调查研究报告

——构建大数据共享开放社会环境

华东师范大学

计算机科学与软件工程学院

构建大数据共享开放社会环境社会实践小组

2016.9

目录

[1.调查的背景和意义 1](#_Toc460452768)

[2.调查对象的基本情况 3](#_Toc460452769)

[2.1 政府部门 3](#_Toc460452770)

[2.1.1中国气象科学研究院 3](#_Toc460452771)

[2.1.2上海市气象局 3](#_Toc460452772)

[2.2 相关企业 3](#_Toc460452773)

[2.2.1中国电信 4](#_Toc460452774)

[2.2.2 浙江物产 4](#_Toc460452775)

[2.3 重点大学 4](#_Toc460452776)

[2.3.1 华东师范大学 4](#_Toc460452777)

[2.3.2 复旦大学医学院 4](#_Toc460452778)

[2.3.3上海大学 5](#_Toc460452779)

[2.3.4 电子科技大学 5](#_Toc460452780)

[2.4 社会群众 5](#_Toc460452781)

[3.调研的范围和方法 6](#_Toc460452782)

[3.1调研范围 6](#_Toc460452783)

[3.2调研方法 6](#_Toc460452784)

[4.调研结果与分析 7](#_Toc460452785)

[4.1数据开放情况 7](#_Toc460452786)

[4.1.1 数据开放概念 7](#_Toc460452787)

[4.1.2 数据开放实现部署 7](#_Toc460452788)

[4.1.3 数据开放数据层探显 8](#_Toc460452789)

[4.1.4 数据开放环境 9](#_Toc460452790)

[4. 2 在大数据共享机制上的困难之处 10](#_Toc460452791)

[4.2.1 不成熟的体制 10](#_Toc460452792)

[4.2.2 “供”与“需”的脱节 11](#_Toc460452793)

[4.2.3 大数据的管理与决策 11](#_Toc460452794)

[4.2.4 大数据人才缺口 12](#_Toc460452795)

[4. 3 对未来的规划 12](#_Toc460452796)

[4. 3.1 政府整合与开发 12](#_Toc460452797)

[4. 3.2 须防患于未然 13](#_Toc460452798)

[4. 3.3 对大数据人才的培养 13](#_Toc460452799)

[5. 调研结果分析与建议 14](#_Toc460452800)

[服务模式建议：共享信息资源型政府服务模式 15](#_Toc460452801)

[管理机制建议：建立健全各项法规 15](#_Toc460452802)

[数据安全措施建议：构建全面的数据安全保护体系 15](#_Toc460452803)

[总结感悟 16](#_Toc460452804)

[6.参考文献和资料 18](#_Toc460452805)

[7.调查实记 19](#_Toc460452806)

[7.1【高校】的采访实录 19](#_Toc460452807)

[7.1.1华东师范大学计算机与软件学院金澈清教授的采访实录 19](#_Toc460452808)

[7.1.2华东师范大学信息化办公室主任 采访记录 21](#_Toc460452809)

[7.2【企业】的采访实录 23](#_Toc460452810)

[7.2.1车辆维修单位汽车公司访谈整理 23](#_Toc460452811)

[7.2.2医疗单位访谈整理 27](#_Toc460452812)

[7.3【政府】的采访实录 28](#_Toc460452813)

[7.3.1国家气象局预报中心主任访谈整理—李集明 28](#_Toc460452814)

[7.3.2上海市气象局科技服务中心主任——刘欧萱 30](#_Toc460452815)

[7.4【社会】的采访实录 31](#_Toc460452816)

[7.4.1 31](#_Toc460452817)

[7.4.2 32](#_Toc460452818)

[7.4.3 34](#_Toc460452819)

[7.4.4 36](#_Toc460452820)

[7.4.5 37](#_Toc460452821)

[7.4.6 40](#_Toc460452822)

[7.4.7 41](#_Toc460452823)

[8.调查问卷 44](#_Toc460452824)

[8.1高校版访谈问卷 44](#_Toc460452825)

[8.2企业版访谈问卷 45](#_Toc460452826)

[8.3政府版访谈问卷 45](#_Toc460452827)

[8.4社会人员版访谈问卷 46](#_Toc460452828)

[9.前期调研 48](#_Toc460452829)

[10.中期小结 50](#_Toc460452830)

[已完成的工作 50](#_Toc460452831)

[11.后期实现 51](#_Toc460452832)

[12．构想 52](#_Toc460452833)

[12.1建设气象大数据开放平台 52](#_Toc460452834)

[12.2举办数据开放创新型应用大赛 53](#_Toc460452835)

[13.附录 54](#_Toc460452836)

# 1.调查的背景和意义

据贵阳大数据交易所发布的《2016年中国大数据交易产业白皮书》数据显示，2014年中国大数据产业规模大约为1038亿元，而2015年产业整体规模达到1692亿元，增长率达63.01%。白皮书预计：到2020年，我国大数据产业市场规模将达到8228.81亿元。2015年，我国政府大数据应用额依然占比最高，达到11.38%；其次是电信和金融股行业，份额分别为9.35%和8.90%；电商、医疗和能源行业紧随其后。中国在大数据邻域仍处于起步发展阶段，但各地各行业发展大数据积极性较高，市场规模增益明显。

国际市场上，英美等国家早就将“大数据战略”上升为国家战略层次。奥巴马政府早在2012年宣布投资2亿美元拉动大数据相关产业发展。英国在2013年年初也宣布，将注资6亿英镑发展8类高新技术，大数据独揽其中1.89亿。“数据是新的石油,是本世纪最为珍贵的财产。”大数据正在改变各国综合国力，重塑未来国际战略格局。

国家《十三五规划纲要》中提出实施国家大数据战略,把大数据作为基础性战略资源。纲要中首次在国家层面推出了“公共数据资源开放”的概念，将政府数据开放列为了中国大数据发展的10大关键工程。纲要设定了两个关键目标：2018年底前将上线国家政府数据统一开放平台，以及2020年底前“逐步实现信用、交通、医疗、卫生、就业、社保、地理、文化、教育、科技、资源、农业、环境、安监、金融、质量、统计、气象、海洋、企业登记监管等民生保障服务相关领域的政府数据集向社会开放”，至此数据“开放”在中国才加入了主流舆论的队伍。

随着[信息化](http://www.chinairn.com/report/20140329/100113737.html)发展进入[移动](http://www.chinairn.com/report/20140325/112116439.html)[互联网](http://www.chinairn.com/report/20140212/100221433.html)和[大数据](http://www.chinairn.com/report/20140312/141637718.html)时代，数据的价值愈发凸现。2009年，美国率先开通政府数据服务网。至今，全球已有63个国家和地区推进政府数据开放，法国、新加坡等甚至将政府数据开放纳入国家大数据发展战略。而通过开放政府数据促进社会转型，带动大数据产业发展，已经成为各国普遍共识。

其实，在2015年前，中国一些地方政府早已开始了政府数据开放的试验。2014年中旬，上海市政府率先实行政府数据资源向社会开放，出自28个市级政府部门、涵盖11个领域的190项数据内容将成为今年重点开放对象。市经信委此次印发的《2014年度上海市政府数据资源向社会开放工作计划》共涵盖公共安全、公共服务、交通服务、[教育](http://www.chinairn.com/yjbg/moref15f18f73ff1.html)科技、[金融](http://www.chinairn.com/yjbg/moref15f12f49ff1.html)服务、[能源](http://www.chinairn.com/yjbg/moref15f15f58ff1.html)环境、健康[卫生](http://www.chinairn.com/yjbg/moref15f1f5ff1.html)、[文化](http://www.chinairn.com/report/20140327/10561515.html)娱乐等11个领域，开放的市级政府部门数量是原先的3倍。其中，地理位置类的数据资源将全面开放，市场监管类数据也成为开放重点，并大力推进交通数据资源开放。此外，北京、广东佛山、贵州等地也走在全国前列。

那这些地方政府的数据开放试验成效如何呢？2015年中，一项名为“中国开放数据探显镜(opendatalens.cn)”的研究，选取了有代表性的7个地方政府开放数据项目进行评估。截至2015年5月20日，这7个地方共发布了1946个数据集，其平均可机读率为84.1%。但若考察各地数据的授权条款，则发现无一满足开放标准。虽有行动，但却游离于国际通行的“开放标准”之外。开放，至少需要满足两个方面。其一，“机器可读”，即为机器所理解而能自动解析的格式。

其二，在法律层面，即数据的授权条款应能保障使用者在信息获取、使用、再传播中不受任何限制。对于中国现有的数据开放工作而言，长期缺乏对“开放”的诠释和定位。

“数据是新的石油,是本世纪最为珍贵的财产。”但大数据开放本身基于数据的“自由”属性，决定着其不作为“垄断性”资源存在于市场，天生具备公共属性。而各级地方政府并没有很好的全局意识，甚至将其视为抢占大数据产业发展先机的优势条件。此外，政府部门是社会信息的主要控制者，其信息又分别被不同部门和区域控制，而各部门地区的数据标准各异，信息资源难以共享。各级部门掌握的仅仅是完整数据链的一个结点，而断裂的数据链称不上高质量的数据。

最主要的矛盾还存在于，关键数据保密与开放之间的那根警戒线。我们需要相关信息保密，确保我们的个人关键信息不泄露，乃至上升到国家安全的层面。二是，数据开放是整个时代的大趋势。若处理不恰当，极有可能导致，该保密的信息没保密住，该公开的信息也没公开。

本次社会实践正是围绕“大数据的社会环境”，通过问卷、走访调查等多种形式各异的方式，了解社会各领域人士对“大数据开放”的理解程度。使用了针对性的主观题为主的问卷，了解政府部门、企业、高校等专业人员以及部分社会人士对“大数据开放”服务生活职业的切实感受，还有他们对数据开放带来的安全问题等热议话题的所思所想。

我们正处于数字时代，开放数据将会带来巨大的生产力。即使短时间内开放数据看不到商业价值，但随着技术的日益完善，将数据转化为资产，转化为巨大的社会生产力，为居民提供更为便利的社会服务，为企业实现更大的商业价值。

本次社会实践也试图找到各个社会组成之间在大数据开放上面的矛盾点。在政府、企业、普通百姓这三个大类之间找到彼此利益最大化的实现点和矛盾点。通过缜密的调研、分析找出解决方法，推动整个社会的大数据开放脚步。

# 2.调查对象的基本情况

此次实践活动从政府部门、大数据相关企业、重点高校以及社会群众四大方面着手进行调研。

本课题调研以问卷调研与试点调研为主，其中对政府、企业的调研以访谈调研形式开展，涉及中国气象科学研究院，上海市气象局，中国电信，浙江物产等多个政府部门及相关企业。对重点高校的调研采用访谈与问卷相结合的形式开展，涉及华东师范大学、复旦大学医学院、上海大学、电子科技大学等重点高校的学校部门、教师与学生。对于对社会群众的调研主要以问卷调查的形式开展，并辅以部分采访，涉及各年龄段，各行业群体。基本情况简介如下：

## 2.1 政府部门

### 2.1.1中国气象科学研究院

中国气象科学研究院，其前身为1956年8月成立的中央气象科学研究所，1978年更名为中央气象局气象科学研究院，1991年更为现名；是中国大气科学领域学科种类最多、规模最大的科研机构，也是中国大气科学研究，特别是应用基础研究和技术开发的中坚和高层次人才的培养基地。中国气象科学研究院以灾害天气、气候与气候系统、大气成分、雷电防护与大气探测、人工影响天气、生态环境与农业气象、数值模式等为研究主攻方向。研究院设有6个职能处，6个研究所（中心）、1个国家重点开放实验室、2个中国气象局部门级重点开放实验室；有正研级科研人员54名，副研级科研人员99名；共培养博士研究生95名、硕士研究生570名。



### 2.1.2上海市气象局

上海市气象局实行上级气象主管机构与本级人民政府双重领导，以上级气象主管机构领导为主的管理体制，在上级气象主管机构和本级人民政府领导下，根据授权承担上海行政区域内气象工作的政府行政管理职能，依法履行《中华人民共和国气象法》和气象主管机构的各项职责。

## 2.2 相关企业

### u=1467316085,3886388750&fm=206&gp=02.2.1中国电信

中国电信集团公司是我国特大型国有通信企业、上海世博会全球合作伙伴，连续多年入选"世界500强企业"，主要经营固定电话、移动通信、卫星通信、互联网接入及应用等综合信息服务。截至到2011年上半年，拥有固定电话用户1.94亿户，移动电话用户（CDMA）6236万户，宽带用户 6174万户；集团公司总资产6322亿元，人员67万人。2016年5月，国务院首批双创“企业示范基地”。

### 2.2.2 浙江物产

浙江省物产集团公司是1996年由原浙江省物资局成建制转体组建的大型国有流通企业，是浙江省政府授权经营管理国有资产的运营机构，是国家120家大型企业试点企业集团和20家重点培育发展的大型流通企业以及浙江省政府确定的36家重点流通企业之一，正厅级规格，注册资本3.5亿元人民币。2014年10月13日，浙江国资国企改革第一单——浙江物产集团启动整体上市。2014年12月26日，浙江省政府批复了物产集团整体上市总体方案，方案明确物产集团整体上市以物产中大作为主体，吸收合并浙江省物产集团有限公司及发行股份购买资产并募集配套资金。

## 2.3 重点大学

### 2.3.1 华东师范大学

华东师范大学是国家教育部直属全国重点大学，是国家“985工程”重点建设高校。该学校是在大夏大学原址上创办的新中国第一所社会主义师范大学，它是全国首批16所重点院校之一，全国重点大学，是国务院批准首批建立研究生院的33所高等院校之一。1996年通过“211工程”预审并被列入国家重点建设大学行列。2006年教育部和上海市决定重点共建华东师范大学，标志着华东师范大学正式跨入“985”重点高校建设行列，学校在建设世界知名的高水平研究型大学的道路上迈出了坚实步伐。

### 2.3.2 复旦大学医学院

复旦大学上海医学院（原上海医科大学）创建于1927年，创立时名为国立第四中山大学医学院，是中国创办的第一所国立大学医学院，颜福庆出任首任院长。1932年，独立为国立上海医学院，为当时中国唯一的国立医学院。1952年，更名为“上海第一医学院”。1959年，被定为全国16所重点高等学府之一。1985年，更名为“上海医科大学”。2000年，上海医科大学和复旦大学合并办学，组建成为新的复旦大学。2011年，新的上海医学院成立，作为复旦大学党政的派出机构，根据学校的授权，在大医口的人才培养、科学研究、学科建设、发展规划、资源配置、对外交流等方面，行使相对独立的管理权限。

### 2.3.3上海大学

上海大学是上海市属、国家“211工程”重点建设的综合性大学，是国家教育部、国家国防科技工业局与上海市人民政府共建高校，是教育部首批“卓越工程师教育培养计划”、“卓越新闻传播人才教育培养计划”重点建设高校，是“国家建设高水平大学公派研究生项目”、“海外高层次人才引进计划”、“教育部来华留学示范基地”高校。学校在全国率先实行学分制、选课制、三学期制、通识教育等先进的教育制度，并与国家文化部、中国科学院、中国社会科学院、中国艺术研究院、中国文联等开展了全面合作。上海大学设有宝山（校本部）、延长、嘉定三个校区，占地面积约3000亩。现有研究生12051人，全日制本科生22799人，外国留学生4087人。学校的战略目标是建设成为世界一流、特色鲜明的综合性研究型大学。

### 2.3.4 电子科技大学

电子科技大学简称电子科大，坐落于有“天府之国”之称的成都市，由中华人民共和国教育部直属，位列“211工程”、“985工程”，入选国家“2011计划”、“111计划”、“卓越工程师教育培养计划”，两电一邮成员，设有研究生院，是一所以电子信息科学技术为核心的全国重点大学，被誉为“中国电子类院校的排头兵”。

电子科技大学原名成都电讯工程学院，是1956年在周恩来总理的亲自部署下，由交通大学（现上海交通大学、西安交通大学）、南京工学院（现东南大学）、华南工学院（现华南理工大学）的电讯工程有关专业合并创建而成。1960年，被中共中央列为全国重点高等学校；1961年，被中共中央确定为七所国防工业院校之一。1988年，更名为电子科技大学。1997年，被确定为国家首批“211工程”建设的重点大学。2000年，由原信息产业部主管划转为教育部主管。2001年，进入国家“985工程”重点建设大学行列。

## 2.4 社会群众

# 3.调研的范围和方法

## 3.1调研范围

本调查研究报告对政府部门（中国气象科学院、上海市气象局）、大数据相关企业（中国电信、浙江物产）、重点高校（华东师范大学、复旦大学医学院、上海大学、电子科技大学）以及社会群众四大方面展开了调研活动。

在调研期间，我们主要调查分析了我国对数据开放共享的政策支持，当前政府部门开展这项工作的情况，当前面对主要矛盾和改革瓶颈，企业对于数据开放共享的态度，群众对数据开放共享的关注了解等问题。

## 3.2调研方法

本次调研采用问卷调研法和现场访谈（试点）调研法。

在调研期间，我们走访了一些政府部门并对相关负责人进行了访谈，调查了我国在数据开放分享领域中的历程，相关政策办法，当前开展情况，主要面对的问题等问题。同时我们也走访了部分企业和高校老师，就当前数据开放在其各自环境中的影响与意义，数据的来源与使用途径等问题进行了访谈。为了了解社会群众及高校学生对数据开放共享的了解程度与参与情况，我们设计了名为《企业版调查问卷》、《政府版调查问卷》、《社会人员版调查问卷》的调查问卷，针对高校学生与社会中各行业的社会群众进行了调查。本次调查的样本范围广、数量多、内容丰富全面、形式快捷新颖，因此具有较高的科学性和可信度。

# 4.调研结果与分析

## 4.1数据开放情况

### 4.1.1 数据开放概念

当前，人们对“开放数据”尚无统一的定义，不同的组织、机构有不同的理解，如“开放数据中心联盟”(Open Data Center Alliance)视之为公司IT基础设施、云计算的应用模式与解决方案；学术出版和学术资源联盟(SPARC)主张它为科学数据的一种新型学术出版模式、理念；W3C倡导的“开放数据运动”则是采用RDF数据模型，在网络空间中不同类型和来源的数据实体之间建立“RDF 链接”，从而用特定的语义网浏览器或搜索引擎为用户在普通的HTML网页和结构化数据中导航，最终实现人人都可自由地获取到其想要数据的目标。

尽管各个领域因数据应用相异而有不同的主张，但以秉持中立(neutral point of view)著称的维基百科对“开放数据”的定义可看作是早先兴起的其他。

“开放运动”(如Open Access)的扩展：它是一种哲学理念及实践，要求相应的数据可以被任何人自由获取，而没有版权、专利或其他机制的限制。这种定义缺陷在于对“数据”这一关键要素及其应用没有明确的规定，只提供了一个指导性的原则，即数据的无限制、无障碍获取。故可考虑改进如下：开放数据，即对Web中产生或存在于Web中的各种类型的数据，按照用户特定的需求和相应的互联网协议、规则、框架进行开发、加工、存储、组织等管理活动，最终目标是实现局域和全部Web空间的数据开放、互通、共享。

### 4.1.2 数据开放实现部署

目前，国际上知名的开放数据评估报告主要针对国家层面而非地方层面，其评估框架和指标体系也并不完全适用于中国国情。我国的政府数据开放首先在地方层面先行探索，国家平台尚未推出，目前尚缺少对中国地方层面开放数据发展现状与问题的评估报告。此外，国际上的评估报告大多同时对供应端和应用端进行评估，而我国的开放政府数据实践刚刚起步，目前的工作主要集中在政府数据的供应端，只有政府将数据开放出来之后，才谈得上对数据应用端展开评估。

因此，我国目前将借鉴国际评估经验，立足我国发展实际，重点针对数据供应端，聚焦于北京、上海、武汉、无锡、湛江、宁波海曙和佛山南海这七个具有引领性、代表性的地方政府的开放数据实践开展评估。

统计图表如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 样本 | 行政级别 | 所在地区 | 开放数据门户名称 | 门户网址 |
| 北京市 | 直辖市 | 华北地区 | 北京市政务数据资源网 | www.bjdata.gov.cn |
| 上海市 | 直辖市 | 华东地区 | 上海市政府数据服务网 | www.datashanghai.gov.cn |
| 武汉市 | 副省级市 | 华中地区 | 武汉市政府公开数据服务网 | www.wuhandata.gov.cn |
| 无锡市 | 准副省级市 | 华东地区 | 无锡市政府数据服务网 | opendata.wuxi.gov.cn |
| 湛江市 | 地级市 | 华南地区 | 湛江数据服务网 | data.zhanjiang.gov.cn |
| 宁波市海曙区 | 副省级市下辖区 | 华东地区 | 海曙区开放数据门户 | data.haishu.gov.cn |
| 佛山市南海区 | 地级市下辖区 | 华南地区 | 数说南海 | data.nanhai.gov.cn |
| 2016.7 | | | | |

### 4.1.3 数据开放数据层探显

从数据层来看，目前中国地方政府的数据开放量较低，按照开放数据定义要求，数据应当符合易操作的可机读格式（即如xls，而非pdf），但目前可机读率仅达84.1%，其中上海目前以398个可机读数据集位列全国第一。

各地方目前普遍发布的都是静态数据（更新频率为一年或不定期），比例高达86.25%。少部分承诺将更新的数据也未能按承诺进行更新。无锡是各地方中按承诺更新比例最高的（62.5%），各地方平均仅有17.21%的数据按承诺得到了更新。

开放数据应当采用开放的数据格式以确保数据文件无需通过指定的应用程序才能访问。基于此定义，CSV是常见的开放格式而XLS则不是。目前仅北京100%提供了开放格式，南海则有98.15%的数据为开放格式，其他地方均未采用开放格式。而且目前仅南海一个地方在发布数据时提供了多种格式（CSV,XLS,TXT,XML,和JSON），以便于不同需求的数据使用者对数据进行加值利用。

在数据使用权利上，仅无锡和海曙确保数据永久免费，未使用“现阶段”字样。但目前未有一个地方能够明确确保使用者享有免费、自由增值利用，以及自由分享传播的权利。

目前各地方对于开放的数据都能提供基本的元数据信息，包括数据名称、发布单位、数据分类以及数据描述等。但南海和海曙未能明确标记出数据的发布时间，仅海曙和南海明确在数据页面标记数据更新时间，而北京、上海、无锡、湛江都将该信息提供在数据目录中而非数据页面，不便于用户查看，武汉则未提供该项元数据。



### 4.1.4 数据开放环境

2015 年 9 月 5 日，国务院印发《促进大数据发展行动纲要》。《纲要》首次从国家层面认定数据是国家基础性战略资源，将大数据行业定位到国家战略层面，大数据成为推动经济转型发展的新动力，成为重塑国家竞争优势的新机遇，成为提升政府治理能力的新途径。

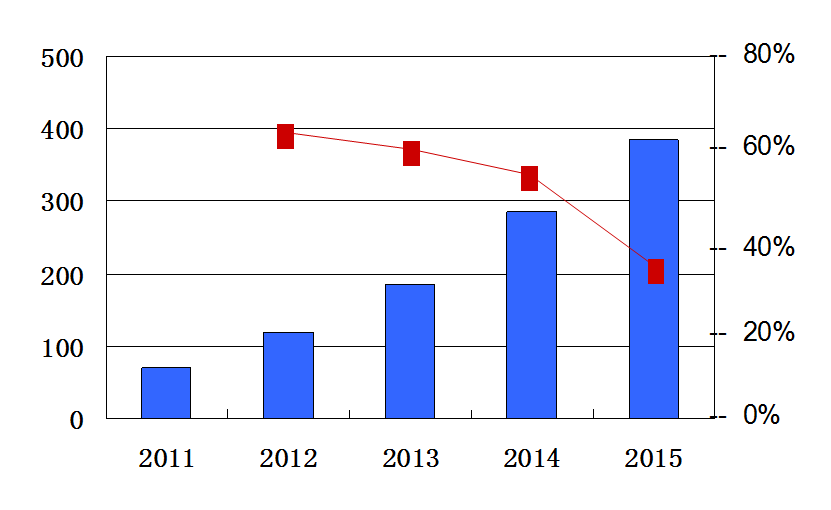
 大数据发展，打破信息孤岛是关键。《纲要》指出，要加强顶层设计和统筹规划，形成公共数据资源合理适度开放共享的法规制度和政策体系。2018 年底前，建成国家政府数据统一开放平台。2020年底前，逐步实现信用、交通、医疗等领域的政府数据集向社会开放。目前，信息孤岛问题依然是阻碍大数据前行的关键要素。目前，60%的主管部门认为数据分布和共享存在难题，这源于不同部门间数据开放标准的不统一，以及在早期建设中各自独立进行和外包导致数据格式标准等的不同。因而建立数据统一平台的前提就是打破信息孤岛，实现数据共享，这对于行业发展至关重要。

大数据行业空间将逐步释放。《纲要》提出，到 2020 年，我国将形成一批具有国际竞争力的大数据处理、分析、可视化软件和硬件支撑平台等产品；并且培育 10 家国际领先的大数据核心龙头企业，500家大数据应用、服务和产品制造企业。相关内容以及时间节点表明大数据产业行业空间将逐步加速释放。

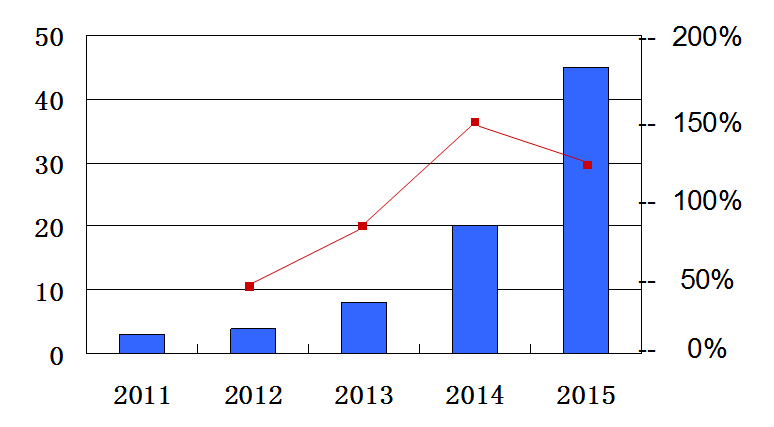
中国与国际的对比：

市场规模/亿美元 同比增长

全球大数据增速



国内大数据增速



市场规模/亿美元 同比增长



## 4. 2 在大数据共享机制上的困难之处

### 4.2.1 不成熟的体制

我们国家的政府在很早就开始了有关数据的工作，具体一点来讲就是四大基础数据库。2002年国务院信息化领导小组发布了《国家信息化领导小组关于我国电子政务建设指导意见》。这个文件将人口、法人单位、自然资源和空间地理、宏观经济基础数据这四个国家基础信息数据库列入国家电子政务建设的重点任务，以能实现信息的汇总便于共享，为决策提供知识上的依据与支持。在具体应用上，各职能部门基本已经成立或正在成立自己的数据的分析处理机构，同时每年面向科研方向开放许多政府数据资源，主要是各高校和科研机构。

我国的官方数据开放已经做了许多工作，每年有大量的官方数据开放给各高校和科研机构，最终目标肯定是进行全面的开放。但是在这个过程中我们遇到了许多的困难，所以现在正在努力的克服困难逐步推进全面开放。开放的形式要视部门和数据的种类不同而采取不同的态度。像实时的天气数据，国家会提供像“风湿温压”这些数据，但其他一些涉及国家经济战略，人口等相关的一些数据开放的形式就是以修饰过的数据进行开放。

推行数据开放共享面对的问题有很多，这也说明了这项工作充满了困难。总的来说，大数据时代开启阶段的瓶颈就是在确保隐私、机密和国家安全的前提下带头开放数据，但如果涉及到国家安全的一些数据，就很难开放给公众，对于一些需要这些数据的科研单位，也会履行很严格的手续才能获取一部分的数据。具体一点说，现在相关法律法规、政策制度和技术标准尚未形成完整成熟的体制，同时长期以来以部门为中心的政务信息化发展模式和资源管控也加大了数据开放的困难。这也就是习总书记所指的改革的“深水区”之一吧，以后还会面对许多的问题。

### 4.2.2 “供”与“需”的脱节

数据应用的前提是数据开放，这已经是共识。有专业人士指出，中国人口居世界首位，但2010年中国新存储的数据为250PB，仅为日本的60%和北美的7%。目前我国一些部门和机构拥有大量数据但宁愿自己不用也不愿提供给有关部门共享，导致信息不完整或重复投资。

例如：工业界和学术界在数据开放方面是是脱节的，工业界不愿意开放数据，而学术界对数据是有很大需求的。工业界的数据很丰富但是欠缺数据分析的能力，他们的确手中有数据，但是不知道怎么办。而学术界呢，非常乐意帮他们解决问题，但是工业界考虑到数据的价值，并不是十分乐意开放他们的数据。而且学术界对数据的应用，更多的是在科研方面而不是商业。

### 4.2.3 大数据的管理与决策

大数据的技术挑战显而易见，但其带来的决策挑战更为艰巨。大数据至关重要的方面，就是它会直接影响组织怎样作决策、谁来作决策。在信息有限、获取成本高昂且没有被数字化的时代，组织内作重大决策的人，都是典型的位高权重的人，要不然就是高价请来的拥有专业技能和显赫履历的外部智囊。但是，在今时今日的商业世界中，高管的决策仍然更多地依赖个人经验和直觉，而不是基于数据。

大数据开发的根本目的是以数据分析为基础，帮助人们做出更明智的决策，优化企业和社会运转。哈佛商业评论说，大数据本质上是“一场管理革命”。大数据时代的决策不能仅凭经验，而真正要“拿数据说话”。因此，大数据能够真正发挥作用，深层次看，还要改善我们的管理模式，需要管理方式和架构的与大数据技术工具相适配。这或许是我们最难迈过的一道坎了。

大数据应用领域仍窄小，应用费用过高，制约大数据应用。国内能利用大数据背后产业价值的行业主要集中在金融、电信、能源、证券、烟草等超大型、垄断型企业，其他行业谈大数据价值为时尚早。随着企业内部的资料量愈来愈大，日后大数据将成为IT支出中的主要因素，特别是数据储存所耗费的成本，很可能造成企业负担，甚至望而却步。因此有远见的CIO必须预先做好准备。

### 4.2.4 大数据人才缺口

如果说，以Hadoop为代表的大数据是一头小象，那么企业必须有能够驯服它的驯兽师。在很多企业热烈拥抱这类大数据技术时，精通大数据技术的相关人才也成为一个大缺口。

虽然，现在“大数据”这个词风靡全球，但是能够真正去实现其价值的人才却不多。绝大多数也只是人云亦云，亦或是略懂皮毛罢了。

大数据建设的每个环节都需要依靠专业人员完成，因此，必须培养和造就一支懂指挥、懂技术、懂管理的大数据建设专业队伍。

可以说，真正启动大数据在企业和社会的全面应用，面临的不仅仅是技术和工具问题，更重要的是要转变经营思维和组织架构，才能真正地挖掘这座大数据“金矿”。

## 4. 3 对未来的规划

### 4. 3.1 政府整合与开发

大数据服务创业公司Connotate对800多名商业和IT主管进行了调查。结果显示，60%受调查者称：“目前就说这些大数据投资项目肯定能够带来良好回报尚为时过早。”之所以如此，是由于当前大数据缺乏必需的开放性：数据掌握在不同的部门和企业手中，而这些部门和企业并不愿意分享数据。大数据是通过研究数据的相关性来发现客观规律，这依赖于数据的真实性和广泛性，数据如何做到共享和开放，这是当前大数据发展的软肋和需要解决的大问题。

例如：2012年美国大选，奥巴马因数据整合而受益。在奥巴马的竞选团队中有一个神秘的数据挖掘团队，他们通过对海量数据进行挖掘帮助奥巴马筹集到10亿美元资金；他们通过数据挖掘使竞选广告投放效率提升了14%;他们通过制作“摇摆州”选民的详细模型，每晚实施6.6万次模拟选举，推算奥巴马在“摇摆州”的胜率，并以此来指导资源分配。奥巴马竞选团队相比罗姆尼竞选团队最有优势的地方：对大数据的整合。奥巴马的数据挖掘团队也意识到这个全世界共同的问题：数据分散在过多的数据库中。因此，在前18个月，奥巴马竞选团队就创建了一个单一的庞大数据系统，可以将来自民意调查者、捐资者、现场工作人员、消费者数据库、社交媒体，以及“摇摆州”主要的民主党投票人的信息整合在一起，不仅能告诉竞选团队如何发现选民并获得他们的注意，还帮助数据处理团队预测哪些类型的人有可能被某种特定的事情所说服。正如竞选总指挥吉姆·梅西纳所说，在整个竞选活动中，没有数据做支撑的假设很少存在。

因此，从社会、国家领域而言，我国亟须在国家层面对大数据给予高度重视，特别需要从政策制定、资源投入、人才培养等方面给予强有力的支持；另一方面，建立良性的大数据生态环境是有效应对大数据挑战、用好大数据的主要出路，需要科技界、工业界以及政府部门在国家政策的引导下共同努力，通过消除壁垒、成立联盟、大数据质量标准、建立专业组织等途径，建立和谐的大数据生态系统。

### 4. 3.2 须防患于未然

大数据时代，传统的随机抽样被“所有数据的汇拢”所取代，人们的思维决断模式，已可直接根据“是什么”来下结论，由于这样的结论剔除了个人情绪、心理动机、抽样精确性等因素的干扰，因此将更精确、更有预见性。不过，由于大数据过于依靠数据的汇集，一旦数据本身有问题，就很可能出现“灾难性大数据”，即因为数据本身的问题，而导致错误的预测和决策。

大数据的理论是“在稻草堆里找一根针”，而如果“所有稻草看上去都挺像那根针”呢？过多但无法辨析真伪和价值的信息和过少的信息一样，对于需要作出瞬间判断、一旦判断出错就很可能造成严重后果的情况而言，同样是一种危害。“大数据”理论是建立在“海量数据都是事实”的基础上，而如果数据提供者造假呢？这在大数据时代变得更有害，因为人们无法控制数据提供者和搜集者本人的偏见。拥有最完善数据库、最先接受“大数据”理念的华尔街投行和欧美大评级机构，却每每在重大问题上判断出错，这本身就揭示了“大数据”的局限性。

不仅如此，大数据时代造就了一个数据库无所不在的世界，数据监管部门面临前所未有的压力和责任：如何避免数据泄露对国家利益、公众利益、个人隐私造成伤害？如何避免信息不对等，对困难群体的利益构成伤害？在有效控制风险之前，也许还是让“大数据”继续待在笼子里更好一些。

大数据的经济价值已经被人们认可，大数据的技术正逐渐成熟，一旦完成数据的整合和监管，大数据爆发的时代即将到来。我们现在要做的，就是选好自己的方向，为迎接大数据的到来，提前做好准备。

### 4. 3.3 对大数据人才的培养

大数据相关人才的欠缺将会成为影响大数据市场发展的一个重要因素。据Gartner预测，到2015年，全球将新增440万个与大数据相关的工作岗位，且会有25%的组织设立首席数据官职位。大数据的相关职位需要的是复合型人才，能够对数学、统计学、数据分析、机器学习和自然语言处理等多方面知识综合掌控。未来，大数据将会出现约100万的人才缺口，在各个行业和领域，大数据中高端人才都会成为最炙手可热的人才，涵盖了大数据的工程师、规划师、分析师、架构师、应用师等多个细分领域和专业。因此需要社会、高校和企业共同努力去培养和挖掘。企业可以与学校联合培养人才，或建立专门的数据科学家团队，或与专业的数据处理公司合作，以解人才之急。

# 5. 调研结果分析与建议

大数据时代

—— 机遇与挑战并存

信息技术与经济社会的交汇融合引发了数据迅猛增长，数据已成为国家基础性战略资源。坚持创新驱动发展，加快大数据部署，深化大数据应用，已成为稳增长、促改革、调结构、惠民生和推动政府治理能力现代化的内在需要和必然选择。

2015年9月5日，国务院印发《促进大数据发展行动纲要》。《纲要》首次从国家层面认定数据是国家基础性战略资源，将大数据行业定位到国家战略层面，大数据成为推动经济转型发展的新动力，成为重塑国家竞争优势的新机遇，成为提升政府治理能力的新途径。

通过本次调查研究后，发现有以下现象值得注意：

1. 政府职能部门有大量的数据，但不知道如何去运用来造福公众。
2. 企业、高校希望得到一手的、实时的数据，可是没有办法获取。
3. 公开数据的内容、程度、范围没有一个确定的标准，无法完全保证数据的安全（尤其是涉及隐私、安全等问题），因而公开的数据使用的实用价值不达预期。
4. 大数据企业之间分工不明确、交流合作不足、协同力度不够。
5. 数据量十分庞大，缺少分析数据的人才。

本着对以上五点的观察和分析，再自身专业相关——计算机科学与软件工程，现提出搭建气象大数据开放平台的建议，旨在搭建一个可以安全提供数据，从而获得有效利用数据去开发的优秀创想的平台，以下是这个平台的初步设想。

气象大数据开放平台

面对对象：全国大学生。

特别适用：有数据分析能力和优秀创想的人。

主要功能：

1. **根据信息整合资源利用平台发布气象、交通等可靠数据。**
2. **使得学生拥有可靠的数据去验证、开拓他们的创想。**
3. **发布国家数据相关政策以及新的有关风向**
4. **提供开放保护式的平台给有同学分享关于大数据的创想、帮助他们组建队伍。**
5. **给大学生创业团队提供参加大赛，去展示自己想法的平台。**

这个平台主要针对在校大学生，意义在于可以安全地提供可靠的、可控的、有效的数据给有创新思维的人，去实现、验证他们原有的想法，从而他们有一个很好的平台将其表现出来，展示自我。同时，也可以使得原本拥有大量数据的政府职能部门、大数据企业，将大量的数据进行有效地利用，从而达到造福社会的最终愿景。

此外，我们还对以下服务模式、数据管理机制、数据安全措施提出建议

### 服务模式建议：共享信息资源型****政府服务模式****

共享信息资源要求改变传统以政府为中心的服务模式，树立以公众需求为中心的新型服务政府。政府职能机构要在网络化的管理平台上进行高效率的协同工作，以跨职能、跨部门的形式组织政府业务流程,例如,并联审批事项、城市突发应急等业务不是某个政府部门能够独立解决的,需要多个职能机构协同工作。为了迅速、有效地开展跨部门的协同业务,对所需基础信息和业务信息要实现信息资源的共享。因此，为适应社会经济发展的需求，必须加强政府各部门之间的协作，实现跨地域、跨部门、跨层次的信息集成和无缝隙信息共享。

### 管理机制建议：****建立健全各项法规****

通过立法促使政府信息公开，采用法律手段抑制各部门对公开政务信息的抵触，避免各级部门利用公共信息资源的谋利行为。在对政府信息资源细化分类的基础上, 界定政府信息资源的产权，以法律为基础形成一套完整的制度, 明确具体政府信息资源的产权分别属于哪些人、哪些组织, 并如何在政府、第三部门和企业之间进行分配和转让。为此，应尽快制定和出台《公共信息资源开发共享管理办法》，以法律形式规定公共信息资源开放共享的内容、程序、标准等，并在此基础上建立公共信息资源共享网站，形成共享服务体系;在数据安全方面，应完善国家数据安全法律法规，同时加强网络安全基础设施建设，建立和完善大数据安全应急机制，从而大幅度降低大数据风险。

### 数据安全措施建议：****构建全面的数据安全保护体系****

一、推进数据安全保护立法进程。加快数据安全立法进程，明确数据保护的对象、范畴和违法责任等，制定关于数据开放共享和跨境流动监管的法律条款。同时，拓宽现有法律的调整范围，将工业互联网、云计算等新技术新应用场景下的数据保护纳入法律调整范畴。

二、出台国家数据安全保护战略。从国家安全、国家战略资源的高度定位数据安全，强化数据战略统筹。制定通信、金融等重点行业的关键数据和用户信息的跨境流动监管政策，推动立法规范我国公民个人信息的境内存储。积极参与国际规则的制定，提升我国在数据保护领域的话语权，为我国开展数据安全保护营造良好的国际环境。

三、加强数据安全保护技术攻关。加强数据保护关键技术手段建设，加快身份管理、APT攻击防御、DDoS攻击溯源等关键技术研发。加快数据安全监管支撑技术研究，提升针对敏感数据泄露、违法跨境数据流动等安全隐患的监测发现与处置能力。

四、健全数据安全标准体系和评估体系。统筹规划数据安全相关标准制定，积极开展通用和专用的数据安全标准研发。强化数据安全相关检测与评估，推动开展数据跨境流动安全评估。

### 总结感悟

数据规模越大，处理的难度也越大，但对其进行挖掘可能得到的价值更大，这就是大数据热的原因。大数据时代，机遇与挑战并存。

中国具备成为数据强国的优势条件，应把握优势，克服挑战，抓住大数据革命带来的“机会窗口”，建设数据强国。从2013年至2020年，全球数据规模将增长10倍，每年产生的数据量由当前的4.4万亿GB，增长至44万亿GB，每两年翻一番。从全球占比来看，中国成为数据强国的潜力极为突出，2010年中国数据占全球比例为10%，2013年占比为13%，2020年占比将达到18%，届时，中国的数据规模将超过美国的数据规模，位居世界第一。中国成为数据大国并不奇怪，因为我们是人口大国、制造业大国、互联网大国、物联网大国，这都是最活跃的数据生产主体，未来几年成为数据大国也是逻辑上必然的结果。

然而，在面对机遇时，我们也应考虑其可能遇到的困难，而未雨绸缪。

大数据的挖掘与利用应当有法可依。去年底全国人大通过的加强网络信息保护的决定是一个好的开始，当前要尽快制定“信息公开法”以适应大数据时代的到来。现在很多机构和企业拥有大量客户信息。应当既鼓励面向群体、服务社会的数据挖掘，又要防止侵犯个体隐私；既提倡数据共享，又要防止数据被滥用。此外，还需要界定数据挖掘、利用的权限和范围。大数据系统本身的安全性也是值得特别关注的，要注意技术安全性和管理制度安全性并重，防止信息被损坏、篡改、泄露或被窃，保护公民和国家的信息安全。

大数据时代呼唤创新型人才。盖特纳咨询公司预测大数据将为全球带来440万个IT新岗位和上千万个非IT岗位。麦肯锡公司预测美国到2018年需要深度数据分析人才44万—49万，缺口14万—19万人；需要既熟悉本单位需求又了解大数据技术与应用的管理者150万，这方面的人才缺口更大。中国是人才大国，但能理解与应用大数据的创新人才更是稀缺资源。

大数据是新一代信息技术的集中反映，是一个应用驱动性很强的服务领域，是具有无穷潜力的新兴产业领域；目前，其标准和产业格局尚未形成，这是我国实现跨越式发展的宝贵机会。我们要从战略上重视大数据的开发利用，将它作为转变经济增长方式的有效抓手，但要注意科学规划，切忌一哄而上。

# 6.参考文献和资料

1. 十三五计划纲要
2. 2016年中国大数据交易产业白皮书
3. 中国数据开放之路
4. 开放数据，大数据时代的必经之路
5. 我国大数据产业遭遇的“三重门”
6. 上海市：实现政府大数据资源开放
7. 中国气象数据网<http://data.cma.cn/>
8. 中华人民共和国国家统计局>>统计数<http://www.stats.gov.cn/tjsj/>
9. 中国国家调查数据库CNSDA <http://www.cnsda.org/index.php>
10. 国土资源---统计指标收录列表<http://tongji.cnki.net/kns55/addvalue/indexlist.aspx?sicode=Z006>
11. 上海・开放数据观察<http://city.opendatachina.com/shanghai/index.html>
12. <http://data.worldbank.org/indicator>
13. 克而瑞信息集团<http://www.cricchina.com/>
14. 广州市气象局<http://www.gz121.gov.cn/>

# 7.调查实记

## 7.1【高校】的采访实录

### 7.1.1华东师范大学计算机与软件学院金澈清教授的采访实录



Q：金老师，您觉得大数据开放的现状是怎么样的？

A：工业界和学术界在数据开放方面是是脱节的，工业界不愿意开放数据，而学术界对数据是有很大需求的。工业界的数据很丰富但是欠缺数据分析的能力，他们的确手中有数据，但是不知道怎么办。而学术界呢，非常乐意帮他们解决问题，但是工业界考虑到数据的价值，并不是十分乐意开放他们的数据。而且学术界对数据的应用，更多的是在科研方面而不是商业。

Q：您觉得事业单位和科研单位之间采用什么样的机制能够让双方互惠共赢？

A：寻找双方相同的诉求点，如果诉求不一致的话，矛盾很难解决。从高校等科研单位的角度来讲，希望通过数据进行纯粹的科学研究来解决一些实际的问题，产生一定的社会影响。从数据拥有者的角度来说，如果能够把自身的问题抛出来，提供自己的数据，由其他感兴趣的科研单位来帮忙解决，那么也能推动数据拥有者自身的工作。

Q：您怎么看待数据开放的安全性问题？

A：数据拥有者对数据开放肯定会有安全方面的顾虑，数据怎么保证不被泄露出去呢？一种是双方之间，比如事业单位和科研单位通过签订具有法律效应的合同，采取保密机制来防止数据的泄漏，另外一种通过数据的加密机制，数据的变换来保证数据即使泄露，也不会产生太大的社会影响。

Q：您怎么看待数据开放的未来走向和发展？

A：数据开放应该是一种趋势、一种潮流，就像软件的开源一样。像上海市的soda大赛，已经开放了非常非常多的数据。国外的话数据开放的程度非常高，国内也正在逐步推进政府数据的开放。

Q：科研过程中哪一方面的数据您认为更有价值一点？

A：首先这和研究方向有关系，什么样的数据才能做出高质量的研究。从这个角度来讲，单一的数据源的话，是很难做出有价值的研究工作的，数据源应该是交叉融合的。有一种说法是数据的多元异构，多元指的是多个数据来源，将数据整合在一起，异构是指不同的数据类型，例如文本、轨迹数据等等。多元异构的数据可挖掘出的信息更多一点。比如交通数据能够得知某个地区拥堵，微博数据能够得知该区域在进行演唱会，这样既能知道一个事件的产生又能知道该事件的产生原因。不同来源的数据有不同的用途。

Q：国家十三五提倡数据开放共享，你认为数据开放的尺度是当数据拥有者有需求，向科研单位定向的开放还是向整个社会免费开放，数据开放的尺度应该是怎么样的？

A：免费开放一般是不可能的，首先要考虑数据拥有者的角色，因为数据拥有者如果是企业的话，会追求数据的利益价值，他们的开放更可能是进行数据处理将中间结果进行开放。如果是事业单位比如说气象局，他们的数据是非利益出发的，但是他们也要考虑其他例如安全性等问题。所以数据的开放程度应该考虑是商业行为还是非商业行为。

Q：您刚才说了企业里面对数据分析的能力是有欠缺的，那么企业对高校的数据方面的人才是不是需求量很大呢？

A：需求量是非常非常大的，从整个IT的发展来看，很多IT公司的核心竞争力就是他们的数据，但是如果公司只是拥有这些数据，不对他们进行分析的话，很难发挥出他们数据的价值。比如说针对一些购买数据，能够分析出购买行为，公司可以根据用户的购买行为来提供对应的服务。而且国内的一些公司，由于中国的人口基数比较大，只要稍微有一点规模，他们的数据量就非常庞大了，在这样一个环境下，数据分析的人才肯定是很欠缺的。

Q：您认为高校在数据人才方面的培养应该是“即插即用”的应用型人才还是更偏向于科研研究呢？

A：从我们研究院的角度来讲，还是很注重学生的实际动手能力的，如果做科研和实际的需求差的非常遥远的话，那相当于只是做一些头脑风暴。所以我们往往还是很重视和企业合作的这样一种模式。也就是说科研应该和需求更紧密的结合在一起。

Q：大数据的人才在您眼里除了比较强的计算能力还需要什么样的素养？

A：我觉得需要的素养还是非常非常多的，数据分析师的能力还是很高的。数据科学我觉得是一门跨学科的东西，里面需要用到统计等数学功底。同时数据分析也应该是严肃的，得出的结论正确性应该有一定的概率能保证的。现在很多人热衷于通过数据分析得出结论，但是没有一定的依据来保证结论的正确性，就像google的流感分析正确性都会出错，说明通过数据分析得出结论是有很高难度的，需要的能力是很高的。如果只是热衷于得出结论而没有比较详细的证明，可能会导致很多的不好的社会影响。

Q：数据密集型科学和大数据您认为两者有什么区别？

A：数据密集型计算是06-08年的词汇，大数据是09年的词汇。我认为两者没有太大的本质区别，如果说有差别的话，大数据是从数据本身的角度出发，数据密集型计算是从解决方案的角度出发。两者我认为是一种前后传承的关系。

Q：我们一直有个疑问就是，大数据的结论得出是有目的性的研究还是在分析数据的时候偶然得出的?

A：这个很难说，看分析者是否有敏锐的观察性。做数据研究有几种途径，第一种是别人已经提出了问题并进行了解决，我通过对数据的分析得出一个更好更高效的解决方式，第二种是领域专家提出问题，我们对数据进行分析来解答这个问题，第三种是观察数据，通过日常的一些观察经验，从数据中得出结论。

### 7.1.2华东师范大学信息化办公室主任 采访记录

1、信息化办公室的主要职能：

1. 校园网的维护
2. 校园卡信息的收集与分析，包括：
   1. 消费信息，通过消费信息的分析从而对其行为活动和生活方式进行分析。例如：某段时间，某学生在校消费明显减少，可以推测其生活可能出现了问题，进而老师可以去关心注意一下该学生的状况。
   2. 图书馆进入次数和时间段
   3. 通过校园网的连接，得到用户所在位置，例如：如果有什么突发意外状况，可以第一时间知道准确位置，并进行救援。
3. 信息系统

2、信息化办公室的历史

华东师范大学信息化办公室2003年开始兴办，主要提供信息服务。现在建立了网络信息中心，主要工作是校园网的建设维护，以及信息共享，通过收集到的数据，分析出结果来服务学生、帮助学生。

例如：学生自助服务。通过日均消费量的统计，如果大幅低于正常水平，则安排教师去关心、了解学生的现状，是否遇到了什么困难，可以如何帮助他。这是人与人之间的关怀帮助，而不是单纯、冷冰冰的数据分析

3、数据分析现阶段遇到了哪些困难和难题？

① 现在支撑的目标是什么？为了解决什么问题？

② 怎么得到支持分析的依据？是什么决策了所提供的服务？

③ 理想中拥有数据所能提供的服务和现实中能够提供的有很大差距，还有许多需要改进。

④ 还处于起步阶段，尚不成熟，所以经验参数不足。

⑤ 数据分析人才缺乏。

人才需要有两方面的能力：

（1）编程实践能力。不仅局限于书本上的知识，还能够熟练应用，用于解决实践的问题。

（2）想象力。拥有丰富的想象力。要善于发现问题，可以从身边的事物上发现突破口。有独立的判断能力。这一点优先级高于低一点。

4、您对“智慧校园”比赛有何建议？

① 缺少一个严谨的概念。何谓“智慧”？

② 这次比赛的主旨“智慧校园，优雅学府”特别符合华东师范大学的学校文化氛围。华东师范大学的学术氛围轻松自由，又有九十余年的历史文化积淀，优雅一词再贴切不过。

③ 如何运用数据去实践，去实现自己别出心裁而有意义的想法。

④ 操作过程中可能遇到困难。

例如：采集数据意味着一些隐私被监控，那么如何去合理地运用这些数据就是一大难题。数据样本集是否愿意数据被用于此也是要打问号的。

⑤ 技术难题。如何去保护数据投资、数据价值，使其可以可持续发展。

5、是否有与其他部门、企业间有合作？

我校是在全国最早提出部门间数据共享的。从2002年开始筹划，于2003年由俞立中副校长主管并开始实施。起先，主要是负责统计数据，从而达到服务学生的目标。后来建立了公共数据库，由管理问题转变为管理系统，更好地为学生提供服务。再到2011年，建立了研究生管理系统。该系统打破了部门间的界限，以老师和学生为中心，在不同的阶段、不同的时期，提供不同的服务与帮助。其因人而异的方式方法十分人性化。

## 7.2【企业】的采访实录

### 7.2.1车辆维修单位汽车公司访谈整理

* **受访人：胡翊鹏职位：浙江物产元通汽车集团有限公司信息部总经理**

Q1：请问您眼中的大数据是怎样的，您是否了解大数据及大数据开放共享？

A：随着信息技术的飞速发展和互联网应用的日益普及，大数据已经无时无刻地围绕着我们，渗透进我们的日常生活。IT领域原先已经有“海量数据”、“大规模数据”等概念，但这些概念只着眼于数据规模本身，未能充分反映数据爆发背景下的数据处理与应用需求，而“大数据”这一新概念不仅指规模庞大的数据对象，也包含对这些数据对象的处理和应用活动，是数据对象、技术与应用三者的统一。只有坚持“对象、技术、应用”三位一体同步发展，才能充分实现大数据的价值。

大数据开放共享作为一个近期比较热门的话题，学界、IT界、业界都有很多观点和讨论，总体来看既有巨大的成长空间，也有亟待夯实基础、规范完善的客观要求。

Q2：请问就您的了解而言，您所在的企业对大数据的利用状况如何？在与业内其他企业之间会存在的商业竞争关系下，您认为可以接受怎样的数据开放程度，如会选择公开哪方面的数据及以什么形式公开？

A：中国2200万的年汽车产销量，相关消费和技术数据汇总起来构成了庞大的海量信息；如何挖掘、利用好这些数据，已经成为众多汽车厂家和经销商关注焦点。

从车联网概念的蓬勃兴起以及汽车网络营销得到普遍认可，汽车流通服务业已经迈入大数据时代，成为不可抗拒的行业趋势。汽车消费链长点多的特点，决定了其是典型的长尾消费类型，过去由于运营理念及技术的限制，客户消费需求是由众多业态分别满足的，客户服务的总体成本很高，而利用大数据技术，打通各个企业的信息瓶颈和业务对接流程，使得形成统一的全天候客户营销服务机制成为可能。

企业要充分利用大数据，数据的统一性、完整性、准确性至关重要，需要适时成立跨业务部门的大数据中心，统一规划业务、财务、客户等数据架构、规范来源、制定维护标准、逐步针对各类业务建立数据模型；大数据要产生效益，需要持续的投入和相应的专业人才，并经过一定时期的数据积累，才能作为业务创新的数据引擎。

从中国的现实情况来看，数据公开分享的基础工作由企业角度来展开，可能是事倍而功半的，而从政府先行将大量数据向社会公开着手，可以比较好地形成数据共享的氛围和事实标准，有利于引导社会和企业向数据共享方向发展，同时在政策法规等方面加强制度建设，逐步引导企业走上共享共荣的发展道路。

企业大数据公开需要大量的投入，也希望政府在这方面给予相应的激励政策，鼓励企业利用自建或者第三方的大数据平台，向社会提供广泛服务。

Q3：大数据并不在“大”，而在于“有用”。价值含量、挖掘成本比数量更为重要。请问您认为该如何在海量数据中选择有用的信息，怎么判断数据是否有价值？

A:数据收集只是大数据应用的第一步，如何分析并应用才是最重要的。

对于很多企业来说，完全意义上的大数据分析，投入过大，其ROI可能是偏低的；将企业资源集中在重点关注的数据区块，即所谓的“中数据”，可能是ROI更为合理的选择。

以客户大数据应用为例，一般需要分为以下几步将多渠道、标准不一、维度不一的客户数据进行整合，建立客户大数据库，并针对性地开展应用。

(一) 整合分散在不同系统、不同渠道的数据，包括客户数据、业务数据、互联网数据、政府信息数据、征信数据等等；

(二) 通过数据清洗，识别出客户的详细信息并通过客户数据分析，进行全网客户识别，产生更为全方位的客户视图；

(三) 将每个客户的特点、爱好、生活习惯，进行细致区分，并以标签化进行用户定义；并根据客户的标签进行分组。

(四) 对客户完成精准细分，针对目标客户开展精准营销和个性化服务。

Q4：请问您认为对于企业而言，大数据开放的利弊分别有哪些？

A：对于企业而言，大数据的开放不一定是必然的选择，每个企业都会从自己的战略和利益出发，决定自己对待大数据开放的态度。

通过数据共享的实现，无疑会实现数据边际价值的进一步提升，但是企业在这方面既面临企业商业机密、客户隐私和数据共享之间客观存在的矛盾，也存在目前国家在数据交换共享、个人隐私方面政策法规不完善、不明晰的现实困难。

企业越大、数据规模越大，这方面的试错成本就会越高，因此在大数据开放方面，初期企业可能会持比较谨慎的态度。

Q5：在大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点的趋势下，请问您所在的企业有制定哪些具体的措施来对大数据进行应用？这些应用给您所在的行业带来了哪些影响？

A：汽车营销企业通过对多渠道的汽车大数据进行融合及挖掘，能够在以下几方面深化应用，实现企业运营管理和客户服务质量的提升：

(一) 客户大数据应用

1. 通过互联网和移动营销等方式，分析客户浏览数据，提升新车潜客挖掘效率，提高成交率；

2. 通过分析客户消费数据，细分客户，开展个性化营销和服务,进一步挖掘客户价值；

3. 通过对客户地理位置数据的分析，在集团内实现服务优化，提高服务响应水平，降低客户流失率；

4. 通过和异业的数据交换分享，实现业务合作与拓展。

(二) 车辆大数据应用

1. 通过对车辆维修数据的分析，提高二手车定价的合理性及业务透明度，避免客户纠纷；

2. 通过对车辆维修及零部件进销存数据分析，进一步改善库存管理；

3. 通过对事故车的维修数据分析，加快事故车报价的速度和准确性，提升事故车业务水平和能力；

4. 技术上及资金上允许的话，运用车联网技术，对车辆OBD数据的实时在线获取，将是未来分析客户行为模式、车辆技术状况等重要的基础数据，并可为UBI（基于驾车习惯的保险）等新式车险业务提供数据支持。

(三) 运营大数据应用

1. 通过对众多成员公司业务财务数据分析，为投资决策、收购并购、预算决算等提供数据支撑；并为不良企业经营提供预警；

2. 通过对4S店员工业务及绩效数据的分析，为复制优秀，分享成功提供数据依据。

目前，汽车营销行业对大数据的收集、分析和利用仍处于探索阶段，如果大数据得以合理应用，必将对行业全产业链产生积极、深远的影响。

Q6：从大数据隐私的视角来说，请问您如何看待企业可能面临的数据攻击，您将如何确保自身以及客户数据的安全？

A：信息采集、收集和整理过程对个人隐私保护，是一个大数据应用无法回避的重要问题。

从技术角度来看，大数据依托的基础技术是非关系型数据库，其并没有严格的访问控制机制及完善隐私管理工具，大数据的来源和承载方式多种多样（包括智能手机、Pad、车联网、各类传感器等），数据分散于各个角落，使得企业将很难定位这些数据和保护所有机密信息；由于非关系型数据库往往允许不断对数据记录添加属性，其安全性及权限控制就变得非常重要，对数据库管理也提出了新的要求；用户数据的收集、存储、管理与使用等目前均缺乏法律规范，更缺乏监管，主要依靠企业的自律，用户往往无法确定自己隐私信息的用途。

要保护数据安全，首先政府一定要出台相关法律有效保护个人隐私权和企业数据财产权，对恶意的黑客行为和隐私泄露行为予以严惩；其次企业也要在数据安全方面多方面加强投入，提升安全性能；企业需要围绕大数据突出的安全和隐私问题，构建数据全生命周期的安全管理体系，结合大数据处理体系的特点，尤其关注分布式环境下的并行计算隔离、分布式集群的数据访问控制以及对敏感、重要数据的分级管控、加密处理和审计追踪等安全保障措施，并加强对安全管理员的专业培训及建立风险自适应的访问控制。

Q7：“阿里巴巴”创始人之一，马云在演讲中曾提到，“未来的时代将不是IT时代，而是DT（Data Technology）的时代”。请问您认为大数据未来的发展中心落在哪，及其发展前景如何？

A：大数据未来的发展中心会在宏观和微观两个方面深化：

宏观方面，大数据将对整个社会的政治经济运行模式产生深远的影响，传统的产业及生产流通组织形式、社会管理运行模式，都将会在大数据技术的支持下，实现新的优化组合；微观方面，每个社会个体的需求，将会被进一步细分满足，并尽可能实现路径最优，突出的例子就是“共享经济”。

要达到上述转变，不仅需要技术的进步，也需要社会、政府、各阶层、各利益集团的共同理解和博弈，从而实现社会运行的整体高效化和低成本化。

Q8：请问您认为现下社会各行各业可以如何合作进行大数据开放共享？企业扮演着怎样的角色，可以发挥怎样的作用？

A：不可否认，目前中国大数据应用已经形成了互联网企业一马当先的局面，但从行业专业分工的角度，互联网企业不可能包揽所有行业，大量的企业会通过自建或第三方平台，为企业内部及社会提供大数据应用服务；同时，政府作为大数据的主要拥有者之一，建立并推广大数据应用也责无旁贷。

从降低大数据应用的总体成本角度考量，互联网企业专注于大数据的存储及计算基础资源建设，其他行业、企业专注于大数据的应用探索挖掘，可能是一条投入小、回报高的可行之路；当然，这其中，也要注意防止互联网企业利用自身的优势地位，形成数据垄断，那将对整个社会的大数据应用发展造成巨大的损害和不利影响。

### 7.2.2医疗单位访谈整理

* **受访人：兰女士职位：衢州市某医院医生**

Q1：请问您眼中的大数据是怎样的，您是否了解大数据及大数据开放共享？

A1：听说过。在报纸上看到李克强总理的报告，再就是平时上网新闻上说的国家搞这个项目。

Q2：这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

A2：作为医疗从业者，我们医院药房进药需要知道最缺哪种药，什么药要进货多少，这些都要根据前几年的数据得出。

Q3：大数据并不在“大”，而在于“有用”。价值含量、挖掘成本比数量更为重要。请问您认为该如何在海量数据中选择有用的信息，怎么判断数据是否有价值？

A3：看它能产生多大的社会效益，经济效益，以及对今后工作有什么帮助吧。

Q4：请问您认为对于医疗单位而言，大数据开放的利弊分别有哪些？

A4：好处肯定有的，比如为患者提供更合理的医疗，但是要注意保护患者的隐私。

Q5：在大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点的趋势下，请问您所在的企业有制定哪些具体的措施来对大数据进行应用？这些应用给您所在的行业带来了哪些影响？

A5：大数据肯定是未来的发展方向，医院里的数据都是有保存的，如果需要可以方便使用。

Q6：从大数据隐私的视角来说，请问您如何看待医院可能面临的数据攻击，您将如何确保自身以及客户数据的安全？

A6：我们医院都是医护人员，未来肯定要和互联网单位合作。

Q7：“阿里巴巴”创始人之一，马云在演讲中曾提到，“未来的时代将不是IT时代，而是DT（Data Technology）的时代”。请问您认为大数据未来的发展中心落在哪，及其发展前景如何？

A7：这个我也不清楚，应该是一线城市吧。我觉得大数据包含的信息量太多了，企业事业单位可以利用它创造价值，将来发展空间很大。

## 7.3【政府】的采访实录

### 7.3.1国家气象局预报中心主任访谈整理—李集明

注：李集明，中国气象科学研究院副院长，中国气象局人工影响天气中心主任，中国气象学会气象灾害与服务分委员会委员，中国气候委员会气候资料与监测分委员会委员，科学数据共享工程工作组成员。

Q1、目前我国政府的数据收集处理存储形式是怎样的，是以部门为单位进行的吗？部门内部中央与地方在储存收集数据中，是各地方收集中央储存处理应用吗？

A：这个问题取决于相应政府部门的组织结构，通俗一点讲就是要分“条管”还是“块管”。所谓“条管”，也就是垂直管理部门，部门的组织结构是自上而下，逐级管派的，这样的部门它的数据储存处理工作就是在中央部门进行，而“块管”是指地方政府分级管理，这样的部门的一部分数据是在各个地方进行处理储存的，当然也会由上级进行一定的管理。

Q2、目前我国政府对于官方收集的各类数据的应用情况是怎样的？是否已经形成或正在形成确定的应用研究模式？

A：我们国家的政府在很早就开始了有关数据的工作，具体一点来讲就是四大基础数据库。2002年国务院信息化领导小组发布了《国家信息化领导小组关于我国电子政务建设指导意见》。这个文件将人口、法人单位、自然资源和空间地理、宏观经济基础数据这四个国家基础信息数据库列入国家电子政务建设的重点任务，以能实现信息的汇总便于共享，为决策提供知识上的依据与支持。在具体应用上，各职能部门基本已经成立或正在成立自己的数据的分析处理机构，同时每年面向科研方向开放许多政府数据资源，主要是各高校和科研机构。

Q3、我国的官方数据开放将会以什么样的形式开展？是逐步推进全面开放还是逐步推进的部分开放？开放的形式是以开放原始数据为主还是以处理修饰过的数据为主

A：我国的官方数据开放已经做了许多工作，每年有大量的官方数据开放给各高校和科研机构，最终目标肯定是进行全面的开放。但是在这个过程中我们遇到了许多的困难，所以现在正在努力地克服困难逐步推进全面开放。开放的形式要视部门和数据的种类不同而采取不同的态度。像实时的天气数据，国家会提供像“风湿温压”这些数据，但其他一些涉及国家经济战略，人口等相关的一些数据开放的形式就是以修饰过的数据进行开放，如果涉及到国家安全的一些数据，就很难开放给公众，对于一些需要这些数据的科研单位，也会履行很严格的手续才能获取一部分的数据。  
  
Q4、当前我国推动数据开放共享面对的主要问题是什么，有哪些需要进一步突破的瓶颈？

A：推行数据开放共享面对的问题有很多，这也说明了这项工作充满了困难。总的来说，大数据时代开启阶段的瓶颈就是在确保隐私、机密和国家安全的前提下带头开放数据，降低公众获取和利用政府数据资源的难度和成本。具体一点说，现在相关法律法规、政策制度和技术标准尚未形成完整成熟的体制，同时长期以来以部门为中心的政务信息化发展模式和资源管控也加大了数据开放的困难。这也就是习总书记所指的改革的“深水区”之一吧，以后还会面对许多的问题。

Q5：当前我国政府各部门间的数据开放情况是怎样的？是否已经推出统一的数据共享开放平台，如果还没有，那么官方有意向推出这样的简政型平台吗

A：在各部门间数据开放中，政府已经做了一些工作，现在也已经有许多部门间开放共享合作的项目在进行，具体的平台可以参见我之前说的四大基础数据库，虽然尚未最终成型，但主要的目的就是提供数据共享服务的。

Q6、对于企业或学校，国家在选择进行数据开放会选择什么样的方式？是以开放平台直接提供数据，还是通过各相关部门或所对应的政府进行洽谈以合同等形式进行开放

A：对于高校的数据开放合作一直在进行过程中，未来会进一步加强与高校的合作，简化流程。开放方式目前主要是以高校参与政府课题，政府与高校签订数据开放协议等形式进行的。同时政府的一些面向全社会公开的数据，也可以由高校或者企业进行使用。

Q7、政府有对企业数据的需求吗？对于与非国企进行的数据开放合作，政府将以什么方式进行合作？进行购买，数据合作共享还是其他方式？

A：当然会有需求，像一些有关人们生活的数据，如网购的相关数据，车辆的相关数据，保险的相关数据都是政府十分需要的数据。和企业之间，主要是以合作的方式进行数据开放合作。像我们熟悉的阿里，腾讯等等，都与政府有着合作关系，方式有很多种，主要是以数据共享为主。

Q8、我校与上海市教育局将合作举办一场面向上海市大学生的大数据解决方案的比赛，目前已在筹备过程中，请您在办赛方式，作品方向，成果后续或是其他方面提出您的宝贵意见。

A：我认为在比赛的主要评判方式上，要将技术水平和创新的思维共同作为评判标准，鼓励参赛者大胆创新，减少比赛的边边框框，才能对应大数据研究的基本特征。另外，希望主办方提供充分的数据，以提供给参赛者更大的发展空间。

### 7.3.2上海市气象局科技服务中心主任——刘欧萱

Q：刘主任，您认为大数据现在开放遇到的瓶颈或者说困难是什么？

A：一方面，某些数据是某个单位所特有的，如果（数据）开放会损害他们自身的利益，另外一方面，什么是安全、有效的（数据）开放？有些数据会涉及国家安全，需要进行风险评估、审核，通过了才能进行针对性的有方向的开放。但是总体来说，我认为（数据）开放是利大于弊的。（数据）开放是为了更好的服务，更好的进步。



## 7.4【社会】的采访实录

### 7.4.1

* **受访人：何运超 职业：中国建设银行郴州分部网络工程师**

Q1：在日常生活中，您是否对大数据战略、大数据开放共享方面有一定的了解。您主要通过什么途径获取有关大数据战略的信息？

A1：听说过大数据，但具体不了解。主要通过新闻媒体报道等。

Q2：在政府相关会议上，也提到了“一些引导社会发展的数据，不涉及国家机密的，都应该像公众开放。”您认为政府向公众开放大数据有哪些利与弊。是否赞同开放呢？

A2：如果向普通民众开放，可能用处不大。普通大众如果不涉及相关行业其实对大数据缺乏了解，政府是否开放数据，他们也不会直接利用。但如果是，相关行业人员，类似于搞IT搞金融的，应该十分便利。在不处理好数据保密安全问题之前，不赞同开放。

Q3：这是一个大数据的时代，请谈一谈您所在的行业对大数据应用以及这些应用对您所在行业的影响。

A3：银行在利率，是否批贷款等等，都有大数据的辐射影响。还有个人信用评级也涉及到海量数据。但作为普通员工具体还是不太了解。

Q4：如果您手上掌握着您所在行业的很多重要数据，涉及您的一些利益，但是高校教育、公众生活等也对这些数据有一定需求，您会采取怎样的方式来处理这些大数据。

A4：我肯定不会向公众开放我手上的数据。开源的信息的安全问题令人堪忧，我很担心会被某些不怀好意的人利用。我宁愿由我自己去处理。

### 7.4.2

* **样本一：受访人：黄和 职业：老师**
* **样本二：受访人：深圳某大数据公司一职员**
* **样本三：受访人：刘柏众 职业：华东师范大学软件工程专业研究生**

Q1：您在日常生活中是否有受益于大数据开放之处，请谈一谈。

A1.1：便捷地获取信息各平台针对性的推送可以迅速找到感兴趣的东西

A1.2：信息检索、智能交通、公共安全

A1.3：平时生活中信息资讯主要来自网络，比如亲戚家小孩高考志愿填报、考研院校专业选择、求职面试、各种人事变动公示等，可以从网络上获取详细全面的历史信息，很有参考价值，算是数据开放的一部分吧。

Q2：这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

A2.1：微课程慕课学籍管理教务管理系统家校通等。在提升学校管理质量和教学质量上具有独特的优势。

A2.2：公司属于大数据行业范畴。公司的发展得益于大数据技术的提升，公司在与其他客户的各种也依仗行业的发展。

A2.3：我所在行业属于计算机的数据库领域。大数据源于互联网，并进一步对整个社会的发展产生了深远的影响，所在的行业恰好也算是大数据和互联网的一部分。大数据时代的到来，首当其冲的是数据的存储和访问，大数据对数据库行业的影响是颠覆性的，传统的数据库无法满足海量数据存储的要求，迫使数据库技术的大变革，促进了新型分布式数据库从理论研究到商用化快速发展。当然，数据库本身就属于大数据应用的一部分，这种影响是必然且相辅相承的。

Q3：在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为政府对公众开放大数据有哪些利与弊。

A3.1：人们通过数据进行分析做出相对科学的决策从而创造价值。开放大数据有利于社会经济发展。弊端主要是对个人信息安全的威胁，包括隐私、财产安全。

A3.2：政府大数据信息共享是构建智慧城市和信息社会的有力保障，也需要政府进行信息共享，同时不能忽略潜在的网络信息安全问题，这也是机遇和挑战。

A3.3：其利是显而易见的，保障了公众对非机密数据的知情权，促进了社会的公平公正，可有效阻止谣言的传播扩散，有得于政府科学民主执政等等；其弊也不容忽视，不法分子可通过对多个来源的非机密数据交叉分析，可推导出一些隐藏的涉及国家机密的数据，可能严重损害国家和人民的利益，另一方面如果数据公开方式途径不当，数据可能被曲解分析，甚至篡改，引起社会恐慌和动荡。

Q4：您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

A4.1：赞同除了前面提到的涉及信息安全的数据外，政府应该向公众开放大数据。比如医疗，交通等和民生息息相关的数据。

A4.2：赞同，相关单位掌握着大量的数据，能够共享本身就有利于行业发展，其实网络信息安全的漏洞也和信息不充分离不开关系.

A4.3：赞同。应该向公共开放政府预算、公职人员财产等细节，方便公众监督，电信、移动、银行等企业人事招聘细节，保证社会公平公正。

Q5：一直以来，社会各界对大数据是否开放，如何开放就争议不断，您对于政府和企业对于大数据的开放有哪些建议。

A5.1：1积极推进数据对外开放，充分发掘和利用数据资源；2在开放数据的同时要加强相关立法；3设置用户获取数据的权限以保护个人隐私等。

A5.2：对于不同的使用人群设置不同的权限.

A5.3：首先，开放是肯定的。其次，如何开放，1）开放时机：应在数据产生后主机主动开放，而不是公众通过申请才能开放，2）开放途径：主要通过网络，但不局限于网络，也可辅以电视广播报纸等，3）教育：从学校开始培养公众对政府企业公开数据的关注意识。

### 7.4.3

* **样本一：受访人：戴明辉 职业：安徽省合肥市肥东县第一中学教师**
* **样本二：受访人：戴鑫睿 职业：电子科技大学学生**
* **样本三：受访人：王晗朝 职业：上海大学学生**

Q1：您之前是否听说过大数据，是否对大数据战略、大数据开放共享方面有一定的了解。如果是，您主要是通过什么样的途径获取有关大数据战略的信息。

A1.1：听说过，在平常的工作中，在区域联考的分析中我们也用过大数据分析手段。通过新闻，网络获取这方面信息比较多。

A1.2:有了解，我是信息安全专业的学生，对这些方面肯定是有了解的。平常课上，以及项目中都会遇到这方面的问题。

A1.3: 听过，主要通过新闻，网络获取。

Q2：您在日常生活中是否有受益于大数据开放之处，请谈一谈。

A2.1：肯定是有的，作为老师，在每一次大型考试中，对于整个皖江片的学生成绩能有一个综合的分析，这要归功于大数据。

A2.2：有啊，比如说淘宝，他会对像我这样的人群进行分析，最后得出我可能需要买什么，或者什么样的东西最适合我，省了很多时间啊，当然钱花的也很多了。

A2.3: 我对这方面接触的不多，说的话还是不太想的起来。理财方面,被推荐的理财方案应该说也是大数据分析出的产物。

Q3:这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

A3.1：就像我之前所说的，他给我们教育行业带来的变革是很大的。学生的成长可以通过数据提现，学校之间的差距也是数据的形式分析出来，这样明显对于学校的进步也有益多了。

A3.2: 如果说信息安全这个行业的话，大数据首先就是对网络安全以及信息安全的重大考验，这么多的数据，保密工作就要更加完善，因为有些数据是绝对不能被暴露的，这对我们行业也是很大的挑战。

A3.3：学生行业的话，说接触很深也没有，但是似乎我们生活中每个方面都有接触。这几年变化还是很大的，很多这方面的报告，对于我们的选择都有益处。

Q4：在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为政府对公众开放大数据有哪些利与弊。

A4.1：首先公众肯定能获取更多的信息了，这是一定的。但是这么多数据透露出来，能否正当使用就是一个大问题了。如果有人想要通过这些数据分析得出一些关键的信息，对于整个社会都会不利。

A4.2: 开放数据的话，对于普通大众来说其实没什么用处，但是对于对这些数据有需求的人来说肯定是一件喜事，他们借此就可以分析数据得出自己所需，但是信息安全还是一个严峻的问题。

A4.3：利在可以有更多的人接触这些数据，那么就会有更多的人能通过这些数据得出更多的信息。弊在这些信息如果到了不法分子手上就成了隐形的炸弹。

Q5:您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

A5.1:我赞同，我觉得银行开房的数据更有用吧，对于人们理财会有帮助。

A5.2：关于个人隐私的一些数据还是要做好保密工作，不过一些宏观的数据也是可以开放的。比如本季度银行业那些方面的增长下降之类的。

A5.3：赞同，涉及民生的方面的大数据对于公众来说更有用，专业的数据可能用处就不是那么多。

Q6：一直以来，社会各界对大数据是否开放，如何开放就争议不断，您对于政府和企业对于大数据的开放有哪些建议。

A6.1：开放应当有限度，不是所有的数据都要开放，而且应当渐渐地开放我觉得才是最好的。

A6.2: 我只能说，一切都从人民利益出发吧。

A6.3：开放是大势所趋吧，我建议还是渐渐开放为好，不应该一次性开放，看看效果再说。

Q7：如果您手上掌握着您所在行业的很多重要数据，涉及您的一些利益，但是高校教育、公众生活等也对这些数据有一定需求，您会采取怎样的方式来处理这些大数据。

A7.1：对于行业的利益必须予以保障，所以我只会开放一部分数据，涉及利益的部分还是要仔细考虑的。

A7.2: 我会把这些数据中关于我行业的信息选择性地开放，如果本行业的利益得不到保障，那样的大数据开放没有意义。

A7.3: 选择不涉及利益的部分开放，我相信这也是开放数据的那些机构所要面临的。

### 7.4.4

* **受访人：兰女士 职业：衢州市某医院医生**

Q1：您之前是否听说过大数据，是否对大数据战略、大数据开放共享方面有一定的了解。如果是，您主要是通过什么样的途径获取有关大数据战略的信息。

A1：听说过。在报纸上看到李克强总理的报告，再就是平时上网新闻上说的国家搞这个项目。

Q2：您在日常生活中是否有受益于大数据开放之处，请谈一谈。

A2：我觉得天气大数据对农业的影响是非常有利的，老百姓的吃饭问题能够得到保证。

Q3：这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

A3：作为医疗从业者，我们医院药房进药需要知道最缺哪种药，什么药要进货多少，这些都要根据前几年的数据得出。

Q4：在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为政府对公众开放大数据有哪些利与弊。

A4：规范利用大数据肯定是对社会有帮助的，老百姓的衣食住行能得到提高，但是大数据肯定不能随意开放，一个破坏社会稳定的人得到这些数据，后果是非常可怕的。

Q5：您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

A5：赞同合理地开放。日常生活离不开银行，他们应该开放理财产品的近年收益数据，让公众更好地选择套餐，理财方式。

Q6：一直以来，社会各界对大数据是否开放，如何开放就争议不断，您对于政府和企业对于大数据的开放有哪些建议。

A6：首先要合理，审核你的开放对象，不能随随便便交给一个完全不了解底细的人。

Q7：如果您手上掌握着您所在行业的很多重要数据，涉及您的一些利益，但是高校教育、公众生活等也对这些数据有一定需求，您会采取怎样的方式来处理这些大数据。

A7：作为基层医院医生我们的数据非常少，但我觉得大数据这是未来的一个趋势吧，将来肯定会出台相关的法律法规，如果要开放就必须按照规定来做。

### 7.4.5

* **样本一：受访人：袁同学 职业：浙江理工大学学生**
* **样本二：受访人：倪女士 职业：中华人民共和国舟山海关公务员**
* **样本三：受访人：陈同学 职业：温州医科大学学生**
* **样本四：受访人：文先生 职业：宁波大学教授**
* **样本五：受访人：陈女士 职业：审计局公务员**

Q1：您之前是否听说过大数据，是否对大数据战略、大数据开放共享方面有一定的了解。如果是，您主要是通过什么样的途径获取有关大数据战略的信息。

A1.1：听说过，有一定的了解，主要通过媒体以及身边人的讲话。

A1.2：听过。有一点了解。通过工作原因了解大数据。

A1.3：听说。有一点了解。报纸书籍。

A1.4：听说过，但不详细。主要通过新闻报道，专家评论中获得。

A1.5：听说过，工作中有用到

Q2：您在日常生活中是否有受益于大数据开放之处，请谈一谈。

A2.1：有，比如日常查询考试成绩，可以帮别人一起查。

A2.2：有，如利用搜索引擎检索现有技术查阅资料等。

A2.3：通过大数据，得以得到相对客观和真实的结论，也可更进一步加深对社会现状和科学理论的理解

A2.4：尚未直接受益，如果全面开放实施应该会有。

A2.5：由于计算机专业技术的问题，受益不明显

Q3：这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

A3.1：身为学生，我接触的基本都是学院内的数据互通，这方便了我们查询成绩课程以后别的资料，非常迅捷。

A3.2：对统计行业来说是大势所趋。

A3.3：作为一个学生，在学术方面常常需要运用大数据来得到一个相对接近真实理论的结论，或者通过大量的实验来验证假设。

A3.4：作为高校教师，大数据的开放与实施对教学与研究将会提供真实的数理依据。

A3.5：大数据将整个审计资源统一起来，大大节约了审计成本

Q4：在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为政府对公众开放大数据有哪些利与弊。

A4.1：利的话在于政府办公透明公开，我们可以随时了解国家的走向以及发展，弊的话在于会有不法分子利用这些东西，泄露国家机密。

A4.2：有助于国民了解国家发展的方方面面，但应注意数据泄露等。

A4.3：利：让民众更加了解政府的决策，甚至可以参与决策，从而对国家和政府更有信心。

弊：可能会针对信息源的真实性、客观性等方面引发争议，也可能产生谣言流传等一系列问题。

A4.4：能让更多的社会研究者，经济研究者，获得真实数据而进行深入研究。

A4.5：更多的公众共享数据资源有利有弊。有利于更多的公众共享数据资源，但也给部分不法分子更多的利用机会。

Q5：您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

A5.1：赞同。公众最关心的应该是自己有关的利益，比如有关家庭一些支出，政府最近的政策制定和地方法律法规。

A5.2：否。电信、移动、银行应该为客户信息保密。

A5.3：赞同。比如服务业应该公开顾客的需求的大数据，以及经调查后得到的双方得益较大的项目等，以达到双方互利共赢的目的。

A5.4：应该向公众开放，公众有知情权，同时能监督政府与企业的工作。

A5.5：赞同，方便公众更直接、更便捷地享用公共资源。

Q6：一直以来，社会各界对大数据是否开放，如何开放就争议不断，您对于政府和企业对于大数据的开放有哪些建议。

A6.1：我觉得应当适当开放，既能够给公众提供一些有用的相关信息，保护了公众的知情权，同时对于一些不适合公开的文件，进行保留。

A6.2：可以开放适当信息，但应注重保密。

A6.3：信息源头要公开透明，为了相对客观公正，要全面地展开调查，做到实时监控进程并向民众传达。

A6.4：除关系到国家机密，企业竞争机密，均应开放。

A6.5：要做好配套工作，防止数据泄密给更多的公众带来负面影响。

Q7：如果您手上掌握着您所在行业的很多重要数据，涉及您的一些利益，但是高校教育、公众生活等也对这些数据有一定需求，您会采取怎样的方式来处理这些大数据。

A7.1：我会一定的开放，但是过于涉及隐私的数据选择保留，公众有知情权，但关于行业利益很重要，要保留。

A7.2：保密。为了自己的利益同时也为了涉及者的利益。

A7.3：我会权衡，既让行业和自身的权益得以保障，也最大限度地让公众受益于大数据。

A7.4：首先我没有。如有，会权衡利益损失与社会研究及公众生活需求的关系而提供数据。

A7.5：在法规和政策许可范围之内，可以提供。

### 7.4.6

* **受访人: 李叶英 职业: 昆明某公司会计**

Q1：您之前是否听说过大数据，是否对大数据战略、大数据开放共享方面有一定的了解。如果是，您主要是通过什么样的途径获取有关大数据战略的信息。

A：听过但是没什么了解

Q2：您在日常生活中是否有受益于大数据开放之处，请谈一谈。

A2：新闻当中会有各种各样的数据展示，天气预报

Q3：这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

A3：本身自己是会计，经常要和数据打交道，所以应该是有利于我们对数据进行记录处理，方便我们对账本的处理

Q4：在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为政府对公众开放大数据有哪些利与弊。

A4：老百姓可以有更多机会接触这些数据，但是一些数据不涉及国家私机密如果开放也会带来困扰，带来更多麻烦

Q5:您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

A5：是，有利于公众对企业的选择和投资调整

Q6：一直以来，社会各界对大数据是否开放，如何开放就争议不断，您对于政府和企业对于大数据的开放有哪些建议。

A6：使用大数据应该小心谨慎，防止不法分子利用

Q7：如果您手上掌握着您所在行业的很多重要数据，涉及您的一些利益，但是高校教育、公众生活等也对这些数据有一定需求，您会采取怎样的方式来处理这些大数据。

A7：在要求范围之内尽量规避不利行为

### 7.4.7

* **样本一：受访人：蒲同学 职业：复旦大学医学院临床医学本科生**
* **样本二：受访人：张先生 职位：上海某工程公司自动化工程师**
* **样本三：受访人：黄女士 职业：上海某银行金融行业工作人员**

Q1：您之前是否听说过大数据，是否对大数据战略、大数据开放共享方面有一定的了解。如果是，您主要是通过什么样的途径获取有关大数据战略的信息。

A1.1: 是，IT行业人物自传，生物信息学，阿里大数据

A1.2：.对大数据概念有一定了解，信息主要来源于网络。

A1.3：有所耳闻。对大数据战略、大数据开发共享有一定了解。通过报纸和网上相关文章对其进行的了解。

Q2：您在日常生活中是否有受益于大数据开放之处，请谈一谈。

A2.1：淘宝“为你推荐”

A2.2：日常生活中，资料查询，电子商务，出行导航等等，已离不开基于海量数据的信息网络提供的服务。

A2.3：有。例如现在出行用的百度地图、高德地图，还有购物时京东上的比价功能，这都得益于大数据开放。

Q3：这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

A3.1：应用:生物基因组和表型组计划

影响：使数据获得和处理更加方便，方便合作，但数据来源有可能不可靠

A3.2：工业设备制造行业无疑受益于大数据的应用，对工业设备制造的研发，设计，制造及质量控制等等都无疑大有益处。

A3.3：我从事的是金融行业，现在很多银行已经开始尝试用大数据驱动业务运营，建立了相应的社交网络信息库，大数据已经渗透到客户花画像、运营优化、精准营销、风险管控四个方面。

Q4：在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为政府对公众开放大数据有哪些利与弊。

A4.1：利：让公众更了解政府工作，也监督政府 弊 ：容易被公众缺乏或过分解读

A4.2：对有关于民生，有利于人们日常生活的数据信息能够更多的开放。

A4.3：首先，大数据便利了我们的生活，让服务更加个人化、私有化。其次，大数据的应用利于维护社会治安、促进和平与稳定。另外，在医疗方面，个人化的医疗服务能有效提高患者满意度。

但是，大数据的共享也有很多的弊端。大数据的计算可以将生活的许多风险都计算在内，孩子的学校表现、股票的走向都会被计算，在一切都被计算好的情况下，就缺失了创意、灵感和惊喜。另外，大数据的开放为商家追踪用户个人信息提供了捷径，但是像淘宝的“猜你喜欢”功能等其他利用用户个人信息的行为逐渐为人们所反感。

Q5：您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

A5.1：是，理财产品销量和收益、每月话费量一类帮助用户做购买决策的数据

A5.2：赞同部分开放数据。如政府关于地方经济发展的具体数据，人文数据，教育数据，财政收支等数据等等。

A5.3：不赞同。大数据的开放要依据情况而定。电信、移动、银行等企业数据都有一定的隐私性，开放大数据可能会侵犯到用户的隐私。

Q6：一直以来，社会各界对大数据是否开放，如何开放就争议不断，您对于政府和企业对于大数据的开放有哪些建议。

A6.1：开放数据的同时可以加以简单的解析；从部分数据开始，渐渐开放更多

A6.2：经济发展，人文教育等大数据的开放，以一定公开平台开放数据，可以让普通民众更了解政府及国家的发展方向及经济目标等。

A6.3：首先，要做好设计工作。建立相关机构进行大数据行业的顶层设计，制定相应的标准。其次，大数据的开放需要政府的政策保障，政府在开放与发展大数据的同时也要做好隐私保护。最后，在进行大数据开放的相关应用时，政府和企业要恪守行规，注意保护用户的隐私。

Q7：如果您手上掌握着您所在行业的很多重要数据，涉及您的一些利益，但是高校教育、公众生活等也对这些数据有一定需求，您会采取怎样的方式来处理这些大数据。

A7.1：建立收费平台，以不高的价格卖出

A7.2：不涉及重要商业秘密或行业规则等，可以提供或部分提供。

A7.3：首先，我会查找法律以及公司相关规定，核查开放该数据是否违规。其次，要对数据进行浏览评估，若开放该数据会对公司利益、行业利益有所损失，或者对用户的权益有所侵害，那么我不会开放该数据。最后，如果以上两条的情况均不存在，那么我会对数据进行分类，将对该行业有用的数据进行公开。

# 8.调查问卷

时间：2016年7月

我们针对不同领域的工作人员设计了不同版本的问卷，包括针对高校、企业、社会人员及政府的，然后将这些问卷分发给多个行业的工作者，例如高校的教授，多个不同企业，气象局的政府人员，以及从事各行各业的工作人员，使我们的调查研究更具普遍性。并且这些问卷由在不同地域的多个同学发出，可保证所得的结果不受地域局限性的影响，得到了最终的访谈结果。

## 8.1高校版访谈问卷

国家《十三五规划纲要》中明确提出实施国家大数据战略,全面实施促进大数据发展行动,加快推动数据资源共享开放和开发应用,助力产业转型升级和社会治理创新。本次社会实践的主题为大数据开放的社会环境调查。我们希望通过前期调查研究、资料整理、技术实现，了解当前大数据共享社会构建的推进情况和困难，并对其提出相应的意见和建议。请您就当下大数据共享开放的情况，回答下列问题。

Q1：这是一个大数据的时代，您作为教育实践者。请您谈一谈大数据在高等学校教育管理和教学中的所起到的作用。高等院校的大数据开发建设工作重心应该是什么？大数据开发管理对教育理念、办学理念带来哪些变化？

Q2：您认为符合社会市场要求的大数据人才需要满足哪些基本素养？如何针对您所提的，对学生进行针对性教育与培养？

Q3:目前国内高等院校大数据教育开发与人才培养方面的发展现状如何？与国外，有何差距？

Q4:您认为现下社会各行各业，大数据开放共享的情况如何？以及未来的发展方向。

Q5:请问您分析一下大数据的开放共享程度加深的利与弊。

Q6:在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为作为高等院校该怎样利用这些开放数据？

Q7:高等院校之于国家社会，在大数据共享开放中，是一个怎样的角色，又起着怎么样的作用。

**分析**：高校作为教育工作者，在现大数据的时代不仅应自身跟上潮流，还应能够带动学生对大数据的了解与应用，应主动开发培养人才，提升基本素养。我们着眼于教育实践者的看法与对现状的了解，询问他们将如何利用开放的数据，将如何培养人才，及高校在大数据共享开放中将发挥怎样的作用。

## 8.2企业版访谈问卷

国家《十三五规划纲要》中明确提出实施国家大数据战略,全面实施促进大数据发展行动,加快推动数据资源共享开放和开发应用,助力产业转型升级和社会治理创新。本次社会实践的主题为大数据开放的社会环境调查。我们希望通过前期调查研究、资料整理、技术实现，了解当前大数据共享社会构建的推进情况和困难，并对其提出相应的意见和建议。请您就当下大数据共享开放的情况，回答下列问题。

Q1：请问您眼中的大数据是怎样的，您是否了解大数据及大数据开放共享？

Q2：请问就您的了解而言，您所在的企业对大数据的利用状况如何？在与业内其他企业之间会存在的商业竞争关系下，您认为可以接受怎样的数据开放程度，如会选择公开哪方面的数据及以什么形式公开？

Q3：大数据并不在“大”，而在于“有用”。价值含量、挖掘成本比数量更为重要。请问您认为该如何在海量数据中选择有用的信息，怎么判断数据是否有价值？

Q4：请问您认为对于企业而言，大数据开放的利弊分别有哪些？

Q5：在大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点的趋势下，请问您所在的企业有制定哪些具体的措施来对大数据进行应用？这些应用给您所在的行业带来了哪些影响？

Q6：从大数据隐私的视角来说，请问您如何看待企业可能面临的数据攻击，您将如何确保自身以及客户数据的安全？

Q7：阿里巴巴创办人马云来台演讲中提到，“未来的时代将不是IT时代，而是DT（Data Technology）的时代”。请问您认为大数据未来的发展中心落在哪，及其发展前景如何？

Q8：请问您认为现下社会各行各业可以如何合作进行大数据开放共享？企业扮演着怎样的角色，可以发挥怎样的作用？

**分析**：企业以经济应用带动国家的发展，但企业中存在明显的商业竞争及数据机密等问题，并且牵涉到群众的切身利益，在群众与各企业的相关行为中，也会将自己的信息资料给予企业。因此，企业要如何在顺应大数据开放的潮流下同时保护自身与客户的安全，便成了重要问题。同样大数据对企业也是潜藏的巨大的商机，就看企业如何利用并促进共同发展。

## 8.3政府版访谈问卷

信息时代的到来对我国政府的决策模式、治理模式和工作方式等都提出了新的挑战。而推进政府大数据战略对实现政府治理有着重要的意义，是政府治理实现的强力助推器。本次社会实践的主题为大数据开放的社会环境调查。我们希望通过前期调查研究、资料整理、技术实现，了解当前大数据共享社会构建的推进情况和困难，并对其提出相应的意见和建议。请您就当下大数据共享开放的情况回答下列问题。

Q1：气象数据可以为农业提供便利，您认为政府部门大数据还可以促进哪些行业的发展？

Q2：数据共享能够促进资源得到更合理的分配，请问您所了解到政府各部门之间数据共享的现状如何？

Q3：国务院《促进大数据发展行动纲要》提出在2018年底前建成国家统一的数据开放平台。您觉得通过各级、各个政府单位协力合作，建立一个大数据信息共享平台对政务有什么帮助？

Q4：诸如卫生部门数据会涉及到患者的个人隐私，您觉得政府部门该如何开发涉及保密的大数据？

Q5：关于政府数据，您觉得是否应向所用者收取费用？

Q6：如今互联网发达，任何人都能从中获取信息。请问过度地开放大数据会有什么弊端？

Q7：目前许多国家正大力支持开放大数据，但是相应政策的制定，数据开放程度的评估还不健全，对于完善政府数据开放，您有什么建议？

**分析**：推进政府大数据战略对实现政府治理有重要意义，且政府的走向更影响着各界工作方向的走向。问卷中我们调查了政府大数据推进情况和困难，了解共享现状以及将采取怎样的措施来促进大数据共享开放。而政府的数据也有很强的机密性，因为数据安全问题也不容忽视。

## 8.4社会人员版访谈问卷

国家《十三五规划纲要》中明确提出实施国家大数据战略,全面实施促进大数据发展行动,加快推动数据资源共享开放和开发应用,助力产业转型升级和社会治理创新。本次社会实践的主题为大数据开放的社会环境调查。我们希望通过前期调查研究、资料整理、技术实现，了解当前大数据共享社会构建的推进情况和困难，并对其提出相应的意见和建议。请您就当下大数据共享开放的情况，回答下列问题。

Q1：您之前是否听说过大数据，是否对大数据战略、大数据开放共享方面有一定的了解。如果是，您主要是通过什么样的途径获取有关大数据战略的信息。

Q2：您在日常生活中是否有受益于大数据开放之处，请谈一谈。

Q3:这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

Q4：在政府相关会议上，也提到了“一些引导社会经济发展的数据，不涉及国家机密的，都应该向公众开放。”您认为政府对公众开放大数据有哪些利与弊。

Q5:您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

Q6：一直以来，社会各界对大数据是否开放，如何开放就争议不断，您对于政府和企业对于大数据的开放有哪些建议。

Q7：如果您手上掌握着您所在行业的很多重要数据，涉及您的一些利益，但是高校教育、公众生活等也对这些数据有一定需求，您会采取怎样的方式来处理这些大数据。

**分析**：最广大的基础还是人民群众，大多数人在日常生活中可能会大数据都是浅层的了解，因此我们也在问卷中询问了社会各界对于企业和政府等有何期望，如何促进发展。同时问卷了解了人们明显感受到的受益之处，用以分析大数据对于公众的作用还有何方面可以再做改进推进发展。

# 9.前期调研

**为了获得更准确的数据，我们采访了社会不同领域的工作人员，对此我们分别准备了相应的问题。**

**对于政府部门，因为它是社会的管理者，起到宏观调控作用，而且掌握大量的数据资源，所以我们将从开放的利弊、监管的规范性、数据共享的现状等方面进行调研。以下是问卷的部分问题。**

Q：数据共享能够促进资源得到更合理的分配，请问您所了解到政府各部门之间数据共享的现状如何？

Q：如今互联网发达，任何人都能从中获取信息。请问过度地开放大数据会有什么弊端？

Q：目前许多国家正大力支持开放大数据，但是相应政策的制定，数据开放程度的评估还不健全，对于完善政府数据开放，您有什么建议？

**对于企业，因为它是数据的利用者，也是财富的创造者，我们将从未来战略，企业竞争，数据隐私以及希望得到的帮助来进行调研。以下是问卷的部分问题。**

Q：请问就您的了解而言，您所在的企业对大数据的利用状况如何？在与业内其他企业之间会存在的商业竞争关系下，您认为可以接受怎样的数据开放程度，如会选择公开哪方面的数据及以什么形式公开？

Q：从大数据隐私的视角来说，请问您如何看待企业可能面临的数据攻击，您将如何确保自身以及客户数据的安全？

Q：请问您认为现下社会各行各业可以如何合作进行大数据开放共享？企业扮演着怎样的角色，可以发挥怎样的作用？

**对于高校，因为它汇聚了社会上精英，并向社会输送大量大数据方面的人才，可以说，高校教育的好坏直接影响到大数据的开发利用，所以我们将从人才培养，发展方向，教育重心等方面进行调研。以下是问卷的部分问题。**

Q：这是一个大数据的时代，您作为教育实践者。您认为高等院校的大数据开发建设工作重心应该是什么？大数据开发管理对教育理念、办学理念带来哪些变化？

Q：您认为符合社会市场要求的大数据人才需要满足哪些基本素养？如何针对您所提的，对学生进行针对性教育与培养？

Q：目前国内高等院校大数据教育开发与人才培养方面的发展现状如何？与国外，有何差距？

**最后，我们还将采访了一部分社会人员，采访对象来自各行各业，从中我们可以了解目前开放大数据这一战略在社会上的认识度，以及听听他们的见解。以下是问卷的部分问题。**

Q：您之前是否听说过大数据，是否对大数据战略、大数据开放共享方面有一定的了解。如果是，您主要是通过什么样的途径获取有关大数据战略的信息。

Q：这是一个大数据的时代，请谈一谈您所在行业对大数据的应用以及这些应用对您所在行业带来的影响。

Q：您是否赞同政府和电信、移动、银行等企业应该向公众开放大数据，如果是，您认为哪一些方面的大数据可能对公众更有用处；如果否，请谈一谈原因。

# 10.中期小结

## 已完成的工作

* 分析调查对象，了解基本情况，确定调查方法以及范围。
* 针对所要调查的社会部门制定相应的问卷。
* 在老师的带领下，我们走访了国家气象局预报中心，上海市气象局科技服务中心，华东师范大学信息办，华东师范大学计算机与软件工程学院。
* 暑假组员们回家采访多个不同领域的从业者，目前得到访谈的行业涉及金融，教育，医疗卫生，IT，政府部门等。
* 将收集到的访谈进行整理汇总。

# 11.后期实现

完善项目报告书，对调研总结结果提出意见和建议，深度了解上海市气象局数据开放共享需求，以此为试点搭建气象大数据开放平台。

# 12．构想

由本次社会实践数据开放双方矛盾提出的两个建设性想法

* 建设气象大数据开放平台
* 举办数据开放创新应用大赛

## 12.1建设气象大数据开放平台

社会实践的过程中，我们访谈了上海市气象局的科技创新主任——刘欧萱女士，了解了气象局对于数据开放的态度和想法。同时我们参考了网络资料，发现上海市政府数据服务网中已经有涉及各个领域如道路交通、教育科技、卫生健康等数据的下载。

通过对类似上海市政府数据服务网这类数据开放网站的参考，以及对气象局这类数据提供者的实地访谈，我们提出了自己的想法——关于建设气象大数据开放平台的建议。

* **建设该平台的目的：**

1. 为渴望获取气象数据的科研单位、高校学生教师、个人兴趣者提供数据获取渠道
2. 为气象局等数据提供者的数据提供开发二次价值的机会，提高数据利用率，产生数据新应用
3. 以气象数据为试点，促进数据开放、数据共享的社会环境建设

* **建设该平台需要注意的事项：**

1. 数据的安全问题
2. 数据开放平台应该设立多级权限设置，实名用户将会有更大的数据量与更多的数据形式提供下载。
3. 实名用户将会签署网上数据安全负责协议。
4. 数据开放平台中的数据应该与数据提供者进行数据安全方面的确认，由数据提供者提供数据授权。
5. 珍贵数据将不公开进行开放，而是由需要者进行申请，由平台进行身份审核后单方面提供下载渠道
6. 数据开放平台本身的防网络攻击能力也需要加强建设
7. 数据的实时性

需要与数据提供者进行合作洽谈，以获取实时的数据进行开放

1. 数据的可视化

除了精确的数据提供下载外，平台也将数据进行可视化，如温度数据，不同温度标注不同颜色，展示在平台上。可视化数据将为非数据研究者提供科普性教育。

* **建设该平台将会遇到的困难：**

1. 初期平台将于气象局进行合作，提供气象数据的开放，平台成熟后将于更多的部门进行联系以获取更多的数据类型提供开放
2. 平台本身的建设需要更多的技术性工作、平台的数据下载流程需要进行设计

* **该平台前景期望：**

平台将会在初期提供气象数据的下载、后期将会提供更多形式的数据。将会有一套完善的机制保障数据安全，一套完整的流程进行数据下载。数据提供者将会获得自身数据的二次利用价值（即数据会产生的创新型应用），数据需要者将会有一个完整的渠道获得数据。同时平台也将为各个部门，各个数据提供者之间的数据共享提供渠道。

## 12.2举办数据开放创新型应用大赛

数据如果不进行分析、挖掘，数据将只会是数据，不会有任何的价值。只有将数据进行一定的加工，获取数据中隐藏的一些数据规律，数据的价值才会被放大。

同时参考上海市SODA数据开放创新型应用大赛，我们提议举行高校级别的数据开放创新型应用大赛，将数据提供给高校生，将命题交给高校生，让高校生在数据中发掘出新的价值。由于高校生的数据挖掘能力与数据分析能力尚不成熟，但是高校生会有新的数据视角新的数据应用角度，因此会产生新的数据创新型应用想法。如果在定期举办的数据开放创新型应用大赛中，能够产生有价值的想法，我们可以再与数据提供者进行协商合作，将该想法进行立项、孵化以产生更成熟、更贴近社会的应用。

同时可以与之前的构想数据开放平台进行结合，一方面提高大赛的知名度，一方面由于可以产生新的数据价值，平台能够吸引更多的数据提供者开放他们的数据，扩大平台的数据量。

同时高校级别的数据开放创新型应用大赛能够提高高校对大数据的认知度，培养更多的数据方向人才。为数据开放、数据共享的社会环境建设献一份力。

# 13.附录

国家大数据战略

——习近平与“十三五”十四大战略

大数据很忠诚，它真实记录人们的每个足迹，深藏功与名；大数据很任性，它的分析有根有据，拒绝流言蜚语；大数据很友好，它提供各种权威参考，它创造绿色经济，让我们的生活更美好。世界已经进入由数据主导的“大时代”。以习近平同志为总书记的新一届中央领导集体，站在时代最前沿，带领全国人民迈入大数据时代。五中全会的“十三五”规划建议提出：“实施国家大数据战略，推进数据资源开放共享。”“学习中国”今天推出《习近平与“十三五”十四大战略》系列文章之“国家大数据战略”篇。

一、大数据引领生活新变化

我们生活在一个充满“数据”的时代，我们打电话、用微博、聊QQ、刷微信，我们阅读、购物、看病、旅游，都在不断产生新数据，“堆砌”着数据大厦。大数据已经与我们的工作生活息息相关、须臾难离。中国工程院院士高文说：“不管你是否认同，大数据时代已经来临，并将深刻地改变着我们的工作和生活。”2015年5月，习近平在给国际教育信息化大会的贺信中说，“当今世界，科技进步日新月异，互联网、云计算、大数据等现代信息技术深刻改变着人类的思维、生产、生活、学习方式，深刻展示了世界发展的前景。”

“大数据”并不神秘。实际上，大数据就在你我身边，虽然你看不到它，不在意它，但它却始终在你身边、且已经并将继续影响着你的生活。习近平和彭丽媛访问英国帝国理工学院时，校方赠送给彭丽媛的羊绒披肩就有大数据的功劳，披肩尺寸就是通过计算机图像分析技术得出的。校方还向习近平展示了运用大数据的方法分析国内人口迁移情况，“一带一路”政策的国际影响力，个性化医疗的推广，以及上海地铁的负载分布和应急办法等。习近平对利用大数据技术帮助提高人民生活质量的做法给予高度评价。

大数据已经广泛应用于日常生活息息相关的诸多领域。必应搜索通过集成以往的飞机票价画出未来票价走势；谷歌利用用户搜索记录判断出美国流感疫情的现状，在时间上比美国疾控中心的预测还快两周；对冲基金通过剖析社交网络推特的数据信息来预测股市的表现；交通部门通过大数据分析出实时路况；交友网站利用大数据分析来寻找爱情，并来帮助需要的人匹配合适的对象；利用穿戴装备（例如智能手表或者智能手环）生成最新的数据，可以根据自身热量的消耗以及睡眠模式来进行追踪身体是否健康。这些无一不体现出大数据与我们的生活紧密相连。大数据掀起的风暴已席卷到生活的各个角落，大数据将会使我们的生活更加美好。

二、大数据孕育发展新思路

数据是发展的产品，发展依靠数据获得新思路。“数谷”是贵阳这座发展相对落后的西部山城的新名片，2014年贵阳在大数据产业的带动下实现地区生产总值2497.27亿元、增长13.9％。2014年“两会期间”，习近平在参加全国两会贵州代表团审议时指出：“贵州搞大数据产业招商，发展电子信息产业，是很好的选择。”2015年6月17日，习近平来到贵阳市大数据广场，走进大数据应用展示中心，听取贵州大数据产业发展、规划和实际应用情况介绍。贵州省以发展大数据作为突破口推动经济社会发展的探索，给习近平留下深刻印象，习近平对当地干部说：“我听懂了，贵州发展大数据确实有道理。”

“大数据经济”的背后逻辑，是中国经济悄然发生的“质变”。中国经济在增速放缓的情况下，就业率不降反增，一个重要原因就是进行了政府的自身改革，推出一系列简政放权举措，引发小微企业、个体户井喷式增长。而这些企业、商户中多数人的经营业务都是依托于互联网展开的，都与数据有关，大数据正在孕育出更多的新业态。李克强指出：“13亿中国人，八、九亿劳动力，这其中如果有越来越多的人依托新业态发展，就会培育出中国经济发展的新‘发动机’，也必将会对社会发展、人民进步造成深刻影响。”

大数据不只是一个产业这么简单。它在社会的各个领域中都无所不在，可以与N个产业“相加”，形成“大数据+”，“互联网+”的本质是连接和数据。中国联通网络技术研究院首席专家唐雄燕说：“数据已经成为一种新的经济资产类别，就像货币或黄金一样，将形成数据材料、数据探矿、数据加工、数据服务等一系列新兴产业。”2014年3月5日，李克强在十二届全国人大二次会议上作政府工作报告时说，要设立新兴产业创业创新平台，在新一代移动通信、集成电路、大数据、先进制造、新能源、新材料等方面赶超先进，引领未来产业发展。这是“大数据”首次进入政府工作报告，也表明其作为一种新兴产业，将得到国家层面的大力支持。

大数据时代，数据正在成为一种生产资料，成为一种稀有资产和新兴产业。任何一个行业和领域都会产生有价值的数据，而对这些数据的统计、分析、挖掘和人工智能则会创造意想不到的价值和财富。习近平强调：“机会稍纵即逝，抓住了就是机遇，抓不住就是挑战。”

三、大数据开辟国家治理新路径

大数据不仅是一场技术革命，一场经济变革，也是一场国家治理的变革。牛津大学教授维克托·迈尔·舍恩伯格在其著作《大数据时代》中说：“大数据是人们获得新的认知、创造新的价值的源泉，还是改变市场、组织机构，以及政府与公民关系的方法。”

大数据时代，互联网是政府施政的新平台。单纯依靠政府管理和保护数据的做法会使政府在面对大规模而复杂的数据时应接不暇、不堪重负。而通过电子政务系统，可以实现在线服务，做到权力运作有序、有效、“留痕”，促进政府与民众的沟通互联，提高政府应对各类事件和问题的智能化水平。“十三五”规划建议指出：“运用大数据技术，提高经济运行信息及时性和准确性。”

大数据正有力地推动着国家治理体系和治理能力走向现代化，正日益成为社会管理的驱动力、政府治理的“幕僚高参”。习近平在参观腾讯公司时说：“互联网在社会管理方面有较大作用，我们怎么去适应它？我看到你们做的工作都是很重要的，比如在这样的海量信息中，你们占有了最充分的数据，然后可以做出最客观、精准的分析。这方面对政府提供的建议是很有价值的。”李克强在考察北京·贵阳大数据应用展示中心时说：“把执法权力关进‘数据铁笼’，让失信市场行为无处遁形，权力运行处处留痕，为政府决策提供第一手科学依据，实现‘人在干、云在算’。”李克强考察山东浪潮集团时指出：“不管是推动政府的简政放权，放管结合，还是推进新型工业化、城镇化、农业现代化，都要依靠大数据、云计算。所以，它应该是大势所趋，是一个潮流。”大数据使得政府决策的基础从少量的“样本数据”转变为海量的“全体数据”。政府树立大数据意识，促进相关数据完全共享，更多地依赖数据进行决策，可以实现从以有限个案为基础向“用数据说话”转变的全新决策。

近年来，“大数据”议题已经成为了国务院常务会议的座上客，“大数据”战略早露端倪。2014年7月23日，国务院常务会议审议通过《企业信息公示暂行条例（草案）》，推动构建公平竞争市场环境。其中要求建立部门间互联共享信息平台，运用大数据等手段提升监管水平。2014年9月17日，部署进一步扶持小微企业发展，推动大众创业，万众创新，其中包括加大服务小微企业的信息系统建设，方便企业获得政策信息，运用大数据、云计算等技术提供更有效服务。2014年10月29日，要求重点推进6大领域消费，其中强调加快健康医疗、企业监管等大数据应用。2014年11月15日，提出在疾病防治，灾害预防，社会保障，电子政务等领域开展大数据应用示范。2015年1月14日，部署加快发展服务贸易，以结构优化拓展发展空间，提出要创新模式，利用大数据、物联网等新技术打造服务贸易新型网络平台。2015年2月6日，确定运用互联网和大数据技术，加快建设投资项目在线审批监管平台，横向联通发展改革，城乡规划，国土资源，环境保护等部门，纵向贯通各级政府，推进网上受理、办理、监管“一条龙”服务，做到全透明，可核查，让信息多跑路，群众少跑腿。2015年7月，国务院办公厅印发的《关于运用大数据加强对市场主体服务和监管的若干意见》提出，要提高对市场主体服务水平；加强和改进市场监管；推进政府和社会信息资源开放共享；提高政府运用大数据的能力；积极培育和发展社会化征信服务。

四、大数据重塑世界新格局

“数据是新的石油,是本世纪最为珍贵的财产。”大数据正在改变各国综合国力，重塑未来国际战略格局。2013年7月，习近平视察中国科学院时指出：“大数据是工业社会的‘自由’资源，谁掌握了数据，谁就掌握了主动权。”

大数据正在成为经济社会发展新的驱动力。随着云计算、移动互联网等网络新技术的应用和发展与普及，社会信息化进程进入数据时代，海量数据的产生与流转成为常态。预计到2020年，全球数据使用量将达到约400亿TB，将涵盖经济社会发展各个领域，成为新的重要驱动力。大数据重新定义了各个大国博弈的空间。在大数据时代，世界各国对数据的依赖快速上升，国家竞争焦点已经从资本、土地、人口、资源的争夺转向了对大数据的争夺。习近平在中央网络安全和信息化领导小组第一次会议上指出：“网络信息是跨国界流动的，信息流引领技术流、资金流、人才流，信息资源日益成为重要生产要素和社会财富，信息掌握的多寡成为国家软实力和竞争力的重要标志。”未来国家层面的竞争力将部分体现为一国拥有数据的规模、活性以及解释、运用的能力，数字主权将成为继边防、海防、空防之后另一个大国博弈的空间。大数据将改变国家治理架构和模式。在大数据时代，用大数据可以通过对海量、动态、高增长、多元化、多样化数据的高速处理，快速获得有价值信息，提高公共决策能力。

鉴于大数据潜在的巨大影响，很多国家或国际组织都将大数据视作战略资源，并将大数据提升为国家战略。2012年3月，美国奥巴马政府宣布了“大数据研发计划”，并设立了2亿美元的启动资金，希望增强海量数据收集、分析萃取能力，认为这事关美国的国家安全和未来竞争力。迄今为止，美国在大数据方面实施了三轮政策，开放了50多个门类的政府数据确保商业创新。欧盟正在力推《数据价值链战略计划》为320万人增加就业机会。日本积极谋划利用大数据改造国家治理体系，对冲经济下行风险。联合国推出的“全球脉动”项目，希望利用“大数据”预测某些地区的失业率或疾病爆发等现象，以提前指导援助项目。截至2014年4月，全球已有63个国家制定了开放政府数据计划，数据开放推动政府从“权威治理”向“数据治理”转变。中国国际经济交流中心副研究员张茉楠撰文指出，中国需要加快形成大数据国家战略。着力规划“大数据战略”中长期路线图与实施重点、目标、路径，统筹布局，加快大数据发展核心技术研发，推进大数据开放、共享以及安全方面的相关立法与标准制定，抢占新全球科技革命和产业革命战略机遇期，重构国家综合竞争优势已经迫在眉睫。

大数据是每个人的大数据，是每个企业的大数据，更是整个国家的大数据。大数据时代拥抱大数据，“十三五”规划建议已经吹响向大数据进军的号角，随着国家大数据战略的实施，基于大数据的智慧生活、智慧企业、智慧城市、智慧政府、智慧国家必将一一实现。