



# 数据新闻制作 简明教程

吴小坤 著

# 数据新闻制作 简明教程

吴小坤 著

 复旦大学出版社

## 图书在版编目(CIP)数据

数据新闻制作简明教程/吴小坤著. —上海: 复旦大学出版社, 2018. 5  
(网络与新媒体传播核心教材系列)  
ISBN 978-7-309-13668-5

I. 数… II. 吴… III. 数据处理-应用-新闻学-教材 IV. G210.7

中国版本图书馆 CIP 数据核字(2018)第 080181 号

## 数据新闻制作简明教程

吴小坤 著

责任编辑/朱安奇

复旦大学出版社有限公司出版发行

上海市国权路 579 号 邮编: 200433

网址: fupnet@fudanpress.com <http://www.fudanpress.com>

门市零售: 86-21-65642857 团体订购: 86-21-65118853

外埠邮购: 86-21-65109143 出版部电话: 86-21-65642845

上海浦东北联印刷厂

开本 787×960 1/16 印张 12.75 字数 189 千

2018 年 5 月第 1 版第 1 次印刷

ISBN 978-7-309-13668-5/G · 1835

定价: 38.00 元

---

如有印装质量问题, 请向复旦大学出版社有限公司出版部调换。

版权所有 侵权必究

# 代序

2013年,我带领团队创建了财新可视化实验室。从最初尝试性地做,到后来拿到亚洲新闻奖、全球数据新闻奖、国际数字媒体创新大赛奖等,总共不到三四年的时间。有人戏称我为“中国数据新闻第一人”,或许是因为我的做法促成了国内对数据新闻的关注,也带动了国内不少媒体开始做数据新闻。很少有人知道,数据新闻最开始在国内出现有赖于三位留学回来的学子,他们创办了数据新闻网,还组织翻译了那本蜚声海外的《数据新闻手册》。后来在与国内高校数据新闻教师的交流中我发现,这本书是课堂必备品。然而,由于国内外环境的差异,数据新闻面临着本土化缺位的问题。我组织过的一项高校调查发现,数据新闻课程在国内还很不完善,优秀的人才和教材都十分匮乏。大家在学习过程中也大多依靠东拼西凑的技能。尽管现有的数据新闻教材多强调理论和丰富的案例,但远不能满足整体流程上手操作的学习需求。

吴小坤老师的这本《数据新闻制作简明教程》理论与实践并重,面向数据新闻的初学者,提供了完整的流程和操作指南。使用这本教程,即便是零基础的学习者,也能掌握数据新闻制作的基本技能。与其他教材相比,这本书更轻松易读,操作性更强。对数据新闻关键要素和基本技能的清晰阐释,体现出作者丰富的教学经验和难得的综合素质。这本书的内容去繁求简,提取了数据新闻中最关键的部分,并提供了进一步学习的方向和知识链接。对于希望能够迅速掌握数据新闻基本方法的学生和教师来说,尤其适用。



当你选择阅读这本书的时候,你或许已经认识到数据新闻是个有趣的事物,它的前面是数据,后面是可视化,这两者是包括新闻传播在内的各行各业未来发展绕不开的方向。随着传媒行业对能够处理数据、掌握前端、写深度故事的人才需求的急速增多,转变才刚刚开始。

2017年11月28日

# 前 言

2014 年年初,我在美国密苏里大学第一次听到“数据新闻”的时候,还并不知道它意味着什么。经过一年的学习,我深深地被这种外形优雅、内涵丰富的新闻样式所吸引。自 2015 年年初在上海大学第一次开设数据新闻课程到 2017 年到华南理工大学继续讲授这门课程,我在教学过程中查阅了数据新闻中英文的各种教程,从中汲取了很多营养。但若换位思考,站在初学者的角度,还是希望有一本简单清晰、易于上手的教程。这样一本教材,也将有益于教师在课堂上按图索骥,引导学生完成数据新闻作品。

出于这样的想法,2017 年年初开始筹划写作这本教材,历时近一年,终于得以面世。这本教材的整体框架得益于我在密苏里新闻学院所学习的数据新闻课程布置,同时根据三年以来自己开设、讲授这门课程的过程中所遇到的一些情况和问题,对内容进行了本土化的补充和调整。教学经验告诉我,即便是从未接触过数据新闻的“文科小白”,跟随这本书的章节细致学习并完成课后练习,也能够很快上手,完成属于自己的数据新闻作品。

依照数据新闻学习的规律,从选题入手,到数据分析,再到可视化设计和呈现,这本教程共分为 10 个章节,每个章节中对该部分所涉及的基本理论、知识和完成作品所需要的工具与技术,都进行了详细的介绍和操作说明。出于篇幅的限制,很多软件和工具的使用并没有深入展开,建议学习者根据相应的介绍继续深入学习。在每个章节相应的部分,都提供了资源链接。书后附录也给出了部分链接,可供参考。

数据是数据新闻特殊性的源泉所在,数据的价值决定了数据新闻的价值。在这本教材中,对数据抓取、分析和处理的部分有所偏重。数据可视化的部分,也以此为基础。本书在内容顺序上做了考究,建议初学者尽量依照



顺序,并能够完成各章节练习。

需要说明的是,与国内既有的数据新闻教材有所不同,这本教程舍弃了案例分析的部分,将案例列表作为附录。但是,对数据新闻学习者来说,向好的作品学习是非常重要的。因此,从第二章开始,建议学习者坚持阅读数据新闻获奖作品,从中借鉴经验。

这些年来从事数据新闻教学的一个深切体会是,新闻领域即将发生更为深刻的变化,这种变化既需要技术也需要专业的新闻人才,更需要不断提升的人才培养能力机制。数据新闻是一个教学相长的领域,年轻的学生对新技术的接收能力快,有想法、有创意,尤其在技术的掌握方面,入门之后的自学很快就能在某个特定的技术应用方面超过老师。我的研究生李喆提供了第五章“使用 Excel 操作数据集”的初稿;另一位研究生王敏提供了第九章中“使用 Echarts 和 Tableau 制作交互图表”部分和第十章“将作品发布出来:HTML、CSS、JS 网页操作实践”的初稿。她们都是通过课程学习后,又自我学习并在某些方面超过老师的很好例子。现在,我很享受作为一线的教学者时常向学生学习的场景。在教学过程中,可以看到学生兴趣高涨,勤奋努力,能力不断提高,这也是一件令人欣慰的事。

数据新闻所涉及的内容多为前沿知识,无论是技术、工具还是理念,都在不断更新。本书尽可能用简单易上手的方式搭建数据新闻所需要的基本知识和技能,难免有疏漏之处,敬请批评指正。

吴小坤

2017年11月1日于华南理工大学

# 目 录

前 言 .....	1
第一章 什么是数据新闻 .....	1
1. 数据新闻的概念来源 .....	2
2. 数据新闻的使用与适用 .....	4
3. 数据新闻的五大特征 .....	6
4. 数据新闻对传统新闻价值的突破 .....	9
5. 数据深化应用的未来挑战 .....	15
第二章 选题判断 .....	18
1. 不是所有的故事都适合做数据新闻 .....	18
2. 先有数据还是先有故事 .....	19
3. 媒体数据新闻选题流程 .....	20
4. 用数据讲故事的选题自查 .....	21
第三章 数据来源 .....	24
1. 查找数据的基本方法 .....	24
2. 常用数据资源分类汇总 .....	27
第四章 数据分析：格式、应用及基本关系 .....	40
1. 数据是什么 .....	40
2. 数据长什么样子 .....	42





3. 新闻中应用数据分析的四种类型 .....	45
4. 数据中的因果关系与相关关系 .....	47
<b>第五章 使用 Excel 操作数据集 .....</b>	<b>51</b>
1. Excel 的基本功能 .....	52
2. Excel 的常用函数 .....	61
3. 神奇的数据透视表 .....	73
<b>第六章 使用更多工具提升数据处理效率 .....</b>	<b>78</b>
1. 从网页上下载数据表的好用工具 tabletools2 .....	80
2. 使用 Open Refine 做数据清洗 .....	81
3. MySQL 数据库系统的简单操作 .....	90
4. 地理坐标拾取和转换 .....	97
5. 使用 Python 和 API 抓取网络数据 .....	97
6. 使用 R 语言和 Selectorgadget 插件抓取数据 .....	103
<b>第七章 简单图式：什么数据适合什么图 .....</b>	<b>109</b>
1. 基本图形 .....	109
2. 根据数据需求选择图形 .....	113
3. 数据表达的图形误区 .....	114
4. 用图表讲故事的基本要领 .....	118
<b>第八章 想要漂亮的图形？设计与配色原理 .....</b>	<b>121</b>
1. 信息图 .....	122
2. 信息图的制作过程 .....	123
3. 如何做出有设计感的信息图 .....	125
4. 实用的信息图制作工具 .....	129
<b>第九章 使用 ECharts 和 Tableau 制作交互图表 .....</b>	<b>132</b>
1. 使用 Echarts 定制可视化图表 .....	133

2. Tableau 导入数据生成可视化图表 .....	143
<b>第十章 将作品发布出来：HTML、CSS、JS 网页操作实践 .....</b>	<b>152</b>
1. HTML、CSS、JS 是什么 .....	153
2. 外部库的链接和使用 .....	163
3. 实用工具 .....	170
<b>附录 .....</b>	<b>175</b>
1. 练习数据下载地址 .....	175
2. 本书中所用到软件下载地址 .....	175
3. 中外数据新闻奖项列表 .....	176
4. 国际知名大赛获奖作品列表(2014—2017) .....	179

# 什么是数据新闻<sup>①</sup>

◆ 数据新闻是基于数据分析和计算机技术的可视化新闻样式，在新闻叙事中使用数据呈现原本仅靠文字所难以呈现的内容，或者通过数据分析发现问题，并进而挖掘出新闻故事。

作为新闻生产的一大方向，数据新闻已在世界范围的新闻媒体中引起广泛关注。由于其真正的流行是在最近五年，其定义也尚未清晰。有学者指出，数据新闻的定义是不稳定的，在数据与新闻之间存在一种紧张而微妙的关系<sup>②</sup>。文献检索发现，该主题下的大多数研究都会先从“数据新闻是什么”这一问题展开，说明数据新闻还是一个较新的概念，学者往往需要先说明概念才能够开展围绕这个概念的分析 and 讨论。本书检索范围包括 18 本国内外相关教材和著作、239 篇 Ebsco 英文文献和 710 篇 CNKI 中文文献，通过人工筛选，提取了 59 种数据新闻的定义。尽管在可能的情况下尽量涵盖了更多的文献，但难免挂一漏万。同时，数据新闻研究文献迅速增多，也给这一概念提供了不断扩大的检验空间。本章主要从数据新闻概念的理论承递、使用和适用以及特征维度出发，对数据新闻的概念界定展

① 本章部分内容发表于《新闻与传播研究》2017 年第 10 期。

② Juliette De Maeyer, Manon Libert, etc, "Waiting for Data Journalism: A Qualitative Assessment of the Anecdotal Take-up of Data Journalism in French-speaking Belgium," *Digital Journalism*, 2015, 3(3), pp. 432-446.



开初步讨论。

## 1. 数据新闻的概念来源

从理论上讲,数据新闻与精确新闻、计算机辅助报道等概念有着一脉相承的关联。这主要体现在新闻报道中对科学性、精确性的追求和新技术手段的运用。20世纪60年代,美国学者、新闻记者菲利普·迈耶(Philip Meyer)提出了精确新闻理论。这一理论的提出,回应了当时新闻界的两个潮流:一个是新新闻主义的发展,另一个是社会调查研究的流行。根据迈耶的定义,精确新闻的“精确”来自采集方法的改变<sup>①</sup>,指记者在进行新闻报道时使用社会科学研究方法来收集新闻素材,完成新闻报道的采写,这些方法包括调查、实验和内容分析等<sup>②</sup>。而精确新闻学的出现,反映出对新闻报道准确性和客观性的追求,它要求记者通过调查和内容分析等科学方法,提高新闻的准确性和客观性。

在新闻报道朝向科学化和精确化发展的同时,计算机技术也逐步应用于社会调查和政府数据库建设中。这一历程包括:从20世纪50年代使用大型机处理政府数据库、发现和调查新闻事实的早期阶段,到70年代至80年代通过PC机普及、商业和政府数据库的进一步开放形成的“以新闻报道为目的,对任何计算机化的信息来源的处理和使用行为”的第二阶段,再到90年代中期以后利用互联网进行新闻采集、分析和制作的第三阶段<sup>③</sup>。计算机辅助报道强化了新闻报道的技术导向,即使用计算机来辅助收集和處理信息的新闻报道方式。安德森(C. W. Anderson)将计算机辅助报道定义为“算法的、社会科学的和精确的报道形式”<sup>④</sup>。在这个定义下,算法开始

---

① 章永宏、黄琳:《重建客观:中国大陆精确新闻报道研究》,中国书籍出版社2013年版,第4页。

② Meyer Philip, *Precision Journalism: A Reporter's Introduction to Social Science Methods*, Amerucac: Rowman & Littlefield Publishers, 2002, p. VIII.

③ 方洁:《数据新闻概论》,中国人民大学出版社2015年版,第17页。

④ Anderson, C. W., "Notes towards an analysis of computational journalism," *Social Science Research Network*, [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2009292](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2009292).

与新闻相结合,并为后来计算新闻学的兴起<sup>①</sup>做了铺垫。

数据新闻是在计算机辅助报道的基础上,从技术角度对新闻业的再度推进。这一概念的首次公开使用是在2011年1月的NICAR(National Institute for Computer Assisted Reporting,美国计算机辅助报道协会)大会上。在该会议上,学者们达成一项共识:“计算机辅助报道”这一概念已经无法阐述这个行业的实践。自那之后,“数据新闻”被全世界的记者广泛使用。在数据新闻林林总总的界定中,最普遍使用的是《数据新闻手册》中的定义,该教材将数据新闻描述为“用数据处理的新闻”,并认为“数据新闻能让记者通过数据和信息图表来报道一个复杂事件的过程。数据可以是新闻的来源,也可以是新闻中讲故事的工具,还可以两者兼顾”<sup>②</sup>。有学者总结道:“数据新闻被广泛定义为基于数据的抓取、挖掘、统计、分析和可视化呈现的新型新闻报道方式,其数据的获取方法和来源更加多元。”<sup>③</sup>学者方洁曾对这三个概念做过比较,她认为,数据新闻并不是由某一位学者提出的概念,这个概念兴起于一个遍布全球的数据记者群,其影响的范围更广;相对于计算机辅助新闻报道,“数据新闻的概念代表着一种新闻发展的形态,其内涵和外延比计算机辅助新闻报道更加广阔”<sup>④</sup>。在这个意义上,数据新闻指明了新闻朝向技术化发展的新阶段。

尽管存在时间先后和方法上的共通性,这三个概念仍然颇为令人困惑。困惑一方面来自上述三个概念之间颇为纠结的承递关系,另一方面来自对数据新闻中的“数据”所指及其与新闻结合的理解。从概念的承递来看,国内存在两种较为明显的观点:一种观点认为,数据新闻是由精确新闻发展而来,数据新闻是“在数据时代的衍生品”,是“精确新闻的进一

---

<sup>①</sup> Lewis, S. C., & Usher, N., “Code, collaboration, and the future of journalism: A case study of the Hacks/Hackers global network,” *Digital Journalism*, 2014, 2(3), pp. 383-393.

<sup>②</sup> Jonathan Gray, Liliana Bounegru, Lucy Chambers, *The Data Journalism Handbook*, O'Reilly Media, 2012, p. 6.

<sup>③</sup> Mark Coddington, “Clarifying Journalism’s Quantitative Turn: A typology for Evaluating Data Journalism, Computational Journalism, and Computer-assisted Reporting,” *Digital Journalism*, 2015, 3(3), pp. 331-348.

<sup>④</sup> 方洁、颜冬:《全球视野下的“数据新闻”:理念与实践》,《国际新闻界》2016年第6期。

步延伸”<sup>①</sup>；另一种观点则认为，数据新闻由计算机辅助报道发展而来，计算机辅助报道起源于精确新闻报道的需要，而数据新闻在计算机方面的应用可以被理所当然地视为计算机辅助报道在大数据时代的提升和发展<sup>②</sup>。两种观点的争议在于，数据新闻在概念界定上是侧重于分析逻辑，还是侧重于技术逻辑。其背后的实质性问题则是，数据新闻与传统新闻之间的传承关系是否存在以及如何存在。对这个问题的理解差异，也直接影响到数据新闻界定和发展思路的不同。

现在，无论是读者还是新闻记者，对数据与新闻的结合已不陌生。事实上，将数据应用于新闻并生成可视化在 20 世纪五六十年代早已有之。当时，著名的南丁格尔玫瑰图和伦敦霍乱地图在特殊的时代环境下，引起政府和公众的关注，将早期信息图形式的数据新闻带入人们的视野。随后，信息图与新闻的握手并不少见，但并未引起足够的重视，更未能形成“数据新闻”的理念。无论从技术应用还是数据丰富度来看，当时的新闻生产在数据获取和挖掘故事方面都受到很大的局限，而这些恰是促成精确报道和计算机辅助报道向数据新闻转向的必要条件。

## 2. 数据新闻的使用与适用

“数据新闻”概念的形成与新闻实践有着非常紧密的关系。从数据中挖掘和发现新闻故事的路径，改变了传统的新闻思维。新闻不再仅仅是信息或叙事，而且也是一个发现问题的过程。在这个过程中，以数据为核心，发现或者解释这个世界上原本不为人所知或者不受关注的事实，通过数据建立关联是数据新闻记者的一项基本的思维训练。较早的实践者是《纽约时报》和《卫报》，它们在 2011 年和 2012 年围绕同性恋、犯罪和骚乱等议题展开大量的数据新闻报道，取得引人注目的成功。在过去几年里，数据新闻的迅猛发展让这种报道形式成为国内外媒体的常备内容。目前，世界范围内

① 郭晓科：《数据新闻学的发展现状与功能》，《编辑之友》2013 年第 8 期。

② 苏宏元、陈娟：《从计算到数据新闻：计算机辅助报道的起源、发展和现状》，《新闻与传播研究》2014 年第 10 期。

的很多主流或新兴媒体都开设了数据新闻专栏,或者将数据新闻作为其探索性产品的重要组成部分,如国外的《洛杉矶时报》、《华盛顿邮报》、FT、FiveThirtyEight、Buzzfeed、ProPublica,我国的财新网、新华网、网易、第一财经等,都取得了很好的成效。

然而,由于其概念的不确定和范畴的相对模糊,不同媒体在尝试数据新闻时往往基于各自的理解,在技术使用和内容表现上差别较大,同时也存在一些因夸大工具性或数据泛化而产生的误解。数据新闻并不仅仅是使用了数据的新闻,也不是由数据构成的新闻。在理解数据新闻这一概念的时候,需要避免一些误区。一个常见的误解是,认为新闻里使用了数据,甚至仅仅是数字,就可以称得上是数据新闻。另外一个与之相对的误解是,认为数据新闻需要使用海量数据,大数据才能撑起数据新闻。这些误解的根源往往源自对“数据”的认识。需要明确的是,数据 $\neq$ 数字,数据 $\neq$ 大数据。根据技术大百科(Technopedia)的解释,“在计算机科学领域,数据指数字化的信息载体。它可以是数字、文字、图形、语音、符号、视频等任何形式,其本身没有意义,需要通过被使用才可能被赋予意义”<sup>①</sup>。然而,数据新闻的既有定义并不能解决其在实践应用层面的模糊性,这一概念需要进一步清晰。

数据只有在使用过程中才能转化为信息,并通过与故事结合呈现其意义。因此,数据新闻具有一定的操作适用性,即并非所有新闻都适合做数据新闻。数据新闻操作层面的适用性反过来能够对其概念框架形成预设。数据适用于新闻的价值和意义主要遵循两条路径:

其一,使用一定量的数据为新闻提供支撑,并以可视化的形式呈现原本仅靠文字所无法呈现的内容。

其二,从数据中寻找并发现问题,进而挖掘出新闻故事。

在第一条路径下,使用数据呈现原本文字所无法呈现的内容是关键。既有作品中有非常多的案例都对此有所体现,例如财新网《周永康的人与财》<sup>②</sup>将6万字的材料中的关系逻辑通过可视化呈现,Buzzfeed的《天空中的

<sup>①</sup> “Definition of Data,” *Technopedia*, revisited in <https://www.techopedia.com/definition/807/data>, 2017-09-04.

<sup>②</sup> 参见 <http://datanews.caixin.com/2014/zhoushicailu/>。

密探》<sup>①</sup>将美国国土安全局(DHS)和美国联邦调查局(FBI)的无人机航行的几万条数据以地图的形式呈现出来。这种做法成为当下数据新闻作品的一个主流方式<sup>②</sup>。在第二条路径下,从数据中发现的新闻故事往往能够打破人们的常规思维,帮助我们看到这个世界不同的一面。从操作层面来看,这条路径的要求更高,实现难度也较大,需要记者具备较强的数据运用能力,并对目标事件的环境和数据非常熟悉。

这两个标准直接指向的是数据新闻与传统新闻在目标要旨上的根本不同,它同样可被应用于辨别什么样的新闻适合做数据新闻,能够指导实践判断,并对数据新闻既有的界定予以补充。在此基础上,我们给出数据新闻的定义:

**数据新闻是基于数据分析和计算机技术的可视化新闻样式,在新闻叙事中使用数据呈现原本仅靠文字所难以呈现的内容,或者通过数据分析发现问题,并进而挖掘出新闻故事。**

这一定义旨在为数据新闻划分边界。基于两条路径的条件规定,有助于解释和衡量长期以来令人困惑的,用大量表格和数据报道呈现的宏观经济分析、产业深度报道、个股评述等是否算作数据新闻的问题。

### 3. 数据新闻的五大特征

本章从已有文献对数据新闻的概念界定中寻找共通性特征,并结合国内外数据新闻奖的评定标准,总结出数据新闻的特征包括:新闻性、统计性、工具性、生产流程标准化、新闻行业创新性。

统计 59 个数据新闻定义中的上述特征可以看出,这些概念多强调工具性和生产流程标准化,其次是统计性和新闻行业创新性,对新闻性的观照却非常少。这一方面显示了数据新闻的技术和工具化特征,以及在此基础上新闻生产流程的改变;另一方面也说明数据新闻在目前的发展阶段通常被

---

<sup>①</sup> 参见 [https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm\\_term=.igqaxKlxA6#.orXzGXJGjW](https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm_term=.igqaxKlxA6#.orXzGXJGjW)。

<sup>②</sup> 这类案例有很多,具体可参看全球数据新闻奖网站,以及财新数字说、DT 财经数据新闻栏目等。



作为传统新闻的延伸,并不需要特别强调新闻性。然而,概念界定所反映出的对工具性的强调和对新闻性的淡化取向,也给了数据新闻的批评者一个强有力的理由。

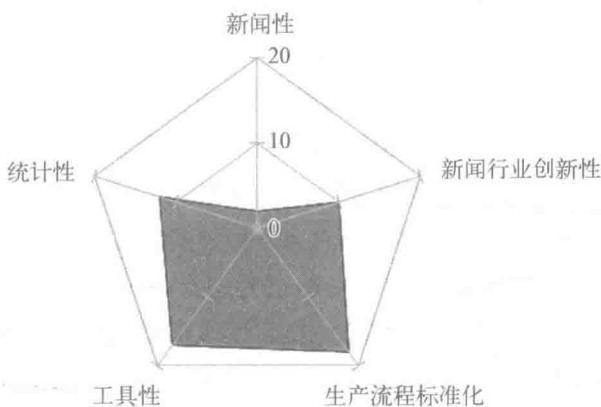


图 1.1 59 个数据新闻定义中特征的分布(2017 年 3 月 26 日统计)

对数据新闻的定义大多都不是单一指向的,它们会同时指向上述特征中的某几个,通常也会有所侧重。国外学者和国内学者的定义皆是如此。阿隆·菲尔霍夫(Aron Pilhofer)指出,“数据新闻是一个概括性术语”,它涵盖了一套正在不断发展的用于讲故事的工具、技术和方法<sup>①</sup>。布拉德肖(Bradshaw)依照新闻学里的“倒金字塔”结构理论,提出了数据新闻的“双金字塔”结构(自上而下是视觉化、叙事、社交化、人性化、个性化、应用化)<sup>②</sup>。《数据新闻手册》将数据新闻生产内容的三个主要阶段划分为获取数据、理解数据、传达数据,这一划分也被国内许多学者沿用、借鉴。与《数据新闻手册》类似,我国学者梁延将数据新闻的生产过程分为四个步骤:获取数据、分析数据、数据可视化和公开数据<sup>③</sup>。他强调,数据新闻就是利用一些开源软件和开放数据创建新闻故事以此揭示比传统新闻更深层的事实。

<sup>①</sup> Aron Pilhofer, “New Approaches to Storytellings,” *The Data Journalism Handbook*, 2012, p. 7.

<sup>②</sup> 任瑞娟、白贵:《数据新闻理论与实践:模式、发现与思考》,《新闻大学》2015 年第 3 期。

<sup>③</sup> 梁延:《大数据视野下“数据新闻”的发展现状、趋势及其困境》,《学术前沿》2014 年第 11 期。