



《大数据导论（通识课版）》

教材官网: <http://dbllab.xmu.edu.cn/post/bigdataintroduction/>

温馨提示: 编辑幻灯片母版, 可以修改每页PPT的厦大校徽和底部文字

第9章 数据开放

(PPT版本号: 2019年秋季学期)



扫一扫访问教材官网

林子雨

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://www.cs.xmu.edu.cn/linziyu>





课程教材

- 林子雨 编著 《大数据导论——数据思维、数据能力和数据伦理（通识课版）》
- 高等教育出版社，2019年11月



提纲

- 9.1 政府开放数据的理论基础
- 9.2 政府信息公开与政府数据开放的联系与区别
- 9.3 政府数据开放的重要意义
- 9.4 国外政府开放数据的经验
- 9.5 国内政府开放数据
- 9.6 政府数据开放的几点启示



高校大数据课程

公共服务平台

百度搜索厦门大学数据库实验室网站访问平台





9.1 政府开放数据的理论基础

9.1.1 数据资产理论

9.1.2 数据权理论

9.1.3 开放政府理论



9.1.1 数据资产理论

- 在大数据时代，数据已经被当作一种重要的战略资源，也可以成为一种资产。数据资产是无形资产的延伸，是主要以知识形态存在的重要经济资源，是为其所有者或合法使用者提供某种权利、优势和效益的固定资产
- 数据资产的类型有很多，常见的数据资产包括书面技术新材料、数据与文档、技术软件、物理资产（主要指通信协议类）、员工与客户（包括竞争对手）、企业形象和声誉以及服务等
- 同其他资产一样，数据资产也是企业价值创造的工具和资本
- 作为现代企业和政府，拥有数据的规模、活性，以及收集、运用数据的能力，将决定企业和政府的核心竞争力



9.1.2数据权理论

- 数据权的概念发起于英国，主要将其视为信息社会的一项基本公民权利，让政府所拥有的数据集能够被公众申请和使用，并且按照标准公布数据。因此，早期的数据权理念强调的是公民利用信息的权利
- 随着数据的进一步开放，大型网络公司对于历史文献资料的数据化，商业集团对于客户资料的搜集，政府部门对于个人信息的调查与掌握，社会化媒体对于社会交往的渗透与呈现，使国家和政府加强了对数据主权的关注，并将其纳入到数据主权的范畴
- 数据主权源于信息主权。信息主权是国家主权在信息活动中的体现，国家对于政权管辖地域内任何信息的制造、传播和交易活动，以及相关的组织和制度拥有最高权力



9.1.2数据权理论

- 数据权包括两个方面：数据主权和数据权利
- 数据主权的主体是国家，是一个国家独立自主对本国数据进行管理和利用的权力
- 数据权利的主体是公民，是相对应于公民数据采集义务而形成的对数据利用的权利，这种对数据的利用又是建立在数据主权之下的。只有在数据主权法定框架下，公民才可自由行使数据权利。公民的数据权利，是一项新兴的基本人权，它是信息时代的产物，是公民个人的基本权利。公民数据权的保护，不仅具有正当合理性，而且已经成为一种人权保障的世界性趋势



9.1.3 开放政府理论

- 开放政府最早出现在20世纪50年代信息自由立法介绍当中。1957年Park的论文“开放政府原则：依据宪法的知情权”中首次提出开放政府理念，其核心是关于信息自由方面的内容。
- 随着很多国家对信息法案的修订，尤其在2009年奥巴马政府公布了《开放政府指令》后，开放政府的理论又被重新提起。2009年1月21日，在关于政府透明和开放化的备忘录上，奥巴马总统指示美国行政管理预算局局长发布一份《政府开放指令》，开放政府由此提出。
- 自2009年开放政府理念被重新提起后，世界各国都在努力使用信息技术革新政府，并在2011年建立了以美国领导的“开放政府联盟”。



9.2 政府信息公开与政府数据开放的联系与区别

- 政府信息公开与政府数据开放是一对既相互区别又相互联系的概念。
- 政府信息公开主要是为了对公众知情权的满足而出现的，信息公开既可以理解为一项制度，又可以理解为一种行为。作为一项制度，主要是指国家和地方制定并用于规范和调整信息公开活动的法规规定；作为一种行为，主要是指掌握信息的主体，即行政机关、单位向不特定的社会对象发布信息，或者向特定的对象提供所掌握的信息的活动



9.2 政府信息公开与政府数据开放的联系与区别

- 政府数据开放是政府信息公开的嬗变必然，将开放对象延伸至原始数据的粒度。政府数据开放强调的是数据的再利用，公众可以分享数据利用创造的经济和社会价值，并且可以根据对数据的分析判断政府的决策是否合理
- 政府数据开放强调的是数据的再利用，公众可以分享数据利用创造的经济和社会价值，并且可以根据对数据的分析判断政府的决策是否合理。政府信息公开更侧重对与公众相关信息通过报纸、互联网、电视等媒体的发布，更强调程序公开，正义公开仍是难点



9.3 政府数据开放的重要意义

9.3.1 政府开放数据有利于促进开放透明政府的形成

9.3.2 政府开放数据有利于创新创业和经济增长

9.3.3 政府开放数据有利于社会治理创新



9.3.1 政府开放数据有利于促进开放透明政府的形成

- 政府开放数据是更高层次的政府信息公开，而政府信息公开也将推动政府民主法治进程
- 如果说政府信息公开还是处于起步阶段，那么政府开放数据则是更高层次的政务公开
- 数据是政府手中的重要资源，政府开放数据的范围、程度、速度都代表着政府开放的程度



9.3.2 政府开放数据有利于创新创业和经济增长

- 美国是气象灾害频发的国家，为减少气象灾害带来的严重损失，2014年3月，美国白宫宣布：将气象数据发布在Data.gov上，随后，与气象相关的企业服务应运而生，包括各种气象播报、气象顾问、气象保险等，形成了一个新的产业链，创造出了极高的经济价值
- 政府数据的再利用，在欧洲也创造出很高的经济价值。2010年欧盟公布的数据显示，欧洲利用政府公开的数据创造出的价值就达到320亿欧元，同时带来了更多的商业和就业机会



9.3.3 政府开放数据有利于社会治理创新

- 政府数据的开放不仅打破了政府部门对数据的垄断，促进了数据价值的最大发挥，同时也构建起了政府同市场、社会、公众之间互动的平台
- 数据分享和大数据技术应用，不仅可以有效推动政府各部门在公共活动中实现协同治理，提高政府决策的水平，也能够充分调动各方的积极性来完成社会事务，实现社会治理机制的创新，给公众的生活带来便利，比如缓解交通压力、增强食品安全、解决环境污染等



9.4 国外政府开放数据的经验

9.4.1 概述

9.4.2 G8数据开放原则

9.4.3 美国开放数据国家行动计划

9.4.4 英国开放数据国家行动计划

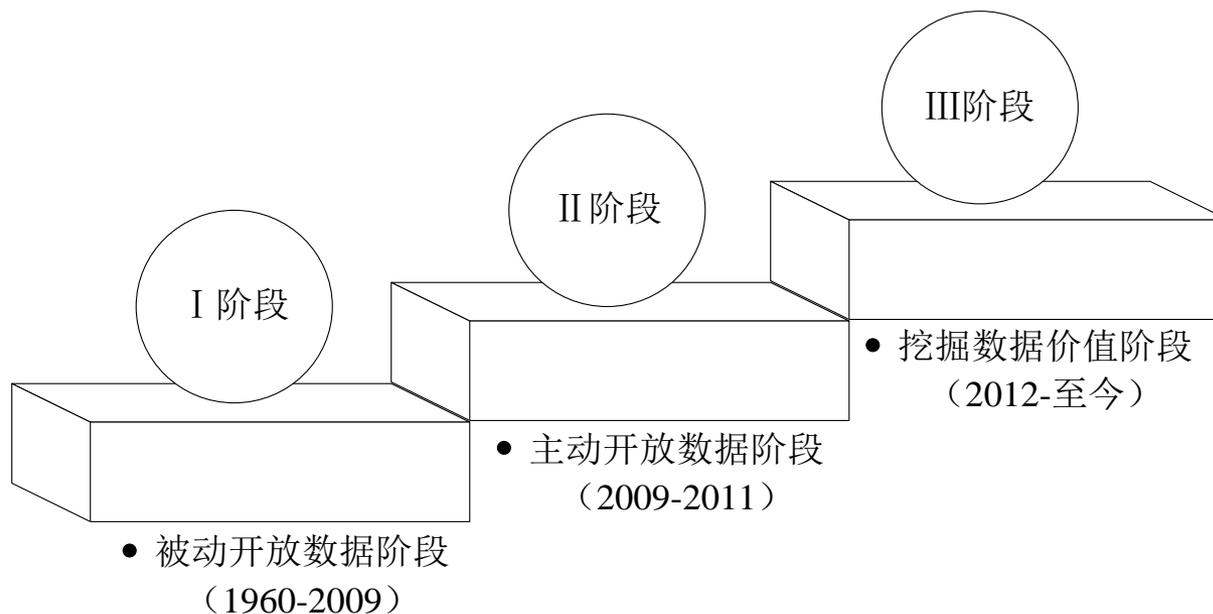
9.4.5 德国政府开放数据行动

9.4.6 日本政府开放数据行动



9.4.1 概述

1. 开放数据发展的三个时期





9.4.1 概述

2. 开放数据的路径探索

国外政府在以下几个方面探索政府开放数据：

- 第一是政府开放数据的范围，即政府开放数据程度上应把握的分寸。
- 第二是政府开放数据的开放许可证，即由谁来负责政府数据的开放。
- 第三是政府开放数据后的数据安全，即如何避免数据关联后带来的新风险。
- 第四是政府开放数据的质量，即如何减少不同部门由于统计口径不同而带来的数据不一致的问题。
- 第五是政府开放数据的形式，即如何让公众读懂这些数据。
- 第六是政府开放数据实施的方法，即通过何种途径、手段来开放数据。
- 第七是政府开放数据的技术，即通过哪种技术来开放数据。



9.4.1 概述

3.开放数据的特点

国外开放数据呈现出以下特点：

- 第一，出台战略和政策，以一定的格式开放政府数据，使开放数据称为默认的规则；
- 第二，纷纷建设开放数据门户，分类开放数据集；
- 第三，注重数据的再利用，采取鼓励措施激发企业和创新者利用数据开发更多应用，促进经济增长和就业；
- 第四，通过示范和典型案例引导数据的开放和开发利用；
- 第五，在开放数据的同时，注重数据的安全、隐私保护和保密，完善相关法律制度。



9.4.2 G8数据开放原则

- 2013年6月17-18日，美、英、法、德、意、加、日、俄召开八国集团首脑峰会，八国领导人在北爱尔兰签署了八国集团开放数据宪章
- 八国集团开放数据宪章明确了**5**大原则、**14**个重点开放领域和三项共同行动计划，其宗旨就是推动政府更好地向公众开放数据，挖掘政府拥有的公共数据的潜力，促进经济增长的创新，提高政府的透明度和责任感。
- “**5**大原则”包括开放数据成为规则、注重数量和质量、让所有人都可用、为改善治理发布数据、为激励创新发布数据。
- **14**个重点开放领域包括：公司、犯罪与司法、地球观测、教育、能源与环境、财政与合同、地理空间、全球发展、政府问责与民主、健康、科学与研究、统计、社会流动性与福利、交通运输与基础设施等，并提供相关的数据集实例
- 三项共同行动计划包括：**G8**国家的行动计划、发布高价值的数据和元数据的映射。



9.4.3 美国开放数据国家行动计划

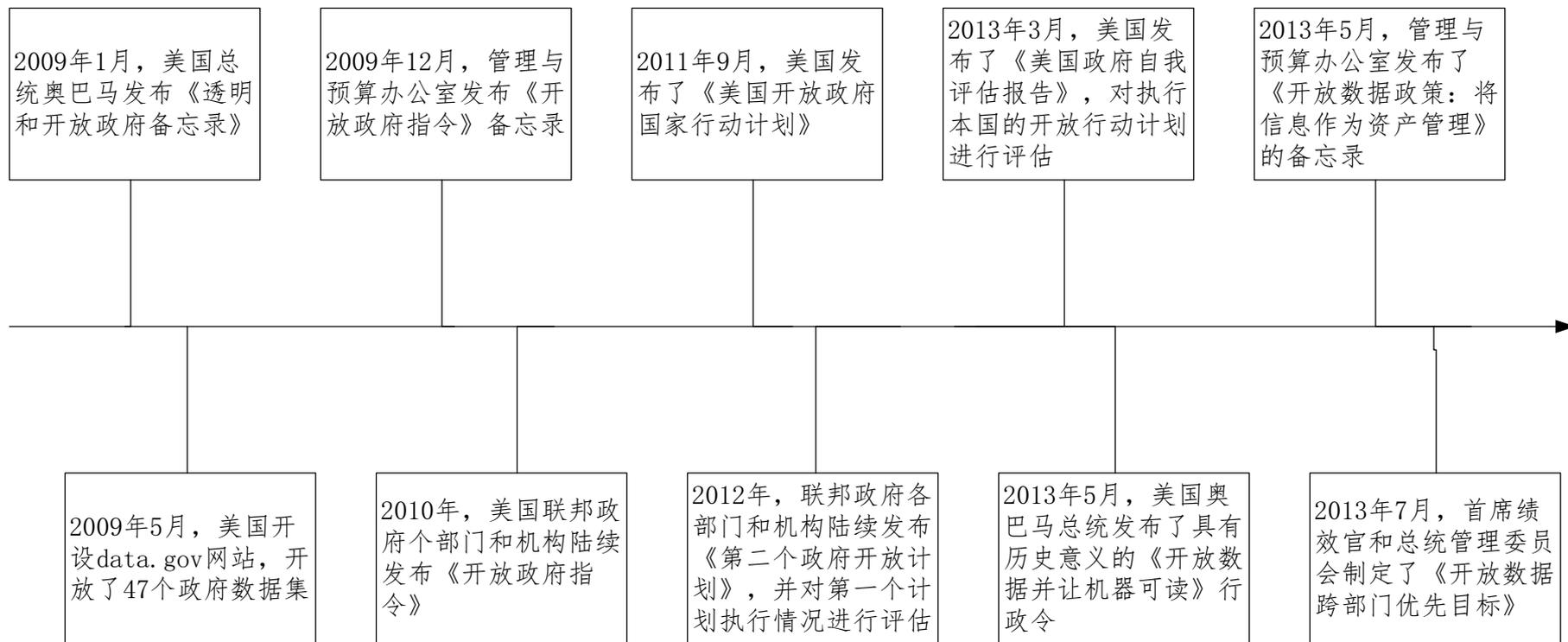


图 美国开放政府数据进展大事记



9.4.4英国开放数据国家行动计划

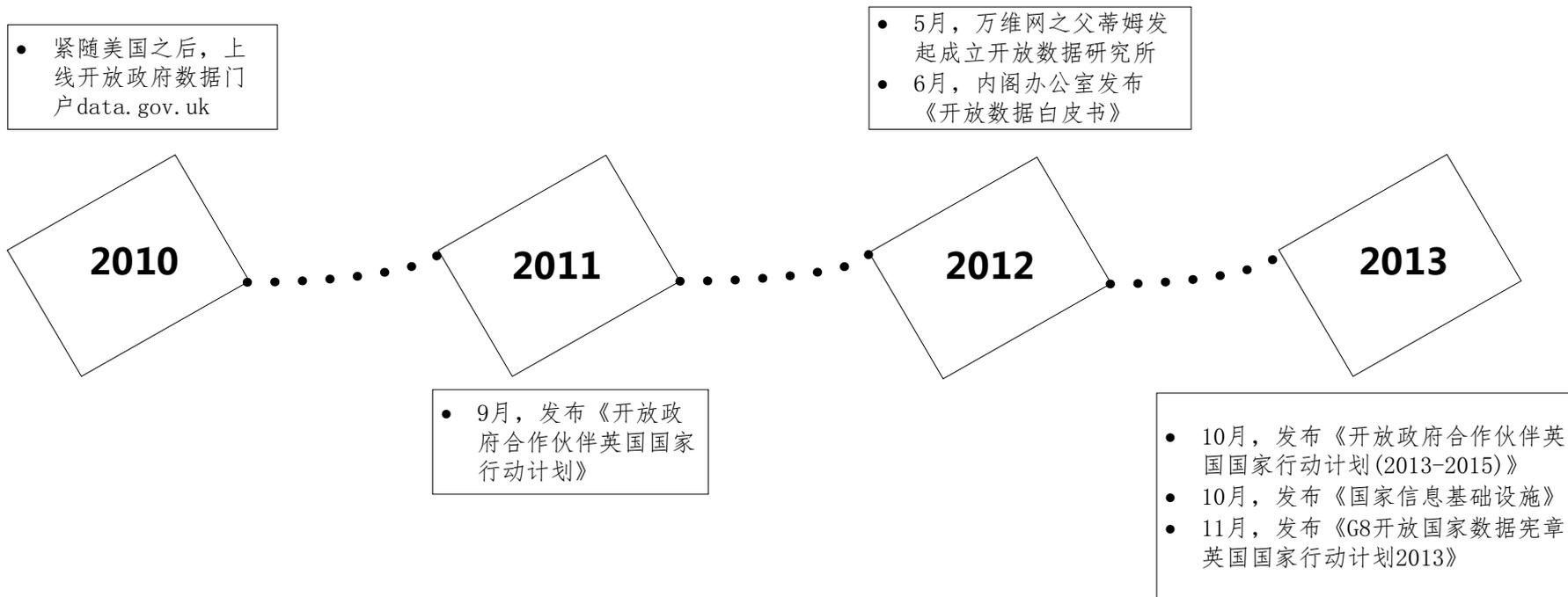


图 英国开放政府数据进展大事记



9.4.5 德国政府开放数据行动

- 2010年9月，德国发布《政府计划：网络型与透明型行政管理》，强调公共行政需要网络连接与透明度。为此，各级政府部门应当在法律范围内界定开放数据，如公共部门的职能、规划目标及其所能搜集或处理的相关数据
- 2010年9月，德国信息技术规划委员会发布“国家电子政务战略”，将“促进开放政府”的项目作为实施电子政务战略计划七个重点项目之一，主要重点在于开放政府数据
- 2014年2月，德国发布《塑造德国的未来》，强调联邦政府与行政部门必须在符合法律的前提下推动开放数据，以标准、机器可读格式，通过开放授权条款来释放数据
- 2013年2月19日，德国推出Beta版政府开放数据门户网站，经过2年实验阶段以后于2015开始正常运行



9.4.6 日本政府开放数据行动

- 日本推动开放政府数据的政策，可追溯至2010年5月信息通信技术战略本部所发布的《新信息通信技术战略》，其旨在促进执政透明化、鼓励公民参与公共政策决策，其核心在于创造以公民为中心的电子政府
- 2012年7月，IT战略本部发布《开放政府数据战略》以“公共数据为公民资产”以及“推动开放政府与促进公共数据利用”为主旨，以机器可读为基本方向，主张公私合作的创新，提供多元价值的公共服务，并在经济、行政效率层面，希望通过开放政府数据来加速整体经济及政府的行政效率
- 2013年6月，安倍内阁正式发布《世界最先进的IT国家宣言》，阐述2013年至2020年期间以发展开放政府数据和大数据为核心的日本新IT国家战略



9.5国内政府开放数据

9.5.1 概述

9.5.2 我国政府数据开放制度体系

9.5.3 当前数据开放存在的主要问题

9.5.4 各地政府数据开放实践



9.5.1 概述

- 通过开放决策过程数据，可以提高政府的公信力；
- 通过开放商业数据，可以驱动经济转型；
- 通过开放各类奖惩信息，可以提高社会信用；
- 通过开放社会统计基础数据，可以促进社会发展科学化。



9.5.2 我国政府数据开放制度体系

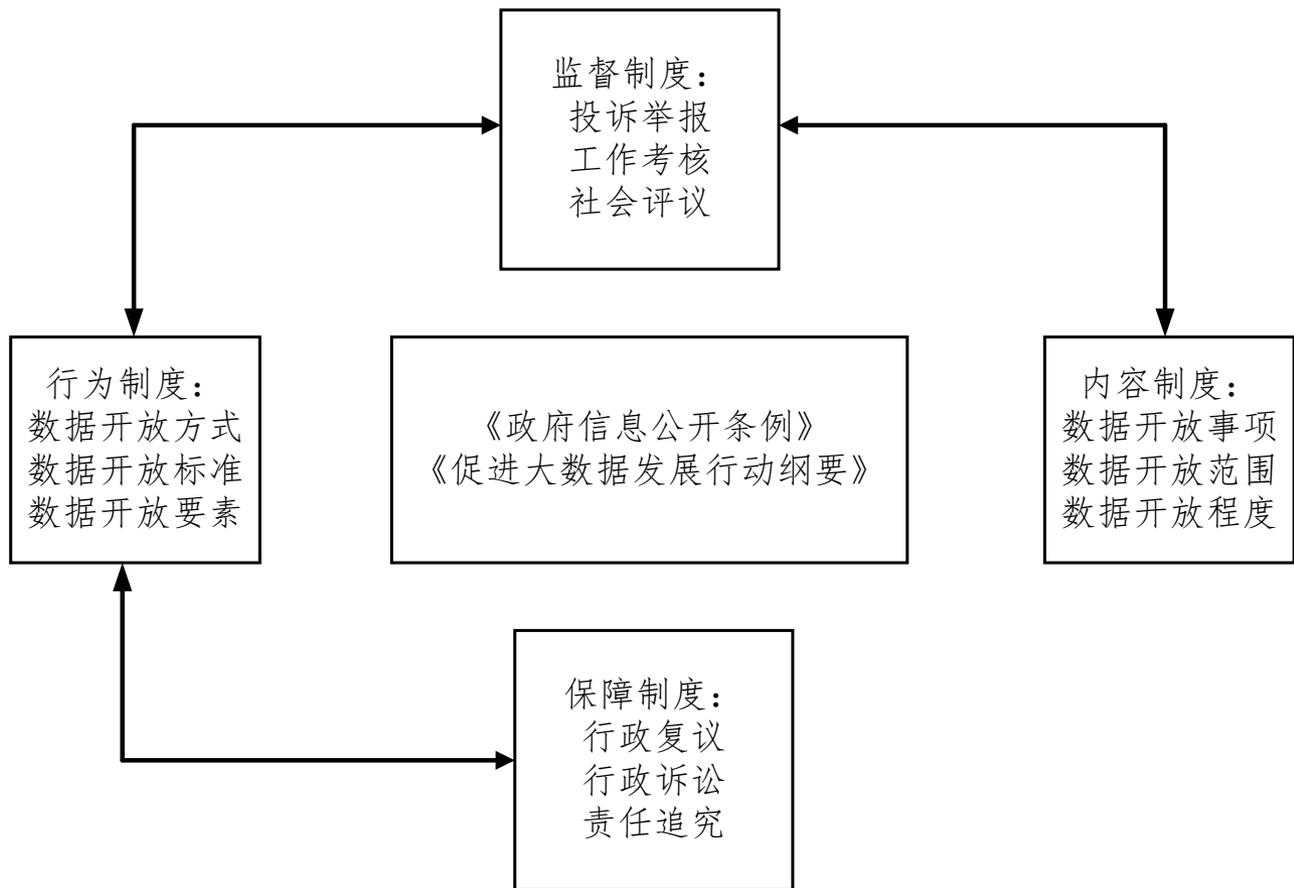
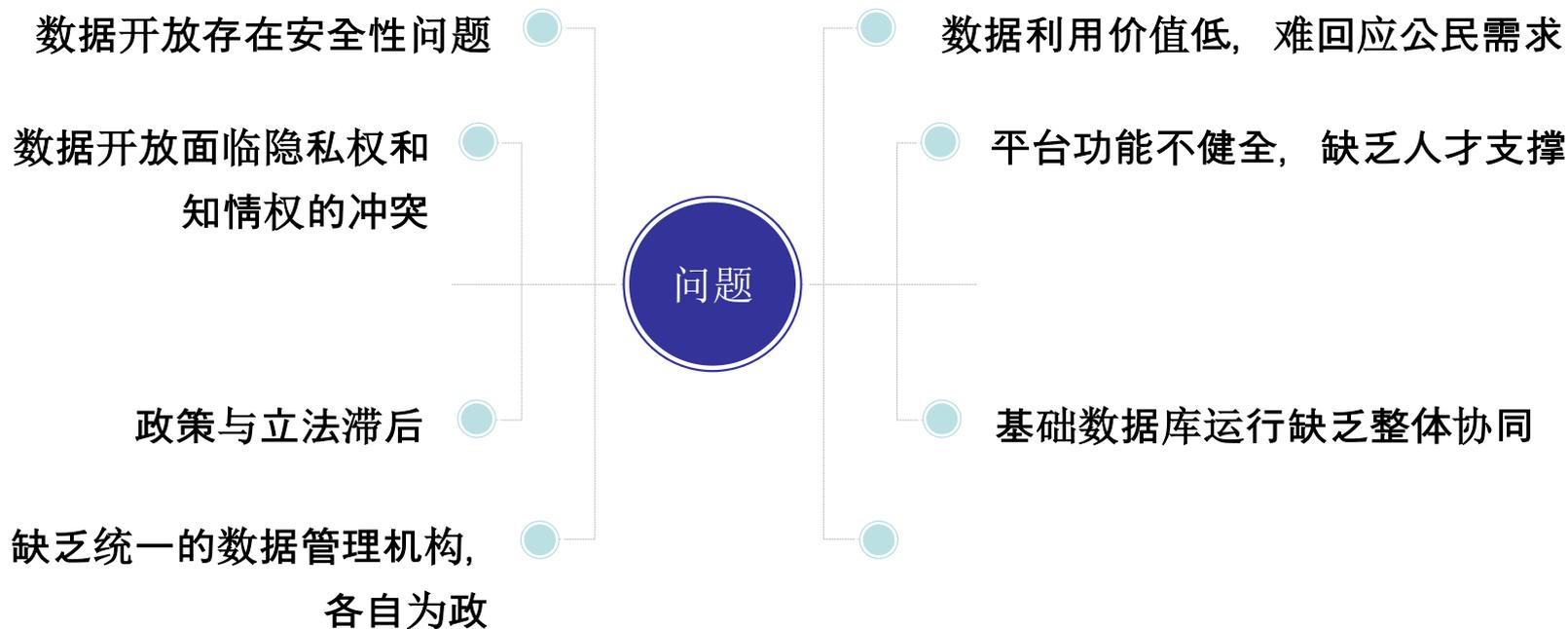


图 我国政府数据开放制度体系



9.5.3当前数据开放存在的主要问题





9.5.4 各级政府数据开放实践

1. 北京政府数据开放实践
2. 上海政府数据开放实践
3. 贵州省政府的数据开放探索和创新
4. 深圳市政府数据开放实践
5. 厦门市政府数据开放实践



1.北京政府数据开放实践

- 北京市政府公开的数据主要发布在各政府部门官网、政府主导建设的对公众开放的数据库以及政府数据开放门户网站（北京市政务数据资源网）上
- 北京市政务数据资源网由北京市经济和信息化委员会牵头建设，北京市各政务部门共同参与，于**2012年10月**开始试运行，提供北京市政务部门可开放的各类数据的下载与服务



2.上海政府数据开放实践

上海在国内率先启动了各政府部门向社会开放政府数据资源。上海市政府数据开放走在全国的前列。

上海市政府开放数据体现以下特点：

- (1) 开放内容扩大
- (2) 查询更为便捷
- (3) 新增互动平台



3. 贵州省政府的数据开放探索和创新

- 2014年，贵州省宣布“云上贵州”系统平台正式上线。作为贵州大数据产业发展的重要基础设施，这一系统平台将为贵州省电子政务云、工业云、智能交通云等“7N”云工程提供云服务，将来贵州还有可能像供电一样向全国提供云计算、云存储和宽带资源等服务。贵州也由此成为国内首个基于云计算建成省级数据统一管理、支持共享平台的省份
- 贵州省政府开放数据的创新之处突出表现在两大方面：
 - (1) 多方参与，依托大数据产业共同打造政府云平台
 - (2) 给政府数据“脱敏”



4. 深圳市政府数据开放实践

•2016年11月，深圳市政府数据开放平台、深圳市统一移动互联网惠民服务平台正式发布。据介绍，深圳市政府数据开放平台为公众打造统一的开放数据访问门户，集中、免费开放政府数据资源。目前，平台已在道路交通、城市建设、公共安全、经济建设等12个领域向社会开放数据，涉及550项数据集、450万条数据，提供数据浏览、查询、下载及API调用等服务





5. 厦门市政府数据开放实践

- 厦门建设了“i厦门”一站式惠民服务平台，各部门只需将已开发的应用软件直接部署在平台上，省钱又省力
- 厦门市建成了城市基础数据库群，并搭建起全市跨部门、跨层级的信息共享、数据同步、业务协同的完整体系，此举不仅促进了政务部门的资源共享，也提供基础数据支撑政府向公众开放信息





9.6政府数据开放的几点启示

- (1) 开放数据是技术、政策、文化三位一体的系统工程
- (2) 发布机器可读的高价值数据和推动数据开发利用是当前数据开放的重点
- (3) 加大数据使用、安全和隐私保护等法律法规和规则的制定，以更好地迎接开放数据带来的挑战
- (4) 推进数据开放过程中应该注重政府和民间的合作



附录A：主讲教师林子雨简介



主讲教师：林子雨

单位：厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

个人网页: <http://dblab.xmu.edu.cn/post/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>



扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），现为厦门大学计算机科学系助理教授（讲师），曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国计算机学会数据库专业委员会委员，中国计算机学会信息系统专业委员会委员。国内高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度和2017年度厦门大学教学类奖教金获得者，荣获2017年福建省精品在线开放课程、2018年厦门大学高等教育成果特等奖、2018年福建省高等教育教学成果二等奖、2018年国家精品在线开放课程。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，主持的教改课题包括1项2016年福建省教改课题和1项2016年教育部产学协作育人项目，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过500万字高价值的研究和教学资料，累计网络访问量超过500万次。打造了中国高校大数据教学知名品牌，编著出版了中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，并成为京东、当当网等网店畅销书籍；建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过100万次。



附录C： 《大数据技术原理与应用》 教材

《大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）》，由厦门大学计算机科学系林子雨博士编著，是国内高校第一本系统介绍大数据知识的专业教材。人民邮电出版社 ISBN:978-7-115-44330-4 定价：49.80元



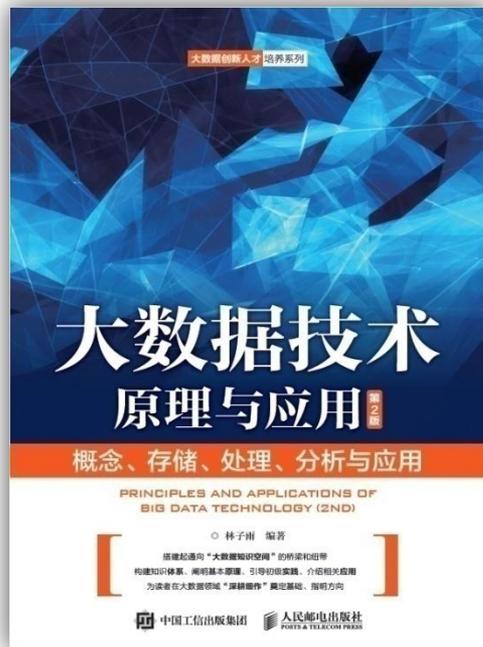
扫一扫访问教材官网

全书共有15章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、Spark、流计算、图计算、数据可视化以及大数据在互联网、生物学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：

<http://dbl原因.xmu.edu.cn/post/bigdata>





附录D：《大数据基础编程、实验和案例教程》

本书是与《大数据技术原理与应用（第2版）》教材配套的唯一指定实验指导书

大数据教材



1+1黄金组合
厦门大学林子雨编著

配套实验指导书



- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，五套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

清华大学出版社 ISBN:978-7-302-47209-4 定价：59元



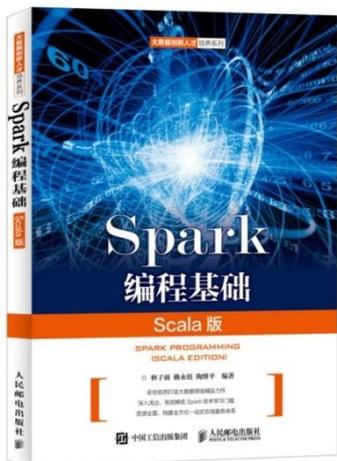
附录E：《Spark编程基础（Scala版）》

《Spark编程基础（Scala版）》

厦门大学 林子雨，赖永炫，陶继平 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-48816-9
教材官网：<http://dmlab.xmu.edu.cn/post/spark/>



本书以Scala作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



附录F：高校大数据课程公共服务平台



高校大数据课程

公 共 服 务 平 台

<http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/>



扫一扫访问平台主页



扫一扫观看3分钟FLASH动画宣传片



附录G：高校大数据实训课程系列案例教材

为了更好地满足高校开设大数据实训课程的教材需求，厦门大学数据库实验室林子雨老师团队联合企业共同开发了《高校大数据实训课程系列案例》，目前已经完成开发的系列案例包括：

《基于协同过滤算法的电影推荐》

《电信用户行为分析》

《实时日志流处理分析》

《微博用户情感分析》

《互联网广告预测分析》

《网站日志处理分析》

系列案例教材将于2019年陆续出版发行，教材相关信息，敬请关注网页后续更新！

<http://dbllab.xmu.edu.cn/post/shixunkecheng/>



扫一扫访问大数据实训课程系列案例教材主页

The background of the slide features a blue gradient with several faint, light-blue silhouettes of people. At the top, there are two groups of people standing and holding hands. On the right side, a person is shown in profile, looking towards the center. In the bottom left corner, two people are shown in profile, one appearing to be speaking or gesturing towards the other. The overall theme is one of community and collaboration.

Thank You!

Department of Computer Science, Xiamen University, 2019