

个人信息去标识化框架及标准化

谢安明¹, 金涛², 周涛¹

1. 北京启明星辰信息技术有限公司, 北京 100081;

2. 清华大学软件学院, 北京 100084

摘要

随着大数据的发展, 个人信息保护面临严峻的安全形势, 去标识化有助于平衡个人信息保护要求和数据共享利用需求。分析了国内外去标识化研究现状, 结合我国法律有关要求和个人信息保护的迫切需求, 提出了个人信息去标识化框架, 给出了规范化的去标识化过程指南。考虑到相关工作的推进, 结合大数据安全标准体系, 提出了去标识化标准规范研究建议。

关键词

个人信息; 去标识化; 框架; 标准化

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2017048

Personal information de-identification architecture and standardization

XIE Anming¹, JIN Tao², ZHOU Tao¹

1. Beijing Venus Information Security Technology Incorporated Company, Beijing 100081, China

2. Tsinghua University School of Software, Beijing 100084, China

Abstract

With the development of big data, there are many difficulties to protect personal information. De-identification removes identifiable information from a dataset so that individual data cannot be linked with specific individuals. De-identification thus helps to balance the contradictory goals of sharing personal information while protecting privacy. De-identification architecture on personal information was proposed. The specifications on implementing a de-identification process were described. On the basis of big data security standard system, some suggestions were proposed to develop de-identification standards.

Key words

personal information, de-identification, architecture, standardization

些属性就可以共享发布,供外部业务系统进行处理分析。

1 引言

个人信息是指以电子或者其他方式记录的能够单独或者与其他信息结合识别自然人个人身份的各种信息,包括但不限于自然人的姓名、出生日期、身份证号码、个人生物识别信息、住址、电话号码等。当前,世界各国政府普遍重视大数据环境下个人信息的安全问题,都着手通过制定法律法规、标准规范和发展信息安全技术来加强对个人信息的保护。

随着信息技术和大数据应用的快速发展,各行各业数据的种类和数量呈指数增长,越来越多的人认识到了数据的价值,意识到了数据开放共享的重大意义。但数据开放共享的同时也带来了个人信息安全保护问题。政府机构、企业和其他组织收集的数据中,通常含有个人姓名、电话、证件号码等信息,如果将收集到的原始数据直接进行发布,会导致严重的个人信息泄露。如何在不泄露用户个人信息的前提下,有效开放共享数据,挖掘大数据的价值,是目前大数据研究领域的关键问题。

近年来,针对个人信息的去标识化研究获得了很多的关注。所谓去标识化,就是指去除一组可识别数据和数据主体之间关联关系的过程。通过这个过程,数据管理者可以删除或改变数据集中的标识信息,使得攻击者很难或不能利用数据集识别出具体的个人主体身份,从而可以将数据集共享到预定范围内使用。去标识化是隐私保护数据发布(privacy-preserving data publishing, PPDP)^[1]的主要工具之一,通过去除数据集中隐私属性和数据主体之间的关联关系,并且具有足够的防止重识别能力后,数据集的某

2 相关工作

2.1 有关法律标准

个人信息能够单独或者与其他信息结合识别自然人个人身份,因而价值巨大,国内外对个人信息保护的法律法规和技术研究一直是信息安全的热点内容之一。

美国通过在敏感领域(如儿童信息、医疗档案、金融数据等领域)分散立法和设立行业规范的方式,在政府引导下,规范各行业内个人信息处理行为。相关法律包括《消费者网上隐私法》《儿童网上隐私保护法》《电子通信隐私法案》《健康保险携带和责任法案》等。

欧盟通过立法来保证个人信息安全的各项基本原则和具体的法律规定。1995年,欧盟通过《个人数据保护指令》,对个人数据提出保护框架。2016年通过《一般数据保护条例(general data protection regulation, GDPR)》,以更替《个人数据保护指令》。GDPR提出了更加严格的数据保护要求,从法律层面保证了对个人信息的保护。

2016年11月,我国正式发布《中华人民共和国网络安全法》,其中就加强个人信息保护做出了规定,指出“未经被收集者同意,不得向他人提供个人信息。但是,经过处理无法识别特定个人且不能复原的除外”;同年12月发布的《国家网络安全战略》也强调要有效保护个人隐私;此外,我国《国务院关于印发促进大数据发展行动纲要的通知》《国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见》《交通运输部办公厅关于推进交

① <https://www.iso.org/standard/69373.html>

通运输行业数据资源开放共享的实施意见》等，都强调了在保证个人信息安全的前提下推动数据的开放共享。在标准制订方面，我国目前已经发布了《信息安全技术公共及商用服务信息系统个人信息保护指南》(GB/Z 28828-2012)，目前正在制订《个人信息安全规范》标准。

2.2 技术及应用规范

② <http://dx.doi.org/10.6028/NIST.IR.8053>

在标识化技术研究方面，PPDP提出在实现个人信息保护的前提下，保证数据的可用性。业界围绕PPDP进行了大量研究，包括各种隐私保护技术、模型、度量方法等。在隐私保护技术方面，周水庚等人^[2]将相关的技术分为基于数据失真的技术、基于数字加密的技术和基于限制发布的技术3种；在隐私发布模型方面，主要研究包括k-匿名^[3]、l-多样性^[4]、t-接近性^[5]、数据重发布匿名技术^[6]、差分隐私^[7]等。

③ http://csrc.nist.gov/publications/drafts/800-188/sp800_188_draft2.pdf

近年来，对去标识化应用实践和规范的研究成为个人信息保护研究热点之一，相关工作主要包括制订国际标准化组织标准、美国健康保险携带和责任法案(health insurance portability and accountability act, HIPPA)、美国国家医学院医学研究所数据共享、美国国家标准技术研究院(National Institute of Standards and Technology, NIST)研究报告、国际芯片卡标准化组织和中国银联的相关技术规范。

④ https://www.emvco.com/wp-content/plugins/pmprom-customizations/oy-getfile.php?u=/wp-content/uploads/documents/EMVCo_Payment_Tokenisation_Specification_Technical_Framework_v1.0.pdf

国际标准化组织(International Organization for Standardization, ISO)和国际电工委员会(International Electrotechnical Commission, IEC)积极展开对个人信息去标识化标准的研究，除制订隐私保护框架、保护措施等标准之外，于2016年形成ISO/IEC 1st CD 20889

草案^①。该草案重点规定了去标识化有关的术语、技术以及应用原则。

1996年，美国政府签署健康保险携带和责任法案。HIPPA在其隐私规则中给出了两种针对健康信息(protected health information, PHI)的去标识化方法：专家决定法和安全港法。其中，专家决定法以专家为主进行数据检查和确定去标识化措施，以保证将隐私泄漏风险降至最低；而安全港法则规定了包含姓名、出生日期、电话号码在内的18种特定类型数据，要求必须对这些数据进行去标识化处理。

美国国家医学院医学研究所在数据共享方面展开了研究，2015年发布临床试验数据共享研究报告，并给出了一个包括11个步骤的去标识化过程^[8]，这些步骤包括确定直接标识符、变换标识符、评估重识别风险等。

美国国家标准技术研究院于2015年发布了NISTIR 8053^②，报告总结了近20年来去标识化研究的成果，研讨了目前的实践工作，包括数据共享模型、不同类型信息的去标识化技术、重标识化风险等，并指出了未来研究方向。接着，NIST于2016年发布NIST SP 800-188第二版草案^③，为政府机构提供数据去标识化技术指导，包括建立和改进去标识化程序、去标识化的技术步骤、去标识化工具的需求以及评价去标识化工具方法等内容。

金融领域对银行卡号的去标识化应用也存在迫切需求。国际芯片卡标准化组织(EMVCo)2014年发布支付令牌化技术框架^④，提出了在支付场景中使用一个不同的号码串替换银行卡主账号的过程规范。接着，中国银联于2016年发布《中国银联支付标记化技术指引》，给出了使用支付令牌代替银行卡号进行交易验证的框架、技术要求和应用场景^[9]。

3 基础知识

个人信息去标识化中涉及一些基本概念,本节在综合有关文献的基础上给出相关定义和说明。

3.1 标识符分类

个人信息去标识化处理的原始数据集(如医疗数据、金融数据等)采取记录集的方式组成,其中每一条记录对应一个数据主体,每一条记录又包含多个属性值。在这些属性中,能够直接用于识别或结合其他信息识别出数据主体的属性包括以下两种。

- 直接标识符:利用该属性能直接识别出数据主体,如姓名、身份证号码等。
- 准标识符:仅利用该属性不能直接识别出数据主体,但联合其他属性或结合背景数据后,能识别出数据主体,如邮编、生日、性别等联合起来在数据集中就可能识别出具体的数据主体,这些属性就是准标识符。

例如,某个学习成绩数据集中,每一条记录包括{姓名,年龄,性别,班级,成绩}等属性,其中{姓名}为直接标识符,{年龄,性别,班级}为准标识符。

3.2 重标识风险

为保护个人信息安全,在对外发布数据集前,应当将数据集中的直接标识符进行删除或屏蔽等处理,并采取适当技术对准标识符进行处理,去除数据和数据主体之间的关联关系,使得攻击者不能从发布的数据中识别出数据主体。

然而,攻击者可以采用数据关联、概率推理等手段试图获知已被去除的数据主体

身份,即把去标识化的数据集重新关联到原始数据主体,这个过程称为重标识或重标识攻击。

重标识风险是指攻击者成功实施重标识攻击的风险,即使数据集已经被最新的去标识化技术处理过,仍然存在重标识风险。

进一步地,攻击者除了能够访问已经发布的数据集外,还可能从其他渠道获得一些关于目标对象的信息,如部分数据主体的出生日期、出生地、邮编等,甚至知道数据集去标识化采用的技术方法,这些攻击者可能获知的关于发布数据的信息统称为背景知识。

因为数据集中存在有意义的数据,可以为应用带来价值,所以需要对数据集进行去标识化处理后才能发布。数据集去标识化后,应当保持数据集的这种有具体含义、有使用意义的特性,这种特性即有用性。

3.3 去标识化原则

对数据集实施去标识化,应遵循以下原则。

- 合法合规:满足我国法律法规、标准规范以及双方约定对个人信息安全保护的有关规定。
- 安全优先:实施去标识化,应优先考虑保护个人信息安全,其次尽量保留数据集的有用性。
- 技管结合:综合利用技术和管理两方面措施实施,达到最佳效果。
- 持续改进:应持续改进方法、技术和工具,定期重新实施重标识风险评估,保证数据安全。

4 去标识化框架

本文提出的去标识化框架包含3个部

分：管理保障、去标识化过程和技术支撑，如图1所示。

该去标识化框架以管理手段和技术机制为保障支撑，旨在规范去标识化的实施步骤，即给出数据集实施去标识化时应遵循的规范流程，以达到安全可靠的去标识化目的。

4.1 去标识化过程

去标识化过程主要包括确定目标、识别标识、处理标识和导出数据4个大步骤，其中，处理标识又可以分为测试验证、转换数据和评估数据3个细分步骤。

4.1.1 确定目标

确定目标的主要工作包括确定去标识化对象、建立安全目标和制定工作计划等内容。

- (1) 确定去标识化对象
- 指确定需要去标识化的数据集范围。在目标数据集中存在标识符时，宜根据法规标准、组织策略、数据承诺、业务背景、

数据用途以及历史发布情况等要素，确定哪些数据子集属于去标识化对象。

- (2) 建立安全目标
- 应根据发布后的应用环境和数据用途，确定重标识风险阈值和数据有用性的最低要求，并在两者之间取得平衡，以有利于控制安全风险。

- (3) 制定工作计划
- 制定个人信息去标识化的实施计划，包括去标识化的目的、安全目标、数据对象、公开共享方式、实施团队、实施方案、利益相关方、应急措施以及进度安排等，形成去标识化实施计划书。

4.1.2 识别标识

确定目标后，应识别出数据集中存在哪些直接标识符和准标识符。识别的方法包括查表识别法、规则分析法和专家判定法。

- (1) 查表识别法
- 预先建立元数据表格，存储需去标识化的直接标识符和准标识符名称，在识别标识数据时，将待识别数据的各个属性名称或字段名称逐个与元数据表中的记录进行对比，以此识别出标识数据。查表识别法适用于数据集格式和属性已经明确的去标识化场景，如采用关系型数据库，在表结构中已经明确姓名、身份证号等标识符字段。

- (2) 规则分析法
- 通过建立自动化程序，分析数据集规律，确立相关标识符识别规则，从中自动发现需去标识化的直接标识符和准标识符。依据相关标准，可以建立身份证号码、银行卡号、邮件地址、姓名、性别等标识符识别规则。规则分析法适用于应用系统未采取规范方式存储数据的情形，例如：未采用常用的字段名称，如使用“备注”字段存储身份证号；数据中存在混乱或错误情

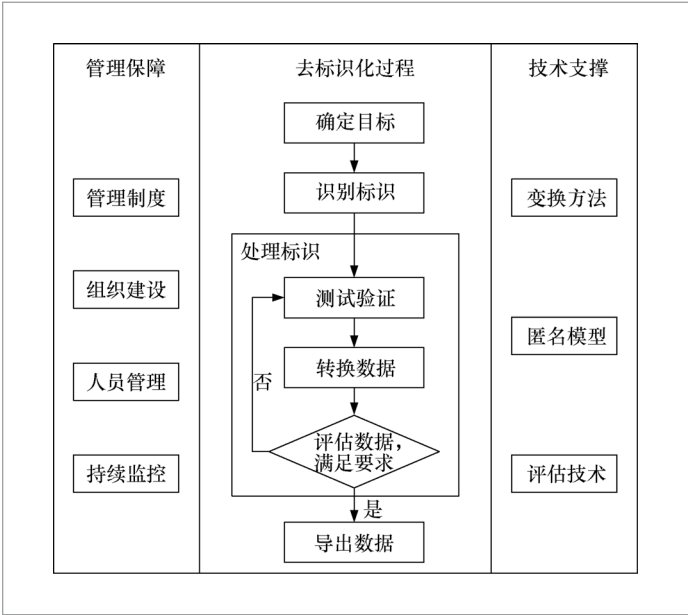


图1 去标识化框架

况,如该“备注”字段前100条记录的值为空,而后10 000条记录的值为用户身份证号码;采用非结构化方式存储数据时,如构建身份证号码识别规则后,可从司法判决书中自动识别出所有的身份证号码。

(3) 专家判定法

通过专家审查,人工发现和确定数据集中的直接标识符和准标识符。专家可在对业务处理、数据集结构、相互依赖关系和背景知识等要素分析的基础上,综合判断数据集重标识风险,直接指定数据集中需要去标识化的标识符。在下列场景时,专家分析法具有明显的优势:数据集中的多个不同数据子集之间存在关联、引用关系时,如通过数据挖掘算法,可联合分析数据集中多个非常见标识符属性后,识别出唯一的用户身份;数据集中有特别含义的数据,或者数据具有特殊值和容易引起注意的值,从而可能被用来重标识时,如超出常人的身高、独特的地理坐标、罕见的病因等。

4.1.3 测试验证

在正式对数据集去标识化之前,应当从数据集中抽取部分数据,形成一个较小的数据子集,采用选择的去标识化模型和技术对该子集数据进行测试,检验去标识化效果。测试验证包括预处理、选择模型和方法、实施去标识化、风险计算4个阶段工作。

(1) 预处理

一般地,预处理是对数据集施加某种变化,形成规范化格式的数据,使其有利于进行后期处理。应根据数据集的实际情况选择预处理措施,或选择不实施预处理。

(2) 选择模型和方法

不同类型的数据需要采用不同的去标

识化技术,所以在去标识化之前,应选择合适的去标识化模型和技术。选择的参考因素包括如下方面:是否可删除、是否需要保持唯一性、是否需要满足可逆性、是否需要保持原有数据值顺序、是否需要保持原有数据格式、是否需要保持统计特征、是否需要保持关系型数据库中的实体完整性、参照完整性或用户自定义完整性等。一般地,对直接标识符要优先选择删除或屏蔽的处理方式。

(3) 实施去标识化

根据选择的去标识化模型和技术,对子集数据实施操作。一般地,按照直接标识符、准标识符、敏感属性的顺序实施去标识化。

(4) 风险计算

查看测试结果数据集,判断数据可用性,计算重标识风险,并与原先设定的风险阈值进行比较,如果满足要求则继续往下执行,否则重新选择模型和方法,并再次测试和评估。

测试验证步骤是一个可选的步骤。在数据集较大、实施去标识化变换较慢时应该选择本步骤。

4.1.4 转换数据

通过第4.1.3节的测试验证过程后,即可应用该选定的去标识化模型和方法,设定去标识化参数,对整体数据集实施去标识化操作。本步骤包括预处理和实施去标识化两个阶段的工作,工作过程和第4.1.3节类似。

4.1.5 评估数据

查看整体数据集去标识化结果,判断数据可用性,并计算重标识风险。将重标识风险与原先设定的风险阈值进行比较,如果满足要求则继续往下执行,否则重新

回到测试验证阶段,并再次测试和评估。

4.1.6 导出数据

最后,在实施机构的管理层批准后,将去标识化处理之后的数据集导出,提供给数据发布系统。

4.2 技术支撑

实现去标识化需要依赖于相关技术的支撑。目前,相关的研究工作已经提出了多种去标识化变换方法、匿名模型和数据评估方法。

(1) 变换方法

对涉及个人信息的标识符数据项进行变换,使用变换后的数据项值代替原有的值,即可达到一定的个人信息保护效果。常用的变换方法包括屏蔽、随机、泛化、加密等几种,见表1。

此外,还有很多去标识化方法融合了其他多种方法和技术机制,因此,很难将其简单地划分到以上某一类中。

(2) 匿名模型

由于攻击者在获取去标识化数据集和相关背景知识后,可能通过关联、推理、链接等攻击手段恢复已去除的主体身份信

息^[8],为进一步提高个人信息保护能力,研究人员提出了多个匿名模型指导数据去标识化和发布,如 k -匿名、 l -多样性、 t -接近性、数据重发布匿名技术、 m -不变性、差分隐私等。

其中, k -匿名模型指在发布去标识化后的数据集时,要求其指定标识属性值相同的数据中,每一等价类至少包含 k 个记录,即对于指定标识属性的任何一个数据值,至少能找到 k 条记录具有同样的值;而 l -多样性模型是在 k 匿名的基础上,又要求每个等价类中敏感属性的值至少有 l 个不同的取值。 l -多样性模型避免了一个等价类中敏感属性取值单一的情况,进一步降低了数据集被重标识攻击的风险。

(3) 数据评估方法

为确保去标识化后的数据数量,还需要相应的评估手段对结果数据集进行个人信息泄漏风险和数据可用性方面的评估。其中,风险度量是指对数据集进行个人信息泄漏风险度量,评估去标识化实施的安全保护效果。现有的风险度量方法中,多采用基于方差统计、信息熵、匿名化程度、披露风险等方法度量信息泄漏风险大小。而可用性度量是指评估衡量去标识化后数据集的可用性,即数据集被去标识化处理

表1 个人信息去标识化方法说明

方法	描述	举例
屏蔽	对标识符数据项进行抑制处理,对其进行删除或者隐藏。屏蔽可以针对整个数据项进行,也可以选择对数据项的一部分进行	屏蔽身份证号“440524188001010014”时,可选择直接删除,也可使用“440524*****0014”代替
随机	使用随机产生或分配的数据代替原来的数据项,随机方法可以包括噪声添加、完全随机产生、数据项重排置换等	中文姓名使用随机生成的姓和汉字表示,如使用随机生成的“辰筹猎”代替“张三丰”
泛化	通过降低数据精度,使用概括、抽象的办法表示原有的数据项。对于数值型数据项,可以使用取整、取最大值等方法对数据进行泛化	如实数数据“1.732”可以泛化为“1”;如“张三”可以泛化为“张某某”
加密	采用密码学方法对数据项进行变换,包括对称加密、非对称加密和杂凑运算等。如果需要保留原有数据项的某些特性,还可以使用保序加密或保留格式加密等算法	如身高“1.73”可以加密为“1.46”

后，数据集满足业务系统需求的程度，一般采用和原始数据集之间的数据差异度、敏感属性距离、关键信息缺损度等指标评估。

4.3 管理保障

含有个人信息的数据集如果没有经过合适的处理，发布后可能造成严重的信息泄漏，甚至导致严重的安全事件发生。在信息安全保障实践中，仅靠信息安全技术是不够的，良好的管理能预防、阻止或减少信息安全事件的发生，我国也提出了管理与技术并重的指导方针。在本文提出的去标识化框架中，针对管理保障，提出了要在管理制度、组织建设、人员管理、持续监控等几个方面开展工作。

(1) 管理制度

管理制度主要是对去标识化工作流程和管理职责进行规范和约束，需要制定相应的管理制度并执行。对生产、开发、测试和其他非生产环境以及外包环境中的业务数据的去标识化工作进行规范，包括去标识化的适用范围、管理对象、岗位及职责、敏感信息、去标识化场景、工作流程等。

(2) 组织建设

个人信息去标识化是一件严肃的工作，应当设置相应的管理岗位，明确工作职责，保证个人信息去标识化工作安全开展。除了最高管理层外，相关的职能部门包括业务部门、数据管理部门和管理监督部门等。其中，业务部门指因业务活动要求而需要使用数据的部门；数据管理部门指负责数据采集、去标识化以及销毁的部门，负责数据管理的日常管理事务；管理监督部门指为了保障整个数据安全管理体系良好运转而设立的数据监管部门。上述部门间在实施去标识化工作时的主要工作流程如图2所示。

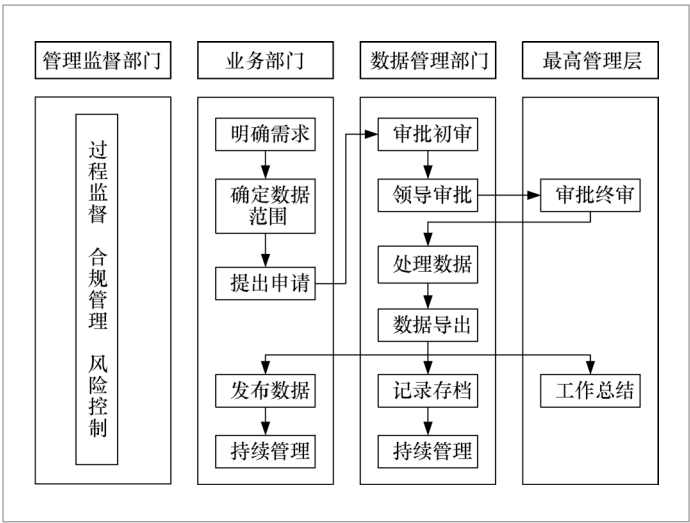


图2 去标识化业务流程

(3) 人员管理

为了确保个人信息去标识化工作的顺利开展，对相关岗位人员应提出知识、技能以及安全意识方面的要求。因此应当在岗位需求、人员招聘、人员培训、离职管理以及保密协议方面制定相应的要求和条款。

(4) 持续监控

数据集在去标识化并发布后，随着时间的推移，原来计算重识别风险的要素可能会发生变化，如重标识攻击能力升级、目标环境变化、发现新的背景知识以及出现新增去标识化数据集等。此时，应再次进行重标识风险评估，并与预期可接受风险阈值进行比较，以采取更好的数据保护措施，更好地保障个人信息安全。

5 标准化建议

随着大数据时代的到来，数据交换共享的需求越来越旺盛。为了更好地实现个人信息保护，许多机构和企业已经开始采用相关的方法和模型对数据集实施去标识化。

标准化有利于实现科学管理,提高管理效率,推动建立最佳秩序。从业界实际工作中提炼出最佳实践,形成指南规范,可以更好地指引和规范去标识化工作,从而有效推动个人信息安全保护。本文从第4节的去标识化框架出发,结合技术研究和最佳实践,对个人信息去标识化的标准规范研究提出如下建议。

(1) 积极推动标准建设

跟进大数据的发展,围绕我国大数据安全标准化体系规划,结合个人信息安全保护要求和去标识化标准需求,积极推动个人信息去标识化标准化工作开展,制订和推广应用相关标准,以更好地规范数据共享和处理等相关生产活动,保障数据安全共享利用。

(2) 尽快规范业界实践

采纳产业界在去标识化方面的应用和实践成果,优先将通用、成熟的成果转化为标准规范,以有效提高产品、工程和服务的质量。其中,综合产业界实践成果,并由清华大学牵头制订的标准《信息安全技术 个人信息去标识化指南》已经进入公开征求意见阶段。此外,还应继续将相关产业界实践提升为标准规范,以引导企业安全、规范实施去标识化工作,避免数据处理乱象。

(3) 引导应用先进技术

吸收国内外在去标识化技术方面的最新研究成果,对成熟、适用和急需的去标识化技术机制、模型以及评估方法进行标准化,引领企业采用先进技术开展个人信息保护工作,有效执行个人信息披露风险度量,确保提升信息安全防护能力。如可综合国内外技术研究成果,推动保序加密和保留格式加密等技术标准制订工作。

(4) 全面推进企业规范

良好的管理能够让技术更好地发挥作用。企业应在业务需求和整体安全策略的指导下,积极应用国家标准,并和管理实情

相结合,建立个人信息去标识化工作规范,包括管理制度、流程规范、运维管控、建设方式和实施规范等方面,全方位保障个人信息安全防护,推进数据高效利用。

6 结束语

大数据时代是机遇与挑战并存的时代,在大数据产生巨大价值的同时,也为个人信息保护带来了难题。个人信息去标识化是个人信息保护研究领域的一个分支,也是近年来新兴的研究方向。目前在去标识化方法、模型和评估方面已经取得了不少成果。本文针对规范化开展个人信息去标识化工作的迫切需求,结合去标识化技术的特点,提出包括去标识化过程、技术支撑和管理保障在内的去标识化框架,给出了规范化的去标识化过程指南,并就去标识化的标准规范工作提出了建议。在未来工作中,可以以该架构为指导,进行具体案例的实施,积极推进有关标准化工作,具有良好的应用前景。

参考文献:

- [1] FUNG B C M, WANG K, CHEN R, et al. Privacy-preserving data publishing: a survey on recent developments[J]. ACM Computing Surveys, 2010, 42(4): 14.
- [2] 周水庚, 李丰, 陶宇飞, 等. 面向数据库应用的隐私保护研究综述[J]. 计算机学报, 2009, 32(5): 847-861.
ZHOU S G, LI F, TAO Y F, et al. Privacy preservation in database applications: a survey[J]. Chinese Journal of Computers, 2009, 32(5): 847-861.
- [3] SWEENEY L. Achieving k-anonymity privacy protection using generalization and suppression[J]. International Journal of

- Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 571-588.
- [4] MACHANAVAJJHALA A, KIFER D, GEHRKE J, et al. l-Diversity: privacy beyond k-anonymity[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 24.
- [5] LI N H, LI T C, VENKATASUBRAMANIAN S. t-Closeness: privacy beyond k-anonymity and l-diversity[C]// The 23rd International Conference on Data Engineering, April 15-20, 2007, Istanbul, Turkey. New Jersey: IEEE Press, 2007: 106-115.
- [6] BYUN J W, SOHN Y, BERTINO E, et al. Secure anonymization for incremental datasets[C]//International Conference on Secure Data Management, September 10-11, 2006, Seoul, Korea. Berlin: Springer-Verlag, 2006: 48-63.
- [7] CYNTHIA D. Differential privacy[C]// The 33rd International Colloquium on Automata, Languages and Programming, July 9-16, 2006, Venice, Italy. Berlin: Springer-Verlag, 2006: 1-12.
- [8] Institute of Medicine of the National Academies. Sharing clinical trial data: maximizing benefit, minimizing risks[M]. Wasting DC: The National Academies Press, 2015: 195-196.
- [9] 周明. 支付标记化技术解读[J]. 金卡工程, 2015(9): 54-56.
- ZHOU M. Interpretation of payment markup Technology [J]. Cards World, 2015(9): 54-56.

作者简介



谢安明 (1977-), 男, 博士, 北京启明星辰信息安全技术有限公司核心研究院研究员, 主要研究方向为大数据安全、密码应用和区块链技术, 参与编写《大数据安全标准化白皮书(2017)》《大数据服务安全能力要求》《个人信息去标识化指南》等标准。



金涛 (1980-), 男, 博士, 清华大学软件学院讲师, 全国信息安全标准化技术委员会大数据安全标准特别工作组秘书, 主要研究方向为流程管理、大数据、大数据安全标准。



周涛 (1979-), 男, 博士, 北京启明星辰信息安全技术有限公司教授级高级工程师、助理总裁, 核心研究院院长, 主要研究方向为大数据安全分析、事件关联分析、入侵检测等。

收稿日期: 2017-08-06

通信作者: 金涛, jintao16@tsinghua.edu.cn