

## 热点聚焦

**编者按：**随着云计算、物联网的发展，我们迎来了大数据的时代。2013 年被国外媒体称为“大数据元年”。大数据如浪潮一般席卷全世界，不仅在信息技术行业备受瞩目，更成为变革科研、商业、政府运作方式乃至人类思维方式的一个热点。人类社会已经进入了一个全新的历史阶段——大数据时代，谁拥有数据资源，谁用活了数据资源，谁就拥有未来。

赛迪顾问统计数据显示：2012 年我国大数据市场规模仅为 4.5 亿元，2013 年中国大数据仅 IT 应用市场规模就达到了 34.3 亿元，展现了良好的增长势头。

进入 2014 年，大数据市场的新产品、新技术、新服务正在不断涌现，我国大数据产业存在重大机遇也面临巨大挑战。我国大数据发展趋势怎样？国内大数据产业发展存在的问题与解决办法是什么？金融、信息技术、消费、教育、医疗等行业如何在大数据中挖掘价值等问题成为国内大数据产业发展关注的焦点。

## 大数据的基本知识

### 什么是大数据？

大数据，或称巨量数据、海量数据；是由数量巨大、结构复杂、类型众多数据构成的数据集合，是基于云计算的数据处理与应用模式，通过数据的集成共享，交叉复用形成的智力资源和知识服务能力。<sup>[1]</sup>

有研究机构如此定义“大数据”：“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。从某种程度上说，大数据是数据分析的前沿技术。简言之，从各种各样类型的数据中，快速获得有价值信息的能力，

就是大数据技术。

### 大数据从何而来？

美国互联网数据中心指出，互联网上的数据每年将增长 50%，每两年便将翻一番，目前世界上 90% 以上的数据是最近几年才产生的。此外，全世界的工业设备、汽车、电表上有着无数的数码传感器，随时测量和传递着有关位置、运动、震动、温度、湿度乃至空气中化学物质的变化，也产生了海量的数据信息。

物联网、云计算、移动互联网、车联网、手机、平板电脑、PC 以及各种各样的传感器，无一不是数据来源或者承载的方式。

### 大数据有多大？

仅以互联网为例，一天之中，互联网产生的全部内容可以刻满 1.68 亿张 DVD；发出的邮件有 2940 亿封之多；发出的社区帖子达 200 万个，相当于《时代》杂志 770 年的文字量……

截止到 2012 年，数据量已经从 TB(1024GB=1TB) 级别跃升到 PB(1024TB=1PB)、EB(1024PB=1EB) 乃至 ZB(1024EB=1ZB) 级别。国际数据公司 (IDC) 的研究结果表明，2008 年全球产生的数据量高达 1.82ZB，相当于全球每人产生 200GB 以上的数据。而到 2012 年为止，人类生产的所有印刷材料的数据量是 200PB，全人类历史上说过的所有话的数据量大约是 5EB。IBM 的研究称，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。而到了 2020 年，全世界所产生的数据规模将达到今天的 44 倍。

### 大数据的“4 个 V”

大数据的 4 个“V”指的是大数据的 4 个特点：第一，数据体量巨大。从 TB 级别，跃升到 PB 级别；第二，数据类型繁多，数据来源于各种各样的渠道。第三，价值密度低，商业价值高。以视频为例，连续不间断监控过程中，可能有用的数据仅仅有一两秒。第四，处理速度快。一般要在秒级时间范围内给出分析结果，时间太长就失去价

值了。这个速度要求是大数据处理技术和传统的数据挖掘技术最大的区别。

由此，业界将大数据的特点归纳为 4 个“V”——Volume(大量)、Velocity(高速)、Variety(多样)、Veracity(精确)。

### 什么是大数据思维？<sup>[2]</sup>

《大数据时代》一书的作者维克托·迈尔-舍恩伯格认为，1-需要全部数据样本而不是抽样；2-关注效率而不是精确度；3-关注相关性而不是因果关系。

阿里巴巴的王坚对于大数据也有一些独特的见解，比如，

“今天的数据不是大，真正有意思的是数据变得在线了，这个恰恰是互联网的特点。”

“非互联网时期的产品，功能一定是它的价值，今天互联网的产品，数据一定是它的价值。”

“你千万不要想着拿数据去改进一个业务，这不是大数据。你一定是去做了一件以前做不了的事情。”

特别是最后一点，充分说明了大数据的真正价值在于创造，在于填补无数个还未实现过的空白。

有人把数据比喻为蕴藏能量的煤矿。煤炭按照性质有焦煤、无烟煤、肥煤、贫煤等分类，而露天煤矿、深山煤矿的挖掘成本又不一样。与此类似，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。

### 大数据的分类<sup>[3]</sup>

大数据可分成大数据技术、大数据工程、大数据科学和大数据应用等领域。目前人们谈论最多的是大数据技术和大数据应用。工程和科学问题尚未被重视。大数据技术指从各种各样类型的数据中，快速获得有价值信息的能力；大数据工程指大数据的规划建设运营管理的系统工程；大数据科学关注大数据网络发展和运营过程中发现和验证

大数据的规律及其与自然和社会活动之间的关系。

## 大数据的实践<sup>[2]</sup>

### 互联网的大数据

互联网上的大数据很难清晰的界定分类界限，我们先看看二巨头的大数据：

阿里巴巴拥有交易数据和信用数据。这两种数据更容易变现，挖掘出商业价值。除此之外阿里巴巴还通过投资等方式掌握了部分社交数据、移动数据。如微博和高德。

腾讯拥有用户关系数据和基于此产生的社交数据。这些数据可以分析人们的生活和行为，从里面挖掘出政治、社会、文化、商业、健康等领域的信息，甚至预测未来。

在信息技术更为发达的美国，除了行业知名的类似 Google，Facebook 外，已经涌现了很多大数据类型的公司，它们专门经营数据产品，比如：

Metamarkets：这家公司对 Twitter、支付、签到和一些与互联网相关的问题进行了分析，为客户提供了很好的数据分析支持。

Tableau：他们的精力主要集中于将海量数据以可视化的方式展现出来。Tableau 为数字媒体提供了一个新的展示数据的方式。他们提供了一个免费工具，任何人在没有编程知识背景的情况下都能制造出数据专用图表。这个软件还能对数据进行分析，并提供有价值的建议。

ParAccel：他们向美国执法机构提供了数据分析，比如对 15000 个有犯罪前科的人进行跟踪，从而向执法机构提供了参考性较高的犯罪预测。他们是犯罪的预言者。

QlikTech：QlikTech 旗下的 Qlikview 是一个商业智能领域的自主服务工具，能够应用于科学研究和艺术等领域。为了帮助开发者对

这些数据进行分析，QlikTech 提供了对原始数据进行可视化处理等功能的工具。

**GoodData:** GoodData 希望帮助客户从数据中挖掘财富。这家创业公司主要面向商业用户和 IT 企业高管，提供数据存储、性能报告、数据分析等工具。

**TellApart:** TellApart 和电商公司进行合作，他们会根据用户的浏览行为等数据进行分析，通过锁定潜在买家方式提高电商企业的收入。

**DataSift:** DataSift 主要收集并分析社交网络媒体上的数据，并帮助品牌公司掌握突发新闻的舆论点，并制定有针对性的营销方案。这家公司还和 Twitter 有合作协议，使得自己变成了行业中为数不多可以分析早期 tweet 的创业公司。

**Datahero:** 公司的目标是将复杂的数据变得更加简单明了，方便普通人去理解和想象。

### **政府的大数据**

美国麦肯锡全球研究院 2011 年 6 月发布题为《大数据：下一个创新、竞争和生产力的前沿》的研究报告，指出“大数据时代已经到来”，数据正成为与物质资产和人力资本相提并论的重要生产要素，大数据的使用将成为未来提高竞争力的关键要素。美国政府于 2012 年 3 月宣布“大数据的研究和发展计划”<sup>[4]</sup>，投资 2 亿美元拉动大数据相关产业发展，将“大数据战略”上升为国家意志。奥巴马政府将数据定义为“未来的新石油”，并表示一个国家拥有数据的规模、活性及解释运用的能力将成为综合国力的重要组成部分，未来，对数据的占有和控制甚至将成为陆权、海权、空权之外的另一种国家核心资产。

在国内，政府各个部门都握有构成社会基础的原始数据，比如，气象数据，金融数据，信用数据，电力数据，煤气数据，自来水数据，道路交通数据，客运数据，安全刑事案件数据，住房数据，海关数据，

出入境数据，旅游数据，医疗数据，教育数据，环保数据等等。这些数据在每个政府部门里面看起来是单一的，静态的。但是，如果政府可以将这些数据关联起来，并对这些数据进行有效的关联分析和统一管理，这些数据必定将获得新生，其价值是无法估量的。

具体来说，现在城市都在走向智能和智慧，比如，智能电网、智慧交通、智慧医疗、智慧环保、智慧城市，这些都依托于大数据，可以说大数据是智慧的核心能源。从国内整体投资规模来看，到 2012 年底全国开建智慧城市的城市数超过 180 个，通信网络和数据平台等基础设施建设投资规模接近 5000 亿元。“十二五”期间智慧城市建设拉动的设备投资规模将达 1 万亿元人民币。大数据为智慧城市的各个领域提供决策支持。在城市规划方面，通过对城市地理、气象等自然信息和经济、社会、文化、人口等人文社会信息的挖掘，可以为城市规划提供决策，强化城市管理服务的科学性和前瞻性。在交通管理方面，通过对道路交通信息的实时挖掘，能有效缓解交通拥堵，并快速响应突发状况，为城市交通的良性运转提供科学的决策依据。在舆情监控方面，通过网络关键词搜索及语义智能分析，能提高舆情分析的及时性、全面性，全面掌握社情民意，提高公共服务能力，应对网络突发的公共事件，打击违法犯罪。在安防与防灾领域，通过大数据的挖掘，可以及时发现人为或自然灾害、恐怖事件，提高应急处理能力和安全防范能力。

### 企业的大数据

企业的 CXO 们最关注的还是报表曲线的背后能有怎样的信息，他该做怎样的决策，其实这一切都需要通过数据来传递和支撑。在理想的世界中，大数据是巨大的杠杆，可以改变公司的影响力，带来竞争差异、节省金钱、增加利润、愉悦买家、奖赏忠诚用户、将潜在客户转化为客户、增加吸引力、打败竞争对手、开拓用户群并创造市场。

对于企业的大数据，还有一种预测：随着数据逐渐成为企业的一

种资产，数据产业会向传统企业的供应链模式发展，最终形成“数据供应链”。这里尤其有两个明显的现象：1) 外部数据的重要性日益超过内部数据。在互联互通的互联网时代，单一企业的内部数据与整个互联网数据比较起来只是沧海一粟；2) 能提供包括数据供应、数据整合与加工、数据应用等多环节服务的公司会有明显的综合竞争优势。

## 大数据应用的典型案例

1. google 流感趋势 (Google Flu Trends) 利用搜索关键词预测禽流感的散布。

2. 麻省理工学院利用手机定位数据和交通数据建立城市规划。<sup>[4]</sup>

3. 统计学家内特·西尔弗 (Nate Silver) 利用大数据成功预测 2012 美国选举结果。<sup>[6]</sup>

4. Amazon 利用用户的购买和浏览历史数据进行有针对性的书籍购买推荐，以此有效提升销售量。<sup>[2]</sup>

5. 梅西百货的实时定价机制。根据需求和库存的情况，该公司基于 SAS 的系统对多达 7300 万种货品进行实时调价。

6. 沃尔玛的搜索。这家零售业寡头为其网站 Walmart.com 自行设计了最新的搜索引擎 Polaris，利用语义数据进行文本分析、机器学习和同义词挖掘等。根据沃尔玛的说法，语义搜索技术的运用使得在线购物的完成率提升了 10%到 15%。这就意味着数十亿美元的金額。

7. 快餐业的视频分析。某快餐企业公司通过视频分析等候队列的长度，然后自动变化电子菜单显示的内容。如果队列较长，则显示可以快速供给的食物；如果队列较短，则显示那些利润较高但准备时间相对长的食品。

8. PredPol Inc.。PredPol 公司通过与洛杉矶和圣克鲁斯的警

方以及一群研究人员合作，基于地震预测算法的变体和犯罪数据来预测犯罪发生的几率，可以精确到 500 平方英尺的范围内。在洛杉矶运用该算法的地区，盗窃罪和暴力犯罪分布下降了 33%和 21%。

9. American Express (美国运通, AmEx) 和商业智能。以往, AmEx 只能实现事后诸葛式的报告和滞后的预测。而传统的 BI 已经无法满足业务发展的需要。于是, AmEx 开始构建真正能够预测忠诚度的模型, 基于历史交易数据, 用 115 个变量来进行分析预测。该公司表示, 对于澳大利亚将于之后四个月中流失的客户, 已经能够识别出其中的 24%。

10. Express Scripts Holding Co. 的产品制造。该公司发现那些需要服药的人常常也是最可能忘记服药的人。因此, 他们开发了一个新产品: 会响铃的药品盖和自动的电话呼叫, 以此提醒患者按时服药。

11. Infinity Property & Casualty Corp. 的黑暗数据 (dark data)。所谓黑暗数据, 就是那些针对单一目标而收集的数据, 通常用过之后就被归档闲置, 其真正价值未能被充分挖掘。在特定情况下, 这些数据可以用作其他用途。该公司用累积的理赔师报告来分析欺诈案例, 通过算法挽回了 1200 万美元的代位追偿金额。<sup>[5]</sup>

12. 塔吉特 (Target) 百货孕妇营销分析。塔吉特百货超市以 20 多种怀孕期间孕妇可能会购买的商品为基础, 将所有用户的购买记录作为数据来源, 通过构建模型分析购买者的行为相关性, 能准确的推断出孕妇的具体临盆时间, 这样 Target 的销售部门就可以有针对性的在每个怀孕顾客的不同阶段寄送相应的产品优惠券。<sup>[2]</sup>

## 大数据的现在和未来<sup>[2]</sup>

先看看大数据在当下有怎样的杰出表现:



大数据帮助政府实现市场经济调控、公共卫生安全防范、灾难预警、社会舆论监督；

大数据帮助城市预防犯罪，实现智慧交通，提升紧急应急能力；

大数据帮助医疗机构建立患者的疾病风险跟踪机制，帮助医药企业提升药品的临床使用效果，帮助艾滋病研究机构为患者提供定制的药物；

大数据帮助航空公司节省运营成本，帮助电信企业实现售后服务质量提升，帮助保险企业识别欺诈骗保行为，帮助快递公司监测分析运输车辆的故障险情以提前预警维修，帮助电力公司有效识别预警即将发生故障的设备；

大数据帮助电商公司向用户推荐商品和服务，帮助旅游网站为旅游者提供心仪的旅游路线，帮助二手市场的买卖双方找到最合适的交易目标，帮助用户找到最合适的商品购买时期、商家和最优惠价格；

大数据帮助企业提升营销的针对性，降低物流和库存的成本，减少投资的风险，以及帮助企业提升广告投放精准度；

大数据帮助娱乐行业预测歌手，歌曲，电影，电视剧的受欢迎程度，并为投资者分析评估拍一部电影需要投入多少钱才最合适，否则就有可能收不回成本；

大数据帮助社交网站提供更准确的好友推荐，为用户提供更精准的企业招聘信息，向用户推荐可能喜欢的游戏以及适合购买的商品。

其实，这些还远远不够，未来大数据的身影应该无处不在，就算无法准确预测大数据终会将人类社会带往到哪种最终形态，但有理由相信只要发展脚步在继续，因大数据而产生的变革浪潮将很快淹没地球的每一个角落。

比如，Amazon 的最终期望是：“最成功的书籍推荐应该只有一本书，就是用户要买的下一本书。”

Google 也希望当用户在搜索时，最好的体验是搜索结果只包含

用户所需要的内容，而这并不需要用户给予 Google 太多的提示。

而当物联网发展到达一定规模时，借助条形码、二维码、RFID 等能够唯一标识产品，传感器、可穿戴设备、智能感知、视频采集、增强现实等技术可实现实时的信息采集和分析，这些数据能够支撑智慧城市，智慧交通，智慧能源，智慧医疗，智慧环保的理念需要，这些都所谓的智慧将是大数据的采集数据来源和服务范围。

未来的大数据除了将更好的解决社会问题，商业营销问题，科学技术问题，还有一个可预见的趋势是以人为本的大数据方针。人才是地球的主宰，大部分的数据都与人类有关，要通过大数据解决人的问题。

比如，建立个人的数据中心，将每个人的日常生活习惯，身体体征，社会网络，知识能力，爱好性情，疾病嗜好，情绪波动……换言之就是记录人从出生那一刻起的每一分每一秒，将除了思维外的一切都储存下来，这些数据可以被充分的利用：

医疗机构将实时监测用户的身体健康状况；

教育机构更有针对性的制定用户喜欢的教育培训计划；

服务行业为用户提供即时健康的符合用户生活习惯的食物和其它服务；

社交网络能为你提供合适的交友对象，并为志同道合的人群组织各种聚会活动；

政府能在用户的心理健康出现问题时有效的干预，防范自杀，刑事案件的发生；

金融机构能帮助用户进行有效的理财管理，为用户的资金提供更有效的使用建议和规划；

道路交通、汽车租赁及运输行业可以为用户提供更合适的出行线路和路途服务安排。

## 大数据时代，政府要领跑<sup>[7]</sup>

广泛采集数据、综合处理数据，实现公共服务的技术创新、管理创新和模式创新，这是大数据时代的必然选择

点开上海市政府数据服务网，房地产开发企业信息、社保卡受理网点、派出所基本情况……有六大领域数据都可下载使用。近日，上海在此前 9 家单位试点的基础上，要求当地所有政府部门都要在年内向公众提供数据产品浏览、查询和下载等服务。大数据时代，政府如何提升服务能力与治理水平？这样的探索无疑有着重要意义。

2013 年被国外媒体称为“大数据元年”，这一概念迅速成为热词。联合国的一份研究报告显示，全球的数据存量在 2010 年已有 1200 艾字节（EB），并将以 40% 的年增长率继续递增。而仅仅是 1 艾字节，就相当于 13 亿中国人人手一本 500 页的书加起来的的信息量。难怪有人说，数据将成为一切组织运行的基本要素，其战略意义甚至将超过土地、人力、技术和资本。

政府部门在数据占有方面，无疑具有天然的优势。有专门的统计部门、干部队伍进行相关工作，有人口普查、经济普查一类的信息资料，日常工作中也积累了大量与社会经济生活息息相关的数据。从大数据的角度看来，那些沉睡在档案袋、文件夹中的数据，有着无比巨大的价值，能产生惊人的效用。比如，某导航公司将上海公开的 2 万多条地理位置信息用于地图编制与更新，服务了上亿用户。北京也结合政务数据和社会力量，开发出“游北京”和“爱健康”两个程序，提供餐饮旅游和卫生保健指南。

遗憾的是，很多数据还在沉睡。一些政府部门，或是缺乏看深一层的眼界，没有“大数据思维”，把自己掌握的丰富信息锁在柜中、束之高阁；或是缺乏迈开步子的勇气，摆脱不了“数据小农意识”，动辄以保密和隐私说事；还有少数人把政务数据当作秘而不宣的资源，甚至出现过倒卖个人数据的案例。这些都极大妨碍了政府在大数

据时代保障公众知情权、提高自身服务能力的步伐。

在新加坡，智能交通综合信息管理平台通过准确预测交通流速和流量，显著提升了高峰时段车辆的通行效率。这个例子说明，政府领跑大数据时代，公开数据还只是第一步。大数据不仅是技术变革，更是一场社会变革，必然伴随公共管理与公共服务领域的变革。很大程度上，大数据就是政府治理现代化的一条技术路径，具有催生管理革命的效果，也必将给政府职能转变和机构改革带来新的气象。

比如，购买一套住房需要填报十几张表格，一些表格 1/3 以上的内容是重复的，完全可以利用政府拥有的基础数据自动生成；比如，“北漂”小伙儿为办护照返乡 6 次补开 5 张证明，如果实现了综合数据联网，完全可以由政府部门内部调取，不必让老百姓急断肠、跑断腿。广泛采集数据、综合处理数据，实现政府公共服务的技术创新、管理创新和服务模式创新，这是大数据时代的必然选择。

党的十八届三中全会提出推进国家治理体系和治理能力现代化，如何实现？不断开拓治理资源、创新治理手段，无疑是重要内容。大数据时代到来，我们正该把握住这一时代潮流，将大数据运用到经济、社会运行的方方面面，在领跑中彰显以人为本的执政理念、与时俱进的执政风格和改革创新的执政品质。

#### 参考文献：

- 1、 带你了解大数据. 中国大数据. 2014-02-06
- 2、 大数据究竟是什么？一篇文章让你认识并读懂大数据. 中国大数据. 2013-11-04
- 3、 大数据. 模式科技. 2013-08-28
- 4、 大数据：价值何在？. 人民日报. 2013-06-18
- 5、 CIO 必须知道的十个大数据案例 . 中国大数据. 2013-12-24
- 6、 大数据成功预测美国大选 . 中国大数据. 2014-2-25
- 7、 大数据时代，政府要领跑. 人民日报. 2014-05-20

## 延伸阅读

### 大数据与云计算的关系

#### 什么是云计算？

云计算(cloud computing)是通过互联网把多个成本较低的计算实体整合成一个具有强大计算能力的完美系统。简单理解就是，运营公司提供服务器、应用程序、存储空间，用户通过网络远程登录服务器，并按照需要使用这些存储空间和应用程序。狭义云计算是指IT基础设施的交付和使用模式，指通过网络以按需、易扩展的方式获得所需资源；广义云计算指服务的交付和使用模式，指通过网络以按需、易扩展的方式获得所需服务。它意味着计算能力也可作为一种商品通过互联网进行流通和交易。（摘自《带你了解大数据》中国大数据 2014-02-06）

#### 大数据与云计算的关系

大数据常和云计算联系到一起，因为实时的大型数据集分析无法用单台的计算机进行处理，必须采用分布式计算架构。大数据的特色在于对海量数据的挖掘，但它必须依托云计算的分布式处理、分布式数据库、云存储和虚拟化技术。两者之间结合后会产生如下效应：可以提供更多基于海量业务数据的创新型服务；通过云计算技术的不断发展降低大数据业务的创新成本。可以说，云计算充当了工业革命时期的发动机的角色，而大数据则是电。

业内是这么形容两者的关系：没有大数据的信息积淀，则云计算的计算能力再强大，也难以找到用武之地；没有云计算的处理能力，则大数据的信息积淀再丰富，也终究只是镜花水月。

#### 大数据与云计算区别

云计算与大数据最明显的区分在两个方面：

第一，在概念上两者有所不同云计算改变了IT，而大数据则改

变了业务。然而大数据必须有云作为基础架构，才能得以顺畅运营。

第二，大数据和云计算的目标受众不同，云计算是 CIO 等关心的技术层，是一个进阶的 IT 解决方案。而大数据是 CEO 关注的、是业务层的产品，而大数据的决策者是业务层。

（摘自《大数据究竟是什么？一篇文章让你认识并读懂大数据》  
中国大数据 2013-11-04）

## 一些与大数据容易混淆的概念

**大数据≠数据中心：**数据中心（Internet Data Center，IDC）是对互联网业务资源进行集中式处理和分发的物理环境。在大数据产业的传输层，是大数据应用的网络基础设施。

**大数据≠数字化信息：**数字化信息是大数据的组成部分，但不是所有的数字化信息都能产生大数据。大数据是数字化信息被生产、消费的过程的记录。

**大数据≠海量数据：**海量是大数据的特征之一，但大数据并不是简单地指量大的数据，而是能够反映人—人、人—物、物—物之间关联关系及过程演化的海量数据。

（中文互联网数据资讯中心 2014-08-29）

## 关于“大数据”的四个误区

### 大数据误区一：只要大就好

国内新兴应用交付厂商太一星晨产品总监冯晓杰表示，数据如果仅仅是大那是没多大用处的！就好像资金的意义在于如何使用周转一样，数据大了，但不使用，让它孤零零地偏安机房一隅，那它就不是大数据了，而是有点“败家子”的意思。

比如不少传统的门户网站，基本上就处于“坐拥金山却无福消费”的境况。每天上亿的用户量，却只是简单的广告呈现，没有通过对数据的分析产生更多价值。

### 大数据误区二：只有技术大牛才懂大数据？

专家表示，其实这样的观点并不全对。对于大数据的应用更多的是一种战略能力，而非细节的执行技能，这种能力是可以帮助决策者能从无尽的数据里看出商机看出价值，从而为企业带来更高的利润。而作为决策者并不用太关心大数据的技术细节层面。

### 大数据误区三：是个公司都得上大数据

专家表示，虽然大数据固然是个香饽饽，但不是所有人都能消化得了，而是要衡量企业的现状，看清楚主次矛盾，或是要考量好投入产出的回报率，大数据并不是适合所有企业的现状。

比如，对于中小型网站来说，一上来就盲目追求先进“高大上”的技术架构，那就有点“宰牛刀杀鸡”的意思。对于这类网站，首要考虑的是商业运作模式和推广，只有等到用户量飙升后，再去考虑技术升级这种大事儿。

### 大数据误区四：我就要海量数据

自从大数据概念火了以后，不少企业在遇到问题的时候，总是会情不自禁的就会想到“是不是我的数据量不够？”“是不是如果有了海量的大数据就能变得更好？”其实，这又是陷入了一个误区。

这又回到了大数据价值和金钱价值的类比概念上。比如用搜索引擎搜索一下“存款 贬值”，那么很快就可以发现类似这样的信息：“五十年前的百万变 13 块”，“一万元存一年赔 19 元”……显然，不流动的钱，是越放越没有价值，而基数越大，可能导致的损失就越大。

金钱如此，大数据亦然。只有像比特币玩家们一样，不停地使用数据，并以无比的热情挖掘数据背后的关系和价值，才能如滚雪球一般，使数据之间的相互关系更丰富更完善。同理，对于企业的大数据

来说，只有充分利用大数据，让大数据充分流动起来，不断的实现增值效果，那么才有机会更大的释放大数据的能量。

因此，专家指出，对于企业决策者来说，看待大数据必须有一个清醒的认识，当在脑袋发热准备花大价钱上大数据之前，都一定得先想明白透彻了：“我真的需要大数据吗？大数据真的能为我所驾驭吗？”

（安防知识网 2014-09-12）

## 中国聚焦：开放“大数据”为人民服务

随着群众路线被重申，中共正把西方兴起的“大数据”理论与经典的“为人民服务”宗旨结合起来，改善它的作风，转变政府职能。这样一来，对于诸如从事实时交通信息服务软件开发的王川久这样的人来说，大数据就意味着实实在在的红利。

“以前我们需要向其他地图企业购买数据，而现在北京市政府推出了大数据平台，公开了部分公交数据，我们就不用每种数据都花钱买了，数据源也更加权威。”身为私营企业北京九州联宇信息技术有限公司创始人兼总经理的王川久说。

他提到的数据平台是北京市经济和信息化委员会去年测试推出的北京市政务数据资源网。王川久的公司正试图借助这一平台开发公交出行应用，提供公交车到站信息查询服务。由于政府数据资源网公开了这些数据，他的公司一年节省了数百万元费用。

信息化不断提速对政府管理提出了新的挑战。“数据挖掘已成为重要的商业创新力量，但如今大量数据都为政府部门掌握甚至垄断。企业如果希望开发和利用数据，就离不开政府的支持。所以我们需要将不同的数据整理好，开放给社会利用，创造更多价值。”北京市经信委相关负责人表示。



这一数据资源平台试运行一年来，已上线公布了 400 余个数据包，涵盖旅游、教育、交通、医疗等门类。网站正面向社会征集地理信息和健康管理方面的应用，已有十几家企业主动联系经信委，希望利用并开发数据。

如今，在世界很多地方，政府数据公开已被视为经济与社会发展的动力。香港特别行政区政府的一站通网站公布了人口、水资源、空气污染、实时交通、食物与卫生等 12 大类、上百个动态更新的数据集。其中，交通信息被市场广泛利用，衍生出多个智能交通应用程序。正因为看好数据红利，包括北京在内的中国各地政府，纷纷探索“大数据”惠民之道。

浙江省交通运输厅今年宣布，将大数据引入交通管理，助力道路治堵；上海市科委发布了推进大数据研究与发展的三年行动计划，在医疗卫生、食品安全、智慧交通等领域探索交互共享、一体化的服务模式；广东则于去年底率先起草了实施大数据战略的工作方案。

然而，各地跃跃欲试的背后，问题也不容忽视。北京市经信委相关负责人坦言，建立大数据平台的过程中，一大困难就是有些部门不愿公开数据，导致数据封闭、“沉睡”，无法被社会开发利用。

王川久就被这一难题所困扰。他的公司希望获得完备的公共汽车行驶信息，但有关部门目前还未能提供。

历史形成的信息分散在各部门的“割据”局面也阻碍了数据的开发利用。此外，一些政府部门还常以“保密”“不宜公开”为由拒绝提供和分享信息。

对此，工业和信息化部信息化推进司副司长董宝青表示，部门的“信息垄断”必须打破。

他说：“数据开放不是一个技术问题，而是意识问题，要求政府有服务的意识和态度。”如今，一些政府部门不愿意公开数据，或是因为“权力”意识大于“服务”意识造成；或是由于“懒政”，认为

“多一事不如少一事”。

“政府不应该以各种理由敷衍各界要求的信息开放，而应该明确信息的公开范围，划清保密信息、商业公开信息、免费公开信息、义务公开信息之间的界限。”他说。

董宝青建议，为促进大数据惠民，国家层面应尽快出台相关的顶层设计方案，协调数据的开放与利用。

国务院 15 日发布了一份关于政府信息公开的意见。这对于大数据的开发利用无疑是一大利好消息。

国家信息化专家咨询委员会委员、国家行政学院教授汪玉凯指出，信息公开是电子政务的灵魂，更是信息时代政府职能转变的重要契机。“政府必须以服务大众的态度跟上时代的步伐、促进社会的全方位进步，而不能成为发展道路上的绊脚石。”

（新华网 2013-10-20）

## 2013 大数据元年

这一年，不管你有没有意识到，大数据都已经来到我们身边。

电商投放广告、物流调度运力、证监会抓老鼠仓、金融机构卖基金、民航节约成本、农民破解猪周期、制片人拍电影……看似毫不相关的事儿，背后都有大数据在发力。随着互联网、移动互联网对各个领域的渗透越来越深，从政府到企业，从群体到个人，数据的积累与日俱增。4G 牌照的发放，又让移动数据通道由“乡村公路”升级为“高速公路”。可以预见，大数据席卷各行各业和人们生活的速度只会越来越快。

业内人士所称道的“大数据元年”，带给我们哪些变化呢？

### “搅局” 金融创新

当传统银行业还在“躺着赚钱”、坐收高利润的时候，门口突然

来了一位“搅局者”——互联网金融。今年，余额宝诞生，微信 5.0 支付功能推出，百度百发理财上线……

以余额宝为代表的互联网金融产品在 2013 年刮起一股旋风，截至目前，规模超 1000 亿，用户近 3000 万。相比普通的货币基金，余额宝鲜明的特色在哪儿？当属大数据。以基金的申购、赎回预测为例，基于淘宝和支付宝的数据平台，可以及时把握申购、赎回变动信息。另外，利用历史数据的积累可把握客户的行为规律。

阿里小贷更是得益于大数据。它依托阿里巴巴（B2B）、淘宝、支付宝等平台数据，不仅有助于识别和分散风险，提供更有针对性、多样化的服务，而且批量化、流水化的作业使得交易成本大幅下降。

传统金融业也在积极拥抱大数据。工商银行正在建设基于线下 POS 机的网络信息数据平台等。平安集团联手百度，共同研究消费者网上的行为习惯，进行产品创新。

### 揪出老鼠仓

早在 2000 年，一篇《基金黑幕》，首次刺破基金的光鲜神话。黑幕的揭开源于数据分析，即 1999 年 8 月 9 日至 2000 年 4 月 28 日期间，国内 22 只基金在上海证券市场的大宗股票交易记录。当时，这项工作仅由一名上交所的研究员完成。

到了 2013 年，“大数据”监管进入公众视野。2013 年 9 月，证监会通报马乐利用非公开信息交易一案，涉及 3 个账户，交易股票 76 只，累计成交 10.5 亿余元，而深交所在监控时发现异常账户重仓的小盘股和马乐操作的账户高度重合。

“深交所有 200 多个异动指标，将投入力量把原形系统项目和股价异动联系起来，构建市场监控综合数据模型，全面提升对内幕交易、市场操纵、证券欺诈文本信息的挖掘和监管。”深交所总经理宋丽萍曾在今年两会期间表示。

此外，2013 年 11 月，国家统计局与阿里、百度等 11 家企业签

署了战略合作框架协议，推动大数据在政府统计中的应用。未来，大数据对于我国深化改革开放所面临的种种问题以及环境、交通、教育、医疗等方面的现实挑战，具有重要战略价值。

### 破解猪周期

一会儿猪肉太便宜了，农民权益得不到保证，一会儿猪价太高了，CPI 上涨百姓生活受到影响。如何才能解决猪周期？关键就要解决农民和市场信息不对称的难题。

新希望集团董事长刘永好提出，我们要把历年的数据集中起来，建立一个动态的养殖、生产和市场的体系。通过大数据和云计算进行猪周期的预测，会发现猪的价格波动周期有一定的规律，大概 3 到 5 年是一个完整的周期，少的时候 2 年多，多的时候 5 年多，而这个周期又受国家的政策变化、天气变化、传染病变化、农民收入变化、原料价格变化等多重因素影响，同时又跟人们的生活水准和购买力有关系。全国所有养猪的农户，如果通过云计算、大数据对庞大的数据进行研究、分析、判断，研究出一个模型，建立信息系统，养猪会变得更加科学化。

其实，农业大数据可以渗透到耕地、播种、施肥、杀虫、收割、存储、育种、销售等各环节，是跨行业、跨专业、跨业务的数据分析与挖掘。

### 赚钱“读心术”

当一个穿着时尚的女士步入商场，售货员可以根据她的穿着打扮，推荐款式。而在网购中，买卖双方彼此“双盲”，从而很难有针对性的服务。

但是，近来这一局面正在改变：网友小甜和闺蜜同时进入同一家电商的网站，但显示的页面居然不一样：小甜被推荐了母婴用品专场，而闺蜜则被视为吃货一枚。

淘宝网有个有趣的功能——购物车自动巡航。这购物车居然是

“活”的，还能猜中主人的心思。一点击自动巡航键，购物车将在全网范围内自动搜罗一批适合用户的商品，并呈现出来。称它懂“读心术”，并不过分。

这些购物预测是怎么做到的？数据是一切的答案。淘宝能够根据每个用户的收藏夹、购物种类、搜索记录、浏览记录，相似推荐等等数据，为每一位消费者量身定制自己的欲望清单，甚至，跟你买过同类型产品的人还在买什么别的，这些信息也会成为预测的一个根据。

前不久，淘宝免费向卖家开放购物车里的数据。卖家在看到有买家把商品放入购物车以后，可以对消费者直接喊话打折。据统计，“购物车营销”促使购物车内商品成交率增长了 2.5 倍。

### 扫描物流节点

2013 年 11 月 11 日下午，杭州。淘宝卖家晨妈收到了快递公司的短信提醒：由于所在区域发货量巨大，建议错峰发货。晨妈决定第二天再发货，结果反而发得更快。

同一日，北京。申通快递员顾提军得知第二天他要派送的快递量达到平时的 3 倍。“不吃不喝跑断腿也送不完啊！”顾提军赶紧找了个朋友帮忙一起送，才算没误事。

在沈阳，消费者陈雅抢购了件跨省发货的衣服，3 天后就到手了。“跟平时差不太多，比去年‘双 11’明显快。”

“双 11”期间，天猫销售额超过 350 亿元。快递量陡增却避免大面积延误的背后，是各大快递公司合作平台的数据雷达实时监控起了作用。11 月 13 日，菜鸟网络的数据雷达监控到圆通快递杭州分拨中心的快件状态 24 小时没有变化。指挥部判断该节点负荷度超高，立刻报告圆通紧急调配资源，压力洪峰迅速得以缓解。

“结合历史数据，根据今年的商家名单、备货量等信息进行分析，我们对‘双 11’订单量作出预测，精确到每个区域、网点，信息与快递公司共享，这样运力布局的调整更加精准。”菜鸟网络副总裁谭

飙介绍。

### 智能化机器

如何让飞行更安全、更具效率，甚至更省钱？航空业长期以来都在追求这个目标。

春秋航空曾轻轻松松省下了可能花掉的 21 万美元。这还要从 GE（美国通用电气公司）在飞机中置入的传感器说起。春秋航空公司飞机上的这些传感器传回的数据显示，一架飞机的二号发动机风扇振动值突然升高，与历史数据比对差异较大，GE 在上海的航空客户支援中心立即向春秋航空发送通知，经排查后发现发动机的两片风扇叶片被外物击伤，一次大故障就这样被避免了，大数据系统可以自己判断哪些需要检修，节约了大范围人工检修的高昂成本。

2013 年，GE 公司负责人表示，未来三年将花费 15 亿美元到“工业互联网”中。

互联网、大数据不仅能让飞机变得更聪明，甚至能让所有的机器拥有智慧。当工业革命带来的机器世界与互联网革命带来的数字世界更深层次地融合时，“工业互联网”的理念应运而生。互联网将人和所有机器连接起来，建立机械王国的神经系统、指挥系统，让机器更好地为人类服务。

### 整合收视点

《纸牌屋》不仅在全美热播，还红到了太平洋彼岸的中国，但这并没有让制作方奈飞公司感到意外，因为该剧导演在喊“开机”之前就知道它一定会火，秘密在于大数据。

奈飞数据科学家分析了用户每天超过 3000 万条播放记录，包括用户在何时、何地、何种设备上观看什么内容，并在记录暂停、倒退、快进、评分、搜索的同时，进行大量截图，试图分析用户在音量、画面色彩甚至场景选取上的喜好。

奈飞公司分析了庞大的播放数据后发现，有相当数量的用户喜欢

《社交网络》导演大卫·芬奇的作品，而著名影星凯文·斯佩西的电影和英国版《纸牌屋》都是点击率比较高的节目。因此奈飞公司决定把这些收视吸金点结合到一起，于是有了美剧《纸牌屋》。

这种方式也得到了中国同行的借鉴，2013年，爱奇艺公司通过对大数据分析开始筹划将一个播放量特别大的商业微电影拍成电影，进而了解大数据指导下的大电影是否能够像预期的那样受欢迎。

（《人民日报》 2013-12-25）

## 阿里助广西建大数据基础设施未来辐射东盟

2014年4月21日，广西壮族自治区人民政府与阿里巴巴集团签订战略合作协议，在云计算和大数据等领域广泛开展合作，共同打造数字互联网广西。阿里巴巴将帮助广西建设大数据时代的基础设施，广西也将成为阿里巴巴未来面向东盟发展的前沿阵地。

阿里巴巴集团董事局主席马云表示：阿里巴巴将为广西提供云计算、智能物流、无线应用、电子商务、创新金融等方面的基础设施服务，助力广西向新经济转型。

云计算方面的合作，将成为大数据基础设施建设的重点。阿里巴巴将利用核心技术，为广西电子政务、城市管理、民生服务、商贸服务等领域提供统一的数字化服务平台，立足广西、影响西南、辐射东盟。广西政府也将推广云计算的应用，推动政务、电商、旅游、交通、医疗、教育等领域的智能化、无线化。

此外，阿里巴巴将在广西打造覆盖全自治区、影响西南和辐射东盟的智能物流骨干网络，以大数据驱动现代物流发展；在广西南宁打造线上线下融合的银泰 O2O 体验中心，提供大数据时代的新型消费业态；助推广西特色农产品的线上销售，推动当地特色农产品的信息化、标准化、规模化发展。

马云表示，广西是中国面向东盟的门户，阿里巴巴与广西开展合作，不仅希望为广西提供云计算和大数据的服务，也希望能够面向东盟市场。阿里巴巴将会把广西作为面向东盟发展的战略地点。

（华龙网 2014-04-22）

## 上海用大数据缓解“城市病”

如今，人们在享受城市化带来的好处时，也被“城市病”——“看病难”、“行路难”等所困扰，尤其在北京、上海、广州等特大城市，类似挑战更为明显。如何破解这些难题呢，大数据技术为缓解“城市病”提供了一剂良方。

### 让城市生活更方便

对于缓解“出行难”，大数据技术有用武之地。上海市交通信息中心建成交通综合信息平台后，汇集了 259 项各类交通数据，为大数据分析提供海量数据基础。在世博会期间，上海对客流量预测的平均准确率超过 90%，世博会之后，交通预测机制也进入常态化。

挂号排起长队、住院一床难求，这让大城市“看病难”饱受诟病。数据显示，上海市医院门诊量每年大幅上升，10 余家三甲医院日均门诊量已破万。

为了减少不必要的重复检查，上海申康医院发展中心用上了大数据。该中心不仅为就诊人员建立健康档案，而且档案能在市区各级医院间共享。就算你去不同医院就诊，医生也能看到你之前的检查记录。这样一来，就避免了重复检查。全市医院“一盘棋”，患者和医生都节省了大量时间。

至于人们极为关心的食品安全，大数据也能助上一臂之力。上海市食药监局结合大数据、物联网等新兴技术，建立了多个监管应用系统。其中，基于移动技术的“餐饮电子化监管”系统，积累了 62 万



余条现场实时执法记录；视频溯源体系覆盖企业已达 2000 多家，每日上报业务数据 6 万多条，从食品的生产到销售，基本做到全流程有监管，有问题可追溯。

### 让城市治理更智能

联合国 7 月发布的一份报告显示，在全球人口最多的城市中，上海排名第三，仅次于东京和新德里。治理大城市，如果没有大智慧，会举步维艰。

“建立一种相互信任和社会监控的机制”，被视为城市治理的关键。那么，如何让信用成为城市的“硬通货”呢？

得益于大数据技术，上海建立了“公共信用信息服务平台”。目前，该平台已基本覆盖上海 138 万家法人单位，以及两千多万常住人口。信用好的人，不交押金就能在图书馆借书，而坐车或地铁逃票、助学贷款不还、欠缴水电煤费用等行为，则会被纳入信用系统，甚至会影响日后向银行贷款。

为保障城市安全运行，上海用大数据构筑“天罗地网”。目前，上海市公安局信息中心已汇集 250 亿条信息，日均数据增量近 5000 万条，包括视频监控、电子警察、治安卡口等系统产生的各类信息。在此基础上，公安部门利用大数据技术，与有关政府部门构建“大联动、大联勤”执法管理平台。

大数据应用带来的产业经济分析不容小觑。比如，用电数据是区域经济冷暖直观的“温度表”，中国（上海）自由贸易试验区设立一年来，国家电网上海浦东供电公司最新数据分析表明，随着新增企业不断入驻，电力数据稳健攀升的背后是投资热潮兴起。与此同时，金融、商务、服务业等新兴产业用电量快速上升，交通、仓储等传统产业用电量略降，数据升降间反映自贸区经济结构的深刻变化。

上海浦东供电公司总经理刘运龙认为，电力数据能直接反映自贸区发展情况，大数据分析实际上为城市产业升级提供了重要

决策参考。

### 让政府决策更科学

在大数据的帮助下，政府可以更全面掌握情况，决策前真正做到“心中有数”。

上海建起“居民经济状况核对信息系统”，对于廉租房等申请家庭，通过对申请对象信息的大数据分析，只要有蛛丝马迹，就能将不符合条件者拒之门外。

在掌握经济信息方面，大数据有不可替代的优势。例如，上海统计局建立“1号店快速消费品价格指数”，从而掌握日常消费品的价格变动，为制定社会救助和保障标准等提供数据支撑。

在制定产业政策方面，大数据的重要性十分明显。上海市财政局通过汇总各行业税收收入、中国人民银行的国库收入数据、市统计局的主要经济指标数据等，摸清产业发展情况，进而调整产业政策和财政扶持重点，从实打实的数据出发，让决策更加科学。

（新华网 2014-10-02）

## 大数据应用的环保案例

### 追求可持续，数据驱动的绿色城市

近年来，拥有陆地面积 1.6 平方公里、居民 1.8 万人的哈姆滨湖城，成为斯德哥尔摩最大的近郊发展项目，其目标是打造成为未来城市发展的标志和典范。

哈姆滨湖城信息中心负责人玛琳娜·卡尔松表示：“如何使用有限的城市资源和能源来满足城市居民的需求，同时又保证资源可持续以及循环利用才是智慧城市最大的挑战。”

在哈姆滨湖城，能看见一排电子垃圾桶，分别用于接收食物垃圾、可燃物垃圾以及废旧报纸等不同类别的垃圾。垃圾桶通过各自的阀门

与同一条地下管道相连，阀门分别在每天自动打开两次，不同类别的垃圾进入地下管道，并以每小时 70 公里的速度被输送到远郊，在电脑的控制下自动分离并输送到不同的容器里，按需要循环利用。整个过程都是通过电脑控制，这个系统提高了垃圾传输和处理速度，以及再利用效率，环境保护程度相应提高了。”

由于斯德哥尔摩特殊的地理环境，很多地区被湖水或海水隔断，哈姆滨湖城这样的近郊地区很容易形成“孤岛”，因此与市中心的交通衔接便十分重要。通过网络提供实时路况信息和出行路线规划，每天大约有 7 万人骑车穿越斯德哥尔摩市区，让人们选择最便捷、最环保、最舒适的出行路线。市政府在交通信号设计上，遵循了自行车优先原则，其次是公交车。在斯德哥尔摩，30%的城市居民选择走路或骑自行车上下班，61%的人乘坐公共交通工具。2.5 万辆汽车为绿色车辆，采用新能源的比例达到 26%。

“公共交通设施是衡量智慧城市的一个重要指标，斯德哥尔摩的公共交通系统已经实现了智慧化，可以随时随地用手机查阅交通工具到达时间，也可以通过短信来买票。”瑞典信息技术研究中心高级科学家马库斯·比隆德说，“越来越多的人选择了公共交通，最终有益于城市环境的可持续发展。”

### **虚拟河流 ， 智慧的水污染治理**

大数据应用在环境保护有两个亮点，首先是 360 天×24 小时的不间断环境变化监测，其次是基于可视化方法的环境数据分析结果和治理模型的立体化展现。通过虚拟的数据，模型可以模拟出真实的环境，进而测试制定的环境保护方案是否有效，这种极具创意的环境治理方式已经在多个国家得到应用。

纽约曼哈顿有一条哈德森河(Hudson River)，北起阿迪龙达克山区，绵延 500 公里南下，入海口在纽约港。在过去 20 年里，居民造成的下水道污物的沉积，以及近代大型工厂倾倒的有毒化学物质，致

使这条生态系统敏感的河流受到严重污染。通用公司的两家工厂还曾将含有多氯联苯的工业污水，直接排放到哈德森河里。多处河段不能作为饮用水水源，渔业年产量锐减 60%。

上世纪 80 年代，环保主义热潮涌起，为了保持、恢复哈德森河的生态系统，纽约州政府发起了一个“新一代的水资源管理计划”。他们在河流的全程都安装了传感器，一些传感器甚至高达两米。这些传感器把水的不同层面、各种物理、化学、生物数据包括河流中的盐度、浊度、叶绿素和颗粒物粒径等信息，实时地通过网络传递到后台的计算中心区，在水面之上的传感器则负责收集河流的风向和风压数据。数据像流水一样不间断地生成，不间断地被处理，并与历史数据进行比对。

后台的计算中心区分为 3 个环节，首先是数据传输环节，传感器将从河中与周边环境收集到的数据以实时连续的方式传送给系统管理层，在接下来的这一环节，关于河流的不同类型数据将被清洗，后台通过消除数据的异源性，使关于哈德森的数据一致化，并具有互通性，然后在分析管理平台对这些数据进行可视化的展现，在科学家的电脑显示屏上，各种数据汇成了一条虚拟的哈德森河，流水何时被污染，化学、物理、生物成分发生了什么变化，一看便知。接下来数据科学家便可利用这些处理过的信息模拟一个哈德森河的环境模型和治理方案，评估不同的治理和人为干预对于哈德森环境的多方影响，以保证在实际治理时的效率和效果。

经过多年努力，哈德森河流已逐渐恢复其清澈的水质和优美的环境。现在每年的父亲节周末，都有持续两天的环保音乐节——清水节 (Clearwater Festival) 来庆祝哈德森河的重生。

### **污染地图，向公众发出邀请**

2006 年，马军创立公众与环境研究中心 (IPE)，主持开发了“中国水污染地图”、“中国空气污染地图”和“固废污染地图”，建立了

国内首个公益性的水污染和空气污染数据库，将环境污染情况以直观、简单、易懂的图表进行展现，在中国公众环境研究中心的网站上，能访问到醒目的中国空气污染地图、水资源污染地图、固废污染地图。地图采用 Flash 技术生成。点击地图后，可以访问每个地区的环境质量信息、地区排放信息等。通过这个公益数据库，任何一个用户都可以进入全国 31 个省级行政区和超过 300 家地市级行政区的相应页面，检索当地的水质信息、污染排放信息和污染源信息，包括超标排放企业和污水处理厂信息。

这是一个新的尝试，它本身实际也是一个鼓励公众参与的过程。企业最终打开了它的大门，以往企业都是封闭在四面墙里面，好像与周围没有什么关系。但实际上企业的排放对周边的社区和环境都会有所影响，应该更多与社区公众进行交流。通过第三方审核，企业打开大门，将他们生产过程当中形成的政府环境监管记录，各种环境管理记录，包括环境管理体系的运行情况，和当地的环保组织有所交流，并通过报告展示给公众。

（《中国环境报》2014-09-10）

## 发达国家推行大数据战略的主要做法

### 一、站位高，合力推进

发达国家已充分认识到大数据的战略价值，将其置于国家战略高度加以推动。美国认为其重要性堪比“信息高速公路”计划。2012年3月29日，美国发布了《大数据研究和发展计划》，同时组建“大数据高级指导小组”，涉及美国国家科学基金、国家卫生研究院、能源部、国防部等6个联邦政府部门，政府还倡议企业、科研院校和非盈利机构集中资源，共同促进大数据发展，在国家战略层面形成了全体动员格局。欧盟认为，大数据是创新工具、创新资料，其开放数据

战略从欧盟全局的高度要求成员国共同推动大数据的发展，协调成员国与欧盟的行动战略，沿着修改补充法律、资金投入、建立公共信息平台等方面推进。新加坡政府认为大数据是“未来流通的货币”，政府各部门合力从基础设施、产业链、人才、技术和立法五大方面系统推进大数据建设，弥补了企业发展大数据的不足，极大促进了大数据的发展和应用。

## 二、突出重点，巨额投入

政府有选择地对大数据进行重点投入。美国是大数据的领跑者，政府投入 2 亿美元重点资助大数据分析以及大数据在医疗、天气和国防等领域的应用。法国政府以新兴企业、软件制造商、工程师、信息系统设计师等为目标，投入 1150 万欧元用于支持 7 个投资项目。英国在经济低迷、财政吃紧的背景下，政府注资 6 亿英镑支持发展 8 类高新技术，其中大数据技术拔得头筹，获得 1.89 亿英镑的资金，其中节能计算是其重要的支持对象。联合国“全球脉动”战略，目标是促进全球经济增长，重点是预测某些地区的失业率、支出削减或是疾病爆发等现象，利用数字化的早期预警信号来提前指导援助项目。

## 三、整合数据，加强公共基础平台建设

政府适度公开、有效整合数据，建立公共基础平台成为了大数据战略发展的重要内容。2009 年，美国 Data.gov 正式上线，按原始、地理数据和数据工具三个门类开放数据。截至 2012 年 11 月，Data.gov 共开放 388529 项原始数据和地理数据。为进一步完善 Data.gov 平台的功能，加入了数据的分级评定、高级搜索、用户交流以及和社交网站互动等新功能。同时，美国还构建了 OGPL 平台，提供开源的政府平台代码并允许任何城市、组织或者政府机构创建开放站点。英国政府发布新的政府数字化战略，旨在使政府服务实现“默认数字化”，即“数字服务简单方便，任何可以使用的用户都会选择数字化服务，而不能使用的用户也不排除在外”。英国承诺 2015 年前开放有关交通

运输、天气和健康方面的核心公共数据库，并将投资 1000 万英镑建立世界上首个“开放数据研究所”。“欧盟开放数据战略”将重点加强在数据处理技术、数据门户网站和科研数据基础设施三方面的投入，旨在欧洲企业与市民能自由获取欧盟公共管理部门的所有信息，建立一个汇集不同成员国以及欧洲机构数据的“泛欧门户”。

#### 四、应用拉动，加快发展

政府高度重视、积极推动大数据的应用，创造需求，加快了大数据产业化和市场化进程。美国最具代表性，大数据政府、民间应用广泛，产业化进程快。政府在公共政策、舆情监控、犯罪预测、反恐等领域已开始依据大数据分析辅助决策。以大数据应用增强社会服务能力。人口、交通、医疗等公共事业部门通过大数据的挖掘，实现了对人口流动、交通拥堵、传染病蔓延等情况的实时分析。如佛罗里达州迈阿密戴德县将数十种关键县政工作和迈阿密市紧密联系起来，帮助政府在制定治理水资源、减少交通拥堵和提升公共安全等方面决策时提供了更好的信息支撑。国防部确定了“从数据到决策、网络科技、电子战与电子防护、工程化弹性系统、大规模杀伤性武器防御、自主系统和人机互动”等 7 个重点研究领域，目的是推进大数据辅助决策，实现由数据优势向决策优势的转化。

（《中国经济时报》2014-03-20）