

大数据:电子数据证据的挑战与机遇

高波

(厦门大学法学院,福建 厦门 361005)

摘要:在大数据时代,挑战与机遇并存。从海量电子数据中寻找那一丝与诉讼相关联的证据犹如海中捞针,使电子数据证据收集的成本增加,电子数据证据偏在成为难以解决的结构性问题。更为重要的是大数据的核心价值及数据挖掘方法对人们思维方式的改变,对电子数据证据的收集与司法应用产生了一定的影响。大数据时代催生了新的电子数据证据的采集方法,大数据挖掘技术也给电子数据证据的发现提供了新思路。在大数据时代最需要做的是努力使电子数据证据收集模式借助于大数据处理技术发挥更大的作用。应对大数据的挑战,把握大数据创造的机遇,完善电子数据证据收集的权利与义务以及对应用新技术收集电子数据证据给予相应的证明力,才能有效保障当事人公平地接近证据,维持当事人在诉讼上公平公正竞争,以促进诉讼及发现真实。

关键词:大数据;电子数据证据;挑战;机遇

中图分类号:DF713 **文献标志码:**A **文章编号:**1008-5831(2014)03-0111-09

2012年被称为大数据元年,大数据开启了一次重大的时代转型,它正在改变我们的生活和认识世界的方式。“这是一场革命,将横扫学界、商界和政界,所有领域都将被触及”^[1]。大数据的冲击迫使人们对电子数据证据的认识活动必须提高到一个新的水平,而认识活动的基本目标是发现“真实”,从纠纷解决的角度看,事实真相的发现对纠纷解决具有永恒的价值^[2]。

大数据的核心价值是预测。通过数据挖掘(Data mining)^①获得大数据的深层含义,世界许多在单纯依靠人类判断力的领域都会被计算机系统所改变甚至取代。大数据分析不仅使亚马逊知道我们喜欢的图书,让淘宝网推荐我们可能需要的产品,甚至在识别犯罪、证据搜索上发挥巨大的作用。基于传统证据收集制度的思维与方式已完全不能适应大数据时代的步伐,面对开放复杂的大数据系统,传统的因果分析难以奏效,因为“大数据时代开启了一场寻宝游戏,而人们对于数据的看法以及对于因果关系各相关关系转化释放出来的潜在价值的态度,正是主宰这场游戏的关键”^[3]。在证据法理论中有一个普遍原则,即必须首先证明有关证据就是提出证据的人所主张的证据,然后才有该证据的可采性问题,这个普遍要求有时被称为证明奠定基础^[4]。当我们进入一个用数据进行预测的时代,我们可能无法解释其背后的原因,但是大数据预测产生的“可能证据”却会对证据法中的“可采证据”产生巨大影响。如何应对大数据时代带来的电子数据证据收集方式与思维模式的改变是保障当事人公平接近证据的重要问题。

一、大数据对电子数据证据收集的挑战

(一)必须面对大数据“数据量”的问题

“看到大数据这个词,大多数人的第一印象恐怕就是 Volume,也就是数据量”^[5]。IDC^②研究表明,数字领

修回日期:2013-10-22

作者简介:高波(1971-),男,蒙古族,内蒙古赤峰人,厦门大学法学院博士研究生,集美大学政法学院副教授,法律系副主任,主要从事诉讼法学、海事行政法研究。

^①最早使用“data mining”术语的人是统计学家。术语“data mining”或者“data dredging”最初是贬义词,意指试图抽取出数据本身不支持信息的过程。当然,现在术语“data mining”的意义已经是正面的了。参见王斌译,Anand Rajaraman, Jeffrey David Ullman 著《大数据:互联网大规模数据挖掘与分布式处理》(人民邮电出版社2012年版,第1页)。

^②全球著名的咨询机构——国际文献资料中心。

域存在着 1.8 万亿 GB 的数据。企业数据正在以 55% 的速度逐年增长。ReadWriteWeb 表示,如今,只需两天就能创造出自文明诞生以来到 2003 年所产生的数据总量^[6]。数据规模越大,处理的难度也就越大。

数据量如此之多,以至于大数据时代不再热衷于追求精确度。达到精确需要有专业的数据库。针对小数据量和特定事情,追求精确性依然可行。但是,在大数据时代追求精确度已经变得不可行,甚至不受欢迎。当我们拥有海量即时数据时,绝对的精确不再是追求的主要目标。但是一般认为在证据收集过程中,“凡认为可能与案件有关联或者有助于证明诉讼问题的事实、法律和其他情况,都在收集、提供之列”^[7]。许多案件中,电子数据证据的收集须耗费极大的时间、人力及物力,甚至于不具有“合理存取”的可能性。以视频监控为例,连续不断的监控流中,对案件事实有重大价值的可能仅为一两秒的数据流;在 360°全方位视频监控的“死角”处,也可能会挖掘出最有价值的证据。在海量电子数据中寻找那一丝与诉讼相关联的证据犹如“在数据的干草中捞到有意义的‘针’,其困难就是‘许多干草看起来也像针’”^[1]。大数据适当忽略微观层面上的精确度尽管会让我们在宏观层面拥有更好的洞察力,但却动摇了证据的客观真实性。

(二) 大数据复杂性影响了电子数据证据的质量

大数据包含了“海量数据”的含义,但是在内容上、规模上都超越了海量数据,换言之,大数据是“海量数据”加“复杂类型的数据”^[8]。大数据的所有数据集,其规模或复杂程度超出了常用技术按照合理的成本和时限捕捉、管理及处理这些数据集的能力。因为复杂性的存在使它只能提示和解释某些事情。海量数据意味着增加了有效使用数据的难度,因此科学评估数据质量和确定有价值的数据集也是一大挑战^[9]。如果电子数据证据收集者认为大数据能够证明的事情比它实际能够做到的多,那就是一件危险的事情。

大数据的复杂性对电子数据证据质量提出了挑战,这种冲击相对于“海量数据”而言更不可小觑,因为“放在天平上的分量不是证据的数量而是由证据产生的盖然性以及案件的全部环境决定的”^[10]。英国学者彼特·莫菲认为在民事案件中的证明标准就是足以表明案件中负有法定证明责任当事人就其主张的事实上的真实性大于不真实性,即要求“或然性权衡”和“盖然性占优势”的标准^[11]。而盖然性占优势证明标准意味着,凡是对于特定事实的存在有说服负担的当事人,必须以证据的优势确立其存在。证据的优势与证人的多寡或证据的数量没有关系^[12]。因此,优势证据不是一项数量标准,而是一项质量标准,反映了证据的可信度和说服力。

通过大数据挖掘获得的电子数据证据,常常会脱离它们所在的上下文的情境,所以有时这些电子数据线索就会被错误地解读。只有怀着一种平衡而和谐的法律责任感来解读电子数据证据,才能在某些人的违法行为刚开始显现时得以成功的遏制。但特别需要指出的是,这些电子数据在没有达到法律要求的质量标准而作为证据使用是一种司法权利的滥用,会极大地损害公平正义的价值追求。因而,电子数据证据收集的任务除了获取越来越多的数据外,如果需要单纯依据电子数据预测作出决策,特定的防护措施就必须到位。

(三) 大数据时代电子数据证据偏在^③的问题凸显

罗森贝克认为,必须对法规范要素在真实的实践中得到实现承担主张责任和证明责任^[13]。《民事诉讼法》第 64 条规定,当事人有责任对自己提出的主张提供证据,即“谁主张谁举证”。但是,20 世纪中期以后,前述基本规则在诸如医疗纠纷、环境污染纠纷、交通事故纠纷和产品质量损害赔偿纠纷等现代型案件中的运用经常导致实质上的不公正现象。于是理论界出现了“证据距离说”、“危险领域说”、“盖然性说”等新说,对特殊类型侵权案件中某些要件事实的证明责任进行了有别于基本规则的分配^[14]。

一般说,如果电子数据证据本身掌握在当事人自己手中,并且易于保存或者制作,那么证据收集活动便较为容易。但是如果电子数据证据被控制在与案件对方当事人或者有关或者无关的第三者的手中,或者电子数据证据难以保存、容易损毁,则电子数据证据收集难度便较大。在大数据背景下,电子数据的持有者通

^③证据偏在是台湾学者提出的概念,认为以公害诉讼、交通事故、商品制造人责任及医疗纠纷等相关事件之处理,如严守“当事人主张有利于己之事实者,就其事实有举证之责任”之原则,难免产生不公平之结果,使被害人无从获得应有之救济,有违正义原则。是以受诉法院于决定是否适用该条但书所定公平之要求时,应视各该具体事件之诉讼类型特性暨待证事实之性质,斟酌当事人间能力、财力之不平衡、证据偏在一方、搜证之困难、因果关系证明之困难及法律本身之不备等因素,透过实体法之解释及政策论为重要因素等法律规定之意旨,较量所涉实体利益及程序利益之大小轻重,按待证事项与证据之距离、举证之难易、盖然性之顺序(依人类之生活经验及统计上之高低),再依诚信原则,定其举证责任或是否减轻其证明度。参见中国台湾地区最高法院 1999 年度(台上字第 836 号)民事判决。

常与计算机网络设备有极大的关联性,该等设备的持有者,往往对该数字数据拥有所有权或管理的权限,而当事人却难以取得该电子数据证据。另外电子数据证据采集具有专业性强、技术性高的特点,并且在通常情况下,当事人如果不具备相应的技术知识就难以正确地收集电子数据证据。显然,电子数据证据偏在是难以解决的结构难题。而证据偏在最大的危险是容易发生证据的销毁与篡改,无论是民事、刑事或行政诉讼案件,掌控电子数据证据的当事人可能为了掩饰自己的行为或其他影响自身权益的事证,往往会通过销毁、篡改证据的方式,让他人难以追诉责任和主张权利。在传统的民、刑事案件中,常见的如伪造文书、变造文书^④,都是具体掩饰罪行、隐匿事证、颠倒是非的方式。在大数据时代若系统欠缺良好的安全机制,则相关事证数据很容易遭到篡改、销毁,由此降低电子数据证据的可信性与不可否认性,案件事实的厘清更显困难,甚至于有是非对错颠倒的结果。此时,受害者主张法律责任时,则将因欠缺可信赖的正确的电子数据,可能连证据能力的阶段都难以通过考验,自然也难以达到保障自身权益和取得胜诉的判决。

大数据时代下,电子数据证据偏在客观存在,只能借助于诉讼中的证据制度加以克服。为达成诉讼中“真实发现”及“公平”的基本价值,必须积极应对大数据时代下电子数据证据偏在的挑战,构建完善的电子数据证据收集的权利与义务关系来对此进行矫正。

(四)通过大数据的“预测”而收集的电子数据证据在适用中存在巨大挑战

大数据的核心价值在于对未来的预测。以刑事的角度看,人们首先想到的是是否可以用大数据来预防及惩罚犯罪。当然如果在犯罪行为发生之前及时制止比事后再惩罚要好得多。但是这很危险,因为人们可以用大数据来预防犯罪,就可能运用司法手段进一步惩罚这个未来的罪犯。即使其未实施的犯罪行为也可以对其起到威慑作用。一般人都会认为,如果只是阻止了某人的犯罪行为而不采取惩罚措施的话,他就可能因为不受损失而再次犯罪。

在大多数情况下,我们的证据分析会放在一个特定的人群之中来进行界定。这是在小数据时代所采用的“按图索骥”的方法。预先设定若干假定的特征,然后在若干证据中找到普遍联系,对适用这种普遍联系的个人深入勘察。这适用于团体内的每一个人,是一条普遍规则。不过这是一种滥用,如巴拉巴西所描述的哈桑·伊拉希(Hasan Elahi)^⑤被美国国土安全局探员调查一样,先入为主地用有色眼镜观察会有严重的缺陷。

大数据规避了“按图索骥”的缺陷,因为大数据区分的是个人而不是群体,所以我们不会再通过“牵连犯罪”给群体中的每个人都定罪。但是大数据“预测”却将我们置于另一个司法难题中,即通过数据挖掘得到的电子数据证据能否直接成为诉讼中的证据。在互联网中用户在搜索引擎中查询过的每一个话题都被逐一记录下来,运用大数据挖掘技术完全可以得到某些有“价值”的信息。例如美国在线(AOL)曾公布了2000万条“去识别化”的查询记录,按大数据的标准,这是一个很小的一组数据,然而这组数据在研究人员手中不仅基本识别出搜索人的身份,更重要的还发现一些可能是犯罪的“线索”,某人大量的查询过“如何杀死老婆”、“死人”、“死者”、“谋杀的照片”等信息^[15]。是否可以从这些搜索记录来判定此人谋杀或预谋杀他的妻子?如果大数据分析完全准确,那么此人的未来行为会被精确地预测,当然,精确的预测是不现实的,大数据分析只能预测一人未来很有可能进行的行为。

中国《刑事诉讼法》第48条规定,可以用来证明案件事实的材料都是证据。解决了证据“事实说”的矛盾与问题^[16],依“材料说”显然“事实”还没有发生,那么“材料”也不能成为证据。当然更主要的问题是对当事人在真正犯罪之前进行惩罚否定了他的自由选择的权利。我们永远不会知道这个受惩罚的人是否真正犯罪,因为我们已经通过预测预先制止了这种行为,如此一来,我们就没有让他按照他的意愿去做,但是我们却依然坚持他应该为自己尚未实施的未来行为付出代价,而通过大数据的预测也永远无法得到证实。这

^④相对于中国的书证,其他国家和地区的“文书”范围更广泛,是指能够通过文字及其他符号与方式表达制作者意思或认识的物品,除通常理解的文件、票据等书面形式物品外,也包括图画、照片、录像及其他能表达信息的物品,后者也被称为准文书。参见张卫平主编《外国民事证据制度研究》(清华大学出版社2003年版,第367页)。

^⑤哈桑·伊拉希(Hasan Elahi)是一位美国多媒体艺术家。他出生于孟加拉国,七岁时移民美国。棕色的皮肤与一个具有民族特征的名字让他自2002年6月起便不断地遭受调查与盘问。后来,他养成一个习惯,每次出国都提前给联邦探员联系,告知其旅行计划,便极少再受到骚扰。2004年1月,哈桑创建www.tackingtransience.net网站上传了自己的照片与行动坐标,公开了其所有的信息,不仅联邦探员随时可以知道他行踪,世界人任何一个想知道他行动的人,在这个网站上都可以查到。参见艾伯特·拉斯洛·巴拉巴西著《爆发:大数据时代预见未来的新思维》(马慧译,中国人民大学出版社2012年版,第2-10页)。

否定了刑事证据法的基本原则之一——无罪推定原则。因为当事人被追究责任,居然是为了当事人可能永远都不会实施的行为,而当事人也无法用证据来证明他未来不行实施的行为。这是大数据预测给证据法基本理念带来的威胁,这不仅仅局限于当事人公平接近证据方面,它还会威胁到任何运用大数据预测对我们未来行为进行责任判定的领域,比如民事诉讼中判定过失以及公司解雇员工的决策。

二、大数据创造了电子数据证据收集的机遇

迄今为止,在审视电子数据证据时,人们惯常的思维模式是寻求一种能够脱离电子数据的虚拟世界而在现实生活中存在的替代物。这种思维的基础是不相信在网络中流动的数字化信息的真实性。诚然,科学技术的发展使伪造电子数据资料变得极其容易,因此人民法院对于电子数据资料,应当辨别其真伪,不要被其表面现象所迷惑^[17]。然而“大数据之所以成为时代变革力量,在于它通过追随意义而获得智慧”^[18]。

(一) 大数据挖掘为电子数据证据收集提供了技术方案

在法律实践中,最重要和最困难的任务之一是事实调查,事实调查涉及证据的分析和收集。“911事件”之后,人们反复宣称这一事件本应能够预测到,因为美国情报机构已经收到一些信息,但他们未能“连接信息点(connect the dots)”,或者无法从大量数据中鉴别出某些有意义的琐事,这些信息从各种渠道流入不同的机构^[19]。从平淡无奇的数据资料中发现、归纳和获取有价值的数据挖掘的需求应运而生^[20]。“现代人从大数据里挖掘价值的过程与古老的沙里淘金有着惊人的一致性,只不过对象从实物变成了抽象的代码”^[21]。在科学研究领域,基于密集数据分析的科学发现成为继实验科学、理论科学和计算科学之后的第四个范例。美国布什政府曾于2002年提出了一项针对所有可获得的数据进行挖掘的计划,目的用于追踪恐怖活动,该计划被称为整体情报预警(Total Information Awareness, TIA), TIA计划无疑在隐私倡导者当中受到极大的关注,虽然最终它并没有被国会通过,但其实这种计划可能已被冠以其他名称而得以真正实施(显然基于“斯诺登事件”分析认为, TIA计划已被PRISM计划“棱镜计划”代替)^⑥。如果浏览了这么多的数据,并且想从这些数据当中发现疑似的恐怖行为,会不会找出很多无辜的行为?答案取决于数据挖掘所使用的算法是否可靠^[22]。

大数据挖掘发现电子数据证据“连接信息点”最早最成功的应用是针对信用卡欺诈。通过搜集用户的刷卡记录,信用卡公司分析这些记录与持卡人信息的特点之间的关系,得出某类持卡人的一个典型消费模式,当这张信用卡被盗刷,或持卡人意图进行信用卡欺诈时,信用卡公司会通过刷卡终端搜集到处于这个模式之外的消费信息,进而标记持卡人并准备后续调查或拒绝交易。如今,利用公共互联网信息监测技术开展“数字知识产权”维权的证据收集亦是成功的应用。即通过“网络基础数据库”及维权搜索系统,可有效地发现侵权行为的源头和证据链条,为权利人定位重点打击侵权主体提供诉讼的有效证据。

(二) 大数据催生了电子数据证据新技术保全方式

依据《民事诉讼法》第64条,中国的证据收集制度确定了两种证据收集的制度:一种是当事人自行收集,向人民法院提供;一种是人民法院收集。一般来说,证据的收集与提供原则上都由当事人来完成,除法律规定的职权调查外,法官原则上没有证据调查与收集的义务。中国的这种证据收集制度,由于缺乏可操作性,导致了实践中的混乱^[23]。诚然,中国法律在证据的收集方面还有一些例外的规定,例如为了克服或避免当事人意志以外的因素导致证据收集权与证明权的不平等而设立了证据保全制度。证据保全制度一定程度上实现了证据收集和证明在实质意义上的平等,同时也使通过保全方式获得的电子数据证据的证明力有所提高。然而,民事诉讼中法官一般不会主动积极地帮助当事人调查证据,而需要由当事人申请,那么从申请到保全行为的实施必然有很大一个时间跨度,在电子数据证据可能随时灭失的紧急情况下,这种取证方式则可能因贻误时机而难以发挥其作用。

大数据中电子数据证据往往以数据流的形式存在,具有很强的时效性,用户只有把握好对数据流的掌控才能有效地利用这些数据。同时,电子数据自身的状态也往往随时空变化而发生演变从而影响其价值。

^⑥2013年6月,前中情局(CIA)职员爱德华·斯诺登将两份绝密资料交给英国《卫报》和美国《华盛顿邮报》,据报道,美国国家安全和联邦调查局于2007年启动了秘密监控项目——“棱镜”计划,可以直接进入美国互联网公司的中心服务器里挖掘数据。参见《谍影重重 斯诺登去向掀起大国风波》载于2013年7月2日亚太日报, <http://www.apdnews.com/xzl/233.html>。2013年6月7日,美国总统奥巴马公开承认该计划。但他强调这一项目经过国会授权,不针对美国公民或在美国的人,目的在于反恐和保障美国人安全,参见中国网《棱镜门中,中国的微妙处境与审慎应对》(http://www.china.com.cn/opinion/node_7185858.htm,访问日期2013年7月24日)。

所以在大数据时代最需要做的是努力使传统工作模式借助于大数据处理技术发挥更大的作用,在大数据技术的基础上探讨如何促进人类自身发展,这才是推动人类社会前进的认识方式。必须充分利用大数据处理技术去收集保全电子数据证据。大数据时代网络就是一个大服务器,或者说是许多的服务器集合。它的特征是24小时不间断地运行,随时都能发送、保存和展示信息。相比自然人,它的最大优势在于可能完成与任何一种网上行为“同步工作”的要求。目前,尚不能指望网络服务商在其所储存的海量信息中寻找电子数据并提供给当事人,并且目前网络服务商也无这样的法律义务。但是,利用大数据处理技术在网络中设立一个虚拟第三方,由其为用户提供包括远程服务器录像保全、电话录音保全、网页实时截图保全、邮件证据保全等多种电子数据证据保全形式,对无形的电子数据进行提取与固定^⑦。在需要时提交给司法机关作为原始证据使用。对个人而言,互联网云技术和智能手机的普及为开发大数据电子数据证据收集应用搭建了最好的虚拟第三方平台^⑧。一般而言,中立的第三方,资信状况比较可靠,由无利害关系第三方提供电子数据证据,理论上电子数据证据“保管链条”比较完整,在很大程度上能够保障电子数据证据的真实可靠性。

三、大数据时代电子数据证据的应对措施

大数据时代主要对电子数据证据的收集产生重大影响,中国法律对当事人及其诉讼代理人收集证据的权利作了概括性的规定以维护当事人及其诉讼代理人的证据收集权利,这主要包括《民事诉讼法》和《中华人民共和国律师法》,此外还有《最高人民法院关于民事诉讼证据的若干规定》(下称《若干规定》)等相关的司法解释。但从整体上看,当事人的电子数据证据收集仍缺乏有力的法律保障,尤其是在大数据时代,急速增长的数据让人焦头烂额,传统证据收集制度无法保障当事人公平地获得电子数据证据,当事人证据权仅仅停留在字面上。建构完善的公平接近证据权利义务体系,成为大数据时代电子数据证据收集的制度机遇。

(一) 基于证据权利与义务关系的应对措施

电子数据证据收集问题对大数据时代纠纷解决的重要性,决定了参与诉讼各方都必须对证据的获取承担相应的义务。而中国现行法律制度关于当事人电子数据证据收集的权利与义务的规定不科学,造成了当事人权利义务的不对等。但在诉讼制度改革中逐步通过立法和司法解释确立当事人举证责任的同时,当事人调查收集电子数据证据的权利仍没有得到实质性改善。目前审判方式改革中出现“当事人的举证责任沉重,而当事人的举证权利稀薄”^[24]。这种不平衡的原因在于,当事人主义下没有把调查收集证据的权利赋予当事人,导致当事人主义的诉讼模式被人为地扭曲。如果在职权主义模式下,当事人收集证据权利的“虚化”尚不足以构成对实现实体正义的威胁,那么在当事人主义模式下,对当事人的证据收集权没有给予充分的程序保障就不能实现实体结果的公平正义,缺乏程序保障的权利事实上已经异化为非权利了^{[24] 1047-1049}。

当我们从更深层次去考虑这一问题时,当事人电子数据证据收集权利与义务的失衡,不仅造成了当事人电子数据证据收集难的问题,而且在不能保证当事人充分收集电子数据证据的前提下,直接适用证明责任进行判决也会造成事实上的不公平^[25]。当事人不能公平接近证据,这一问题的存在已经影响到程序公正与正义,更有甚者将严重阻碍实体公正的最终实现。在大数据时代下必须完善电子数据证据收集权利与义务制度,这是当事人公平地接近证据、实现其诉讼权利必不可少的手段。

1. 设置电子数据证据证明请求权

《若干规定》第75条规定,有证据证明一方当事人持有证据而“无正当理由”拒不提供的情况,并且对方当事人主张的该证据内容是对证据持有人不利的,那么就可以推定该主张成立。也就是说除非有正当理由而拒绝提出数字证据的情况,法院是不得推定对证据持有人不利的主张成立,或者是该证据应证的事实为真实,此规定在中国台湾地区被称为证据持有人享有的“隐匿事证自由权”。中国法律并未明确规定“正当理由”是什么,一般理解认为,持有证据的当事人享有某种值得保护的利益,足以与证据收集权能受保障的当事人的利益相抗衡的,自应允许其拒绝提出文书,方属妥当,或者是享有“合理隐匿权”^[26]。

民事诉讼法在赋予“隐匿事证自由权”的同时,也应当赋予当事人享有证明请求权,所谓证明请求权,是指对于在认定事实上所必要的证据,应享有提出证据以证明事实的权利。如果不赋予当事人“证明请求

^⑦例如网络上出现“公证云”服务平台,用户使用智能手机、个人计算机终端通过公证云平台所提供的各种取证手段,以一定的形式将无形的电子数据加以固定并保存在公证云平台所提供的平台用户数据库内及公证机构监督的公证数据库内,以便在厘清事实时使用。参见 <http://www.cunnar.com/html/knowledge.html>。

^⑧例如由熊志海、黄永洪研究的“一种基于电子数据证据在线保全的第三方认证保全系统及方法”已申请发明专利。

权”,而持有证据一方享有“隐匿事证自由权”,这样的权利分配,则是回归到举证责任分配的基本原则性规定,当事人需要对其有利的法律规范要件负举证责任。而经济学认为人是理性自利者,愿意自我承认不法情事者,恐怕只有圣贤人才会如此,所以当然不会在没有法律规范的要求下,就自愿地将相关电子数据证据双手奉上。因此,理性自利的当事人当然会践行“隐匿事证自由权”,拒绝将相关事证提出,此一结果在大数据时代电子数据证据偏在的情况下,更显现出不公平不公正的结果。也因此特定情况下,举证责任分配的原则均有例外。如中国台湾地区民事诉讼法第277条规定:“当事人主张有利于己的事实者,就其事实有举证的责任。但法律别有规定,或依其情形显失公平者,不在此限。”在大数据时代电子数据证据极易产生证据偏在的情况下,有扩大当事人文书提出义务范围的必要,即“知悉事实、持有证据之当事人,在经请求时,必须说明事实、提出证据”的义务^⑨。应当赋予当事人“证明请求权”,才不会因为证据偏在的情况而影响案件真实的发现。

2. 借鉴文书提出命令制

文书提出命令制是赋予举证当事人据以搜集他方所持文书作为证据的机会,可要求持有文书的对方当事人或第三人开示与诉讼有关联的文书资料,以贯彻当事人间武器平等原则,保障其公平接近证据的证明权,并维持当事人在诉讼上的公平公正竞争,促进诉讼及发现真实。很多国家立法都对当事人收集书证的程序作出了相关规定,只是在具体程序规定上存在一定的差异。如法国民事诉讼法规定了强制提出书证制度,即法官可以应他方当事人的请求,要求提交其持有的某项证据材料,如果该当事人拒绝提交,法官有权科处逾期罚款^[27]。在德国,文书提出命令是当事人所享有的主要证据收集方法^[28]。而依据中国台湾地区民事诉讼法第342条、第343条规定,当事人有权“申请命他造提出文书”,即在申请状中载明他造有提出文书义务,而文书由他造所执者,则应由当事人申请法院命他造提出。法院认为应证事实重要,并且申请人请求正当,即可裁定命他造提出文书。当然,综观各国的文书提出命令制,对于提出范围亦有明确规定,以防止申请人通过滥用这种权利来达到拖延诉讼等不正当目的。因为当事人在诉讼上负有提出文书义务属于一种有限度的公法义务,不宜作出任何更为扩大化的理解,它应当以法律明确规定的具体适用范围为限度^[29]。

以民事诉讼制度自充实、保障当事人证据收集权能的角度观察,在大数据时代电子数据证据偏在现象凸显,当事人一方因证据的结构性偏在而无法取得另一方所持有的电子数据证据,有碍其诉讼上的主张与举证,违反当事人间的实质平等,电子数据证据制度自当借鉴文书提出命令制度设置能够使当事人取得他人持有的证据制度。对处于弱势地位的当事人提供了保护手段,使诉讼双方当事人能够获得充分提供审理所需要的电子数据证据的能力或渠道。

3. 建立电子数据证据妨碍排除制度

受国家司法权管辖的任何人,只要其了解案件的事实情况,就有向法院作证的义务。《民事诉讼法》第72条规定,凡是知道案件情况的单位或个人都有出庭作证之义务。这是国家使证人的作证一般义务化表现。日本在修订民事诉讼法时将这种义务扩展到文书提出范围上,即实现了文书提出行为的一般义务化^⑩。因而一般认为“当事人应对法院用以查明事实真相的诉讼程序中所运用的证据加以保存负有普遍性的义务”^{[29]265}。这种义务或来自于法律规范的强制;或来自当事人自愿承担;或产生于案件提交法院,其权益受到威胁,或合理预见到以后会形成讼案时。当客观存在这种证明协力义务,而义务人无正当理由拒绝履行义务时,则对案件待证事实具有不可替代性的证据材料或证据方法将构成妨碍遭致他人产生不利的裁判后果,即构成证明妨碍。如果法院此时还是适用证明责任原则作出判决,从而导致负有证明责任的当事人败诉,那么对当事人不公平。于是就应当用证明妨碍来“避免通过证明责任作出裁判”^[30]。大数据时代,数据容量增长的速度引发了数据存储和处理的危机,从而大量的数据会被删除^{[21]6}。传统的证据发现和电子数据证据发现最重要的区别是电子存储信息的数量和删除的难度不同。由于巨大的存储容量的电子系统,越来越多的数据被保存,但当新技术致使以前的系统过时,恢复这些数据却不容易^[31]。

如前所述,如果以篡改电子数据证据的内容或毁损数据的方法构成证明妨碍,那么就会使负有证明责

^⑨参见中国台湾地区最高法院1999年度[台上字第1361号]民事判决。

^⑩日本通过1996年的《民事诉讼法》修改,确立了争点中心审理主义,因此在技术层面就必须赋予当事人能够进行充分提供审理所需要的证据的能力和渠道,对文书提出命令进行强化并予以普遍化成为重点。参见张卫平主编《外国民事证据制度研究》(清华大学出版社2003年版,第371-373页)。

任的当事人陷于证据缺乏的境地,致使案件事实真伪难判。大数据时代应当在电子数据证据规则中建置证明妨碍机制。如果法律规范有要求当事人设置电子数据储存设备及保存相关电子数据的具体规范,当诉讼进行中或预知诉讼发生可能之时,却以妨碍对方使用为目的,故意将相关涉及诉讼的电子数据证据损毁、隐匿或设置障碍使用等行为,则可以判定为证明妨碍,可以考虑实行行政罚款,并认可对方关于该文书的主张或依该文书应证的事实为真实,或者是调整举证责任等措施^①,以有效遏止当事人此种违反协力义务的行为,从而达到发现真实的目的。

(二) 基于证据规则的应对措施

1. 拓宽公平接近证据的理解

大数据预测的准确性越来越高,它几乎能预测未来行为的发生,那么通过大数据预测的结果在人们犯错之前,就可以提前采取某些措施。因为预测结果披着“科学”的外衣几乎不可反驳,人们很难为自己开脱。但是这种基于预测得出的惩罚不仅违背自由意志原则,同时也否定了人们会突然改变选择的可能性。当法官给当事人判定责任时,必须牢记人类意志的神圣不可侵犯。人类的未来必须保留部分空间,允许我们按照自己的愿望进行塑造。否则,大数据将会扭曲人类最本质的东西,即理性思维和自由选择。通过大数据挖掘对未来可能行为进行惩罚是对公平正义的亵渎,因为公平正义的基础是人只有做了某事才需要对它负责,毕竟,想做而未做不是犯罪。社会关于个人责任的基本信条是:人只为其选择的行为承担责任。

身处大数据时代,我们必须拓宽对公平接近证据的理解,必须把对个人动因的保护纳入进来,就像目前我们为程序公正所做的努力一样。如若不然,公平的信念就可能被完全破坏。要确保个人动因能防范“数据独裁”的危害——我们赋予数据本不具备的意义和价值。通过保证个人动因,我们可以确保司法对人为行为的评价是基于真实行为而非单纯依靠大数据分析。从而,只能依法对人们过去的真实行为进行追究,而不可以追究仅用大数据预测到的当事人未来的行为。在司法评判当事人过去的行为时,应尽量避免防止单纯依赖大数据分析。如果必须依据大数据分析才能获得的电子数据就必须公开用来进行预测分析的数据和算法系统;具备由第三方专家鉴定的可靠、有效的算法系统;明确提出个人可以对其预测进行反驳的具体方式。

2. 因果与相关——大数据时代电子数据证据收集的思维转变

寻找因果关系是人类长久以来的习惯。人的行为与结果之间的因果关系也成为司法实践中判定责任一关键要素。然而在大数据时代,我们无须再紧盯事物之间的因果关系。但我们又陷入了一个历史的困境,那就是我们活在一个“上帝已死”的时代。当大数据时代由探求因果关系变成挖掘相关关系时,我们怎样才能不损坏建立在因果推理基础之上的保障电子数据证据的收集与认定呢?

2002年诺贝尔经济学奖得主普林斯顿大学的丹尼尔·卡尼曼(Daniel Kahnman)证明了人有两种思维模式,一是不费力气的快速思维,通过这种思维方式几秒钟就能得出结果;二是比较费力的慢性思维,对于特定的问题,就是要考虑到位^[32]。运用快速思维模式会使人们偏向用因果关系来看待周围的一切,即使这种关系并不存在。这是我们对已有的知识和信仰的执著。在古代,这种快速思维模式很有用。它能帮助我们在信息缺乏时却必须快速做出决定。如最初的神示证据制度,就是人们相信超自然力量与案件真相的发现存在某种因果关系。但是,通常这种臆想的因果关系并不存在。如今在日常生活中,由于惰性或者是习惯性直觉的存在,我们很少慢条斯理地思考问题,所以还会经常臆想出一些因果关系,最终导致对世界的错误理解。在小数据时代,很难证明由直觉而来的因果联系是错误的,因为我们大部分的习俗和惯例都建立在一个预设好的立场上,那就是我们用来进行决策的信息必须是少量、精确并且至关重要的。但是,当数据量变大、数据处理速度加快,而且数据变得不那么精确时,之前的那些预设立场就不复存在了。

当人们已经习惯了从因果关系的视角来理解世界时,大数据总是被滥用于因果分析,而且人们往往非常乐观地认为,只要有了大数据预测的帮助,法官进行个人责任判定就会更高效。然而,通过大数据获得的电子数据证据不能告诉人们因果关系。相应地,进行个人责任推定需要行为人选某种特定的行为,他的选择是造成这个行为的原因。但是大数据技术并不是建立在因果关系基础之上的,所以它完全不应该用来帮助法官进行个人责任的推定。

^①参见中国台湾地区民事诉讼法第282条之一“因妨碍他造证据使用之时之举证责任减轻。”

相关关系分析是分析因果关系的基础。正因如此,英美证据法将相关性作为现代证据法律制度的基本原则。美国联邦证据规则 401 认为,证据具有与没有该证据相比,使得某事实具有更可能存在或者不可能存在的任何趋势^[33]。只有相关的证据才有助于陪审团取得理性的成果,即建立在陪审团成员们运用其推理能力获得的成果。而在中国证据法中被普遍认可的证据的客观性、相关性与合法性中,证据的相关性是最基本的属性,它也是证据理论的核心问题之一。其实通过找出可能相关的事物,我们可以在此基础上进行进一步的因果关系分析,如果存在因果关系的话,可以再进行进一步找出原因,这种便捷的机制通过严格的实验降低了因果分析的成本。通过大数据挖掘可以从相关联系中找到一些重要的变量,这些变量可以用到验证因果关系的实验中去。因为在大多数情况下,一旦完成了对大数据中的电子数据证据的相关关系分析,知道了“是什么”后就会继续向更深层次去研究电子数据证据之间的因果关系,找出背后的“为什么”。因此在大数据时代电子数据证据的收集完全可以从相关关系出发,形成初步证据链条后,再依传统的证据收集方式进行证据调查。大数据挖掘的电子数据证据提供的不是最终答案,只是参考答案,为司法提供暂时的帮助。

四、结语

总之,大数据技术及相应的基础研究已经成为学术界的热点,大数据科学作为一个横跨信息科学、社会科学、网络科学、系统科学、心理学、经济学等诸多领域的新兴交叉学科正在逐步形成。伴随着大数据技术的进步,新的事实确认方式已经开始在司法领域挑战传统的事实认定法^[34]。大数据的核心价值是预测,最基本的方法是数据挖掘,大数据不仅是一种提供信息的资源库,同时也是发现电子数据证据的一种工具。如今许多证据不再是以纸面的方式储存,而逐步转换成以数字的形态呈现,电子数据将不再只是证据法的一小分支,而将快速地成为一个主要的形态。大数据对电子数据证据的影响已见端倪,这种影响将会更广泛和深远。当电子数据证据制度在面对大数据的影响,而有所扞格不入的结果时,就必须适时地加以调整。我们在使用大数据挖掘与分析电子数据证据的时候应当怀有谦恭之心,只有铭记人性之本才能充分保障当事人能公平接近证据的权利。

参考文献:

- [1] Big Data's Impact in the World [EB/OL]. [2013-07-28]. www.nytimes.com/2012/.../big-datas-impact-in-the-world.html.
- [2] 何家弘. 证据法学研究[M]. 北京:中国人民大学出版社,2007:228.
- [3] 舍恩伯格·维克多·迈尔,库克耶·肯尼思·大数据时代——生活、工作与思维的大变革[M]. 盛杨燕,周涛,译. 杭州:浙江人民出版社,2013:20.
- [4] 罗纳德·J·艾伦. 证据法:文本、问题和案例[M]. 北京:高等教育出版社,2006:205.
- [5] 城田真琴. 大数据的冲击[M]. 周自恒,译. 北京:人民邮电出版社,2013:4.
- [6] 耿秋,孟剑. 大数据时代 机遇? 挑战? [J]. 中国新时代,2012(6):60-61.
- [7] 梁玉霞. 什么是证据——反思性重塑[C]. 北京:中国检察出版社,2001.
- [8] 涂兰敬. 专家观点:“大数据”与“海量数据”的区别[J]. 网络与信息,2011(12):37-38.
- [9] BIZER C, BONCZ P, BRODIE M L, et al. The meaningful use of big data: four perspectives——four challenges[J]. ACM SIGMOD Record,2012,40(4):56-60.
- [10] 沈达明. 英美证据法[M]. 北京:中信出版社,1996:46.
- [11] PETER MURPHY. A practical approach to evidence[M]. Blackstone Press Limited, 1992:105.
- [12] 摩根. 证据法之基本问题[M]. 台北:世界书局,1982:48.
- [13] 莱奥·罗森贝克. 证明责任论[M]. 北京:中国法制出版社,2002:104.
- [14] 陈荣宗. 举证责任分配与民事诉讼法[M]. 台北:三民书局有限公司,1984:24.
- [15] ABELSON HAL,LEDEEN KEN,LEWIS HARRY. 数字迷城:信息爆炸改变你的生活[M]. 李卉,王思敏,张魏,译. 北京:人民邮电出版社,2011:55-56.
- [16] 卞建林. 中华人民共和国刑事诉讼法最新解读[M]. 北京:中国人民公安大学出版社,2012:65.
- [17] 全国人大常委会法制工作委员会民法室. 《中华人民共和国民事诉讼法》条文说明、立法理由及相关规定[M]. 北京:北京大学出版社,2012:108.

- [18] 姜奇平. 大数据的时代变革力量[J]. 互联网周刊,2013(1):34-37.
- [19] 特伦斯·安德森,舒姆戴得,特文宁·威廉. 证据分析[M]. 北京:中国人民大学出版社,2012:64-73.
- [20] TAN P. Introduction to data mining[M]. Pearson Education India, 2007:4.
- [21] 郭晓科. 大数据[M]. 北京:清华大学出版社,2013.
- [22] ANAND RAJARAMAN,JEFFREY DAVID ULLMAN. 大数据:互联网大规模数据挖掘与分布式处理[M]. 北京:人民邮电出版社,2012:4.
- [23] 常怡,杨军. 我国证据收集制度的反思与重构——兼论民事审判方式的改革[J]. 汕头大学学报,2002(1):49-56.
- [24] 赵信会. 民事诉讼中的证据调查制度[J]. 现代法学,2004(6):87-92.
- [25] 汤维建. 论美国民事诉讼中的证据调查与证据交换——兼与我国作简单比较[C]. 北京:人民法院出版社,2000.
- [26] 汤维建. 民事诉讼法全面修改专题研究[M]. 北京:北京大学出版社,2008:105.
- [27] 刘玉中. 民事诉讼上证据收集之研究[D]. 台北大学,2004.
- [28] 张卫平. 外国民事证据制度研究[M]. 北京:清华大学出版社,2003:494.
- [29] 何家弘,张卫平. 外国证据法选译(上下)[M]. 北京:人民法院出版社,2000:485.
- [30] 毕玉谦. 民事诉讼证明妨碍研究[M]. 北京:北京大学出版社,2010.
- [31] 高桥宏志. 民事诉讼法:制度与理论的深层分析[M]. 北京:法律出版社,2003:466.
- [32] NICHOLS L R. Spare the rod, spoil the litigator—the varying degrees of culpability required for an adverse inference sanction regarding spoliation of electronic discovery[J]. Ky. LJ. 2010, 99: 881.
- [32] 丹尼尔·卡尼曼. 思考,快与慢[M]. 胡晓姣,李爱民,何梦莹,译. 北京:中信出版集团股份有限公司,2012:5.
- [33] 王进喜. 美国《联邦证据规则》条解[M]. 2011年重塑版. 北京:中国法制出版社,2012:56.
- [34] 达马斯卡. 漂移的证据法[M]. 北京:中国政法大学出版社,2003:200.

Big Data: Challenges and Opportunities on the E-evidence

GAO Bo

(School of Law, Xiamen University, Xiamen 361005, P. R. China)

Abstract: In the era of big data, it seems like seeking for a needle in a haystack to find a tiny evidence associated with proceedings from the mass of electronic data, and the complexity of data is also a challenge to the quality of E-evidence, as not only increases the cost of collecting E-evidence, but also makes the imbalance of E-evidence a difficult structural problems to solve. The era of big data has spawned new methods of collecting E-evidence, and the mining technology of big data also provides new ideas to the discovery of E-evidence. In the era of big data, it is thus necessary for us to make the collection mode of evidence play a bigger role by using big data processing technology. Confronting the effects of big data, we should improve the rights and obligations of collecting E-evidence, give corresponding weight to the collection of E-evidence by using new technologies, which can effectively protect the parties equitably access to evidence, maintain equitable competition, and promote litigation and the discovery of authenticity in the civil proceedings.

Key words: big data; E-evidence; challenges; opportunities

(责任编辑 胡志平)