

数据资源与哲学、社会科学研究

周世禄

社科院 计算机网络中心

【摘要】信息化是当今世界经济和社会发展的的大趋势，信息技术正在以前所未有的速度广泛应用于社会经济生活的各个领域，各行各业都在充分运用信息技术创新工作，提升工作能力和水平。本文根据社科院网络信息化建设十余年的实践经验，从更广阔的角度观察信息化工作，就信息化建设中数据的获取、组织、共享，数据的保值、增值，数据的使用和管理等问题，提出了个人观点；对如何将信息化建设融入到社会科学研究，信息化建设如何支持社会科学研究，e-社会科学研究等进行了初步的探讨，对我院今后信息化建设的发展和完善工作，提出了个人的意见。

【关键词】信息化、数据、e-社会科学

前 言

信息化是当今世界经济、社会、文化发展的大趋势，各行各业都在充分运用信息技术创新工作，提升工作能力和工作水平，信息技术正在以前所未有的速度广泛应用于社会政治、经济、文化生活的各个领域，正在和必将产生出对社会进步强大的杠杆推动作用，信息化社会将会以一种新的社会发展阶段出现在人类历史的长河中。

以全院网络成功开通为标志的社科院网络、信息化建设已经走过了十二个年头，十二年来「在院党组的正确领导下，在院网络管理部门的积极努力下，在全

院同志的大力支持下，我院信息化建设克服困难，稳步推进，实现了从无到有、从奠基初创到快速发展的历史性跨越，取得了令全院同志瞩目和全国哲学社会科学界肯定的成就」（引自院信息化建设十周年总结大会上陈佳贵副院长的讲话）。十二年来在「网络是基础，信息是生命，应用是目的，管理是保证，人才是关键」的信息化发展策略指导下，我院**信息化基础设施建设不断完善；信息开发成果显著；网站建设特色鲜明；办公自动化稳步推进；对外信息服务和学术交流逐步展开；管理和安全工作不断加强**。我院的信息化工作已经成为我院科研、管理不可或缺的支撑平台。十二年的信息化建设，全院完成了从「要我建设」到「我要建设」的转变；信息化建设和科研、管理工作相互交织、相互促进、共同发展已经成为共识；信息开发、信息应用自觉的纳入了各单位「一把手」工作的重点，各单位出现了抢着建「名网」，争着建数据库的热情；院信息化建设的工作侧重也适时地从基础设施建设转变到了数据内容建设和成果展示、信息应用、信息服务、信息安全保障上来。在这种形势下，在制定我院信息化发展「十二五」规划的时候，让我们坐下来静静地回顾一下过去，借鉴一些前人在信息化建设中的经验；发挥后发优势少走些弯路，规划组织好自己的数据建设，做好数据服务，提升数据的使用价值十分必要。

一、 社科研究中的数据支撑

我们说高水平的社会科学研究学者要有敏锐的洞察能力，正确的分析能力，迅速的决断能力，这一切都来自科学的数据支持，来自于对数据准确的把握和积累。可以说真实、完整、准确、及时的科研数据是社会科学研究的基础和命脉。

（一）信息技术带来数据建设新需求

当今的世界是一个信息化的世界，高速发展的信息技术对科研数据建设提出了新的内涵，「一支钢笔一张纸，数据全在脑子里」的社科研究模式已经一去不复返了，信息技术在为科学研究插上了飞翔的翅膀的同时，也对数据资源建设提出了新的要求和挑战。

IDC 在《数字宇宙膨胀：到 2010 年全球信息增长预测》研究报告中指出，

2006 年全球每年制造、复制出的数字信息量共计 1,610 亿 GB，中国数字信息量为 127.1 亿 GB，占全球信息量的 7.9%；受「富媒体」、用户创建内容和 16 亿网民三大因素推动，到 2010 年，全球数字信息量预计为 9,880 亿 GB，而中国的数字信息量预计为 900.5 亿多 GB，占到全球信息量的 9.1%。

毫无疑问，16 亿网民是个保守的估计，根据中国互联网最新统计报告中公布的数字，截止到 2010 年 6 月中国网民数量已经达到了 4.2 亿人。随着信息技术的发展，网络客文化的发展，SNS 网络模式的发展，网络信息量将会以更高的速度膨胀。高速膨胀的信息环境，既改变了人们的认知方式和认知习惯，极大的扩充了我们了解真实社会的视野，同时也增加了对信息资源获取、整合、处理的难度和成本。信息、数据的搜集、组织和处理显得比以前任何时候都更为重要，这对以数据、信息为原料的社科研究带来了新的机遇和挑战。如何在浩瀚的数据海洋中「抽丝剥茧」、「去粗取精，去伪存真」，如何获取、组织、整理和应用对我有用的数据资源成为对社科学者学术水平的新考验，信息技术的使用和处理能力成为考量科研能力的新标准。面对数据爆炸，一些知名的老年学者呼唤像以前「学术秘书」那样的「信息学术秘书」协助他们做好对数据的前期准备和处理工作。

（二）区域性研究、跨学科研究中的数据共享需求

社科院是一支学科众多、研究领域齐全的哲学社会科学研究的国家队，研究范围涵盖了哲学、人文社会科学的方方面面，很多学科研究有很高的学术水平，在国、内外有很大的影响。院内有些研究领域划分比较细，研究所专业性比较强，比如对历史问题的研究按时间顺序分为历史所、近代史研究所、当代中国研究所等；按地域划分有世界历史研究所；按研究领域的划分有中国边疆史地研究中心、考古研究所等。还有相当多的研究所从事着区域性研究，比如主要研究边疆问题的「中国边疆史地中心」、主要研究民族问题的「民族学与人类学研究所」、主要研究城市发展问题的「城市发展与环境研究所」等，这些研究所研究的内容涉及到了相关的法律、政治、经济、文化等众多方面。不仅如此，即便是专业划分比较细的研究所也在向着跨学科、跨领域方面发展。要研究美国问题，只了解美国是不行的，还要了解美国周边国家的情况，与美国密切接触国家的情况，甚至是

美国敌对国家的情况；仅了解美国政府是不行的，还要了解美国的历史，美国的法律，美国的企业和美国人民等等诸多方面。这不仅对我们的研究机制、体制提出了挑战，对我们学者的研究方式、研究内容提出了新课题，也对我们如何进行数据、信息组织、数据建设，如何提供多学科的数据支持提出了新的要求。下面以人口问题为例，谈谈个人对数据建设的看法。

全世界都非常关注中国的人口问题，因为今天中国出现的很多情况都与其庞大的人口数量有关。我国在 1954 年发表了第一次人口普查公报，全国人口总数量从原来普遍认为的 4 亿 7 千万被修正为 6 亿 193 万 8,035 人，并预测人口增长率大体上会保持在千分之二。时任北京大学校长的马寅初先生基于这组数据进行推算，提出我国如果不对人口进行积极的控制会出现严重的问题。十年过去后，1964 年我国进行第二次人口普查，全国人口达到了 7 亿，同年**国家成立了国务院计划生育委员会**。1972 年，人口接近 9 亿，政府提出了「**实行计划生育，使人口增长与国民经济发展相适应**」的战略思想¹。在 2000 年我国进行第五次人口普查时，全国人口数量为 12 亿 9,533 万²，2002 年 9 月 1 日我国实施了《**中华人民共和国人口与计划生育法**》，根据 2009 年统计局公布的数据，我国人口约 13.35 亿。当前我们面临城市化的巨大压力，城市发展没人不行，但大量的农民进入城市后，政府要解决人员合理安置，住房、就业；子女上学；医疗、保险；养老等一系列问题，人口问题与城市建设、环境保护、经济发展、就业、社会稳定等诸多问题联系在一起，密不可分。比如对中国目前 2.11 亿流动人口的管理，以及由此带来的人员的流动和迁徙问题，那一个城市会成为下一个流动人口聚居区，如何做到提前准备，防患于未然，做到可持续性发展，这些问题都需要进行多学科、多领域的共同研究，这些研究要求以真实、准确、综合、共享的数据分析做基础。

国际上在多学科数据建设和共享方面的一些做法值得借鉴。下面是我们在 2006 年对欧洲社科研究的考察时了解到的情况：（该文引自小组的考察报告）

国际社会调查计划，简称 ISSP (The International Social Survey Programme)，是一个持续的跨国年度调查项目，覆盖了社会科学研究领域中的重要主题。1983 年，美国的一般社会调查 (the General Social Survey, GSS)、英国社会态度调查

(the British Social Attitudes Survey, BSA)、德国社会调查 (the Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, ALLBUS)和澳大利亚的国家社会科学调查 (the National Social Science Survey, NSSS) 等四个调查的资助机构探讨合作问题, 在各自的年度普选调查上进行合作。大家同意将一个 15 分钟的比较问题集作为一个补充模块, 并同意扩展共同的背景变量。「中心数据库」将负责把各国的数据集合并为一个跨国数据集。每一个参与组织将为自己的运作提供资金。除了作为一个跨国分析的资源外, 这一合作计划还要能够提供时间比较。为此, 主要的题目将每 5 年重复一次。以此为基础诞生了「国际社会调查计划」(ISSP)。

调查从 1985 年开始, 第一批模块包括了「政府的作用」。随后, 中心的议题是社会网络、社会的不平等、家庭和变化中的性别角色、宗教、工作方向和环境等内容。参与机构的代表每年任命一个小组来分别设计模块。想要参加 ISSP 的组织必须符合的要求, 比如使用随机抽样, 使用 ISSP 的标准扉页纸数据, 资金能够保证运作的连贯性, 和存储能够合并到国际数据集的数据。文书工作的地点定期轮换等。刚开始有 6 个国家参与, 现在有世界各地的 40 个国家。主要调查内容包括政府的角色、社会网络、社会不平等、家庭性别角色、宗教、环境、民族身份和公民权等。

网址：http://www.gesis.org/en/data_service/issp/

(三) 数据建设中的资产性和可持续性的需求

随着社会的发展, 特别是信息时代的到来, 为信息资产赋予了新的内容。国务院印发的《2006—2020 年国家信息化发展战略》指出, 「进入 21 世纪, 信息资源日益成为重要生产要素、无形资产和社会财富」, 充分肯定了信息是一种资产。在知识经济时代, 特别是在信息沟通越来越方便简捷的今天, 数据、信息资产意义更是超越了传统资产。以前我们更多的是在谈数字资源的建设, 很少去谈数据资产建设, 我认为**数据建设的任务更应该是把资源转变为资产, 并让资产创造出财富。**

我院十几年的信息化建设实践经验告诉我们, 信息化建设、特别是数据库建

设需要长期的、巨额的经费投入，但数据、信息的高投入却不等同于高价值。数据讲求量，信息讲求质，无用或无效的数据同样会花费相当大的费用，产出和保存的却可能是信息垃圾。所以，我们要走出「死库人必死，活库逼死人」怪圈，就需要更加关注提高数据质量、提升信息价值等问题，做好数据资产的保值和增值。

有学者给出数据库系统总拥有成本的数学模型³：

$$T_{C(\min)} = T_F + \sum_{N=1}^X T_N(T_F)$$

其中， T_C 是总的拥有成本， T_F 是建设时期的一次性投资成本， T_N 是系统在第 N 年的维护成本， X 是系统的生存周期。由于系统维护与一次性投资有关，作者把维护成本看成了建设成本的函数。从公式中我们可以看出，数据库的成本是不断变化着的，后期成本的累加将会成为数据库建设中不可忽视的重要因素。

笼统的讲，影响数据价值主要体现在以下几个方面：首先，数据、信息的价值与其内容质量（如：完整性、准确性、及时性和可用性等）有关，和数据、信息所提供的功能有关，也就是说，越有用的数据价值越高。其次，从表现的形式上讲，共享程度越高、使用的人次越多、使用频度越高、服务时间越长其数据的价值越高。试想，如果不被访问的数据、信息存储在高端的服务器和存储设备上，高昂的维护成本必然会造成资源的浪费。同样，一组相关性很差的数据能反映出的信息量和价值受到了极大的制约。另外，拥有数据、信息资产，并不说明可以让这些信息产生使用价值，脱离了工作实践、脱离了实际应用，为了完成任务而生产、管理的数据和信息资产是缺乏使用价值的，是得不偿失的。

为此，在进行数据资源建设中，要求我们尽可能的降低数据、信息化建设的综合成本，最大化的保持和提升数据、信息的价值。不断把数据资源变成数据资产，努力把数据资产转变成数据财富，逐步走向市场化运营的模式，保证信息化建设健康、稳定、可持续发展。

二、我院数据资源建设现状

(一) 科学研究产生出了巨大的原创性数字资源。

我院现有 55 个所局单位，主管的社团有 107 个；院属非实体研究机构近 90 个；每年国际间的学术交流 1,300 余批次，互往人员达 2,000—3,000 人次；每年举办的 50 人以上的国际学术会议近百场。每年结项完成的科研课题超过 300 项。每年大约为 5 亿汉字量的原创资源。

以 2008 年为例，我院完成专著有 379 种，约 1 亿 3,942 万字；学术论文 5,473 篇，约 6,200 余万字；研究报告 1,545 篇，超过 4,100 余万字；学术资料 132 种，约 3,609 余万字；译著 101 种，约 3,023 万字；译文 644 篇约 1,200 万字；论文集 113 种约 3,500 万字；影视资料 100 余部约 3,300 小时；古籍整理 53 种；教材 43 种；工具书 27 种；一般性文章 4,100 余篇，超过了 1,800 万字，还有 93 种核心学术报刊、期刊，学术科普读物等大量数据资源。

(二) 大量自建的数字资源，在学术研究中起到了积极的作用。

我院科研数据库的建设由来已久，一些研究所在上个世纪 90 年代初期即开始了数据库的建设，跟据 2009 年对我院信息化建设的统计，各所上报的可以使用或曾经使用过的自建数据库 348 个，其中，计算机网络中心 19 个，文献中心 43 个，经济学 32 个，文史哲学 105 个，社会政法学 33 个，国际研究学部 93 个，其他单位及职能机构 23 个，外购数据资源近百个。从统计的结果来看，除经济学部大多使用关系型数据库外，院自建的数据库以非关系型数据库居多；关系型数据库主要采用了微软公司的 SQL Server 软件，非关系型数据库的建设以国产天宇软件为主。外购的数据基本上是免费向全院开放，所内自建的数据库基本上是自给自足，在所内授权访问。这些数据资源的建设在我院学术研究和科研活动中发挥了积极的作用。在调查中我们也了解到，如今我院的数据库建设也正在朝着走出去的战略方向发展，出现了一些像超星杂志、论文库、图书馆藏，国民党档案数据库，满铁剪报数据库等联合开发，共建共享，互通有无的数据库建设模式，这对我院数据建设的健康发展会起到积极的推动作用。

调查中我们也看到了一些问题，各所自建的数据库的数据规模普遍都不大，

数据库建设基本上是以课题组为单位，一些数据库是应课题任务的需求而建的，这也造成了有些数据库因为课题的完结，课题组的重组，以及课题完结后没有后期维护经费的支持，失去了数据更新和维护能力形成死库。比如国际学部曾经使用过的 93 个数据库，在 2010 年 4 月份的统计表中显示，仅有 60 余个数据库还在维护，大约占到总建设数量的三分至二左右，在这 60 余个数据库中记录数量不足 1,000 条的 37 个，约占 61%；超过了一万条数据记录的大约有 10 个，占六分之一左右。院里自建的数据库，成规模的还是凤毛麟角。

（三）数据建设中我们应该着手做的几件事情：

1. 加快数据标准建设，处理好各所数据资源建设的高度热情与数据标准建设滞后的矛盾。

我们已经建设了 300 余个大小不一，内容疏同的数据库，对学术研究起到了积极的支撑作用。但因为标准不统一，库与库之间关联性不强，数据库基本停留在了资源、资料的水平，没有形成能够共享的合力，没有合成为高效的资产，更没有转变为数据财富。没有规矩不成方圆，没有统一标准建设就没有真正意义上的数据资源建设和资源财富的积累。所以我们要抓紧我院信息建设的标准化建设工作，在统一的标准下开发我院的信息资源，推动我院信息产业发展。

2. 发挥院、所两个积极性，处理好各所数据资源建设与院重点数据资源建设经费投入的关系。

首先，数字资源的应用需求在研究所，准确的数据资源来源也在研究所，这决定了数据资源建设中的主力军是研究所的科研人员，他们才是数据资源建设和应用的主体。其次也应该看到我们的数字资源建设，一方面是数据建设经费严重不足，一方面却是对经费使用的疏于管理。小而不全的数据建设，建设经费投入不低，使用效果不高，维护成本难以保障；数据建设的高度分散性、不系统性，造成了数据的效用和价值都比较低。院要出规章、出制度，有组织、有规范、整合多学科的数据资源，形成可以服务于全院各学科的海量数据库。

3. 建立新的数字资源建设的体制机制保障。

数据资源建设源头在研究所里的研究人员，组织和管理在院里的相应机构，服务于全院或全国的社科研究人员，提升的是整个社会科学的科研、学术水平，受益的是全社会的老百姓。没有一支专业的数据建设的人才队伍，没有适合的体制、机制保障，数据资源建设只能是停在口头上的一句空话。多年来的信息化实践告诉我们，数据资源建设必须调动全院科研人员的积极性，采取共建、共享的模式，遵从互通、互利的原则，形成合理的数据资源产业，才可能保证数据资源建设健康、有序和可持续发展。为此，我们要加强数据资源共建、共享等政策性研究，加强数据建设组织管理，加强数据资源对科研的支撑力度，让大家真正尝到甜头，投入到数字资源的建设中来。好的消息是，院领导已经决定并开始「数据中心」的组建工作。

4. 搭建安全的信息服务平台。

中科院院长路甬祥说「在当今世界，信息化是科学技术发展的重要基础，没有先进的信息化环境，不可能跟上世界科学技术快速发展的步伐，更不可能在一些领域取得重大的突破。」数据资源建设的目的在于应用，搭好数据资源服务平台是数据应用的基础。

院里要搭好安全的信息服务平台，提供有效的数据资源服务。好的信息服务平台，首先要求有科学合理的建设结构；其次要有良好的用户界面，具有分布式查询能力和面向对象技术；同时还要保证系统在管理大量的异构数据资源时具有较好的灵活性和可扩展性；最好能够做到资源内容的动态推送，科研需要什么就推送给他什么；最后一点也很重要，平台要有足够的开放性，能够灵活适应信息技术高速变化和发展，如云计算、移动办公等。

三、展望

社会科学家的职责是：研究借鉴过去已经发生的历史经验，剖析现在社会中存在的问题，把握将来可能出现的状况，做出切实可行应对建议和对策。所谓上

医医未病之身，正如古时的扁鹊，此时就显示出了名家的水准。要对社会事件做出准确的判断，需要真实、完整、准确、及时的科研数据作为研究工作者的食粮，否则也只能是巧妇难为无米之炊。

中科院副院长江绵恒认为：「整个信息化并不单单是一个技术问题，更大程度上是一个观念的问题，是管理理念的问题，是我们创新的科研模式的产出问题。」

我院信息化建设已经走过了十二年，现在发展到了一个调整期，整合资源，融合资源，升华资源。今后的一段时间，我院信息化建设将会围绕着服务来展开，来讨论。信息服务要建立的体系还有很多：如信用体系、认证体系、标准体系、安全体系，可能还会有支付体系、物流体系等方面的建设。

根据需要，我们还将建设全院的信息基础平台、信息服务平台、功能服务平台、交易服务平台、物流服务平台等，还需要我们从人才、资金、技术、法律、政策等多方面的关注，需要从节能减排、资源综合利用、信息安全等多方面的考量。

参考文献

- 1 新华网《我国计划基于政策产生及发展过程》
- 2 《2000年第五次全国人口普查主要数据公报》
- 3 情报学报第2期《数据库系统的总拥有成本》
- 4 社科院信息化建设十年论文汇编

【作者简介】

周世禄 男

职 称：社科院 计算机网络中心 研究员

職 務：社科院 计算机网络中心 副主任

研究领域： 计算机网络、信息化建设、网络与新媒体传播

个人简介： 1982年毕业于清华大学，1994年调入社科院，2003年始任计算机网络中心副主任，主要负责网络中心的技术工作。1996年以来，组织、参与了院计算机网络和网络管理系统、网络安全系统、邮件系统、域名系统等信息系统的论证和搭建工作。现任国家金财工程专家咨询委员会委员、中国计算机用户协会网络分会副理事长、社科院研究生院新闻系硕士生导师。

联络电话： 86-13901285350

联络邮箱： zhousl@cass.org.cn