

我国农业元数据标准初探

满芮 王健

(中国农业科学院农业信息研究所, 北京 100081)

摘要: 我国农业元数据标准起步较晚, 但已从理论研究步入应用研究的阶段。文章讨论了两项农业元数据标准, 分析了当前面临的问题, 并提出了对策建议。指出应建立适应用户根本需求的体系结构, 逐渐完善其功能, 为实现农业数据共享奠定基础。文章就我国农业元数据标准及其体系进行了论述, 以ASTICM为例, 阐述了标准体系建立的原则及意义, 并提出了当前发展的问题及解决办法。

关键词: 农业; 元数据; 元数据标准; 科学数据; 农业科技

中图分类号: G250

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2016.02.002

Analysis of Status Quo About Metadata Standard in the Agricultural Field

MAN Rui, WANG Jian

(Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081)

Abstract: China's agricultural metadata standards a late start, but from the stage of theoretical research into applied research. This paper discusses two agricultural metadata standards, analyses the current problems, and put forward some suggestions. It pointed out that the establishment of basic needs to adapt to the user's system configuration, and gradually improve its function, only lay the foundation for the realization of agricultural data sharing. In this paper, China's agricultural system and metadata standards were discussed, with ASTICM to explain the significance of the principles and standards established and raised the issue of current developments and solutions.

Keywords: agriculture, metadata, metadata standard, scientific data, agriculture science and technology

1 引言

元数据是一个不成熟且不含有表意功能的数据。依据元数据的英文“metadata”前缀“meta-”可知, 其本质是“与……一起”^[1]。因此, 可以理解元数据的意义是一种信息的资源, 或者是得到某种信息的一类途径。它是对数据的说明, 提供的是准确理解和精确解释数据所需的信息。元数据的著录是指信息的提供、组织者即

著录者依据元数据标准中的元素对具体的信息数据施行描述、记录的过程。对于描述一个对象, 元数据标准所涵盖的数据有数据项集合、各数据项语义定义、著录规则和计算机应用时的语法规规定等^[2]。元数据标准制定的意义在于数据的应用, 使其加工后具有规整化、标准化, 可以实现开放、共享以及深一层的交流。在农业领域, 元数据是唯一一类可以对农业信息实体进行描述和标识的数据, 并且能够帮助用户发现和获取相关

作者简介: 满芮* (1985—), 女, 中国农业科学院农业信息研究所助理研究员, 研究方向: 科学数据管理与共享。王健 (1971—), 男, 中国农业科学院农业信息研究所研究员, 研究方向: 农业专业信息搜索理论与技术、信息 (包括空间信息) 查询与检索、信息需求、开放数据组织与共享等。

基金项目: 社科基金项目“科学数据用户相关性标准与使用模式实证研究”(14BTQ056)。

收稿时间: 2015年12月29日。

信息资源对象的数据^[3]。

我国是一个农业大国,拥有十分复杂而丰富的农业信息资源,具有数量极大、数据变化极快、分布极广的特点。针对农业元数据标准的研究十分重要,对制定我国农业信息元数据标准及框架、科学有效地管理我国农业信息、奠定信息职能管理有着极其重要的意义。

随着计算机科学技术的发展,尤其近年来互联网的全面展开,交叉领域的元数据标准问题已全面凸显,尤为明显的就是数据集的管理与服务。元数据标准的著录者和使用者最关注的就是如何更全面、更快速地发现、访问、获得以及使用数据。这样,数据的内容、质量等元数据资源变得十分重要。

以实现信息资源的开放、共享、管理为目标,通过标准的数据集对信息资源进行描述,是当前元数据标准的科研工作者以及使用者共同关注的问题。但由于描述的数据对象以及用户需

求不尽相同,出现了大规模的元数据标准体系。学者们认为“元数据是关于数据的数据”或者说“描述数据的数据”,这个概念广泛地存在于各个领域中对数据的描述现象。元数据标准的建设关乎信息发掘、开放、共享等问题,是最终实现合理、高效利用信息资源的坚实基础与后盾。国内外数据信息研究机构陆续开展了相关的研究。表1列举了目前国际上常用的元数据标准。这些可应用的著录对象包含了当前已存在的大部分信息类型^[4-5]。

我国元数据标准的研究起步较晚,但已初具规模,详见表2。目前,国内相对成熟的元数据标准已展开相关应用,同时各个领域也表现出了多方位的发展趋势。事实上,从应用环境来看,国际乃至国内都尚未使用统一的元数据标准。为了解决元数据的多元化以及协调发展的问题,需要搭建有效的机制从而进一步与国际接轨,统一发展元数据标准^[3]。对此,本文讨论了

表1 国际常用的元数据标准

| 名称 | 发布机构 | 应用领域 | 元素 |
|---|---|------------------------|---|
| Dublin Core | Dublin Core Metadata Initiative | 描述网络资源 | 名称、制作者、主题及关键词、说明、出版者、发行者、时间、类型、格式、标识、来源、语言、相关资源、范围、版权 |
| Categories for the Description of Work of Art | Art Information Task Force | 描述艺术品相关数字图像资源 | 作品对象、分类、方位布置、题名、形态、版本、尺寸、材质技术、制作方法、物理描述、碑铭标志、条件检查历史、保存历史、创造性、收藏历史、版权、风格、主题、上下文、展出借出历史、视频文件、文本参考、反向、编目历史、所在地 |
| Core Categories for Visual Resources | Visual Resources Assionation | 描述建筑、艺术、古玩等三维实体资源 | 类型、题名、作者、时间、身份号、文化、主题、关系、描述、来源、版权、式样、尺寸、材质、技术、所在地、风格、时期 |
| Federal Geographic Data Committee | Federal Geodata Commission | 描述地理空间数据 | 主要子集:标识、数据质量、空间数据组织、空间参照系统、实体和属性、发行、元数据参考。次要子集:引用文献、时间、联系 |
| Government Information Locator Service | The federal Government of the United States | 描述政府公用资源 | 标题、创作者、投稿者、时间、地点、语种、文摘、规范主题索引、非控主题词、空间域、时间段、联系方式、附加信息、目的、处理程序、参照、来源日期、有效性、获取条件、使用权限、进度号码、处理标识、来源控制标识、记录来源、使用语种、最后更新时间、记录检查时间 |
| Text Encoding Initiative | Association for Computing Machinery | 描述电子文本的标记定义、记录结构 | 题名、版本、长度、出版、丛书、附注、来源元素、项目过程描述、抽样、编辑、标签、参照、分类、特征体系、变化声明元素、制作信息、语言使用、文本类别、文本参数、参加者、背景描述、日期、责任说明、变化项目元素 |
| Encoded Archival Description | Library of Congress and MARC Standards Office | 描述文本文档、电子文档、可视化材料、声音记录 | 取用限制、增加、采访信息、其他可取得的形式、鉴定、档案描述、编排、书目、传记/历史、收藏历史、描述规则、EAD识别、语言资料、法律状态、附注、其他描述资料、原件位置、来源、其他检索工具、实体描述、实体技术、处理信息、相关资料、范围与内容、分别资料、单元日期、单元识别、单元题名、使用限制 |

表2 国内常用的元数据标准

| 名称 | 发布机构 | 应用领域 | 元素 |
|-------------------|----------------|---------------------|---|
| 中国科学院科学数据库核心元数据标准 | 中国科学院计算机网络信息中心 | 描述科学数据库的建设、管理、共享、服务 | 名称、URI、主题、描述、目的、类型、数据量、数据来源、提供者、贡献者、更新频率、数据集时间、语种、URL、关联数据集、数据集范围、数据志、评测报告、数据格式、技术要求、收费策略、权限声明、订购指南、访问时间、联系方式、分发信息元数据示例、元数据标准、元数据时间、元数据联系信息、元数据参考信息元数据示例、指示信息、检索点、实体、关系、学科范围、时间范围、空间范围、联系人名称、联系地址、其他联系方式、联系时间 |
| 中文元数据方案 | 国家图书馆 | 描述中文数字资源的建设、保存、共享服务 | 名称、主题、版本、摘要、内容类型、语种、内容覆盖范围、内容创建者、其他责任者、内容创建日期、出版、版权所有者、资源标识符、关联资源、数字资源制作者、数字资源制日期、数字资源制作地、权限声明、公开对象、操作许可、原始技术环境、加工处理历时、维护历史、认证指示符、基本抽象格式描述 |
| 中文元数据标准框架 | 北京大学 | 描述各类数字资源对象 | 元素名称、主要责任者、关键词、资源描述、其他责任者、日期、资源类型、资源形式、来源、语种、相关资源、时空范围 |
| 基本数字对象描述元数据标准 | 中华人民共和国科技部 | 描述各类数字对象的管理 | 名称、主要责任者、主题、描述、出版者、其他责任者、日期、类型、格式、标识符、来源、语种、关联、时空范围、权限 |

我国农业元数据的两项标准，分析了目前存在的问题，并提出对策建议。

2 农业元数据标准范例

目前，我国已经建立且相对成熟和应用较为广泛的有两项农业元数据标准。

(1) 中国农业科学院农业科技信息核心元数据标准 (ASTICM)

这是一套关于农业科技信息的元素集合，其目的是为了农业科技信息提供通用的元素和相关规范，在各个层面上用来支撑农业科技信息的整合、检索与数据交换。值得一提的是，此元数据标准已完成了数据的开放，使用者可以依照模型，结合自己所需进行定义。依据元数据的描述，所有信息可以拥有自己的“标签”^[7]，易于数据信息的管理；依据对元数据著录信息而进行检索的信息实体，使得数据信息有依据可寻，有效处理了信息难以被发掘的问题；由于数据信息涵盖了其位置、类别等，获取信息更加方便，大大提高了信息的利用率；经过著录的信息都采用兼容的元数据进行描述，解决了信息的整合与共享问题。中国农业科学院农业科技信息核心元数据标准的建立，解决了部分农业科技数据信息管理中的问题，

优化了从中央到地方各级政府、高等院校、科研单位对农业领域信息的管理；解决了数据的发现、获取、共享问题，服务更多农业领域的各级使用者；针对农业领域数据信息交叉性强、涉及范围广的现实，有效地规范了信息源不相同、格式不一致的问题。ASTICM是一个针对农业科技信息资源的元数据标准，规定了通用的元数据元素与模块。其主体包括：信息资源描述信息、分发信息、参考信息等三个主要复合性质的元素。模块和范围信息以及联系信息的辅助部分，不可单独使用，主要是供其他模块的特定元素所需时进行相关引用。

关于中国农业科学院农业科技信息核心元数据标准 (ASTICM)、中国科学院科学数据库核心元数据标准 (SDBCM) 以及都柏林元数据标准 (DC) 的关系如图3所示。

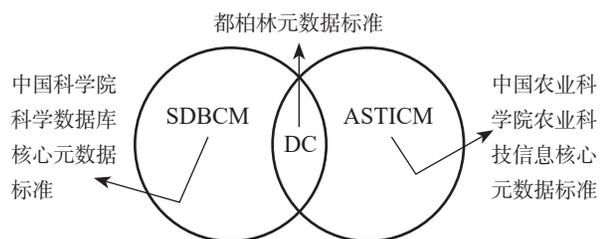


图3 都柏林标准、中科院核心元数据标准与中国农科院农业科技信息核心元数据标准关系

(2) 农业生产技术信息元数据标准(APTIM)

此标准用于描述农业生产技术信息的元素集合,为农业相关的生产技术提供描述元素与规范。其中参考了中国农业科学院农业科技信息核心元数据和都柏林核心元数据,就信息自愿的描述、分发以及元数据参考三个主要模块进行了定义,与此同时还包括范围信息和关联信息两个辅助模块^[8]。

有了元数据的描述,信息均被贴上标签,这样信息的可管理性将有显著提高,农业元数据标准的建立,解决了目前农业信息不便管理的问题;可通过对元数据检索实际的信息实体,信息的发掘变得“有理有据”,解决了信息不易发掘的问题;元数据中含有信息的位置与类型等数据,所以获取信息变得更加容易,解决了信息不易获取的问题,有效地提高了信息的利用率;经过著录的信息都具有兼容性,实现了信息的共享和在高层次上的整合,解决了信息的共享与整合问题。

3 农业领域元数据标准所面临的问题

(1) 元数据标准体系有待完善

当前,我国农业领域关于元数据标准的工作尚未全方位展开。虽然已有了上述两项元数据标准,但仍缺乏专门针对我国农业特点的、面对农业大环境发展与需求的元数据标准体系。当然这也说明农业元数据标准体系建设还有更大、更广的发展空间。农业信息数据所涉及的范围既广泛又复杂,包括气象类、生物类、环境类等,而且生态环境具有多变性,生物种类具有多样性,这一切都使得农业信息的数量极为巨大,种类繁多,内容繁杂并且具有衍生性、交叉性,为建设农业领域的元数据标准带来了巨大的挑战。但是,建立我国农业领域的元数据标准体系又势在必行。要实现多学科、跨领域的农业信息共享,所有的基本数据只有实现透明化,才能为建立元数据标准体系打好坚实的基础。

(2) 元数据标准的内容精准性不够

我国农业领域的信息资源由于地域广阔,南

北方对于同一作物的生长数据并不相同,同理,同一数据在不同的时间范围内也不尽相同,例如,新疆地区上午8点和海南地区上午8点所产生的数据资源并不可通用。元数据标准的制定关键要规定向使用者提供具有通用性的数据属性,例如:如何采集、处理、应用、存储等。可是目前以上这些属性还未规整到我国农业领域的元数据中,无法满足使用者最基本的需求。在当前已经初步形成的元数据标准中所反映的数据要素包括:数据集的标题、摘要、关键词,并且元数据的实体涵盖了信息模块,若是当成在共享层的元数据内容,也只是资源域的元素之一。对于农业领域的科研人员或者长期接触农业科学数据的工作人员是可以理解的,但对于普通使用者从理解到使用都具有难度,而事实上普通使用者才是主要的用户群。因此,我国农业领域之所以还不具有完善的元数据标准体系,其内容难以精准地集合起来是一个至关重要的问题。也许某一类小范围的科学数据目前已可以建立相关的元数据标准,但纵观大局,搭建内容涵盖广阔精准、适用性强的元数据标准体系还有一段长远而艰巨的道路要走。

(3) 元数据标准的功能不够明确

元数据标准的研究无论是国际上还是国内,都已具有初步规模,体系的建设基本都有雏形,其功能已不再局限于对信息的描述与检索上。随着各个领域对元数据标准的探求,已逐渐转为诸如评价、搜索以及数据管理、维稳、质量把握等功能。但是,功能的深入并没有运用到广大用户,我国农业领域现有的元数据标准更多服务于科研工作者,真正的用户在基层。元数据标准只有针对基层用户的功能才是至关重要,才是最有意义也是最具发展潜力的。当前的元数据标准是以农业资源发掘与应用的视角对其进行标识、使用及维护,其功能更侧重在发现数据、使用数据,并非评价与验证,这使基层用户难以把握元数据功能。

(4) 元数据标准的应用尚未达到预期

目前,元数据标准大多应用于科研领域,

用户大多是农业领域的科学工作者，但是应用很不规范。关于元数据标准目前所记录的相对完整的资源并未用到实处，著录者根据自身需要，以自我意识为出发点完成元素项，有一些数据、资源、项目处理起来相对繁琐，由于体系还没有完全建立，著录者忽略了难以处理的部分恰恰造成这些元素不完整，使得元数据标准难以发挥其全部作用。这是元数据标准尚未达到预期的根本原因之一。由于元数据存在质量不高、内容不充分、无法实现用户的全面需求等问题，现有的元数据标准只能服务于少部分人群。举个典型的例子，元数据的内容里明确规定了需要由数据志来反映数据的质量，可在元数据标准的应用中著录者往往忽略此内容，使用者的评价以及使用数据的重要参考对象就是数据的质量信息。因此，对于处理再加工过程中相对复杂的数据，使用者不应因困难而放弃，而应对此加以重视。

4 对策与建议

我国农业元数据标准的发展已经从理论性研究进入应用性研究的阶段，而在实践方面还处于起步阶段，发展前景广阔。因此，需要增加与国内外前沿机构的合作学习，深层次地拓展农业领域元数据的标准工作。元数据标准体系结构的建立会更加复杂，配合各类使用者的根本需求，功能将逐渐完善，严格规范著录流程，农业领域的元数据标准研究的深入发展将成为数据信息资源一项有序的组织工作，为实现农业数据的开放与共享奠定基础。对此，提出如下对策与建议。

4.1 健全元数据标准的内容，扩展其功能

我国农业领域元数据标准的核心问题是如何健全元数据的内容，元数据发挥的作用包括发掘、认知、管理、用户评价、加工再利用等。当前农业方面的元数据在发掘和管理两个分支发挥了一定的作用，但在用户评价、加工再利用分支上还需增强，所以需要全方位地健全元数据内容，充分扩展其功能。其相应措施涵盖以下三项内容。

(1) 针对我国农业领域元数据的属性与特征

进行系统分析。一般元数据可以向使用者呈现数据的大小、语言形式、格式等外部特征以及产权的归属等，也会呈现学科领域的情境类的信息特征。农业方面的数据涉及范围广博，具有交叉性、衍生性的特点，结合气象类、生物类、环境类信息，数据的类型不仅有实验型数据，而且有统计型数据、观测型数据。因此，在描述其属性时，需要增补描述项。例如：当描述实验产生的数据时，增补实验情境、信息背景、实验仪器与方法相关的数据。这样，使用者对综合的数据理解起来就更加直观透彻，使用时也更加准确灵活。

(2) 增补第三方元数据。农业情报以及数字图书馆领域的发展势不可挡，因此，评价类型的元数据产生是必然结果。其数据内容包括了相关领域学者对信息的专业评价、普通使用者对信息的反馈等。教育信息资源的发展恰好应用于此，使用者在应用数据资源时经常以这些评价类型的数据为基础，学名为第三方元数据或注释类元数据^[9-10]。

(3) 根据我国农业领域的现状，以农业方面的使用者为切入点，实实在在地分析需求，深层次调研相关的农业用户使用数据的行为，精确分析农业领域数据资源的使用者在查找以及评测数据的着重点^[11]，明确元数据的内容项在使用者发掘和评测时发挥的作用。只有在解决了以上问题的前提下，方可改善我国农业领域元数据标准的现状，健全其内容，发挥其功能。

4.2 针对使用者需求，注重元数据标准体系建设

为了加强我国农业领域元数据标准体系的建设，目前已完成了农业科学数据共享公共元数据标准的搭建，但这不足以满足各领域、各阶层、各类使用者的需求，应完善对农业信息类针对性强的元数据标准。对已见雏形的数据进行系统规范的分析加工，结合用户切实的需求，在原有标准的基础上扩充专用类型的元数据标准。我国农业领域对元数据标准的使用者大致可分为三类：一是农业政府决策类相关用户，他们对于元数据标准的需求在于如何结合多领域元数据制定国家

政策,规划其发展;二是农业科研工作者,这类人群对于元数据标准的理解相对透彻,结合某一深层次的科研范围,其需求也更加具有针对性、专业性;三是数量最多的广大的农业基层工作者,他们需要元数据标准来执行具体工作,其需求更为基础务实。

4.3 规整元数据标准的应用

在元数据标准的应用中有一项重要的环节即著录环节,这个过程既基础又关键,它严重影响着数据的质量,因为在未来标准建设道路上,需严格规整著录环节。首先,应将著录环节制定相关原则以及著录标准来指导相关人员开展此项工作,不能因个人意愿来决定完成质量,更不能对难以著录的数据选择避开;其次,应对元数据著录标准的工作人员提出要求,需对元数据以及元数据标准有深度的理解,尤其是对每一个元素项都要有所提炼;最后,应参照类似农业行业标准构建流程的平台或者管理系统,建立一个既实用又简练的著录平台(软件系统),用于管理、组织元数据,提供某一类别中所有的元素描述,提供获得信息的入口,容易开发,方便应用,同时专家还可以通过此平台(或者软件系统)第一时间对元数据进行评审与核查。综上所述,利用最便捷的手段规整元数据标准的应用,提高数据质量,扩充应用领域和深度,为我国农业领域元数据标准的发展发挥作用。

5 结语

对已制定的农业科技信息核心元数据标准,政府及管理部门应组织专家论证,进一步修改和发布,进行大范围的推广、试点,以促进该标准

的应用。鼓励各级政府部门、科研院校大力应用农业元数据标准,在此基础上建立专门的应用方案,加速农业科技信息元数据标准体系的建设。将农业信息元数据标准体系的研究纳入国家级科研项目,为后续研究的顺利进行有所保障,早日制定完整的农业领域各类元数据标准,提升我国农业元数据标准的管理水平以及应用范围。

参考文献

- [1] Metadata for managing scientific research data [EB/OL]. [2013-12-20].http://www.niso.org/news/events/2012/dcmi/scientific_data/.
- [2] 朱虹.农产品流通信息元数据标准化研究[J].中国标准化,2014(7):54-59.
- [3] 崔运鹏.农业科技信息核心元数据标准框架研究及其著录信息管理系统[J].中国农业科学,2007,40(4):685-692.
- [4] MICHAEL J, DAVID M. Universal meta data models [M]. U.S.A.: Wiley Publishing, Inc., 2004.
- [5] 冯项云.国外常用元数据标准比较研究[J].大学图书馆学报,2001(4):15-21.
- [6] 吴显义.我国元数据研究现状分析[J].情报科学,2003,22(1):55-58.
- [7] 钱平,苏晓鹭,崔运鹏.农业科技信息核心元数据标准的研究[J].农业网络信息,2006(2):18-21.
- [8] 毕强.元数据标准及其互操作研究[J].情报理论与实践,2007,30(5):666-670.
- [9] YOU S. Evaluative metadata in educational digital libraries: how user evaluative metadata in the process of document selection[J]. TCDL Bulletin, 2010, 4(2): 1-11.
- [10] 马蕾.数字文献元数据标准比较分析[J].情报理论与实践,2003,26(1):72-74.
- [11] 孟宪学.农业科学数据核心元数据[M].北京:中国农业科学技术出版社,2005.

(上接第6页)

势[J].中国图书馆学报,2008,(1):65-69.

[7] 张士男,宋文.《科图法》SKOS描述方案设计[J].现代图书情报技术,2010(6):7-11.

[8] 王刘安,常春.《英文超级科技词表》编制中概念优选词的选择方法研究[J].图书情报工作,2014(13):

107-112.

[9] 王莉,梁冰,白海燕.以数据空间理念建立关系发现应用:NSTL智能检索平台的实践[J].数字图书馆论坛,2014(6):48-54.

[10] 章成志.自动标引研究的回顾与展望[J].现代图书情报技术,2007(11):33-39.