

# 从“大数据”到“智能数据” ——剖析德国国家智能数据创新实验室的定位和模式

宋 娥

(中国科学技术交流中心, 北京 100045)

**摘 要:** 面对全球范围的“大数据之争”, 德国具备信息技术产业及市场快速增长、信息技术服务业和软件业世界领先、传统制造业数字化基础良好等优势条件, 亦有明显的结构性缺陷。为落实“数字变革战略”, 实施“数字议程”, 提升德国在大数据领域的核心竞争力, 德国联邦政府联合科技界和产业界代表, 启动了全新的国家尖端研究平台——“智能数据创新实验室”, 提出“智能数据”创新理念, 通过推动产学研合作, 对大数据进行进阶的智能化处理, 以抢占未来的战略发展高地。本文对该平台的定位和模式进行了剖析, 供国内有关部门参考。

**关键词:** 德国; 大数据; 智能数据; 创新实验室

**中图分类号:** TP311.13 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2015.04.003

2013年12月17日, 由德国基民盟、社民党组成的第18届大联合政府共同签署了以“塑造德国未来”为题的“联合执政协议”<sup>[1]</sup>, 确定了面向未来的一揽子科技创新发展规划, 其中一个重要举措是在科技领域制定“数字变革战略”, 改进复杂数据研究的获取和应用等问题, 实施“数字议程”, 在数字经济方面建设数字增长强国, 力争欧洲第一。在2013年底召开的德国国家IT峰会<sup>①</sup>上, 德国联邦教研部部长、“教育与科研面向数字化未来”分会主席约翰娜·万卡(Johanna Wanka)指出, “当前社会数字化信息量暴涨, 我们需要新的管理工具, 将其作为知识元并开发利用其价值。”<sup>[2]</sup>大会最终通过决议, 为相关科研机构和企业提供平台, 开发有效的智能分析工具, 将产业界的海量数据的研究和应用相结合, 使最新研究成果更加有效地向产业界转移, 提

升德国和欧洲企业在大数据领域的核心竞争力。

2014年初, 德国联邦教研部宣布, 2014年斥资2 000万欧元, 针对大数据课题启动新的科研项目, 主要涉及工业生产中的大数据处理以及生命与地球科学、数据保护、隐私与IT安全等重要课题。基于此, 德国全新的国家尖端研究平台——智能数据创新实验室(SDIL)<sup>[3]</sup>于2014年1月8日在卡尔斯鲁厄理工学院(KIT)<sup>[4]</sup>正式启动。

## 1 SDIL定位: 智能分析海量数据, 提升德国竞争力

数字信息被称为“21世纪的石油”。数字经济在不断发展的信息技术产业的培育下羽翼渐丰。2012年, 德国数字经济总产值达2 220亿欧元<sup>[4]</sup>。德国联邦政府深刻认识到, 在该领域占得先机, 是德国

**作者简介:** 宋娥(1984—), 女, 中级翻译, 主要研究方向为德国科技创新体系和技术创新政策国别研究。

**收稿日期:** 2015-04-01

<sup>①</sup> 为推动信息技术产业发展, 抢占未来战略发展高地, 德国联邦政府自2006年起, 每年举行一次国家IT峰会。峰会由德国联邦政府总理、相关内阁成员及科技、产业界精英构成, 是德国政府与科研、产业界沟通的重要平台, 也是确定信息技术产业战略布局的重要机制。

未来确保工业强国地位的新机遇，主要原因如下：

### 1.1 信息技术发展对德国数字经济有利，但存在明显缺陷

数字经济的本质是信息化。近年来，信息通信技术产业已发展成为德国第3大支柱产业。据德国信息通信新媒体协会（BITKOM）统计，2012年，德国信息通信技术产业产值达1 260亿欧元，超过了机械制造、化工、电子等传统行业。但是，德国的信息通信技术产业存在结构性缺陷：一是仅有信息技术服务业和软件业跻身于全球前列，其硬件技术、消费性电子产品技术则长期依赖亚洲和美国；二是即使在软件和信息通信技术服务业领域，其主要客户也是企业，服务供应方也以中型企业为主，很少有企业开展国际业务。

### 1.2 信息技术发展推动德国传统制造业升级

德国是世界上最具竞争力的工业生产基地，制造业占其国民经济的比重为26%，在先进制造和设备供应方面全球领先。但是，德国传统制造业面临资源紧缺、能源转型、员工年龄结构改变、全球化等挑战，须向更加灵活、高效和人性化的方向转变。德国联邦政府将信息通信技术与制造业相结合视为提升传统制造业的战略发展方向，于2012年正式推出“工业4.0”战略，拨款2亿欧元作为第一步研发经费，意在成为第四次工业革命的领跑者。目前，德国90%的工业生产流程应用了信息通信技术，其信息自动化及软件密集型嵌入式系统处于世界领先地位，尤其在汽车和机械制造领域，因此，专家认为，德国面对这场深远变革，具备明显优势。

### 1.3 智能数据领域尚无王者，产学研合作助德国抢占先机

大数据无疑会在数字化社会中发挥极大的作用，但面对体量巨大、种类繁多、价值密度低又相对自我封闭的海量信息，最为重要的是，对大数据进行挖掘和分析、使单纯数据生成有用知识的能力。但是在此方面，世界各国尚无重大突破。德国提出“智能数据创新”理念，建设SDIL，就是希

望在该领域实现跨越式发展，以便在联邦政府的支持下，在全欧范围内促进产学合作，尽快提升产业竞争力。

## 2 SDIL 的目标与研发机制

### 2.1 SDIL 的目标

利用德国在信息技术服务业和软件业的国际领先优势，结合其发达的传统制造业的良好基础，SDIL 希望集政府、企业和科研机构三方力量，探索将海量数据进行智能化联结的可行性模式，在智能数据技术方面实现突破，获得先发优势，实现大数据领域的跨越式发展。

#### 2.1.1 服务科技界

SDIL 服务于科技界的两大目标群体：一是信息科学相关研究，针对大数据的采集、传递、存储、计算、分析等，研究新技术、新算法（如人工智能研究、数据安全系统研究）以及对实时数据源方案的检测等；二是基于实时数据，以数字基础设施为工具的专项目标研究。

SDIL 为科技界提供以下服务：一是用于处理大数据的一流基础设施及工具；二是来自产业界、公共部门及互联网公开的数据源；三是被存储于开放源代码数据库中的代码构件。除提供一流的基础设施外，SDIL 的重要价值在于，平台参与者不仅能调用互联网上公开的数据以及出于计算需要上传至平台的数据，而且可以通过数据源目录获得实时数据源，甚至可以申请调用以高安全性模式加密存储的数据<sup>①</sup>。

#### 2.1.2 服务产业界

参与的企业在提供数据时，将自己在业界遇到的计算问题带到科研中并希望得到解决，这有利于在研究项目框架下，加强SDIL等科研机构与产业界的对话；与此同时，SDIL 希望向数据提供方开放其研究成果，以便加快技术转移和知识转让。参与的企业原则上享有对所提供数据的完全控制权，包括向个别研究机构授权调用其数据，规定数据使用周期以及何时删除该数据等。参与SDIL的企业

<sup>①</sup> 包括征得数据提供方同意可被用于科研分析的数据源，以及在数据提供方和研究团队建立的VPN（虚拟专用网络）上且在企业许可范围内提供的数据源。

享有如下优惠待遇：

（1）加入针对某一特定战略研究领域的“数据创新工作组”，可参与欧洲最前沿的研究机构合作项目，及时掌握智能数据领域的最新趋势。

（2）在提供数据的同时，也提出亟待解决的问题，以期通过相关研究合作解决问题（如检测数据异常、安全漏洞等）。企业可免费获得解决方案，其向研究团队提供的数据可被视为“虚拟货币”。

（3）上传数据时，需接受“数据管理员”审查，确保上传数据质量，尽量排除无价值数据；可在“数据管理区”了解关于数据上传和互联的最新信息。

此外，SDIL 重视为中小企业搭建平台，支持其与平台创始伙伴（多为大企业和科研机构）合作并建立联系；同时帮助宣传其产品，例如，如果中小企业的软件产品被证明有助于提高 SDIL 的整体研发能力，作为回报，SDIL 可帮助宣传推介其产品。

### 2.1.3 服务社会，惠及民生

德国联邦教研部部长万卡在 2014 年汉诺威国际电子、信息通信技术博览会上称：“在数据量急速增长的数字社会，必须有正确的数据使用方法，要有先进的技术来保障人们安全自主地在数字世界中活动。”数字化社会一旦形成，人类生活中的一切都可基于数据来描述。智能数据的理念是对海量数据进行分析，找出背后的规律和方法，并做出模糊判断。这样的分析系统将具备人工智能前所未有的基础能力——学习能力，根据环境（数据）变化不断增长其智能性，甚至具备推而广之的扩展性。

德国人工智能研究中心（DFKI）在 SDIL 落成典礼上表示：DFKI 为在苏黎世举行的大型焰火典礼设计了一种智能手机应用软件，使用时会自动报告地理位置信息，方便系统管理者实时统计观礼人数，并预测人数变化及人流行为模式等，以免再次发生类似杜伊斯堡踩踏事件的灾难性事故<sup>①</sup>。技术服务的本质对象是“人”，因此，SDIL 的最终目标也是开发出最佳的智能数据分析系统，真正及时、准确地判断人的需求，刻画和预测“人”的行

为模式，从而提升社会服务品质。

## 2.2 SDIL 的研发机制<sup>[6]</sup>

### 2.2.1 基础设施

KIT 提供高性能的 IT 基础设施，使科技界可以对大数据进行更高效的分析处理。作为国家级大型研究中心，KIT 还承担着欧洲核子研究中心（CERN）全球分布式网络数据中心 GridKa 及其他国家大数据基础设施的日常运营，并且与产业界联系密切。

### 2.2.2 实时数据源

为使研究尽可能贴近现实，由参与企业合作提供其实际运营中的数据信息，为 4 个数据创新工作组牵头开展的 4 大重点领域科研工作提供支持。这些数据是对互联网上公共部门数据信息的有益补充。

### 2.2.3 数据存储

如今的网络传输速度已远远不能满足海量数据的处理。SDIL 将那些以科研为目的的数据信息进行高安全性的集中式存储。目前，许多高度现代化的软件产品可满足这一要求，例如，内存数据库和分析工具。除了数据源目录外，SDIL 还提供一个开放源代码数据库，可存储研究过程中产生并可被科研机构继续利用的代码构件。

## 2.3 SDIL 的基本原则

数据保护是 SDIL 的首要工作原则，具体要求如下：所有数据处理都遵守“德国数据保护法”；KIT 提供的所有数据信息均以高安全性模式存储，无第三方授权不可调取；企业数据源只有在数据提供方明确授权时才可被调取；对多个数据提供方提供的数据信息进行处理后得到的结果，若未能明确定义知识产权，原则上不上传至平台。

## 3 SDIL 的组织架构和工作流程

### 3.1 SDIL 的组织架构

SDIL 由参与平台的产学研各界合作伙伴共同管理运营，其基本组织架构如下（参见图 1）：

#### 3.1.1 指导委员会

SDIL 的指导委员会由 4 个数据创新工

<sup>①</sup> 2010 年 7 月 24 日，德国西部鲁尔区杜伊斯堡市举行“爱的大游行”电子音乐狂欢节，因人群疏导不畅发生踩踏事件。据官方公布数据，踩踏事件造成 19 人死亡、342 人受伤，其中 1 名遇难者是中国公民。

作组的牵头单位与 KIT、软件供应方共同组成，共同决定SDIL 的资源配置和发展路线。

3.1.2 数据创新工作组

SDIL 优先服务于战略重点领域，现已规划

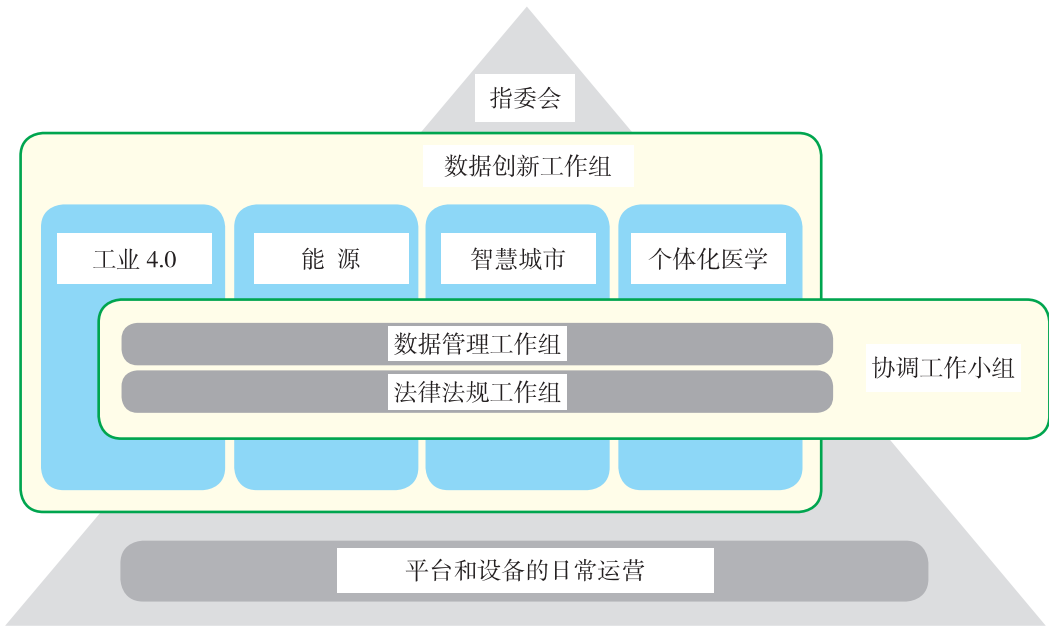


图 1 SDIL 的组织架构<sup>[6]</sup>

好 4 个数据创新工作组：工业 4.0（DFKI 和 Bosch 牵头）；能源（KIT 和 EnBW 牵头）；智慧城市（Fraunhofer IAIS 和 Siemens 牵头）；个性化医学（FZ Jülich 和 Bayer 牵头）。每组按“1+1 模式”由 1 家企业和 1 家科研机构共同牵头负责，目标是讨论并确定该领域研究重点，以此对项目申请书中的相关内容进行调整，同时，在欧洲范围内征募企业和科研机构作为合作伙伴，加强各组的能力建设。

（1）工业 4.0 工作组

工业 4.0 是指将工业制造和信息通信技术紧密结合，以促进效率优化和产业升级，而与之密不可分的“物联网”技术，则是在生产制造领域，使机械、生产设备和仓储系统愈加紧密地共享信息，启动操作和互相控制，将设计、生产、制造和服务之间的流程更加优化。数据海量增长是工业 4.0 革命及“物联网”发展的重要驱动力。而后者又为信息技术的进步拓展了新的空间，将 ERP 系统、数据库、互联网以及与工厂、供应链和产品相关的实时信息联系在一起。

该工作组主要基于大数据分析探索工业 4.0 的发展方向，例如，对生产性资源进行前瞻性维护，发现生产流程中的异常环节等。项目课题面向欧盟

所有研究机构和企业（包括工业制造领域的应用型企业、自动化及 IT 行业企业）。

（2）能源工作组

能源行业正在发生根本性的变革。能源转型、智能电表标准化（欧盟要求）、客户导向的新商业模式——这些变革都为 IT 基础设施带来了新挑战。借助对结构化及非结构化数据的智能分析（例如通过移动终端设备、互联网门户网站和社交媒体上的应用程序生成的数据），能源供应方可以优化交易过程，开发出新的商业模式和服务业态。因此，大数据分析可以优化消费行为预测，使供应方更好地控制能源交易。通过大数据，针对不同消费族群的收费也趋于合理化，更容易了解用户满意度，最终提高客户忠诚度。

该工作组主要基于大数据分析研究能源，例如，基于智能电表的数据、按需对能源价格进行实时调整。项目课题面向欧盟所有研究机构和企业（包括能源领域的应用型企业、自动化及 IT 行业企业）。

（3）智慧城市工作组

在城市发展与交通管理方面，大数据也拓展了新的应用可能。借助交通通信的综合解决方案及智



能交通管理系统，在快速发展的人口密集区，交通流量可以得到更有效的控制。地铁、公交车、出租车、交通摄像头等设备所产生的数据量惊人，现有的信息技术常常无法对未来的交通运输情景进行预测、模拟或进一步的数据分析。如果能对交通信息进行实时分析和正确评估，并与历史数据进行比对，就可以提前预判交通拥堵、道路危险等问题，交通流量、二氧化碳排放及行驶时间都将显著减少。

该工作组主要基于大数据分析，对城市生活领域进行研究，例如，交通控制、垃圾处理、灾害预防等。项目课题面向欧盟所有研究机构和企业（包括：相关领域的应用型企业、自动化及 IT 行业企业）。

（4）个体化医学工作组

现代医学也产生巨大的数据量，例如核磁共振等现代检测手段产生的图像分辨率越来越高，医疗技术越来越依赖信息技术控制，医疗档案越来越翔实和庞大，人类基因组相关知识也不断扩充，在个

性化治疗癌症中越来越频繁的使用软件分析临床、分子学及药理学方面的数据信息，并针对每个病人实时采取相应的有效治疗方案，大大提高治疗效果。

该工作组主要基于大数据分析，对个体化医学领域进行研究，例如，按需调节病人护理、信息化控制的临床技术、基于互联网的病人咨询等。项目课题面向欧盟所有研究机构和企业（包括医院、相关领域的应用型企业、自动化及 IT 行业企业）。

3.1.3 合作伙伴

该平台由 KIT 总体协调运营，创始伙伴包括拜耳、博世、微软（德国）、SAP、西门子、Software AG 和德国人工智能研究中心（DFKI）、弗劳恩霍夫学会、尤利希研究中心等 10 家企业和研究院所。此外，还有大众、英飞凌、通快（Trumpf）、德国信息通信新媒体协会、德国信息学会（GI）等 20 余家机构对平台建设给予支持。

3.2 SDIL 的工作流程

SDIL 的工作流程如图 2 所示。



图 2 SDIL 的工作流程<sup>[6]</sup>

### 3.2.1 上传数据源

SDIL 拟将可公开数据源和加密数据源制作成一个大型目录,因此,必须使这些数据不受任何具体科研项目时间的影响,在平台上运行。为了尽可能降低企业投入,同时保持数据源质量,SDIL 专门设置了“数据管理员”功能,协助企业上传数据并生成必要的元数据。数据上传的流程如下:

(1) 申报。数据供应方先与数据管理员取得联系,按模板填写数据源简介。管理员随后与相关数据创新工作组的牵头单位对该数据源的采用度进行审核。

(2) 分类。数据供应方在数据管理员协助下,使用元数据描写数据源,管理员随后将其分类集成至数据源目录中。同时,供应方还可设置与其所提供数据相关的问题,请研究机构运用这些数据找到解决方案。此外,供应方还要与管理员共同确定数据存取的安全级别(公开、有限公开、不公开)和生命周期(存储期限,必要时更新)。

(3) 上传数据。数据管理员授权数据供应方为 SDIL 使用者,将其数据按规定时间上传至平台,同时将供应方信息加入数据源目录<sup>①</sup>。

### 3.2.2 项目申请

数据创新工作组确定领域内的科研重点后,公布在 SDIL 门户网站上,科研机构即可按以下流程申请项目:科研单位提交项目内容简介,并与 SDIL 共同评估与数据源供应方合作的可能性。项目简介模板由 SDIL 提供,科研机构可按要求填写所需资源、时间节点以及所需数据源。每个工作组定期对计划的执行进行评估,并将平台资源分配给申请者。每年可集中评估 3~4 次。工作组会将评估通过的结果及资源分配情况通报相关科研单位,未通过评估的申请者可填写费用报销表,请 SDIL 承担相关费用。

### 3.2.3 项目执行

工作组遴选项目并配置相应的平台资源后,研发活动即可开始。研究团队应与数据源供应方密切交流。目前的平台能力可同时开展 30 个科研项目。

### 3.2.4 发布研究成果

SDIL 发布研究成果需遵循以下原则:(1) 与

某一数据源供应方提供数据相关的科研成果(如检测数据异常、安全漏洞等),仅可向供应方通报,绝不可对外公布,避免损害其企业名誉;(2) 除此之外,一切科研成果均可按惯例公开。

基于上述原则,SDIL 建立了一个大数据领域的大型知识库,包括如下内容:(1) 基于 SDIL 平台的科研工作得出的一般共性知识,如数据管理的方法和原则;(2) 科研机构基于 SDIL 平台发表的出版物目录及科研成果清单;(3) 被作为开放源代码存储于 SDIL 的相关研发活动的代码构件;(4) 介绍产学研合作的成功案例,吸引更多机构参与合作。

### 3.3 SDIL 的工作安排

SDIL 在 2014 年上半年将必要的基础设施安装完毕,相关基础设施和数据源即可向科研项目开放。其具体工作安排如下:

(1) 2014 年第 1 季度启动 4 个“数据创新工作组”;由工作组确定各领域的科研重点,并实施研究项目;工作组还将进一步吸纳欧洲范围内的产学研各界的合作伙伴加盟,以增强其研究实力。

(2) KIT 会与软件供应方负责基础设施设计及安装工具,并启动试点项目以测试 SDIL 功能。再次,KIT 牵头的“法律法规”工作组负责制定关于 SDIL 分工合作的基本合约条款,特别是数据源供应方与运营方 KIT 之间的必要协议。

(3) DFKI 牵头的“数据管理”工作组将设计出必要的理念,用元数据来描述数据源,保证数据源品质。

## 4 小结

在“大数据时代”,德国要确保其欧洲领先地位,打造“数字强国”,核心任务是紧紧围绕其信息通信技术战略“数字德国 2015”<sup>[7]</sup>和“数字议程(2014—2017)”<sup>[8]</sup>设定的目标,在政府的支持下,凝聚全欧洲产学研各界的力量,尽快提升产业竞争力。

德国借 SDIL 提出“智能数据创新”理念,立足于对海量数据的智能分析,进一步挖掘数字化的价值创造潜力,以克服自身的结构性缺陷。

<sup>①</sup> 为避免影响性能,数据通常被上传至 SDIL 的中央服务器,必要时数据源才按需上传至其它分散式存储区。

SDIL 将创新型服务作为研发重点，特别是将中小企业纳入行动范围，力求开发出实操性强、安全性高、成本低的平台服务，并紧密结合工业、交通、能源、医疗四大领域对典型项目进行孵化推广，以克服大数据在制度、技术和法律方面的障碍，完善大数据相关的法律法规和基础框架，为工业发展营造更充分的基础条件和自由发展空间，其成功模式和定位值得我国有关部门认真研究和借鉴。■

参考文献：

- [1] Die Bundesregierung. Deutschlands Zukunft Gestalten: Koalitionsvertrag zwischen CDU, CSU und SPD. 18. Legislaturperiode (2013).
- [2] BMWi. Archiv. Arbeitsgruppe 6: "Bildung und Forschung für die digitale Zukunft". Nationaler IT Gipfel 2013. <http://www.it-gipfel.de/IT-Gipfel/Navigation/archiv,did=592412.html>.
- [3] BMBF. Smart Data Innovation Lab. Karlsruhe: BMBF. 2014-01-08.
- [4] KIT. Presseinformation 004/2014. Neue Plattform für Smart-Data-Spitzenforschung[EB/OL].(2014-01-08) [2014-03-15]. [http://www.kit.edu/kit/pi\\_2014\\_14408.php](http://www.kit.edu/kit/pi_2014_14408.php).
- [5] Expertenkommission Forschung und Innovation. Gutachten zu Forschung, Innovation und Technologischer Leistungsfähigkeit Deutschlands. Berlin: EFI, 2014.
- [6] Smart Data Innovation Lab. <http://www.sdil.de>.
- [7] Federal Ministry of Economics and Technology (BMWi). ICT Strategy of the German Federal Government: Digital Germany 2015. Berlin: BMWi, 2010-11.
- [8] The Federal Government. Digital Agenda 2014–2017. Berlin: Federal Government, 2014-08.

## From Big Data to Smart Data: In Case of Smart Data Innovation Lab of Germany

SONG Song

(China Science and Technology Exchange Center, Beijing 100045)

**Abstract:** In the face of the global big data competition, Germany is characterized with the advantages of rapid growth of IT industry and IT market, world's leading IT services and software industry and a thorough grounding for the digitalization of traditional manufacturing industry, but there are obvious structural defects in it. To implement the ICT Strategies of the German Federal Government "Digital Germany 2015" and "Digital Agenda", and to enhance Germany's core competitiveness in the field of big data, the German federal government launched a new national platform for cutting-edge research — "Smart Data Innovation Lab"(SDIL) in the support of research and industry representatives. With "Smart Data" as an innovative concept proposed, SDIL aims at promoting the cooperation between industry, academy and research, developing an advanced intelligent processing of big data, and grabbing the anticipated opportunity. This report analyzes the definition and the elements of the SDIL- model.

**Key words:** Germany; big data; smart data; innovation lab