

引用格式: 韩媛,张红英,梁楠. 大数据在地质资料管理与服务中“落地”问题分析[J]. 中国地质调查,2016,3(3): 67-70.

大数据在地质资料管理与服务中“落地”问题分析

韩媛,张红英,梁楠

(中国地质调查局西安地质调查中心,西安 710054)

摘要: 大数据已成为继云计算、物联网之后新的IT热词。大数据与每个人的生活息息相关。地质资料数据包括了基础地质、矿产、地球物理、地球化学等丰富的信息,对国家、社会具有巨大的潜在应用价值和保存价值。地质资料管理与服务如何与大数据技术更好地结合,产生更大的价值,值得研究。以基层地质资料馆为例,分析地质资料管理与服务的现状及需求,探索如何借鉴大数据思维,改进地质资料管理与服务工作,实现地质数据高效管理,有效地挖掘,使宝贵的地质资料数据发挥出更大的社会使用价值。

关键词: 大数据; 地质资料; 信息服务; 数据共享

中图分类号: P537; P539

文献标志码: A

文章编号: 2095-8706(2016)03-0067-04

0 引言

大数据是指大小超出了传统数据库软件工具的抓取、存储、管理和分析能力的数据库群^[1]。大数据已成为继云计算、物联网之后新的IT热词,各行各业高度关注的热点。大数据具有数据量大(Volume)、数据种类多样(Variety)、实时性强(Velocity)、蕴藏商业价值大(Value)的特性,简称具有“4V”的特点。

地质资料数据是指在地质调查、矿产勘查和科研工作中形成的,以文字、图表、声像、标本、样品测试分析、岩矿心等不同形式存在的地质、矿产和实物信息等。按地质专业分,包括了基础地质、矿产、地球物理、地球化学、重砂、遥感、水文、环境等方面丰富的信息。地质资料数据对国家、社会具有保存价值和巨大的潜在应用价值。据不完全统计,目前全国各类地质工作所形成的地质信息及其数据总量在PB级以上。从地质资料数据种类、数据量级、高商业价值等特点来看,它完全符合大数据的特征。

从全球来看,地质资料管理与服务的主要发展趋势是由主动借阅服务向知识服务转变^[2]。地质资料除应用于地质找矿、灾害预防、科学研究等

方面之外,在城市规划、建设、国防安全等方面,还有更加广阔的应用前景。做好大数据时代下地质资料管理与应用工作,实现地质资料高效管理、合理分析,深挖地质资料内在价值,是地质资料管理由被动借阅向知识服务转变的具体体现,使地质资料得到更充分、更有效的利用,为社会发展提供更有力的支撑。

1 地质资料管理服务现状及存在问题

1.1 地质资料管理及服务现状

我国省、部两级政府部门共有33个地质资料馆藏机构,石油天然气和海洋地质资料委托保管机构35个。地质资料分散存储在各级保管单位,由保管单位完成每年资料的汇交、检查、保管和借阅服务^[3]。以往,由于国家政策、历史沿革、技术手段等诸多原因,地质资料管理长期存在资料信息分散、综合研究力度不够、数字化信息化程度不高、服务渠道不畅、服务能力不强等问题,形成信息孤岛,使地质资料信息的巨大潜在价值未能得到充分发挥。

近年来,随着《找矿突破战略行动纲要(2011—2020年)》的不断推进,地质资料管理与服务工作在“十二五”期间得到飞速的发展。以中国地质调

查局西安地质调查中心(下简称西安地质调查中心)地质资料管理服务工作为例,主要体现在以下几个方面:

(1)地质数据资源得到不断积累。近十年来,随着国家对西北地区地质调查工作投入的倾斜,地质调查项目激增,地质资料数据增长迅猛。同时开展馆藏重要地质档案图文数字化,岩石矿物古生物标本收集整理等工作,丰富了地质信息内容。

(2)开展成果地质调查资料信息集成整合,形成系列服务产品。近年来西安地质调查中心地质资料馆,建立了包括地质成果地质资料、原始资料的案卷级、文件级目录数据库,地质资料元数据库,西北地区地质工作程度“一张图”数据库,西北地区1:25万、1:20万、1:5万等不同比例尺的区域地质图空间数据库,西北地区1:20万区域地球化学数据库,西北地区矿产评价成果基本信息数据库,西北地区地质灾害信息服务数据库等一系列不同专业、不同比例尺的基础地学数据库。这些数据库为提升西北地区地质资料管理与服务信息化水平奠定了基础。

(3)研发、部署一系列地质信息平台。西安地质调查中心自主研发“地学空间信息检索平台”,该平台实现了广域网节点集群地质资料信息资源汇聚,并在地质大数据分布式计算环境中成功部署,解决了因大数据效应导致的地质资料数据存储、迁移、检索及分析等诸多环节面临的效率瓶颈。部署了“中国地质调查信息网格”、“地质调查协同部署平台”、“地质调查项目运行监管系统”等一系列地质调查项目管理、业务信息系统。在地质调查项目管理和具体实施过程中,“数字地质填图系统”、“探矿工程数据编录系统”、“数字地质调查信息综合平台”及“资源储量估算与矿体三维建模信息系统”等数字地质调查系统也得到广泛的应用。

(4)开展非结构化数据存储与处理等技术的探索。通过“西安结点非结构化地质数据集成于服务应用示范”、“地质资料集群产业化试点研究”、“地质资料集群示范”等项目的开展,初步进行非结构化数据存储、处理及分析等技术的探索,取得初步成效。

1.2 存在问题

随着近年来地质信息化工作的不断推进,地质资料管理与服务信息化水平得以提高。从数据层面、技术层面为大数据时代下地质资料信息管理与

服务研究奠定了良好的基础,但还存在一些问题:

(1)地质资料汇交不及时,信息公布不及时。地质工作本身周期较长,一般为2~3a,在这过程中地质资料信息一般都不对外公开。在项目结题后,由于诸多原因,还有些项目组没有及时汇交地质资料。在资料汇交后有部分资料馆藏单位也未能将信息及时公布,这就造成无法快速、全面地收集数据。

(2)地质资料信息管理与服务信息化水平虽然不断提高,但在数据的管理与分析方面还是缺乏大数据思维,数据分析能力也有待提高。近年来建成的数据库多为传统的关系型数据库,现行的信息系统也都比较孤立,数据再利用率并不高,且缺乏统一规划,对资料价值的深挖程度低,交叉综合利用率也不高,使得数据未能产生更大的价值。

为使地质资料数据更有效地利用,发挥更大的价值,需要借鉴大数据思维,从制度、标准、技术等方面,进一步推进数据共享、提高数据分析与挖掘能力,才能使地质资料数据不断产生潜在价值,实现数据价值的最大化。

2 大数据时代下的地质资料管理与服务对策与建议

2.1 大数据对地质资料管理与服务的作用

目前,我国地勘单位众多,地质资料的储存较为分散。地质资料数据储存形式较为多样,包含了文字、图表、声像、电磁介质等形式的原始地质资料、成果地质资料和岩矿心、标本、光薄片及样品等实物地质资料等不同类型、不同结构的资料。要想使地质数据资料信息在大数据时代能发挥更有效的作用,首先要进行数据的汇聚、共享,其次要以用户需求为导向进行数据的挖掘。

要做到以上2点,从技术方面,可通过先进的大数据相关信息技术,实现对异构数据的管理,对过多节点数据的共享,实现数据之间的关联;从管理保障方面,建立符合大数据理念的管理机制,制定完善制度标准,推进数据共享,重视自身数据人才的培养,并与外界加强合作。

2.2 大数据平台架构思想

地质资料管理与服务大数据分析平台的系统架构如图1所示。系统包含3个层次:数据源、平台层和应用层。数据源是大数据平台的基础,包括

地质工作过程中产生的文本、图形、视频、数据库及实物标本等一系列的地质数据及 workflow 数据;大数据平台层是大数据平台的核心层,提供大数据存储和综合分析功能,实现数据的有效汇聚、有效管理,实现数据查询、分析及可视化展示等功能;应用层是根据分析预测的结果,为用户提供大数据解决方案。

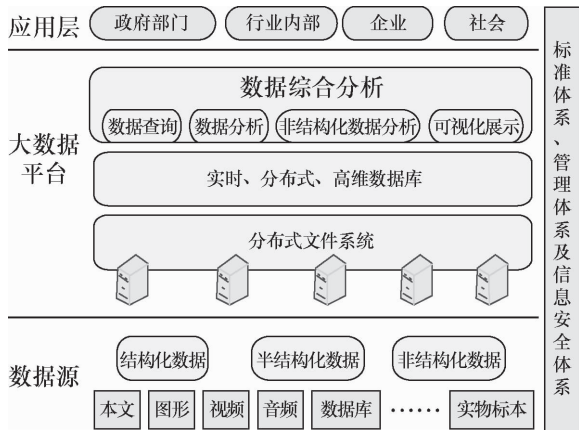


图1 地质资料管理与服务大数据分析平台架构

Fig.1 Architecture of large data analysis platform for geological data management and service

2.3 地质资料信息的收集

大数据基于数据的汇聚,一旦数据本身不准确,就可能导致错误的分析、预测和决策结果。所以准确、全面的数据是大数据的基础。地质资料采集应进一步加强数据质量与实效的控制,扩大数据采集范围。

传统地质资料数据采集多偏向专业数据,在大数据背景下,地质工作各个业务流程所产生的数据都是有用信息,对这方面数据的收集、采集应予以重视。以往由于地质资料的汇交不畅、地质资料数字化程度不高等原因,众多有价值的地质资料没有及时收集,造成地质数据信息滞后。应加强地质资料的汇交监管,加快地质资料数字化进程,加强如数字地质填图掌上机等基于物联网数据信息的实时数据的采集与积累,提高地质调查数据采集的时效性。此外,应建立更广泛的地质资料信息范畴。相较中国较为分散及单一的地质资料信息范畴,地质资料信息在美国囊括了包括气象、地球特性、生态系统、环境问题、地理分析、地质过程、水文过程、自然灾害、自然资源、海洋及海岸线、行星、动植物、

科技应用、水生资源等多个方面信息^[4]。更广泛的地质资料信息,使基于大数据的分析、预测结果更加准确。

2.4 制定完善制度标准,推进数据共享

近年来,地质信息服务的不断推进,地质数据的共享程度有了一定的提高。但由于历史、政策等一系列原因,数据开放程度仍有待提高。如数据储存分散,形成的“数据孤岛”现象较为严重;由于地质数据取得的成本高、困难大及涉密等问题,数据共享程度依然偏低。大数据体系下,在政策允许范围内,应制定一系列制度,例如加强地质资料目录的共享制度。近年来,全国地质资料馆先后建成了全国地质资料目录中心,汇总了各省(区)地质资料馆藏目录,从很大程度上提升了全国地质资料信息的共享程度,但并没有建成各省地质资料馆对全国地质资料目录中心的实时更新的制度,造成数据资源共享不及时。建议建立健全全国地质资料目录中心实时更新制度,推进数据共享。此外,地质资料汇交制度,公益性地质资料的界定及共享制度,非公益性地质调查数据的交换原则等,也应进一步完善,促成数据共享、汇集。避免因数据的重复采集而造成资金的浪费,实现信息高效率利用。

大数据需要建立规则 and 标准。数据产生价值需要共享和交换,意味数据彼此的接口可以建立。一方面为满足用户对数据重复使用、关联分析及自由加工等需求,数据的标准和相关原则应该得到统一;另一方面制定合理、合法的规则和标准,使数据提供者的利益得到相应的保护。所以大数据的标准体系建设就成为非常重要的环节。

2.5 以用户需求为导向,加强数据深挖

大数据时代下,地质工作逐渐向需求驱动型的模式转变,了解用户需求就显得尤为重要。在地质资料管理与服务中,通过加强对用户行为数据的收集和分析,及时了解“谁,何时、何地、关心何内容?他们有什么共性?他们对数据的评价如何?”等信息,进而挖掘用户的兴趣,准确地向用户“推荐”地质资料,加强地质资料处理与服务的针对性。并且通过这种积累和分析,一方面基于用户实时反馈,及时改进策略,不断提升数据分析、处理能力,提高服务的准确度;另一方面,在地矿领域本体库的辅助支持下,可智能地理解用户需求,实现语义级查询检索,也可以通过互联网地矿信息快速收集整理,智能分析地矿行业社会舆情,甚至及时做出趋

势预判^[5]。因此,需要根据地质调查工作的特点,以用户业务需求为驱动,开发数据应用模式,创建数据挖掘模型,把握数据、信息与工具的业务命脉,创造数据的持续可利用价值^[6]。

2.6 重视自身数据人才培养,并与外界加强合作

大数据技术的关键在于对含有一定意义的较准确的庞大数据进行专业、智能化的处理,从中分析、挖掘出有价值的信息。数据挖掘和分析如果不够准确,会直接影响大数据分析和预测效果,有时甚至是与实际背离,所以强大的数据分析挖掘能力非常重要。就地质行业本身而言,现阶段数据和业务的结合还较为松散,这就需要更多既懂专业又具有大数据思维的专业人员,他们不但了解业务,也熟悉数据框架体系,同时能对现有专业数据指标进行思考和有效分析。

在互联网时代,地质行业是传统行业。传统行业要想在大数据时代有所作为,必须多方合作。一是获取更多更广泛的数据;二是寻求与大数据领域优秀的公司合作,进行数据的深度分析与挖掘。但要注意的是,地质数据专业性较强,所以还不能完全依赖大数据领域的公司进行分析,必须采取紧密的合作模式,才能开发出符合行业特点的大数据分析工具,得出正确、有效的分析预测结果。

2.7 采取先试点、后推广的策略

大数据在应用实施中,应采取先试点、后推广的策略^[7]。选择一些有代表性方向进行试点,投石问路。避免一哄而上,避免因顶层设计出现问题或模型错误造成损失。通过大数据试点,各方面的工作不断完善,大数据的作业模型不断改进,

为大数据的全面铺开打好基础。逐步实施、改进大数据架构,最终落实大数据应用,支撑行业的战略决策。

3 结论

大数据要想落地,必须有2个条件:一是丰富的数据源;二是强大的数据挖掘和分析能力。大数据时代下地质资料管理与服务也要从这两方面着手,做好数据的收集、汇聚工作,通过各种手段提升数据的挖掘和分析能力。制定合理的标准、政策方针,采取先试点、后推广的策略。使大数据在地质资料管理与服务工作中“落地”,使地质资料数据产生更大价值。

参考文献:

- [1] Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovation, competition, and productivity [R]. [s. l.]: McKinsey Global Institute, 2011.
- [2] 连健, 丁克永, 吴小平, 等. 地质资料管理与服务国外发展形势跟踪研究[J]. 中国矿业, 2013, 22(7): 63-67.
- [3] 贾文珏, 周舟. 地质资料信息集群化共享服务平台关键技术研究[J]. 国土资源信息化, 2012(5): 19-22.
- [4] 芦书文, 朱卫红. 美国地质资料信息服务发展研究[J]. 中国矿业, 2013, 22(9): 77-82.
- [5] 屈红刚. 大数据视角下的地矿工作发展与变革研究[J]. 国土资源信息化, 2014(5): 34-39.
- [6] Rajaraman A, Ullman J D. 大数据: 互联网大规模数据挖掘与分布式处理[M]. 王斌, 译. 北京: 人民邮电出版社, 2012.
- [7] 赵刚. 大数据咨询: 迈向成功的第一步[EB/OL]. 赵刚的博客, 2013.

Analysis of the big data “landing” in geological data management and service

HAN Yuan, ZHANG Hongying, LIANG Nan

(Xi'an Center of Geological Survey, China Geological Survey, Xi'an 710054, China)

Abstract: Big data has been a new IT hot word after cloud computing and internet of things and it connects everyone's life. Geological data includes basic geology, mine, geophysics and geochemistry. All of these data have significant application value for the nation and society. Combining the big data with geological data and service is worth investigating. The basic geological library was taken as an example in this paper to analyze the geological data management and present service and demand situation. Also the exploration of big data thinking, improvement of geological data management and service, realization of the efficient management and effectively extraction of data are considered in this research. This work gives a brief introduction of the great application value of geological data for whole society.

Key words: big data; geological data; information service; data sharing