

共享世界的支撑技术



时空大数据与云平台 (理论篇)

吴信才 著



科学出版社

共享世界的支撑技术

时空大数据与云平台

(理论篇)

吴信才 著

科学出版社

北京

内 容 简 介

本书结合现代 IT 技术、地理信息系统软件新技术、云计算、大数据、物联网等先进技术,详细介绍时空大数据概念与发展,基于云环境的时空大数据平台体系 T-C-V 软件结构和组成,重点阐述时空大数据中心、时空信息云服务中心和云应用集成管理中心三大部件,最后以全空间一张图平台为例,介绍该平台的实践情况。

本书采用泛在技术加案例建设方式,以软件平台设计流程为顺序编写,通俗易懂,为读者了解和学习云环境下的时空大数据平台提供了有力的参考。本书内容丰富、针对性强,可作为地理信息系统、软件工程、测绘工程等专业本科生和研究生的学习参考书,也可以作为城市规划、国土管理、市政工程、环境科学及相关专业研究和开发人员的应用宝典。

图书在版编目(CIP)数据

时空大数据与云平台. 理论篇/吴信才著. —北京: 科学出版社, 2018.4
(共享世界的支撑技术)

ISBN 978-7-03-056664-5

I. ①时… II. ①吴… III. ①数据处理 ②计算机网络 IV. ①TP274 ②TP393

中国版本图书馆 CIP 数据核字(2018)第 040029 号

责任编辑: 苗李莉 李 静 / 责任校对: 何艳萍

责任印制: 肖 兴 / 封面设计: 图阅社

科学出版社出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2018 年 4 月第 一 版 开本: 787×1092 1/16

2018 年 4 月第一次印刷 印张: 11

字数: 252 000

定价: 69.00 元

(如有印装质量问题, 我社负责调换)

前 言

随着云时代的来临,大数据(big data)吸引了越来越多的关注。截至2012年,数据量已经从TB($1\text{TB}=1024\text{GB}=1048576\text{MB}$)级别跃升到PB($1\text{PB}=1024\text{TB}=1048576\text{GB}$)、EB($1\text{EB}=1024\text{PB}=1048576\text{TB}$)等级别。整个人类文明所获得的全部数据中,有90%是过去两年内产生的。而到2020年,全世界所产生的数据规模将达到今天的44倍。由于移动互联网、物联网技术的应用,时空大数据的增长速率非常惊人。时空大数据与非空间数据相比,具有空间性、时间性、多维性、海量性、复杂性等特点,其云计算方法和挖掘技术是目前测绘科学技术的前沿领域之一。时空大数据的价值在于时间、空间、对象之间的关联关系,主要研究和探索数据与现实中对象、行为、事件间的对应规律,研究时空大数据高效表达与组织、时空大数据多维关联与协同计算,揭示大尺度事件的演化推理机理。

当前正在建设的智慧城市是在数字城市基础框架上,运用物联网、云计算、时空大数据集成等新一代信息技术,将现实的城市与数字城市进行有效融合,促进城市规划、建设、管理和服务智慧化的新理念和新模式。作为智慧城市的地理空间信息的载体,城市时空大数据云平台应管理、分析时空数据体系,实现土地、规划、交通、林业、水利、税务、民政、公安等公共部门及民用企业互联互通,形成时空地理信息资源共享交换机制,为政府决策、城市管理、社会公益服务等各行业提供基础的地理信息服务。

武汉中地数码科技有限公司(以下简称:中地公司)积累了几十年的地理信息系统(geographic information system, GIS)软件研发和工程应用经验,采用创新性的T-C-V软件结构,引入虚拟化技术、云计算技术等成熟、先进的技术手段,首创共享世界的支撑技术思想,基于MapGIS I²GSS(international internet GIS service sharing)云平台技术,开展了时空大数据与云平台的研发和工程实践。时空大数据与云平台实现了时空大数据分析、挖掘、自动化的运维管理、智能的集群部署和应用场景的个性化定制,提供基于云计算环境下的时空信息服务,开展了多领域智慧应用。

本书围绕中地公司的时空大数据与云平台展开描述,研究和探索云计算环境下时空大数据与云平台的存储、分析、应用模式与服务体系等问题,极具必要性和商业价值。本书共分为9章,系统地阐述了时空大数据与云平台的概念、体系、组成和应用。

本书由长江学者吴信才教授策划并撰写,参与本书数据资料的搜集及书稿整理的人员还有万波、吴亮、黄波、黄颖、刘永、黄胜辉、陈小佩等,他们长期从事GIS软件的研究、开发与服务,具有丰富的经验,使本书融入了技术团队在近年取得的研发成果。

由于时间紧,水平有限,不妥之处在所难免,敬请读者提出宝贵意见。

作 者

2017年10月

目 录

前言

第1章 绪论	1
1.1 共享世界的认知	1
1.2 时空大数据认知	6
1.3 时空大数据的发展	11
1.4 时空大数据与云平台构建的必要性	15
第2章 面向大数据与云服务的 T-C-V 软件结构	18
2.1 GIS 架构技术发展的四个阶段	18
2.2 面向大数据云服务的 T-C-V 软件结构	23
2.3 T-C-V 软件结构对共享世界的价值	31
第3章 时空大数据与云平台架构	37
3.1 平台体系概述	37
3.2 平台架构要素分解	39
3.3 时空大数据与云平台关键要素	42
第4章 云环境下时空大数据中心体系	46
4.1 时空大数据处理内涵	46
4.2 时空大数据中心体系架构	53
第5章 云环境下时空大数据平台构建	58
5.1 时空大数据库构建	58
5.2 时空大数据分布式数据资源管理	75
第6章 时空信息云服务中心	81
6.1 时空信息云服务中心概述	81
6.2 时空信息云服务分类	82
6.3 时空信息云服务定制	85
6.4 分布式云服务协调管理	89
第7章 时空云应用集成管理中心	104
7.1 云应用集成管理中心概念模型	104
7.2 云应用集成管理中心特点	110
7.3 云应用集成管理中心组成	114
第8章 全空间一张图实践	132
8.1 全空间一张图建设目标	132
8.2 全空间一张图整体框架	133
8.3 全空间一张图大数据中心建设	135

8.4 全空间一张图云服务中心建设.....	141
8.5 全空间一张图集成管理中心建设.....	155
8.6 全空间一张图的价值.....	161
第9章 结语.....	164
9.1 总结.....	164
9.2 机遇与挑战.....	166
参考文献.....	170

1.1 绪论.....	1
1.2 研究背景.....	1
1.3 研究意义.....	1
1.4 研究内容.....	1
1.5 研究方法.....	1
1.6 研究思路.....	1
1.7 研究目标.....	1
1.8 研究范围.....	1
1.9 研究重点.....	1
1.10 研究难点.....	1
1.11 研究创新点.....	1
1.12 研究结论.....	1
1.13 研究展望.....	1
1.14 研究总结.....	1
1.15 研究致谢.....	1
1.16 研究参考文献.....	1
1.17 研究附录.....	1
1.18 研究索引.....	1
1.19 研究目录.....	1
1.20 研究后记.....	1
1.21 研究声明.....	1
1.22 研究版权.....	1
1.23 研究隐私.....	1
1.24 研究安全.....	1
1.25 研究合规.....	1
1.26 研究透明.....	1
1.27 研究负责.....	1
1.28 研究诚信.....	1
1.29 研究公正.....	1
1.30 研究公开.....	1
1.31 研究合作.....	1
1.32 研究共赢.....	1
1.33 研究发展.....	1
1.34 研究进步.....	1
1.35 研究繁荣.....	1
1.36 研究富强.....	1
1.37 研究民主.....	1
1.38 研究法治.....	1
1.39 研究和谐.....	1
1.40 研究文明.....	1
1.41 研究健康.....	1
1.42 研究幸福.....	1
1.43 研究美好.....	1
1.44 研究希望.....	1
1.45 研究未来.....	1
1.46 研究梦想.....	1
1.47 研究理想.....	1
1.48 研究信念.....	1
1.49 研究力量.....	1
1.50 研究智慧.....	1
1.51 研究勇气.....	1
1.52 研究毅力.....	1
1.53 研究恒心.....	1
1.54 研究决心.....	1
1.55 研究信心.....	1
1.56 研究耐心.....	1
1.57 研究细心.....	1
1.58 研究专心.....	1
1.59 研究恒心.....	1
1.60 研究决心.....	1
1.61 研究信心.....	1
1.62 研究耐心.....	1
1.63 研究细心.....	1
1.64 研究专心.....	1
1.65 研究恒心.....	1
1.66 研究决心.....	1
1.67 研究信心.....	1
1.68 研究耐心.....	1
1.69 研究细心.....	1
1.70 研究专心.....	1
1.71 研究恒心.....	1
1.72 研究决心.....	1
1.73 研究信心.....	1
1.74 研究耐心.....	1
1.75 研究细心.....	1
1.76 研究专心.....	1
1.77 研究恒心.....	1
1.78 研究决心.....	1
1.79 研究信心.....	1
1.80 研究耐心.....	1
1.81 研究细心.....	1
1.82 研究专心.....	1
1.83 研究恒心.....	1
1.84 研究决心.....	1
1.85 研究信心.....	1
1.86 研究耐心.....	1
1.87 研究细心.....	1
1.88 研究专心.....	1
1.89 研究恒心.....	1
1.90 研究决心.....	1
1.91 研究信心.....	1
1.92 研究耐心.....	1
1.93 研究细心.....	1
1.94 研究专心.....	1
1.95 研究恒心.....	1
1.96 研究决心.....	1
1.97 研究信心.....	1
1.98 研究耐心.....	1
1.99 研究细心.....	1
1.100 研究专心.....	1

第1章 绪 论

随着云计算、物联网的发展，我们迎来了大数据的时代。大数据之父维克托·迈尔-舍恩伯格提出：大数据的核心要义在于共享。

作为具有强大变革能力的大数据，不仅引发技术革命、经济变革，更引发了人类的认知变革。从对现实世界认知，到概念世界的认知，到数字世界的认知，无不经历着一次又一次的认知变革。然而大数据是共享型资源，带来的变革是全方位的，影响的人越来越多，资源交换共享越来越多，催生了“共享世界”的出现。地理信息共享作为一种智力资源将共享价值变现，利用云计算、大数据的共享理念，产生经济效益，通过地理信息共享让人人享有地理信息，让人人低成本参与地理信息共享。

1.1 共享世界的认知

1.1.1 现实世界认知过程

所有动物都有认知（cognition）的本能，尤其人类更为突出。德国哲学家康德认为：时间和空间是人类的“先验认知”，所谓“先验”就是先于经验的，即还没有任何知识的时候，时空就已经存在每个人的意识当中，人对所有事物的认知要放在时间和空间的框架中来，时空即成为了所有知识的共同基础，正是这样，人类才能清晰的认知事物之间关系。认知包括感觉、知觉、思考、想象、推理、求解、记忆、学习和语言等。人类生活在地球上，人类的一切活动无不与空间认知（spatial cognition）息息相关。在信息快消时代，什么时间、什么地点、发生了什么事情、为什么发生在这里、事发地点的环境及其周围环境的关系，这些都是人们比以往任何时间都更为关心的问题。如何将时间、地点、事务、环境等因素融合起来，综合分析，更有利于人们做出判断和决策，这同时也是一个将客观世界的地理现象转化为抽象表达的数字世界相关信息的过程，这个过程涉及三个层面：现实世界、概念世界和数字世界，如图 1-1 所示。

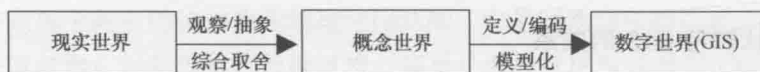


图 1-1 认知过程

(1) 现实世界是存在于人们头脑之外的客观世界，事物及其相互联系就处在这个世界之中。人们的认知来源于现实世界。认知理论早在 20 世纪 70 年代就被引入地图学，并用“刺激-反应”的关系模式来研究用图者在读图时的心理——物理反应。现实世界是非常复杂的，人们所见所得从早期的纸质地图记载，逐渐演进为计算机系统输入输出。人们借助计算机系统将现实世界各种各样的地理现象进行模拟再现，然而机器无法代替

人脑，许多复杂的现象、海量的信息无法识别。要正确认识和掌握现实世界这些复杂、海量的信息，需要进行去粗取精、去伪存真的加工。因此，对现实世界的认识是一个从感性认识到理性认识的抽象过程。

(2) 概念世界是现实世界在人们头脑中的反映。客观事物在概念世界中称为实体，反映事物联系的是实体模型。实体模型又分为逻辑模型和物理模型。人们研究发现，人脑就是一个信息加工系统，人类对外界的知觉、记忆、思维等一系列认知过程，可以看成是对信息的产生、接收和传递的过程，计算机和人脑两者的物质结构大不一样，但计算机软件所表现出的功能和人的认知过程是类同的，即两者的工作原理是一致的，都是信息加工系统：输入信息、进行编码、存储记忆、做出决策和输出结果。因此，空间认知是一个信息加工过程，将各种空间认知抽象成概念模型。

(3) 数字世界是概念世界中信息的数据化。现实世界中的事物及联系在这里用数据模型描述。计算机通过对各种地理现象的观察、抽象、综合取舍，得到实体目标（有时也称为空间对象），然后对实体目标进行定义、编码结构化和模型化，以数据形式存入计算机内。空间认知处理现实世界的空间属性，依赖于位置、大小、距离、方向、形状、格局、移动，以及事物间关系的认知。其中时间、空间、属性，形成了地理空间信息的三要素。人们对地理信息的获取、认知亦来源于这三个要素。因此空间数据表示的基本任务就是将图形模拟的空间物体（时间、空间、属性）表示成计算机能够接受的数字形式，完成逻辑数据模型向物理数据模型的转变，如图 1-2 所示。

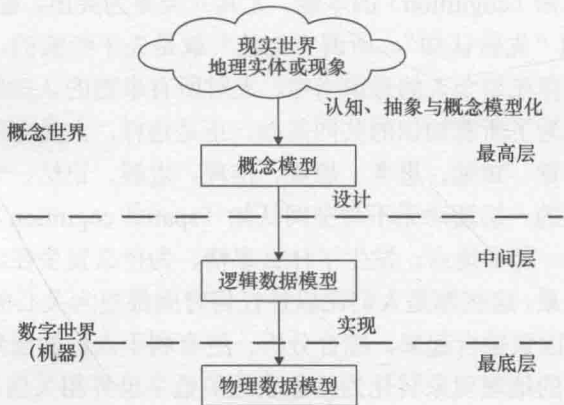


图 1-2 认知模型

1.1.2 空间认知与 GIS 的关系

1) 空间认知对 GIS 的意义

关于空间和场所的空间认知表达了人与环境及人与地球之间的关系，地理学家希望空间认知能够有助于理解他们感兴趣的传统现象，如人们选择在哪里购物取决于他们对于距离和道路连接的认识。

目前对 GIS 认知问题的关注程度不够、理解不够是地理信息技术有效性的一个主要障碍。一定程度上，空间认知和表达过程代表了 GIS 发展的过程。认知研究将直接导致 GIS 系统的改进，改进后的系统将充分体现人类的地理感知，空间认知有助于提高 GIS

使用和设计，以及其他地理信息产品，它们部分依赖于人们对于空间关系表述的理解。因此，认知问题的研究对设计更有效的 GIS 是有帮助的。

2) GIS 与空间认知的关系

地理空间认知是地理学的一个重要研究领域，是地理认知理论之一和 GIS 数据表达与组织的桥梁和纽带，研究地理空间认知对 GIS 的建立具有重要作用。认知、空间认知、地理空间认知，以及 GIS 之间有着紧密的关系，如图 1-3 所示。

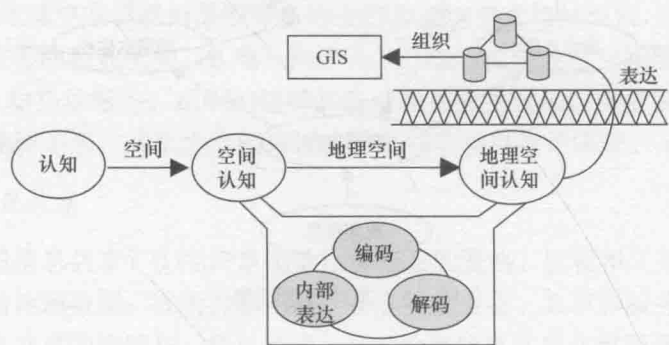


图 1-3 GIS 与空间认知关系

地理知识的描述需要地理思维，它们与 GIS 相结合会产生基于知识的 GIS 和基于 GIS 的专家系统两种结果。这两种系统都是以地理认知为基础的。地理认知、地理思维和 GIS 的关系可用图 1-4 来描述。

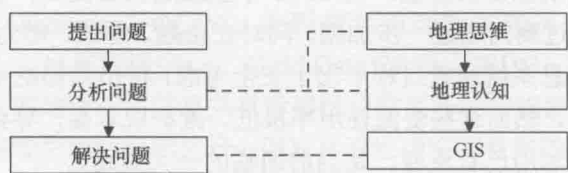


图 1-4 地理认知、地理思维和 GIS 的关系

GIS 空间数据组织的对象来源于现实世界的地理现象（客观实体），因此，必须对现实世界进行抽象和表达，以建立现实世界的 GIS 数据模型，抽象的过程是人们对现实世界进行认知的过程，表达的过程是人们对现实世界进行计算机再现的过程。

发生在地理空间上的认知称为地理空间认知，它是对地理空间信息的表征，包括感知过程、表象过程、记忆过程和思维过程，实质是对地理现象或地理空间实体的编码、内部表达和解码的过程。本体论是一种对于现实世界的概念结构进行系统化描述的方法论。本体论通过研究概念世界、GIS 语言世界和地理现实世界，产生有关地理世界结构更好的理解，为 GIS 的发展提供更合理的概念模型，从而避免现有的数据模型与人类空间认知机制的巨大反差。GIS 是信息获取、储存、处理和输出的信息系统，将 GIS 本身称为内部世界；相对应地将 GIS 研究的自然和人文环境称为外部世界。外部世界的物质称为实体，对实体描述的信息称为实体对象，如图 1-5 所示。通过对实体对象的研究实现对外部世界及其运行规律的揭示。GIS 通过构建空间认知体系，具有了获取、存储和

处理外部世界空间信息的能力，从宏观和微观两个层面，揭示外部世界的特性及其运行规律。

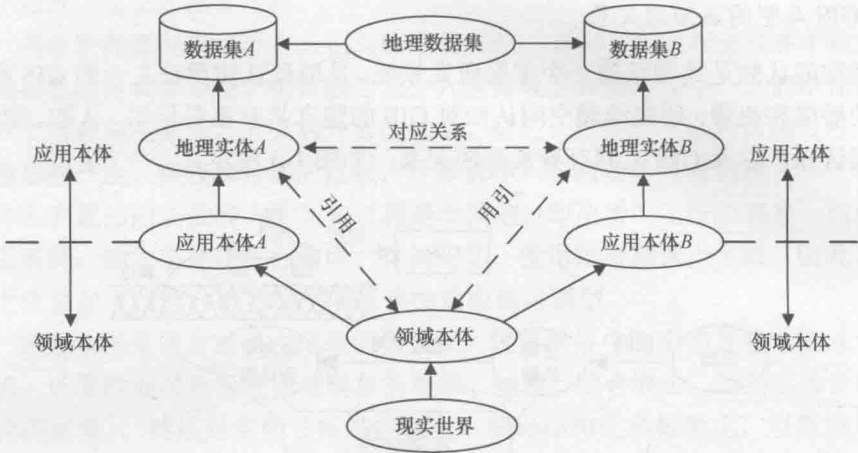


图 1-5 本体概念模型

1.1.3 地理信息共享世界

1) 共享世界的出现

自 21 世纪全球经济体进入高速发展期以来，经济高度繁荣，生产力快速发展，导致生产关系不匹配生产力的发展，需求增长跟不上产能增长，全球产能过剩。在全球产业博弈中，产能过剩现象愈演愈烈。自 2008 年金融危机爆发后，金融风暴席卷全球，需求萎缩严重，产能过剩问题进一步加剧。同时在金融、通信、电力、地理信息等领域，许多大中型企业在信息系统建设过程中出于安全考虑，付出高昂的价格来购买足够的计算资源、存储资源等，然而这些资源使用率极低、维护成本高，导致资金、资源供需不对等。如何充分有效利用已有资源，是当前面临的一大难题。

进入 2010 年后，网络技术、手机通信的发展与普及，使得共享世界离我们越来越近，不再是空中楼阁遥不可及。人们借助网络、地理信息、计算机技术尝试着共享金融、共享物品、共享空间、共享出行、共享知识，逐步探索出共享世界的要义：提高资源的利用率，对重组量化后的资源进行再分配，满足供需双方的需求，实现人力、物力、智力等多方面资源共享、优势互补。由于涉及生产资料的使用与共享，“共享世界”的出现需以“共享经济”为基础。“共享经济”这一定义是由杰里米·里夫金明确给出的，他认为，协同共享是一种新的经济模式。共享经济要经过几个阶段：第一个阶段叫物力共享，把空间、物体拿出来共享；第二个阶段叫人力共享，像滴滴把人力和自己的技能共享出来；第三个阶段叫脑力共享，这是一个智力时代，每个人都有盈余的时间。物力共享是物的交换价值，人力共享是人的价值，真正脑力共享时代，共享经济才成为了一个更大的概念，脑力共享不是齿轮带动齿轮的物理概念，而是一个相互作用与反应的化学概念。21 世纪云计算、大数据的出现，催生了这种“化学反应”。大数据的新模式、新架构、新文件系统的诞生打破了原有的存储、计算方式方法，使得资源供需不再是难题。云端计算资源、存储资源、网络资源的共享促进了共享世界的发展。大热的云计算降低

了技术开发成本,数据访问的灵活性、移动性、扩展性等性能得以极大提升;大数据使得海量异构数据的快速处理成为可能,且能实现供需双方精准匹配,根据用户的搜索和行为推送用户需求;地理信息技术能够实时定位消费者地理位置,全球卫星导航系统(global navigation satellite system, GNSS),如GPS、北斗等,定位精度提高,大数据的位置标签更加精准,使空间共享可见可得。

共享世界的思维打破传统经济对商业组织的依附,整合线下各种角色、应用和资源,削弱传统经济对商业组织的依附。地理空间信息强调使用权,淡化拥有权,使用比拥有更有价值。同时地理空间信息商品或服务的生产者和消费者边际模糊,产销者(prosumer)出现,使得从过去的商业机构(B端)向个人(C端)提供地理空间信息服务,转向更多的C端向C端提供服务。组织的结构和雇佣关系发生变化,更多以“商业伙伴”形式出现,没有雇佣关系。共享世界因为地理信息共享得以真正实现。

2) 地理信息共享

什么是地理信息共享?①用户享用非己有信息的资格、权利和义务;②生产为实现信息共享必须的地理数据;③为共享信息准备必要的设备、共享和服务环境;④政府对发展和协调信息共享的调控权。简单地说,目前地理信息共享在我国是指在政府宏观调控下,依据一定的规则和法律,实现地理信息的流通和共用。

地理信息标准化和规范化是实现地理信息共享的前提条件之一。因此,在国际上,特别是北美、西欧等许多技术比较发达的国家及众多的国际组织十分重视地理信息的统一技术标准、统一政策和统一技术体系问题,最终让国家乃至全球的地理信息具有共享性。例如,ISO/TC211(国际标准化组织地理信息技术委员会)、开放地理空间信息联盟(Open Geospatial Consortium, OGC)、美国联邦地理数据委员会(Federal Geodata Commission, FGDC)、世界数据中心(World Data Center, WDC)等众多机构的建立及推广,使得发达国家的地理空间信息共享管理已经发展得比较成熟,形成自身的特点:一是由政府决定对地理空间信息的生产,国家具有地理空间数据的主导地位;二是要突显出地理空间信息的公益性原则,明确规定提供免费共享信息或廉价提供;三是地理空间信息的应用已经渗透到生活乃至家庭、社会的每一个角落及每一个行业;四是有严格的法律,标准更加完善。20世纪80年代以来,我国科研界就对地理信息共享问题非常重视,主要方面是研究统一的技术标准和相关政策的制定。1983年,国家科学技术委员会高技术基础司组织开展跨部门“资源与环境信息系统国家规范研究”,并于1984年出版了《资源与环境信息系统国家规范与标准研究报告》。1991年,国家测绘局组织开展了10余项地理信息共享与标准等问题的研究。1996~2005年的10年中,我国的地理信息共享的标准化进行了大量的研究与实践,并取得了丰硕的成果,对于地理信息共享政策与法律法规也进行了一些研究。

由于信息不对称、资源不对等问题,导致我国地理信息共享在较长一个时期滞留在政企部门,甚至在政企部门之间还存在“信息孤岛”。长期的资源累积和资源不对等,阻碍了地理信息面向社会的数据共享,造成科技信息资源的巨大浪费,触发了地理信息分享需求的增长。社会的需求引导了地理信息分享的方向。用户需求驱动消费,消费定制生产,各类地理信息数据共享、服务共享、应用共享如雨后春笋般涌现,个人需求得到极大满足,地理信息上中下游产业链因为共享再次盘活。

在此形势下，我国传统地理信息产业要适应全球共享经济模式，需分三步走。

第一步：开放式共享，通过网络，让地理空间信息以最快、最直观的方式实现共享。

第二步：融合式共享，在开放式共享基础上，提供一种更为开放的地理空间信息体系架构和生产模式，建立不同时空陌生人之间的黏性，融合全球所有的人力、智力、物力，根据不同的需求不断产生、交易、迁移、聚合、重构成各种各样的地理空间信息应用，让有限的资源变成无限的可能。

第三步：可持续发展式共享，在融合式共享的基础上，建立一个功能完备的生态系统，如图 1-6 所示，包含地理空间信息发展涉及的需求、生产、交付、服务、集成等各个环节，打破传统经济对商业组织的依附，整合线下各种角色、应用和资源，让每一个人都可以在这个无限的、充满想象的、可不断扩展的生态圈中自由享有自己关注的信息、服务，构建属于自己的关注圈子甚至完整的行业生态圈。

一个新的功能完备的生态系统因地理信息共享世界而诞生。



图 1-6 共享世界生态系统图

1.2 时空大数据认知

1.2.1 传统数据与大数据

大数据时代的来临，让人们更加关注和探究大数据的来源。1997 年美国国家航空航天局(National Aeronautics and Space Administration, NASA)研究员 Michael Cox 和 David Ellsworth 在 IEEE 第 8 届国际可视化学术会议中首先提出了“大数据”术语；2008 年 9 月 *Nature* 学术杂志发表的一篇名为 *Big Data: Wikiomics* 的文章，使大数据在科学研究领域得到了高度重视；2012 年 3 月美国政府发布《大数据研究和发展倡议》，旨在提高和改进人们从大数据中获取知识的能力。大数据引起了主要国家和全社会的重视，一场大数据引发的变革渗透到各个角落。面对大数据时代的挑战与机遇，国际上的专家学者针对大数据处理开展了一系列的探索和研究。2012 年在印度新德里举行的首届大数据分析国际会议上，与会代表达成共识，认为大数据的表达、检索、挖掘是大数据处理面临

的三大挑战。目前，对大数据的研究已经逐渐发展成为信息科学的主要研究趋势。2014年我国大数据市场规模达到 84 亿元人民币，依照中国信息通信研究院的估算，2015 年达到了 115.9 亿元，增速 38%。2016~2018 年中国大数据市场规模增长还将维持在 40% 左右的高增长率，这背后的根本变化在于数据源、传感器将成为主要的数据来源：遥感卫星很快会实现“准实时”的对地观测，数以万计的无人机时刻不停的获取数据，移动通信中每个人都在实时产生位置信息，物联网的发展将带来更大量级的数据内容。大数据在经过了近几年的发展后，已经被民众接纳，获得了社会的广泛认可。

然而什么是大数据呢？来自维基百科的定义：大数据是一个复杂而庞大的数据集，以至于很难用现有的数据库管理系统和其他数据处理技术来采集、存储、查找、共享、传送、分析和可视化。

2013 年 5 月召开的第 462 次香山科学会议——数据科学与大数据的科学原理及发展前景 2013，给出了技术型和非技术型两个定义。

(1) 技术型定义：大数据是来源多样、类型多样、大而复杂、具有潜在价值，但难以在期望时间内处理和分析的数据集。

(2) 非技术型定义：大数据是数字化生存时代的新型战略资源，是驱动创新的重要因素，正在改变人类的生产和生活方式。

相对于传统数据，大数据具有数据量大、数据种类多、要求实时性强、数据所蕴藏价值大的特点，见表 1-1。大数据是具有体量大、结构多样、时效强等特征的数据；处理大数据需采用新型计算架构和智能算法等新技术；大数据的应用强调以新的理念应用于辅助决策、发现新的知识，更强调在线闭环的业务流程优化。

传统数据——从数据来看，传统数据管理针对的是过去一段时间内已知范围内的易于理解的数据；从处理工具来看，传统数据管理要求高效、高吞吐处理数据，并未有严格的时限要求；从数据算法来看，传统数据管理统计分析主题关系早已确立且不变。总体而言，传统数据呈现的特点主要归纳为两点：静态、已知。

大数据——大数据技术针对的是实时产生的大量结构化及非结构化数据；从数据处理来看，大数据技术要求实时处理数据；从数据算法来看，大数据技术探究的是建立算法模型，基于实时数据不断优化。大数据呈现的特点主要归纳为两点：动态、未知。

表 1-1 传统数据和大数据比较

	传统数据	大数据
数据量	GB	TB PB
数据产生速率	每小时、每天……	实时、非常快速
数据结构	结构化数据	半结构化、非结构化
数据源	集中的分布式	分布式
数据集成	简单	困难
数据库	RDBMS	HDFS、NOSQL
数据接口	交互式	批处理、实时

1.2.2 大数据与时空大数据

大数据时代产生了大量的具有时空标记、能够描述个体行为的时空大数据，如手机数据、出租车数据、社交媒体数据等。这些数据为人们进一步了解社会经济环境提供了

一种新的技术手段。近年来, 计算机科学、地理学和复杂性科学领域的学者基于不同类型数据开展了大量研究, 试图发现海量群体的时空行为模式, 并建立合适的解释性模型。通过时空大数据可以解决许多以往难以解决的复杂问题。麦肯锡在《大数据: 下一个竞争、创新和生产力的前沿领域》的研究报告中认为, 医疗保健、零售业、公共领域、制造业和个人位置的数据构成了目前 5 种主要的大数据流, 上述无论哪种数据流都具有显著的地理编码与时间标签。从这个角度看, 时空信息不仅是大数据的重要组成部分, 更可被看成是大数据本身(边馥苓等, 2016)。

时空数据是数据的一种特殊类型, 它是指带有空间坐标的数据, 这类数据通常是地图文件, 用点、线、面及实体等基本空间数据结构来表示。一个地图文件通常只包含一种类型的空间数据结构, 如面(代表国家或者地区)、线(代表道路或者河流)或点(代表特定的地址)。如果想要比较复杂的地图文件, 其中包含多种空间数据结构的话, 通常需要多个地图文件叠加来获得。除了地图信息, 时空数据还包括地图信息的背景数据, 用来描述地图文件上的对象属性。例如, 一个地图文件包含街道, 那么就需要相应的背景数据来描述该街道的大小, 名字或者一些分类信息(分行道、单行道、双行道、禁止通行等)。因此, 结合时空数据的特征, 时空大数据可定义为: 是指用来表示空间实体的地理位置和分布特征等方面信息的数据, 表述了空间实体或目标事件随地理位置的不同而发生的变化。

时空大数据由于其所在空间的空间实体和空间现象在时间、空间和属性三个方面的固有特征, 呈现出多维、语义、时空动态关联的复杂性, 时空大数据一方面具有一般大数据的大规模、多样性、时效性和价值性的特点, 另一方面还具有与对象行为对应的多源异构和复杂性, 与事件对应的时空、尺度、对象动态演化, 对事件的感知和预测特性, 如图 1-7 所示。

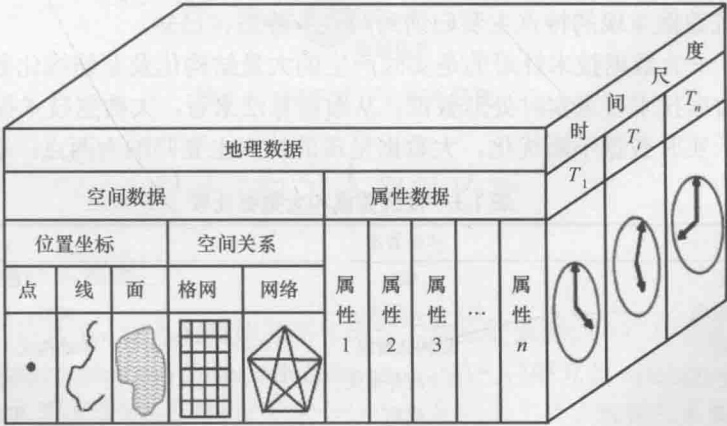


图 1-7 时空大数据数据维度表达

1.2.3 时空大数据特征

几乎所有的大数据, 都需要而且可以与时空数据融合。
时空大数据是一种结构复杂、多层嵌套的具有空间和时态特性的多维数据, 它有效记录了事物的空间位置和时空变化过程, 并准确地表达了事物的历史、当前和未来状态,

如城市变迁、疾病扩散、环境变化、地质演化、移动对象位置变更等。

时空大数据作为大数据行业一个重要分支,是指同时具有时间和空间维度的数据,现实世界中的数据超过 80%与地理位置有关。时空大数据包含对象、过程、事件在空间、时间、语义等方面的关联关系。空间性是时空数据区别于其他数据的标志性特征。

由于时空大数据的固有特点,因此呈现出多维、语义、时空动态关联的复杂性。特别是随着智慧城市和物联网的迅猛发展,无所不在的传感器网将产生极其大量的数据,使得世界进入真正的大数据时代,其中大量的与时空位置有关的时空大数据需要被存储、分析和处理。由于应用范围的日益广泛,时空大数据成为了大数据领域越来越重要的分支之一。作为典型的特定领域内的大数据,它具有大数据的主要特点:数据规模极大、数据间关联性复杂、类型多样化、时效性高等。同时,时空大数据还具有空间信息领域内的特征,具体特征包括以下七点。

(1) 无所不在 (ubiquitous)。一方面,在大数据时代,数据获取将从空天地专用传感器扩展到物联网中上亿个无所不在的非专用传感器。例如,智能手机,它就是一个具有通信、导航、定位、摄影、摄像和传输功能的时空数据传感器;又如城市中具有空间位置的上千万个视频传感器,它能提供 PB 和 EB 级连续图像。这些传感器将显著提高数据获取能力。另一方面,在大数据时代,GIS 的应用也是无所不在的,它已从专业用户扩大到全球大众用户。

(2) 多维动态 (multi-dimension and dynamics)。大数据时代无所不在的传感器网以日、时、分、秒甚至毫秒计产生时空数据,使得人们能以前所未有的速度获得多维动态数据来描述和研究地球上的各种实体和人类活动。例如,智慧城市需要从室外到室内、从地上到地下真三维高精度建模,基于时空动态数据的感知、分析、认知和变化检测在人类社会可持续发展中将发挥越来越大的作用。通过这些研究,GIS 将对模式识别和人工智能作出更大的贡献。

(3) 互联网+网络化 (internet+networking)。在越来越强大的天地一体化网络通信技术和云计算技术支持下,形成互联网+空间信息系统,将时空大数据从专业应用向大众化应用扩展。原先分散的、各自独立进行的数据处理、信息提取和知识发现等将在网络上由云计算为用户来完成。

(4) 全自动与实时化 (full automation and real time)。在网络化、大数据和云计算的支持下,地球空间信息学有可能利用模式识别和人工智能的新成果来全自动和实时地满足军民应急响应用户和诸如飞机与汽车自动驾驶等实时用户的要求。

(5) 从感知到认知 (from sensing to recognizing)。在大数据时代,通过对时空大数据的数据处理、分析、融合和挖掘,可以大大地提高空间认知能力。例如,利用智能手机中连续记录的位置数据、多媒体数据和电子地图数据,可以研究手机持有人的行为学和心理学。

(6) 众包与自发地理信息 (crowd sourcing and volunteered geographic information)。在大数据时代,基于无所不在的非专用时空数据传感器(如智能手机)和互联网云计算技术,通过网上众包方式,将会产生大量的自发地理信息来丰富时空信息资源,形成人人都是地球空间信息员的新局面。但他们的非专业特点,使得所提供的数据具有较大的噪声、缺失、不一致性、歧义等问题,引起数据有较大的不确定性,需要自动进行数据

清理、归化、融合与挖掘。当然,如能在网上提供更多的智能软件和开发工具,将会产生好的效果。

(7) 面向服务(service oriented)。地球空间信息学是一门面向经济建设、国防建设和大众民生应用需求的服务科学。它需要从理解用户的自然语言入手,搜索可用来回答用户需求的数据,优选提取信息和知识的工具,形成合理的数据流与服务链,通过网络通信的聚焦服务方式,将有用的信息和知识及时送达给用户。从这个意义上看,地球空间信息服务的最高标准是在规定的时间(right time)将所需位置(right place)上的正确数据/信息/知识(right data/information/knowledge)送到需要的人手上(right person),实现服务代替产品,以适应大数据时代的需求。

1.2.4 时空大数据基本要求

在大数据时代,如果要想实现多源异构时空大数据的融合,进行数据分析、决策,成为“智慧的时空数据”,必须构建一个独立于具体数据的基础框架,搭建一个云平台,才能够实现数据、资源、应用之间的融会贯通,可以称之为新一代的时空数据模型,需满足以下三个条件。

1. 以顶层设计、共享先行为原则规划时空大数据接入和共享机制

第一,具有全局性和独立性。所谓全局性,有两个含义:其一是可以作为承载所有类型数据的共同框架,其二是成为跨越数据集之间的全局索引。所谓独立性,是指独立于具体的数据内容而存在,不依赖特定的数据类型。第二,继承传统 GIS 数据。新的时空数据模型需要兼容和继承传统 GIS 数据,可以方便调用、抽取和整合数据。第三,适应各类传感器数据。新的时空数据模型需具备组织、管理和调度传感器数据的能力。传感器数据是时空大数据的主要来源,包括遥感数据、各类监测传感器和泛物联网数据。第四,具备时空一致性和稳定性。时空定义模式要保证不同时空尺度下时空定义的一致性和稳定性,使时空范围可以作为数据汇聚和融合的线索。第五,适应大数据 IT 架构。数据模型需要适应分布式计算、分布式存储和机器学习。第六,普适性。时空属性是所有数据的共同特征,新的时空数据模型可为任意类型的数据添加时空标签。

2. 以应用导向、创新驱动为原则实现时空大数据与应用领域的深度结合

时空大数据非常之多,且增长速度迅猛。只有有序的挖掘才能出现价值。目前时空大数据分为五大形态:第一是基础地理信息数据,即按照国家标准精准测绘所形成的政府数据;第二是专题地理信息数据,如地理国情所获取的大量与经济、社会有关的数据;第三是专业地理信息数据,如农业部门、国土部门等专业部门的数据;第四是传感地理信息数据,也是时空大数据里面重要且有活力的部分,并涵盖很多的领域;第五是社交地理信息数据,如采集的共享单车所走的轨迹等数据。其中第四类和第五类数据活跃度最高,也是最有挖掘和变现价值的时空数据。以应用为导向,盘活时空数据,实现时空大数据与应用领域的深度结合。

3. 以资源池化共享、虚拟存储为原则实现时空大数据资源的有效融合

实现存储资源的服务化,实现基础资源的充分统筹共享,提高存储资源利用率,减

少信息化建设成本,具体而言:包括搭建高效的数据库业务平台和灵活的云计算资源池平台,同时采用存储虚拟化功能实现现有异构存储资源的共享和统一管理,形成灵活的存储资源池。只有实现了现有资源整合与虚拟化,提高存储资源利用率,才能为构建时空大数据提供一个高性能可匹配业务不断发展的动态资源平台,同时也为客户将现有传统的数据中心架构向时空大数据与云平台过度奠定了基础,提供了平台和架构保障。

1.3 时空大数据的发展

1.3.1 时空大数据的发展历史

人类历史上,从未有哪个时代和今天一样产生如此海量的数据。数据的产生已经完全不受时间、地点的限制。从采用数据库作为数据管理的主要方式开始,人类社会的数据产生方式大致经历了运营式系统阶段(被动产生)、用户原创内容阶段(主动产生)、感知式系统阶段(自动产生)三个阶段,而正是数据产生方式的巨大变化才最终导致大数据的产生。在感知式系统阶段,人类保存数据的能力(数据存储时长、数据存储容量等)增强,使用数据的能力(应用、二次开发等)增强,挖掘数据的能力(决策分析、风险规避等)增强。地理信息行业作为海量时空数据的生产者,从数据采集到数据在各行各业中的应用可能都会在智能化时代中由机器完成。随着科学进步,人类对时空服务的需求正在从事后走向实时和瞬间、从静态走向动态和高速、从粗略走向精准和完备、从陆地走向海洋和天空、从区域走向全球、从地球走向深空和宇宙。地理信息的存在性需求,决定了它在智能化时代中不会消失,但必须完成从信息化到智能化的转型。这一转型极大地提高和完善了时空数据的生命周期。时空大数据的时代到来了!

国际上,欧美等国家早在 20 世纪,特别是北美、西欧等地区许多技术比较发达的国家,以及众多的国际组织投入大量人力和资金开展研究,未雨绸缪构建时空数据体系。因此在诸多地理信息共享标准上占得先机。国际地理信息产业市场主要分布在北美和西欧,美国在地理信息市场所占份额居于全球领先地位,拉丁美洲、东欧、中东和亚太地区地理信息产业市场也正在蓬勃兴起。预计 2016~2020 年,全球地理信息产业市场的年复合增长率将达到 18.25%(爱尔兰商业咨询公司 Research and Markets “2016~2020 全球地理信息系统分析市场”报告)。在我国,地理信息企业规模和公众市场不断扩大,产业发展质量得到显著提升。2016 年我国地理信息产业总产值预计达到 4360 亿元,同比增长 20.1%。截至 2016 年 10 月底,全国测绘资质单位数量突破 1.7 万家,较 2015 年年末增加 6.9%(11 月 1 日,2016 年中国地理信息产业大会,国家测绘地理信息局副局长宋超智讲话)。我国科研界对地理信息共享问题在 20 世纪 80 年代逐渐提上日程,主要方面是研究统一的技术标准和相关政策的制定。1984 年出版了《资源与环境信息系统国家规范与标准研究报告》。1996~2005 年的 10 年中,我国的地理信息共享的标准化进行了大量的研究与实践,并取得了丰硕的成果;对于地理信息共享政策与法律法规也进行了一些研究。21 世纪计算机技术的飞速发展推动了全球智能化、智慧化进程。现在很多地方在建立智慧城市(smart city),需要大量地理信息数据做支撑。智慧城市这一概念发端于 80 年代的信息城市(information city),经历了 90 年代的智能城市(intelligent