

机构知识库：颠覆与创新



张晓林

中国科学院文献情报中心

China IR 2015

2015.09.24



机构知识库：颠覆与创新

- 1. 机构知识库的发展趋势
- 2. 机构知识库的严峻挑战
- 3. 机构知识库的创新需要
- 4. 中国：你能做什么？

1. 机构知识库的发展趋势



主要国家纷纷
推动信息共享
促进经济发展

多数科研资助
机构出台
开放获取政策

机构知识库
及其内容规
模持续增加

机构知识库
互操作富内
容多功能

2. 机构知识库的严峻挑战

机构知识库的空壳化挑战

- Elsevier IR API
- CHORUS
-
- 机构知识库的内容权挑战
 - 超长的时滞期
 - “追责” PDF版？
 -
- 机构知识库的主导性挑战
 - 从支付图书馆订购转换到支付开放出版？
 - ArXiv, LinkedIn, ResearchGate, Academia, CiteULike
 -

1. 机构知识库的严峻挑战

机构知识库的有效性挑战

- 有人存缴吗？政策遵从度检验
- 如何制约？存缴遵从度制约
-

■ 机构知识库的规范性挑战

- 标识、描述、引用、许可等的规范性？
- 管理措施的规范性？
-

■ 机构知识库的有用性挑战

- **Openness for computing**
- **Richness of content and their supplementary data**
-

1. 机构知识库的定位变迁

■ 从支持科研与教育的视角

- 科研工作流就是科研信息流（杨国帧）
 - 如何支持信息驱动的科研管理、科技创新和科研知识管理？
- IR是机构知识创造与传播流程中的有机环节
 - [独立]机构知识库已死（IR is dead），Tom Cramer, 2010
 - 让IR消失在科研流程中（Invisible IR），W. Horstman, OR2013
- 支持开放数据应用？
 - 可获取、可理解、可评价（RS: Science as an Open Enterprise）
 - 开放许可、可计算能力、规范格式、开放接口（UK ODWB,2012）
- 支持科研可重复性（Reproducibility）？
 - Borgman, C.著，青秀玲译。科研数据共享的挑战。2013
 - Reproducibility Initiative
 - <https://www.scienceexchange.com/reproducibility>



1. 机构知识库的定位变迁

- 从机构知识管理角度
 - 基于证据的科研管理与评价
 - 知识成果保存评价
 - 知识成果利用影响力评价
 - 知识审计（知识能力、知识竞争力、知识生产率）
 - 基于知识的知识产品再创造与扩展利用
 - 衍生与集成知识内容的生产？
 - 知识转移转化管理？
 - 科学普及与科学教育？
 - 基于知识的科技战略管理
 - 知识图谱、知识能力发现与对接
 - 知识创新能力综合评价
 - 研究与创新路线图挖掘？
 -



机构知识库发展趋势与可能挑战

- 1. 机构知识库的定位变迁
- 2. 机构知识库支持非文本内容
- 3. 机构知识库支持科研过程
- 4. 机构知识库支持机构知识管理



2 机构知识库与非文本知识内容

- IR支持NTM:
 - 支持各类非文本内容的存缴、描述、保存和使用
- 技术问题
 - 数据格式、内容元数据、数据集标识与引用、数据格式或元数据转换、数据保存包组织等
- 保存管理问题
 - 复杂数据几何管理、数据转换管理、数据溯源管理
- 使用管理问题
 - 数据权益描述、数据转移管理、使用许可管理
 - 使用的细粒度管理（包括“开放数据”的使用管理）



2 机构知识库与非文本知识内容

- IR支持NTM:
 - 全流程管理/服务前置
 - 只管理“最后成果”导致“没有成果”
- 支持科研数据管理规划
 - 资助机构DM要求、规划模板、规划范例、咨询评估工具与服务，例如，PURR
- 支持科研数据管理流程
 - 项目组将数据管理计划转化为基于科研过程的数据流程模块，支持项目组在各阶段保存各类数据，管理各阶段数据的上传、共享、转换、转移、关联、修改等
- 支持科研数据管理的工具与服务
 - 采集、清理、审核、转换、提取、融汇、可视化...



2 机构知识库与非文本知识内容

- IR支持NTM:
 - NTM的管理和利用更加丰富
- 支持NTM的开放处理与管理
 - 嵌接开放科研或社交网络，支持对数据采集、分析、判读、扩展、关联的众包（Crowd-sourcing）和公民科研。
- 支持NTM的内容挖掘与分析
 - 嵌接各类NTM内容识别、分析、挖掘等的工具与服务，支持细粒度的语义化的检索、分析和利用。
 - 支持音视频的对象识别、场景划分、内容与故事构建...
 - 对数据进行荟萃分析（Meta-analysis）
- 支持关联数据服务
 - 发布关联数据、建立关联服务



机构知识库发展趋势与可能挑战

- 1. 机构知识库的定位变迁
- 2. 机构知识库支持非文本内容
- 3. 机构知识库支持科研过程
- 4. 机构知识库支持机构战略规划



3 机构知识库支持科研过程

- 信息驱动的科研管理
 - IR内容与科研管理系统的有机链接
 - Michael Day, 集成IR和科研信息管理系统, 2010
 - RIOXX元数据集中的项目信息元素
 - IR支持数据管理流程
 - 支持数据管理计划编制 (PURR)
 - 支持科研过程协同数据管理 (PURR)
 - IR enhanced content?
 - Depositing and linking supplementary materials?
 - Integrating aggregated content (via OAI ORE)?
 - Supporting open linking into research processes
 - Standard Identifiers
 - Standard metadata, enriched metadata
 - Standard openAPI, linked open data

3 机构知识库支持科研过程

■ 信息支持的科研教育创新过程

- **Make IR metadata as Open data**
- **Support Open Use of IR full-text content**
- **Provide tools for re-using IR content**
- **Manage rights in data and rights in open data use**

Level	Format	Licence
★	Make your data available on the web (in any format)	and Open Licence
★★	Make it available as structured data (for example, Microsoft Excel instead of image scan of a table)	and Open Licence
★★★	Make it available in an open, non-proprietary format (for example, CSV or XML instead of Microsoft Excel)	and Open Licence
★★★★	In addition to using open formats, use Uniform Resource Locators (URLs) to identify things using open standards and recommendations from W3C, so that other people can point at your stuff	and Open Licence
★★★★★	In addition to using open formats and using URLs to identify things, link your data to other people's data to provide context	and Open Licence

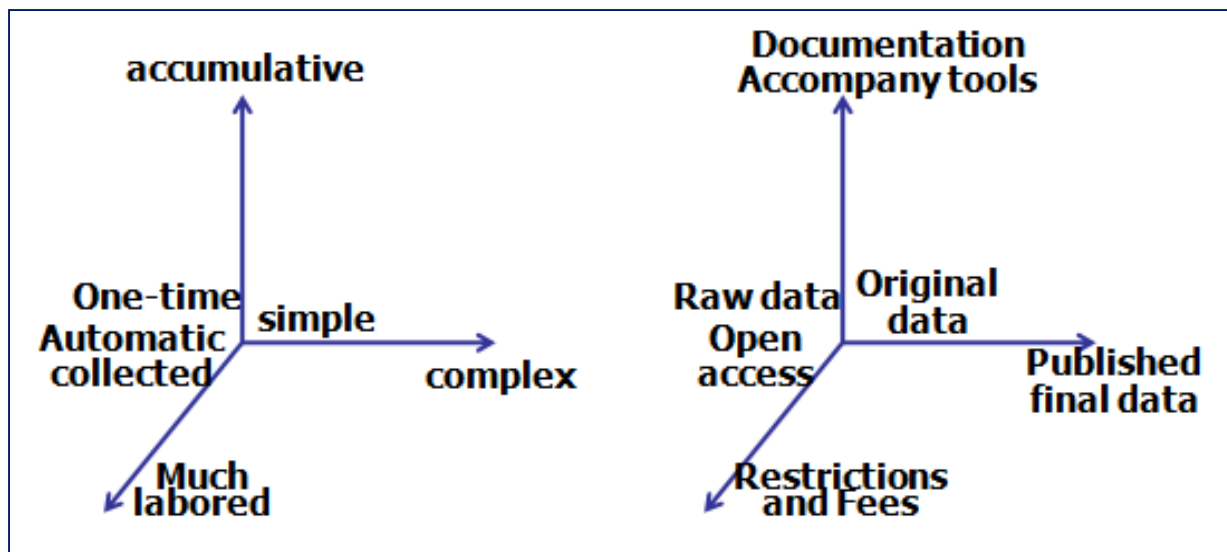
开放数据应用管理

- **Measures**
 - **Complete open**
 - **Open while monitoring**
 - **Open after registration**
 - **Open after agreements**
 - **Open for individual use**
- **Factors to consider**
 - **Frequency, Data scale, Type of data**
 - **Nature of data or use**

3 机构知识库支持科研过程

- 支持科研可重复性

- 如何支持科研可重复性？（Borgman/青秀玲）



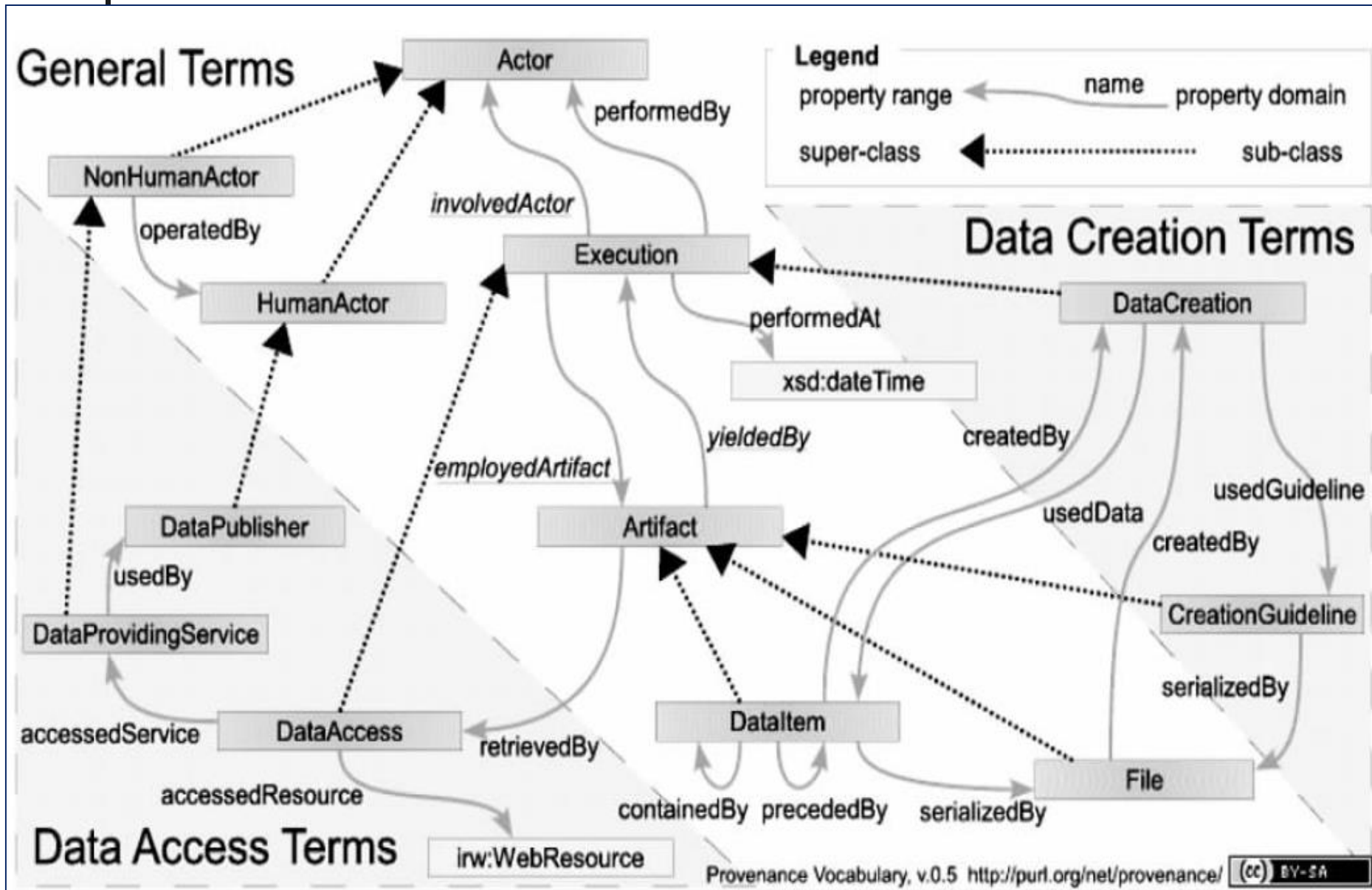
- 支持数据全流程管理，支持数据溯源管理

- Metadata/Ontology, Data+Process tools, IR/DR support
 - Data provenance (Buneman, 2011)
 - A Survey of Data Provenance Techniques (Simmhan 2005)
 - 语义网环境下数据溯源表达模型研究综述（沈志宏等）

IR: 颠覆与创新-张晓林-2015



Data Provenance Vocabulary





机构知识库发展趋势与可能挑战

- 1. 机构知识库的定位变迁
- 2. 机构知识库支持非文本内容
- 3. 机构知识库支持科研过程
- 4. 机构知识库支持机构知识管理



4 机构知识库支持机构知识管理

- (1) 基于证据的科研成果管理与评价
 - **How to evaluate the rate of deposit**
 - 与综合文摘索引系统的比对分析？
 - 与论文引证系统的对比分析？
 - 与科研管理/学位管理系统申报数据的比对分析？
 - **How to evaluate the impact of research results**
 - IR内容→引文数据→引文网络（谁来引用？引用文章影响）？
 - IR内容下载及其细粒分析（谁在下载、下载趋势）？
 - 其他Article level metrics？
 - **How to conduct knowledge accounting（知识审计）**
 - 基于产出的知识能力统计
 - 基于产出成果发布与利用的知识利用能力统计
 - 知识生产率
 - 考虑人员、经费、时间、项目层次等基数的知识产出与影响力分析



4 机构知识库支持机构知识管理

(2) 基于知识的知识产品再创造与扩展利用

■ 衍生知识内容的生产

- 横向的文集汇编工具（相关相似判断、集成结构）
- 纵向的文集汇编工具（相关相似判断、演进结构）
- Overlay Journal生成工具？
- 内容提炼、演示文档或专著生成工具？

■ 知识转移转化管理

- 知识优先权管理
- 专利关联分析（引用、创造者、主题关联？）
- 专利可转化性分析（研究宽度与厚度、研究可持续性）

■ 科学普及与科学教育

- 教学资料包组织工具？
- 教学或科普系统辅助工具（内容关联、嵌入、提取、PPT化？）
- “教育可用”或“科普可用”资料？（视频、PPT等？）



4 机构知识库支持机构知识管理

■ (3) 基于知识的科技战略管理

■ 机构知识图谱

- 谁在做什么？（重复？特色？合作机遇？可利用全文分析）
- 还有谁在做这些？（扩展图谱和关联分析；合作效益？）
- 我们遗漏了什么？（空白？偏差？矛盾？机遇？）

■ 知识创新能力评价

- 谁做得怎么样？（内部和与外部的比较评价？）
- 谁在引领和突破？（与竞争对手和业界态势的全面比较）
- 创新生产率评价？（与项目管理结合，考虑投入资源？）

■ 研发规划与路线挖掘

- 科研布局支撑能力分析？（谁究竟有能力做什么？）
- 基于态势异常点的研发规划分析？
- 基于IR内容的研发动态与路线挖掘分析？
-



机构知识库发展趋势与可能挑战

■ 参考文献:

- Borgman, C.著, 青秀玲译。科研数据共享的挑战。现代图书情报技术, 2013年第5期, 1-20.
- Brase, J. and Blumel, I. Information supply beyond text: non-textual information at the German National Library of Science and Technology (TIB) – challenges and planning. *Interlending & Document Supply, Volume 38 Issue 2, 2010.*
- Buneman, P. and Davidson, S. B. Data provenance – the foundation of data quality. September 1, 2010
- CALL: International Journal of Digital Libraries Special Issue on Digital Scholarship. <http://duraspace.org/call-international-journal-digital-libraries-special-issue-digital-scholarship>
- Cramer, Tom. Digital Library 2.0: Trends in Management, Access & Preservation. Chief Technology Strategist, Stanford University Libraries, September 22, 2010
- http://fez.schk.sk/eserv/changeme:6587/DL_20100923_Jasna_Cramer.pdf



机构知识库发展趋势与可能挑战

■ 参考文献:

- Day, Michael. Models for integrating institutional repositories and research information management systems, UKOLN, *Position paper for the Workshop on CRIS, CERIF and Institutional Repositories, CNR, Rome, 10-11 May 2010*
- eScidoc, <http://escidoc-project.de/homepage.html>
- Goportis Conference 2013 on Non-Textual Information Strategy and Innovation Beyond Text. March 18-19, 2013, Hannover, Germany.
<http://www.nontextualinformation2013.de/>
- Hey, Tony等著, 潘教峰等译。科学研究的第四范式：数据密集型知识发展。科学出版社, 2012。
- Horstmann, Wolfram. Invisible Repositories, Re-Use and Re producible Research. Open Repositories 2013, July 9-11, 2013 <http://or2013.net/sessions/invisible-repositories-re-use-and-reproducible-research>



机构知识库发展趋势与可能挑战

■ 参考文献:

- Lyon, Liz. Reflections on open scholarship: process, product and people. 2nd International Digital Curation Conference, Nov. 2006.
- Purdue University Research Repository. <https://purr.purdue.edu/>
- Simmhan, Y. L. et al. A Survey of Data Provenance Techniques. Indiana University Technical Report IUB-CS-TR618, 2005
 - <http://www.cs.indiana.edu/l/www/ftp/techreports/TR618.pdf>
- Ronald, L. L. On the Threshold of Cyberscholarship. Journal of Electronic Publishing, 11(1), Winter 2008.
- TIB Competence Center for Non Textual Materials. <http://www.tib-hannover.de/en/services/competence-centre-for-non-textual-materials/>



机构知识库发展趋势与可能挑战

谢谢!

<http://ir.las.ac.cn>