

Preserving Academic Content in a Digital Age

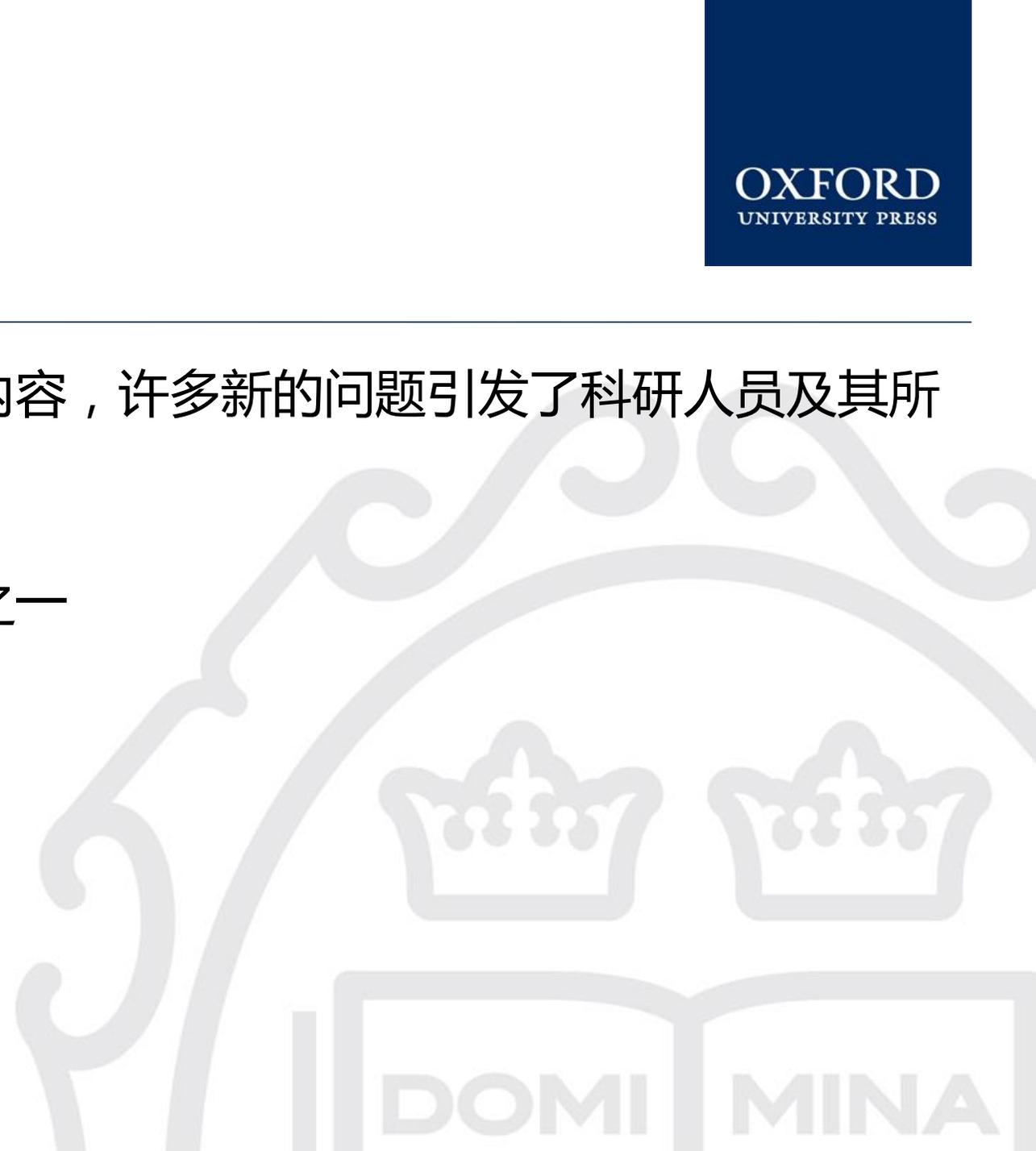
数字化时代的学术内容数据保存

James Phillpotts

Director of Digital Operations Innovation

Oxford University Press

- 随着学术出版逐渐转向以在线形式发布内容，许多新的问题引发了科研人员及其所属科研机构以及出版商们的思考
- 学术内容的保存是其中备受关注的问题之一
- 确保学术内容可以长期访问十分重要



- 本报告包括以下内容：
 - 学术内容的保存为何如此重要
 - 有哪些措施可以确保学术内容的保存
 - 以牛津大学出版社（OUP）为例，学术出版商如何进行学术内容的保存

背景介绍

- 学术内容的数字保存比内容本身更令人关注
- 各类“原生数字”内容的长期保存已成为一个重要问题
- 文化遗传正在被数字化——在某些情况下，这些数字化形式已成为了主要资源
- 学术内容也是如此



学术内容的数字化保存为何如此重要？

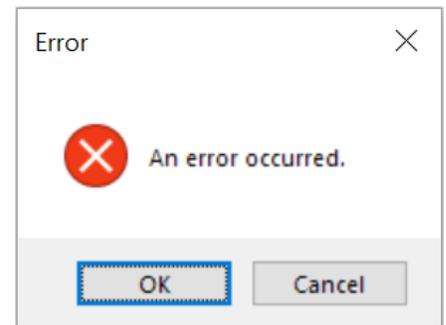
- 如今，学术期刊的在线版本（而非印刷版）已成为其内容传播的主要方式
- 期刊的印刷版本同样面临长期保存的挑战！然而，随着在线版成为主要版本，数字形式的内容保存成为人们关注的主要问题
- 科研作者和机构图书馆都十分关注在线学术内容的持续可访问性



© Flickr, Meaghan K CC BY-NC-ND 2.0

学术内容的数字化保存为何如此重要？

- 这不仅仅是理论性问题
- 有形的印刷版本可能会遗失、损毁或破朽；数字内容不会有这些问题，并且可以更容易地被多个读者同时访问
- 然而，数字内容本身可能很脆弱
- 这是因为学术内容今天可以在线访问并不代表明天仍然可以访问
- 机构图书馆需要确保他们采购的内容可以安全地长期保存
- 影响这一问题的潜在因素包括资金缺失、管理不善、抑或是保存方法不够完备



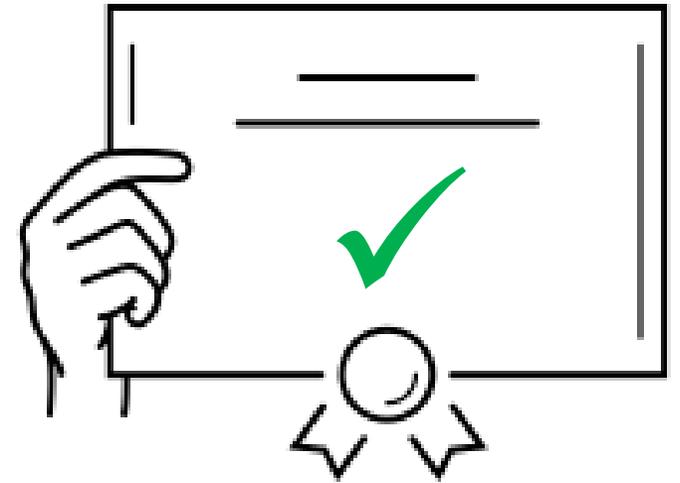
学术内容的数字化保存为何如此重要？

- 出版商可能终止业务或其平台遭受持续的灾难性故障
- 期刊可能停刊或部分内容从其出版商平台上删除
- 这些意味着学术内容将不再可用，并且
 - 以前可访问这些内容的科研人员将无法再次访问这些内容
 - 其原始研究的学术产出无法被此后的学者所获知
 - 如果相关内容已被引用，内容的不可用将使得引用证据无法得到核实从而导致引文链的断裂
 - 科研人员需要访问特定的重要学术内容——其它在线资源可能提供有关某一主题的信息，但学术记录是经过同行评审的、经过验证的信息来源



出版商的职责

- 学术研究和出版是全球性的活动
- 学术内容的出版商希望因其专业性和高质量的内容得到国际认可
- 学术期刊的出版已有300多年的历史，具备完善的实践规范，包括近年来在数字内容在线访问方面日益重要的规范
- 全球领先的出版商在专业性方面都有其特质



出版商的职责

- 对期刊出版而言，这些特质包括：

- 国际化的编委会
- 严格的同行评审、投稿和剽窃检查流程
- 出版道德规范
- 版权管理
- 明确的识别规范（ISSN、DOI）
- 实现内容的引用、可发现性和可访问性

- 以及

- 确保内容的长期数字化保存，使得内容始终可用，让作者和机构图书馆放心



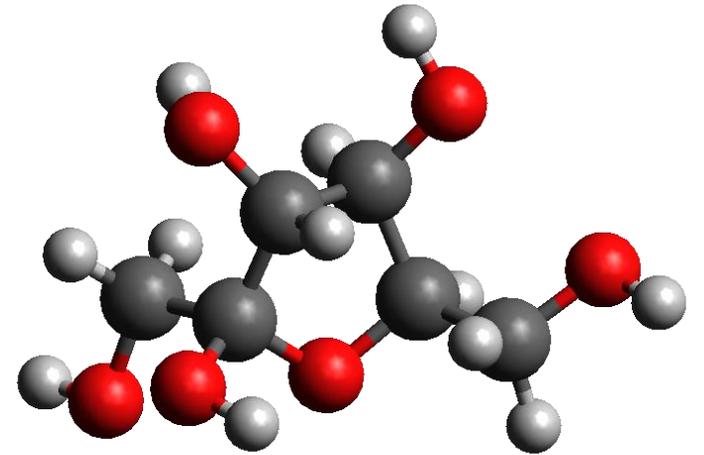
出版商的职责

- 保存的需求一直存在，但其要求和期望已经改变
- 对于印刷版本，保存的责任在于内容的接收方。出版商出版和分发内容，其他人员（比如机构图书馆）负责确保资料得到保存和可供使用
- 因此，机构图书馆长期以来一直保存着学术内容的产出，时至今日仍然如此。但数字保存有着不同的考量
- 如今，机构图书馆付费以获得各个独立运行的学术资源的访问权限，而不是购买有形的副本并管理它们
- 这解决了科研人员要求方便地远程访问大量学术内容时面临的很多问题！
- 出版商有责任帮助确保这些内容的长期保存



不断变化的要求

- 与此同时，科研产出（“学术内容”）的性质也在改变
- 科研产出的形式更加多样化，不再仅仅是传统期刊文章中的文字和图片
- 以在线形式提供更多的科研产出成为可能，例如：
 - 多媒体和动态对象（比如三维模型）
 - 数据集和可重复代码
- 补充性内容，比如附带的播客或出版后的评论也将为科研人员带来更多有价值的内容
- 这些都是有效保存面临的挑战

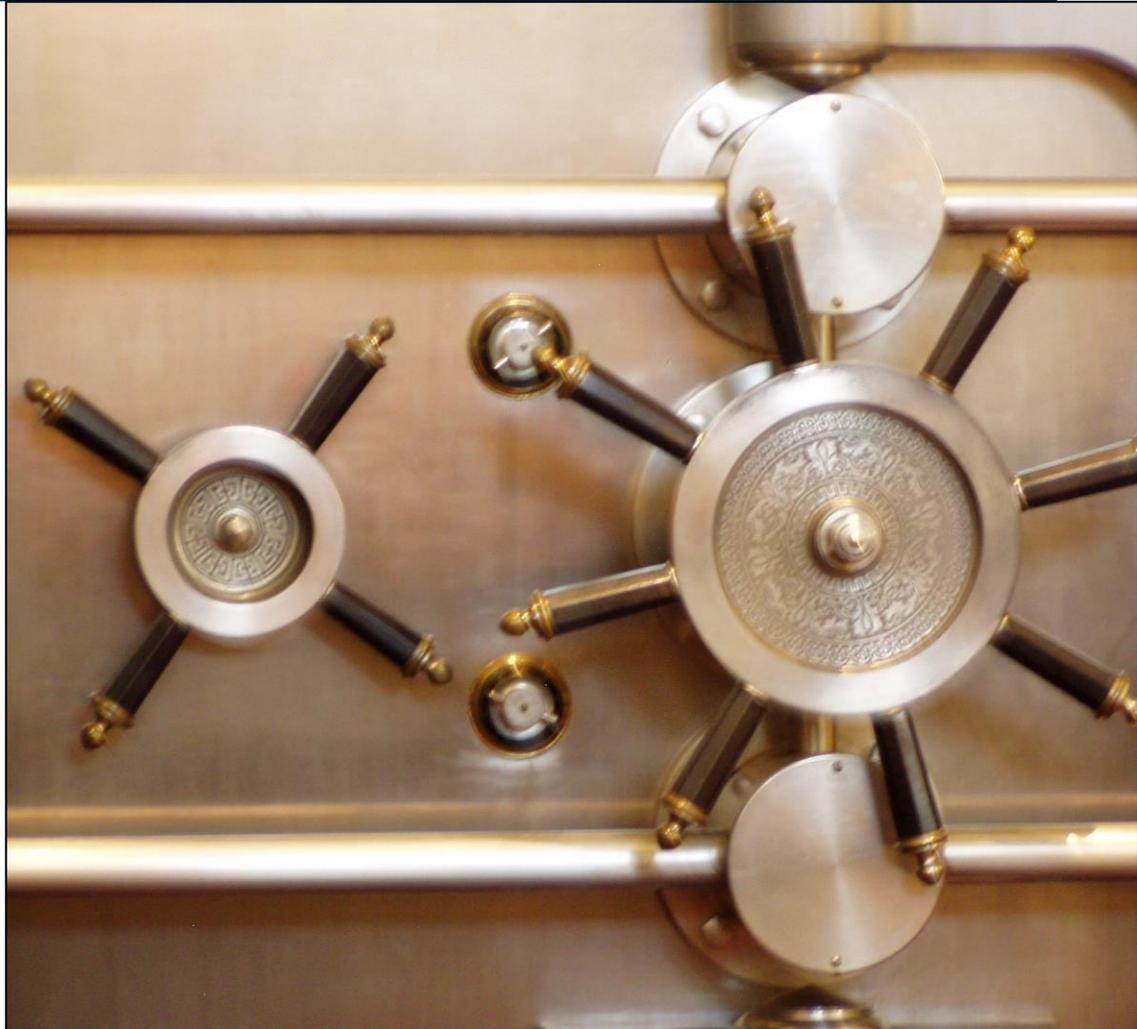


不断变化的要求



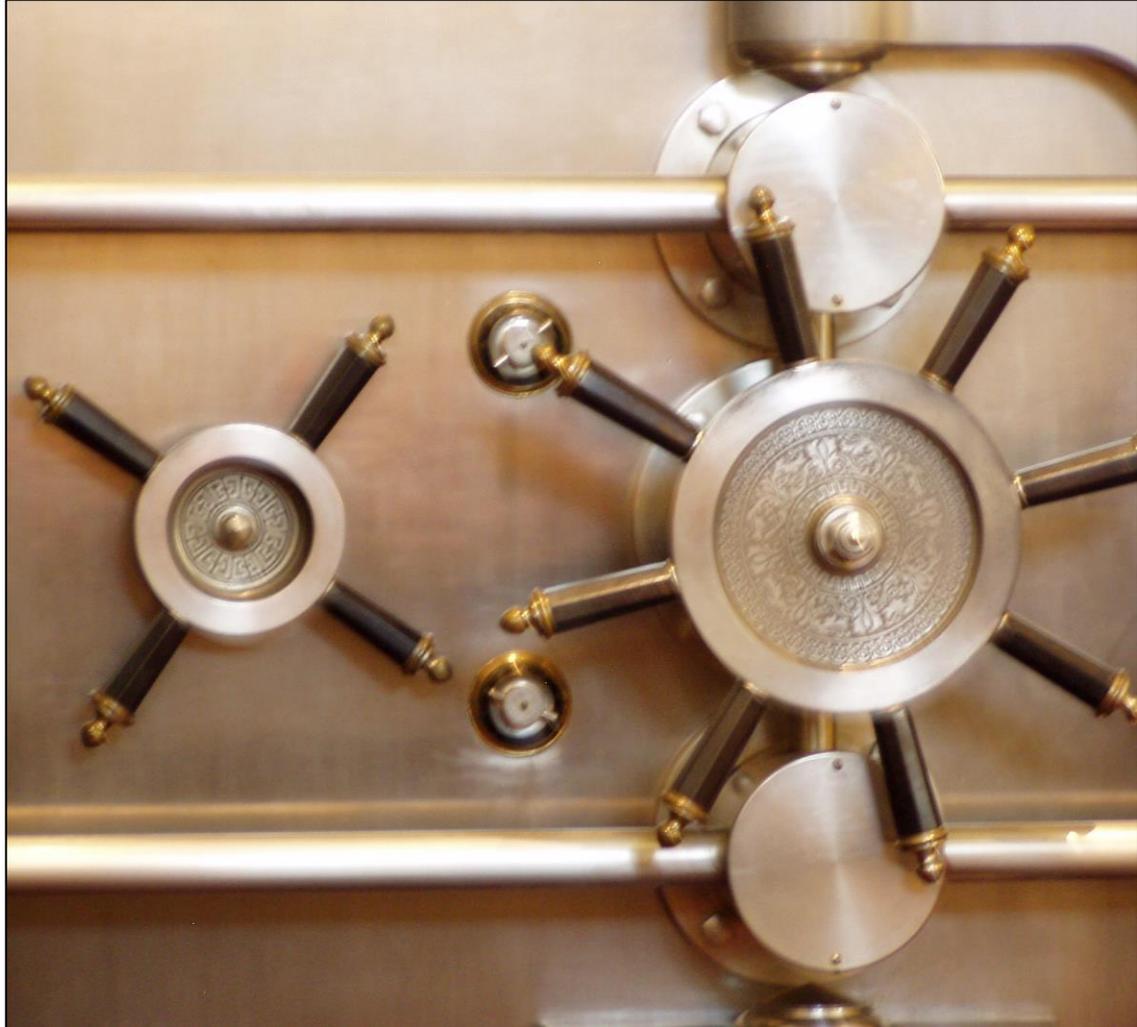
什么是保存？

- 在数字时代，学术内容的保存十分重要。但什么是“保存”？
- “数字保存是指为确保数字资源在尽可能长的时间内可持续访问而必需的一系列管理活动”（来自数字保存联盟的定义）
- 数字保存既是一个持续的过程，也包括支撑这一过程的服务



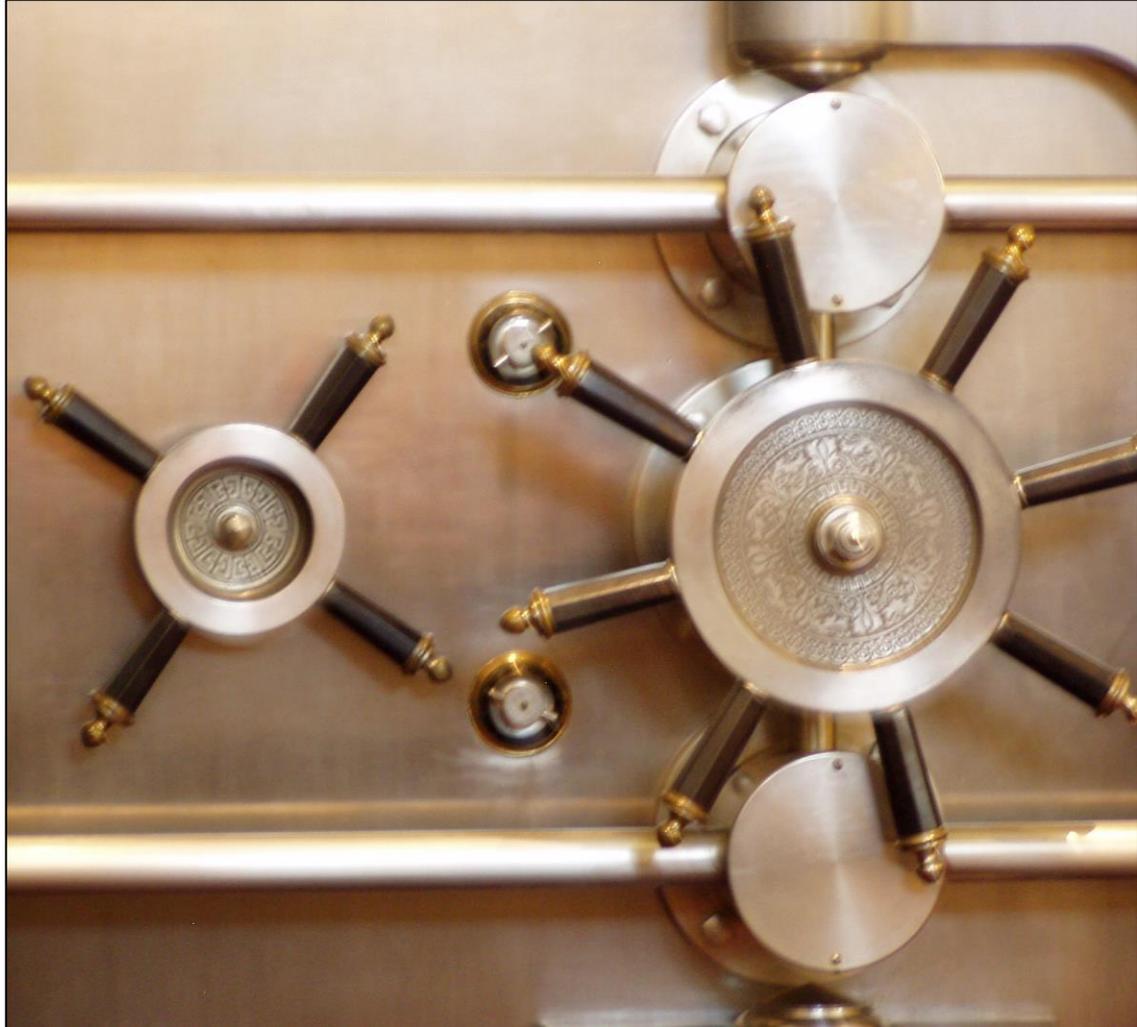
什么是保存？

- 备份是最简单有效的方式
 - 每个内容至少有三份拷贝
 - 至少一个拷贝是离线的或是在云端
 - 云存储已十分常见并成为备份的有效方式
- 备份**不是**长期保存
 - 它们不能保证可访问性
 - 它们不能保证有效性
 - 那么什么才是**保存**？



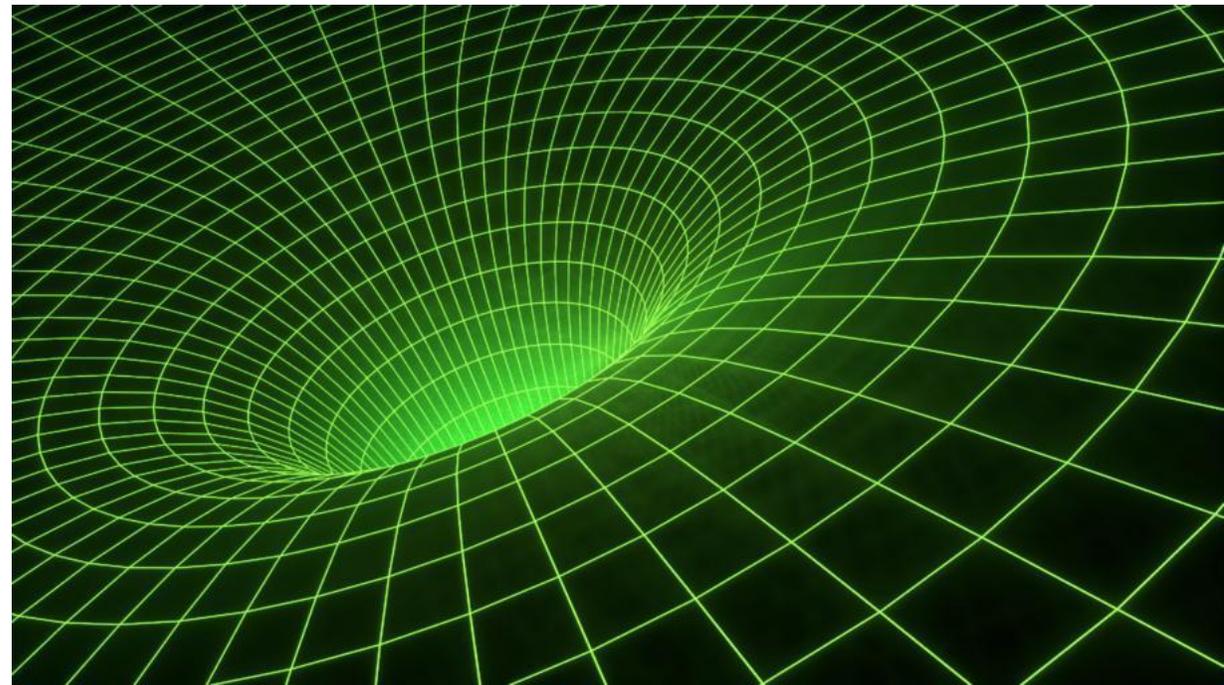
什么是保存？

- 内容的多个分布式副本
- 无限期保存内容
- 以适合长期保存的格式存储内容
- 使用适当的保存元数据存储内容
- 审核和维护内容
- 一旦“触发事件”发生，则提供对内容的访问



什么是保存？

- “触发事件”十分关键
- 当访问内容的方式不再可用时，将引发“触发事件”，例如：
 - 无法再访问特定内容
 - 期刊停刊
 - 出版商平台持续故障
 - 出版商的运营完全终止
- 当“触发事件”发生时，保存服务将确保相关内容的可用性



什么是保存服务？

- 特定的服务提供有保证的保存
- 与这些服务项目的合作是所有国际领先出版商的专业性体现之一
- 出版商将内容提供给这些服务
- 这些服务确保内容按照安全的保存规范保存
- 除非发生“触发事件”，否则不能通过服务项目访问内容（“不公开存储”）

什么是保存服务？

- 解决了科研人员关心的问题：
 - 他们可以确信学术内容持续可用
- 解决了机构图书馆关心的问题：
 - 内容与出版商独立，得以有效地保存
 - 机构不必储存自己的内容副本，以减轻长期保存带来的问题（尽管他们可能因其他原因而仍需这样做）
- 为出版商的职责提供支持：
 - 它们与出版商提供“有效”内容访问权限的初衷或职责并不冲突

什么是保存服务？

- 与出版商和机构图书馆合作的主要保存服务示例：
 - CLOCKSS
由相应社群管理的学术内容不公开存储，在全球12家领先图书馆保存所有内容副本，运行LOCKSS以确保数据保持有效。CLOCKSS“触发事件”发生时，其内容将提供开放访问
 - Portico
由相应社群支持的不公开存储。对于参与组织，Portico还可以支持实现退订后的访问权限，以及“触发事件”



什么是保存服务？

其它保存方案示例：

- LOCKSS (Lots of Copies, Keep Stuff Safe)
 - The Global LOCKSS Network (GLN)
 - Private LOCKSS Networks (PLNs)
- Trusted Digital Repositories (TDRs)
 - Canadiana
 - Chronopolis
 - HathiTrust
 - Scholars Portal
- 国家保存计划，例如：
 - British Library
 - Library of Congress
 - Koninklijke Bibliotheek
 - Chinese National Library of Science

The Keepers Registry

- 由一个或多个保存服务保存的期刊索引
- 机构图书馆员可以使用Keepers Registry检查其保存的馆藏书目
- 由ISSN国际中心运行和维护

The Digital Preservation Coalition (DPC)

- DPC致力于促进和实现数字内容和服务的灵活长期访问
- 他们制作了一本数字保存手册，“为随着时间的推移管理和持续访问数字资源提供了国际权威的实用指南。所有参与数字资源生产和管理的人都将对此感兴趣”

The Transfer Code of Practice

- 提供“一致的准则，帮助出版商确保期刊在各方之间转移时图书馆员和读者能够继续访问期刊内容，并保证在转移过程受到最小的影响”
- 期刊转移有可能引发“触发事件”，这一准则列明了期刊转出和转入出版商的职责，以确保其根本诱因得以解决

OUP的保存政策

内容保存是现代出版的一个关键要素。它使得我们在无法提供正常服务、维护已停刊的期刊内容和履行我们协议中对客户的永久访问承诺时，提供内容的持续访问。

通过在全球范围的出版工作，推动研究、学术和教育进一步臻于卓越是OUP使命的核心。保存我们出版的内容十分重要，它可以使科研人员能够持续访问经过完整编辑加工的高质量内容（无论是过去、现在还是将来的出版物），从而自信而精准地审视他们周围的世界。

随着学术界不断创新并加速其科研产出，对OUP而言，至关重要的是为未来的技术和出版形式做好准备。



OUP的保存政策

在OUP，我们致力于：

1. 学术研究的永久保存和可用性
2. 与外部保存服务的伙伴关系
3. 确保持续访问转移的第三方内容
4. 出版社内严格的内容存档流程和系统的投入
5. 持续的存档维护、适当的存储内容格式的转换和升级
6. 参与内容保存相关的业界关注焦点和行业举措



OUP的保存方案

严格的内部存档

- 大量的投入
- 我们正将我们的内容存储迁移到基于云端的通用方案
- Atlas系统：
 - 存储和分发任意文件类型和文件大小
 - 在任意必需的数据模型中存储和分发XML内容
 - 按自动计划将内容分发到外部节点
 - 使用FTP、SFTP、S3或API分发内容
- 所有内容都将遵循这一方案，以确保全面的内部存档和对保存服务以及其它合作伙伴的有效交付



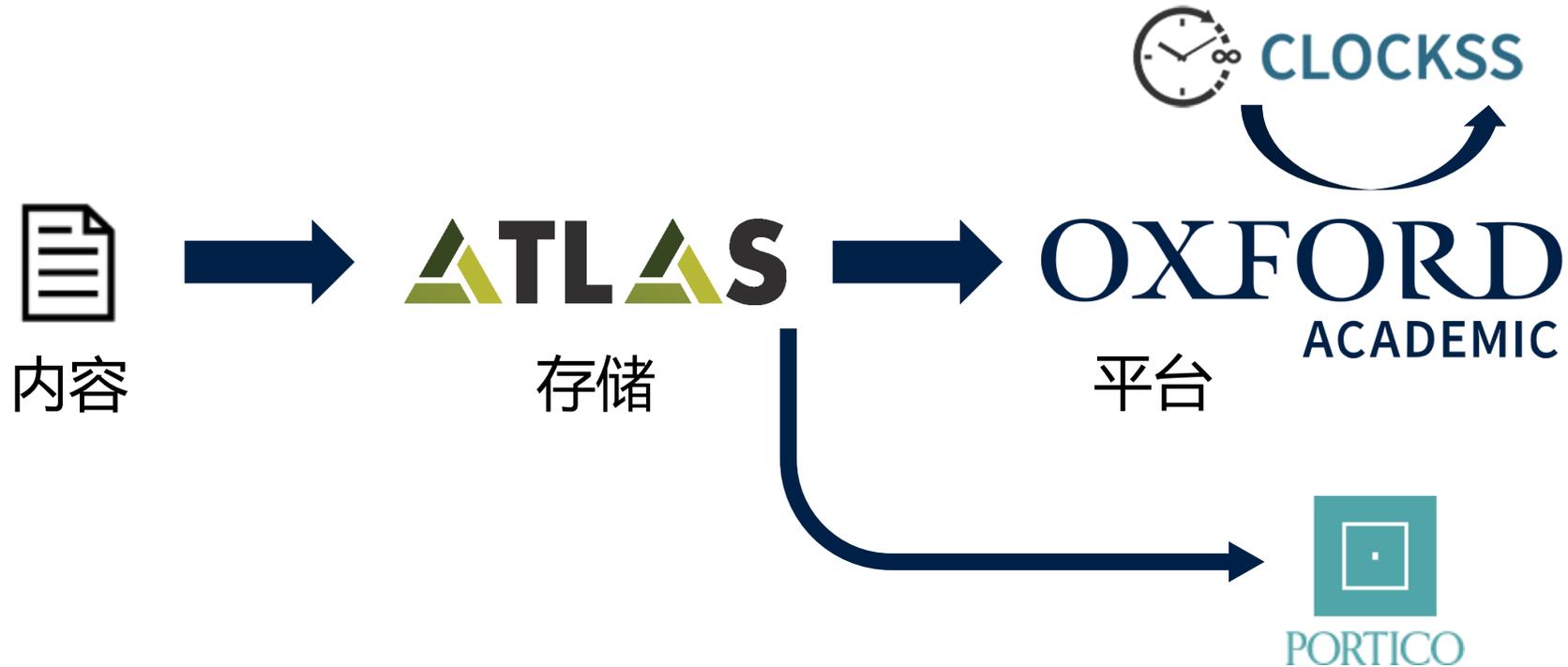
OUP的保存方案

与外部保存服务的伙伴关系

- 核心保存协议：
 - 期刊内容：CLOCKSS
 - 期刊和在线产品（书籍内容）：Portico
- 交付至国家存储
- 支持Transfer Code of Practice
- 行业参与，包括
 - CLOCKSS委员会成员
 - Transfer委员会成员
 - 参与关注机构图书馆教育和关注事项的项目（例如NASIG Digital Preservation Committee）

OUP的保存方案

牛津期刊内容保存流程示例



OUP的保存方案

当前的焦点

- Atlas上线
- 将内部存档迁移至此新解决方案

重点内容

- 直接交付至保存服务
- 审核以确保保存服务的全面覆盖
- 提升图书内容数据模型的互操作性（移至BITS）
- 持续评估和完善我们的方案和合作伙伴关系

结语

- 随着学术出版逐渐转向以在线形式发布内容，许多新的问题引发了科研人员及其所属科研机构以及出版商们的思考
- 学术内容的保存是其中备受关注的问题之一
- 确保学术内容可以长期访问十分重要，无论是访问内容的科研人员，为内容投入的机构，还是提供内容的出版商，都很重要
- 内容保存面临的挑战随着科研产出的不断演变而变化
- 解决学术内容的长期保存问题是OUP等出版商专业性的核心
- OUP通过严格的内部存档、与成熟的国际保存服务伙伴合作以及积极参与行业组织，直接参与了以上内容



谢谢！