

数字对象使用及重用评估的 伦理考量及准则

由数字内容重用评估框架工具包（D-CRAFT）[项目组](#)和[顾问团](#)编拟

2023 年 6 月定稿

引言

美术馆（Galleries）、图书馆（Libraries）、档案馆（Archives）、博物馆（Museums）和资料库（Repositories）等机构（合称 GLAMR）的当前工作必须基于对身份认同、社会经济特权及白人特权、可访问性、隐私以及其他重要标准的批判性认识展开。伦理考量应当规范从业人员、用户以及可能创建或捐赠数字内容的社群之间的相互交流。这些考量可引导各方开展积极协作，增强所有利益相关者的权能。

数字内容重用评估框架工具包（Digital Content Reuse Assessment Framework Toolkit，简称 D-CRAFT）项目组（以下简称“项目组”）编拟了《数字对象使用及重用评估的伦理考量及准则》（以下简称《准则》）。D-CRAFT 是由博物馆和图书馆服务协会（Institute of Museum and Library Services，简称 IMLS）联邦拨款资助多年的项目，旨在开发资源，为可持续地衡量、评估文化遗产知识组织所持有的数字资产的重用情况提供建议做法^[1]。2017 至 2018 年间，项目组使用 IMLS 的首笔资助创建了项目“开发数字对象重用衡量框架”，成功地为这一工具包创造了使用案例，为美术馆、图书馆、档案馆、博物馆和资料库所持有的数字资产的重用评估提供了资源和建议做法^[2]。D-CRAFT 充分借鉴参考了前述项目。

项目组编拟本《准则》旨在重点考量在 GLAMR 行业工作并受其影响的人士所产生的社会和政治影响，以及在工作流程中统一道德观和价值观、明确问责制的重要性。当今世界，由 COVID-19 引发的全球大流行对黑人和棕色人种群体的冲击尤为严重，美国旷日持久的白人至上主义愈演愈烈，导致乔治·弗洛伊德和其他许多黑人惨遭谋杀，社会显性和隐性结构继续压迫、边缘化、非人化那些历史形成和新近出现的被弱势化群体，项目组和从业人员必须认识到我们在这些结构中的集体角色^[3]。

¹ IMLS 资助项目，“D-CRAFT”：<https://www.ims.gov/grants/awarded/lq-36-19-0036-19>

² IMLS 资助项目：“衡量重用”：<https://www.ims.gov/grants/awarded/lq-73-17-0002-17>；“衡量重用，2017-2018”：<https://reuse.diglib.org/project-history/>

³ 本文使用“历史形成和新近出现的被弱势化群体”这一术语来代表广泛多样的个体、身份和社群。最初，项目团队使用术语“黑人、土著人和有色人种（BIPOC）”来代指相关群体。有些人对 BIPOC 一词提出质疑，认为它没有明确地包括特定群体，或者说是在“区别对待”黑人和土著人社群，而不是以其为中心。

就压迫历史形成和新近出现的被弱势化群体而言，GLAMR 机构不乏相关事例，包括：选择、获取、组织和创建实体和数字馆藏时以白人中心；参与构建充斥白人至上主义、殖民主义和瑕疵的知识组织体系并使其延续；以及排他性的数字化实践导致了充满偏见的决策以及歧视性的馆藏数字化^[4]。出于以上原因，本《准则》的制定就是为了在推动其发展的 D-CRAFT 资助项目之外发挥作用^[5]。

制定伦理准则的目标

本《准则》面向的是评估数字对象的使用及重用的从业人员。本《准则》所指的“从业人员”系指在数字图书馆、数字资料库、数字博物馆藏、数字展览、数据存储库或数字档案馆等数字环境中整理、维护或保存数字对象的个人。

本《准则》旨在为从业人员在使用和重用评估工作中提供决策参考。创建这些准则的关键在于充分考量用户隐私，并重点关注历史形成和新近出现的被弱势化群体的关切事项和想法。本《准则》侧重于评估，而非应当如何合乎道德地使用、重用数字内容。

遵循本《准则》进行评估，从业人员可能需要抛弃那些不以隐私保护、包容和公平为中心的评估策略。尤其是新手，他们可能很难在 GLAMR 机构中适应，因为这些机构已经根据特定的评估方法创建了项目和馆藏。所以本《准则》试图让从业人员重新审视当前的实践策略，以更开放的姿态接受新的考量因素，并在其机构中倡导任何形式的变革。

本《准则》为数字图书馆从业人员以及管理 GLAMR 社群所托管、聚合、策划的内容的人士提供了指引，还搭建了对机构施加积极影响的框架。本文中所介绍的事例并不代表整个 GLAMR 行业。

另一些人则认为这个词充满包容性。为确保项目组和本《准则》尽可能地包容所有社群，包括工薪阶层和穷人、LGBTQIA2S+、残障人士、神经多样性者等等，本准则采用了这个含义更为广泛的术语。

⁴许多学者和社群成员都对这个广泛的议题进行了有意义的思考：Olson, H. A. (2013). *The power to name: locating the limits of subject representation in libraries* (命名的力量：发现图书馆中主题呈现的局限性). Springer Science & Business Media; Howard, S. A., & Knowlton, S. A. (2018). *Browsing through bias: The Library of Congress classification and subject headings for African American studies and LGBTQIA studies* (在浏览中发现偏见：美国国会图书馆对非裔美国人研究和 LGBTQIA 研究的分类和主题词表). *Library Trends*, 67(1), 74-88; Tomren, H. (2003). *Classification, bias, and American Indian materials* (分类、偏见和印第安人材料). 未公开出版著作, San Jose State University, San Jose, California; Bone, C., & Lougheed, B. (2018). *Library of Congress subject headings related to indigenous peoples: changing LCSH for use in a Canadian archival context* (美国国会图书馆与土著民族相关的主题词表：更改主题词表以在加拿大档案语境中使用). *Cataloging & Classification Quarterly*, 56(1), 83-95; Baucom, E. (2018). *An exploration into archival descriptions of LGBTQ materials* (LGBTQ 资料的档案描述探讨). *The American Archivist*, 81(1), 65-83; Christensen, B. (2008). *Minoritization vs. universalization: Lesbianism and male homosexuality in LCSH and LCC* (少数化 vs. 普遍化：美国国会图书馆主题词表和分类中的女同性恋和男同性恋). *Knowledge Organization*, 35(4), 229-238.

⁵目前项目组（包括顾问团）的成员来自美国和加拿大的公共及私立学术机构。

适用范围

本《准则》的适用范围包括 *评估* 数字对象的使用或重用，这有别于具体使用或重用数字对象。项目组对“使用”和“重用”的定义如下：

- 使用：与外部用户表现出潜在兴趣且/或认定有特定价值的数字对象进行被动交互
- 重用：与外部用户表现出潜在兴趣/或认定有特定价值的数字对象进行主动交互

以上定义与八个“参与层面”构成了“使用-重用”矩阵，针对与数字对象的交互哪些应被视为使用、哪些应被视为重用进行了释义和确定。对“使用-重用”矩阵的完整解释，包括其定义的演变，不在本文的讨论范围之内，但可见于别处^[6]。

术语

- **可访问性**：在本文中，这一术语指的是网络可访问性或数字环境中的可访问性。
- **ALA**：美国图书馆协会（American Library Association）。
- **顺性别**：认同其出生时指定性别的人士。
- **顺性别异性恋**：既是顺性别又是异性恋的人士。
- **核心价值观**：本文提出的、在对数字图书馆资料的使用和重用进行伦理评估时必须予以考虑的概念。
- **数字对象**：由数字图书馆收集、保存和传播的经过数字化处理或原生就是数字形态的材料和相关元数据，包括但不限于：文化遗产文物实物、研究成果、学术成果和数据。
- **DPLA**：美国数字公共图书馆（Digital Public Library of America）。
- **GDPR**：[通用数据保护条例（General Data Protection Regulation）](#)，欧盟的一项数据保护法案。
- **GLAMR**：美术馆、图书馆、档案馆、博物馆和资料库的合称。
- **本《准则》**：本文即《数字对象使用及重用评估的伦理考量及准则》，简称《伦理评估准则》，提出了一个以核心价值观为基础、对数字对象进行使用及重用评估的框架。
- **异性恋身份认同**：“用于描述对出生时指定为‘相反’性别的人产生浪漫和/或性吸引力的个人的性身份的术语”^[7]。
- **IDEAS 原则**：代表包容、多元、公平、可访问和社会正义
- **神经典型者**：“表现出大多数人的典型思维和行为模式”。^[8]

⁶ 见 Kenfield, A. S., Woolcott, L., Thompson, S., Kelly, E. J., Shiri, A., Muglia, C., Masood, K., Chapman, J., Jefferson, D., & Morales, M. E. (2022). Toward a definition of digital object reuse (对数字对象重用的释义). *Digital Library Perspectives*, 38(3), 378–394. <https://doi.org/10.1108/DLP-06-2021-0044>

⁷ 异性恋身份认同，（无日期），见“Homosaurus 词汇表”，《跨性别者数字档案》，2023年4月20日检索自 https://homosaurus.org/v3/homokit0000624?from_q=heterosexual

⁸ 神经典型者，（无日期），见《牛津学习者词典》，牛津大学出版社，2023年4月20日检索自 <https://www.oxfordlearnersdictionaries.com/us/definition/english/neurotypical>

- **从业人员：**在数字图书馆、数字资料库、数字博物馆藏、数字展览、数据存储库或数字档案馆等数字环境中整理、维护和/或保存数字对象的个人。虽然通常指代的是正式的专业职位，但热心公民、新手、业余爱好者和自学成才的人士也可以是从业人员，尤其是对社区档案馆、公众科学研究和生态博物馆的收藏而言。
- **隐私：**隐私的概念为“不被打扰的权力”，指的是不被他人侵扰或干涉的自由。隐私也指控制自己的个人数据是否被他人收集、使用、分享和保存的能力。缺乏隐私意识会给数据关联人士带来严重的负面后果，包括情感伤害、声誉损害、身体伤害以及遭受歧视。
- **从业人员责任：**从业人员在规划和从事使用和重用评估（包括结果的分析和传播）时遵守准则的务实做法。
- **透明性：**一种长期性的、反复的方法，以与社区利益相关者（包括创作者、可能在馆藏材料中出现的个人和社群、读者和策展人）之间保持负责任的、包容的、渠道顺畅的沟通。

核心价值观

本文中提出并推荐的核心价值观是由图书馆^[9]、档案馆^[10]和博物馆^[11]工作人员以及信息和数据整理者/维护者^[12]构成的 GLAMR 行业内独立而又互补的实践团体提出的职业道德准则。

核心价值观告知从业人员以何种方式收集、使用并重用数据，以及在评估数据主体的使用和重用时应如何保护馆藏用户的隐私和保密性。技术发展十分迅猛，所以从业人员应有心理准备，持续性关怀用户隐私和最佳利益的最佳做法将会频繁变化。为代表性不足的群体倡导权益至关重要，这不仅需要从业人员深刻认识自己的职业责任，还要在公认的职业道德准则、价值观或其他规范性文件不够深入的情况下，承担起拓展边界的责任。

本《准则》以下列核心价值观为依据，还包括对下列价值观在“实际应用”中的评论。实际应用旨在加深从业人员对这些关键议题的认知，并考虑其实践是否合乎道德。对于部分核心价值观而言，不存在明确的实际应用或解决方案。有鉴于此，该部分被命名为“实际考量”，具体探讨了未来工作中的机会以及这些概念能够如何提升 GLAMR 从业人员的集体实践。

1. IDEAS 原则（即包容、多元、公平、可访问、社会正义）
2. 隐私

⁹ 参见美国图书馆协会《职业道德规范》<http://www.ala.org/tools/ethics> 以及国际图书馆协会与机构联合会（IFLA）制定的《图书馆员及其他信息工作者的伦理准则》<https://www.ifla.org/faife/professional-codes-of-ethics-for-librarians>。

¹⁰ 参见美国档案工作者协会（SAA）的《核心价值声明和道德准则》<https://www.ifla.org/faife/professional-codes-of-ethics-for-librarians> 以及国际档案理事会（ICA）的《道德准则》<https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics>

¹¹ 参见美国博物馆联盟（AAM）的《道德准则》<https://www.aam-us.org/programs/ethics-standards-and-professional-practices/code-of-ethics-for-museums/> 以及国际博物馆协会（ICOM）的《道德准则》<https://icom.museum/en/resources/standards-guidelines/code-of-ethics/>

¹² 参见国际信息科学与技术协会（ASIS&T）的《职业行为规范》以及美国计算机协会（ACM）的《道德准则》<https://www.acm.org/code-of-ethics>

3. 传统知识、文化遗产和知识产权
4. 职业发展和培训
5. 透明性
6. 认知

I. IDEAS 原则

长久以来，西方世界里创立推荐做法和标准时，历史上的西方人、白人、神经典型者、顺性别异性恋者、男性、健全人和非工人阶级背景的声音一直是标准，如今到了信息时代这一点基本上仍未改变。我们的目标是让《伦理评估准则》重新聚焦于如何真正体现公平、多元和包容。重新聚焦有助于支持尊重所有人及其贡献的最佳做法，无论这些个体的种族、族裔、部落归属或认同、国别、移民身份、性别认同或表达、性取向或性向认同、经济背景、阶层、年龄或语言是什么，也无论其是健全人还是残疾人。

实际考量

从业人员应当有不断自我反思的意识并付诸行动——在 GLAMR 机构里，他们是积极推行多元化与包容性，还是充当了压制者的角色？从业人员的目标是在进行评估时，将那些历史形成和新近出现的被弱势化群体的观点和声音予以放大。换句话说，本节的目标是让从业人员将 IDEAS 原则融入日常工作流程，并建立起（或完善）他们对这些身份领域的个人认知。

本文的“多元”一词用于描述人类存在的巨大差异，包括但不限于：与种族、族裔、社会经济地位、阶级、性别认同、性取向、健全或残疾状态、土著民族身份、国别、公民或移民身份、语言以及文化、政治、宗教和其他归属关系相关的群体性和社会性差异。我们必须对人类的多元化以及在教育、个性、技能、经验和知识基础方面的差异进行持续研究，否则无法确保全体人类都得到适当代表。

“包容”一词意指主动、有意和持续性地参与多元化（如前所述）并确保尊重差异。我们以开放的心态接受在文化、信仰、实践和贡献方面的新观念，并充分接纳那些历史上或在新近时期遭受限制、排斥并成为少数派的群体。

权力动态导致了上述社群的进一步边缘化，与那些掌握权力和领导地位的群体相比，他们未能被充分考虑。我们主张，不仅要留意特权人士的特权，注意他们如何在有意识和无意识偏见的情况下使用特权，而且还要关注那些通常处于领导位置边缘的群体，以颠覆过去的父权结构，平等、尊重地对待所有相关人士，无论其出身如何。

同样重要的是对与社会正义及其理论和实践相关的价值观进行批判性审视，包括但不限于承认人类内部在种族主义、暴力、智识、人权、贫困、饥饿和环境保护等方面存在的分歧。

压迫系统的主体来自于集体主义，典型例子包括父权制、白人至上主义、种族主义、性别歧视、体能歧视、资本主义催生的经济等级制以及更多。本《准则》有助于确定一系列公正且合乎道德的标准，在可访问性、公正、公平、多元和包容方面为所有人奠定公平基础。

II. 隐私^[13]

隐私是一项基本人权^[14]。在 GLAMR 语境中，隐私保护也被视为研究和创造的必要条件^[15]。因此，保护个人用户的个人数据、征求用户同意以及在服务及评估方面保持公开透明，践行注重隐私保护的行业价值观，是对 GLAMR 从业人员的职业要求^[16]。在此背景下，D-CRAFT 旨在促进数字图书馆和资源库的数据主体的重用评估；然而，数据的收集和评估不应以牺牲用户的隐私权为代价，也不应放弃去争取用户的知情和明确同意。本《准则》力求重点关注那些因针对性的、系统性的监控结构和技术而遭受严重侵害的群体。有时，进行评估所需的数据类型会给用户带来相当大的隐私风险。不同的群体将经受不同的隐私侵犯，个人受到侵犯的严重程度也因人而异。

由于数据的使用或重用存在关联到具体个人的可能性，因此收集及评估时必须考虑收集特定用户数据可能产生的危害。在某些情况下，遵循数据隐私保护和信息安全的最佳实践能够降低隐私泄露风险，我们将在随后对部分做法加以解释。不过从业人员必须认识到，部分数据或评估要求会给用户隐私带来巨大风险，以至于以合乎道德的方式进行收集或评估几无可能。在规划和决策过程中，从业人员衡量数据收集和评估的潜在隐私风险和伦理影响时，应牢记这一点。

数据和隐私

本《准则》所讨论的数据包括个人和非个人的。针对个人数据有着不同的定义，尤其是在数据隐私监管方面。我们将参考《通用数据保护条例（GDPR）》和美国国家标准与技术研究所（NIST）对个人数据的定义。GDPR 对个人数据的定义如下：

“如果数据主体可以被直接或间接识别，特别是通过参照标识符，如姓名、识别号码、位置数据、在线标识符或若干特殊特征之一，来揭示这些自然人的身体、生理、遗传、精神、商业、文化或社会身份，则数据主体是可识别的。在实践中，数据主体还包括以任何

¹³ 有关隐私和技术伦理的更多信息，请参考数字图书馆联盟（DLF）的隐私和技术伦理工作组：

<https://osf.io/bdyvq/>

¹⁴ 联合国《世界人权宣言》：<https://www.un.org/en/about-us/universal-declaration-of-human-rights>

¹⁵ Jones, Kyle ML 及其他人，"“We're being tracked at all times”: Student perspectives of their privacy in relation to learning analytics in higher education（“我们无时无刻都在被追踪”：学生对高等教育学习分析中隐私的看法）。” *Journal of the Association for Information Science and Technology* 71.9 (2020): 1044-1059.

¹⁶ Zimmer, Michael. "Librarians' attitudes regarding information and internet privacy（图书馆员对信息和互联网隐私的态度）。” *The Library Quarterly* 84.2 (2014): 123-151.; Newman, Bobbi, and Bonnie Tijerina, editors. *Protecting Patron Privacy: A LITA Guide*（保护读者隐私：图书馆与资讯技术协会指南手册）。Rowman & Littlefield Publishers, 2017. Lee, Una, and Dann Toliver. *Building Consentful Tech*（构建知情同意的技术）。2017.

方式分配给或可能分配给一个人的所有数据，如电话号码、信用卡号码、员工编号、账户数据、车牌号码、外观、客户号码或地址都是个人数据。由于定义中包含‘任何信息’，因此必须假定‘个人资料’一词的外延无限广泛。”^[17]

国家科学技术研究所（NIST）将个人身份信息（PII）定义为：

“由机构保存的关于个人的任何信息，包括(1)可用于区分或追踪个人身份的任何信息，如姓名、社会安全号码、出生日期及地点、母亲的婚前姓名或生物识别记录；(2)关联或可能关联到个人的任何其他信息，例如医疗、教育、财务和就业信息。”^[18]

还应注意的是，NIST 建议相关机构考虑 PII 定义之外的数据的敏感性或机构性/个人性风险。

两种定义中都包含不同类别的个人数据：

- 直接标识符（姓名、电子邮件地址、位置）
- 间接标识符（年龄、种族/族裔、性别认同）
- 行为数据（包括内容使用/重用历史和浏览活动等活动生成的数据）

大量研究表明，间接标识符和行为数据可用于识别个人；因此，处理直接标识符的谨慎做法也适用于这些类型的个人数据^[19]。

数据去标识化方法，如去除直接标识符、数据聚合、混淆处理和数据截断，能够降低、但不能完全消除在数据集中重新识别到个人的风险。一种特殊的去标识化方法，假名化——即用与可识别个人无关的标识符替换个人标识符——也不能避免被重识别的风险。匿名化——打破数据和可识别个人之间的联系——对于个人数据来说非常困难，甚至难以实现^[20]。在收集、存储或提供重用评估数据之前，从业人员必须考虑任何去标识化计划的重标识风险。与重标识相关的因素包括整个数据集中的个体数量、与外部数据源重叠的数据元素、由这些数据元素的独特组合定义的小类档案、弱假名化手段以及数据集和外部数据源之间的人口重叠模式^[21]。

¹⁷通用数据保护条例（GDPR），“个人数据”：<https://gdpr-info.eu/issues/personal-data/#:~:text=The%20data%20subjects%20are%20identifiable.%2C%20genetic%2C%20mental%2C%20commercial%2C>

¹⁸ E McCallister, T Grance, and K A Scarfone, “Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) (保护个人身份信息 (PII) 机密性指南)”, Gaithersburg, MD: National Institute of Standards and Technology, 2010, <https://doi.org/10.6028/NIST.SP.800-122>.

¹⁹ Latanya Sweeney, “Simple Demographics Often Identify People Uniquely (简单的人口统计数据通常足以识别出唯一个体)”, Data Privacy Working Paper (Carnegie Mellon University, 2000, <https://dataprivacylab.org/projects/identifiability/paper1.pdf>); Arvind Narayanan and Vitaly Shmatikov, “How To Break Anonymity of the Netflix Prize Dataset (如何打破 Netflix 奖项数据集的匿名性),” *ArXiv:Cs/0610105*, 2007 年 11 月 22 日, <http://arxiv.org/abs/cs/0610105>.

²⁰ Boris Lubarsky. “Re-Identification of ‘Anonymized’ Data (对“匿名”数据的重识别).” *Georgetown Law Technology Review*, 2017 年 4 月 12 日. <https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/>.

²¹ <https://doi.org/10.5334/egems.270>

如果收集时尚未明确评估需求，则不得收集数据。从业人员会发现，通常情况下可以收集非个人数据或以对隐私侵犯较小的方法来满足核心评估需求。从业人员应制定政策，规定机构不得为评估目的而收集个人数据，因为个人数据可用于综合重用评估的情况十分鲜见。每个机构还应各自如何保护个人和非个人数据做出决策、发布文件并公之于众。

实际考量

评估工具考量

许多从业人员面临双重需求：既要开展评估，又要谨慎评估所使用的评估工具，以便在符合隐私规范和职业价值观的前提下合乎道德地使用。如前所述，有些时候在开展特定评估时，对用户造成的隐私风险太过巨大以至无法进行。风险计算的一部分便是考虑使用什么工具来收集和分析用户数据用于评估。部分可用于重用评估的工具默认收集个人数据。从业人员应该关闭所有的个人数据收集设置；但是，有些时候并不存在这样的选项。重用评估工具所收集的个人数据示例如下：

1)一些网络分析工具会追踪互联网协议（IP）地址。**Matomo** 通过默认设置屏蔽地址的最后八位字节来匿名化 IP 地址，并允许执行者进一步掩盖此信息。在将于 2023 年 7 月 1 日发布的 **Google Analytics 4** 之前，**Google Analytics** 默认设置为允许 **Google** 收集每位网站访问者的 IP 地址，执行者必须设置 IP 匿名化功能才能防止 **Google** 追踪这些信息。

Google Analytics 4 采取与 **Matomo** 类似的做法，通过掩盖地址的最后一个八位字节来实现 IP 地址的匿名化。即使没有 IP 地址，分析工具仍然能够收集可用于创建用户数字指纹的其他数据，例如浏览器、操作系统和其他设备信息。

2)**Qualtrics** 等调查软件默认收集每个受访者的 IP 地址和位置数据。此外，调查表通常使用姓名和电子邮件地址分发给已知的个人名单，默认情况下，这些数据字段会与个人回复一起存储。从业人员必须启用“匿名回应”设置，以从 **Qualtrics** 的个人回应中删除 IP 地址和位置数据。然而，这一设置并未对参加调查所提交的个人数据进行去标识化处理，因此便产生了重识别的风险。

每个事例中的工具都有一些数据收集和去标识化设置，但这些设置并没有完全去标识化或阻止该工具对个人数据的全面收集。从业人员必须根据为评估而收集的数据类型来权衡每个评估工具的隐私风险，并考虑如何结合隐私保护功能和隐私保护最佳实践，如只收集对评估绝对必要的数据（或数据最小化），来降低这些风险。

另一个值得关注的问题是托管和维护评估所使用工具的第三方。从业人员面临着依赖第三方工具、而这些工具可能并不遵守本《准则》的困境。在实践中，创建这些工具的组织可能会与本《准则》背道而驰，如缺乏可靠的数据隐私保护和他安全实践，或者未经用户同意向其他第三方转售或共享数据。**GLAMR** 社群应倡导或资助并开发收集数据且能保护隐私的技术。虽然从业人员

应当在与供应商订立合同时针对如何加强用户隐私保护而进行洽谈，但这并不能保证用户不会遭受来自使用供应商工具的相关方的隐私侵犯。因此，从业人员应当明确地寻找不侵犯用户数据隐私权、隐私侵犯较少的评估技术。GLAMR 机构应该利用他们在市场上的地位，联合起来推动供应商在隐私保护和其他领域实施变革。

话虽如此，从业人员无法控制每一种可以收集数据供组织内其他人或外部第三方进行重用评估的工具。从业人员可能无法控制的重用评估数据收集方法的示例包括：反向图像查找、链接分析和引文分析。从业人员应向用户明确说明使用了哪些工具、由谁使用、为何使用以及每个工具会自动进行哪些数据收集和追踪行为，以此教育公众去了解这些工具中存在的隐私限制和风险。

知情告知和用户同意

在创建、发展评估框架的过程中，从业人员必须有意识地以弱势化社群为中心。尤其是在收集用户个人数据时，从业人员必须尽一切可能征得用户的知情同意。取得知情同意的流程应经过谨慎设计，以确保被评估者了解数据收集及评估的实际过程。这一流程应当由那些最易受到隐私侵犯、在历史上往往遭受社会不公平对待的人士制定并审查。在规划评估时，应仔细、连贯地制作文献，如用于外展宣传、同意流程、关系映射和权力分析^[22]的文件。制定、审查这些评估规划材料时应考虑社群的人口统计数据或数字图书馆馆藏的变化。社会和身份会随时间推移而演变，这些材料也应接受重新审视。今天创建的内容可能会在之后的时间里带来危害或变得无关紧要。

从业人员应尽最大可能不去收集个人数据用于重用评估，而是应考虑收集最少可行数据的做法，即只收集使评估得以实现的最少数据量。如果收集了个人数据，则应考虑进行去标识化、匿名化或假名化处理，以保障数据关联人士的隐私。在参与数据收集和评估之前，从业人员必须告知用户并征得用户对其数据收集和评估的同意。除了作为许多基础隐私框架的核心原则——例如经济合作与发展组织的隐私保护原则^[23]——数据隐私保护法规通常需要某种形式的知情同意。例如，GDPR 要求用户的同意是自主给予的、具体的、知情的和明确的，用户也可以在任何时候撤回同意^[24]。从业人员应尽可能只在用户给予收集和评估知情同意的情况下收集用户数据。

一旦为数字馆藏评估订立了数据追踪、使用及隐私保护政策，从业人员应将这些政策与用户分享。这些通知应以通俗易懂的语言书写，避免使用过多的专业术语和法律术语。这些信息应在公共网站上以隐私声明的形式呈现给用户。从业人员还应在数据收集发生之前通知用户。例如，从业人员可能会在下载点发布通知提醒用户，重新使用从馆藏中下载的数字材料可能会接受出于评估目的的追踪。这个通知应该赋予用户选择权，如果他们不同意追踪，就可放弃继续下载。理想情况下，用户应该能够选择接受数据追踪，而不是选择拒绝追踪而被迫放弃对材料的潜在使用/重

²² Hagan, Teresa, and David Smal. "Power-Mapping—I. Background and Basic Methodology (绘制权力图—I. 背景和基本方法)." *Journal of Community & Applied Social Psychology* 7, no. 4 (1997): 257–67. [https://doi.org/10.1002/\(SICI\)1099-1298\(199709\)7:4<257::AID-CASP428>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1099-1298(199709)7:4<257::AID-CASP428>3.0.CO;2-P).

²³经济合作与发展组织隐私保护原则：<http://oecdprivacy.org/>

²⁴ GDPR “同意的条件”：<https://gdpr.eu/gdpr-consent-requirements/>

用；但是，如果由于技术限制而无法做到这一点，从业人员必须尽职尽责，在隐私声明中和首次使用数字材料之前向用户清楚地传达这一限制和约束。

如果一位用户想撤回最初对其数据收集及重用评估的明示（opt-in）同意，应该有便利的机制让其执行此操作。任何默示（opt-out）同意也应触发对过去收集的用户相关数据予以删除的操作。根据 GDPR，这被称为“删除权”或“被遗忘权”^[25]。如果可对汇总追踪进行设置，也应当在数据收集时告知用户。如果无法设置此选项，同样也需要告知。

在某些情况下，可能无法获得知情同意；例如，当用户的身份未知时，调查其身份本身就会侵犯其隐私。在这些情况下，我们建议从业人员向参与者提供一份通俗易懂的通知，告知评估的益处和隐私风险。评估工具应始终尊重浏览器中的全局隐私控制（即 GPC：“请勿追踪”标示的后继者），供用户明确选择默示同意退出数据收集和追踪。是否同意以重用评估为目的收集个人数据不应成为使用数字材料的障碍，同意流程也不应迫使用户必须放弃权力或能动性才能换取数字材料的使用或重用。

从业人员不应将收集到的用作重用评估的数据进行货币化或转售，也不应与将数据货币化或转售的第三方共享数据。从业人员应尽一切可能与第三方洽谈，在与其订立的协议中取消第三方对用户数据进行货币化或转售的权利。参与转售数据提升了营利性收益，但与此同时也剥削了少数群体。无论是发布错误信息（无意识分享不实信息）或虚假信息（有意分享不实信息），还是有目的地保持缄默，都有可能对真相进行有针对性的操纵，从而进一步剥夺少数群体的权力。这些侵犯隐私的表现削弱了关系和信任，如果不加以解决，可能导致 GLAMR 本来试图聚合的群体出现分裂。

D-CRAFT 鼓励从业人员继续集结全行业力量，以批判性的姿态介入这一复杂议题^[26]。

III. 传统知识、文化遗产和知识产权

本《准则》旨在为数字文化遗产、传统知识、文化资料和数据等数字对象的重用评估提供便利。我们对数字对象重用的构想受到了西方的知识产权、版权和合理使用概念的显著影响。不过，也存在一场为“传统知识、遗传资源和传统文化表现形式（民俗民情）”提供有效保护的^[27]运动，要求它们与殖民意识形态下创作的作品享有同样的知识产权保护。

实际考量

²⁵ GDPR 第 17 条“删除权”：<https://gdpr-info.eu/art-17-gdpr/>

²⁶ 欲了解隐私和技术伦理相关的更多信息，请参考数字图书馆联盟（DLF）隐私和技术伦理工作组：<https://osf.io/bdyvq/>

²⁷ 《传统知识和知识产权——背景简介》（无日期），世界知识产权组织，于 2023 年 4 月 21 日检索自 https://www.wipo.int/pressroom/en/briefs/tk_ip.html

本《准则》在很大程度上针对的是数据对象的使用和重用的评估工作。然而，围绕传统知识、文化遗产和知识产权的伦理准则与这些材料的获取、管理、所有权和文化渊源紧密地交织在一起，可能会对数字化和评估产生影响。因此，本节包含了围绕获取、管理材料的伦理问题所展开的讨论。

这一流程的关键环节是核实材料的获取是否以尊重土著民族文化协议的适当方式进行，以及获取这些材料是否征得了土著民族的许可。历史上有许多在未获得材料所属的土著社群许可或知情的情况下获取材料的事例，并且还存在着只为发挥最大影响力、罔顾文化正确性的其他例子。例如，摄影师 Edward Curtis^[28]的大部分照片为了迎合“消失的种族”这一主题，被刻意抹去了与现代社会的任何联系，在特意安排下，拍摄对象的着装也并不完全与真实生活相符。

在处理来源于土著民族和弱势化群体的材料，包括进行重用评估时，有必要针对其价值观、信仰和关切事项开展额外研究，以便相关群体能与以数字形式展出或计划展出这些材料的资料库之间建立起互信交流。评估可能会发现违反土著民族文化协议的行为以及在隐私和公平方面存在的问题，需要对应解决。事实上，每个土著民族社群对传统知识和“认知方式”的评估都存在不同的定义。要建立并维持基于信任的关系，从业人员必须谨慎考虑相关问题。

对数字对象的使用和重用进行评估、却罔顾创建材料的社群的特定协议，这是一个西方概念。理想情况下，应在材料进行数字化处理之前就确定其获取方式是否符合道德规范和社群协议。但是，如果在数字材料的重用评估过程中发现土著社群没有参与其中，则必须采取措施与土著社群建立适当联系。是否继续允许使用或重用相关材料将由所涉及的土著社群决定。

为了保护土著民族的权益，科学研究数据全球联盟（Research Data Alliance）下属的国际土著社群数据主权倡议团体（International Indigenous Data Sovereignty Interest Group）针对土著民族数据治理制定了 CARE 原则，即“集体利益、控制权、责任和道德操守”。这些原则基于加拿大的 OCAP 原则，即“所有权、控制权、访问权和占有权”，确保所涉及的土著群体积极参与、受到尊重，保障了土著群体对其文化材料和财产的管理权。OCAP 承认土著民族对其所有的文化材料的所有权、控制权、访问权和占有权，并主张土著民族对数据收集过程、材料占有以及信息用途加以控制。

CARE 原则的主要目标为：(1)通过加强土著民族对数据的使用来促进土著民族自决；(2)尊重“科学数据管理”的 FAIR 原则：可发现、可访问、可互操作和可重用^[29]，确保按照土著民族的条件共享数据。

对以数字方式共享的材料进行评估时，必须与这些材料原本所属的土著民族合作。这种做法的首要法则是“没有我们的参与，不要做关于我们的决定”。西方的版权概念不能应用于本质上由这些民族或民族内的家庭所管理的材料。个人所有权，特别是那些被认为是神圣的或具有深刻文化意义的材料的所有权，并不是大多数土著民族所奉行的概念。特定的图像、故事、舞蹈和服饰可能由民族内部的家庭或成员为后代传承而保留，但并不被视为由群体内的任何个人所拥有。

²⁸ Mathilde Arrivé, « Beyond True and False? (超越真假) », *Études photographiques* [En ligne], 29 | 2012, mis en ligne le 18 juin 2014, consulté le 02 février 2023.

<http://journals.openedition.org/etudesphotographiques/3474>

²⁹ Wilkinson 及其他人, 2016. <https://doi.org/10.1038/sdata.2016.18>

Anderson 指出，“土著民族必须以核心身份参与获取、使用其知识和知识实践的适当框架的制订工作”^[30]。针对“土著民族要求法律承认其文化体系及管理知识获取及使用的方式，并尊重其作为土著文化中以及之外的有价值的知识的保管者/所有者/培育者的身份”，她进行了具体阐述。

2019年6月，加拿大政府根据工业、科学和技术常务委员会（Standing Committee on Industry, Science and Technology）对加拿大版权法案审核的结果，发布了《版权法定复核报告》^[31]。常务委员会根据土著社群代表所做的陈述发言提出了修订建议，其中就有“承认加拿大法律在版权法范围内及范围之外对传统艺术和文化表现形式进行有效保护。”其他建议还包括“土著社群参与制定国内和国际知识产权法”以及“授予土著民族管理传统艺术和文化表现形式的权力，一个重要手段即是在《版权法》中增设一项不可克减的权利条款”。这些建议均以土著社群的证言为基础。报告承认，“传统观念认为作品源于一位或少数几位作者个人，因此所有权授予个人；但土著代表证人认为，传统艺术和文化表现形式是社群共有的，这与传统的版权所有权观念截然相反。”

这种公共所有权的思路应被纳入负责任的数字内容管理。在对土著社群的文化材料进行数字化、共享及重用评估时，确保向其征求意见并让其充分参与，对于真实呈现、充分尊重其传统知识以及实现土著自治管理至关重要。在从业者可能遇到的任何评估中，将通过非法手段或在未经验证的来源链下获得的圣物的高分辨率替代文件进行共享，均应包含对潜在伦理风险的考量。由于管理库无法控制材料是以何种方式获得、是否以合乎道德的方式获得以及读者在随后将如何处理内容，因此共享这些数字对象可能违反特定土著社群的圣物处理协议。

因此，从业人员的责任不止评估知识产权状况，还要在材料及其用户之间建立关联。第一步是确定出处。从业人员应该考虑以下一些问题：材料是如何被纳入资源库的？它们是如何分类的？哪些资源可以用来验证其来源和分类？

通过与身为文化资料管理者的土著社群合作、尊重土著民族对资料拥有主权这一原则，从业人员在向读者提供数字材料时，便不会在无意识中违反土著民族知识产权或著作权法规。

IV. 职业发展及培训

职业发展和培训对于获得新技能、磨练现有技能是必要之举，还能催生可能对从业人员观念产生积极影响的不同声音，并有助于萌生可提升从业人员日常操作和工作的设想。这在评估工作中尤为重要，因为评估结果经常作为依据来影响决策，包括制定数据驱动政策、确定数字化工作的优先级以及评估资金需求。就数字对象重用评估工作以及如何合乎道德地开展工作设定职业发展目标并安排培训，是全新的研究和实践领域。因此，从业人员必须扮演好他们作为数字文化遗产艺术品管理者的专业角色，同时也要考虑这些艺术品重用评估的伦理问题。

实际考量

³⁰ Anderson, J. (2010). *Indigenous/traditional knowledge & intellectual property*. Duke University School of Law Center for the Study of the Public Domain. https://web.law.duke.edu/cspd/pdf/ip_indigenous-traditionalknowledge.pdf

³¹ 《版权法定复核报告》，2019。 <https://www.ourcommons.ca/DocumentViewer/en/42-1/INDU/report-16/>

从业人员不应在不了解所选方法、框架和/或评估工具的情况下尝试进行重用评估。评估培训应遵循从业人员的自我反思和批判意识。例如，培训可以整合反偏见模块来支持从业人员的自我意识和其他权力动态——无论是显性还是隐性。任何基于评估活动的政策变更或建议都应着眼于从业人员的专业知识水平以及对所选评估方法和工具的熟悉程度，并记录存档。同样，在评估中采用算法和/或人工智能（AI）应格外谨慎。这类工具许多都是专有的，其代码不会提供给开发工具的公司之外的任何人。这一点值得关注，因为研究表明，计算机算法^[32]和人工智能往往反映了人类程序员的偏见，并且也存在训练数据集缺乏多样性^[33]等其他问题。人工智能和计算机算法可以成为从业人员的有利工具，但其立场不能被视为客观，也不能将其作为唯一的评估手段。此外，从业人员应在不侵犯相关方隐私的情况下提供被引为决策支持证据的评估结论以及数据、方法、工具和分析结果。

V. 透明性

透明性始于与社群中的利益相关方（包括创作者、馆藏材料中可能描绘的个人和群体、用户和读者以及策展人）进行负责任、包容、渠道畅通的交流。本《准则》所载的核心价值观体现在整个数字生命周期中透明性相关的主题、概念和应用中。因此，本节专门面向对透明性在自身工作中的作用特别感兴趣的从业人员，这些内容也为其他核心价值观提供了参考。保持透明即就数字图书馆从业人员管理、处理信息的流程和政策进行沟通。事实上，这些流程和政策应该参考并遵循本《准则》之前提及的隐私和 IDEAS 相关部分的原则。

实际应用

为确保长期保持透明性，本《准则》强烈建议从业人员在评估数字资产重用时对流程和行为进行记录。政策、准则或最佳实践等文件应公之于众，并以通俗易懂的语言撰写、尽量避免专业术语，以方便大众阅读。这可能还涉及到将文献翻译成其他语言。公共可访问性是指任何人都能访问，而不仅是对材料或项目有着直接投入的利益相关者。从理论上讲，将信息公开符合本《准则》的要求。从技术上讲，一些数字图书馆从业人员可能会发现很难确定或找到所有相关群体。解决此问题的潜在方法是在数字对象本身所在的同一网页上提供这些政策的链接。

提升文献可访问性的方法可能包括将其翻译为多种语言、转换为可数字访问的格式、使用非专有软件编排并存储文献以及/或者针对较长的文件协议提供报告摘要或概述，以确保所有社群成员都有与信息互动的入口。许多机构和组织共享其可访问性标准，我们可以参照这些标准^[34]。可能对

³² For more information see Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism* (算法的压迫性：搜索引擎如何强化种族主义). 纽约大学出版社。

https://openlibrary.org/works/OL19734838W/Algorithms_of_oppression

³³ 更多信息见 Buolamwini, J. & Geburu, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification (性别图谱：商用性别分类技术中的种族准确率差异)，第一届公平、责任和透明度会议论文集，见机器学习研究会议论文集 81:77-91:

<https://proceedings.mlr.press/v81/buolamwini18a.html>.

³⁴ 例子包括：数字图书馆联盟“数字无障碍访问群组”：https://wiki.diglib.org/Digital_Accessibility_Group/；美国档案工作人员学会《残障人士无障碍获取档案指南》：<https://www2.archivists.org/groups/reference-access-and-outreach-section/guidelines-for-accessible-archives-for-people-with-disabilities>

信息产生影响的流程或协议如若出现修订和更新，也应保留更新日志^[35]等文件。因此，建议在对具体机构或团体而言效率最高的时限内返回现有文档并更新流程，使其反映最新的操作、标准、协议和准则。

如果个人的工作场所没有保留这样的通讯交流，这些通讯交流没有接受定期审查及修订，并且/或者如果这些做法没有得到坚持，请考虑迈出实现透明的第一步。如果对管理敏感信息或明确传达评估程序的方法没有清晰的理解，对从业人员或机构而言，透明性将变得更加难以实现和维持。规范性的文件也应体现透明。此外，从业人员应在不侵犯相关方隐私的情况下，公布被引为决策支持证据的评估结论以及数据、方法和分析结果。

VI. 认知

本《准则》所持立场为：人类或人类所创造的机器和工具，如计算机算法和人工智能，无法实现真正的中立、公正或客观。所有的行为者都会受到其所在环境、境遇和生活经验的影响，这些影响会塑造我们从人际交往到物质生活的方方面面，数字图书馆从业人员的工作也不例外。在评估重用时，我们应明确对自己的交叉身份和参考框架的认知。针对进行重用评估的任何工具或方法也应如此，这在使用专有软件、算法或人工智能时可能无法实现。

我们并非力求“中立”，而是以实现本文中所定义的透明和公平为目标。

实际应用

从业人员应认识到，在进行评估时，他们并非、不能、也不应是中立或不偏不倚的。恰恰相反，在评估数字对象的重用时，从业人员有责任倾听用户心声、倡导用户权益并保护用户的个人数据——受到重用数据评估影响的、历史形成和新近出现的被弱势化群体尤其应得到如此对待。虽然本文承认，个人并非也不应被认为是“不偏不倚的”或“中立的”，但从业人员有责任认清自己身份。从业人员应考虑评估技术固有的局限性，并在依据重用评估做出决策或建议之前，将这些工具中的偏见影响纳入考量。

在评估数据重用以及依据这些评估制定的任何政策、准则和协议时，从业人员应意识到自身身份。文献编制中可以透明地体现出这一点，如在后记中加入个人的履历资料^[36]。

³⁵ 更新日志是一份编辑记录，包括对数字对象（如网站、电子文档、软件代码）进行增补、修订和删除。可以将其视为一种来源元数据。[如何维护更新日志](#)网站对“更新日志”的定义是“一个由人工编辑、以时间为倒序的列表，用于记录项目中每个版本的显著变动。”[信息管理标准研究推进联盟（CASRAI）的研究数据管理词汇表](#)提供了一个更通用的定义：“追踪每个变更从提交到审核、批准、实施、结束的进度。日志可以使用文档或电子表格手动管理，也可以使用软件或基于网络的工具进行自动管理。”

³⁶ 加入履历资料以及以其他形式展现过程透明性的一例：Light, Michelle, and Tom Hyry. “Colophons and Annotations: New Directions for the Finding Aid (后记和注释：检索工具的新方向).” *The American Archivist* 65, no. 2 (2002年9月): 216–30. <https://doi.org/10.17723/aarc.65.2.l3h27j5x8716586q>.

延伸阅读

- Caswell, M., & Cifor, M. (2016). From Human Rights to Feminist Ethics: Radical Empathy in the Archives (从人权到女权主义伦理: 档案中的激进同理心). *Archivaria* 81, 23-43. <https://www.muse.jhu.edu/article/687705>.
- 数字图书馆联盟 (Digital Library Federation, 简称 DLF) 隐私和技术伦理工作组: <https://osf.io/bdyvq/>
- Olson, H. A. (2013). *The power to name: locating the limits of subject representation in libraries* (命名的力量: 发现图书馆中主题呈现的局限性). Springer Science & Business Media.
- Howard, S. A., & Knowlton, S. A. (2018). Browsing through bias: the Library of Congress classification and subject headings for African American studies and LGBTQIA studies (在浏览中发现偏见: 美国国会图书馆对非裔美国人研究和 LGBTQIA 研究的分类和主题词表), 《图书馆趋势》, 67(1), 74-88.
- Tomren, H. (2003). Classification, bias, and American Indian materials (分类、偏见和印第安人材料), 未公开出版著作, 圣何塞州立大学, 加利福尼亚州圣何塞。
- Bone, C., & Lougheed, B. (2018). Library of Congress subject headings related to indigenous peoples: changing LCSH for use in a Canadian archival context (美国国会图书馆与土著民族相关的主题词表: 更改主题词表以在加拿大档案语境中使用), 《编目和分类季刊》, 56(1), 83-95.
- Baucom, E. (2018). An exploration into archival descriptions of LGBTQ materials (LGBTQ 资料的档案描述探讨), 《美国档案工作者》, 81(1), 65-83.
- Christensen, B. (2008). Minoritization vs. universalization: Lesbianism and male homosexuality in LCSH and LCC (少数化 vs. 普遍化: 美国国会图书馆主题词表和分类中的女同性恋和男同性恋). Knowledge Organization, 35(4), 229-238.
- 世界知识产权组织, 《传统知识与知识产权——背景简介》, 访问日期: 2020 年 11 月 8 日 <https://www.wipo.int/publications/en/details.jsp?id=3858&plang=EN>.
- Dressler, V. (2018). Framing Privacy in Digital Collections with Ethical Decision Making (以伦理决策界定数字收藏中的隐私) (San Raphael, CA: Morgan and Claypool). <https://doi.org/10.2200/S00863ED1V01Y201807ICR064>
- 全国网络隐私论坛 (National Web Privacy Forum) 的项目成果: <https://www.lib.montana.edu/privacy-forum/>
- 企业数据伦理: 高级分析及人工智能时代的数据治理转型 <https://ssrn.com/abstract=3478826>

致谢

我们要感谢我们的受薪[专家](#)为以下章节贡献了重要的书面意见:

Monique Manatch——传统知识、文化遗产和知识产权

Becky Yoose——用户隐私

Scott Young——数据隐私