

# 大数据背景下中科院科研管理数据中心建设探讨<sup>①</sup>

杜义华

(中国科学院 计算机网络信息中心, 北京 100190)

**摘要:** 介绍大数据概念与技术发展, 结合大数据思维分析中科院资源规划项目中信息资源中心的建设, 提出构建中国科学院“科研管理数据中心”的必要性、内涵定位, 详细设计其资源构成、应用发展, 并从理顺机制、统筹组织方面具体探讨其建设推进措施. 对政府部门、企事业单位的数据中心规划建设有一定参考意义.

**关键词:** 数据治理; 资源整合; 大数据; 政府数据中心; ARP

## Analysis of Research Management Data Center Construction in Chinese Academy of Sciences in Age of Big Data

DU Yi-Hua

(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Under the concept and developmental history “Big Data” as well as the big data analysis of the ARP-IRC (Information Resource Center of Academia Resource Planning System) construction, the paper highlighted the necessity of constructing Research Management Data Center in Chinese Academy of Sciences. It also elaborated on the concept and content of the center and how to push forward its construction from the perspective of operating and organizing mechanisms. The paper is hopefully to shed light for government sectors and enterprises on data center planning and construction.

**Key words:** data governance; resource integration; big data; government data center; ARP

随着信息化系统的建设应用, 各企事业单位已有一定数据沉淀, 但数据的分析展示和决策支持应用效果并不理想. 近年来“大数据”炙手可热, 学术界热烈探讨、产业界钟爱有加, 各国各级政府积极推动. 如何突破瓶颈、落地数据的整合建设和分析应用, 是大数据时代信息化工作者们面临的困惑. 笔者长期参与中国科学院的科研管理信息化工作, 以下结合大数据概念和 ARP-IRC 现状, 提出、设计和探讨建设中科院科研管理数据中心.

## 1 “大数据”概念与技术发展

### 1.1 “大数据”的概念

“大数据”指需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率

和多样化的信息资产. 其产生背景是数据量呈指数级爆发式增长, 机器数据日益重要, 数据成为一种自然资源且不被利用就是成本, 而其规模和复杂性现有数据库软件难以处理. 特点包括 Volume(数据体量大)、Velocity(速度快实时)、Variety(类别多样)、Value(价值密度低, 商业价值高)等.

### 1.2 “大数据”相关技术

大数据相关产品与技术包括 Hadoop(Mapreduce)、NoSQL(Redis 和 MongoDB)、数据挖掘(深度机器学习)、Knowledge Graph、自然语言处理(sentiment analysis 情感分析, 开放域信息收取, 分类聚类)、大数据可视化、R 统计分析语言等.

目前主要形成有离线批处理、实时流处理和交互式分析三种计算模式, 分别以 MapReduce 和 Hadoop 系统、

<sup>①</sup> 基金项目:中国科学院计算机网络信息中心一三五规划重点培育方向专项(CNIC\_PY\_1412)

收稿时间:2014-05-07;收到修改稿时间:2014-05-28

Yahoo 的 S4 系统和 Twitter 的 Storm 系统、谷歌的 Dremel 为代表<sup>[1]</sup>。

在“管理”大数据方面，多从架构和并行方面考虑，解决高并发数据存取的性能要求及数据存储的横向扩展；在“理解”大数据方面，特别是对非结构化信息的自然语言理解、多媒体内容理解等，仍缺乏实质性的突破和进展。

### 1.3 “大数据”应用现状

大数据发展得到广泛关注，联合国“数据脉动”计划、美国“大数据”战略、英国“数据权”运动、日本“面向 2020 年的 ICT 综合战略”、韩国大数据中心战略等先后启动。

大数据应用主要实现有面向消费者的互联网普遍服务、通过用户行为分析的精准营销，并逐步在各行各业，如电信、金融、能源、交通环保、工商税务、食品安全、医疗卫生等公共服务领域开展有个性应用。其使用最广泛的是新闻、微博、搜索、购物等网络数据，其次有时间和位置数据、文本数据、RFID 数据、传感器数据、车载信息服务数据、遥测数据、视频监控数据，社交通信数据<sup>[2]</sup>。

全球来看，对大数据认识、研究和应用还仍处于初期阶段，大数据的获取、整合、传递、加工、算法等技术都有待成熟，大数据应用真正落地还需数据足够丰富和开放、数据分析工具强大、数据化决策管理理念适配。

## 2 “ARP-IRC”建设应用情况

中国科学院资源规划项目(Academia Resource Planning, 简称 ARP), 是从院所两级治理结构出发, 以科技计划与执行管理为核心来设计开发的资源配置管理平台, 已在全院一百多个院属单位实施应用十余年, 形成具有十大应用系统和两大应用平台的院所两级联动的整体系统<sup>[3]</sup>。

ARP 系统架构下的信息资源中心(Information Resource Center, 以下简称 ARP-IRC), 按照经典的数据中心流程架构, 将业务系统数据 ETL, 形成数据仓库管理, 进行在线分析、前端应用, 实现了围绕 ARP 系统(主要是所级系统)信息从“产生-采集-传输-存储-处理-应用”各个环节的处理, 初步形成全院有一定历史积累的科研管理数据, 共汇聚有经转换处理后人事、财务、科研管理等基础数据 500 多张表 3 亿多记

录, 并开发部署有院所级信息资源服务平台、院领导综合服务平台等提供查询报表、态势分析服务。

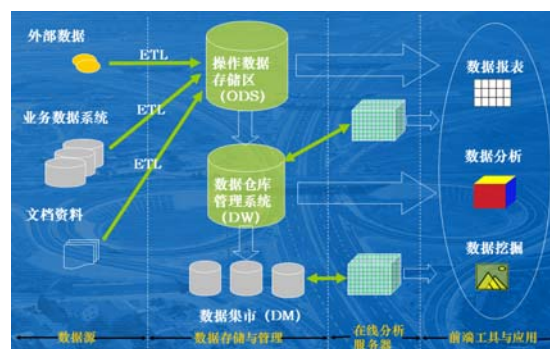


图 1 ARP-IRC 数据流程图

各级领导对利用 ARP 数据提供科学决策支持寄予厚望, 但因数据的不足和管理决策的模式, 应用效果仍不理想。数据的不足体现在其仅局限于 ARP 已有核心模块的主题数据, 缺少其它相关系统数据或还没有相关业务系统, 缺少 ARP 内、外的各类非结构化资料等。一方面是技术原因, 由于 IRC 是 10 年前是针对 ARP 所级系统数据(同一 VPN 网络中异地分布但结构相同的关系数据表)的交换汇聚设计, 随着院信息化项目建设和 IT 技术发展, 此架构、技术及运行模式已无法满足需求和适应发展; 一方面是管理原因, 项目工作人员主要是运维系统和完成工程任务, 实际进行资源建设整理的力量极少, 没有义务也无法主动全局推进。

## 3 大数据思维对“ARP-IRC”的启示

大数据思维主要包括总体样本取代随机样本、对不精确的容忍度增加、相关关系取代因果关系<sup>[4]</sup>。当前“ARP-IRC”中积累数据远称不上“大数据”, 其体量有限、类别唯一、关系型、价值密度较高等也与大数据特征正好相反。但两者在处理流程和面临问题有一定相似, 大数据是发展方向, 大数据的概念、思维可引入参照。

### 3.1 数据的来源、规模与类型

#### 3.1.1 “信息资源”特征

信息资源是人类活动各个领域所产生的有使用价值的信息集合, 是一种具有特殊内涵和特殊配置形式的社会资源。具可增长性、共享性, 不同于单一消耗性资源。具可传播性, 不断传播的过程, 就是其价值不

断得到实现的过程。当信息不再稀缺甚至泛滥的时候, 稀缺的是关于信息的信息。

信息资源具有开发、利用和价值转化性。信息无时不有无处不在, 信息收集归纳是需有专人负责的目的行为。如同物质资源的开采、发掘、加工一样, 信息资源的价值转化与利用需一系列过程运作才能实现<sup>[5]</sup>。

### 3.1.2 “大数据”的基础和源头

大数据的基础与源头是时时连接在线的海量人和物产生的数据<sup>[6]</sup>。移动互联、社交网络、电子商务、物联网的快速发展, 使得全球超过一半的人时时连接在线, 商品的物流、空气的动态变化等也连接在线。不同于图书馆的知识、档案馆的档案, 大数据是时时并发、开放、不断拓展和动态利用的。

大数据的来源包括运营式系统(应运营管理决策的需要被动产生的)、互联网系统(互联网环境下“应人而生”, 结构复杂, 无秩序, 弱一致性)、感知式系统(物联网环境下源源不断自动产生的)<sup>[7]</sup>。

### 3.1.3 总体数据与局部数据

理论上讲大数据应该是总体数据, 由于技术、商业、保密和其它原因, 除少数原始拥有者, 其它人实际上是局部数据。局部数据指总体数据的一个子集, 但不是通过随机方法从总体中抽出来的, 而是用各种便利或现存的方法得到的, 可能比样本数据大很多。

统计分析中, 如果只考虑质量, 不计成本、效率等因素, 总体数据最靠谱, 其次是样本数据、最不靠谱的是局部数据。

### 3.1.4 对“ARP-IRC”发展启示

信息资源是企业生产及管理过程中所涉及的一切文件、资料、图表和数据等信息的总称。“ARP-IRC”目前是中国科学院一部分运营式系统的一部分数据, 数据规范, 有秩序, 一致性较强。“ARP-IRC”数据范围有待扩展。

绝大多数业务信息均是非结构化数据, 非结构化数据中含有更丰富更直观的信息, 可对分散在各类工作人员中的文档报告资料、数字化后科研档案音视频等汇聚利用, 并尽量形成完整的总体数据。

科学院管理信息化和科研信息化过程中, 同样有大量互联网应用和物联网应用, 实时感知数据、各类运行日志数据等需有目标的组织收集归纳。

## 3.2 数据的建设方式与资产价值

### 3.2.1 数据的建设模式

“大数据”是把所有可能采集到的数据全部集中在一起分析处理, 其不仅是“多”(海量)、“广”(格式种类多), 更是“深”, 即用传统思维来看似乎不相关、但可能有被忽略的复杂联系的数据。建设应用上常是先有数据后有模式, 从量变到质变。

### 3.2.2 数据活性与价值

决定数据资产价值的有数据的规模、活性, 以及收集运用数据的能力。大数据的重要特征之一是数据类型多和价值密度低。让数据产生价值, 不是大数据自身能解决的, 需将数据组织成数据资源体系, 进行数据的层次类别划分、相关性标注。只有数据所有者围绕核心业务需求构建起数据间的关联关系, 才能提高那些不同来源获取的结构化与非结构化数据的活性。

### 3.2.3 资产拥有/控制者

根据拥有和控制的角度, 数据资产可分为第一方数据(数据的生产者)、第二方数据(乙方为甲方服务时产生, 所有者拥有对数据的掌控权, 但受制于获取路径, 在使用、交换或交易的过程中会有一些限制)、第三方数据(通过网络爬虫、甚至黑客手段获取的数据, 因第一方、第二方数据常无法通过交易或授权渠道获得)<sup>[8]</sup>。

### 3.2.4 对“ARP-IRC”建设启示

目前 IRC 数据是传统方式上根据需要制定计划、采集整理的, 即使海量也不同于“大数据”概念。相对核心管理数据的重“质(专)”, 对于互联网等信息还需“量(广)”。

数据建设应用需要一个周期, 不可能“立等可取”。可以先大量收集汇聚, 再逐步消化应用; 建设分析引擎平台和知识服务平台。

实现数据的保值增值, 除扩大数据规模, 提升收集运用数据的能力, 重要的是提高数据活性, 如对于一些非结构化数据, 真正的价值来将其同其它更具结构性的信息联系起来。

数据管理和应用的主体应是科学院。相对院属单位自身或面向院领导, 院内各类数据均是第一方数据; 若相对网络信息中心, 汇聚的数据则多为第二方数据(各种操作访问日志、电话工单运维记录除外), 不便于采集和发挥利用。

汇聚建设重点应是院内资源和中科院相关资源。IRC 应用中通过网络采集的基金委、专利局的一些数

据为第三方数据. 对于需要的院外相关资源目前可适当收集整理, 但随着政府和社会数据公开共享机制健全, 此类数据应更多的去合作共享而不是花大量精力且仍无法及时完整的采集获取.

### 3.3 数据分析方法与预测模型

#### 3.3.1 处理流程

大数据的处理方法很多, 普遍适用的流程可以概括为四步, 分别是采集、导入和预处理、统计和分析, 最后是数据挖掘.

#### 3.3.2 预测模型

大数据的核心在于预测. 大数据不是商业智能(BI), BI为基于样本数据因果关系的精确分析, 大数据基于全部数据的相关关系的模糊分析. 如电子商务网站通过数据预测顾客是否会购买推荐的产品; 信贷公司通过数据预测借款人是否会违约; 执法部门用大数据预测特定地点发生犯罪的可能性; 交通部门利用大数据预测交通流量.

预测模型的构建是关键. “本质而言, 所有模型都是错误的, 只是有些模型更有用”(统计学家 George E. P. Box). 一个模型是否正确不是最重要的, 重要的是决策者对这个模型有多大的把握, 决策者能否从这个模型中获利.

#### 3.3.3 对“ARP-IRC”应用启示

“ARP-IRC”现有的数据处理流程与其类似, 辅助决策方面面临同样困境, 即: 决策者(管理人员)希望(大)数据能够为他们提供帮助, 但是却不知道(大)数据能够为他们提供何种帮助; 研究(工作人员)制订的基于数据的模型不一定能说服决策者据此决策且帮助改善决策.

大数据本质上是“一场管理革命”. “ARP-IRC”当前先切实做好已有业务系统的数据梳理汇聚, 加强查询分析(商业智能)类应用, 同时努力务实构建模型, 在正确的时间关注正确的业务问题.

## 4 从“ARP-IRC”到院“科研管理数据中心”

遵循信息化应用规律, 顺应大数据发展趋势, 建议构建新一代院“科研管理数据中心”.

不同于“ARP-IRC”仅对 ARP 系统的关系型主题数据抽取和将“管理云”各子云集数据调用, 科研管理数据中心用于集成全院各类相关数据, 并强调整合、整理和应用, 进行“广泛收集、全面整合、深度萃取”.

### 4.1 构建必要性

实现数字化科学院、科学决策支持系统等需要院全局数据. ARP 项目统一全院人财物项目等业务管理系统, 实现了基础核心管理数据间的关联整合. 但科学院在十一五、十二五信息化专项, 以及其它项目支持下, 院属单位又分别建设有“科技云”、“教育云”和其他各类平台系统, 客观上又形成了系列数据孤岛. “ARP-IRC”是科学决策提供辅助服务方面不可或缺的关键数据, 但 ARP 项目不是院科研管理活动信息化的全部, 还需与科研信息化、科技数据资源、继续教育与科教融合、科学传播服务及院外相关资源进行集成共享.

在“大数据”背景下, 建设能全面支撑科研决策、全方位展示中科院的“海一云”信息化管理环境, 其所面对的数据内涵已经不能仅仅局限于系统内的日常交易数据, 还应包括来源于新媒体环境下的交互数据、多媒体数据、日常工作中的文档资料以及仪器装备的感知数据等.

### 4.2 数据中心定位

按字面理解, 数据中心易狭义的偏向“基础设施与存储环境”, 关注存储空间、能源效率、灾备安全等. 此处的“数据中心”应类似“信息中心”, 可理解为情报机构的一种, 偏向信息资源的利用, 用于整合中国科学院各类相关信息, 使人人皆可访问并从中受益, 重点是挖掘隐性数据.

科研管理数据中心定位于“院数字资产的保管与增值应用”. 即: 全面整理汇聚科研管理数据, 为院内外各级用户提供权限内的服务, 包括盘活已有信息化系统的数据资源, 用数据需求驱动业务系统的开发调整, 支撑数据中心作为院科研管理各项业务和服务应用的枢纽.

### 4.3 存储管理技术

数据中心的问题包括数据管理、数据技术和数据应用. 在技术架构发展方面, 方向是云存储+分布式‘云’数据库模式, 提供云接入和应用服务, 实现海量分布式存储和处理、高并发读写、高性能获取、负载均衡、资源共享、在线扩展迁移, 以及足够的安全、简单的管理、标准的访问接口.

作为云数据中心, 业界有两种“云”中的分布式数据库, 一种是基于 key/value 的键值非关系型并行数据库, 提供商如 Google BigTable, Amazon SimpleDB, MS

SDS, 开源/独立有 Hadoop Hbase, oldemort,Cassandra. 另一种是关系型数据库/数据仓库分布式解决方案, 如 OLTP&DW(Oracle,DB2)、DW(Teradata, Netezza, Greenplum), 开源有 Hadoop CloudBase 等. 两者各有特点和应用领域, 关系型分布式数据库可能更适合企业私有云数据中心的建设.

#### 4.4 资源构成分析

##### 4.4.1 “科技信息资源”构成

中国科学院是国家在科学技术方面的最高学术机构和全国自然科学与高新技术的综合研究与发展中心. 科技信息资源的构成包括:

- ① 面向科研工作者的知识传播信息, 如电子图书、电子期刊、数据库.
- ② 面向科技部门、企业、科研机构的科技项目管理信息, 包括科技发展规划、科技法规、研究进展、项目申请、学术交流计划等科技管理信息, 和科技活动概况、科技投入产出、高技术产业等科技统计数据.
- ③ 面向其他科研活动和科技服务而开发的区域条件平台, 如各行业领域的科技文献、科学数据、仪器设施、资源条件、实验基地情况、专业技术资料.

##### 4.4.2 院“科研管理”数据构成

围绕院的“科研管理”活动产生的各类信息, 来源除“业务/运营数据”外, 还包括文档资料、交互信息、感知数据. 其数据构成:

- ① 基础管理数据库: 以组织单元、业务工作、人才、项目、成果、资产等为主线的全院科研活动的数据库全集(涉密除外);
- ② 政务信息资源: 围绕院管理工作的各事件/人物不同时空点的多媒介全息信息;
- ③ 综合数据库: 各类统计数据、图书(年鉴年报、发展报告、要情专报等);
- ④ 部门业务工作数据: 专门系统, 以及院各局处需要的、关注的院外工作数据;
- ⑤ 外部科技资源信息: 科技项目管理信息、科技统计数据及科技产出数据.

#### 4.5 应用发展设计

构建数据中心的目的是为了应用, 除满足管理业务需求、数据服务报告外, 还需形成统一的云服务平台和逐步成为面向院内外用户综合应用的核心.

形成统一“数据中心”云服务平台. 为院内用户提供统一入口和权限管理, 支持 PC 端和移动智能终端,

数据和功能来源数据中心或调用分布式各业务系统.

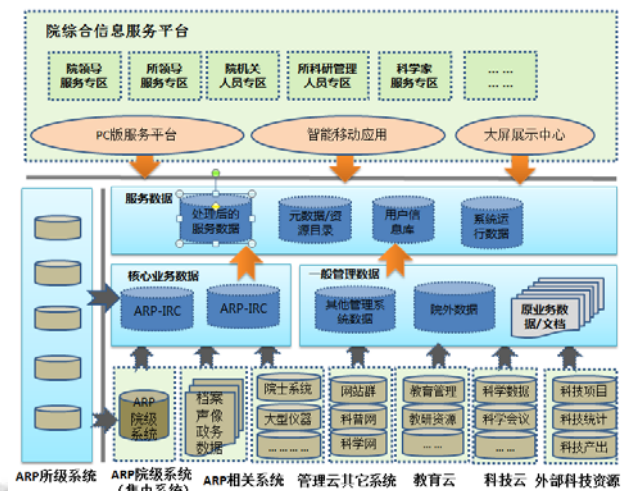


图 2 科研管理数据中心功能架构图

逐步成为面向院内外用户综合应用的核心. 在提供信息服务、可视化展示与交互、辅助决策支持基础上, 联合开发指挥调度或跨部门专题应用, 设计接口标准服务于综合(协同)办公平台, 进一步衍生发展为“公众服务平台”, 如面向公众或科技人员的“院资源导航门户”(科研主页、个人主页)、深度整合的科普应用(虚拟科学院、仪器装备)等.

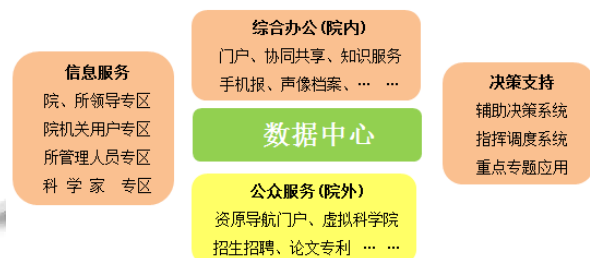


图 3 科研管理数据中心应用拓展示意图

## 5 数据中心建设推进措施探讨

发达国家推行大数据战略的特点主要是: 站位高、合力推进; 突出重点, 巨额投入; 整合数据, 加强公共基础平台建设; 应用拉动, 加快发展<sup>[9]</sup>.

院“科研管理”数据也许不算“大数据”, 但其建设治理也是一项基础性工作, 应用服务则是院信息化好坏的集中体现, 其难点不仅是技术问题, 还有组织、制度和投入等配套环境问题.

### 5.1 理顺机制体制

#### 5.1.1 明确责任主体 提升数据中心地位

传统信息中心多是服务性部门或自负盈亏机构,处在结构底层,制约其积极作为。ARP-IRC 是依托项目运行,资源配置上优先用于工程任务和业务系统开发,工作人员无法深入了解院的中心业务工作,无法协调信息化建设和机构部门间壁垒,无法把面向业务的数据转换成面向管理的数据,很难成真正为决策和管理的支撑,有待提升地位升级角色后走到信息化发展的舞台中央。

数据中心偏指导管理和来料加工,工作应与各业务系统(含 ARP)的建设运维相对独立。

### 5.1.2 明确职能分工 落实运转机制

数据中心(或称信息中心)职能是数据管理、监测评价和服务推送<sup>[10]</sup>,包括发挥引领渗透作用,顶层设计,用数据驱动业务;进行数据的感知集成,在简单数据收集、存储管理上进行资源的有机整合加工;进行数据监测评价和分析应用。

各个职能的每个环节都需要依靠有激情、对数据敏感的专业人员完成,须组建、培养和造就一批熟悉信息技术、了解院中心工作、并有协调指挥的职能和能力的专业队伍。

### 5.1.3 培养数据文化 建立规则与生态系统

建立标准规范。数据建设是一项统筹的、有序的、动态的、可持续发展的系统工程,必须建立良好的运行机制,以保障各环节正常运转、促进各环节正规有序。建设标准则是实现各级各类业务系统网络互连、信息互通、资源共享的基础。

形成资源建设管理的主体力量、支撑力量、外围力量。建立全员关心、全员参与的数据文化。树立数据建设与维护人人有责,每个人应关心数据建设、参与数据管理,工作中优先考虑数据中心的建设和应用。

确立“数据是资产,不是包袱”、“没有垃圾数据,只有暂未清洗好、待利用数据”的理念和共识。

## 5.2 统筹组织推进

### 5.2.1 制订“科研管理数据”治理计划

制定策略来协调多个职能部门的目标,从而优化、保护和利用科研管理数据,将其作为一项资产管理,包括元数据、数据质量、数据安全、信息生命周期等管理。

其重点之一是引导管理人员开放提供工作数据;指定各数据的来源主体和数据要求;规范管理部门的

业务工作方式,避免随意开发系统或填报采集带来的数据分散、重复和不一致。

### 5.2.2 核理指标和数据源头 指导应用系统建设

制订院科研活动管理信息全集指标体系。通过分析为院所领导、科学家、社会公众服务的资源需求,分析院各类重点业务工作中需要或产生的数据资源,分析院各类业务管理系统或信息化项目中已有数据源资源,结合起来初步形成较完整指标体系。

通过数据需求驱动业务系统建设。按照“各类信息均应有且只有唯一的来源出处”原则,对缺少来源的数据项考虑新建业务系统,或增改调整已有某业务系统,对存在多处来源且可能不一致的,核理业务系统间关系,统一设计流程、整合共享资源。

对于需要的第三方数据,加强与相关机构如科技部、基金委、专利局、中国网的合作。

### 5.2.3 跟踪研究相关关键技术和解决方案

“数据中心”是知识技术密集型工作。在(大)数据产业发展、IT 技术更新、社会专业化细分背景下,更需有深入了解各类技术、全面熟悉中科院业务的“通才”和“大师”。技术运用水平和集成质量,直接影响整个数据中心建设的成败。

跟踪业界动态和数据价值链中各环节关键技术的研究进展,制订不同发展阶段的解决方案,如:选择适合的“云”分布式数据库技术和硬软件架构;研究数据交换集成技术,支持复杂环境下不同类型信息资源的汇聚融合;构建非结构化数据的采集管理平台,制订数据汇聚集成规范和共享服务接口标准;结合数据分析、移动互联、物联网等技术,实现信息资源的人性化展示和综合性服务。

### 5.2.4 搭建服务平台和共享平台 边建设边应用

数据的应用与数据的质量(管理)是相辅相成、相互推动关系。解决数据问题的最好办法是应用,而不是因为有问题就不去用。应用功能的设计实现需管理人员与技术人员共同进行。细分受众,边建设边应用。开放资源,让各级人员参与核实、充实或协助推动。

核心业务数据要准确、及时、完整。只有不断地使用,才能够发现存在的问题和不足并针对性地解决。“大”数据,在一定时间内无法使用传统软件工具分析处理,重在积累整理但也可标引处理做“小”后逐步展示,变无形建设为有形成果应用,增强工作信心。

## 6 结语

辅助决策、综合展示是信息化好坏的集中体现。在“大数据”背景和管理应用需要下,整合构建中科院“科研管理数据中心”是大势所趋。从“ARP-IRC”到“科研管理数据中心”,不仅是数据范围扩大、数据量增加和汇聚管理技术升级,而且是一种理念“进化”,需要开放思维、加大投入,甚至变革组织行为模式,需要从源头进行数据的治理集成,群策群力开发云服务平台,逐步建成各类系统数据交换共享的枢纽和面向院内外用户综合应用的核心。

### 参考文献

- 1 唐永建.“大数据”现状及发展浅谈.[http://www.zbeic.gov.cn/art/2013/12/9/art\\_2344\\_171250.html](http://www.zbeic.gov.cn/art/2013/12/9/art_2344_171250.html).
- 2 Franks B. 驾驭大数据.北京:人民邮电出版社,2013.
- 3 及俊川.十年辛勤耕耘,服务一流管理——纪念 ARP 项目实施 10 周年.科技促进发展,2012.11-17.
- 4 维克托·迈尔·舍恩伯格.大数据时代.杭州:浙江人民出版社,2013.
- 5 陈叶娜.信息资源社会价值评价研究[硕士学位论文].成都:成都理工大学,2008.
- 6 黎争.数据源看大数据.IT 经理世界,2013,14.
- 7 孟小峰,慈祥.大数据管理:概念、技术与挑战.计算机研究与发展,2013,50(1):146-169.
- 8 中云网.大数据时代的数据资产管理.<http://www.china-cloud.com/dashujuzhongguo/disanqi/2013/1206/22073.html>.
- 9 王晓明.发达国家推行大数据的战略经验.中国经济时报,2014 年 03 月 20 日 5 版.[http://jjsb.cet.com.cn/show\\_176532.html](http://jjsb.cet.com.cn/show_176532.html).
- 10 魏强.大数据时代下信息中心的挑战与机遇.[http://news.ccidnet.com/art/1032/20130828/5150663\\_2.html](http://news.ccidnet.com/art/1032/20130828/5150663_2.html).