\\_5470944.htm](http://www.gov.cn/zhengce/zhengceku/2020-01/20/content_5470944.htm).
- [2] 中共中央办公厅. 国务院办公厅印发《数字乡村发展战略纲要》[EB/OL]. [2020-03-03]. [http://www.gov.cn/gongbao/content/2019/content\\_5395476.htm](http://www.gov.cn/gongbao/content/2019/content_5395476.htm).
- [3] 我国政府部门数据开放的认知度研究 [EB/OL]. [2020-02-15]. [https://www.sohu.com/a/251619739\\_99983415](https://www.sohu.com/a/251619739_99983415).
- [4] 苗珍珍, 翟军, 林岩, 等. 英国政府开放教育数据的实践与启示[J]. 中国教育信息化, 2017, 2(3): 13-18.
- [5] 翟军, 李昊然, 孙小荃, 等. 美国《开放政府数据法》及实施研究[J/OL]. 情报理论与实践:1-8 [2020-07-03]. <http://kns.cnki.net/kcms/detail/11.1762.G3.20200317.1150.002.html>.
- [6] 屠健. 我国政府开放数据政策执行因素研究[J]. 图书馆研究与工作, 2020(3): 57-61.
- [7] 相丽玲, 李彦如, 陈梦婕. 中外政府数据开放运行机制的实证分析[J]. 现代情报, 2020, 40(1): 134-143.
- [8] 王萍, 周霞, 宋婧馨, 等. 国内外政府数据开放平台研究述评[J/OL]. 情报理论与实践, 1-15 [2020-07-03]. <http://kns.cnki.net/kcms/detail/11.1762.G3.20200325.1306.008.html>.
- [9] 刘博浩. 我国开放政府数据质量评价研究[D]. 河南: 郑州大学, 2019.
- [10] 王娟. 国内外政府开放数据质量研究述评[J]. 图书馆理论与实践, 2019(12): 27-31.
- [11] 徐咪咪. 我国政府开放数据的元数据标准主题研究[J]. 江苏科技信息, 2020(2): 7-9.
- [12] 顾嘉琪, 袁莉. 基于公众需求的政府数据开放服务质量提升研究[J]. 情报杂志, 2020, 39(6): 196-202.
- [13] JACOBSEN A, AZEVEDO R D M, JUTY N, et al. FAIR Principles: Interpretations and Implementation Considerations[J]. Data Intelli-

- gence, 2020, 2: 10–29.
- [14] 山东省大数据局[EB/OL]. [2020-04-08]. <https://baike.baidu.com/item/%E5%B1%B1%E4%B8%9C%E7%9C%81%E5%A4%A7%E6%95%B0%E6%8D%AE%E5%B1%80/22915313?fr=aladdin>.
- [15] 各省市大数据管理局[EB/OL]. [2020-04-08]. <http://www.cuobiezi.net/t/show/58038>.
- [16] DOI[EB/OL]. [2020-04-08]. <https://baike.baidu.com/item/doi/6776783?fr=aladdin>.
- [17] 韩宏军. 政府数据开放许可使用模式研究[J]. 图书情报导刊, 2017,2(9): 69–72.
- [18] 服务协议条款[EB/OL]. [2020-04-08]. <http://data.guizhou.gov.cn#!/serviceTerms>.
- [19] 迟玉琢. 科学数据能力研究:内涵、框架和影响因素[J]. 农业图书情报学报, 2020, 32(1): 23–29.