

· 学科进展与展望 ·

国家自然科学基金委员会数据资源规划研究与实践

付小龙¹ 袁芳¹ 张民社² 闫术卓² 宋树仁¹

(1 清华大学计算机与信息管理中心, 北京 100084; 2 国家自然科学基金委员会信息中心, 北京 100085)

[摘要] 信息流的畅通和数据资源的有效利用是信息化可持续发展的基础。本文介绍了基于顶层设计的数据资源规划方法, 提出数据资源规划的工作步骤。结合国家自然科学基金委信息化现状和未来业务发展需要, 归纳、分析、聚类、抽象出国家自然科学基金委的全局业务和全域数据, 建立了国家自然科学基金委的业务模型和数据模型。提出自然科学基金委的数据资源管理规范, 从而形成了国家自然科学基金委数据资源的整体规划。

[关键词] 顶层设计, 数据资源规划, 信息流分析, 数据建模, 数据规范

1 引言

国家自然科学基金委员会(以下简称自然科学基金委)经过多年信息化建设, 积累了大量自然科学基金项目申请、项目资助、项目成果, 以及相关研究人员和科研机构信息, 形成了国内基础研究方面的核心数据资源, 支撑着自然科学基金委的业务运转。如何深入有效利用这些珍贵的数据资源, 为基础科学研究、科学基金资助管理提供深层次的信息化服务与支持, 使自然科学基金管理更加规范、科学, 是自然科学基金委新一代信息系统建设需要重点解决的问题。为此, 本文采用基于顶层设计的数据资源规划方法, 立足自然科学基金委当前信息化建设, 面向未来信息资源建设和利用, 从全局角度、自顶向下地对自然科学基金委自然科学基金管理和日常办公业务进行分析与设计, 建立自然科学基金委的全局数据模型, 提出自然科学基金委的数据资源管理规范, 形成了自然科学基金委数据资源的整体规划。

2 数据资源现状分析

自然科学基金委的信息系统建设起步较早, 经过十几年的努力, 目前已经基本建成了基于网络化的电子政务系统及自然科学基金管理业务系统, 这些系统的建设及使用对提高自然科学基金委的政务办公和科学基金管理的水平发挥了重要作用。但由于早期信息系统开发是根据需求的发展, 由不同的团队在不

同时间逐步开发完成, 没有实现整体规划和统一部署, 缺乏有机联系和信息沟通机制, 影响了信息共享和信息资源的进一步深化挖掘利用, 具体体现在:

(1) 信息重复存储: 相同信息在多个信息系统中重复存储, 共性基本信息在每个业务系统中均有副本, 项目申请者和承担单位需要花费大量时间重复填写, 并且相同信息的分散存储无法保证数据一致性, 管理成本也随着数据量的增加日益增高。

(2) 存在信息孤岛: 不同信息系统的业务数据间不关联、信息不共享互换以及信息与业务流程和应用相互脱节, 同时没有建立数据交换共享的管理机制。

(3) 缺乏全局数据定义和描述: 由于早期的信息系统采用了分散开发、单项应用、数据自采自用的方法建设, 没有形成从整体业务流程的角度规划信息系统业务数据及数据流转模型, 是造成数据重复存储和信息孤岛的主要原因。

(4) 缺乏数据管理策略和工具: 目前缺乏对数据质量保证和质量控制、数据标准规范、数据所有权等问题的有效的管理策略和管理工具; 缺乏数据使用和数据业务规则的明确定义和规划; 对于自然科学基金项目特殊的保密要求, 缺乏数据利用的审计管理。

(5) 缺乏数据文档和元数据: 目前信息系统的开发文档相对比较完整, 而用于描述数据的文档和元数据则非常有限, 缺乏对数据建模、数据迁移、数据转换和数据结构变化的定义和描述, 使得数据维护工作日益困难。

本文于 2011 年 6 月 7 日收到。

自然科学基金委“十二五”期间,要求采用“统一规划、统一系统、统一门户、统一编码、统一数据、信息一次录入,统一共享使用”等原则建设新一代信息系统,进一步提升信息化基础设施的建设水平,拓展和完善数据中心的服 务能力。因此,加强对自然科学基金委数据资源的管理和规划,通过对数据资源进行深入分析规范、统一并管理好这些数据资源,是建设新一代信息系统的基础,是解决目前信息孤岛问题,盘活数据资源资产的主要手段,也是充分有效利用数据资源,进一步提高自然科学基金委在基础研究领域的主导地位需要考虑的关键问题之一。

3 基于顶层设计的数据库资源规划方法

信息化顶层设计^[1]是从全局的视角出发,站在整体的高度,以信息化的思维,全面分析机构的各项

业务,建立该机构的业务模型、组织模型、功能模型、数据模型和用户模型,设计技术架构,规划基础保障环境,并结合该机构的信息化现状,设计出信息化总体技术方案与路线图,完整顶层设计实施包括3个部分5个环节。基于顶层设计的数据库资源规划方法主要采用顶层设计中的数据建模和数据架构设计方法进行数据库资源的分析、规划与设计,进行信息流分析、职能域数据模型设计、全域数据模型设计等。

基于顶层设计的数据库资源规划方法的内容和特点,本文在分析过程中还引入了专业的建模工具,利用工具中模型生成、模型间链接和模型同步等相关技术,对各类结构化模型对象进行统一管理,采用工具完成各类模型的关联和转换、模型冲突的检测。基于建模工具的数据库资源规划方法如图1所示。

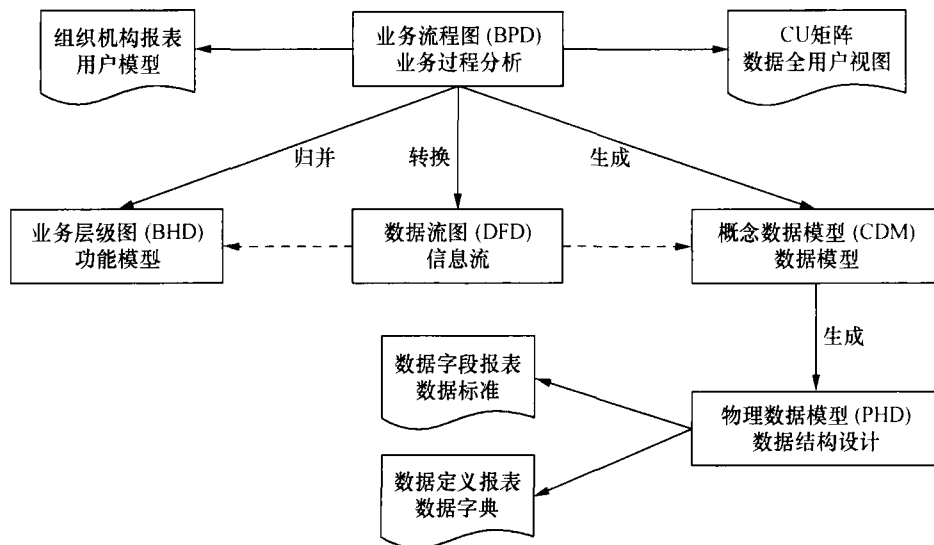


图1 基于专业建模工具的数据资源规划方法

4 自然科学基金委数据库资源规划实践

数据库资源规划^[2]是通过对自然科学基金委在自然科学基金管理和日常办公活动中的各项业务活动及信息流向分析,梳理业务流程以及所产生的各种数据库资源的类型、访问模式、访问范围等,并通过归纳、分析和聚类,抽象出自然科学基金委全域的数据模型,设计长期数据库结构,从而形成自然科学基金委的数据库资源规范和信息交换及信息共享规范等。

(1) 信息流分析。自然科学基金委的主要职能是管理国家自然科学基金,根据国家发展科学技术的方针、政策和规划,有效运用国家自然科学基金,支持基础研究,坚持自由探索,发挥导向作用,发现和培养科学技术人才,促进科学技术进步和经济社会协调发展。因此其核心业务分为自然科学基金的

管理和电子政务办公两部分,一部分是完成自然科学基金管理的主体业务,一部分是完成基金委内部管理和日常运作电子政务办公。信息流分析从组织机构设置入手,详细分析了科学基金管理和电子政务办公两部分业务的流程,分析业务活动的时序关系、层级关系,形成业务流程图;分析各职能域内的数据流向,绘制一级数据流图和二级数据流图;分析各职能域间的数据流向,汇总各职能域的数据流图,最终形成自然科学基金委全域数据流、自然科学基金管理顶层数据流、电子政务办公顶层数据流3条主信息流以及自然科学基金管理和电子政务办公各项业务活动产生的158条信息流,并据此形成了自然科学基金委的整体业务模型。

其中,全域数据流图描述了自然科学基金管理

和电子政务办公两项业务的主要业务过程、产生的信息以及之间的关联,并对两者之间的业务以及信

息交换进行分析,是自然科学基金委业务的宏观模型,如图 2 所示。

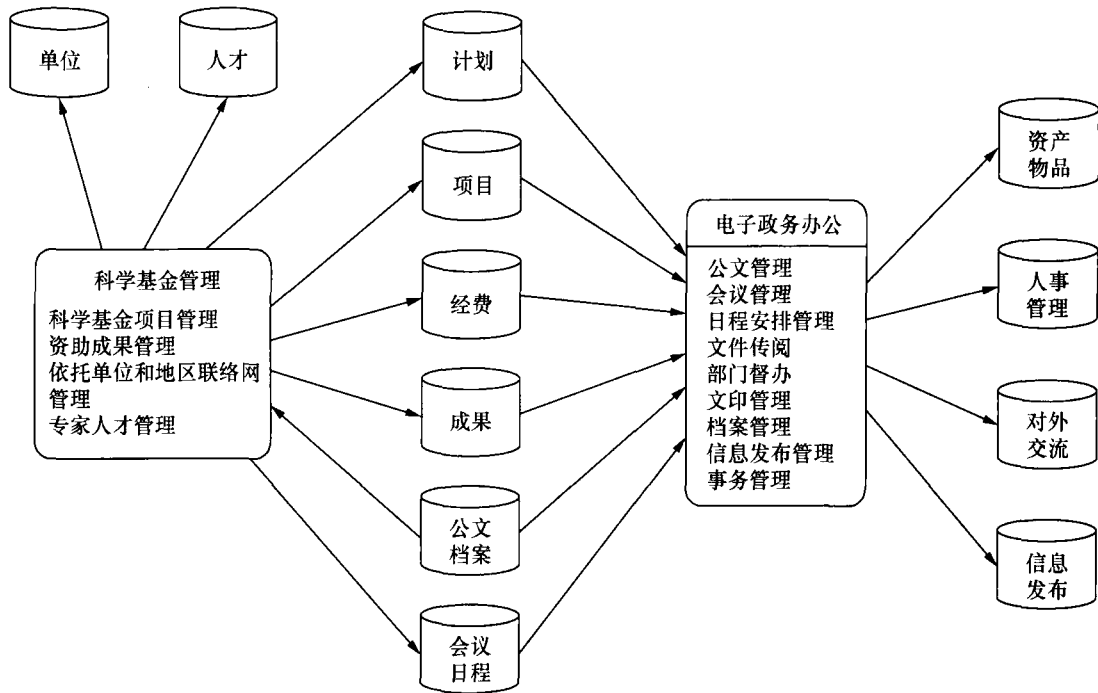


图 2 全域数据流图

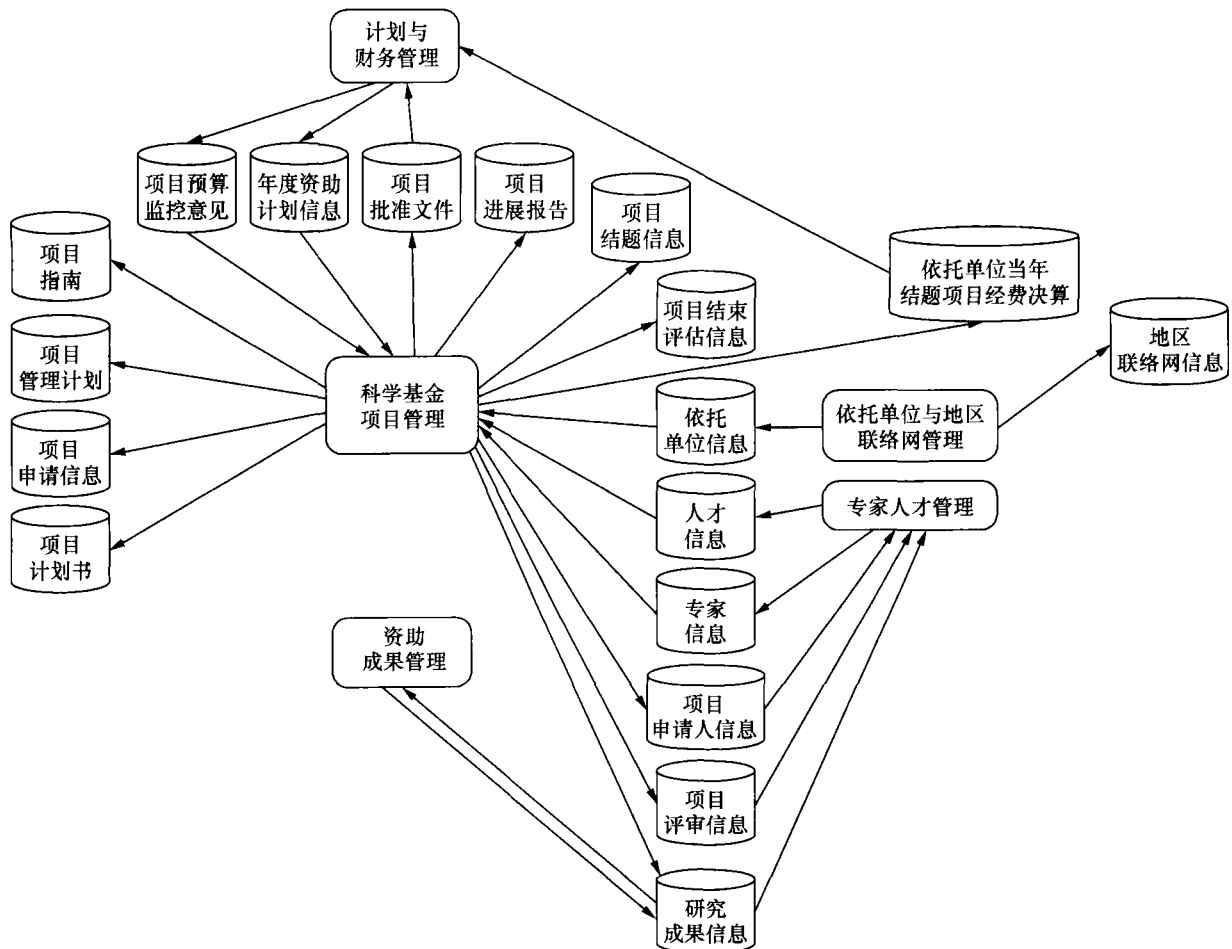


图 3 自然科学基金管理顶层数据流图

自然科学基金管理是自然科学基金委的核心业务,自然科学基金的管理采用项目资助的方式进行,涉及到计划、项目、经费、机构、人才、成果等实体。自然科学基金管理顶层信息流描述了这些实体在科学基金全过程管理过程中的活动及其之间的关联关系,如图3所示。

电子政务办公为科学基金管理自然科学基金委日常公共提供服务支撑,其中涉及公文档案、会议日程、资产物品、人事、对外交流、信息发布等信息。电子政务办公顶层信息流图描述了电子政务办公业务的各项活动、产生的信息及其之间的关联关系,如图4所示。

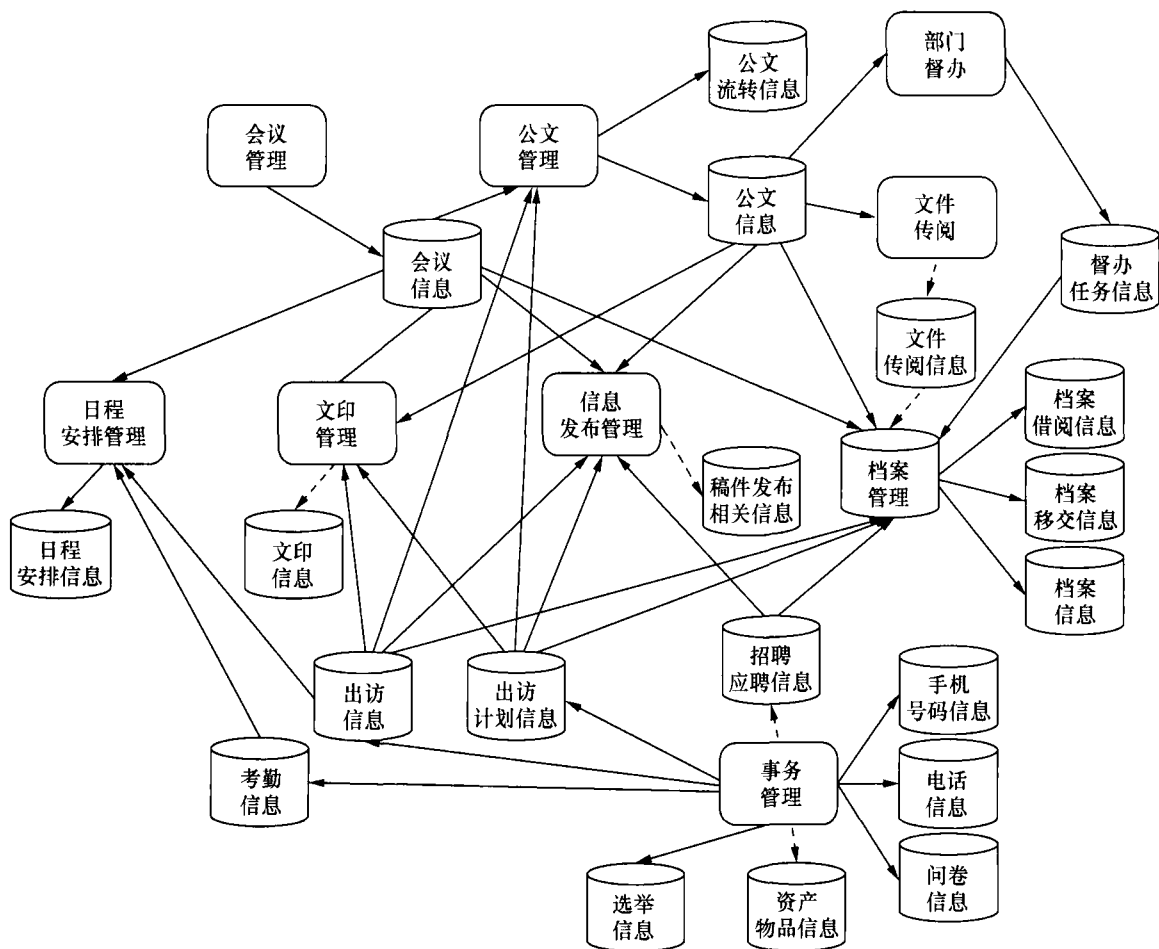


图4 电子政务办公顶层数据流图

(2) 数据模型设计。基于信息流分析产生的数据流图,通过梳理各数据流具体的数据输入和输出,对各类用户视图进行分解,对全域数据流进行平衡性分析,通过结构化组织和层级关系的描述设计形成全域数据模型;依据数据流的密级程度和业务活动特点等进行归并和抽象出主题数据库;并分析其中各项数据实体间的关系设计基本表,形成长期数据结构。

全域数据模型描述科学基金管理和电子政务办公所涉及的所有基本数据表,采用实体关系图来表达所包括的数据实体以及各实体之间的关联关系。在全域数据模型中,根据数据之间关联关系的密集程度,形成了以项目、经费、科研机构、人才、成果、会议等实体为核心,周边关联若干辅助信息的情况,可

以清晰地描述数据的分布状况。

依据对自然科学基金委各项业务的信息流分析,对信息流中数据进行数据梳理和建模,并总体分析数据和数据之间的相关度,在全域数据模型的基础上对相关度高的数据进行归并、归纳和抽取,从而形成描述了科学基金管理和电子政务办公信息化过程中涉及到的主体数据对象的12个主题数据库。主题数据库面向业务主题组织数据,实现基础信息的共享(不是信息私有或部门所有),数据一次一处输入、多次多处使用,来保证数据的准确性、及时性和完整性。同时,各个主题数据库之间存在着紧密关联关系,使自然科学基金委各主题数据库能够集成起来构成统一的数据库,而不是一个个分离的“数据孤岛”,如图5所示。

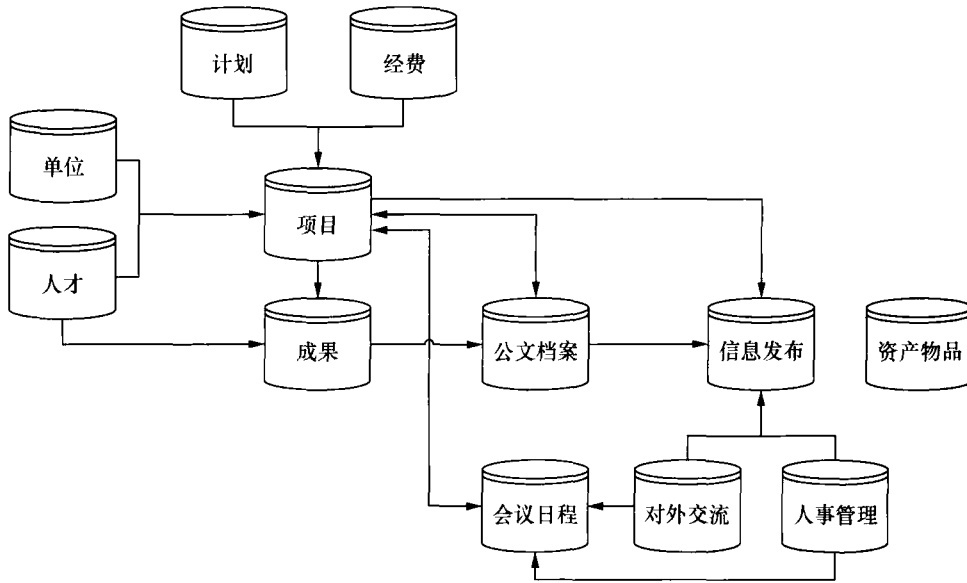


图 5 主题数据库模型

在全域数据模型和主体数据库模型的基础上,结合自然科学基金委信息化现状,依据“面向管理、面向科研和面向未来”的 3 个原则,通过识别业务主键、定义字段属性、部署关联关系,设计了数据表的物理结构,形成长期稳定的数据结构,共设计数据表数据表 396 张,数据字段 2822 个。

(3) 数据标准规范设计。依据全域数据模型、主题数据库和长期数据结构设计,遵循已有国际标准、国家标准、部委或行业标准,结合自然科学基金管理的业务需求,抽象其中的原子数据项,设计自然科学基金委的数据标准和编码标准,并统一管理数据标准所描述的数据项名称、属性、有效期、数据采集部门、权威发布部门等相关管理信息,形成元数据规范,并设计了全局共享的代码标准集。

5 数据资源规划成果的应用建议

国家自然科学基金管理是以信息收集、处理、传递、存储、利用的信息流过程为核心的。自然科学基金委数据资源规划通过对自然科学基金管理和电子政务办公两项核心业务进行了信息流分析,梳理了信息的运作流程,通畅了业务活动的信息流,打通了科学基金管理 with 电子政务办公之间的信息传递,实现了全域信息流闭环管理,为自然科学基金委的业务流程优化提供参考,设计了清晰合理的业务流程体系,为进一步促进科学基金管理水平和效率的提高奠定了基础。

稳定的数据模型是信息系统建设的重要基石,可以实现信息的有效共享,保证数据的准确性、及时性和完整性。通过对自然科学基金委数据模型的设计,梳理了全域的信息点,确定了自然科学基金管理

和电子政务办公涉及数据的来源,为数据的科学采集奠定了基础;统一了数据定义和描述,厘清了数据之间的关系,为数据的充分共享提供了标准化保障;抽取出核心数据的主要线索,为数据的有效利用提供了思路;为自然科学基金委新一代信息系统建设中的数据库设计提供参考。

数据标准规范是在数据建模、信息采集、加工处理、数据交换的过程中对数据进行科学、规范的定义和分类。通过建立自然科学基金委数据标准的管理规范,使信息有序流通,最大限度地实现信息优化管理和信息资源共享,帮助使用者方便、快捷、规范地建立应用系统的数据结构,满足信息系统的建设需求,同时保证信息的一致性与权威性、发挥信息资源的综合效益,为自然科学基金委新一代信息系统建设的数据标准设计提供参考。

6 结束语

数据资源的规划、设计与应用是适应信息集成阶段信息化建设的新需求而产生的,是当前信息化建设现阶段重点关注的问题。自然科学基金委数据资源规划完成的工作仅仅是数据资源管理工作的开始,还需要持续不断地研究并深入实践,才能充分发挥数据资源的效益,实现自然科学基金委信息化的可持续发展,实现自然科学基金科学、高效地管理,进而推动国家基础科学研究的不断进步。

参 考 文 献

- [1] 刘启新,蒋东兴,付小龙等. 校级统一信息系统顶层设计方法. 大连海事大学学报, 2009, 35 卷(增刊): 65—68.
- [2] 袁芳,付小龙,蒋东兴. 数字校园数据管理体系的研究与实践. 大连海事大学学报, 2009, 35 卷(增刊): 21—24.

RESEARCH AND PRACTICE ON DATA RESOURCE PLANNING OF NATIONAL NATURAL SCIENCE FOUNDATION

Fu Xiaolong¹ Yuan Fang¹ Zhang Minshe² Yan Shuzhuo² Song Shuren¹

(1 Computer and Information Center, Tsinghua University, Beijing 100084;

2 Information Center, National Natural Science Foundation of China, Beijing 100085)

Abstract The smooth of information flow and the efficient utilization of data resource is the base for informatization to improve continuously. In this paper, a data resource planning method based on top design is introduced, and detailed steps about data resource planning are put forward as well. Considering the informatization's current status and the future development of National Natural Science Foundation of China, the business model and data model is built up. And a data resource management specification is formed. As a result, the paper gives a introduction on the practice of the data resource planning of National Natural Science Foundation of China.

Key words top design, data resource planning, information flow analysis, data modelling, data specification

· 资料 · 信息 ·

民航联合研究基金投入增加

国家自然科学基金委员会与中国民用航空局日前协议决定,从2011年开始,第三期民航联合研究基金双方每年共同投入从1000万元增加到2000万元,此举将进一步促进民航科技的创新能力,以适应“十二五”期间民航快速发展的需要。

自2003年至今,民航联合研究基金已运作了两期,前两期持续6年的资助超出了预期的目标,在推进民航基础科学和应用基础科学研究、加快民航类科技人才培养、提高民航自主创新和集成创新能力方面均已初显成效。第三期将在充分吸收前两期民航联合研究基金成功运作的经验基础上,拟资助“重点项目”3—5项,平均资助强度300万元/项,资助期限5年;“面上项目”平均资助强度约50万元/项,资助期限4年,资助项目数根据申请和评审情况确定。这些研究将涉及航空安全类、空中交通管理类、航空器维修类、机场类、飞行类、信息技术类以及民航运输管理类、节能减排等多个领域。

2003年,国家自然科学基金委员会与中国民用航空局签订并启动了联合资助基金项目,目的在于发挥自然科学基金的优势,推动知识创新与技术创新的结合,吸收全国范围内的科技工作者参加以我国民航事业发展为背景的基础研究,提升我国民用航空科技源头创新能力。

民航联合研究基金设立两期资助计划以来,共

受理了711个项目,批准资助了139个项目;发表学术论文734篇,获批准专利8项,科研成果已推广应用9项并出版专著14部。支持立项的研究内容按照领域基本划分为航空安全类、空中交通管理类、航空器维修类、机场类、飞行类、信息技术类以及民航运输管理7大类。在2004—2009年的6年中,全国共有29家研究机构作为依托单位参与了研究,其中民航局直属机构5家,占17.2%;非民航直属机构24家,占82.8%。在批准立项资助的139项目中,重点项目7个,面上项目132个,基本达到了民航联合基金设立的目标。值得一提的是,仅第二期(2007—2009年),民航联合研究基金共培养人才525人,其中博士98人,硕士354人。

民航联合基金项目申请、评审和管理参照国家自然科学基金委员会相关类型项目管理办法执行。申请受理和评审与国家自然科学基金面上项目、重点项目同步。民航联合基金设立双方联合管理办公室负责具体项目管理,联合方民航局机构设在人事科教司,联系电话010-64092631/2606,联合方国家自然科学基金委员会机构设在信息科学部,联系电话010-62327817。

(中国民航大学 供稿)