关于
机构知识库（IR）可持续发展
的思考与探讨

李箐
IEEE 中国区资讯经理
美国雪城大学信息学院博士生

2010年5月12日

◆IEEE

什么是机构知识库(IR)?

*An institutional repository (IR) is "**an electronic system** that captures, preserves, and provides access to the digital work products of a community."*

- Foster & Gibbons (2005)

IR is "**a set of services** that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members."

- Clifford Lynch (2003)

IR的起源

- ◩ Open Access 概念的兴起
  - 来自研究团队的推动

- ◩ Open Access 软件的发展
  - E.g. DSpace, E-print, Fedora etc

- ◩ 电子资源的购买费用
  - 来自图书馆的担忧

- ◩ 最初的IR存档对象主要针对e-prints

◆IEEE

IR在美国高校的前期发展-分布

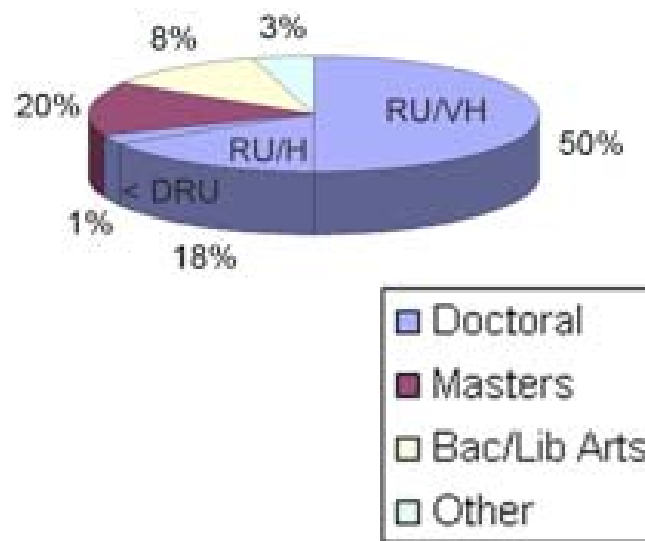美国高校大部分IR由图书馆启动并承担主要工作，包括建立IR基础设施、制定政策以及内容收集等。

Fig 1.1 IRs by institutional Carnegie classification

McDowell, C.S. (2007)
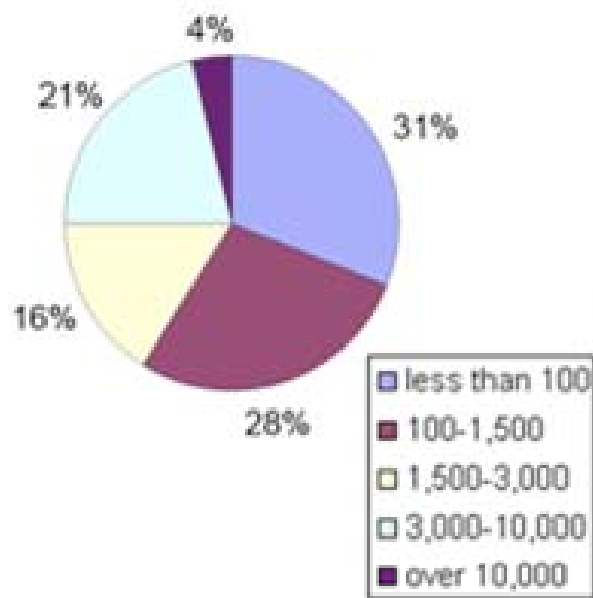
◈IEEE

IR在美国高校的前期发展-规模

▣ IR内容数量是最重要的指标之一

Fig. 2.2
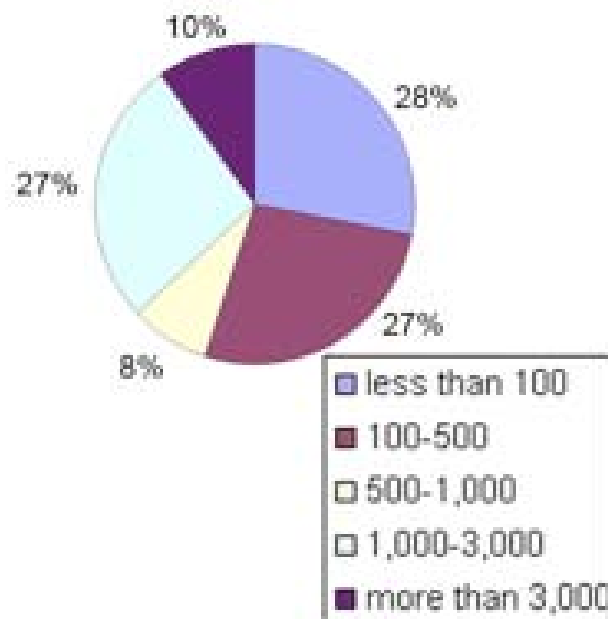Repository size by items Nov. 2006

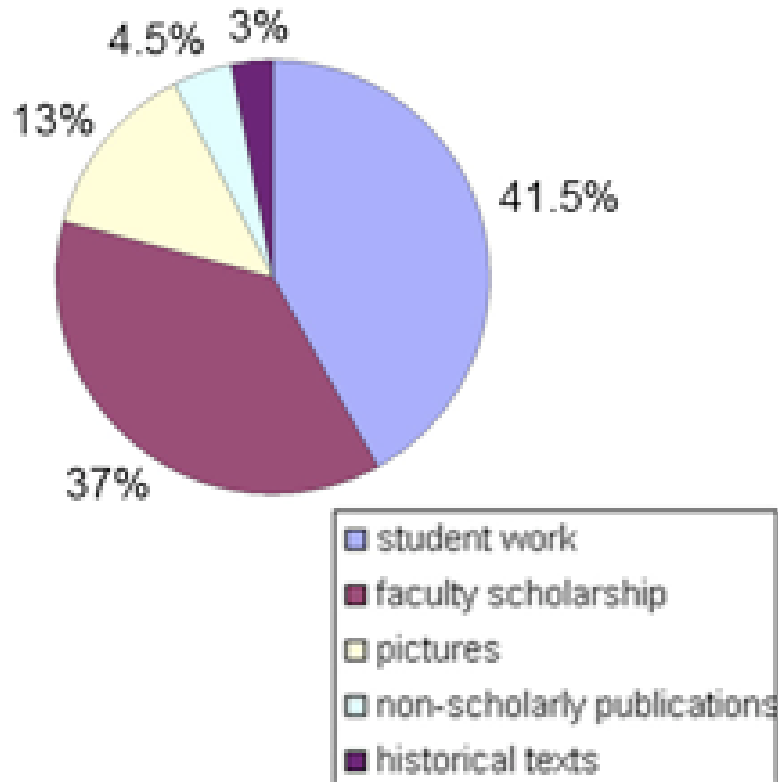Fig. 3.3 Annual growth by items

McDowell, C.S. (2007)

IR在美国高校的前期发展-数量困境

- 内容收集是美国高校IR面临的最大问题，很多学校发现连收集最初的100篇都相当难

"The phrase 'if you build it, they will come' does not yet apply to IRs"

-Foster & Gibbons (2005)

IEEE

IR在美国高校的前期发展-内容

4.5% 3%
13%
41.5%
37%

- ■ student work
- ■ faculty scholarship
- □ pictures
- □ non-scholarly publications
- ■ historical texts

Fig. 4.1 Estimated content types by %

e-prints比例：其中同行评审内容-pre- and post-prints, e-journal articles, and e-books大约只有13%.

IR在美国高校的前期发展-质量困境

- Faculty Output在美国大多数高校IR前期建设中没有起到预期作用

- 高质量内容如同行评审文献平均只占总内容的13%

- 高质量内容的存档难度使得IR并没有太多的实现最初的OA运动的目标

McDowell, C.S. (2007)

◆IEEE

为什么**IR**会面临双重困境？

IR如何能够可持续发展？

IEEE

IR is not FREE

■ 短缺的资金
  – 很多**IR**由一笔特拨基金启动
  – 如何将一个短期的项目转为长期的可持续项目
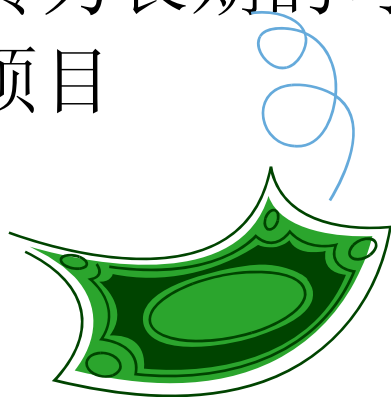
■ 高额的建设维护费用
  – 2002年，MIT图书馆做出估计其**IR**每年花费**28.5万**美元用于员工薪酬、运营以及设备维护。这样每年每个文档的成本是**71**美元。

(Henty, M. 2007)
Barton, M.R. & Walker, J.H. (2002).

IEEE

IR contents are not FREE either

- IR存档内容

  - E-prints (pre-print/post print)

  - Finished work – faculty/student paper

  - Unfinished work/data

问题一:
为何**IR**原本关注的**e-prints**没有成为主角? 本校人员的**e-prints**可以直接存档吗?

数字版权问题

- IR面对的实际问题：哪些数字资源可以OA存档？
  - 各个出版社都有不同的Self-Archiving规定
  - 为避免侵权纠纷，IR需核查相关出版文献 Self-Archiving规定

*"Aspects of copyright that were identified as issues were **the cost of checking individual items to ensure no breach of copyright** and ongoing difficulties in obtaining the relevant permissions from authors and publishers"* – Henty, M. 2007

◆IEEE

SHERPA/ROMEO体系

- SHERPA 项目建立于2002年，全称为**Securing a Hybrid Environment for Research Preservation and Access**)

- SHERPA所制定的ROMEO体系在开放获取研究中非常著 名，该体系收集了出版商允许研究者自我存档（Self-Archiving ）研究成果的政策（http://www.sherpa.ac.uk/romeo.php）

Pre-Print vs Post-Print

ROMEO colour	Archiving policy
green	can archive pre-print *and* post-print or publisher's version/PDF
blue	can archive post-print (ie final draft post-refereeing) or publisher's version/PDF
yellow	can archive pre-print (ie pre-refereeing)
white	archiving not formally supported

参与**RoMEO**体系 的**724**家出版商

SHERPA/RoMEO Colours

http://www.sherpa.ac.uk/romeo/browse.php

- Green (206 = 28%)
- Blue (176 = 24%)
- Yellow (72 = 10%)
- White (270 = 37%)

SHERPA/RoMEO 09-May-2010

Total = 724 publishers

IEEE is a RoMEO "Green" Publisher

http://www.sherpa.ac.uk/romeo/search.php?id=38&fIDnum=|
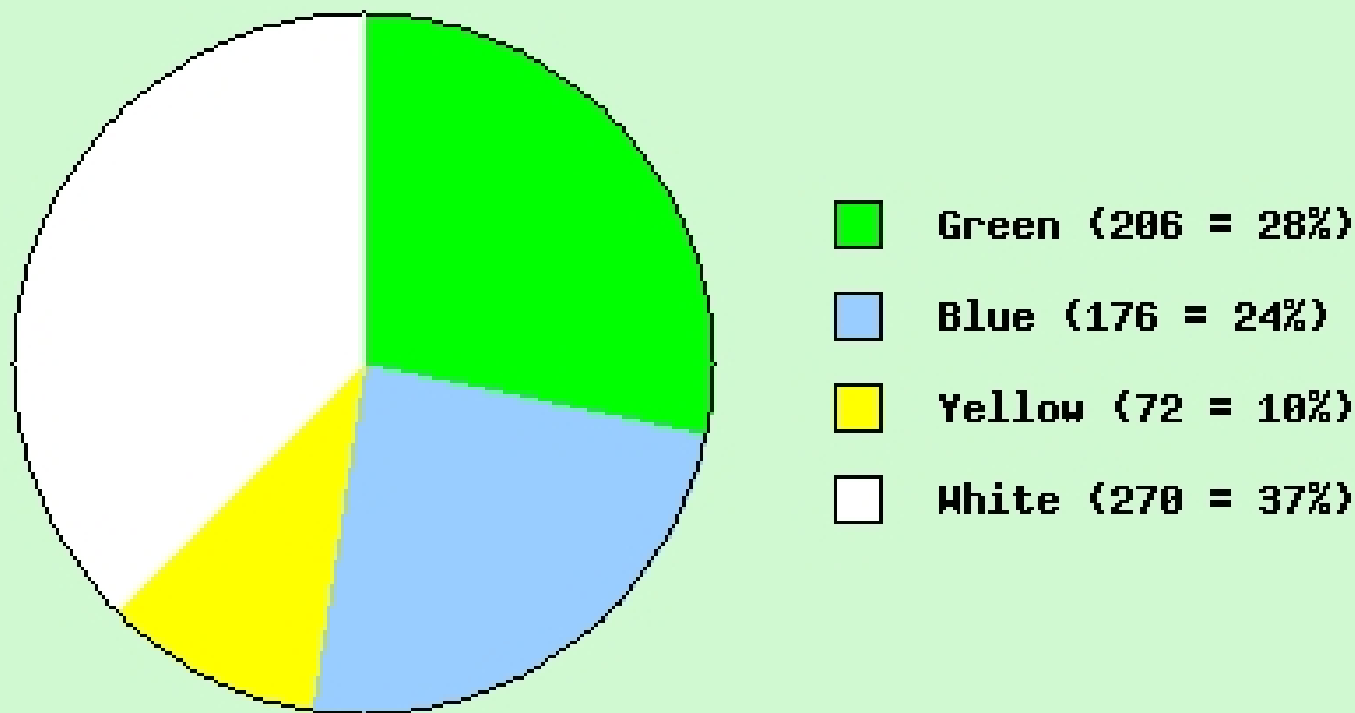
This summary is for the publisher's *default* policies and changes or exceptions can often be negotiated by authors. *All information is correct to the best of our knowledge but should not be relied upon for legal advice.*

Publisher:	Institute of Electrical and Electronics Engineers (IEEE)
Author's Pre-print:	✔ author **can** archive pre-print (ie pre-refereeing)
Author's Post-print:	✔ author **can** archive post-print (ie final draft post-refereeing)
Publisher's Version/PDF:	✔ author **can** archive publisher's version/PDF
General Conditions:	- Authors own and employers publicly accessible webpages - Preprint - Must be removed upon publication of final version and either replaced with citation or Published version with citation - Preprint - Set-phrase must be added once submitted to IEEE for publication (see policy) - Preprint - Set phrase must be added when accepted by IEEE for publication (see policy) - Preprint - IEEE must be informed as to the electronic address of the pre-print - Postprint - Publisher copyright and source must be acknowledged - Post-print- Publisher's version/PDF may be used (preferred)(see conditions for Pre-prints)
Mandated OA:	Compliance data is available for 6 funders
Copyright:	Electronic information dissemination policy (pdf)- section 8.1.9 (pdf) - copyright form (pdf) - other copyright information
RoMEO:	This is a RoMEO **green** publisher
Updated:	23-Apr-2010. Suggest an update for this record

问题二：
为何致力于服务科研工作的**IR**却难以得到用户的踊跃支持？

IEEE

IR建设目标和用户实际需求的差异

- *"I didn't want to go through all of that effort to put up something that was already available".*

- *"...if we stopped 100 of our researchers in the street and asked them what a digital repository was, what was its purpose and how would you access it, and how would you use it and is it important to you, then you'd get a very wide range of opinions. But you get a lot of people going, "a digital what?"*

- "Meeting real needs"

– Henty, M. 2007

用户参与-强制存档**IR**

- 最直接的方式是定制政策要求学校师生必须将其研究成果存档

- 正方经验
  - "Repository managers should invest in promotion and follow–up for two–three years after a mandatory policy is promulgated, after which the behavior becomes routinized"

  - "No especial activities need to be undertaken to convince researchers to deposit research articles soon after publication – this seems to happen naturally under mandatory policies"

- Sale, A. 2006

◆ IEEE

用户参与-强制存档**IR**

- 反方忧虑：
  - "Could the repository service cope with the ensuing demand? Would this create ill-feeling? Would researchers actually comply? And if not, what then?"

  - "I certainly think if mandatory deposit doesn't kind of become generalized, that repositories aren't sustainable, because it takes a lot of work to get people self-archiving. [But] once they're self-archiving, they'll keep on doing it".
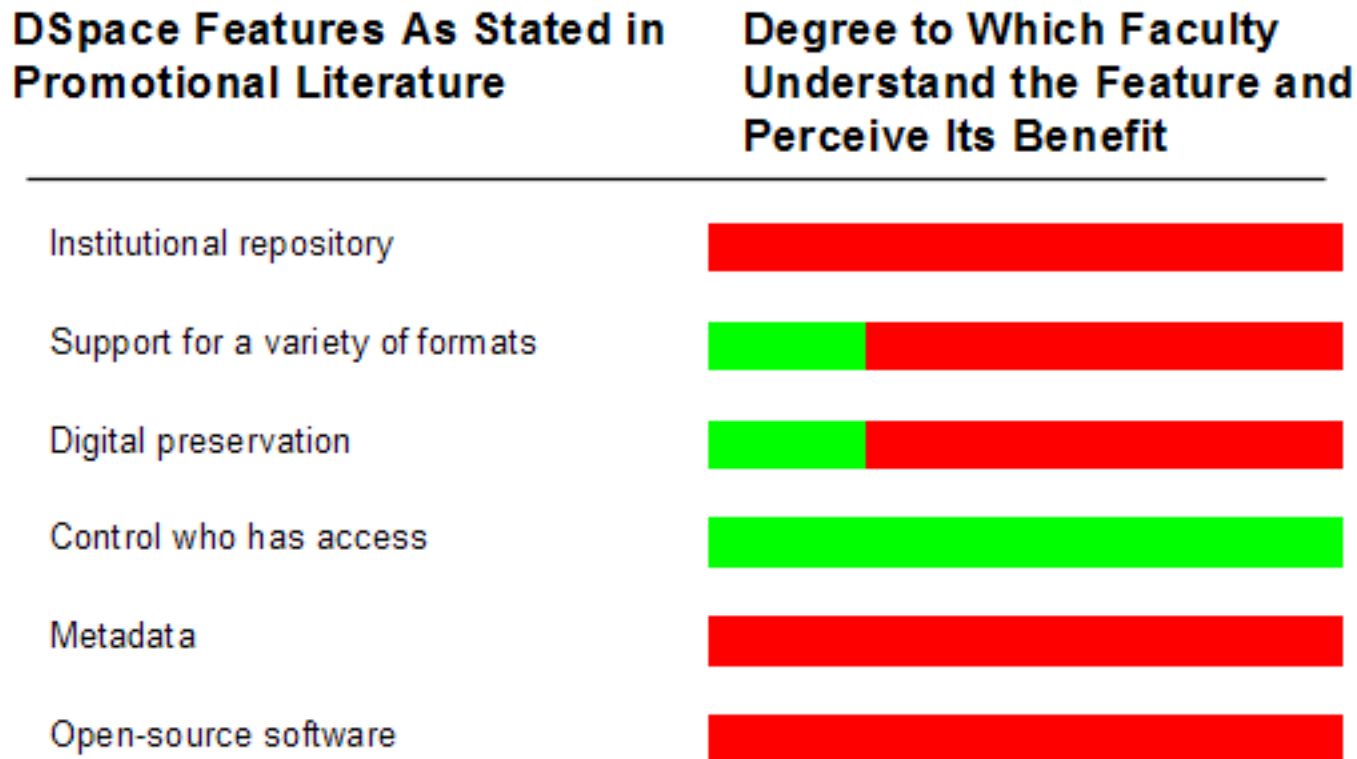
– Henty, M. 2007

IEEE

用户参与-自由存档**IR**

▣ 调查显示最多只有**15-20%**的师生会选择主动将其研究成果在IR进行存档 (Sale, 2007)

▣ 2003年在 Institute of Museum and Library Services基金支助下， Rochester大学图书馆开展研究了解**IR**所提供服务和教师实际需求之间的差异，尤其是了解教师在实际科研中到底做什么，需要怎样的支持

-Foster & Gibbons (2005)

◈IEEE

启动**IR**之前，需要问自己的几个问题

用户能否听懂我们的话？

- 研究发现在推动**IR**的过程中教师和科研人员反映不热烈的原因之一是沟通方式错误

DSpace Features As Stated in Promotional Literature	Degree to Which Faculty Understand the Feature and Perceive Its Benefit
Institutional repository	
Support for a variety of formats	
Digital preservation	
Control who has access	
Metadata	
Open-source software	

-Foster & Gibbons (2005)

用户的愿望我们能够满足吗？

- 方便地与共同作者进行合作
- 管理同一文档的不同版本
- 按照他们自己特定的体系管理文档

- 方便别人访问自己著作以及自己访问别人著作
- 掌握内容的所有权、安全性和访问权限
- 确保不侵犯知识版权

- 有专人负责IR和相关工具的使用和培训
- 不要增加研究人员的工作负担　　　*-Foster & Gibbons (2005)*

换句话说，
IR能否成为用户工作流程的一部分？

- 除了刚才提到的对科研文献的支持外

- 科研数据
  - 2007年以来E-Research/E-Science的兴起使得科学数据的存储、管理和访问提上日程，进一步加大对IR的挑战

◆IEEE

什么是融入科研工作流程的**IR**？

- 案例：FLOSSmole项目-研究者自己启动的IR
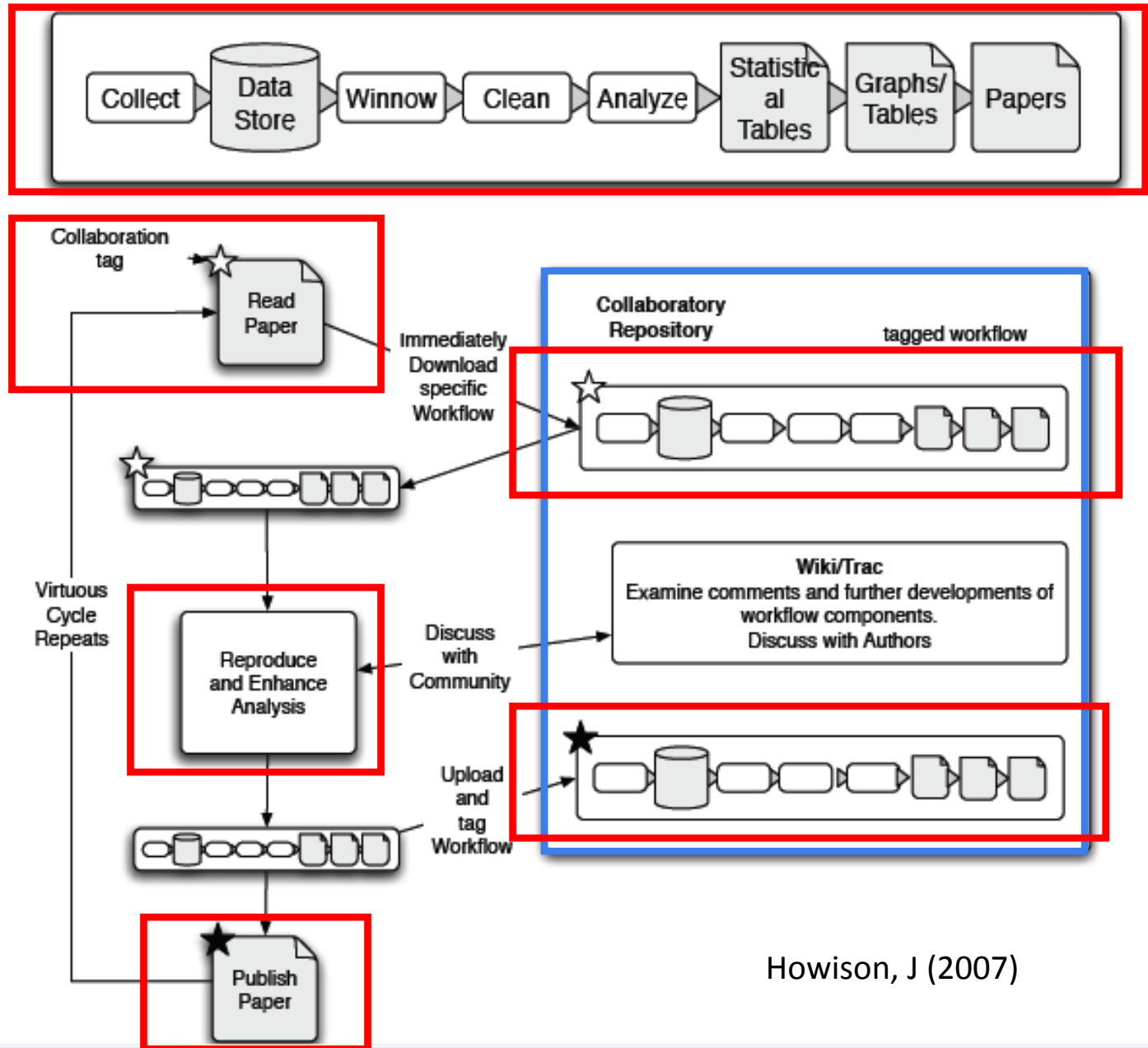 - 美国雪城大学(Syracuse Univ.) FLOSS研究组致力于建立一个协作知识库，为开源代码软件研究者提供数据、论文等各类相关资料的存储、管理和访问服务

科研人员常见难题：
*无法拿到数据
*数据难以理解
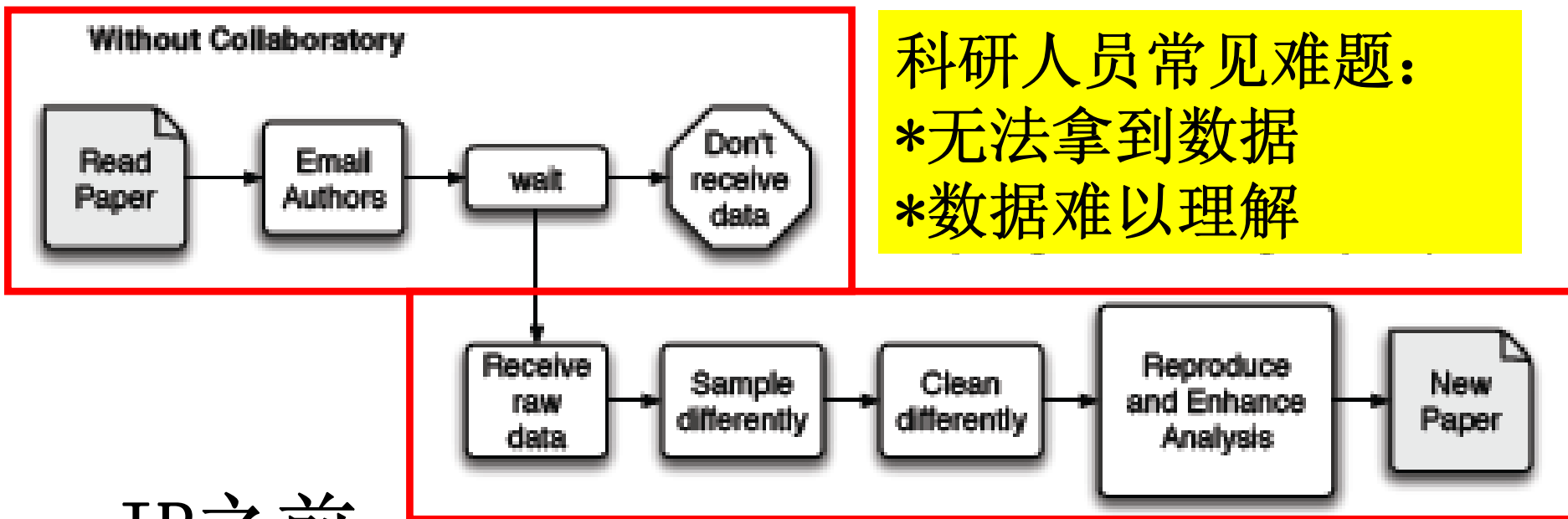
IR之前

(Howison, J. 2007)

Howison, J (2007)

IR之后

论文与数据之间的无缝连接
●原始数据

●同一数据的不同处理版本

●分析结果

●同一论文的不同版本

抛砖引玉

希望今后与图书馆各位老师共同探讨**IR**可持续发展模式

◆IEEE

谢谢各位老师！
欢迎您参加**5月13日11：40-12：20的IEL**
数据库新平台培训！

李箐
IEEE Client Services/UPP Manager
qing-li@ieee.org
http://www.ieee.org/go/clientservices

◇IEEE

参考文献

- Barton, M.R. & Walker, J.H. (2002). *MIT Libraries' DSpace Business Plan Project: Final Report to the Andrew W. Mellon Foundation*, p. 33. http://libraries.mit.edu/dspace-mit/mit/mellon.pdf

- Clifford Lynch (2003)"Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL Bimonthly Report* 226 (2003). http://www.arl.org/newsltr/226/ir.html

- Foster, N.F. & Gibbons, S. (2005). D-lib Magazine, 11:1. http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/january05/foster/01foster.html

- Gandel, P., Katz, R. & Metros, S. (2004). The Weariness of the Flesh: Reflections on the Life of the Mind in an Era of Abundance. *Educause Review*, 39:2, pp. 40-51. http://www.educause.edu/apps/er/erm04/erm042.asp

- Hentv, M. (2007). D-lib Maganize, 13:5/6. http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/may07/henty/05henty.html

- Howison, J (2007) e-Social Science 2007. http://floss.syr.edu/Presentations/HSDposter_8.ai.pdf

- McDowell, C.S. (2007). D-Lib Magazine, 13:9/10.http://www.dlib.org/dlib/september07/mcdowell/09mcdowell.html

◈IEEE