



北京理工大学

国内外数据平台调研分析

报告人：楚恒亚

指导老师：崔宇红

2015年12月6日

研究背景

1

研究方法

2

结果分析

3

建议与思考

4

Contents

目录

研究背景

01

“大数据 (Big Data) ”频繁出现在Science , Nature , The Economist等国际著名杂志上

02

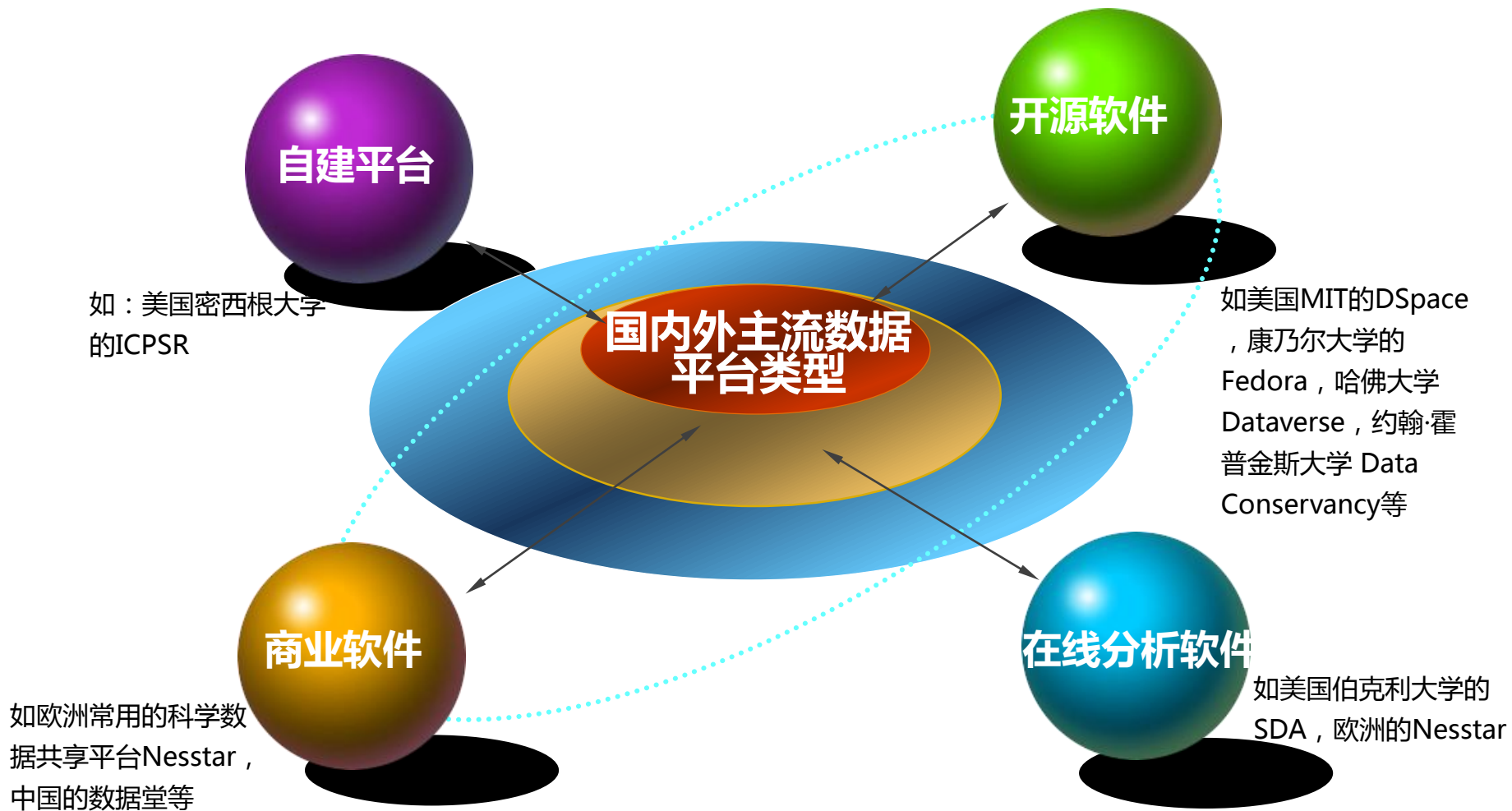
研究数据共享越来越被认为是学术和科学研究的重要组成部分。

03

数据发布和数据共享需要高效的数据管理和健壮的数字基础设施。数据知识库 (data repository) 可以在科研项目成立以来整个生命周期内从数据收集、分析、记录、出版、策展 (curation) 、保存和管理等工作等很多点进行部署。



研究背景



研究背景

1

研究方法

2

结果分析

3

建议与思考

4

Contents

目录



研究方法

文献调查

通过对相关研究数据平台文献的收集、整理，紧紧围绕调查目的，选择适当的统计方法和指标，对国内外数据平台的特征和功能进行解析，不同数据平台之间进行差异性分析。

网站调查

通过登录数据平台的官方网站，进行实地考察，了解其平台功能、特征、相关政策规范和标准等，条件允许情况下注册账号，对平台功能和服务进行更详细的调研。



主要参考文献

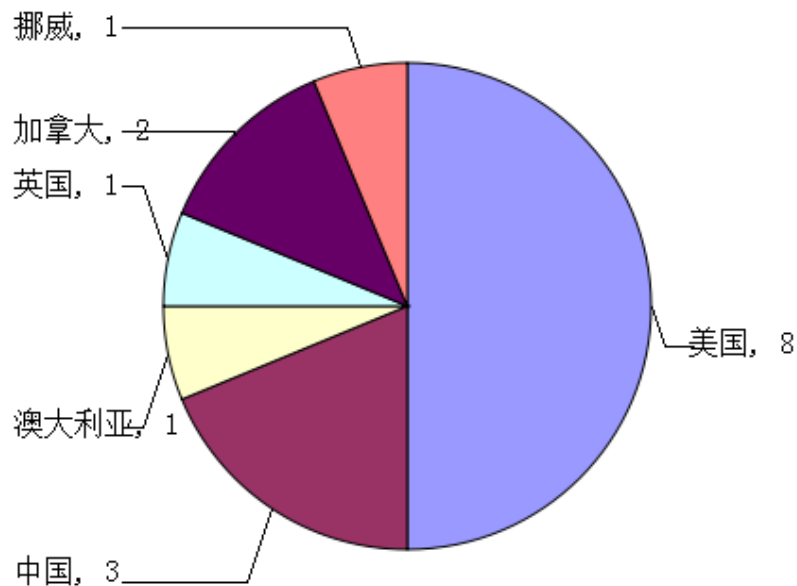
1. Claire Austin, Susan Brown, Nancy Fong, et al. Research Data Repositories: Review of current features, gap analysis, and recommendations for minimum requirements[J]. Minneapolis : IASSIST Quarterly (IQ) , 2015
2. 张计龙, 殷沉琴, 张用, 等. 社会科学数据的共享与服务 ——以复旦大学社会科学数据共享平台为例[J]. 大学图书馆学报, 2015, (1) : 74-79
3. 覃丹. 英美社会科学数据管理与共享服务平台调查分析[J]. 图书情报工作, 2014, 58(16): 67-75.
4. 赖剑菲, 洪正国. 对高校科学数据管理平台建设的建议[J]. 图书情报工作, 2013, 57(6): 23-27
5. 殷沉琴, 张计龙, 张莹, 等. 社会科学数据管理服务平台系统选型研究-以复旦大学社会科学数据平台为例[J]. 图书情报工作, 2013, 57 (19) : 92-96
6. 孟祥, 保钱鹏. 高校社会科学数据管理的国际经验及其借鉴——以UKDA和ICPSR为例[J]. 情报资料工作, 2013, (2):77-80
7. 王玉林, 曾咏梅, 崔然, 等. 我国高校图书馆面向社会开放现状调查[J]. 图书与情报, 2012 (6): 26-32
8. Crosas M. The dataverse network? : an open-source application for sharing, discovering and preserving data[J]. D-lib Magazine, 2011, 17(1): 2
9. King G . An introduction to the Dataverse Network as an infrastructure for data sharing[J]. Sociological Methods & Research , 2007 , 36(2):173 - 199



调研国内外数据平台

ICPSR https://www.icpsr.umich.edu/icpsrweb/landing.jsp	SDA http://sda.berkeley.edu/archive.htm
DSpace* http://www.dspace.com/en/inc/home.cfm	数据堂 http://www.datatang.com/
Fedora* https://getfedora.org/	ANDS http://www.ands.org.au/index.html
Data Conservancy https://archive.data.jhu.edu/dvn/	Nesstar* http://www.nesstar.com
DataCite https://www.datacite.org	UK Data Archive http://www.data-archive.ac.uk/home
Dataverse* http://dataverse.org	IRIS http://www.iris.edu/hq/
复旦大学社会科学数据平台 http://fisr.fudan.edu.cn/	Ocean Networks Canada http://www.oceannetworks.ca/information
中国人民大学中国社会调查开放数据库 http://nsrc.ruc.edu.cn/	Canadian Astronomy Data Centre http://www.canfar.phys.uvic.ca/canfar/

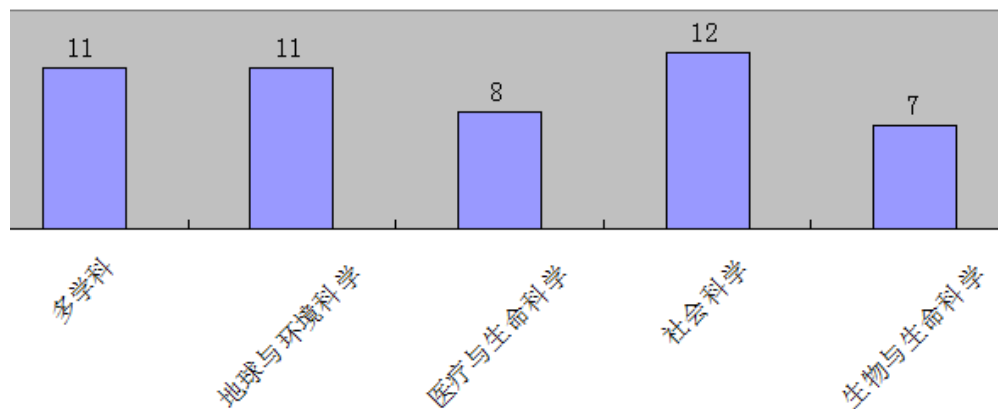
数据平台国家分布



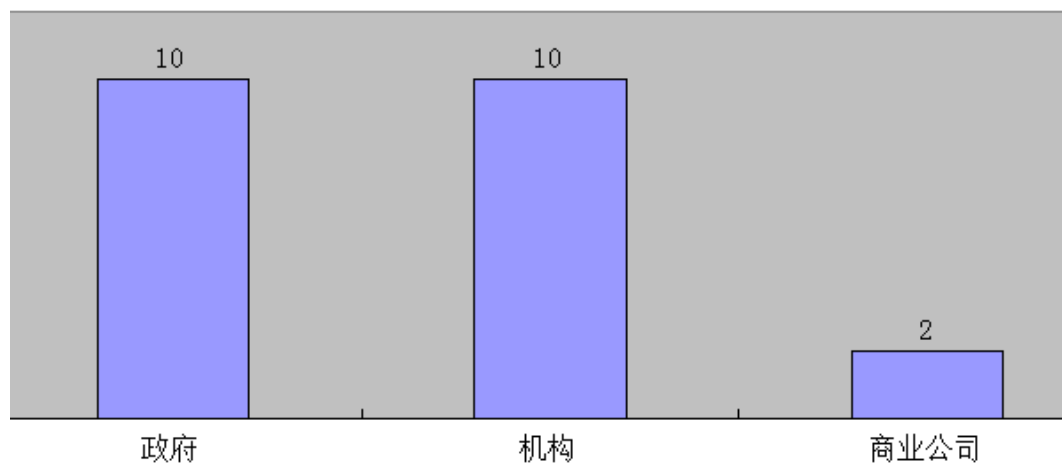
数据平台类型分布



学科领域分布



隶属关系



数据平台比较特征分类



研究背景

1

研究方法

2

结果分析

3

建议与思考

4

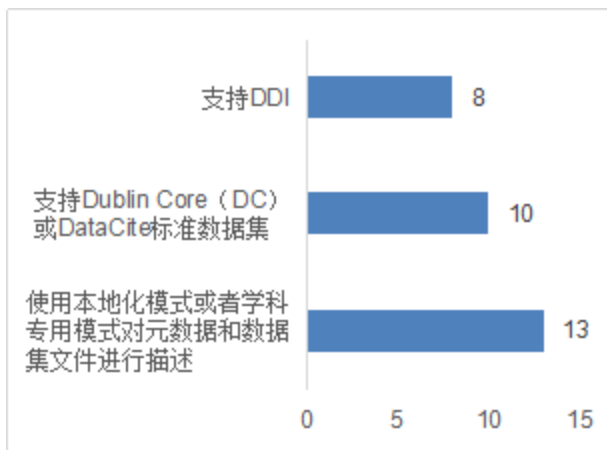
Contents

目录



01

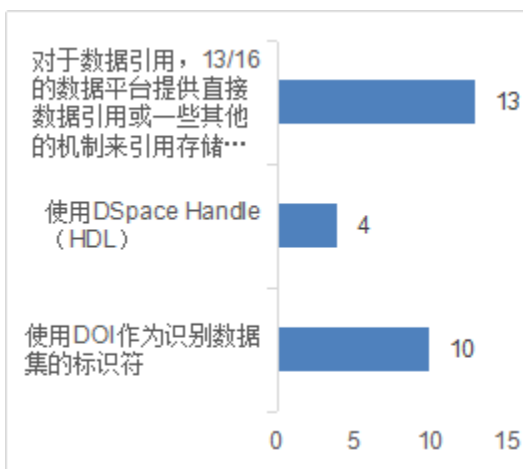
元数据



虽然很多平台使用标准化的元数据，一些主要平台使用不规范，系统内部不能跨模式搜索，平台系统之间资源不能互操作，不能实现互联互通，只有很少的数据集和元数据能够实现。

02

永久标识符



03

版本控制

通过比较发现，版本控制是一个重要的问题，且依然没有解决。绝大多数数据平台允许数据上传后进行修改编辑，而只有一半（8/16）的数据平台提供版本声明或版本控制系统，数据平台使用数据时间戳来识别改变的文件，这种做法并不能解决版本控制问题。（**没有提出系统的、持续的方式来解决数据集的版本问题**）



04

开放获取

05

数据使用

06

数据Curation

无需注册，匿名需“同意”使用条款进行开放获取资源

11

提供数据分享、术语表的使用及所有权具体政策

11

提供数据集创作共用许可或其他开放许可

12

关于跟踪调查的平台的数据使用量，75%（12/16）表示，通过下载量或其他统计数据来显示数据的使用情况。

有近一半（7/16）的平台提供数据策展活动方面的信息，但策展活动的策展服务细节和范围很不明确。



07

互操作性

通过API进行在线访问数据和元数据

9

开放档案计划 (OAI) 协议

7

0 2 4 6 8 10

08

长期存储

提供数据的长期存储和保存

10

标准文件传输和复制系统的使用, 冗余归档节点的分布式网络

4

存在多个冗余副本

10

09

认证

认证状况

1

WDS

2

DSA

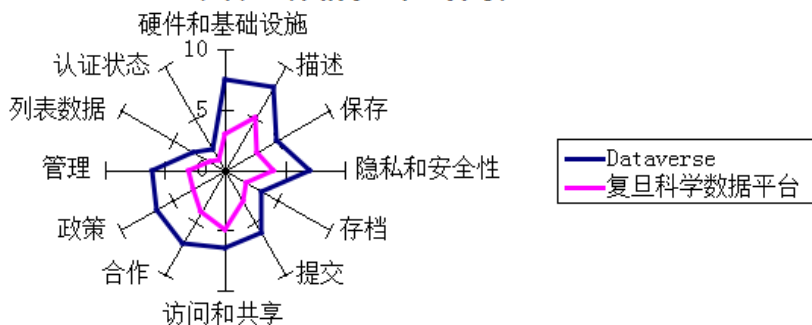


数据平台差异性分析

同类型
对比分析

不同类型
对比分析

平台整体情况对比分析



将开源软件、商业软件、在线分析软件三种类型的平台整体情况进行比较分析，可以发现不同类型的平台侧重点不同，开源软件更注重数据管理、数据服务、数据共享和数据交换，实现系统之间的互联互通；商业软件更注重数据的利用及其价值，对数据进行增值，把数据的价值最大化，推动相关技术、产业和应用的创新；在线分析软件更注重数据分析功能、创造变量功能、数据下载、数据搜索。同时也可以发现这三类平台仍有很大的发展空间。

国与美国还有很大差距。



研究背景

1

研究方法

2

结果分析

3

建议与思考

4

Contents

目录



功能需求和目的明确

首先确定目标，目标不同，设计的科学领域就不一样，结构模式就不同，构建什么类型的平台侧重点就不同。（定位）

1

数据管理政策

制定完善地平台数据管理政策体系，包括数据管理政策，数据共享政策，用户隐私保护政策等，有效地实现了平台的服务功能。

2

数据管理规范

统一数据标准有利于跨系统的数据交换与共享，实现数据共享和互联互通，同时可以与国际数据平台无缝连接，实现国际平台合作交流与共享。

3

加强合作交流

学习欧美数据平台经验，加强沟通合作，同时结合中国的实际情况，构建中国特色的数据管理与共享平台。

4

构建数据管理与共享平台的几点建议



思考



Thanks!

