图书馆大数据平台敏感数据保护研究

陈臣
（兰州商学院信息中心，兰州 730020）

摘要：首先界定了敏感数据的内容，并概述了敏感数据的学术研究现状；进而对图书馆敏感数据的来源及其面临的安全性问题进行分析总结。提出了从大数据采集、存储、服务和应用四个层面进行敏感数据保护的基本思路，以及图书馆大数据平台敏感数据的安全管理策略。

关键词：图书馆；大数据平台；敏感数据；保护
中图分类号：G250.76
DOI：10.3772/j.issn.1673-2286.2015.07.010

1 前言

伴随大数据时代的到来，大数据已成为图书馆准确获取知识、预测事物发展趋势、掌握读者个性化特征和分析辨识事物真相的科学依据。大数据具有海量、高速、多样和价值的特点[1]。随着图书馆服务模式和读者阅读方式的改变，图书馆的数据总量和数据类型呈现快速增长的态势，大幅增加了图书馆数据管理、分析和决策的难度。此外，图书馆在数据中心基础设施构建、用户服务模式组织、读者客户关系管理、服务市场竞争环境分析中会涉及敏感数据，这些敏感数据是关系图书馆服务可信、可持续发展和读者阅读满意度的关键因素。图书馆应实现敏感数据在采集、存储、处理、分析和决策过程的安全管理，才能确保敏感数据安全、可用和可控。

2 敏感数据研究现状分析

敏感数据主要指当其被窃取、传播、不当使用、未经授权被他人存储与共享、非法数据价值挖掘后，可能会对国家、政府、企业和数据所有人造成严重侵害的数据。我国许多学者对敏感数据的保护、匿名发布、安全管理和远程传输等问题进行了深入的研究。余永红等提出基于分布式安全数据库服务的隐私保护方法，较好地平衡了数据库查询性能与隐私保护之间的矛盾[2]。骆永成对数字图书馆敏感数据匿名发布所涉及的若干关键技术进行了研究，指出通常的数据匿名发布技术，不足以解决数字图书馆敏感数据发布多种场景下的隐私保护问题[3]。闫玺玺对开放网络环境下敏感数据的安全与防泄密技术进行了研究，认为敏感数据在生成、存储、传输、分发、共享等各个环节下均应处于安全状态[4]。张元等对涉密应用系统敏感数据的远程传输方法进行了研究，认为利用二维条码传真方式传递敏感数据，利用技术手段对数据进行加密处理，在保证数据安全的前提下，提高了数据传输的时效性[5]。刘明辉等研究了云环境下敏感数据所面临的安全风险，分析了云环境下的敏感数据的安全需求，给出了云环境下的敏感数据保护的技术方案[6]。

3 图书馆敏感数据来源与其面临的安全性分析

图书馆的敏感数据主要来自于图书馆信息化设施构建、服务系统运营监控、系统安全性和可用性审计、读者个体特征与阅读行为数据采集、视频监控与传感器网络数据采集、个性化服务推送，以及与第三方服务商的敏感数据共享等方面，主要涉及图书馆的安全运营管理、企业商业秘密、知识产权保护、关键业务信

息、业务合作方信息、读者个体特征和隐私、系统管理员与读者的账号和密码、读者阅读行为记录、读者个体位置信息等。

随着大数据时代的到来，图书馆敏感数据的安全性面临了一些新的问题，表现在以下三个方面。

3.1 读者敏感数据的所有权和隐私权易受侵犯

隐私权是自然人享有的对其个人与公共利益无关的个人信息、私人活动和私有领域进行支配的一种人格权。读者敏感数据主要包括读者的个体特征、阅读行为、阅读社会关系、阅读账号和密码、个体地理位置、移动路径、阅读终端参数和阅读方式等。读者敏感数据具有强烈的读者身份和人格属性，读者对个人敏感数据的所有权和隐私权有迫切的保护需求。图书馆在为读者提供大数据服务的同时，实际上也获取了读者敏感数据的采集权、使用权、经营权和传播权。如何将敏感数据的所有权和使用权进行科学划分，是保护读者隐私权和提高图书馆大数据应用效率的重要问题。同时，图书馆大数据决策过程中的读者个人隐私泄漏、隐私数据非法共享，也是影响读者敏感数据所有权和隐私权保护的另一个重要问题[7]。

3.2 图书馆敏感数据面临更多的安全威胁

随着黑客技术的发展，高级持续性威胁成为黑客获取图书馆信息系统控制权限和窃取敏感数据的主要手段。在发动攻击之前，黑客会通过精确收集图书馆的业务流程和目标系统的安全漏洞，主动挖掘被攻击对象系统和应用程序的漏洞。在隐匿自身攻击方法和行为的同时，对图书馆敏感数据长期、有计划地窃取、篡改和伪造。其次，图书馆敏感数据在多系统中零散分布，图书馆难以准确判断敏感数据所处的系统位置、控制对象、访问权限和移动路径，难以借助第三方安全管理系统实现敏感数据的访问控制、入侵防御和补丁管理[8]。最后，图书馆网络内部的非法用户、病毒机、管理员误操作给敏感数据的安全管理带来了严重隐患。

3.3 读者隐私数据的权利边界消失

伴随大数据应用的深入，读者敏感数据呈现网络化、透明化和共享化的趋势，导致读者对隐私数据的可控性和权利边界消失，隐私数据易于被截获、窃取、篡改和非法使用。当读者隐私数据被采集和标准化之后，便以数据化形式被存储在图书馆的大数据库中，图书馆通常会依据读者个性化服务和维护企业利益需要去访问、挖掘这些数据。此外，如果这些数据涉及到国家的安全，政府相关部门也会通过“立法”等合法途径，对个人敏感数据进行监控和检查。大数据时代，图书馆成为读者敏感数据的采集、存储、管理和使用者。传统的图书馆安全保护方法已失效，读者丧失了个人敏感数据被使用的知情权、编辑与管理权、删除不当权等，无法有效控制个人敏感数据的采集、存储、共享和传播。同时，图书馆采集的读者数据具有多源头、多对象、多类型和碎片化的特点，单一信息可能不会泄露读者的“隐私”，如果将若干个信息进行整合、重组、挖掘和分析，可能会获得读者的个人敏感数据[9]。

4 图书馆大数据平台敏感数据保护思路与管理策略

4.1 图书馆大数据平台敏感数据保护思路

对于图书大数据平台敏感数据保护应用从数据采集、存储、服务和应用四个层面进行，其保护思路见图1。

大数据采集层负责数据的采集、分类、整理和传

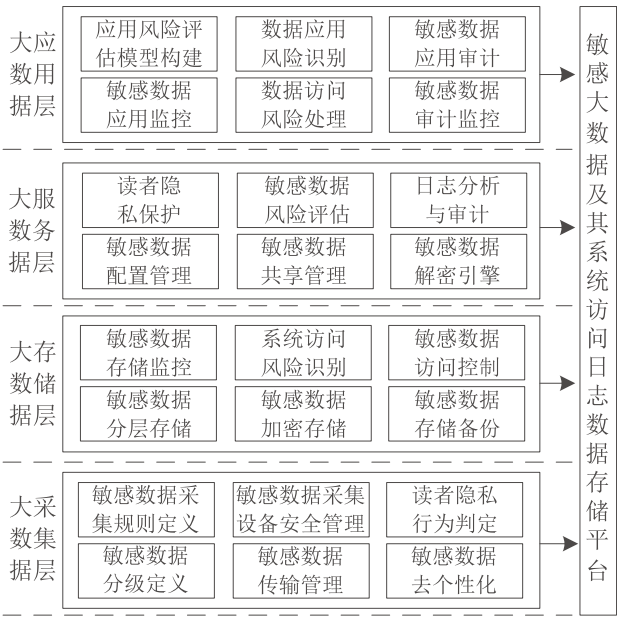

图1 图书馆大数据平台敏感数据保护思路

输。该层通过对敏感数据的采集规则定义、采集终端设备安全管理、读者隐私行为判定、敏感数据分级定义、敏感数据的传输管理、敏感数据的个性化特征去除等方式，来保护数据采集层的安全。

大数据存储层负责敏感数据的安全存储与管理。该层通过对敏感数据存储的实时监控、存储风险管理、访问控制、分层与加密存储、安全备份等方式，来确保敏感数据的安全存储。

大数据服务层基于大数据存储层的支持，为大数据应用层提供数据分析和决策。该层通过对敏感数据风险评估、读者隐私数据保护、敏感数据配置管理、敏感数据共享管理、敏感数据应用日志分析等，来实现大数据服务层的安全管理。

大数据应用层负责图书馆大数据的分析、判断和决策，通过安全威胁识别、安全风险评估、应用过程的安全监控、敏感数据的安全审计、数据访问风险的处理等，来保证敏感数据的应用安全。

4.2 图书馆大数据平台敏感数据的安全管理策略

4.2.1 图书馆大数据应采用分层存储与管理的策略

首先，图书馆应依据大数据的数量、敏感性，将大数据划分为不同的安全等级，并按照安全级别将大数据分流存储于大数据平台和敏感数据存储系统上，在敏感数据存储系统上执行更加安全、智能和自动化的大数据存储策略，为敏感数据生命周期全程提供安全的保障。其次，对于和图书馆大数据决策、读者个性化服务无关的读者特征数据，应采取敏感信息的数据脱敏和去除个性化特征等措施，在不降低大数据价值总量、可用性和可控性的前提下，降低读者隐私和敏感数据被侵犯的风险。第三，图书馆应通过访问者身份限制和访问权限管理的方式，对系统管理员、图书馆员和读者等用户进行身份和权限管理，严格控制上述人员访问大数据的权限、内容、种类、方式和数量。

4.2.2 大数据备份与灾难恢复平台建设应坚持高效、可控的原则

首先，图书馆安全防御系统应具有对新型威胁和新攻击方式数据的精确收集和识别能力，可确保敏感数据在生命周期全程安全、可控和可用。其次，敏感数据灾难恢复平台应注重与虚拟化环境的集成，实现重复数据删除、敏感数据深度压缩、敏感数据优先级备份、映像快照的持续保护、细颗粒度的单一对象数据恢复能力。第三，敏感数据灾难恢复平台还应支持多种操作系统和兼容多应用软件，提升敏感数据备份与灾难恢复过程数据传输的效率，并大幅度降低数据备份和灾难恢复的时间窗口，实现敏感数据的瞬间一键备份和恢复[10]。

4.2.3 加强读者隐私数据的保护

首先，读者应与图书馆签署个人数据采集、共享、传播和应用隐私保护协议的方式，确定读者完全拥有个人大数据资源的所有权、知情权和选择权，并由读者决定个人数据采集、使用、公开和删除的内容、程度、权限和时限。其次，图书馆在对读者隐私数据进行采集和使用时，必须向读者展示其隐私数据采集、使用、共享、删除的对象、途径、方法和内容，并将敏感数据管理平台的管理权交付读者。同时，读者可登陆大数据管理平台并以自助管理的方式，实时对图书馆采集、访问、存储和传播自身敏感数据，以及对自身敏感数据分析、决策和应用的程度进行动态管理，确保个人敏感数据不被侵犯和滥用。第三，图书馆还应结合自身敏感数据的应用环境和生命周期特点，通过敏感数据发布匿名保护、社交网络匿名保护、数据水印、数据溯源、风险自适应的数据访问控制、数据内容的可信验证等技术，保证敏感数据使用安全、可控和可用。

4.2.4 利用大数据技术加强图书馆敏感数据的保护

随着数据中心系统负荷和用户需求的不断增长，图书馆每天会产生海量的服务器运行、安全监控、访问者身份与行为记录、安全日志等大数据信息。这些大数据资源科学反映了图书馆的系统运行状况、安全威胁方式与内容、访问者危险行为和安全防御系统的漏洞，可为图书馆安全防御系统建设和敏感数据保护提供可靠的大数据决策支持。通过对长时间海量安全数据的分析，可发现图书馆在服务系统组织、安全运营管理、访

问者行为规范和安全防御系统可靠性等方面存在的问题，完成对未知风险、攻击威胁的精确预测和科学防御。其次，基于对访问者行为和访问设备特征数据的分析，图书馆可精确判断访问者的身份、访问目的和访问方式，有效防止恶意用户通过破解用户密码和伪造设备标识，而非法获得敏感数据的访问权限。此外，基于大数据的用户身份认证能够在网络全局对访问者进行认证，实现不同管理、服务系统的统一身份认证，大幅提高用户身份、密码认证效率和降低系统资源的损耗[11]。

5 结语

图书馆大数据为图书馆服务模式变革、服务能力提升、读者个性化服务需求发现、服务的可持续发展提供可靠的大数据支持。但是，图书馆也面临着诸如大数据总量增长、大数据环境和数据结构复杂、大数据安全威胁增多和敏感数据易受侵犯的问题。如何构建科学、高效的安全防御体系，防止图书馆敏感数据被侵犯，是关系图书馆服务可信度和读者阅读满意度的重要问题。图书馆不断提升敏感数据保护的水平，为图书馆用户服务提供可靠的大数据决策支持。

参考文献

[1] 和婷.大数据思维对图书馆信息服务工作的启示[J].图书馆建设,2014(1):64-68.

[2] 余永红,柏文阳. 基于敏感数据加密的分布式安全数据库服务研究[J].计算机应用研究,2010,27(9):3510-3513.

[3] 骆永成. 数字图书馆敏感数据匿名发布若干关键技术研究[D].上海:东华大学,2011.

[4] 闫玺玺. 开放网络环境下敏感数据安全与防泄密关键技术研究[D].北京:北京邮电大学,2012.

[5] 张元,朱之贞,胡建明,等. 涉密应用系统敏感数据远程传输方法的研究[J]. 云南大学学报(自然科学版),2013,35(S2):152-155.

[6] 刘明辉,张尼,张云勇,等. 云环境下的敏感数据保护技术研究[J].电信科学,2014(11):2-8.

[7] 马晓亭. 大数据时代图书馆个性化服务读者隐私保护研究[J].图书馆论坛,2014(2):84-89.

[8] 江颉,顾祝燕,高俊骁,等. 基于敏感等级的云租户数据安全保护模型研究[J]. 系统工程理论与实践,2014,34(9):2392-2401.

[9] 林伟胜,陈森利,许卓伟,等. 保护隐私的轻量级云数据共享方案[J].计算机应用研究,2015,32(8):11-17.

[10] LiuBing.Web数据挖掘[M].余勇,薛贵荣,韩定一,译. 北京:清华大学出版社,2009:327-329.

[11] 李国杰. 大数据研究的科学价值[J].中国计算机学会通讯,2012,8(9):8-15.

作者简介

陈臣，男，1974年生，副教授，研究方向：大数据、数字图书馆建设，E-mail：chen_4325@126.com。

Research on Sensitive Data Protection for Big-Data Platform of Library

CHEN Chen
（Information Center, Lanzhou University of Finance and Economics, Lanzhou 730020, China）

Abstract: Firstly, the scope of sensitive information of big-data platform of library was discussed, and the general situation about sensitive data was analyzed in this paper. And then, we indicated the source of the sensitive data, and the safety risk of sensitive data of library was carried on the summary analysis. Finally, the basic idea of sensitive data protection and the sensitive data security management strategy for library was proposed in this paper.

Keywords: Library; Big Data Platform; Sensitive Data; Protection

（收稿日期：2015-04-26；编辑：雷雪）