

科研数据共享建设研究

仲妍, 钟艳玲, 郭文思

(军事科学院防化研究院, 北京 100028)

摘要: 科研数据是教学和科研工作中极有价值的资源, 为实现科研数据的集成融合与共享应用, 按照制度化、标准化、集中化思想, 研究建立覆盖建设、运行、管理全流程的数据监管体制, 以及符合利益主体权利和义务的数据共享机制。以标准化管理、服务为特征的数据资源管理平台架构, 以及面向专业科研应用、综合科研应用的科研数据应用架构, 推动科研数据由分散管理向集中管理转变, 数据质量由主观控制向标准控制转变, 数据维护由简单工具向综合平台转变, 数据应用由孤立应用向体系应用转变, 彻底解决各信息系统中涉及的基础数据、共用数据存在内容描述不一致、表达方式不相容、组织方法不清晰等问题, 为信息系统互联互通和数据综合分析应用提供制度保障和技术支撑。

关键词: 科研数据; 数据共享; 监管体制; 共享机制

中图分类号: TP392

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.201015

中文引用格式: 仲妍, 钟艳玲, 郭文思. 科研数据共享建设研究[J]. 电子技术应用, 2021, 47(9): 64-68, 74.

英文引用格式: Zhong Yan, Zhong Yanling, Guo Wensi. Research on scientific research data sharing construction[J]. Application of Electronic Technique, 2021, 47(9): 64-68, 74.

Research on scientific research data sharing construction

Zhong Yan, Zhong Yanling, Guo Wensi

(Institute of Chemical Defence, AMA, PLA, Beijing 100028, China)

Abstract: Scientific research data are valuable resources in teaching and scientific research work, in order to achieve scientific research data integration and sharing application, according to institutionalization, standardization and centralization, this paper has carried out research on construction of data supervision system covering whole process of construction, operation and management, and data sharing mechanism that conforms to the rights and obligations of interested parties. Data resource management platform architecture characterized by standardized management and service, and the scientific research data application framework for professional scientific research applications and comprehensive scientific research applications, can promote scientific research data management from decentralized to centralized, data quality control from subjective to standard, data maintenance from simple tool to comprehensive platform, data application from isolated application to system application. Research results will help thoroughly solving problems such as inconsistent descriptions, incompatible expressions and unclear organizational methods of basic data and shared data in various information systems, and provide institutional guarantee and technical support for information system interconnection and data comprehensive analysis application.

Key words: research data; data sharing; regulatory system; sharing mechanism

0 引言

应用大数据、人工智能技术进行科研数据分析, 例如, 专业发展方向预测、技术关联关系分析等, 都需要科研数据积累到一定程度, 数据呈现出大数据 4V 特性。只有数据形成规模之后, 才能够应用数据挖掘技术发现隐含在大数据量之下的数据关联, 分析出数据的隐含价值, 并进行专业方向与决策分析。而数据的来源(即数据生产者)主要是科研人员、实验仪器设备、科研管理系统、科研实验系统等科研相关人员、设备和信息系统。科研人员和设备的数据也必须进行电子化、格式化后,

方可进入信息系统, 才能被进一步分析利用。

目前, 由于科研与教学信息系统独立建设, 导致数据难以共享。要解决这个难题, 需要解决以下问题: (1) 系统数据如何保证共享; (2) 如何应用信息技术实现数据共享。解决第一个问题就是要解决数据共享正规化建设, 包括数据监管体制构建和数据共享机制研究; 解决第二个问题就是要应用数据及信息技术建设数据共享平台。

1 数据监管共享研究

在各数据应用领域、各个数据汇聚行业中, 都涉及

数据共享与数据保护等问题^[1-7]。因此,一方面需要分析数据监管体制,通过监管干预的方式梳理管理流程中数据拥有者、使用者和管理者之间的层级关系;另一方面需要分析数据共享机制,实现数据权益分层定义,实现数据权力保护。

1.1 数据监管体制研究

信息系统的建设目标诸如开展相关科学研究、获取科技情报信息、开展战场环境和装备模拟仿真研究等,信息系统建设层出不穷,也增加了信息系统和数据建设的异构性,同时对统一的信息系统监管提出了新需求。需要通过对科研数据的产生、汇聚、运用等数据全流程环节的分析,构建数据监管体制。

信息系统的建设、管理、使用单位称为信息系统建设归属部门。数据生产之后需要具有独立职能的科研数据中心建设和维护单位(称为数据信息化建设部门),来负责数据的标准体系建设和配套基础设施建设。数据应用除了供本信息系统所有者使用外,应该在数据共享的基础上应用数据分析技术发现数据更大的价值。这就需要建立统一的具有数据管理权限的单位(称为数据规划机构),来统筹数据资源建设相关的科研经费、科研项目,同时制定数据共享管理规定、要求和机制等规章制度。

通过以上分析可以总结,信息系统作为科研数据的承载者,从建设、运行、管理的流程上来看,需要确立一系列信息管理系统单位的管理职能,建立统一的数据监管体制,自下而上包括信息系统生产者(数据生产者)、信息系统建设归属部门、数据信息化建设部门和数据规划机构等。监管体制建设层级如图1所示。

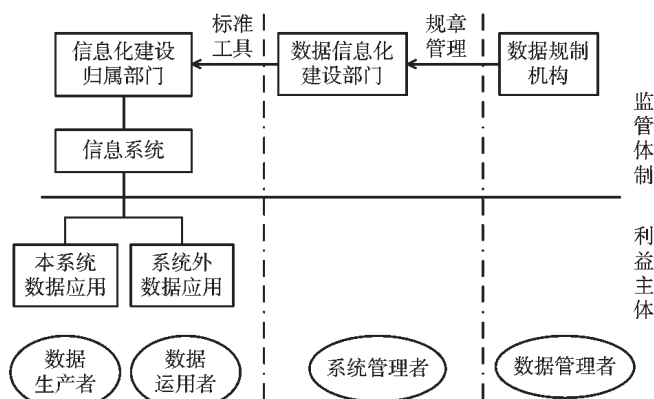


图 1 监管体制

1.2 数据共享机制研究

数据生产者想要保护自己的数据,数据运用者则有对已有数据应用的现实需求,两者存在矛盾。只有同时保证数据生产者和运用者的不同利益价值体现,才能使生产者更愿意支持数据共享、共用,也更能为用户提供更好的决策支持。除了数据的产生和运用外,数据由信息系统统一汇聚至数据信息化建设和数据规划机构,因此这两个部门的管理人员也要纳入信息主

体。为实现网络互联环境下的数据共享,需要完善数据共享机制,保护各信息主体的合法权益^[8-12]。

将数据共享的利益主体与信息主体统一起来,分为数据生产者、数据运用者、系统管理者和数据管理者(如图1所示)。通过借鉴文献[12]中关于利益主体的权利和义务研究思路,建立数据共享机制,为形成数据共享规章制度提供支撑,同时为应用信息技术实现数据共享提供功能补充。各利益主体的权利和义务见表1。

表 1 各利益主体的权利和义务

利益主体	义务	权利
数据生产者	提供信息系统中产生的科研实验数据、情报数据等	可以获取相应的数据共享价值,如记入管理与评价体系、数据被引用记录、获取相应的报酬等
数据运用者	遵循数据使用的审批流程、数据应用的引用描述、数据应用的成果展示等	可以在共享数据上形成独立的学术成果
系统管理者	制定数据共享和信息系统建设的标准规范,提供统一的数据共享平台与工具	可以对共享数据进行认定与评鉴
数据管理者	制定科研数据共享规章制度	开展科研数据共享服务应用支持

2 数据共享技术应用

在数据监管体制管理之下,以共享机制为理论指导和顶层设计,应用信息技术实践数据共享平台,辅以技术手段实现信息系统间数据的互联互通。

2.1 主数据建设

主数据是实现数据共享的核心数据资源,是沟通信息系统的媒介。通过主数据的建设可以实现基础数据的统一建设、统一维护,实现数据复用,减低信息系统建设成本。按照数据监管体制中的职能部门分工,由数据信息化建设部门统一组织实施主数据建设,由数据管理者筹划建设和推行管理及应用机制,由数据生产者和数据管理搭载信息系统开展应用。

2.1.1 标准规范建立

首先,要梳理科研数据现状,在厘清主数据遵循的相关国标、国军标和行业标准的基础上,建立科研主数据的相关标准规范。建设步骤如下:

(1)清理目前现有科研信息系统、实验数据平台、实验仪器设备等中的科研数据界定,建立现有数据体系;

(2)按照数据管理流程,梳理数据各环节建设需要遵循的标准,明确信息系统标准框架中标准缺项,根据可参照的国家标准、国家军队标准、行业标准等编制相应的明细表,制定数据资源及信息系统建设标准体系框架;

(3)分析现有信息系统间数据交互情况,包括系统间集成关联、各系统数据体系及流向、数据接口等,通过关联判断、抽取分析,明确信息系统的主数据应用需求,列出科研主数据清单及与各信息系统之间的应用关联关系;

(4)制定主数据规范,开展主数据建模;制定信息系统间数据交互规范和服务协议,包括常用接口支持、统一数据适配。

2.1.2 主数据建模

主数据模型主要包括:(1)基于标准体系架构的主数据范围及领域划分模型;(2)主数据编码模型;(3)适合教学和科研需要的主数据组装数据模型;(4)架构抽象数据服务模型及接口模型;(5)数据服务协议模型。

2.1.3 主数据采集

科研数据孤立存在于各类信息系统,没有统一的采集手段,致使长期以来不能有效利用各类信息系统产生的数据资源。为解决上述问题,依据建立的主数据模型,在不影响原系统正常运行前提下,采用 ETL 采集、离线采集等技术将现有系统中建立基础较好的共用数据抽取出来,并按照主数据模型进行标准化映射,最后由信息系统建设归属部门、数据规制机构作为使用方和监管方共同审核并确认形成主数据库。

2.2 主数据管理平台架构设计

在严格遵循主数据标准规范体系^[13-15]的基础上,重点开展主数据管理^[16-18],实现主数据服务,构建主数据管理平台框架,如图 2 所示,针对各类信息、功能和主题服务资源,进行主题、信息和功能微服务模型统一的开发、封装、组装和部署,实现服务模型的统一集成管理、共享和使用,按照服务契约流程沟通服务形态、编程模式、通信模式、服务模式和运行模型,并实现所有服务模型一键化配置管理和下发模式。

2.2.1 主数据管理

主数据管理包括数据仓储、元数据管理、模型管理、主题管理、版本管理和数据接口等功能。数据仓储记录主数据的历史版本转变过程,具备主数据迁移功能;元数据管理提供了数据资源体系及关系数据的结构定义和维护,通过对模型及模型结构定义,灵活扩展数据资源管理范畴,以统一数据结构定义方式为数据清理提供依据;模型管理通过元数据定制化组装主数据模型;主题管理提供主数据的主题管理,可通过模型建立不同主数据主题类别;版本管理提供主数据从产生到消亡全过程的版本管理功能;数据接口提供统一操作接口,允许主数据产生系统通过这些接口对主数据进行增加、删除、修改、查询操作,并提供合理的主数据同步机制^[19-20]。

2.2.2 主数据服务

主数据服务包括服务意见收集、服务注册发布、服务信息推送、服务基本管理、服务消息路由和服务监控等功能。服务意见收集根据用户服务使用体验,对服务进行评价和排序;服务注册发布可以把各科研信息系统提供的服务接口在平台中进行注册,审核确认后提供服务发布;注册发布的服务包括系统已有服务和依据服务规范生成的服务;服务消息推送根据各信息系统相关需求,定制推送相关服务描述信息;服务基本管理提供服务编目、申请、审批、搜索等基本功能;服务消息路由使用智能化查询、搜索方法,通过监听服务接口请求消息,实现服务请求的消息路由;服务监控通过监控服务运行状态、访问情况等,用以统计分析主数据的使用频次等,

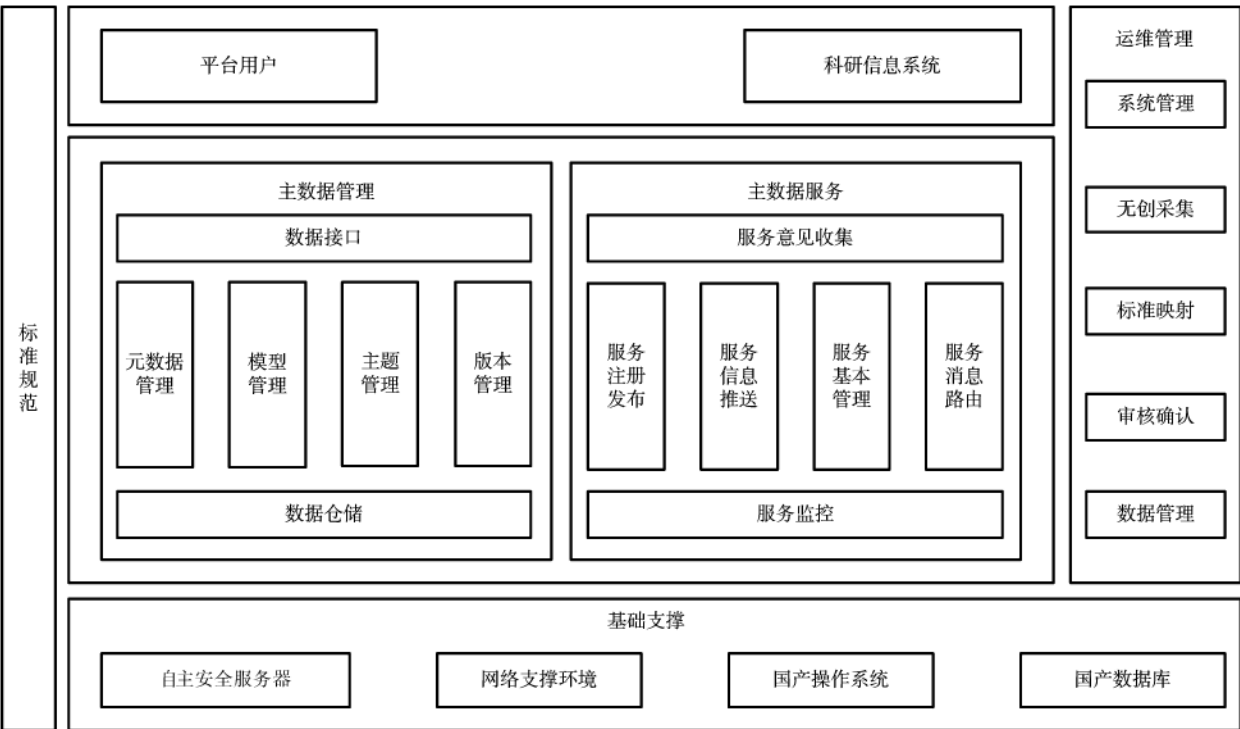


图 2 平台框架图

并设计主数据更新策略。

3 科研数据应用架构

科研数据的应用涉及科研各层次、全领域,按照领域可划分为专业科研应用和综合科研应用,总体应用架构如图3所示。

在科研领域,科研建设、管理、训练和保障各要素相互交织、相互关联。通过科研能力画像、科研设备画像、科研对象分析、科研活动分析等方式进行规律分析和模型提取,找出隐藏在科研数据之下的规律,支撑学术交流、科研计划、科研监管、科研成果和科研评价等专业应用开展,提升科研发展、保障、人员、资源、趋势和应用等综合应用水平。

科研数据应用必须与科研数据建设同步展开,共同推进。以科研数据应用为抓手,通过应用促进科研数据建设发展,围绕数据采集、整理、分析、发掘、展现等环节,扩大科研数据服务范围,提高科研数据服务质量。

3.1 专业科研应用

(1)学术交流。为实现学术论坛、学术会议、学术讲座、学术咨询和学术调研等活动的计划编制、通知公告、资料归档及查阅提供数据支撑。

(2)科研计划。为实现项目指南、综合论证、立项批复、科研大纲和周期计划等文书的编制汇总、审批下发等提供数据支撑。

(3)科研监管。为实现人员监管、设备监管、经费监管、过程监管、绩效监管和协作监管等全要素综合监管提供数据支撑。

(4)科研成果。为实现成果归档、成果应用、成果鉴定和成果报奖等全流程成果管理提供数据支撑。

(5)科研评价。为实现立项评价、过程评价、质量评价、能力评价和学科评价等全方位考核评价提供数据支撑。

3.2 综合科研应用

(1)科研发展。构建科研发展数据集,支撑开展科研态势融合展示、科研数据建设规划和科研发展顶层规划,为科研发展提供信息支撑与辅助决策支持。

(2)科研保障。面向科研保障的各类复杂决策问题,构建科研保障数据集,支撑开展科研保障监控分析、科研保障配置分析、科研保障绩效评估和科研保障合规审查,提高科研保障精细化水平。①科研保障监控分析:提供主题关联的重要科研保障可视化展现;②科研保障配置分析:监控主要科研保障变化情况,预测科研保障变化趋势;③科研保障绩效评估:建立科研需求分析、能力分析、方案优化等模型,构建科研保障评估模型库,基于各级各类数据资源,推动实现基于数据的科研保障绩效评估;④科研保障合规审查:基于规划计划、运行实施、过程反馈和执行结果等各类数据,依据各类标准和评价指标,进行合规审查。

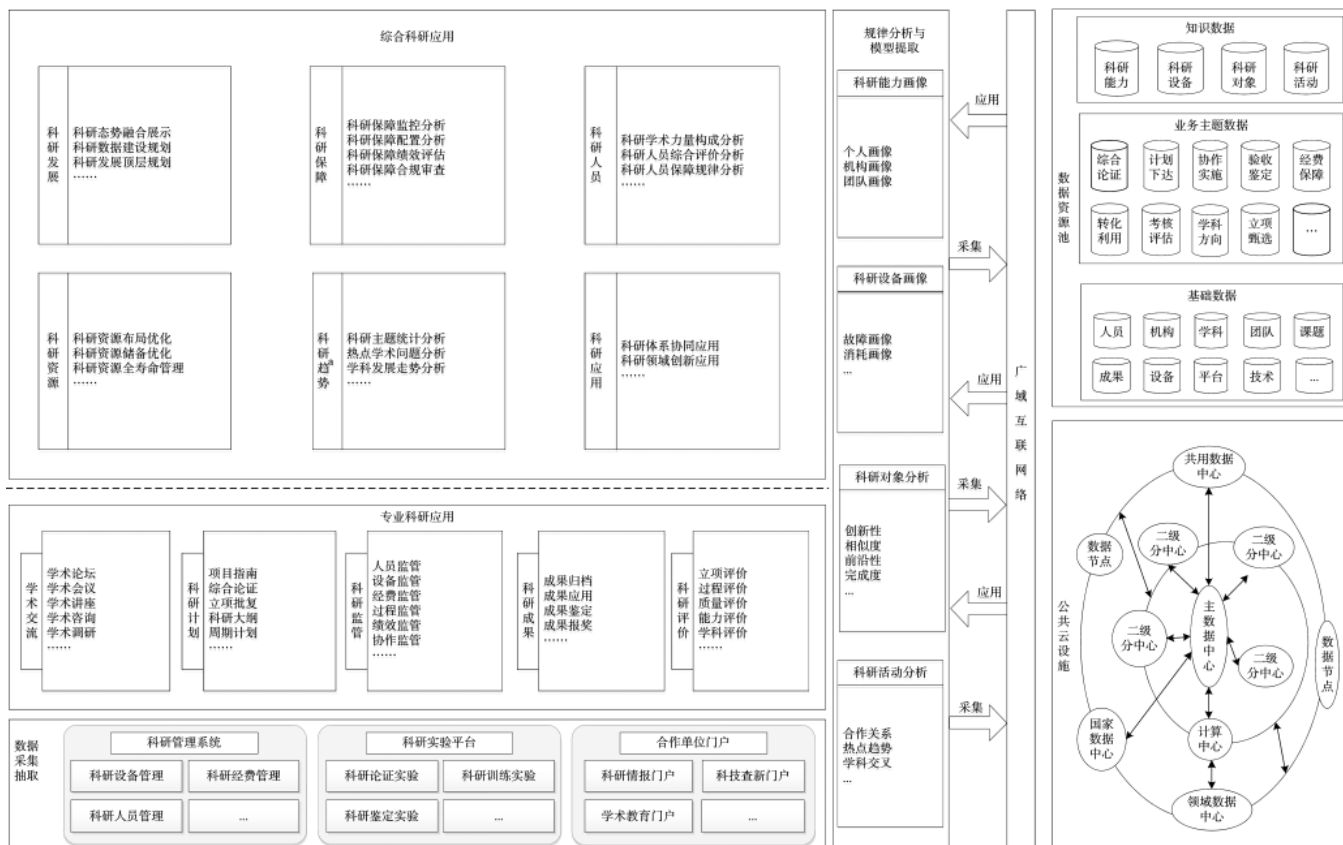


图3 科研数据总体应用架构

(3)科研人员。基于科研人员基础数据,整合多渠道获取的人员科研活动、保障情况、考核业绩等数据,构建科研人员数据集,支撑开展学术力量构成分析、科研人员综合评价分析和科研人员保障规律分析。

(4)科研资源。综合科研方向变化、资源投入规模、资源损耗规律、资源急需程度、资源管理流程等因素,构建科研资源数据集,支撑开展科研资源布局优化、科研资源储备优化和科研资源全寿命管理。

(5)科研趋势。基于各类科研信息服务平台,结合科技发展动态,综合发达国家科研发展情况等,构建科研趋势数据集,支撑开展科研主题统计分析、热点学术问题分析和学科发展走势分析,通过大数据技术手段,分析学科的关注程度、发展趋势、影响程度等。

(6)科研应用。按照科研数据顶层设计和科研领域数据建设应用总体规划,在运用科研数据资源和应用成果基础上,基于领域特色需求和专业科研活动,开展多源汇聚和关联分析,形成与科研体系衔接配套的数据资源体系,支撑科研体系协同应用、科研领域创新应用等。

4 结论

本文研究了科研数据共享相关的制度机制、标准规范、数据模型、管理平台、应用架构等,通过综合运用数据资源无创采集、标准映射与集中管控,以及多源数据融合归一化维护管理、数据管理平台定制化组装等,有助于解决跨系统跨部门科研数据的统一采集、联动更新问题,实现数据源头一致、数据一致、更新一致,便于数据资源集成,有效管理数据资源、控制数据质量,提升应用效能。后续将在此基础上开展平台原型建设。

参考文献

- [1] 杨帆.金融监管中的数据共享机制研究[J].金融监管研究, 2019, 94(10): 53-68.
- [2] 张贵香,刘桂锋,梁炜.我国科研数据管理理论与服务研究进展述评[J].情报理论与实践, 2020(6): 187-193.
- [3] 司莉,邢文明.国外科学数据管理与共享政策调查及对我国的启示[J].情报资料工作, 2013(1): 61-66.
- [4] 顾立平,陈新兰,张萧月,等.开放科研数据中的数据价值提升策略[J].图书馆论坛, 2020, 40(9): 115-124.
- [5] 邢文明,吴方枝,司莉.高校图书馆开展科研数据管理与共享服务调查分析[J].图书馆论坛, 2013, 33(6): 19-25.
- [6] 刘韞照,刘文江.世界一流高校科研数据管理政策研究[J].大学图书馆情报学刊, 2020, 38(3): 31-41.
- [7] 段雨泽,王其发.数据共享机制探讨[J].科学与信息化, 2019(30): 57.
- [8] 李真.P2P 网贷信息用征信:金融分析与法律构建[J].当代经济管理, 2015(7): 85-91.
- [9] 黄源,施棚婕,李晨英.基于科学数据管理流程的科研机构职责分析[J].数字图书馆论坛, 2020, 188(1): 20-26.
- [10] 金贞燕,孙华丽.科研数据管理相关权益分析[J].情报资料分析, 2019, 40(2): 45-51.
- [11] 祝风云.基于利益相关者视角的研究数据管理平台比较研究[J].图书馆学研究, 2020(9): 35-42.
- [12] 司莉,李璐.我国高校科研数据共享中的知识产权与利益协调机制[J].图书馆, 2018(7): 18-24.
- [13] 邹丹.铁路主数据管理平台关键技术研究[J].铁路计算机应用, 2017, 26(1): 31-35.
- [14] 司莉,曾粤亮.机构科研数据知识库联盟数据治理框架研究[J].图书馆论坛, 2018(8): 61-67.
- [15] 王建民,王晨,刘英博,等.大数据系统软件创新平台与生态建设[J].大数据, 2018(5): 104-112.
- [16] 陈君.主数据管理平台建设研究[J].铁道工程学报, 2016, 212(5): 134-136.
- [17] 甘彬,冯敏.基于 SOA 架构的主数据管理系统研究[J].自动化技术与应用, 2020, 39(3): 42-44.
- [18] 雷洁,赵瑞雪,李思经,等.知识图谱驱动的科研档案大数据管理系统构建研究[J].数字图书馆论坛, 2020, 189(2): 19-27.
- [19] 李萌,魏伟.基于 SOA 的主数据管理架构设计及实践[J].兵工自动化, 2015, 34(8): 49-51.
- [20] 谢婧,钱力,师洪波,等.科研学术大数据的精准服务架

(下转第 74 页)

(上接第 63 页)

- metrics to determine the suitability and cost-effect of Self-* capabilities for autonomic computing systems[J]. IEEE Access, 2019, 7: 139759-139772.
- [12] SINGH S, CHANA I. QoS-aware autonomic resource management in cloud computing: a systematic review[J]. ACM Computing Surveys(CSUR), 2015, 48(3): 1-46.
 - [13] BORTOLUSSI L, GALPIN V, HILLSTON J, et al. Hybrid semantics for PEPA[C]//2010 Seventh International Conference on the Quantitative Evaluation of Systems(QEST), Williamsburg, USA, 15-18 Sept., 2010: 181-190.
 - [14] HILLSTON J. Fluid flow approximation of PEPA models[C]//

Proceedings of the 2nd International Conference on Quantitative Evaluation of Systems, Torino, IEEE Computer Society Press, September, 2009: 33-42.

- [15] HUANG Y, KINTALA C, KOLETTIS N, et al. Software rejuvenation: analysis, module and applications[C]//Twenty-Fifth International Symposium on Fault-Tolerant Computing. Pasadena, CA, USA. 27-30 Jun., 1995: 381-390.

(收稿日期: 2020-08-04)

作者简介:

孙日明(1983-),男,硕士,工程师,主要研究方向:计算机相关技术。



扫码下载电子文档

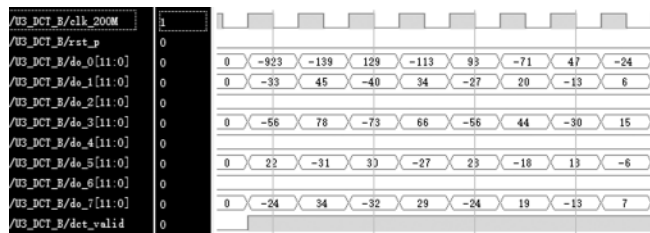


图8 FPGA 计算 DCT 变换结果

从图6和图8中可以看出,DCT变换在MATLAB平台上的仿真结果与在FPGA中实际计算结果完全一致。在设计验证过程中,针对不同的数据源进行重复性测试,实验结果表明,在图像分辨率为 5120×5120 、帧率为 116 f/s 、数据率为 3 GB/s 的连续帧图像数据源下,处理过程中数据无丢失,平均每帧图像的处理时间不超过 10 ms 。针对不同的数据率图像源进行测试,经过多次重复实验,得到DCT变换单元的性能测试结果,如表2所示。

表2 DCT变换单元性能测试结果

分辨率/ppi	帧率/(f/s)	数据率/(GB/s)	时钟频率/MHz	单帧处理时间/ms
5120×5120	116.1	3	200	8.633
5120×5120	103.8	2.72	200	9.66
5120×5120	85.58	2.24	200	11.7

5 结论

本文所设计的基于FPGA的高速图像DCT处理器能够实时处理高分辨率、高帧率时序图像。采用并行流水结构设计,在DCT处理器中设计两个DCT变换核,可同时处理16个像素,实现了数据率高达 3 GB/s 图像DCT变换,此时平均单帧图像处理时间在 10 ms 以内,满足高吞吐率、高精度、低时延要求。设计的DCT处理器具备灵活性和可移植性,可根据需要选择相应模块集成到图像JPEG压缩编码器中,实现高速图像实时压缩,具有较大的应用价值。

参考文献

- [1] NIKARA J A, TAKALA J H, ASTOLA J T. Discrete cosine and sine transforms—regular algorithms and pipeline architecture[J]. Signal Processing, 2006, 86(2): 230–249.
- [2] 唐永亮. 高速高精度离散余弦变换的设计与实现[D]. 天津: 天津大学, 2008.
- [3] CHEN W H, SMITH C H, FRALICK S C. A fast computa-

tional algorithm for the discrete cosine transform[J]. IEEE Transactions on Communications, 1977, 25(9): 1004–1009.

- [4] 徐洁. 基于FPGA的JPEG图像压缩系统的实现[D]. 大连: 大连理工大学, 2013.
- [5] 马林, 李锦明, 张虎威, 等. 高速CMOS图像存储与实时显示系统设计[J]. 电子技术应用, 2017, 43(6): 7–10, 14.
- [6] 杨志勇, 董振兴, 朱岩, 等. 星载高速大容量存储器文件化块管理设计[J]. 电子技术应用, 2017, 43(6): 11–14.
- [7] 张春田, 苏育挺, 张静. 数字图像压缩编码[M]. 北京: 清华大学出版社, 2006.
- [8] DUHAMEL P, GUILLEMOT C. Polynomial transform computation of the 2D-DCT[C]//International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1990: 1515–1518.
- [9] LEE Y, CHEN T, CHEN M, et al. A cost effective architecture for 8×8 two-dimensional DCT/IDCT using direct method[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1997, 7(3): 459–467.
- [10] 张慧慧. 基于FPGA的JPEG压缩编码器的设计与实现[D]. 太原: 中北大学, 2017.
- [11] 曹青云. JPEG图像压缩编码器的研究与设计[D]. 上海: 复旦大学, 2013.
- [12] BABIC D. Discrete cosine transform algorithms for FPGA device[M]. Zagerb, 2003.
- [13] YU S, SWARTZLANDER E E. DCT implementation with distributed arithmetic[J]. IEEE Transactions on Computers, 2001, 50(9): 985–991.
- [14] 来蒲军, 桑红石, 陈朝阳. 一种用于JPEG IP核的DCT设计[J]. 计算机与数字工程, 2007, 35(3): 148–150.
- [15] Automated Imaging Association. Specifications of the camera link interface standard for digital cameras and frame grabbers—v2.0[S]. 2012.

(收稿日期: 2020-08-05)

作者简介:

刘思军(1993-), 男, 硕士研究生, 主要研究方向: FPGA高速信号处理、图像算法处理。

秦明伟(1979-), 通信作者, 男, 博士, 副教授, 主要研究方向: 软件无线电与自适应通信、高速信号采集与实时处理、复杂算法硬件加速技术, E-mail: qmw_qyq@163.com。

刘多强(1991-), 男, 硕士研究生, 主要研究方向: 图像算法处理。



扫码下载电子文档

(上接第68页)

构设计[J]. 数据分析与知识发现, 2019, 25(1): 63–70.

(收稿日期: 2020-10-16)

作者简介:

仲妍(1982-), 女, 博士, 助理研究员, 主要研究方向: 数

据工程。

钟艳玲(1979-), 通信作者, 女, 硕士, 工程师, 主要研究方向: 数据工程, E-mail: 331965476@qq.com。

郭文思(1983-), 女, 硕士, 工程师, 主要研究方向: 数据工程。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所