

# 复杂场景下科学数据中心运行状况评价模型研究

王瑞丹<sup>1</sup> 陈祖刚<sup>2,3</sup> 石蕾<sup>1</sup> 卢逸航<sup>2,4</sup> 李国庆<sup>2,3</sup>

(1. 国家科技基础条件平台中心, 北京 100038; 2. 中国科学院空天信息创新研究院, 北京 100094;  
3. 国家对地观测科学数据中心, 北京 100094; 4. 中国科学院大学, 北京 100049)

**摘要:** 科学数据成为科研工作乃至国家发展的重要战略资源, 国内外纷纷设立了一大批科学数据中心。随之而来的是, 科学数据中心建设发展相互隔离, 水平高低不同的现象十分明显, 迫切需要制定一套适应国家数据战略需求的科学数据中心运行状况评价体系对不同层次科学数据中心进行系统而有效的评价。在分析科学数据中心学科领域、层次级别、地理空间位置分布、发展阶段、隶属系统等异构性特征和学科动态发展、国家政策变化、科学数据中心工作重点等动态特征的基础上, 提出科学数据中心运行状况评价模型。该评价模型能有效应对科学数据中心多样化、差异化和动态变化等复杂场景, 提供公平合理的科学数据中心量化运行状况评价值, 在我国科学数据中心管理工作中具有重要的应用价值。

**关键词:** 科学数据; 科学数据中心运行状况; 评价模型; 动态性; 异构性

**DOI:** 10.3772/j.issn.1674-1544.2022.02.004

**CSTR:** 15994.14.issn.1674.1544.2022.02.004

**中图分类号:** C931.6

**文献标识码:** A

## Research on the Evaluation Model of Operating Status of Scientific Data Centers under Complex Scenarios

WANG Ruidan<sup>1</sup>, CHEN Zugang<sup>2,3</sup>, SHI Lei<sup>1</sup>, LU Yihang<sup>2,4</sup>, LI Guoqing<sup>2,3</sup>

(1. National Science and Technology Infrastructure, Beijing 100036; 2. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094; 3. National Earth Observation Data Center, National Science and Technology Infrastructure, Beijing 100094; 4. University of Chinese Academy of Sciences, Beijing 100049)

**Abstract:** Scientific data has become an important strategic resource for scientific research and even national development. A large number of scientific data centres have been set up at home and abroad. As a result, there are many scientific data centres with different levels of quality and independent to each other. There is an urgent need to develop an evaluation model of operating status of scientific data centers that meets the needs of national data strategies to evaluate scientific data centers at different levels in a systematic and effective manner. This study proposes an evaluation model for the operation of scientific data centres, taking into

**作者简介:** 王瑞丹 (1965—), 女, 国家科技基础条件平台中心研究员, 研究方向为科技资源管理; 陈祖刚 (1989—), 男, 中国科学院空天信息创新研究院助理研究员, 研究方向为对地观测科学数据共享与知识挖掘研究; 石蕾 (1982—), 女, 国家科技基础条件平台中心研究员, 研究方向为科技资源管理与共享; 卢逸航 (1996—), 男, 中国科学院大学硕士生, 研究方向为对地观测科学数据共享与知识挖掘研究; 李国庆 (1968—), 男, 中国科学院空天信息创新研究院研究员, 研究方向为高性能地学计算和网络化大数据工程 (通信作者)。

**基金项目:** 国家自然科学基金面上项目“基于动态与异构场景的科学数据中心评价方法研究”(72074017)。

**收稿时间:** 2022年1月8日。

account the heterogeneous characteristics of scientific data centers in terms of disciplinary fields, hierarchical levels, geographical and spatial distribution, development stages and affiliation systems, as well as the dynamic characteristics of disciplines, national policy changes and data center work priorities. The evaluation model can effectively respond to complex scenarios such as diversification, differentiation and dynamic changes of scientific data centers, and provide a fair and reasonable quantitative evaluation value of operating status of scientific data center, which is of great application value in the management of scientific data centers in China.

**Keywords:** scientific data, operating status of scientific data centers, evaluation models, dynamism, heterogeneity

## 0 引言

科学数据通常是指在自然科学、工程技术科学等领域,通过基础研究、应用研究、试验开发等产生的数据以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据<sup>[1]</sup>。科学数据中心是科学数据管理与共享领域内实施具体事务解决实际问题、促进科学数据资源管理与共享水平持续提高的专门机构。随着科学研究进入数据驱动的第四范式,科学数据成为科研工作乃至国家发展的重要战略资源<sup>[2]</sup>,科学数据中心成为科研工作取得国际重大创新性突破的关键基础设施。

在此背景下,国内外设立了一大批科学数据中心。在国外,有美国国家空间科学数据中心(National Space Science Data Centre, NSSDC)、英国数据保存中心(Digital Curation Centre, DCC)、英国数据档案中心(UK Data Archive)以及澳大利亚国家数据服务中心(Australian National Data Service, ANDS)<sup>[3]</sup>。在我国,科技部、财政部于2019年6月发布《国家科技资源共享服务平台优化调整名单》,公布了20个国家科学数据中心和30个国家生物种质与实验材料资源库,从此我国国家级科学数据中心开始进入全面建设的阶段<sup>[4]</sup>。

在大数据、互联网快速发展的今天,科学研究产生的科学数据资源越来越多,学科之间交叉融合越来越强,传统学科边界越来越模糊,增加了科学数据中心管理的复杂性和多样性。然而,科学数据中心处于高速发展期,科学数据中心建设发展相互独立,水平高低不同的现象十分

明显,存在大量科学数据管理不规范、不完整、质量无法保证的现象。而许多科学数据中心对数据挖掘和再分析的能力不足,只是简单地把数据收集起来,难以发现数据背后的价值信息和增值渠道<sup>[5]</sup>。迫切需要建立科学数据中心的运行状况评价体系,来促进科学数据中心提升管理水平,保障科学数据的真实性和可靠性,支撑国家重大决策和经济发展,充分发挥各类科学数据中心的作用和价值。

近10年来,多种类型科学数据中心建设资助机构积极开展研究并建立了对科学数据中心运行评价制度,通过制定评价指标和流程,引入第三方专业评估机制等方式,对科学数据中心的运行质量和效果进行重点考察,并探索将评价结果作为机构卓越性和影响力的重要衡量依据<sup>[6]</sup>。目前,科学数据中心运行评价体系主要有科技部2011年实施的国家科技基础条件平台运行服务绩效考核、中国科学院科技数据应用效果评价、世界数据系统(WDS)CoreTrustSeal数据中心认证体系等。虽然一些部门、单位或国际组织已经形成了一些数据中心评价考核指标并开展了评价实践,但各个评价体系差异明显,不具备可比性和通用性,不利于科学数据中心的协调统一发展,也不利于形成高水平的国家级和世界级科学数据中心,而现有的评价指标的量化程度较低,缺乏系统性的研究和成果,评价指标体系没有考虑科学数据中心类型、学科、发展阶段、社会环境等特征造成的差异性。因此,急需构建新型的适用于各种不同的科学数据中心的评价模型和体系,实现对科学数据中心公平、有效、合理的评价,以促进科学数据中心管理能力和科学数据支

撑科学研究能力的提升。

本文将在综合考虑科学数据中心差异性特征的基础上，提出一套适用于不同类型不同级别的科学数据中心运行状况评价模型，以提升科学数据评价体系的技术水平和能力。

## 1 研究现状

### 1.1 科学数据中心运行状况评价方法

科学数据中心的运行状况评估方式与方法主要分为机构自评和第三方评估两种。机构自评即由数据仓储机构自身根据设定好的评估体系和指标进行自我检视和评价的一种评估方法。该评估方法可随时开展，有利于其判断自身发展是否达到预期目标以及判断其水平高低的标准。在评价活动正式开始时，需要被评价方提交正式的自我评价报告，根据评价指标要求，逐项对其运行绩效进行描述。核心可信数据仓储认证（CoreTrustSeal Certification, CTS）<sup>[7]</sup>和开放数据晴雨表（The Open Data Barometer, ODB）<sup>[8]</sup>等评估方案都采取了机构自评这种方式。前者为主动的自评，后者为被动的自评。

第三方评估则是指由与评价组织者和被评价主体无直接关系的第三方按照一定的标准对被评价主体做出评价。这种评估方式能够避免利益相关部门的利益羁绊，最大程度地实现评估结果的独立、客观与公正。在针对开放数据的评估中，第三方评估有专家调研、同行评审、大众评估和网络系统分析等方法。ODB在最近一次评估方案中，通过定制的在线调研平台于2017年12月至2018年3月进行专家调研<sup>[8]</sup>。联合国开放政府数据评估（Open Government Data Assessment, OGD A）也采用了专家调研这种方法，但使用了面对面访谈形式进行专家调研<sup>[9]</sup>。CTS认证过程中，机构自评提交后进入同行评审阶段，由CTS董事会召集人员来完成<sup>[7]</sup>。OGD A大众认知调查对至少50位包括私人机构、政府机构、筹资/投资机构、社会服务机构、科研学术机构、民用技术和媒体等方面的代表进行了问卷调查<sup>[9]</sup>。欧洲开放数据监测评估项目（OpenDataMonitor）通过

网站平台收割欧洲公共行政机构开放数据中心的元数据，描述欧洲开放数据资源概况，实现数据目录、数据和元数据的开放以及评估结果的可视化<sup>[10]</sup>。2011年，科技部、财政部联合发布了《关于开展国家科技基础条件平台认定和绩效考核工作的通知》<sup>[11]</sup>，并向社会公布了《国家科技基础条件平台认定指标》和《国家科技基础条件平台运行服务绩效考核指标》<sup>[12]</sup>。

### 1.2 国内外科学数据中心评价方法的应用

在国外，CTS的认证方式为首先根据评估指南Core Trustworthy Data Repositories Extended Guidance进行机构自评，再通过同行评审，评估方案组织方CoreTrustSeal委员会才会向数据中心颁发CoreTrustSeal认证<sup>[7, 13]</sup>。开放数据准备度（Open Data Readiness Assessment, ODR A）制定并发布了详细的评估办法“Open Data Readiness Assessment Methodology”和用户指南“Open Data Readiness Assessment Users’ Guide”，对8个评估维度进行了详细的解释，并针对每个评估维度，列出了正面和负面的例证，给出了非常详尽的评价参考<sup>[14]</sup>。OGD A综合了UNDESA的“开放政府数据公众参与指南”（Open Government Data for Citizen Engagement Guidelines）、世界银行的ODR A、互联网基金会的“开放数据晴雨表”以及世界经合组织（OECD）的“OUR数据指数”等评估方案的特点，从8个评估维度出发，采用了包括文献研究、媒体监测、桌面研究、调研、自评（工具）、实地考察以及利益相关者专题讨论等评估方法<sup>[15]</sup>。ODB以专家调查为主导，政府自评和二手数据调研作为辅助，制定发布的专家评估手册（Open Data Barometer Leaders Edition Research Handbook - v1.0），对专家调查提出了非常详细的建议。它建议专家通过问卷中的信息检索指导、开放数据专家、非营利组织成员、记者和政府专员等途径寻求评估问卷中所涉及问题的信息和证据<sup>[8]</sup>。基础风险评估的数字仓储审核方法（Digital Repository Audit Method Based on Risk Assessment, DRAMBORA）自我评估过程包含3个步骤：①评审员创建机构

简介,描述仓储的义务、目标、活动和资产;②评价中存在的风险;③评审员提出针对已识别风险的管控办法,并通过一个在线评估工具包 DRAMBORA interactive toolkit 实现评估<sup>[16]</sup>。美国 NASA 的 EOSDIS (Earth Observation System) 系统对其数据中心的评价主要分为两部分:一是对其系统性能数据进行提取,主要考核其科学数据处理系统、地球科学数据与服务、元数据、常规元数据仓储、用户注册系统等方面;二是接受美国科罗思咨询集团(CFI Group)开展的客户满意度调研,该集团使用的基于科罗思方法的美国客户满意度指数 ACSI 是美国国家标准指数之一<sup>[17-18]</sup>。

在国内,国家人口健康科学数据共享平台根据《国家人口健康科学数据共享平台运行服务实施方案》形成了稳定的运行考评机制,其评价方式包括线上评测、现场评测、组织行为评测和自评4种方式。考核评价指标分为两套核心评价量表,即绩效考核指标表、资源综合评价指标和评估内容<sup>[19]</sup>。国家农业科学数据中心使用两种方式进行运行评价:按一套可量化的体系进行自动化评估以及按一套定性评估指标体系由领域专家进行会议评估。其由系统自动提取的定量评价指标有3个一级指标,即环境类指标、数据库类指标、用户类指标;定性评价指标有3个一级指标和8个二级指标,重点考查数据整合与中心建设、共享服务效益以及人才队伍与运行机制等3个方面<sup>[20]</sup>。国家生态系统观测研究网络制定了其章程、运行服务绩效考核评估办法、评估细则、野外研究站遴选考核与评估办法等规章制度,其年度考核采用自动统计、人工抽查、专家评估、用户反馈等形式,对各项评估指标进行加权统计,综合打分。每3年依据前两年的考核结果对国家生态站进行一次综合测评,每5年进行一次综合评估<sup>[21]</sup>。中国科学院科学数据库每年开展一次定量与定性相结合的考核,考核方式包括系统采集数据、提交考核报告、口头汇报、专家评审等方式,评价指标体系含有4个一级指标,有12个二级指标,有16个三级指标,主要考核数据

中心在数据资源、在线系统、运行维护、服务成效等方面的表现<sup>[22]</sup>。国家科技基础条件平台认定和绩效考核方法秉承科技平台“整合、共享、完善、提高”的理念,以管理机制、科技平台资源整合状况及运行状况和长效发展能力作为重点考察内容;绩效考核指标以认定指标为基础,突出科技平台的共享作用,重点考察科技平台的服务数量与服务成效,重视用户评价的反馈<sup>[23]</sup>。

### 1.3 评价方法的分析

经过对当前评估案例的调研,可以发现当前的评估方案具有的特点是:一个评估方案通常采用几种评估方法,但评估指标是固定的;定性评估与定量评估相结合,没有一个综合性的可比较结论。综合来看,现有科学数据中心评价体系的不足之处包括:①科学数据中心是同质的,科学数据中心的运行背景是静态的,因而评价体系的指标内容和权重等要素都是固定的,而科学数据中心实际上是在非常复杂的场景下运行发展,如科学数据中心在学科类别、层次等级、建设时长等方面存在巨大的差异性,且科学研究不断进步,科学数据中心发展的政策、学术环境也是动态变化的,用统一的、简单化的、一成不变的考核评价方法体系,显然不能适用于科学数据中心公平和长期的评价。②现有的评价方法的量化程度不足,以定性或者半量化的方法为主,这就导致了科学数据中心评价时存在较多的灰色地带,也增加了评价的工作量。因而,迫切需要针对科学数据中心的复杂场景开展科学数据中心运行状况评价模型的研究。

## 2 模型构建

### 2.1 科学数据中心的异构和动态特征分析

科学数据中心运行发展的复杂场景可以分为异构性和动态性条件(特征)两个方面。

(1)异构是指一种包含不同成分的性质。科学数据中心的异构特征体现在很多方面,如科学数据中心的地理空间位置分布、层次级别、学科领域、隶属系统等。对科学数据中心运行状况进行公平合理评价需要充分考虑性质不同的诸因素

的影响。本文选取不同科学数据中心间差异最明显的学科、级别、分布区域、成熟度、隶属系统作为科学数据中心的异构特征。

(2) 动态是指随时间发生变化的性质。科学数据中心的动态特征是随着时间发展, 科学数据中心内部条件、外部环境发生变化的特征。如随着科学数据中心建设的深入, 科学数据中心的工作重心是动态变化的, 而随着学术研究的深入和支持政策变化, 科学数据中心的考核指标也应该是动态变化。本文通过评价指标的权重变化体现工作重心的变化, 选取考核指标的属性体现考核指标的动态变化。

## 2.2 科学数据中心的动态异构评价指标体系模型

在开展科学数据中心运行状况评价时, 需要以具体的考核指标获取运行状况数据, 因而考核指标的设计至关重要。本文提出的考核指标不同于现有的考核指标体系, 所提出的考核指标体系需要明确考核指标的动态性、异构性和指标采集数据的定量或者定性特征。

设科学数据中心的考核指标体系为  $P_1, P_2, P_3, P_4, \dots$ , 而对于多级考核指标体系, 可以在现有指标体系的基础上划分为二级或者多级的指标。如  $P_1$  可以继续划分为  $P_{11}, P_{12}, \dots, P_{1n}$ 。考核指标可以根据科学数据中心的管理者意图、当前科研条件与水平、科学数据中心职责内容等具体设立。在充分借鉴国内外现有科学数据中心考核评价指标体系内容的基础上, 根据我国科学数据中心发展的特殊阶段, 结合我国《科学数据管理办法》对科学数据中心职责的要求, 本文提出了一套科学数据评价指标体系。这套指标体系从科学数据中心的资源整合能力、平台服务、分析与挖掘、运行管理、支撑条件、可持续发展和国际影响力等 7 个方面对科学数据中心进行评价, 具体包括 30 个二级评价指标, 具体情况见表 1。对于每一个具体的考核指标, 进一步明确这个考核指标的动态性、异构性和数据性质。譬如, 多数考核评价指标体系都把科学数据中心的注册用户量作为考核指标。很明显, 这个指标既受科学数据中心所属的学科、级别、成熟度、隶属系统等异构特

征的影响, 也受科学数据中心工作重心等动态特征的影响, 注册用户量可以直接用一个整型数表达, 是一个实数。因而, 科学数据中心动态异构评价指标体系的模型见表 1。

## 2.3 科学数据中心运行状况评价模型

由表 1 可知, 科学数据中心评价的指标集合为  $P: \{P_1, P_2, P_3, P_4, P_5, \dots\}$ , 每个评价指标都有相应的异构和动态特征, 其中动态特征的权重是指这个评价指标的权重随时间的变化而变化, 属性是指这个评价指标的内涵随时间的变化而变化。评价指标集合  $P$  中的内容也是随时间的变化而变化。把异构特征的学科、级别、区域、成熟度、隶属度分别表示为  $S, L, D, M, U$ ; 把动态特征的权重和属性分别表示为  $W, A$ 。

在一般情况下, 科学数据中心的运行状况评价有一个周期, 有些科学数据中心管理机构 1 年评价 1 次, 也有一些机构 3 年评价 3 次, 甚至 5 年评价 5 次。为了在一个较长时间段上描述科学数据中心评价工作, 设在一个较长时间段内对科学数据中心评价的时间点为  $t_i (i=1, 2, \dots, m)$  ( $t_i$  指一个评价周期的结尾, 可以是 1 年或者  $n$  年的结尾), 评价对象 (即各个科学数据中心) 为  $U_k (k=1, 2, \dots, g)$ 。

设特定时间点为  $t_i$ , 其评价指标体系以及指标的权重都是确定的, 其评价体系主要体现为异构性。通过实时抓取、科学数据中心填报、考核指标信息值量化建模等步骤, 设所有评价对象采集的数据集为:

$$X_i = U \begin{matrix} P_1 & P_2 & \dots & P_n \\ \left[ \begin{array}{cccc} X_{11}^i & X_{12}^i & \dots & X_{1n}^i \\ X_{21}^i & X_{22}^i & \dots & X_{2n}^i \\ \dots & \dots & \dots & \dots \\ X_{g1}^i & X_{g2}^i & \dots & X_{gn}^i \end{array} \right. \end{matrix} \quad (1)$$

### 2.3.1 异构特征评价内容转换模型

$X_{in}$  是一个实数。如何把通过各个评价指标  $P_j$  采集的具有异构特性的实数、区间数和语言变量转化为合理的  $X_{in}$  值, 是本文提出的评价模型要解决的关键问题之一。根据考核指标采集的原始数

表1 科学数据中心动态异构评价指标体系模型

类别	考核指标	数据性质	异构特征	动态特征
资源整合	清单任务完成度 (P1)	实数	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	资源增量 (P2)	实数	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	资源数量 (P3)	实数	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	资源质量 (P4)	区间数	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	资源合作网络 (P5)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	数据资源汇交 (P6)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
平台服务	资源服务量 (P7)	实数	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	注册用户量 (P8)	实数	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	数据出版量 (P9)	实数	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	专题服务情况 (P10)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	科技支撑效果 (P11)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	社会效益 (P12)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input checked="" type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	经济效益 (P13)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input checked="" type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	行业效益 (P14)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input checked="" type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	服务对象满意度 (P15)	实数	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	数据分析与挖掘	数据管理能力 (P16)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统
	数据分析挖掘能力 (P17)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	工具应用能力 (P18)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
运行管理	组织机构运行 (P19)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	标准规范制度建设与落实 (P20)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	信息化管理能力 (P21)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	资质认证 (P22)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
支撑条件	工作场地 (P23)	语言变量	<input type="checkbox"/> 学科 <input type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input type="checkbox"/> 内容
	信息设施能力 (P24)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
可持续能力	期内新增资助强度 (P25)	实数	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input checked="" type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	人才团队与人才培养 (P26)	语言变量	<input type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input checked="" type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
国际影响力	国际化人才 (P27)	语言变量	<input type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input checked="" type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	国际用户 (P28)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	国际应用 (P29)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
	国际交流 (P30)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input type="checkbox"/> 区域 <input type="checkbox"/> 成熟度 <input type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容
其他	其他 (P31)	语言变量	<input checked="" type="checkbox"/> 学科 <input checked="" type="checkbox"/> 级别 <input checked="" type="checkbox"/> 区域 <input checked="" type="checkbox"/> 成熟度 <input checked="" type="checkbox"/> 隶属系统	<input checked="" type="checkbox"/> 权重 <input checked="" type="checkbox"/> 内容

注：每个评价指标的数据性质分为3种：一是实数，如300等；二是区间数，如500~800、A~D等；三是语言变量，如一段文字、图文结合的文档等。

据的不同，本文提出把采集的原始数据结合评价指标的异构特征，构建将原始数据转换为合理值的异构函数的方法。即，对于实数类型的原始值  $P_{ih}$ ，通过式(2)转换为  $X_{ih}$ 。

$$X_{ih} = F(P_{ih}, S, L, D, M, U) \quad (2)$$

其中， $F$ 是转换函数。

对于区间数据类型的原始值  $P_{ih}$ ，首先需要把区间数转化为具体的实数，进而结合异构特征构建转换函数。即，

$$P_{ih} \xrightarrow{f} Q_{ih} \quad (3)$$

其中， $Q_{ih}$ 是一个实数， $h$ 是区间数  $P_{ih}$ 转换为实数的规则或者函数。最后  $X_{ih}$ 获取的方法为：

$$X_{ih} = F(Q_{ih}, S, L, D, M, U) \quad (4)$$

对于语言变量类型的原始值  $P_{ih}$ ，首先需要通过专家评审打分、专家系统、人工智能模型等方式把语言变量类型转化为可比较大小的实数，进而结合异构特征构建转换函数获取  $X_{ih}$ 。即，

$$P_{ih} \xrightarrow{f(\text{专家评审打分|知识系统|人工智能模型等})} Q_{ih} \quad (5)$$

### 2.3.2 动态特征评价内容转换模型

获取 $t_i$ 时刻各个科学数据中心的评价指标体系的量化值后，为了把所有考核指标的值转化为一个综合值，需要确定每个评价指标的权重值，即 $t_i$ 时刻的权重。这个权重可以随着时间的变化而变化。 $t_i$ 时刻的权重可以采用层次分析法<sup>[24]</sup> {Saaty, 1990 #26} 等方法建立。该方法将一个复杂的多目标决策问题看作一个系统，将目标分解为多指标（或准则、约束）的若干层次，并通过计算判断矩阵的特征向量，获得每一层次的各元素最优权重。

在 $t_i$ 时刻，对 $U_k$ 科学数据中心经过异构建模获取了其指标量化值 $(X_{k_1}^i, X_{k_2}^i, \dots, X_{k_n}^i)$ ，经过层次分析法等方法获取了指标的权重 $(W_{k_1}^i, W_{k_2}^i, \dots, W_{k_n}^i)$ ，那么 $t_i$ 时刻科学数据中心的综合评价价值 $a_{ki}$ 为：

$$a_{ki} = \sum_{j=1}^n X_{kj}^i \cdot W_{kj}^i \quad (6)$$

在其他时刻，科学数据中心的考核指标体系是需要动态调整的即 $P$ 是变化的，各个考核指标的权重 $(W_{k_1}^i, W_{k_2}^i, \dots, W_{k_n}^i)$ 也是要调整的，相应时刻考核评价指标体系的建立需要在上一个时刻指标体系的基础上，综合考虑管理者意图、学科发展状况、国家政策环境、科学数据中心业务内容等进行调整。在一般情况下，相邻两个时刻的考核评价指标体系不做巨大的变动。考核指标确定以后，可再次应用层次分析法，确定此时刻评价指标的权重。

### 2.3.3 科学数据中心评价综合值

对于一个评价时间段的多个时间点，经过为每个时间点构建评价指标体系模型、评价内容采集、异构评价内容建模、指标权重动态确定等步骤，最终可以获取各个评价对象综合评价值。这些评价价值的时间序列构成如式（7）的矩阵：

$$A = \begin{matrix} & t_1 & t_2 & \dots & t_m \\ \begin{matrix} U_1 \\ U_2 \\ \dots \\ U_g \end{matrix} & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{g1} & a_{g2} & \dots & a_{gm} \end{bmatrix} \end{matrix} \quad (7)$$

其中， $a_{ij}$ 表示评价对象 $U_i$ 在 $t_j$ 时刻的科学数据中心评价综合值。

假设需要获取一个评价周期多个时刻评价指标的综合值，那么每个时刻对应的评价价值可以采用加权求和的方式，作为这个评价周期的综合评价价值，实现动态异构场景下科学数据中心的评价。

## 3 结论与讨论

科学数据中心运行状况评价是对于科学数据中心运行的成效进行监督的重要手段，也是促进科学数据中心提高管理服务水平，支撑科技创新和经济发展的重要方法。我国科学数据中心具有涉及的学科领域多、管理层次级别多、空间分布区域多、发展阶段不同、隶属行政系统不同等复杂场景，如何在此复杂场景下开展科学数据中心的公平、合理评价，提高科学数据中心建设管理服务的积极性，激发数据中心的活力，是本文需要解决的首要问题。

本文把我国科学数据中心的复杂场景特性总结为数据中心的动态性和异构性，提出了考虑动态异构特征的科学数据中心运行状况定量化评价模型。该模型首次考虑了科学数据中心的异构性和动态性特征，提出了异构和动态建模方法，最终构建了定量化的科学数据中心运行状况评价指数，实现任意科学数据中心在任意时刻的定量化评价，大大提高了科学数据中心运行评价方法的科学性、先进性、公平性、实用性，必将促进科学数据中心管理工作的进步，提升科学数据中心的总体发展水平。

本文是首个面向复杂场景下科学数据中心定量化评价的模型，也是科学数据中心合理评价的顶层设计方案。该模型的有效实施依赖于3个重要基础：①科学数据中心各个评价指标数据的有效采集，包括数据中心运行状况信息的完整和准确填报两个方面。②异构特征转换模型的建立，异构特征转换函数能否兼顾公平与效率，反映出科学数据中心建设的成效，是该评价模型有效实施的关键。③评价指标及其权重的动态生成，决

定了该模型能否反映某一阶段数据中心工作重心,也决定了该模型在引领科学数据中心发展的能力。

本文作为科学数据中心运行状况评价方法的前瞻性研究,将会在接下来的科学数据中心评价工作中实现具体应用。未来将进一步提升模型的精细化程度,采用人工智能、自动化系统、公众纵包等技术和方法提高模型的自动化程度,并面向部分科学数据中心,开展评价数据的采集,构建异构指标转换函数及模型,采用层次分析法,建立专家委员会,利用专家智能开展指标权重定期评价,实现科学数据中心评价指标体系权重的持续更新,利用异构函数转化后的评价指标值以及相应的指标权重,实现科学数据中心运行状况的定量化评价。

## 参考文献

- [1] 苏楠. 一种文件复制方法, 装置及可读存储介质: CN111858766A [P]. 2020.
- [2] 徐波, 王瑞丹, 陈祖刚, 等. 科学数据中心综合运行评价体系赋权研究[J]. 中国科技资源导刊, 2021, 53(4): 96-103.
- [3] 王瑞丹, 杨静, 高孟绪, 等. 加强和规范我国科学数据管理的思考[J]. 中国科技资源导刊, 2018, 50(2): 1-5.
- [4] 科技部, 财政部. 科技部 财政部关于发布国家科技资源共享服务平台优化调整名单的通知[EB/OL]. (2019-06-05)[2021-06-16]. [http://www.most.gov.cn/xxgk/xinxifenlei/fdzdgdgknr/qtwj/qtwj2019/201906/t20190610\\_147031.html](http://www.most.gov.cn/xxgk/xinxifenlei/fdzdgdgknr/qtwj/qtwj2019/201906/t20190610_147031.html).
- [5] 王卷乐, 王祎, 卜坤, 等. 世界数据系统 CoreTrustSeal 数据中心认证实践: 以 WDC 可再生资源与环境数据中心为例[J]. 农业大数据学报, 2019(3): 71-81.
- [6] 王明明, 王卷乐, 赵强, 等. ICPSR 科学数据中心的建设经验与启示[J]. 中国科技资源导刊, 2017, 49(6): 100-107.
- [7] DSA-WDS. DSA-WDS core trustworthy data repositories requirements[EB/OL]. [2021-06-16]. <https://www.coretrustseal.org/why-certification/requirements/>.
- [8] World\_Wide\_Web\_Foundation. Open data barometer leaders edition research handbook - v1.0[EB/OL]. (2017-10-02)[2021-06-16]. [https://docs.google.com/document/d/1I24b20z9n3-FF8UCZjzqv3pB6-ENmJYROZ\\_0cyQIAWM/edit#heading=h.gjdgxs](https://docs.google.com/document/d/1I24b20z9n3-FF8UCZjzqv3pB6-ENmJYROZ_0cyQIAWM/edit#heading=h.gjdgxs).
- [9] UN\_DESA. Open Government Data (OGD) readiness assessment[EB/OL]. (2019-10-02)[2021-06-16]. <https://publicadministration.un.org/en/ogdassessment>.
- [10] OPENDATAMONITOR. Open data monitor [EB/OL]. (2019-10-02)[2021-06-16]. <https://opendatamonitor.eu/frontend/web/index.php?r=dashboard%2Findex>.
- [11] 科技部. 关于开展国家科技基础条件平台认定和绩效考核工作的通知[EB/OL]. (2019-10-02)[2021-06-16]. [http://www.most.gov.cn/xxgk/xinxifenlei/fdzdgdgknr/qtwj/qtwj2011/201108/t20110805\\_88852.html](http://www.most.gov.cn/xxgk/xinxifenlei/fdzdgdgknr/qtwj/qtwj2011/201108/t20110805_88852.html).
- [12] 黄铭瑞, 李国庆, 李静, 等. 国家科学数据中心管理模式的国际对比研究[J]. 农业大数据学报, 2019(4): 14-29.
- [13] DILLO I, LEEUW L D. CoreTrustSeal[J]. Mitteilungen der vereinigung österreichischer bibliothekarinnen & bibliothekare, 2018, 71(1): 162-70.
- [14] World\_Bank\_Group. Open data readiness assessment users' guide[EB/OL]. (2019-10-22)[2021-06-16]. [http://opendatatoolkit.worldbank.org/docs/odra/odra\\_v3.1\\_userguide-en.pdf](http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_userguide-en.pdf).
- [15] HU Y, BAI X, SUN S. Readiness assessment of open government data programs: a case of shenzhen [C]// Proceedings of the 17th International Digital Government Research Conference on Digital Government Research. Shanghai: Association for Computing Machinery, 2016: 97-103.
- [16] DRAMBORA. Digital Repository Audit Method Based On Risk Assessment, DRAMBORA [EB/OL]. (2019-10-22)[2021-06-16]. <https://www.dcc.ac.uk/tools/drambora>.
- [17] NASA\_EOSDIS. EOSDIS Distributed Active Archive Centers (DAACs)[EB/OL]. [2021-06-16]. <https://earthdata.nasa.gov/eosdis/daacs>.
- [18] NASA\_EOSDIS. NASA EOSDIS annual customer satisfaction survey 2019[EB/OL]. (2019-10-22)[2021-06-16]. [https://podaac.jpl.nasa.gov/announcements/2019-10-22\\_NASA\\_EOSDIS\\_Annual\\_Customer\\_Satisfaction\\_Survey\\_2019](https://podaac.jpl.nasa.gov/announcements/2019-10-22_NASA_EOSDIS_Annual_Customer_Satisfaction_Survey_2019).
- [19] 李赞梅, 胡志民, 孙海霞, 等. 国家人口与健康科学数据共享平台资源综合评价指标构建研究 [J]. 中国数字医学, 2018, 13(3): 2-4.
- [20] 国家农业科学数据中心. 农业科学数据质量检查与控制规范(NADC002)[EB/OL]. (2019-01-01)[2021-06-16]. <https://www.agridata.cn/detail.html?type=standard&pt=%2FWord%2F%E5%86%9C%E4%B8%9A%E7%A7%91%E5%AD%A6%E>

(下转第64页)



发展科普人员队伍。科普投入是一个长期的持续性行为,建立以政府投入为主导,社会资本广泛参与的多元化投入体系,同时加大科普经费的利用效率,形成科普促进科技创新的可持续激励机制。科普人员是科学知识和理论的传播者,是科普事业的管理者。人才是第一资源。要进一步提升科普创作人员和志愿者素质水平,加强科普人员队伍建设,不但是提升我国科普综合能力的内在要求,也是提高公民科学素质的重要手段。

### 参考文献

- [1] 张思光,吴坎坎,武向平.我国公务员科学素质建设的现状、问题及建议:基于广东、湖北、新疆的问卷调查分析[J].中国科学院院刊,2020,35(9):1123-1132.
- [2] 邱成利.推进我国科普资源开发与建设的若干思考[J].中国科技资源导刊,2015,47(3):1-6,14.
- [3] 张瑞才.承担职责使命创新社会科学普及[J].学术探索,2016(1):1-7.
- [4] 王康友,尹霖,谢小军,等.把科学普及这一翼打造得更强大[J].科普研究,2016,11(3):5-9.
- [5] 姚晓辉.提升公众科学素养刻不容缓[J].人民论坛,2018(23):80-81.
- [6] 李健民.科技创新与科学普及融合发展的思考[J].安徽科技,2019(7):5-7.
- [7] 刘嘉麒.让科学普及与科技创新比翼高飞[J].科学教育与博物馆,2018,4(6):361-363.
- [8] 任嵘嵘,郑念,赵萌.我国地区科普能力评价:基于熵权法——GEM[J].技术经济,2013,32(2):59-64.
- [9] 吴华刚.我国省域科普资源建设水平指标体系的构建及评价研究[J].科技管理研究,2014,34(18):66-69.
- [10] 张立军,张潇,陈菲菲.基于分形模型的区域科普能力评价与分析[J].科技管理研究,2015,35(2):44-48.
- [11] 李建忠.我国各省科技服务绩效评价研究:以科学普及为例[J].科技管理研究,2015,35(8):43-46,51.
- [12] 张越,符洋,黄虹.基于因子分析的湖南省区域科普能力测度评价[J].科技管理研究,2019,39(18):82-87.
- [13] 赵艳君,谢铭.陕西省科普能力指标体系构建与评价[J].中国科技信息,2019(23):97-98,100.
- [14] 胡萌,朱安红.江西省科普效果指标体系及综合评价研究[J].科技广场,2012(12):21-24.
- [15] 王宾,李群.基于DEA分析的中国科普投入产出效率评价研究[J].数学的实践与认识,2015,45(15):214-220.
- [16] 刘广斌,李建坤.基于DEA方法的地区科普资源配置及利用效率评价[J].科普研究,2017,12(6):69-76,108-109.
- [17] 李卉,熊春林,尹慧慧.基于规模与效率的地区科普能力评价研究[J].科技与经济,2019,32(3):11-15.
- [18] 李婷.地区科普能力指标体系的构建及评价研究[J].中国科技论坛,2011(7):12-17.
- [19] 佟贺丰,刘润生,张泽玉.地区科普力度评价指标体系构建与分析[J].中国软科学,2008(12):54-60.
- [20] 佟贺丰.科普投入的国内外对比研究及对策分析[J].科普研究,2006(4):3-8.
- [21] 董全超,胡峰,马宗文.基于典型相关分析方法的我国科普投入产出研究[J].科普研究,2019,14(2):61-67.

(上接第34页)

- 6%95%B0%E6%8D%AE%E8%B4%A8%E9%87%8F%E6%A3%80%E6%9F%A5%E4%B8%8E%E6%8E%A7%E5%88%B6%E8%A7%84%E8%8C%83637304045984360850.html#.
- [21] 国家生态系统观测研究网络.国家生态系统观测研究网络平台运行服务绩效考核评估细则[EB/OL].(2019-10-22)[2021-06-16].[http://www.cnerm.org/getContent.action?id=1750&class\\_id=76&article\\_class\\_id=2](http://www.cnerm.org/getContent.action?id=1750&class_id=76&article_class_id=2).
- [22] 中科院计算机网络信息中心.中国科学院科学数据库[EB/OL].(2019-01-01)[2021-06-16].<https://www.cas.cn/ky/kycc/kxsjk/>.
- [23] 许东惠,吕先志,袁伟,等.国家科技基础条件平台运行服务绩效考核指标体系研究[J].中国基础科学,2013,15(1):40-43.
- [24] SAATY T L.How to make a decision: the analytic hierarchy process[J].European journal of operational research,1990,48(1):9-26.