科技人才元数据标准框架研究*

□ 王运红 彭洁 李大玲 吴晓莉 / 中国科学技术信息研究所 北京 100038

摘要：为了促进科技人才信息资源的描述规范化、科技人才信息共享、交换和利用，需要研制科技人才元数据标准。文章阐述了科技人才元数据标准研制的目的和意义，对科技人才元数据描述的实体进行定义，研究了科技人才元数据的特点，进而确立了科技人才元数据标准的三层框架，设计了科技人才元数据的核心元素和扩展元素，以及扩展规则。期望通过科技人才元数据标准的研制与推广，使当前结构各异的科技人才元数据更加规范，为科技人才的信息利用和数据共享提供标准化的支撑。

关键词：科技人才，元数据，元数据标准，元数据标准框架，元数据元素
DOI：10.3772/j.issn.1673—2286.2013.08.009

1 引言

当今世界已进入知识化和信息化时代，世界范围内的综合国力竞争变为人才的竞争，科技人才的重要性日渐凸显。2010年以来，随着中央人才工作会议的召开和《国家中长期人才发展规划纲要（2010-2020年）》（以下简称《人才规划纲要》）的颁布实施，党和国家对科技人才的重视提到了前所未有的高度，科技人才工作全面推进。近年来，我国科技人才数量增长迅速，2010年全国R&D全时人员总量已达229.13万人，科技人力资源总量达到318.37万人[1]。《人才规划纲要》中提出“到2020年，研发人员总量达到380万人年，高层次创新型科技人才总量达到4万人左右”[2]。面对海量的科技人才信息，各级政府部门、高等院校、科研院所、科技企业和中介机构都已开展科技人才信息化、标准化建设工作。

“科技人才”是我国独有的概念，主要体现在政策文件之中。科技人才相关的定义包括科技人力资源、R&D人员、科学家、工程师、科技人员、科技工作者以及科学工作者等。根据国家教委1982年对全国人才预测时的定义，科技人才主要指获得中专以上正规学历人员以及获得技术员及技术员以上专业技术人员和经营管理人员[3]。R&D人员主要是OECD在《研究与发展统计》和《主要科学技术指标》中使用，采用《弗拉斯卡帝手册》中的定义，指所有被雇用直接从事R&D的人员，还包括那些提供直接服务①的人员，如R&D管理人员、行政人员和文职人员等。按职业分为研究人员、技术人员和同等人员、其他辅助人员三大类，间接服务人员②则不包含在内[4]。在本文研究中，科技人才元数据描述的实体对象——“科技人才”的概念与OECD定义的“R&D人员”基本相当。

科技人才元数据标准是科技人才信息资源组织、描述、管理、长期保存和保护的基础，是实现国内各类科技人才信息系统数据库之间数据共建共享与互操作的重要前提。由于目前科技人才信息数据库由不同的单位设计和维护，其元数据结构及元素定义存在很大的差异，阻碍了科技人才信息数据库之间的共建和共享。因此，建立一套适合中国特色的科技人才元数据标准，有助于各类科技人才信息资源的描述、管理、长久保存，规范科技人才信息数据库的建设和管理，对促进科技人才数据共享，满足科技人才数据处理、交换、存储、维护和信息发布，提高科技人才信息服务能力和水平，具有重要的意义。

2 国内外研究情况概述

国外出现了一些服务与科技

*基金资助：中国高层次科技人才数据库建设（ZD2012-7-5）；科技人才信息宏观监测机制研究（2009GXS4K047）。
①直接服务人员：为为R&D活动提供技术服务、行政管理等工作人员，属于R&D人员的范畴。
②间接服务人员：为R&D活动提供间接服务的人员，如餐饮服务、健康理疗等，不属于R&D人员的范畴。

人才的元数据标准，互联网上关于人的元数据互操作是FOAF（The Friend of a Friend）项目。朋友的朋友（The Friend of a Friend，FOAF）作为一个“实验性的链接信息工程”项目，由Dan Brickley和Libby Miller于2000年年初创建[5]。但是该元数据方案主要是对社会网络中人的简单描述，对于描述科技人才缺乏整体和全面性，不能满足科技人才丰富的信息描述需求。

国外政府和研究机构的科技人才和专家库建设很多，如加拿大的SCIENCE.CA数据库、以色列的生物技术科学家数据库、德国“德国学者组织”建立的在美国德国人才网络、印度政府的海外专家人才数据库等；针对学术的专家学者库如Thomson ISI的专家学者数据库，作为人才中介的商业公司建设的人才资源库多不胜数。部分国家开始尝试建立全国性、统一的科技人才库，如巴西尝试建立科技人才统一电子履历表。因为各国科技人才数据库建设目标各不相同，即使在同一国家内，各数据库的数据内容、详略程度、规范性等也各不相同，目前还没有发现国外有统一的科技人才描述标准和规范。

我国针对科技人才元数据的研究工作尚处在摸索和实践阶段。通过对我国科研管理部门和高校、研究机构所建设的高层次科技人才数据库情况调研获知，科技部依托国家科技计划建立了各类评审专家库和课题人员信息库，中组部、人事部、教育部、中科院、社科院和自然科学基金委员会等一级预算部门也都建立了高级专家库、“百千万工程”人员库、留学归国人员库、基金项目评审库等一批专业人才信息库。各地市的科协、科委都开始建立相关的科技人才库。另外还有一些面向招聘的网站、猎头公司创建的人才招聘数据库。但这些人才数据多为内部资源，没有形成统一的科技人才数据描述规范和标准，各单位建立的科技人才数据的描述及属性缺乏规范，标识规范各异，直接影响了科技人才信息资源的开发、整合和利用。

国内的科技人才数据的相关标准有中华人民共和国水利行业标准《人才管理数据库表结构及标识符标准（SL453-2009）》[6]。该标准是用于规范管理水利人才、实现水利人才信息资源共享的技术标准。该标准主要是规范水利人才的数据库表结构和分类标识符，在国内和其他行业未有使用和有效地推广。目前国内尚没有科技人才元数据相关的国家标准和其他标准。

针对以上情况，有必要构建较通用的适合我国科技人才信息化建设的元数据标准。我们在工作中承担国家公益基金研究项目——“中国高层次科技人才数据库建设”和国家软科学项目——“科技人才信息宏观监测机制研究”，开始了对科技人才元数据标准的研究，并进行示范应用。在研究科技人才元数据标准的过程中，主要参考了中国科学技术信息研究所的《中国高层次科技人才信息数据库建设标准及分类编码标准》，中华人民共和国水利行业标准《人才管理数据库表结构及标识符标准（SL453-2009）》，加拿大的SCIENCE.CA数据库，Thomson ISI的专家学者数据库。

3 科技人才元数据标准框架研究

3.1 科技人才信息的特点

本文将科技人才信息分为两类：一是科技人才基础信息，二是与科技人才管理相关信息。

科技人才基础信息包括：（1）科技人才基本信息，表征其自然属性和社会属性的一些信息，如姓名、性别、学历、学位、毕业院校、工作单位、技术职称、党派、职务以及在学术团体任职情况等；（2）科技产出信息，这是科技人才较为特殊的信息，但表征了科技人才的特征，主要包括著作、论文、专利、著作权、标准等，科技成果完成和获得科技奖励情况，参与的科研项目等；（3）其他相关信息，包括政治素质、学术道德及科研信用信息等[7]。

在科技人才基本信息和科技产出信息中，科技人才的专业、从事研究的领域和方向、科研活动和产出（著作、论文、专利、成果、项目、会议等）、科技奖励和科技荣誉、科研诚信道德等是科技人才不能或缺的信息，表征了其作为科研工作者的独有属性。

科技人才管理相关信息，包括科技人才的人事信息、人才统计等。这类科技人才信息目前主要由相关的科技管理部门拥有，但比较分散，有助于了解整个国家、区域、行业、领域科技人才的分布、规模、结构等，为科技人才宏观监测、安全预警、人才政策制定提供决策支持[8]。

本文研究的科技人才元数据是围绕科技人才基础信息来设计的描述性元数据，是用于描述或标

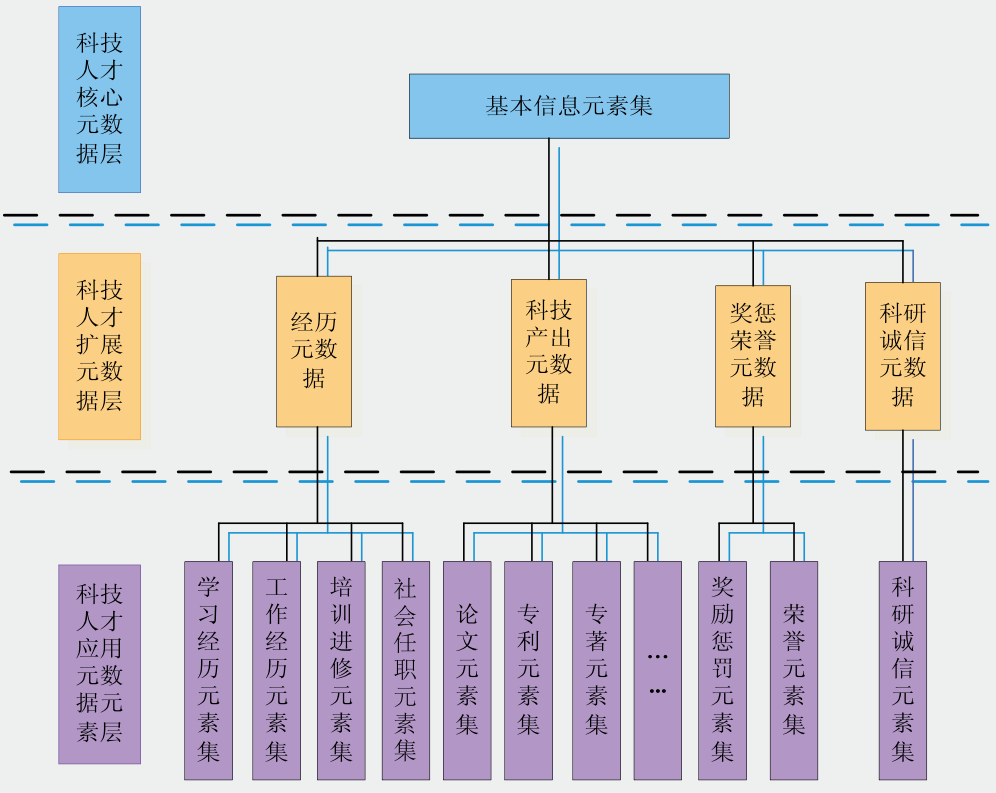

图1 科技人才元数据三层架构图[9]

识数字对象内容的元数据。科技人才管理相关信息不作为本文的研究范围。

3.2 科技人才元数据标准框架设计

科技人才元数据标准框架作为规范科技人才资源所用的元数据标准，需要遵照特定的规则和方法，它是抽象化的元数据结构。在科技人才元数据标准框架中，包含了三层：第一层为科技人才核心元数据；第二层为扩展元数据；第三层为应用领域的专用元数据，这是在第一、二层元数据基础上的扩展，如图1所示。

在这个框架中，为满足在不同系统中的需求和应用，可以扩展和修改某一个元数据集的元素。对某一个扩展元素集合的修改和补充，不会影响其他的元数据集合。在每个集合下的子集元数据标准基于它所属的集合模式进行扩展。一个子集标准只能唯一从属于一种集合，而该集合模式可以扩展出多个专用的子集标准，如经历元数据，可以扩展出工作经历、教育经历、培训进修经历、社会任职等。科技产出元数据可以扩展出论文、专利、项目、成果等。

4 科技人才元数据设计

4.1 科技人才核心元数据设计

在设计科技人才的元数据核心元素时，参考了都柏林核心元数据集和其他元数据集的设计策略[10-16]，科技人才元数据核心元素如表1所示。

在本标准框架中，元素名为英文（原文），并全部小写，以便于计算机标记和编码，并保证与其他语种的应用保持语义一致性；标签为中文，便于人们阅读。标签只是元素名称的一个语义属性，在具体的应用领域，为突出科技人才资源的个性和元数据的专指性，更好地体现该元素名称在具体应用中的语义，允许赋予其适合的标签，但语义上与原始定义不允许有冲突，不允许扩大原始的语义。

为促进不同领域和机构数据的互操作，很多元数据元素建议其元素的值取自受控词表。同样，为了某些特定领域内的互操作性，也可以开发利用其他受控词表。

表1 科技人才核心元数据元素描述

元素名	标 签	定 义	注 释
researcherID	学术ID	赋予科技人才的唯一标识符	指科技人才的唯一身份标识，是对科技人才重名的解决机制。建议采用符合规范标识体系的字符串进行标识，并使用全局统一的编码规则。
name	姓名	科技人才的姓名全称	科技人才的姓名全称，可以是中文，也可以是外文姓名。描述可以包括但不限于以下内容：姓名、曾用名、外文名，或者关于科技人才的多姓名描述。
gender	性别	科技人才的性别属性	采用国标编码体系：《人的性别代码》（GB/T 2261.1-2003）
ID	身份证号	身份证号或者护照号码	科技人才的有效身份证件号码
birthdate	出生日期	科技人才出生日期	日期可以用来表达任何级别粒度的时间信息。建议采用一个编码体系，例如ISO 8601 [W3CDTF]）的W3CDTF。信息不全的日期建议采集到年份。
nationality	国籍	科技人才的国籍	指科技人才持有护照的国家，科技人才为该国的合法公民。可以包括但不限于以下内容：目前国籍、曾经国籍，以及科技人才的双重国籍身份描述。建议采用《世界各国和地区名称代码》（GB/T 2659-2000）
researchFields	研究方向	研究方向	包括描述科技人才学习或者从事的研究方向。建议参照国家标准：《学科分类与代码》（GB/T 13745-2008）。
degree	最高学位	获得的最高学位	科技人才获得的最高学位，采用国标编码体系：《中华人民共和国学位代码》（GB/T 6864-2003）。
educationalLevel	最高学历	获得的最高学历	科技人才获得的最高学历，采用国标编码体系：《学历代码》（GB/T 4658-2006）。
professionalTitle	专业技术职称	科技人才目前获得的专业技术职称	指经国务院人事主管部门授权的部门、行业或中央企业、省级专业技术职称评审机构评审的工程系列专业技术职称。采用标准《专业技术职务代码》（GB/T 8561-2001）。
contact	联系方式	科技人才的联系方式	科技人才的当前有效联系方式，包括通讯地址、邮政编码、电话、传真和电子信箱。
dataSource	数据来源	科技人才信息的起始来源	科技人才信息的原始数据来源，以便数据交换、信息共享和即时更新。

4.2 科技人才扩展元数据设计

在科技人才的核心元数据集的基础上，可以设计科技人才信息应用的扩展元数据（见表2），可以根据实际业务，对扩展元数据进行补充。扩展元数据为科技人才元数据结构中的第二层，应用元数据元素集为第三层，形成以父、子、孙节点的元数据集合架构的树状结构。元数据的元素标准及参考设计读者可以参考《科技人才元数据（工作组讨论稿）》。

表2 科技人才扩展元数据描述

元素名	标　签	定　义	注　释
workExperience	工作经历	科技人才的工作经历	描述科技人才的工作经历，包括但不限于以下内容：任职机构、开始时间、结束时间、职务、工作描述（含海外）。
parttimeJob	社会（学术）职务	科技人才在学术组织、团体等的任职情况	描述科技人才在重要的学术机构或团体的任职情况。包括但不限于以下内容：国家、组织/团体名称、开始时间、结束时间、职务、工作描述。
education	教育经历	指科技人才的教育经历	描述科技人才获得学历学位的教育经历，建议以大专以上为教育起点。包括但不限于以下内容：就读院校、开始时间、结束时间、学习专业或者内容、获得证书或者资质。（含海外）
training	培训进修经历	指科技人才的进修经历	描述科技人才未获得学历学位的教育经历，包括但不限于以下内容：就读院校、开始时间、结束时间、学习专业或者内容、获得证书或者资质。（含海外）
specialty	专长	指科技人才的技术等专长	描述科技人才工作中的技术等专长，包括但不限于以下内容：专业技术、语言、水平等级。
rewardPunish	奖惩信息	科技人才获得的奖励信息	包括但不限于：奖励名称、授奖单位、等级、时间。可参考《奖励、纪律处分信息分类与代码 第2部分:荣誉称号和荣誉奖章代码》（GB/T 8563.2-2005）或自建词表。
honor	荣誉信息	指科技人才获得的荣誉信息	包括但不限于：荣誉名称、授予单位、时间。可参考《奖励、纪律处分信息分类与代码 第2部分:荣誉称号和荣誉奖章代码》（GB/T 8563.2-2005）或自建词表。
project	参与课题项目	科技人才参与或主持的项目信息	包括但不限于：项目名称、级别、分类、参与角色。
book	出版专著	科技人才出版的专著或者编著图书信息	建议采用国标《普通图书著录规则》（GB/T 3792.2-2006）进行描述。
paper	发表论文	科技人才在国内刊物发表文献信息	指在国内学术期刊上发表的文献信息，建议采用《信息与文献 都柏林核心元数据元素集》（GB/T 25100-2010），并可进行扩展。
patent	发明专利	科技人才发明专利情况	描述科技人才发明的专利，包括中国专利局获得的专利。中国专利项目可参照《专利文献著录项目标准（试行）》（ZC 0009-2006）。 描述科技人才在海外获得审批的专利，包括欧专局和美专局等机构获得的专利。可参照WIPO ST.36标准，或者美专局的标准数据源。
researchIntegrity	科研诚信	科技人才在科研活动和学术研究中的诚信记录	描述科技人才在科学研究、项目申报、学术活动中的诚信情况记录，主要记录学术道德上发生的不端行为和结果，为科技人才在科研活动中的管理和其他工作提供参考。

4.3 科技人才元数据扩展规则

业务应用的多样性意味着通用元数据难以适应所有的需求，为满足特殊用户的业务需求，扩展元数据是必须的。但在元数据扩展中需要遵循统一的扩展规则，这样既能满足个性化需求，又能使扩展后的数据进行共享和集成。

对于扩展的每一个元数据元素，应定义其名称、标识符、数据类型、定义、约束/条件等。扩展规则简要说明如下：

- 扩展的元数据元素不应用来改变现有元数据元素的名称、定义或数据类型。
- 扩展的元数据可以定义为实体，可以包含扩展的和现有的元数据元素，作为其组成部分。
- 允许对现有元数据元素施加更加严格的约束/条件（如：可选的元数据元素在扩展后可以是必选的）。
- 允许对元数据元素的域施加更严格的限制（如：域为“字符型”的元数据元素，在专用标准中可以限定为适当值的列表）。
- 允许对认可的域值的使用加以限制（如：现有元数据元素的域值有五个值，在扩展后可以规定它的域只包含其中三个值，要求用户从这三个域值中选择一个）。
- 允许对代码表中值的数目进行扩展。

5 总结

科技人才的信息化和数据共享工作面临着多种数据标准共存的难题，科技人才元数据不规范和没有统一标准已经制约了科技人才信息的有效利用和共享。利用科技人才元数据的标准化来统一管理分散的数据资源，并通过网络实现数据共享与服务是解决这个问题的有效途径。

本文在介绍了科技人才元数据标准研究的国内外情况基础上，结合国内外的研究现状和实际工作的需求，阐述科技人才元数据标准研制的重要意义。同时研究了科技人才元数据描述实体对象，并分析科技人才元数据的特点，进而确立科技人才元数据标准的三层架构，并设计了科技人才元数据的核心元素和扩展元素，制定扩展规则。

本研究成果已经获得科技人才元数据标准的国家标准项目资助，并已应用于在中国科学技术信息研究所的高层次科技人才数据库建设、科技人才交流开发服务中心的创新型科技人才申报评审系统、国家科技计划专家库、重庆生产力促进中心建设的西南地区科技专家数据库中。

今后本研究将继续深入开展，对科技人才元数据的核心元素和扩展元素进行改进和完善，尤其是面向创新型科技人才和创业型科技人才的扩展元素设计上进一步研究、探讨，使科技人才元数据标准能指导更多的实际工作。同时将继续研究科技人才元数据共享互操作的难题，为不同结构的数据源提供集成映射方案。

参考文献

[1] 国家统计局,科学技术部.中国科技统计年鉴2011[M].北京:中国统计出版社,2011.
[2] 中共中央国务院.国家中长期人才发展规划纲要(2010-2020年)[Z].2010-06-06.
[3] 中国科学技术协会调研宣传部,中国科学技术协会发展研究中心.中国科技人力资源发展研究报告[M].北京:科学技术文献出版社,2010:13-19.
[4] 张玉勤,译.研究与试验发展调查实施标准弗拉斯卡蒂手册[M].高昌林,校.北京:中国科学技术出版社,2008:76-80.
[5] About FOAF [EB/OL]. [2012-12-05]. http://www.foaf-project.org/about.
[6] 水利部人才资源开发中心.人才管理数据库表结构及标识符标准(SL453-2009)[S].北京,2009-08-07.
[7] 肖琥,陈凌,冯项云,等.中文元数据标准框架及其应用[J].大学图书馆学报,2001,19(5):29-35.
[8] 中国科学技术信息研究所.科技人才信息整合机制研究报告[R].2010.
[9] 王运红,赵伟,李大玲.科技人才信息共享中的元数据结构设计[C]//北京:Scientific Research Publishing, USA,2011:356-359.
[10] Dublin Core Metadata Element Set, Version 1.1 [EB/OL]. [2012-11-05]. http://dublincore.org/documents/dces/.
[11] Preservation Metadata for Digital Objects [EB/OL]. [2012-10-12]. http://www.oclc.org/research/activities/past/orprojects/pmwg/presmeta_wp.pdf.
[12] DIGITAL LIBRARIES: Metadata Resources [EB/OL]. [2000-11-05]. http://archive.ifla.org/II/metadata.htm.
[13] 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.
[14] 金更达,何嘉荪.电子文件元数据标准设计框架研究[J].档案与建设,2005(9):4-7.

[15] 曹蔚光,王申康.元数据管理策略的比较研究[J].计算机应用,2001(2):2-5.
[16] 冯项云,肖珑,廖三三,等.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.

作者简介

王运红（1971-），研究方向：科技资源信息化、科技人才管理。E-mail: wangyh@istic.ac.cn

Study of the Framework of Scientific and Technological Talents Metadata Standard

Wang Yunhong, Peng Jie, Li Daling, Wu Xiaoli / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: For the information standardization and sharing, it is necessary to study the metadata of scientific and technological talents. This article first introduces the purpose and importance on researching metadata standard of scientific and technological talents. Scientific and technological talents and its information features are both defined and explained, then the metadata standard Three-tier framework, the core and the extend elements were designed in this paper. The metadata standard of scientific and technological talents was researched and drafted, which can make the different metadata applications more regular. This study will provide data standardization for scientific and technological talents information utilization and data sharing.

Keywords: Scientific and technological talents, Metadata, Metadata standard, Metadata standard framework, Metadata element

（收稿日期：2013-01-16）