

基于系统动力学的数据监护模型 仿真研究

宋秀芬¹ 邓仲华²

(1. 湖北警官学院, 武汉 430034; 2. 武汉大学信息管理学院, 武汉 430072)

摘要: 本文基于系统动力学原理分析数据监护的运行机制, 以寻求较优的数据监护系统结构与功能。采用系统动力学方法, 剖析数据监护的构成因素、影响因素及其相互关系, 揭示数据监护的运行机制。从国家与组织支持力度、组织策划能力、发展障碍、服务能力4个方面提出数据监护能力提升策略与建议。

关键词: 数据监护; 系统动力学; 数据质量; 数据服务; 开放存取

中图分类号: G203

DOI: 10.3772/j.issn.1673-2286.2019.01.001

目前传统机构知识库数据服务不到位, 基金委员会、出版商与科研机构以数据政策形式要求科研人员共享数据以实现数据增值。数据监护作为一个新兴研究领域应运而生, 是围绕数据生命周期而展开活动以确保其有用性、完整性、可靠性与真实性。数据监护改变了传统数据私有的观念, 打破信息壁垒, 能够实现数据有效交流与共享, 提升数据价值, 并加速科研进程。数据监护过程包括: 策划—元数据创建—数据评估与选择—数据摄取—数据维护—数据存储—数据访问、利用和再利用—数据转换^[1]。

国外学者^[2]于2001年开始研究数据监护, 以英美为首的机构或社区联盟已从数据监护试点项目转向实践应用, 在教育培训、政策制定、标准开发、平台构建等方面成效显著。数据监护研究主要涉及四大主题: 概念与模型、关键活动(数据共享、访问和再利用)、基础设施(信息化基础设施、学术团队、协作机制、发展规划、数据政策、标准及规范)、相关责任机构(档案馆、学术型图书馆、机构知识库与数据中心)^[3]。

国内学者^[4]于2011年开始研究数据监护。近几年, 国内学者在借鉴国外研究成果的基础上, 研究内容逐渐由引进介绍、理论研究及经验总结与启示等理论研究转向平台建设、教育培训等实践研究, 主要分布在5

个方面^[5], 即数据监护对图书馆的影响^[6-7]、国外数据监护职业发展与教育^[8-9]、国外数据监护研究综述^[10-11]、国外研究成果和实践经验借鉴^[12]及数据监护与知识库^[13]结合研究。总体上来看, 我国数据监护研究还处于起步阶段。

从研究方法上看, 国内外学者目前大都从定性角度研究数据监护, 鲜有结合具体理论模型, 以量化的方式从数据监护的影响因素及其相互关系的角度进行研究。数据监护是一个具有信息反馈功能的复杂系统, 其本身包括哪些构成要素及主要影响因素, 各要素之间如何牵动与联动, 如何实现资源优化配置来寻求较优数据监护系统结构与功能, 都尚未有定论。

为解决以上问题, 需要摸清数据监护系统运行机制。本文以系统动力学为研究视角, 将数据监护过程视为一个系统, 通过剖析数据监护的影响因素及各因素间的因果关系, 采用Vensim PLE 7.3.4版本软件作为可视化建模工具。通过主体与影响因素分析、模型边界与基本条件假设、因果关系分析、系统流图绘制、关键变量函数式的确定等步骤来构建数据监护模型; 通过变量函数与初始条件设定、模型有效性检验、既定参数下的仿真结果分析、多因素敏感性分析等步骤来仿真模型, 以量化方式呈现模型参数对模型输出的影响程度

来寻求较优系统结构与功能。

1 数据监护与系统动力学

1.1 数据监护

数据监护 (data curation或digital curation) 是对数字化科研数据整个生命周期进行维护、保存与增值的活动^[14-15]。数据监护中“数据”专指科研数据, 科研数据伴随着科学研究而产生, 科研数据的类型分为观测数据、实验数据、仿真数据与派生数据^[16-17]。数据监护活动3个关键环节或重要活动包括数据共享、开放存取与数据再利用^[18], 其系统包括组织架构、数字对象管理、技术与基础设施三大部分^[19]。

1.2 系统动力学

系统动力学 (system dynamics, SD) 是利用系统科学理论和计算机仿真来研究系统反馈结构与行为的一门科学。它是一门分析研究信息反馈系统的学科, 也是一门认识系统问题和解决系统问题的综合性交叉学科; 是系统科学与管理科学的一个分支, 跨越自然科学与社会科学^[20]。系统动力学方法适合于研究综合交叉学科的系统反馈结构与行为、理论与仿真结合方面^[21]。系统动力学的5种基本方法包括系统框图、因果关系图、系统流图、方程及仿真平台^[22]。

1.3 基于系统动力学研究数据监护可行性

数据监护是一个由众多因素构成的复杂的、具有多重信息反馈的非线性系统, 各因素相互作用决定了系统结构及系统最终输出, 协调好系统内部各种资源以寻求较优系统结构与行为。因此, 数据监护的运行机制研究需要利用系统论知识与原理来分析影响因素间相互关系, 以及拟合敏感因素影响效果及其变化趋势。

系统动力学模型能够清晰地描述研究对象的复杂关系及多要素间相互关系的动态属性。这是对结构描述和变化趋势的模拟, 适合研究复杂系统动态变化的问题。另外, 系统动力学属于定性与定量相结合的方法, 可以用来处理复杂系统问题以及分析系统中非线性因素的作用机制, 以寻求较优的系统结构。根据系统动力学方法的用途, 系统动力学方法具备处理非线性问

题、信息反馈、时间滞延、动态性复杂的能力, 这些与数据监护系统内部运行机制问题相吻合, 从动态与系统的视角分析数据监护影响因素的作用效果具有其可行性。鉴于此, 笔者尝试将系统动力学知识引入数据监护研究, 通过构建系统模型来仿真模拟数据监护系统的影响因素、作用效果及整体服务能力 (数据监护效应) 的变化趋势, 为数据利益相关者配置数据监护系统提供参考与借鉴。

2 数据监护的系统动力学框架分析

2.1 数据监护的主体分析

根据数据认证标准DSA (Data Seal of Approval)^[23]与数据监护协作模型^[24], 数据监护系统涉及3个利益相关者, 即数据作者、数据用户与数据监护人员。数据作者负责科研数据创建质量, 数据监护人员负责数据存储质量, 数据用户负责数据使用质量。因此, 数据监护系统的作用主体包括数据作者、数据用户与数据监护人员。

从数据监护协作模型 (见图1) 可以看出, 数据监护系统是由工具、政策与技能及数据利益相关者构成的统一体。数据监护系统正常运行需要数据用户、数据作者及数据监护人员协同工作, 利用政策与机制明确利益相关者的职责与义务, 作用主体在数据监护中的角色与职责如下。

(1) 数据作者。数据作者负责科研数据质量。数据作者根据标准与政策将数据及元数据存储到数据知识库中。数据作者的工作包括数据文件与相关元数据提供、文件格式转换、协同工作、数据共享、数据使用跟踪等。

(2) 数据监护人员。数据监护人员负责数据存储质量。数据监护人员遵守标准、法律规定、合同、政策条例, 承担数据评估、处理、存档、发布、长期保存使命, 利用技术基础设施来促进数据发现、利用与引用, 保障数据与元数据的完整性、真实性与有效性。因此, 数据监护人员承担服务可持续性、数据安全性、数据可见性、技术适用性、数据长期有效性的任务。

(3) 数据用户。数据用户负责数据使用质量, 遵守数据政策、法规、访问许可与准则来访问与再利用数据。数据用户的职责包括数据引用、使用协议许可遵守、衍生物链接提供等方面。

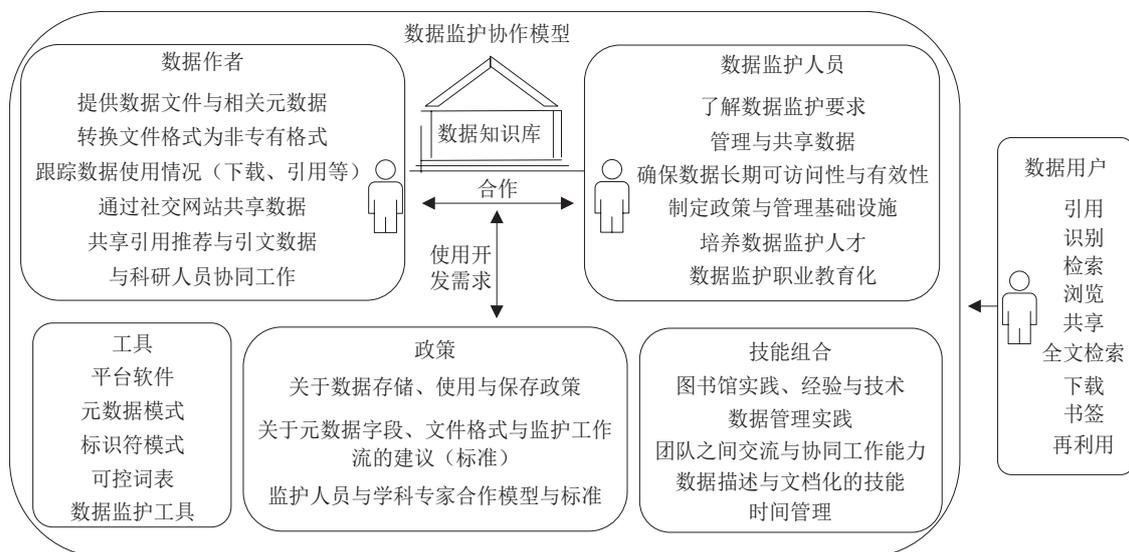


图1 数据监护协作模型

2.2 数据监护的影响因素分析

在明确数据监护作用主体的基础上，要确定数据监护影响因素，为数据监护系统动力学模型的构建奠定基础。数据监护系统动力学模型使用变量“数据监护效应”代表数据监护系统的整体能力^[25]。

根据机构数字资产评估工具（assessing institutional digital assets, AIDA）^[26]、《可信数字知识库属性与责任》（*Trusted Digital Repositories: Attributes and Responsibilities*）^[27]研究报告、《可信数字知识库审计与认证：标准与清单》（*the Trustworthy Repositories Audit & Certification: Criteria and Checklist*）^[28]，提炼出数据监护的影响因素。其关键影响因素包括数据管理、组织架构与技术基础设施三大部分。组织架构是数据监护实现与可持续发展的基础，数字对象管理是数据监护工作的重心和目标，而技术基础设施则为数据监护提供保障，三大部分的具体指标相互联系，形成不可分割的有机整体。

数据管理包括科研协助、质量控制与访问管理等内容。科研协助涉及数据管理计划的板块与数据记录收集工具提供、在线数据处理与分析、在线帮助与指导等；质量控制包括验证机制、内容真实性与完整性审核、存档信息包创建、描述信息的获取、描述元数据获取与创建等方面；访问管理涉及用户权限管理、出版物与衍生出版物统计与跟踪、政策与存储协议的一致性、对象副本传播的管理等方面。

组织架构涉及资源配置、组织规划与政策标准执行等内容。其中资源配置包括在岗员工技能与经验要求、合理员工配置、财务状况说明、资金来源说明等；组织规划涉及财务计划、业务计划、应急与继任计划、员工职业发展规划、人才培养计划等；政策标准执行包括遵守可持续性服务规定、免责声明、存储协议、创作共用许可、附加许可、数据收集、存储、更新与访问控制政策等。

技术基础设施包括系统功能、技术标准匹配与风险管理等内容。系统功能包括元数据管理（保存、导出与控制）、数据管理（检索、浏览、在线处理与分析、发现、关联等）、数据跟踪、安全保障、支持与帮助等方面；技术与标准匹配包括软硬件匹配、技术与标准匹配、标准与政策的一致性、技术与流程适用性、技术与服务适用性；风险管理包括离线备份、风险分析、服务持续性分析、技术退化的应对策略、安全需求分析、员工安全责任明确等方面。

3 数据监护的系统动力学模型构建

3.1 模型边界的界定与基本假设

3.1.1 模型边界的界定

系统框图描述系统与组成部分之间关系及重要变量与有关变量之间的关系，用于系统分析与系统结构分

析的初步阶段,有助于明确系统边界。本文的研究内容主要针对单个数据监护系统,不考虑数据监护系统之间的互操作与相互影响。由数据监护的影响因素分析结果可得,数据监护系统主要包括组织架构、数据管

理、技术基础设施三大要素模块,这三大要素模块彼此相互作用与影响,共同决定数据监护系统输出。系统边界及影响要素见图2。

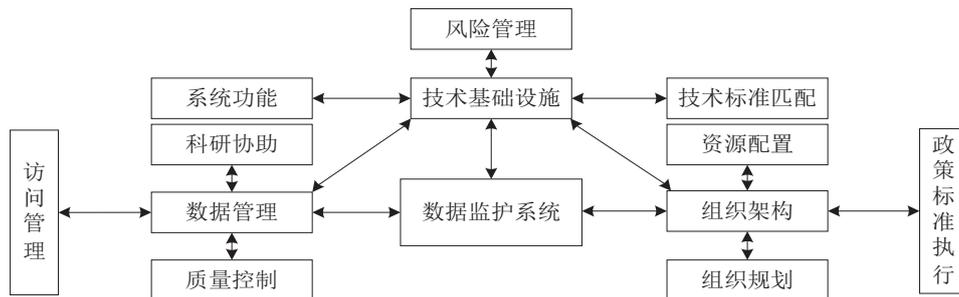


图2 数据监护系统框图

3.1.2 模型的基本假设

鉴于社会系统的复杂性,无法将系统的所有要素都考虑,为了限定数据监护系统研究范围和确定系统关键因素,系统模型的基本假设如下。

假设1:模型中将辅助变量“数据监护效应”定义为受系统模型中众多因素影响的结果,或为系统模型的输出,或为数据监护系统整体能力。

假设2:模型中辅助变量“数据监护效应”主要受组织架构能力、数据管理能力与技术基础设施能力三大要素影响,技术基础设施影响作用较大,整个系统的因变量与自变量的相互作用关系符合柯布-道格拉斯生产函数。

假设3:组织架构能力、数据管理能力、技术基础设施能力三大要素的初始值为20、最大值为100,状态变量的取值范围为 $[0, 100]$,根据生产函数中要素产出最大值的计算,辅助变量“数据监护效应”的取值范围为 $[0, 127]$ 。

假设4:当数据监护效应大于80时,领导支持、投资力度降低到日常维护费用或政策支持费用水平,而数据质量维持在较高水平;当数据监护效应小于80时,领导支持、投资力度与数据质量随着数据监护效应提升而增长。

假设5:国家与机构政策支持数据监护系统构建及发展,每年以最低标准来支持数据监护发展,从而促进数据监护系统平稳发展。不考虑政策重大变动及其他非正常情况所导致的系统崩溃情形。

假设6:系统重点考虑数据监护人员的作用,数据

用户与数据作者对数据监护系统产生间接影响,数据用户与数据作者都具有较高的数据素养与共同目标。

假设7:服务持续性、数据质量、环境依赖性、数据安全随着数据监护整体能力取值范围不同,相应设定条件有所不同,也就是说,数据监护整体能力决定了服务持续性、数据质量、环境依赖性与数据安全的水平。

3.2 因果关系分析

因果关系图又称为因果回路图,是一种定性描述系统变量之间关系的工具。数据管理、组织架构、技术基础设施是影响数据监护的关键模块。其中数据管理模块是数据监护系统的关键,包括数据质量控制、科研协助与访问管理等方面;组织架构模块是数据监护的前提,主要指资源配置效率、组织规划能力与政策标准执行程度等方面;技术基础设施是数据监护的基础,包括技术与标准匹配、系统功能与风险管理等方面。

从数据管理能力的因果关系(见图3)来看,数据管理能力有利于提升数据监护效应。正反馈回路:数据管理能力—数据监护效应—数据质量—质量控制—数据管理能力。数据管理能力正向影响数据监护效应,数据监护效应正向影响数据质量高低,长期来看,数据质量高低影响质量控制,质量控制与科研协助正向影响数据管理能力。负反馈回路:数据管理能力—数据安全—开放存取—访问管理水平—数据管理能力。数据管理能力与数据安全为正向关系,数据安全与权益保护负向影响开放存取,开放存取与访问管理水平是正向关系,而访问管理水平正向影响数据管理能力。

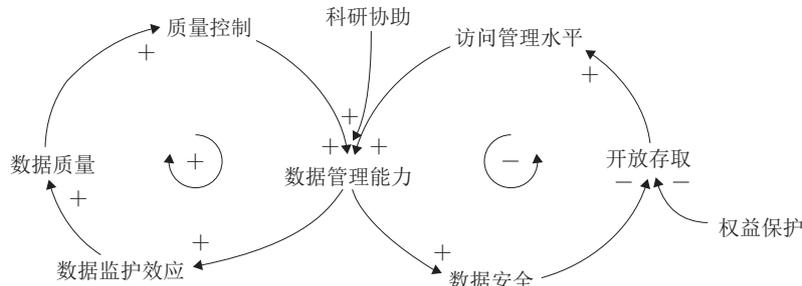


图3 数据管理能力的因果关系图

从组织架构能力的因果关系(见图4)来看,组织架构能力是关于数据监护的准备与前提。正反馈回路:组织架构能力—数据监护效应—组织领导支持—资源配置效率—组织架构能力。组织架构能力正向影响数据监护效应,数据监护效应、组织服务要求与组织领导支持是正向关系,组织领导支持提升资源配置效率,资源配置效率与组织规划能力正向影响组织架构能

力。负反馈回路:组织架构能力—环境依赖性—组织联盟—政策标准一致性—政策标准执行程度—组织架构能力。组织架构能力与环境依赖性为负向关系,环境依赖性与政策支持正向影响组织联盟,资源限制与组织联盟为负向关系,组织联盟影响政策标准一致性,政策标准一致性与技术壁垒影响政策标准执行程度,最终影响组织架构能力。

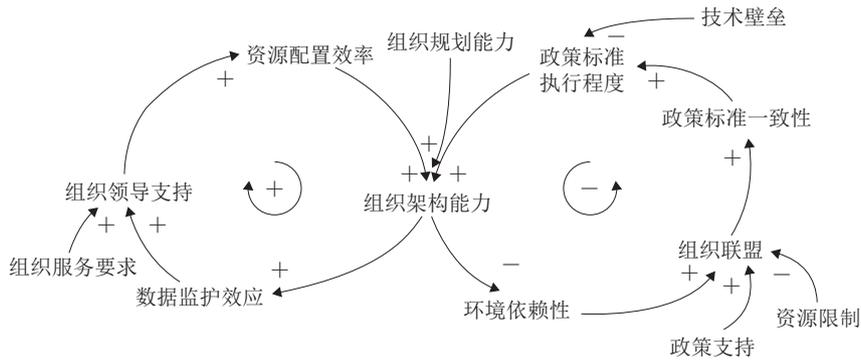


图4 组织架构能力的因果关系图

从技术基础设施能力的因果关系(见图5)来看,技术基础设施是数据监护的基础保障。正反馈回路:技术基础设施能力—数据监护效应—投资力度—技术更新—技术与标准匹配—技术基础设施能力。技术基础设施能力正向影响数据监护效应,数据监护效应、政策支持与投资力度是正向关系,投资力度影响技术更新,技术更新正向影响技术与标准匹配,技术标准匹配与系统功能正向影响技术基础设施能力。负反馈回路:技术基础设施能力—服务持续性—风险管理难度—技术基础设施能力。技术基础设施能力与服务持续性是正向关系,服务持续性、技术退化与防灾能力影响风险管理难度,风险管理难度负向影响技术基础设施能力。

技术更新正向影响技术与标准匹配,技术标准匹配与系统功能正向影响技术基础设施能力。负反馈回路:技术基础设施能力—服务持续性—风险管理难度—技术基础设施能力。技术基础设施能力与服务持续性是正向关系,服务持续性、技术退化与防灾能力影响风险管理难度,风险管理难度负向影响技术基础设施能力。

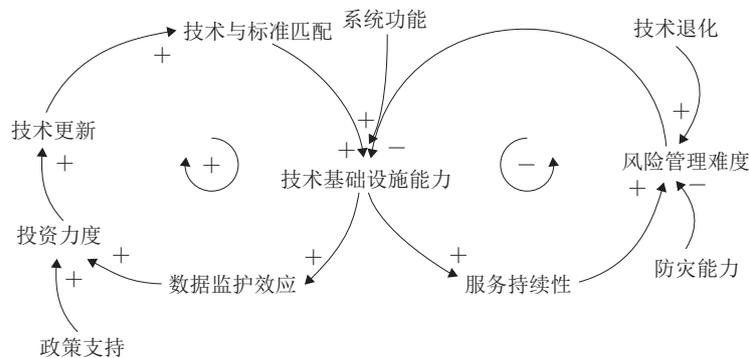


图5 技术基础设施能力的因果关系图

从3张因果关系图看出,数据对象管理能力、技术基础设施能力、组织框架能力是影响数据监护效应的关键因素,借助因果关系图对其影响机理进行详细分析。数据管理能力主要由质量控制、科研协助与访问管理水平三方面影响;组织架构能力主要由组织规划能力、资源配置效率、政策标准执行程度三方面影响;技术基础设施能力主要由系统功能、技术与标准匹配及风险管理难度三方面影响。

3.3 系统流图构建

在因果关系图的基础上,数据监护系统动力模型分为组织架构、数据管理、技术基础设施三大子模块。数据监护的系统模型中包括3个水平变量、6个速率变

量、17个辅助变量及10个常量,其中3个水平变量包括数据管理能力、组织架构能力、技术基础设施能力,6个速率变量包括数据管理能力增强、数据管理能力减弱、组织架构能力增强、组织架构能力减弱、技术基础设施能力增强、技术基础设施能力减弱,17个辅助变量包括数据监护效应、开放存取、技术与标准匹配、技术更新、投资力度、政策标准执行程度、数据安全、数据质量、服务持续性、环境依赖性、组织领导支持、访问管理水平、组织联盟、质量控制、资源配置效率、风险管理难度、政策标准一致性,10个常量包括系统功能、组织服务要求、政策支持、资源限制、防灾能力、技术退化、组织规划能力、科研协助、技术壁垒、权益保护。在系统变量设定的基础上,数据监护的系统流图绘制见图6。

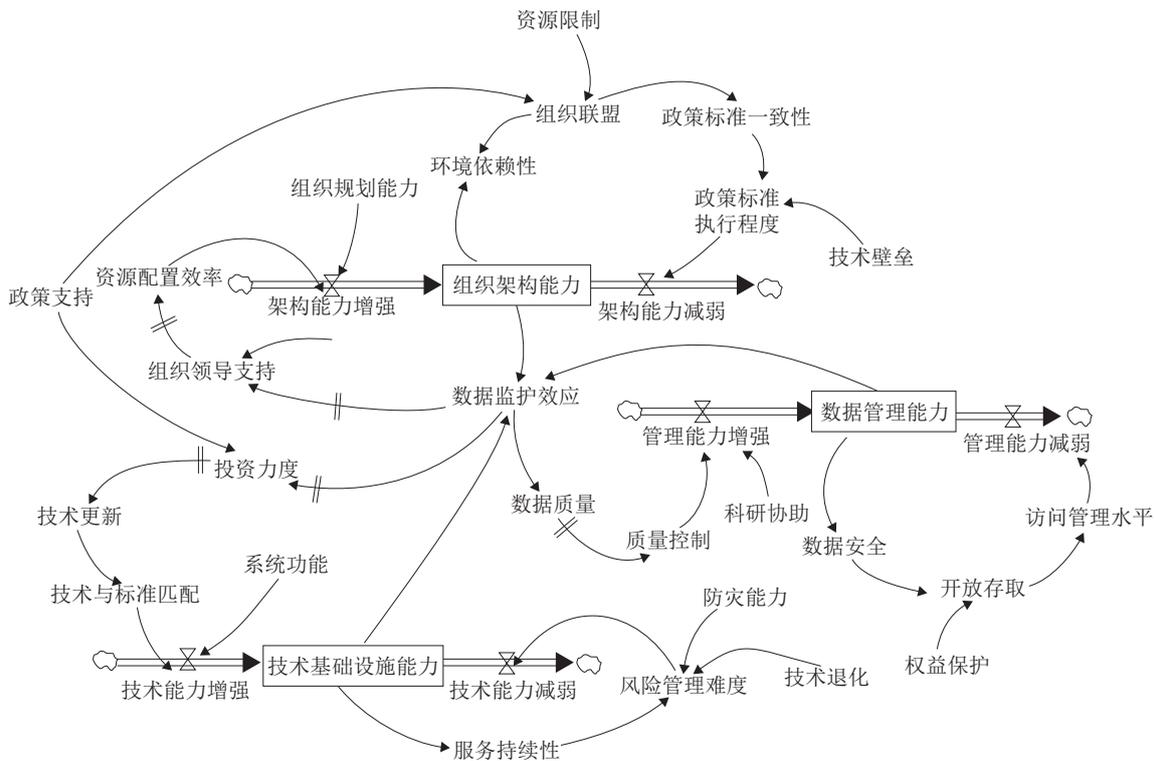


图6 数据监护的系统流图

3.4 关键变量函数式确定

柯布-道格拉斯生产函数是由美国数学家柯布 (C.W. Cobb) 与经济学家保罗·道格拉斯 (Paul H. Douglas) 于20世纪30年代共同研究与创造出来的。生产函数是探讨投入与产出关系的经济学模型,反映的是在现有技术水平下某一特定要素投入组合的最大产出^[29],是

用来预测工业系统生产情况的一种数学模型,在数理经济学与计量经济学研究与应用中占有重要地位。

根据数据监护系统流图,数据监护效应为组织架构、数据对象管理与技术基础设施三大模块间动态与互动、有机耦合的结果。数据监护与要素模块(组织架构、数据管理、技术基础设施)的作用关系符合柯布-道格拉斯生产函数。

数据监护系统模型中自变量与因变量的关系符合柯布-道格拉斯生产函数^[30-31]。将数据监护效应抽象为三大模块产出,借助经典的柯布-道格拉斯生产函数对其描述,柯布-道格拉斯生产函数见公式(1)^[32]。该生产函数表示在既定技术水平下生产要素组合 $\chi_1, \chi_2, \chi_3 \dots \chi_n$ 在每个时刻下最大产出为Y。公式(1)中 $\chi_1, \chi_2, \chi_3 \dots \chi_n$ 代表生产要素(劳动、资本、土地、才能等)的投入量,根据柯布-道格拉斯生产函数中规模假定, $\alpha + \beta + \lambda > 1$ 代表规模报酬递增的生产函数, $\alpha + \beta + \lambda = 1$ 代表规模报酬不变的生产函数, $\alpha + \beta + \lambda < 1$ 代表规模报酬递减的生产函数, α, β, λ 分别代表各个因要素产出的弹性系数, μ 代表随机干扰影响, $0 \leq \mu \leq 1$, A_0 代表综合技术水平, Y代表工业产出。

$$Y = \mu A_0 \chi_1^{\beta_1} \chi_2^{\beta_2} \chi_3^{\beta_3} \dots \chi_n^{\beta_n} \quad (1)$$

数据监护是一个多变量共同作用的、螺旋式上升的动态过程。因变量(数据监护效应)抽象为组织架构能力、数据管理能力、技术基础设施能力三大要素的产出,见公式(2)。

$$DC = \mu A I^\alpha M^\beta T^\lambda \quad (2)$$

因变量DC代表数据监护效应(即数据监护整体能力), μ 代表随机干扰影响, A代表受3个要素模块影响数据监护效应指数(3个模块影响数据监护效应的转化率), 3个自变量I、M、T分别代表组织的架构能力、数据对象管理能力、技术基础设施能力, α, β, λ 分别代表组织架构能力、数据管理能力、技术基础设施能力的作用系数。在不影响数据监护效应的趋势变化与可比性条件下,公式(2)中 $A=0.8, \mu=1$;为了体现组织架构能力、数据管理能力、技术基础设施能力3个要素模块规模报酬递增效应,取 $\alpha + \beta + \lambda > 1$,公式(2)中 α, β, λ 分别取0.3、0.3、0.5。

4 模型检验与模拟

4.1 系统主要方程式与初始值设置

系统动力学模型初始值的3种设置方法包括历史数据拟合法、平衡态赋值法、特殊增长规律赋值法。数据监护的3个模块是根据国际数据监护标准的指标体系划分的,由于数据监护的3个模块抽象程度较高,系统变量很难根据现实数据进行赋值。因此,本系统模

型采用平衡态赋值法,对模型中各变量赋以相对基准初始值,利用模拟仿真手段对参数设定前后的变化情况来分析变化趋势的可比性与真实性,并利用系统模型预测功能来分析数据监护的发展趋势与敏感参数的影响效果^[33]。

在模拟运算与模拟过程中对所有变量均进行无量纲化处理,数据监护效应的取值范围为[0, 127](根据生产函数中3个要素产出最大值的运算结果),3个水平变量组织架构能力、数据管理能力、技术基础设施能力初始值设为20,取值范围为[0, 100];其他速率变量、辅助变量与常量的取值范围均设定为[0, 1];数据监护效应公式 $DC = \mu A I^\alpha M^\beta T^\lambda$ 中参数取值为: $\mu=1, A=0.8, \alpha=0.3, \beta=0.3, \lambda=0.5, I=20, M=20, T=20$ 。变量初值代表数据监护系统模型的初始状态,根据不同实际情况设定初始值,参数值的设置体现了要素模块的规模效应报酬递增。系统模型的仿真时间单位设定为季度,仿真时间为40个季度,步长为1个季度,仿真开始时间与仿真结束时间分别为0与40个季度。表1列举了数据监护系统模型的变量名称、类型、取值范围及方程式或说明,其中对状态变量、辅助变量、速率变量设置了方程式。另外,根据系统初始状态设置了常量初始值,系统功能=0.2,组织服务要求=0.2,政策支持=0.2,资源限制=0.1,防灾能力=0.3,技术退化=0.3,组织规划能力=0.2,科研协助=0.2,技术壁垒=0.1,权益保护=0.2。

4.2 模型检验

在模型量纲一致性、行为异常与边界适当性等检验的基础上,需要对模型进行有效性检验。模型有效性检验以理论检验为主,着重检验模型结构有效性、一致性与适用性。在已有系统流图的基础上,检验模型不同时期各变量的变化情况,通过分析数据监护中各个模块中水平变量仿真结果与理论预期值差异情况来检验系统动力学模型的有效性。

图7选取了不同时期的3个模块水平变量与辅助变量“数据监护效应”的数值指标,比较分析4个变量的代表性数据。对图7中4个变量(1个辅助变量与3个水平变量)不同时期的数据比较分析看出:数据监护效应、技术基础设施能力、组织架构能力这3个系统变量的变化趋势基本相同,从0~25个季度,3个变量处于上升趋势,数据监护效应、技术基础设施能力、组织架构能力在第25个季度分别达到最大值92.0、75.3、79.5。从第25

表1 系统模型中主要方程

类型	变量	取值范围	方程式或说明
状态变量	组织架构能力	20 [0, 100]	组织架构能力= INTEG (组织架构能力×(架构能力增强-架构能力减弱), 20)
状态变量	数据管理能力	20 [0, 100]	数据管理能力= INTEG (数据管理能力×(管理能力增强-管理能力减弱), 20)
状态变量	技术基础设施能力	20 [0, 100]	技术基础设施能力= INTEG (技术基础设施能力×(技术能力增强-技术能力减弱), 20)
速率变量	架构能力增强	[0, 1]	架构能力增强=组织规划能力+资源配置效率
速率变量	架构能力减弱	[0, 1]	架构能力减弱=1-政策标准执行程度
速率变量	管理能力增强	[0, 1]	管理能力增强=科研协助+质量控制
速率变量	管理能力减弱	[0, 1]	管理能力减弱=1-访问管理水平
速率变量	技术能力增强	[0, 1]	技术能力增强=技术与标准匹配+系统功能
速率变量	技术能力减弱	[0, 1]	技术能力减弱=风险管理难度
辅助变量	数据监护效应	[0, 127]	数据监护效应=INTEGER (0.8×(组织架构能力^0.3)×(数据管理能力^0.3)×(技术基础设施能力^0.5))
辅助变量	开放存取	[0, 1]	开放存取=1-数据安全-权益保护
辅助变量	技术与标准匹配	[0, 1]	技术与标准匹配=技术更新
辅助变量	技术更新	[0, 1]	技术更新=SMOOTH (投资力度, 4)
辅助变量	投资力度	[0, 1]	投资力度=SMOOTH (IF THEN ELSE (数据监护效应/100>0.8, 政策支持, 数据监护效应/100), 6)
辅助变量	政策标准一致性	[0, 1]	政策标准一致性=组织联盟
辅助变量	政策标准执行程度	[0, 1]	政策标准执行程度=政策标准一致性-技术壁垒
辅助变量	数据安全	[0, 1]	数据安全=IF THEN ELSE (数据管理能力/100<0.8, 数据管理能力/100, 0.8)
辅助变量	数据质量	[0, 1]	数据质量=IF THEN ELSE (数据监护效应>80, 0.8, 数据监护效应/100)
辅助变量	服务持续性	[0, 1]	服务持续性=IF THEN ELSE (技术基础设施能力<80, 技术基础设施能力/100, 0.8)
辅助变量	环境依赖性	[0, 1]	环境依赖性=1-IF THEN ELSE (组织架构能力/100>0.8, 0.8, 组织架构能力/100)
辅助变量	组织领导支持	[0, 1]	组织领导支持=SMOOTH (IF THEN ELSE (数据监护效应/100>0.8, 组织服务要求, 数据监护效应/100), 4)
辅助变量	访问管理水平	[0, 1]	访问管理水平=开放存取
辅助变量	组织联盟	[0, 1]	组织联盟=环境依赖性+政策支持-资源限制
辅助变量	质量控制	[0, 1]	质量控制=SMOOTH (数据质量, 5)
辅助变量	资源配置效率	[0, 1]	资源配置效率=SMOOTH (组织领导支持, 4)
辅助变量	风险管理难度	[0, 1]	风险管理难度=服务持续性-防灾能力+技术退化

个季度以后,这3个变量呈现小幅度上下波动趋势,第34个季度达到最低值分别为73.0、57.2、53.3,这9个季度中3个变量的波动幅度分别为20.6%、25.0%、33.3% (见表2)。这是因为随着时间推移技术基础设施受到技术退化、风险管理难度、基础设施老旧等因素影响来带动技术基础能力的小幅度下降;资源限制(人才紧缺、技能缺乏、技术壁垒等)、资源配置率、组织架构等影响组织架构能力小幅度下降,继而出现继续上升趋势,技术基础设施与组织架构能力长期呈现上下波动趋势;另外,数据监护效应变化趋势受到技术基础

设施、组织架构能力与数据管理能力影响,而技术基础设施能力对数据监护效应具有重要影响,数据监护效应的变化趋势与技术基础设施能力同步且相似。因此,数据监护效应、技术基础设施能力、组织架构能力的变化趋势符合预期发展态势。

数据管理能力在前5个季度前处于平稳状态,从第5个季度到第34个季度,数据管理能力呈现上升趋势,第34个季度达到最高值78.1,数据管理能力从第34个季度开始基本上处于平稳状态,也就是说,数据管理能力刚开始两年内基本不变,中间呈现上升趋势,随着时

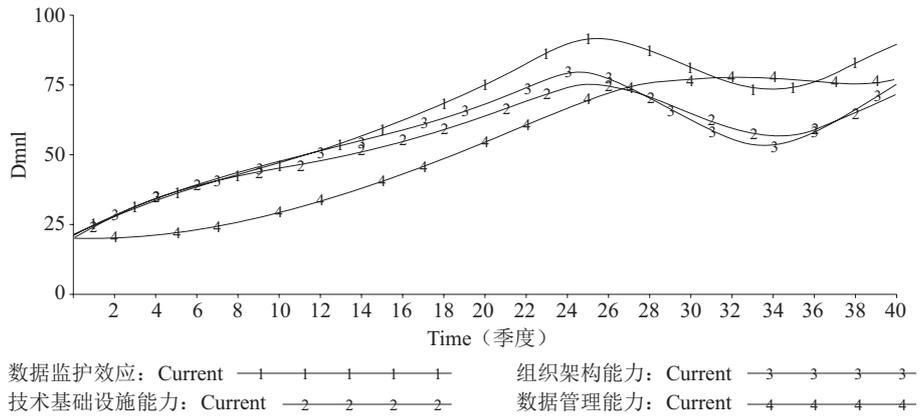


图7 1个辅助变量(数据监测效应)与3个水平变量的变化趋势

表2 数据监测效应与主要影响要素的仿真数值
(单位: 无量纲化处理)

时间 (季度)	数据监测 效应	技术基础 设施能力	数据管理 能力	组织架构 能力
0	21.0	20.0	20.0	20.0
1	25.0	24.2	20.2	24.2
2	28.0	28.2	20.3	28.2
14	56.0	51.5	38.1	54.3
15	59.0	53.4	40.6	56.4
16	62.0	55.4	43.2	58.6
24	90.0	74.5	67.2	79.5
25	92.0	75.3	69.9	79.5
26	91.0	74.7	72.0	77.1
34	73.0	57.2	78.1	53.3
35	74.0	57.2	76.9	54.8
39	86.0	67.9	76.1	71.1
40	90.0	71.3	76.8	75.5

间推移,数据管理能力处于平稳发展态势。数据管理能力受质量控制、科研协助与访问管理的影响,在时间转移过程中累积经验、学习技能与改进方法,从而逐步提高数据管理能力,最终促使数据管理能力趋于平稳状态。因此,数据管理能力发展趋势符合现实情况。

通过对模型有效性分析,模型仿真输出值与实际情况预期值之间拟合情况良好,因此判定模型是科学有效的。本模型较强地表达与识别出不同时期数据监测效应的变化特征,与现实中数据监测能力提升的各个阶段表现出来的特征基本相符,能够据此判断数据监测效应发展趋势与敏感参数的影响效应。

4.3 模型模拟

数据监测的系统动力学模型中的诸多变量均对数据监测效应产生不同程度影响,为保持数据监测系统结构合理与不断优化,既要注重各模块内部单一因素的作用,也要注重整个系统框架下多个模块与多因素间的组合及联动作用。由于篇幅限制,本文着重进行多因素灵敏性仿真分析,多因素灵敏性分析包括两种情况,即限制性因素联合作用分析与促进性因素联合作用分析。

4.3.1 减弱系统限制性因素联合作用

系统限制性因素包括资源限制、技术壁垒、技术退化、权益保护4个,4个因素各增加0.05(见图8)。其中标号2曲线代表初始曲线,标号1曲线代表资源限制、技术壁垒、技术退化与权益保护限制性因素各增加0.05后的曲线,当限制性因素共同作用时,通过增加限制性因素的参数值来观察系统整体与局部变化。由图8看出,4个变量的总体变动趋势没有发生根本性变化,相比系统初始曲线,4个变量呈现更加平稳上升趋势,曲线拐点相应滞后。在这种情况下,数据监测效应增长缓慢,曲线拐点滞后14个季度。数据监测效应的曲线发生显著性变化,这种影响效果远大于单因素作用效果。

由此可见,系统限制性因素对数据监测效应产生明显负向影响。在实践中,数据监测效应提升需降低资源限制、技术壁垒、技术退化、权益保护因素的负向影响,实现较短周期内抵达数据监测效应峰值,长期内实现数据监测效应平稳。

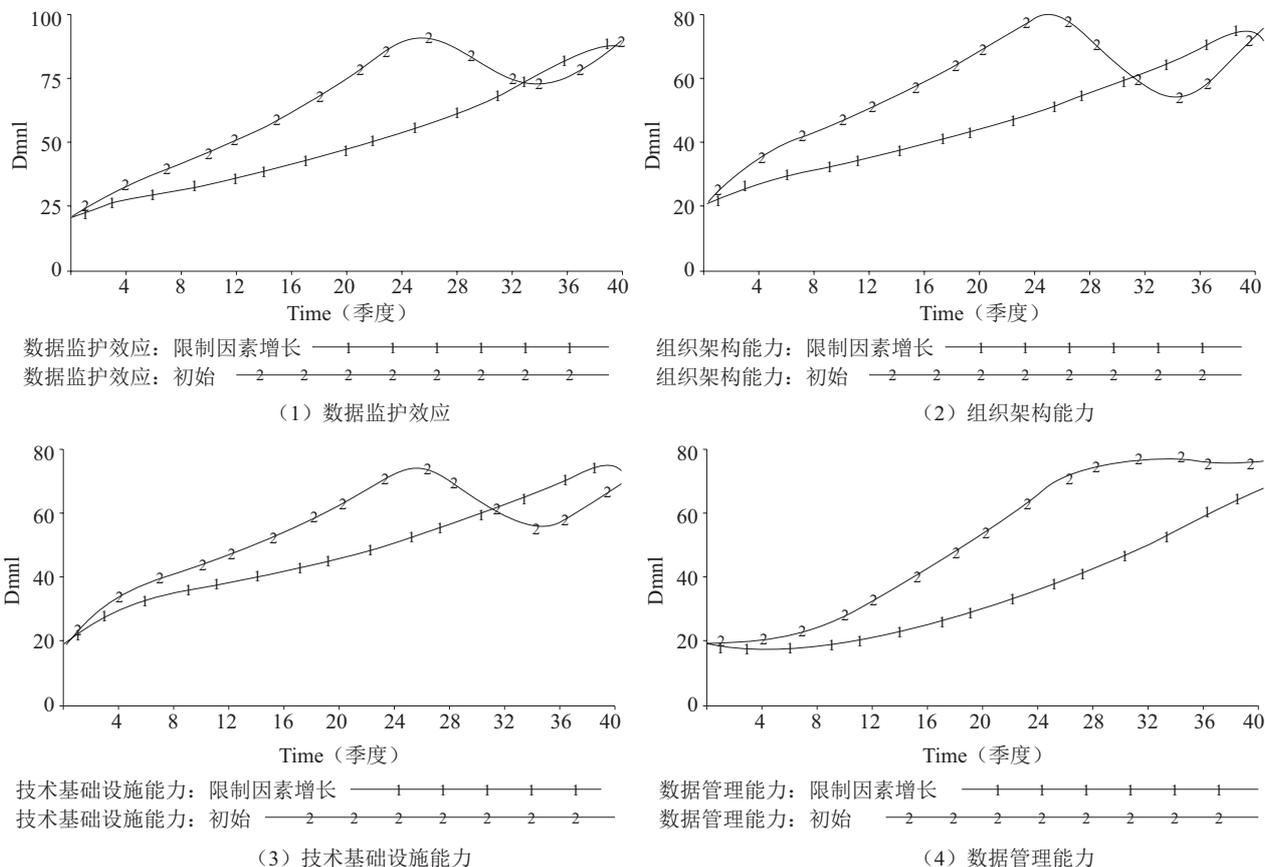


图8 数据监护的限制性因素增加0.05前后曲线

4.3.2 增强系统促进性因素联合作用

系统6个促进性因素包括防灾能力、系统功能、组织服务要求、组织规划能力、科研协助、政策支持。6个促进性因素均降低0.05,其中标号2曲线代表初始曲线,标号1曲线代表防灾能力、系统功能、组织服务要求、组织规划能力、科研协助与政策支持促进性因素均降低0.05后的曲线(见图9)。

促进性因素共同作用时,通过降低促进性因素的参数值来观察系统整体与局部变化。系统中4个变量变动趋势发生根本性变化,这种影响效果远大于单因素影响:促进性因素值增大0.05后,4个变量在40个季度的仿真周期内均呈现更加平稳上升趋势,不再出现先上升后降的趋势,改变了促进性因素的参数值的初始仿真结果,整体上大幅度拉低了数据监护效应,改变数据监护效应初始状态曲线的变动态势,上升趋势更为平缓,曲线拐点相应滞后,仿真周期内数据监护效应未出现峰值,一直处于平稳上升趋势。这一结果表明,系统促进性因素对数据监护效应具有促进作用,

减弱系统促进性因素作用对数据监护效应产生显著抑制效果。

由此可见,系统促进性因素对数据监护效应产生明显正影响。在实践中,数据监护效应提升需从防灾能力、系统功能、组织服务要求、组织规划能力、科研协助、政策支持6个方面采取措施。系统中4个变量变动趋势产生根本性变化,6个因素共同作用远大于单个因素的影响效果,延长了峰值到达时间,降低了峰值。在实践中,针对6个影响因素的复合作用采取政策措施来提高数据监护效应,并维持长期发展平稳性。

从图8与图9可见,系统4个限制性因素作用效果小于6个促进性因素,一方面是由于因素之间的复合作用不同,另一方面是由于因素的当前值与数量的不同,两方面共同决定限制性因素作用效果小于促进性因素作用效果。在实践中,数据监护效应提升需要减弱限制性因素影响(资源限制、技术壁垒、技术退化、权益保护)并增强促进性因素(防灾能力、系统功能、组织服务要求、组织规划能力、科研协助、政策支持)作用效果,一定周期内促使数据监护效应达到较高峰值,长期

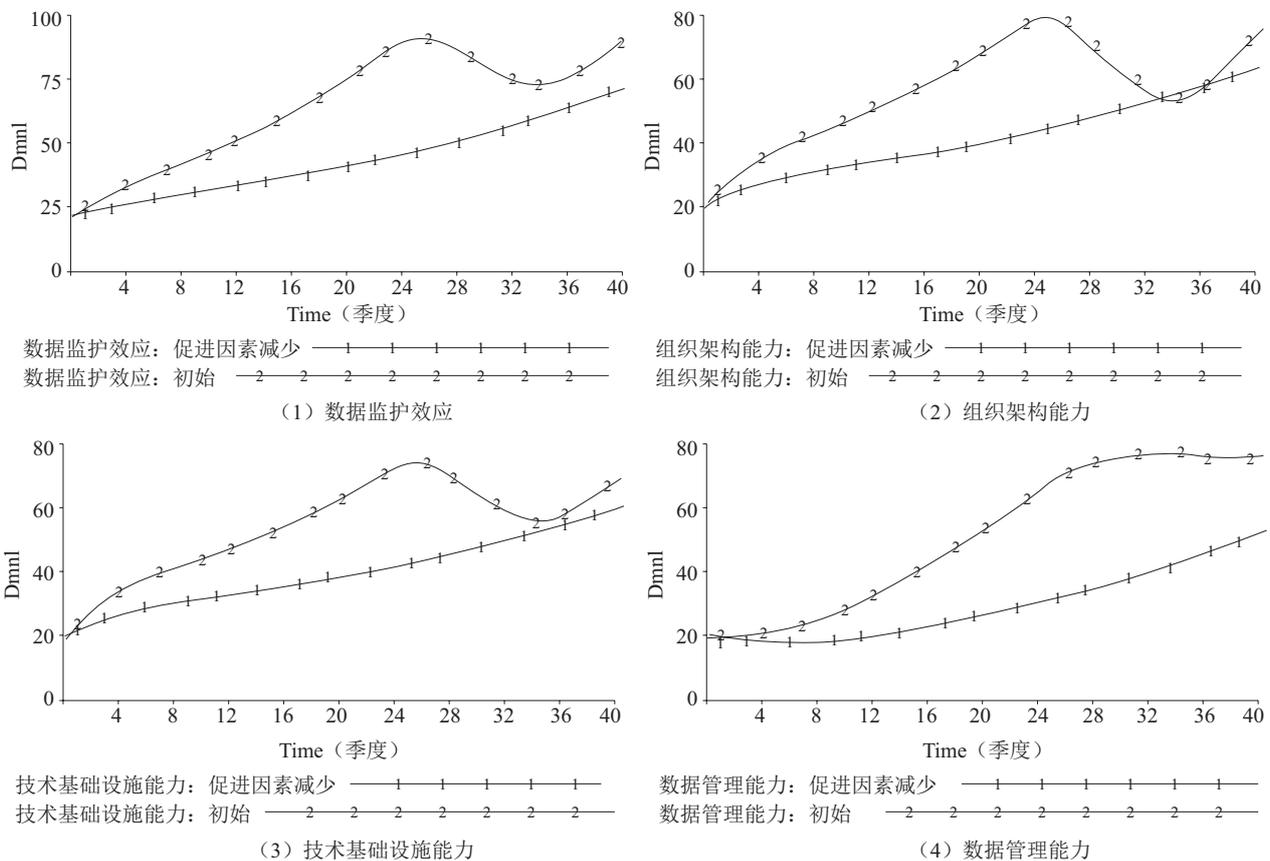


图9 数据监护的促进性因素降低0.05前后曲线

内促进数据监护效应呈现小幅度上下波动趋势，也就是说，短期达到最大值，长期实现平稳发展。

仿真结果表明，技术壁垒、技术退化、资源限制、政策支持、防灾能力等系统可控性外部变量与数据监护效应存在较强的联动关系，影响效果较强，应从可控外部变量着手采取相应的政策措施与建议来提升数据监护效应。

5 对策建议

通过对数据监护系统的仿真结果运用分析，发现可以从国家与组织支持力度、组织策划能力、发展障碍、服务能力方面的出台政策方案来提升数据监护能力。数据监护能力提升策略是对政策方案的延伸与深化，其具体的提升策略涉及政策支持力度、知识产权保护、数据监护人员技能、平台功能、数据可信度、数据互操作、数据长期保存能力、嵌入式科研流程服务质量8个方面。

(1) 加大政策支持力度。政策支持力度体现在国

家政策与机构政策支持两个方面。国家政策体现为国家对高校或科研机构数据监护实践的资金、人才、基础设施、数据开放存取等方面支持；组织政策支持体现为组织对数据监护系统构建与开发经费的支持，对数据共享的鼓励，对人才培养、机构领导与员工数据素养提升、协作能力提升等方面的重视^[5]。

(2) 建立知识产权保护机制。从国家与数据监护平台这两方面建立数据产权保护机制。国家须完善数据保护的政策体系，政策体系涉及数据共享政策、数据使用政策、数据治理政策等；数据监护平台的具体数据政策包括用户条款、数据存储条款、数据使用条款、数据备份及保存条款等，利用知识产权保护机制合理协调数据利益相关者间关系。

(3) 提升数据监护人员知识技能。数据监护人员包括数据作者、技术支持人员、图书馆员、数据科学家、信息专业人员等。通过教育培训（专业课程、实习、研究会等形式）来提升数据监护人员知识技能的具体措施有：第一，图书情报院校开设数据监护相关的正规课程；第二，岗前就业准备相关实习、见习课与实地

考察；第三，建立国内外数据监护交流社区^[13]。

(4) 完善数据监护平台功能。数据监护平台功能有待完善的方面包括质量管理、访问管理、嵌入式服务、数据互操作、平台透明性、作者标识符采用、数据在线处理、数据关联等。在借鉴国外数据监护平台构建的最佳实践基础上，需要利用技术创新、专业技能、政策等措施来提升数据监护平台功能。

(5) 提升数据可信度。数据可信度来自平台功能与利益相关者信任两方面。其中，平台功能包括政策体系、服务质量、系统化流程；利益相关者信任包括组织信任（完整性、透明性、组织信誉、识别性）、社会因素（同事影响）、机构保障（可持续性、机构信誉、第三方认证）、专业水平及学科影响。因此，从平台功能与利益相关者信任两方面来提升数据可信度。

(6) 增强数据互操作性。数据互操作性提升须从两方面采取措施：一方面，遵守国际通用标准来开发、构建及发展数据监护平台以实现数据互操作；另一方面，数据集成与无缝化服务需要有效接口和相应的开放系统以增强系统可操作。因此，从标准化平台、接口及开放系统来增加数据互操作性。

(7) 增强数据长期保存能力。数据长期保存能力是数据可持续利用的关键因素，数据长期保存能力主要表现在组织规划能力与防灾能力两个方面。防灾能力提升须从风险与威胁分析、风险评估、防灾准备、灾难恢复、自我评估等方面来采取措施；规划能力提升须制定与实施长期服务计划、继任与应急计划、人员配置计划、自我评估计划、经营计划、防灾计划等，并在实践中不断地更新与维护计划以保证计划实施的有效性。

(8) 提升嵌入式科研流程服务质量。嵌入式科研流程服务质量提升的措施是构建将数据监护、科研活动与数据知识库融为一体的新型管理与服务模式。数据监护人员利用数据监护平台实时为科研人员提供咨询与指导，以及数据、工具、方法与相关资料等服务，协助科研人员解决科研过程的各种问题。二者的协同工作将加快科研进程，提升科研效率。

本文建立在单个高校或科研机构数据监护过程的系统动力学模型基础上，分析了仿真模型的有效性和主要参数的灵敏度，模型较好地数据监护过程进行拟合，揭示了单个机构数据监护的特征及机理，为高校或科研机构开展数据监护实践提供依据。由于真实的数据监护是一个较复杂与抽象的深化过程，本模型存在一定的局限性，只对理论数据进行仿真模拟；另外，

本模型结构较简单，只考虑了有限的变量与参数，较为复杂的、动态的系统模型需要进一步研究与探讨。

参考文献

- [1] HIGGINS S. The DCC curation lifecycle model [J]. *International Journal of Digital Curation*, 2008, 3 (1): 134-140.
- [2] BEAGRIG N, POTHEN P. The digital curation: Digital archives, libraries and e-science seminar [EB/OL]. [2018-09-30]. <http://www.ariadne.ac.uk/issue30/digital-curation>.
- [3] POOLE A H. How has your science data grown? Digital curation and the human factor: a critical literature review [J]. *Archival Science*, 2015, 2 (15): 1-39.
- [4] 杨鹤林. 数据监护: 美国高校图书馆的新探索 [J]. *大学图书馆学报*, 2011, 29 (2): 18-21.
- [5] 宋秀芬. 数据知识库的数据监护研究 [D]. 武汉: 武汉大学, 2016.
- [6] 刘雄洲, 王菲. 国外数据存管实施现状及其对国内高校图书馆的启示 [J]. *图书馆*, 2012 (5): 81-83.
- [7] 陈定权, 诸葛列伟. 数据监护发展现状及对国内高校图书馆的思考 [J]. *图书馆论坛*, 2014, 34 (3): 97-102.
- [8] 叶兰. 国外数据监护教育与职业发展研究 [J]. *大学图书馆学报*, 2013 (3): 22-28.
- [9] 高珊, 卢志国. 国外数据馆员的能力需求与职业教育研究 [J]. *图书馆*, 2015 (2): 65-69.
- [10] 李文文, 陈雅. 国内外Data Curation研究综述 [J]. *情报资料工作*, 2013, 34 (5): 35-38.
- [11] 高红文, 陈清文. 国外数据监管研究综述及启示 [J]. *图书馆学研究*, 2013, 10 (3): 2-4.
- [12] 殷沈琴, 张计龙, 窦方. 欧洲科学数据监护的标准与实践——UKDA 案例研究 [J]. *图书馆杂志*, 2013, 32 (6): 76-80.
- [13] 宋秀芬, 邓仲华. 基于数据监护的机构知识库研究 [J]. *图书馆学研究*, 2016 (2): 44-48.
- [14] DCC. What is digital curation? [EB/OL]. [2018-09-16]. <http://www.dcc.ac.uk/digital-curation/what-digital-curation>.
- [15] 邓仲华, 宋秀芬. 信息资源云的数据监护研究 [J]. *图书馆学研究*, 2014 (17): 45-52.
- [16] BEAGRIG N, POTHEN P. The digital curation: Digital archives, libraries and e-science seminar [EB/OL]. [2018-09-30]. <http://www.ariadne.ac.uk/issue30/digital-curation>.

- [17] SIMBERLOFF D, BARISH B C, DROEGEMEIER K K, et al. Long-lived digital data collections: enabling research and education in the 21st century [R]. Technical Report NSB-05-40, National Science Foundation, 2005: 13-20.
- [18] 邓仲华, 宋秀芬. 英美数据监护研究 [M]//陆伟, 查先进, 姜婷. 情报学研究进展. 武汉: 武汉大学出版社, 2017: 189-236.
- [19] Center for Research Libraries. TRAC Metrics [EB/OL]. [2018-10-25]. <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.
- [20] 贾仁安. 组织管理系统动力学 [M]. 上海: 复旦大学出版社, 2014: 5-28.
- [21] 王其潘. 系统动力学 [M]. 上海: 上海财经大学出版社, 2009: 47-80.
- [22] 钟永光, 贾晓菁, 钱颖. 系统动力学 [M]. 北京: 科学出版社, 2011: 30-48.
- [23] Data Seal of Approval. Data Seal of Approval [EB/OL]. [2018-09-29]. <http://www.datasealofapproval.org/en/>.
- [24] LEE D J. Research Data Curation Practices in Institutional Repositories and Data Identifiers [D]. Tallahassee: Florida State University, 2015: 23-105.
- [25] 孙晓华. 产业集聚效应的系统动力学建模与仿真 [J]. 科学与科学技术管理, 2008, 29 (4): 71-76.
- [26] University of London Computer Centre. AIDA self Assessment Toolkit [EB/OL]. [2018-12-25]. <http://aida.jiscinvolve.org/wp/toolkit/>.
- [27] BEAGRIE N, BELLINGER M, DALE R, et al. Trusted Digital Repositories: Attributes and Responsibilities [R/OL]. [2018-11-25]. <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.
- [28] DALE R, AMBACHER B. Trustworthy Repositories Audit & Certification (TRAC): Criteria and checklist [R]. OCLC Online Computer Library Center, 2007: 18-30.
- [29] 高鸿业. 经济学: 微观部分 [M]. 北京: 中国人民大学出版社, 2011: 100-120.
- [30] 李杰兰, 陈兴鹏, 王雨. 基于系统动力学的青海省可持续发展评价 [J]. 资源科学, 2009 (9): 1624-1631.
- [31] 孙晓华. 产业集聚效应的系统动力学分析 [J]. 科技管理研究, 2007 (12): 248-249.
- [32] 高鸿业. 西方经济学 (微观部分): 第五版 [M]. 北京: 中国人民大学出版社, 2000: 103-104.
- [33] 狄国强, 曾华艺, 勒中坚, 等. 网络舆情事件的系统动力学模型与仿真 [J]. 情报杂志, 2012, 31 (8): 12-20.

作者简介

宋秀芬, 女, 1982年生, 博士, 讲师, 研究方向: 数据监护, E-mail: songxiufenlook@126.com。
邓仲华, 男, 1957年生, 教授, 博士生导师, 研究方向: 知识组织。

Research on Modeling and Simulation of Data Curation Based on System Dynamics

SONG XiuFen¹ DENG ZhongHua²

(1. Hubei University of Police, Wuhan 430034, China;

2. School of Information Management, Wuhan University, Wuhan 430072, China)

Abstract: This paper deeply analyzed the operating mechanism of data curation based on the system dynamics theory, which was to find the superior system structure and function of data curation. It analyzed component factors, influencing factors as well as their interactive relationship, and revealed the operating mechanism of data curation with system dynamics method. It put forward some suggestions to improve data curation capability in four aspects, including national and organizational support strength, organizational planning ability, obstacle to the development and data service capacity.

Keywords: Data Curation; System Dynamics; Research Data Quality; Data Service; Open Access

(收稿日期: 2018-12-20)