

主持人：谢谢我们吴总的致辞。我们也希望他们虽然已经到了退休年轻，依然持续关注我们行业的发展。我们再一次以热烈的掌声感谢这些前辈对我们CMRA的贡献！接下来就到了我们峰会正式的内容。第一个环节就是政府数据需求与第三方评估的环节。有请第一位嘉宾是国家统计局统计科学研究所姜澍处长，他带给我们的题目是《数据创新在政府统计中的应用》，大家掌声有请！

姜澍：大家上午好！我是来自国家统计局统计科学研究所，主要是做研究和统计科学研究的。今天很高兴能来参加这个会议。这个会议举办的非常恰逢其时，非常高质量。无论从会议的选题、会议的组织、筹办以及选址等等都非常好。非常高兴能来发言。我这次来感觉我们协会有历史有温度又团结，聚焦前沿，一直在思考和前进。对我来说感受了新理念新朋友，再一次表示感谢！

我今天是来自政府统计部门，从政府统计的角度来说一说政府统计在数据创新的想法和思路，看看是否有合作的机会共同推进政府的统计工作。

中国政府统计工作的现状。

主要的工作统计政府统计现在主要的统计工作有哪些？包括统计调查方法、普查、抽样调查和全面定期统计报表等统计调查等方法。以及统计采集方式联网直报、电子采价器、住户

记账器等电子终端、计算机辅助电话调查的CATI系统以及统计数据发布。就是各行各业的统计调查、收集、汇总、整理以及提供调查的数据。再有，我们要组织实施能源、投资、消费、价格、科技、劳动力、社会发展这些基本情况。就是我们常规的这些传统的社会经济的指标也是需要统计局来做的，这是大家关注这些信息。统计数据的发布，分析功能以前也是经常说，但是这两年加重了，特别是对于重点问题的分析，中等收入以及专题性的分析。周期性的普查是基础，我们常做的是经常性的抽样调查是主体，从统计的调查方法上来说，大家也无外乎是这样的，现状就是这样的。而统计的采集方式现在这几年，从2008年、2009年，我们的工程调查里面2008、2009年开始推CAI工程，大家可以搜一下政府统计的数据。从那个时候开始我们更新我们数据采集的手段和方法。有部分的省份可以直接向数据中心报送数据。还有一个是根据国家统计局审计的数据节点来报送数据。然后在这个报送节点上进行取数，是这样的。就是采用电脑化的终端，这样过程控制比较容易一些，人为的一些数据的干扰就得以避免。然后再有就是在社情民意我们广泛采用计算机辅助调查的CATI系统，这及从互联网发展以及信息化的方面做了很多的更新。以前主要是通过布置报表以及手中填报等等这些，现在是用网络的设备。

政府统计它的流程来说，它标准化流程，这个是通过我们设计管理以及政府的统计管理司确定的7个流程制度设计、数据

采集、数据处理、数据存储、数据质量评估、数据发布与传
统、数据分析，一个就是制度设计，因为我们通常如果无论是何种何样的调查，电报还是抽样调查还是什么样的调查，我们首先要有制度、有标准，就是说要设计好这个报表制度，内容比如说有说明报表的目录、表示这些都很细，还有指标的体系等等这些都要通过制度来研究和顶层设计。第二部分就是数据采集，采集的话，就是向相关的用户进行数据采集，然后再是数据处理，数据处理是对采集上来的数据进行查询、较验、汇总、整理、验证，这些都属于数据处理，我们政务统计更多的是通过汇总上报的形式，就是从目前的统计工作来说。数据存储就拖用说了，还有一个很数据质量的评估，要准确及时可比一致以及适用来获得经济性，这些都要进行分析。然后最后是数据发布，找不到有一些数据，目前也是在不断的完善这个法律制度，比如说最近开放的微观数据实验室，大家做研究或者是调查，大家找不到一些数据，但是这是一个记录我们开放一部分的微观数据。然后向市民、公众提供。除此之外，我们也有各种制度，比如说像主要的精准的社会经济指标，大家都可以在网站上可以看到它会哪一天发布。最后还有一个数据分析，像我刚才说的现在比如说像新经济的发展，像一些专题性的，中等收入这些中央领导、政府决策比较关心的话题，统计局也会做相关的调查，除了平时的调查之外也会做一些分析。供决策来用简单回顾了一下政府统计的职能。

我们来自各行各业，我们从政府统计这方面来说，我们聚焦了很多方面，大数据是政府统计数据创新的途径之一。比如说三新的调查制度，新产业、新模式、新方式的调查，最重要的一个特征就是大数据，希望把大数据作为创新的途径之一补充完善现有的调查制度的体系和指标。大数据从特征开始，政府统计已经关注到这种，因为当时的领导高瞻远瞩，希望我们能够研究这个大数据应用，待会儿我也会介绍一下我们的一些想法。我来这个会议之前，刚从卢旺达参加了联合国大数据全球工作组的会议，刚刚从那个会议上回来，现在联合国大数据峰会，因为大数据有很多商业等等各个方面，因为这个工作组本身是政府的大数据是联合国统计署，大数据如何在政府应用。现在重点分了几个组，最关注的就是手机的数据，手机主要就是通过各个国家能够给出一些尝试，包括人口有统计，这个大家应该好理解，然后还有第二类的数据就是卫星图象和地理空间数据，我们通过会议也发现很多欧洲的公司，具体它的方式我不是很理解。但是它的公司的总营业额就是通过到非洲去采集卫星图象和地理空间的数据，用非洲的数据来做农业的统计，他们常年有研究员在非洲驻扎，来做农业的分析。第三个就是扫描数据，我们可以用它来做价格的统计。再有就是像网络社交媒体数据，这次会议有介绍利用网络社交媒体来做情绪分值和指数来分析这方面，这个会议还设了一个宣言大家可以通过治疗来找到，所以联合国作为政府统计工作的高层领导

部门，他关注的数据在未来有很大的应用空间。而且这个会议大家可以看到它是第五届了，前几届大家不断摸索和聚焦最后形成数据的需求和方法的探索也是我们未来中国统计的应用主流方向，随着方法的成熟以及先进国家经验的分享，以后在中国政府统计上也是一个很大的一个路径。

联合国全球平台，他还建了一个UN Global Platform，这集政府统计、企业、学术界Work together learn together，在这个平台上可以申请一个帐户，由某一个授权方来授权大家组成一个工作组，比如说研究手机的使用这些方面来做的，进行分享数据、方法和技术服务。大家如果有兴趣可以在这个平台上国家统计局的授权大家可以一起来做这个事情，这是一个很好的平台。

我们政府统计2012年就关注到大数据，也定了一些原则和工作的目标。在整个统计局在推进的时候，还是有方法和工作的，主要的目标任务主要是构建大数据统计标准和方法体系。大数据的杂乱无章，比如说语言、文本这些非结构化的数据，这些是重要的任务，还有就是扩大政府专业的统计的数据源以及改革统计调查方法和数据采集手段，然后挖掘数据的水平，比如说可视化。因为大家来自各行各业。我说一些适合政府统计的收集。

比如说联合国基于Twitter数据对价格的调查和分析。这是用马来西亚的数据进行分析的，会发现他们之间有关联性的，

可以用Twitter上的数据进行价格预测和分析。然后每日网上价格指数来获取这些价格信息来编制价格指数，然后预测美国经济危机的发生。然后英国、荷兰、挪用这些国家利用网络抓取数据编制CPI，这个数据是英国的食品类的CPI和它的网上的价格一个比对，大家可以看到。英国还利用大数据来测算数字经济，这个也是一个很厉害的。随着我们经济社会发展，政府统计有时候很难跟上社会发展的形式来补充相关的指标，大数据是一个很好的数据来源，可以弥补这方面的空白。数字经济就是一个，英国利用了大数据来分析一下数字经济公司的数量和非数据公司的差异，最后发现数字经济公司的质量增长的速度并没有非数字经济的快。然后欧盟它利用谷歌搜索数据预测失业，他加入了谷歌的搜索数据以后，提高模型，关于就业、失业预测的精准度，提高了。就是在政府统计拿不出支撑新经济发展的指标发言数据的情况下，财新智库基于大量的网络搜索数据建立了新经济指数，这也是一个很好的例子。还有阿里巴巴网购全网商品价格的指数。还有中科院做的基于百度搜索数据的消费者信心指数，这个跟我们统计局相关的数据也做了一些比对。然后北京市统计局应该有搜索首都和非首都的政策的要求，所以他有一部分的资金来专门做这样的项目，就跟北京移动合作来进行人口统计，这个是利用手机移动信号来进行人口统计，这个是北京统计局和北京移动合作，现在联通也加入了，利用这个手机信号我这个非功能区我人口的疏解效果，大

家可以通过手机信号量的多少就可以观测出来。北京这样大的城市，每天要花很多的时间上班，在东城住，在西城上班还是指数分离是怎样的情况？这也是可以分析出来的。还有就是简单手机信号的量跟各区域的人口数量跟我们人口数据的资料进行对比，可以发现他们在数量上不是绝对的量，但是跟我们的调查来看，从各区的排名上，第一大人口区域是哪里？第二大人口多的区域在哪里？排名上是样。所以手机信号还是很有前景的，大家可以看到它实际上做这个非常有意义，但是背后也有很多的资金做支持的。还有应用源，双泉堡等等这些低端人口可能会住人多一点。

我们科研所也做了一些分析，比如说经济雷达，主要是利用百度搜索数据来做预测，做GDP预测。当时跟百度公司合作，然后我们来做统计分析预测当前的GDP，问题肯定会有，但是像预测的情况比较短期以及满足需求等等，预测本身这个事情就受很多政策和环境各方面的影响，效果我们只能说预测，但是可以通过分析做到。我们当时找了一段时间，2011年第一季度到2015年二季度期间，请百度公司那段时间记录的数据然后我们再进行分析。在我们局里面人口司做了新经济就业的识别和稳定性分析。也是利用大数据来分析新经济的的企业以及分析他就业人口的特征，比如说人员的分布主要在哪些方面，这个年龄是多大的？因为新经济就业的人普遍都是比较年轻，学历怎么样？稳定性如何。

还有就是就业景气程度这一块。从就业网站的数据，就业网站相对比较直观，招聘网站可以建立管理的指数，跟我们自己的调查失业率再做一些验证。做做一些预测。这些我们做了一部分。然后也是我们之前看到的一些问题。所以政府统计关注的大家可以看到都是跟国民经济和社会发展相关的这些部分的指标情况。要么可以利用大数据来提高我们的统计效率，在你正式的指标没有出来之前，我们可以用大数据做短平快的分析以及预测，可以提高决策的时间。因为政府做决策，从决策的使用到效果的收回都有一定的时间，大数据可以提高你的效率，缩短你的时间，在正式数据之前我认为是这样的，这是一个需求。第二就是政府统计没有覆盖的方面，三新的调查以及新经济的识别等等这些方面没有标准，没有标准的情况下大数据是一个很好的补充。我们也希望能够做到这些方面。

我们各专业有设想，比如说贸经司可以进行一些贸经统计，农业部门我们也做了一些尝试，而且农业统计是国家统计局在大数据统计方面做的比较深的一些项目，做的有一些成功的试样电话。交通运输的统计，利用公路和航运来做分析，还有投资统计，投资可以跟一些网络的数据或者是跟一些重型机器的制造也可以做一些开工率的预测，他有传感性，他可以知道自己的机器卖出去以后开了多少？还有人口普查以及流动人口，我们自动化城市管理以及精准化管理。还有价格，除了网络价格，还有CPI统计等等这些都可以，还有房地产价格，生产

资料价格，像与山东卓创开展合作，可以看到企业的资讯公司在价格咨询上的情况。虽然大家调查的范围和渠道以及时间并不一样。但是它是一个资源和来源。

大数据可靠吗？经常我们会说，特别是对于政府统计需要真实、客观，用数据做决策的角度来说，真的不能完全对他的分析结果非常满意。因为大数据的计量和计算方法是完全一样的，不能完全取代政府统计，但是我们可以利用它来做一些分析。政府统计对它来说非常有兴趣而且愿意去尝试，但是也需要不断的去谨慎的对待分析它的结果去验证的。显而易见大数据是最容易计量的，从我们的角度来看，它非常有价值，有待分析，但是我们前面所讲的一系列，一个是我们政府统计的关注点，还有就是未来我觉得政府统计可以跟各位做大数据公司可以做的点。我们认为大数据不能取代政府统计，但是可以补充和验证政府统计，政府统计本身除了应用大数据还应该更加关注大数据标准的制定，与多方合作，包括企业、高校和学术界为政府统计做好相关方面的工作。

最后我想总结一下，政府统计数据创新和跟大家合作以及我们自己去这个事我们目前现在最大的三个困惑。一个就是数据的可获得性不高，北京市统计局对于人口的统计这个要花几千万，这个对于政府来说是很大的，如果大数据跟BAT等公司合作，如果要购买服务方式不堪重负。还有就是难以接触数据源的技术能力有限，缺少相应的人才，政府统计也需要创新需要加强与各方的合作。谢谢大家！