

# 跨域经纬·创新翻转

—圖書館與數位人文的對話

汉学数字人文通用平台的需求设计与实现

THE CONSTRUCTION OF UNIVERSAL DIGITAL HUMANITIES  
PLATFORM FOR SINOLOGY



刘 炜 [wliu@libnet.sh.cn](mailto:wliu@libnet.sh.cn)

上海图书馆

# 传统人文的“数字化范式转型”

学科	目的	对象	方法	技术	工具
诗/语言学	诗以正言，义之用也	研究与语言相关的现象和规律	文本（包括语音）及通过文本的交互	基于文本各种表现和行为的分析	各种文本分析（包括语音/语法/语义分析）
书/政治文献学	书以广听，知之术也	治术/知识的源流及发展演变规律	书目控制（搜集整理）	数据库	君王/作者作品的内容分析
礼/社会学	礼以明体，明者著见，故无训也；	社会与人类群体行为的理解	各类实证方法，调查研究	逻辑归纳，统计分析	调查管理，统计分析
易/哲学	易为五常之原，易不可见，则乾坤或几乎息矣，与天地为终始也	对世界的根本认识	概念化，逻辑	本体/语义技术	概念语义分析
乐/艺术	乐以和神，仁之表也	审美载体的创造和研究	多媒体	多媒体技术	各种
春秋/历史	春秋以断事，信之符也	记载和解释人类活动进程	基于各种史料的解释和重建	内容分析	分析、模拟、统计等
新兴人文和数字人文：人类学、考古学、文化分析、文化组学（Culturomics）等					

# 方法论 共同体

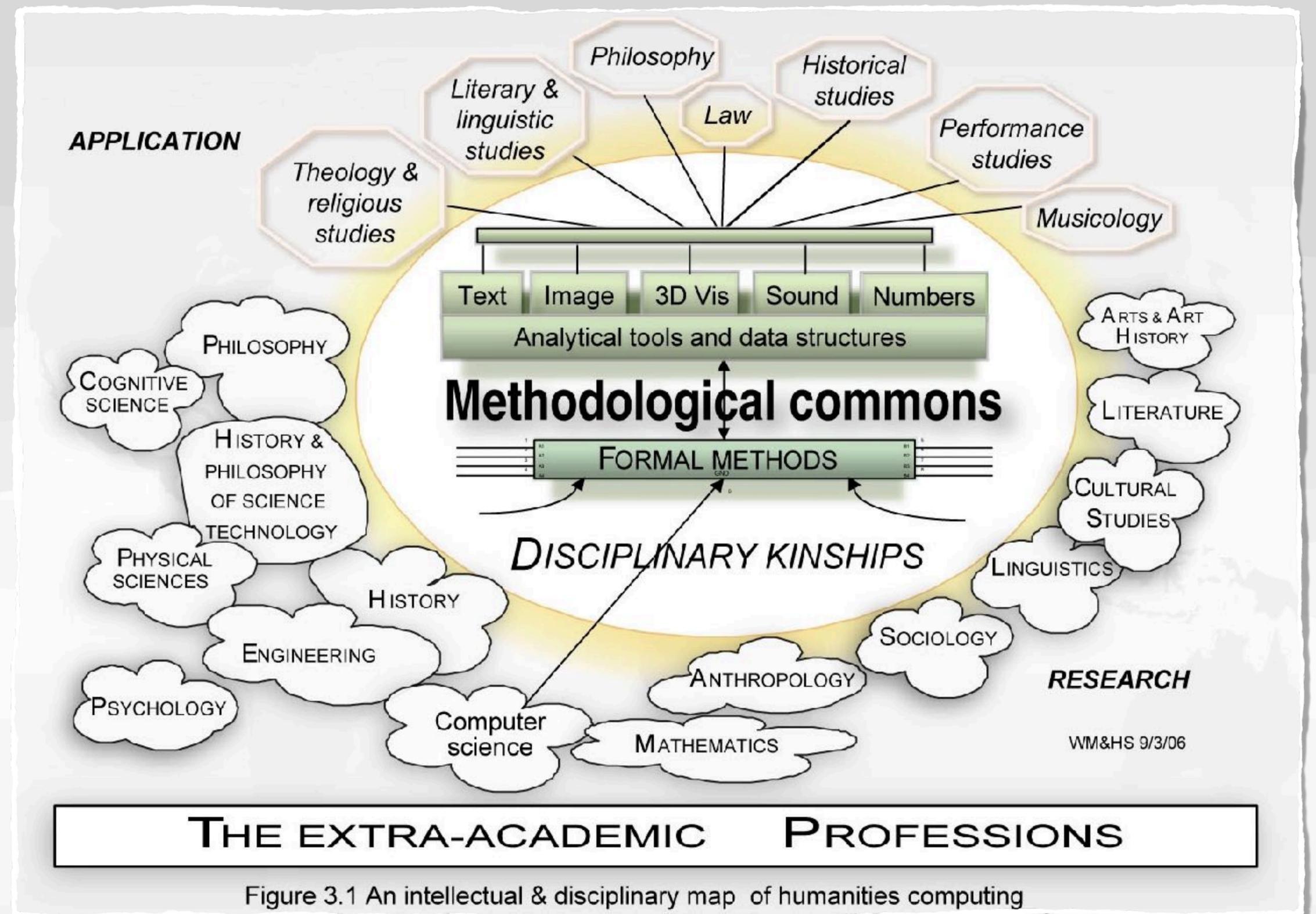


Figure 3.1 An intellectual & disciplinary map of humanities computing

# 人民大学对于数字人文的定义

- 数字人文是计算机学科和人文学科交叉研究所衍生的一个新领域，它以涉及 针对计算工具与所有文化产品的交叉领域为研究对象，系统地研究数字与人文相结合的普遍规律和应用方法。
  - 以人文科学的基本问题为研究对象
  - 以不断发展进步的信息技术和数字技术等为主要工具
  - 以数字资源构建、信息资源管理等数据基础设施建设为基础
  - 以计算分析和案例分析方法等为主要研究手段
  - 通过建立描述学术活动理论、方法和功能的框架以及各种类型的项目实践
- 探讨数字技术与人文科学跨学科对话中的方法、过程、特征和相互关系，以及数字人文作为一个整体与社会环境之间的互动与关联，并从中探索、归纳和总结出获得成效、提高效率的一般理论、方法和规律，以推动知识创新和服务。

# 海外中文数字人文资源情况

- 东京大学东洋文化研究所汉籍善本全文影像资料库, <http://t.cn/aoxNuq>
- 東京大學東洋文化研究所雙紅堂文庫全文影像資料庫, <http://t.cn/zTyT450>
- 日本国会图书馆古典籍资料, <http://t.cn/zOm2To6>
- 日本學習院大學東洋文化研究所所藏古典籍, <http://t.cn/8k3IHye>
- 京都大學人文科學研究所在線版四庫提要, <http://t.cn/aNmUFV>
- 京都大學人文科學研究所東方學デジタル圖書館, <http://t.cn/aoHBzA>
- 早稻田大学古典籍総合, <http://t.cn/hGhxOv>
- 京都大學電子圖書館貴重資料畫像數據庫, <http://t.cn/zjyfHfq>
- 中國哲學書電子化計劃, <http://t.cn/zRO0E0e>
- 中央研究院汉籍电子文献瀚典全文检索系统, <http://t.cn/aolwf9>
- 台灣網上書上網—數位典藏與學習電子書庫, <http://t.cn/a9lJX9>
- 故宮博物院和東吳大學數位古今圖書集成, <http://t.cn/a3PqU3>
- 首都图书馆古籍插图库, <http://t.cn/SlqnOw>
- 《工部局董事会会议录》28册, <http://t.cn/8kUQCOh>
- 哥伦比亚大学东亚图书馆Chinese Paper Gods, <http://t.cn/aKkyal>
- 康奈尔大学Wason Pamphlet Collection即华生中国收藏, <http://t.cn/8kI36fl>
- 哥伦比亚大学东亚图书馆古巴华工调查录, 6卷, <http://t.cn/8kjiZVX>
- 台灣大學圖書館公開取用電子書, 110074筆, <http://t.cn/hbJgKn>
- 輔仁大學歷史學系等南京教區契約文書數位典藏計劃, <http://t.cn/8kyFO2y>
- 賴永祥長老史料庫, <http://t.cn/8FKtJAH>
- 劍橋大學圖書館館藏中國叢書綜合目錄檢索, <http://t.cn/8kZMTQk>
- 剑桥大学图书馆电子目录, 北平图书馆善本书胶片, <http://t.cn/8F87bo3>;
- 剑桥大学图书馆电子目录, 景印离藻堂四库全书荟要, <http://t.cn/8F87bou>
- 金陵圖書館《民國叢書》目錄索引, 第一輯至第五輯, <http://t.cn/zOp8Nex>
- 中國大陸各省地方志書目查詢系統, <http://t.cn/hcZpmI>
- 東洋学文献類目檢索[第7.3a版], <http://t.cn/zTX4c28>
- 漢字データベースプロジェクト#, 能查《说文》和《康熙字典》等, <http://t.cn/SiZBD7>
- 日本全国汉籍检索系统, <http://t.cn/aCMEfl>
- 台湾大学台湾历史数位图书馆之研究工具集, <http://t.cn/z85Tcsv>
- 清代人口史研究資料庫, <http://t.cn/zWv04zn>
- 清代檔案人名權威資料查詢 <http://t.cn/zR6uYkd>
- 中研院國際電腦漢字及異體字知識庫, <http://t.cn/hgsrrH>
- 中研院語言學研究所等搜文解字, <http://t.cn/8kx2foj>
- 中央研究院語言學研究所漢字知識本體, <http://t.cn/8FA0ncX>
- 缺字系統, <http://t.cn/zThmsVi>
- 中國文化大學中華百科全書1983年版, <http://t.cn/akc9jE>
- 中國文化大學中文大辭典#全10冊, <http://t.cn/zO3QT0L>
- 网络资料版华人基督教史人物辞典, <http://t.cn/aO5ZEd>
- 金陵图书馆建有十余个具有地方特色的数据库, <http://t.cn/8ktw4VO>
- 中国际中国学杂志《通报》文章目录汉译, <http://t.cn/aYRwYx>
- 近現代漢語學術用語研究, 即“ModernChineseScientificTerminologies”, <http://t.cn/zRYoShP>
- 海德堡大学汉学文典, <http://t.cn/zRjMO6v>
- 海德堡大学早期《申报》索引, <http://t.cn/zRlkFch>
- 中文和合本聖經查詢系統, <http://t.cn/h0XGb>
- 中華福音神學院圖書館中文基督教期刊論文索引, <http://t.cn/8FyIYix>
- 澳門期刊論文索引, <http://t.cn/8FLGer7>
- 台湾NCL中国文化研究论文目录, <http://t.cn/8kJECgC>
- 台湾民间传说主题资料库, <http://t.cn/8F24jwG>
- 台湾人文及社会科学引文索引资料库, <http://t.cn/8F240WC>
- 台灣文史哲論文集篇目索引系統, 近6萬篇篇目, <http://t.cn/8kixsvj>
- 台北学研究主题资料库, <http://t.cn/8F240Wo>
- 臺灣大學法律學院臺灣日治時期統計資料庫, <http://t.cn/zROOF7V>
- 中華佛學研究所中國佛教寺廟志數位典藏, <http://t.cn/zOm0lCZ>
- 中国佛教数字目录, <http://t.cn/>

# 中文数字人文资源/项目/工具/平台情况\*

## A.tools & platforms 平台與工具：18组

1. 上海圖書館古籍聯合目錄及循證平臺 (劉焯, 夏翠娟)
2. Central Library (魏家惠, 莊惠茹, 王宏德)
3. Chinese Text Project (Donald Sturgeon)
4. 10000 rooms (Tina Lu)
5. 古籍數位人文研究平臺 (陳志銘)
6. 中華書局籍合網 (洪濤)
7. Docusky (胡其瑞, 杜協昌)
8. MARKUS (Hilde De Weerdt, Mees Gelein)
9. 法鼓文理學院數位人文項目 (洪振洲)
10. Christian Wittern
11. 中研院數位研究平台 (王祥安)
12. 國家圖書館出版社數字化項目 (王濤)
13. China National Knowledge Infrastructure CNKI (謝)
14. China Biographical Database project (王宏甦)
15. Asia Network (何浩洋)
16. Kitamoto Asanobu
17. 搜韻 (陳逸雲)
18. 臺灣大學圖書館數位人文項目 (陳光華)

## B.text & image repositories 文字圖像數據庫：18组

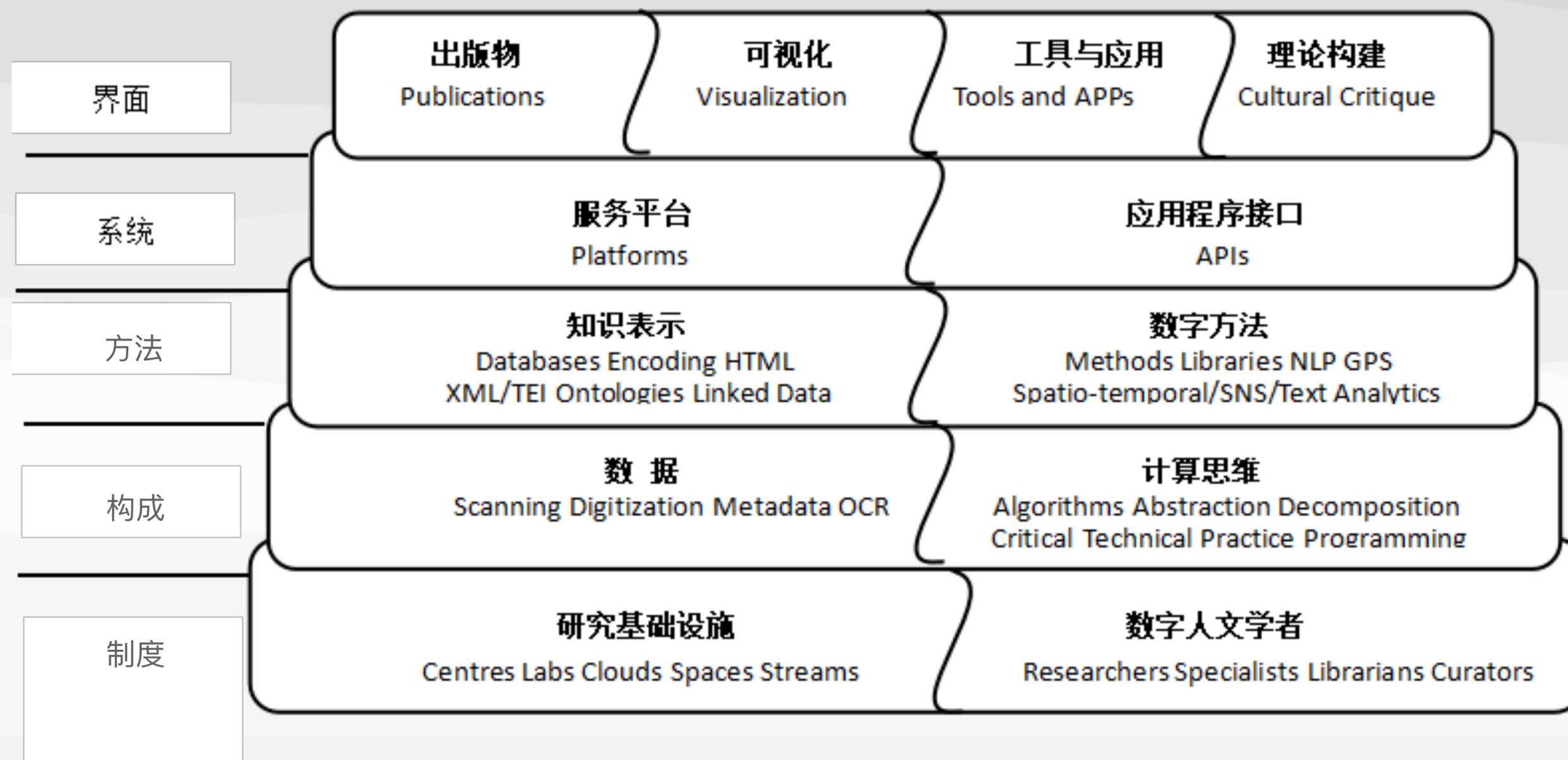
1. CADAL (黃晨, 徐建剛)
2. 中文在線 (朱厚權, 杜佳)
3. 中研院漢籍電子文獻 (劉錚雲)
4. 香港中文大學圖書館 (Louise Jones)
5. Overview of the Activities of Digital Humanities Initiative at University of Tokyo (Masahiro Shimoda)
6. 陳澄波畫作與文書 (謝國興, 王麗蕉)
7. Berlin State Library (Matthias Kaun)
8. EastView (Robert Lee, Sasha Chen)
9. Dataverse Peking University Library (聶華, 朱本軍)
10. 臺灣學術經典 (伍翠蓮)
11. 徽州文書電子化項目 (王蕾)
12. 北京時代瀚堂科技有限公司 (王宏源, 王心怡)
13. 上海圖書館近代報紙/期刊數據庫、海派圖片庫 (楊敏)
14. 中國地方歷史文獻數據庫 (趙思淵, 蔣勤)
15. 中華經典古籍庫籍合網 (洪濤, 程瑞雪)
16. 書同文 (張軸材, 張弛宜, 張虎)
17. 華藝數位股份有限公司 (吳宗憲, 鄭學淵)
18. 國學大師 (鄺勇)

## C.Databases 數據庫：17组

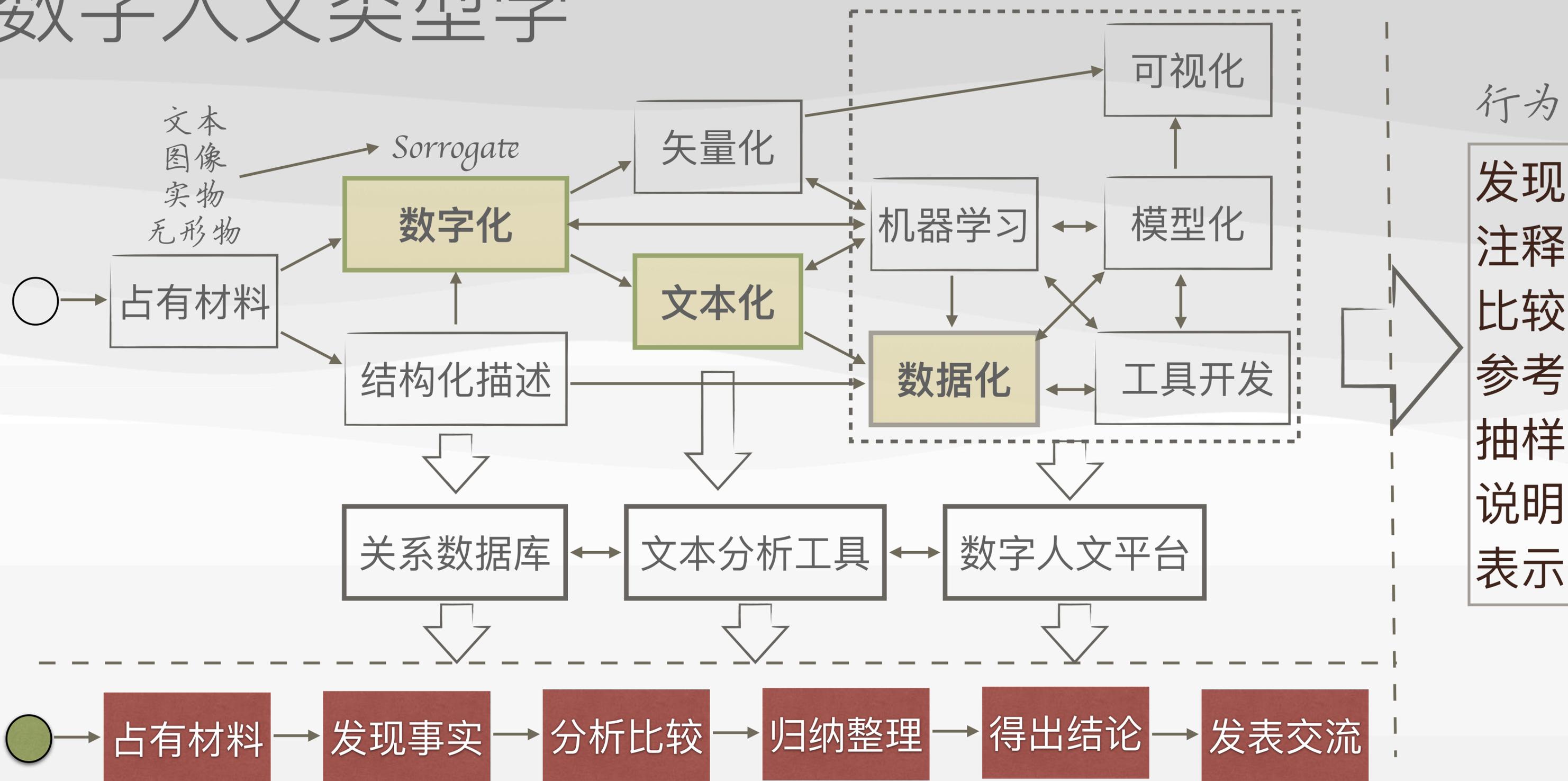
1. Utah Genealogical Society (Steve Nickle)
2. 浙江大學學術地圖發佈平臺 (張豐)
3. China Qing Government Employee Database (James Lee, C)
4. 中南民族大學 數字人文資源研究中心 (王兆鵬, 邵大為, 李)
5. Leiden University Centre for Digital Humanities (Paul)
6. 政治大學數據庫項目、從學者角度對 cyberinfra. 的設想 (
7. 中央研究院人社中心GIS專題中心 (廖泫銘)
8. 臺灣歷史人物傳記資料庫 (張素玟)
9. Dictionary Databases (Charles Muller, Wenlin, Pleco)
10. Ming Qing Women's Writings (方秀潔, 施松)
11. 南京大學數字人文項目介紹 (王濤, 陳靜, 史梅, 陳剛)
12. Local Gazetteers Research Tools (陳詩沛)
13. 復旦大學中國歷史地理信息系統 (張曉紅, 孫濤)
14. 中央研究院 (歷史語言研究所/數位文化中心)
15. 中國海關數據庫、數據庫版權的法學討論 (劉冬)
16. 上海圖書館家譜數據庫 (劉焯, 夏翠娟)
17. Kansai University Open Research Center for Asian Stud

\*来自哈佛上海中心2018年3月14-16日汉学基础建设数字人文会议

# 数字人文领域全景 (堆栈2.0)



# 数字人文类型学





# 历史人文大数据平台

The DH Big Data Center of Shanghai Library



高级检索

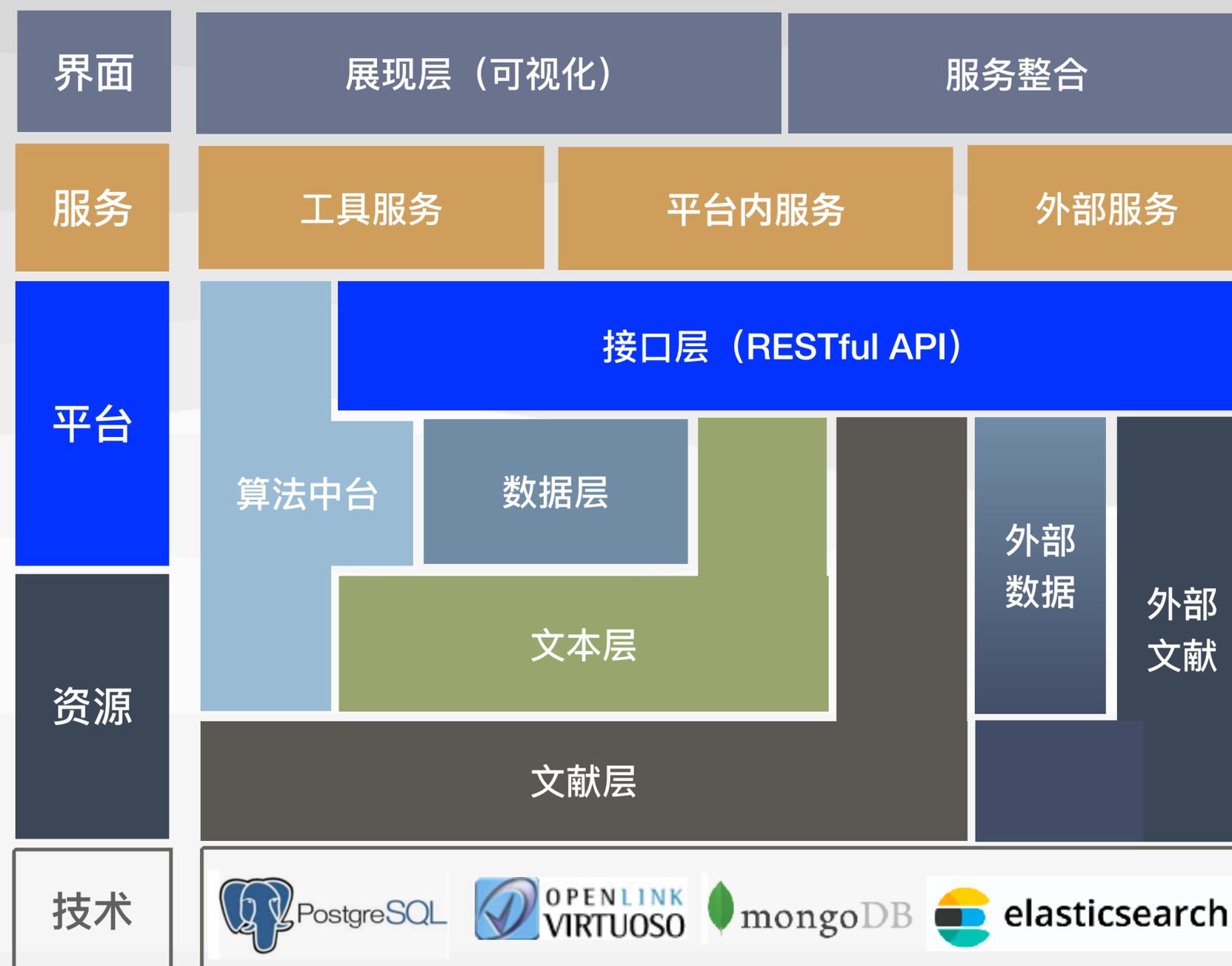
截止2020年10月1日，已收录报纸300种、篇目30000条；期刊100种、篇目5000条；整理知识单元1000,000个

---

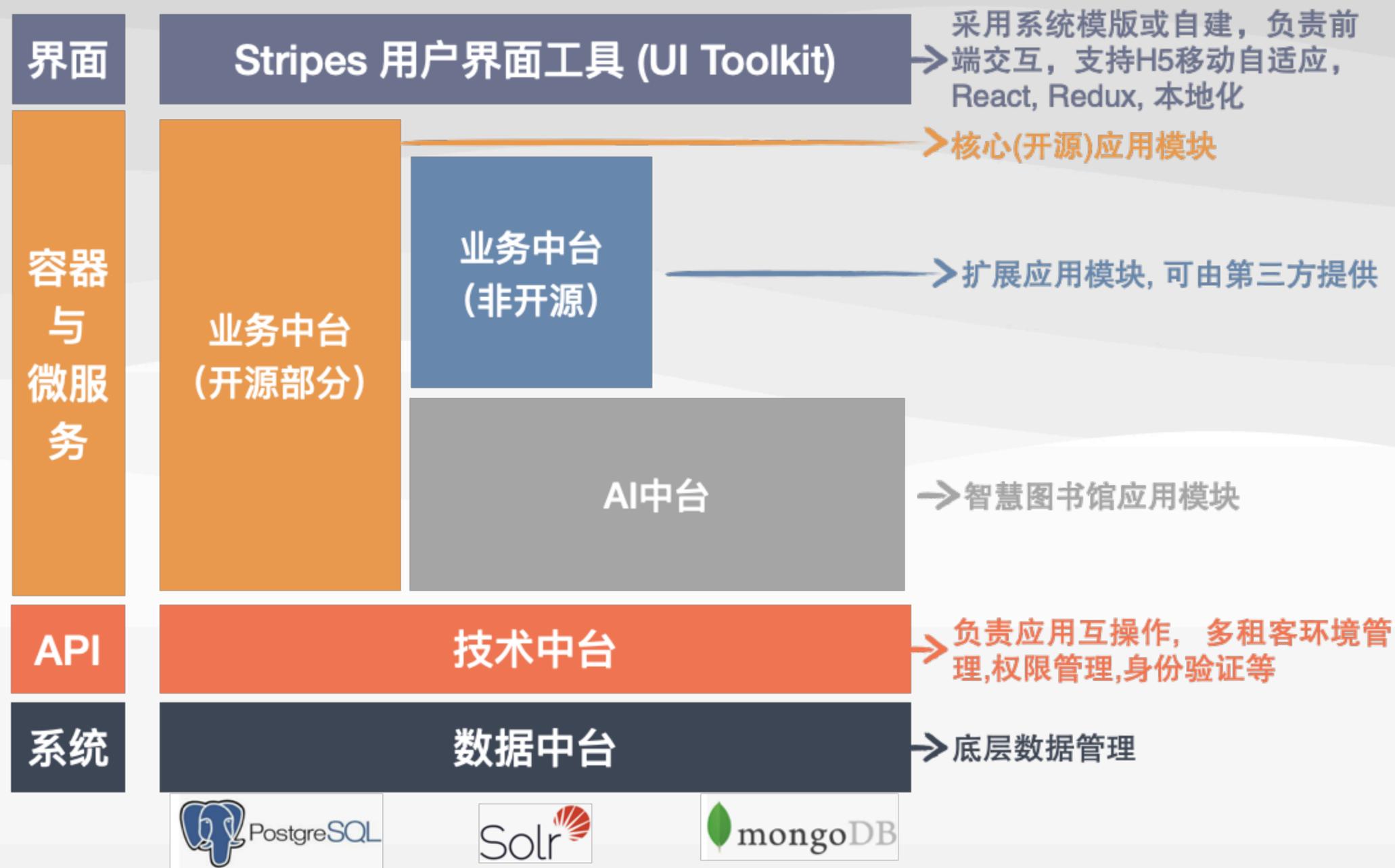
# 数位人文平台技术趋势

- 从烟囱到广场：IT体系架构的进化
  - 从功能到语义：知识服务的新架构
  - 从孪生到宇宙：诞生中的崭新世界
-

# 从烟囱到广场：IT体系架构的进化



# 云原生服务平台



# 系统架构的应用场景

## ① 知识单元的标识及其管理。

所有对人文研究具有独立意义的实体或信息单元，如文献，或人、地、时、事、物、事件、概念，以及各类属性和取值词表等，都需要有独立的标识（即ID），并统一ID编码标准，通常用http URI，其相互之间的关系如有必要可以通过建立本体知识库来管理。当然建立过程可以采用自动抽取加人工辅助校验方式。

## ② 支持多种协议的跨网域搜索发现或获取链接。例如OAI-PMH规范，各类Restful+JSON的API规范、联邦检索页面分析规范等等。

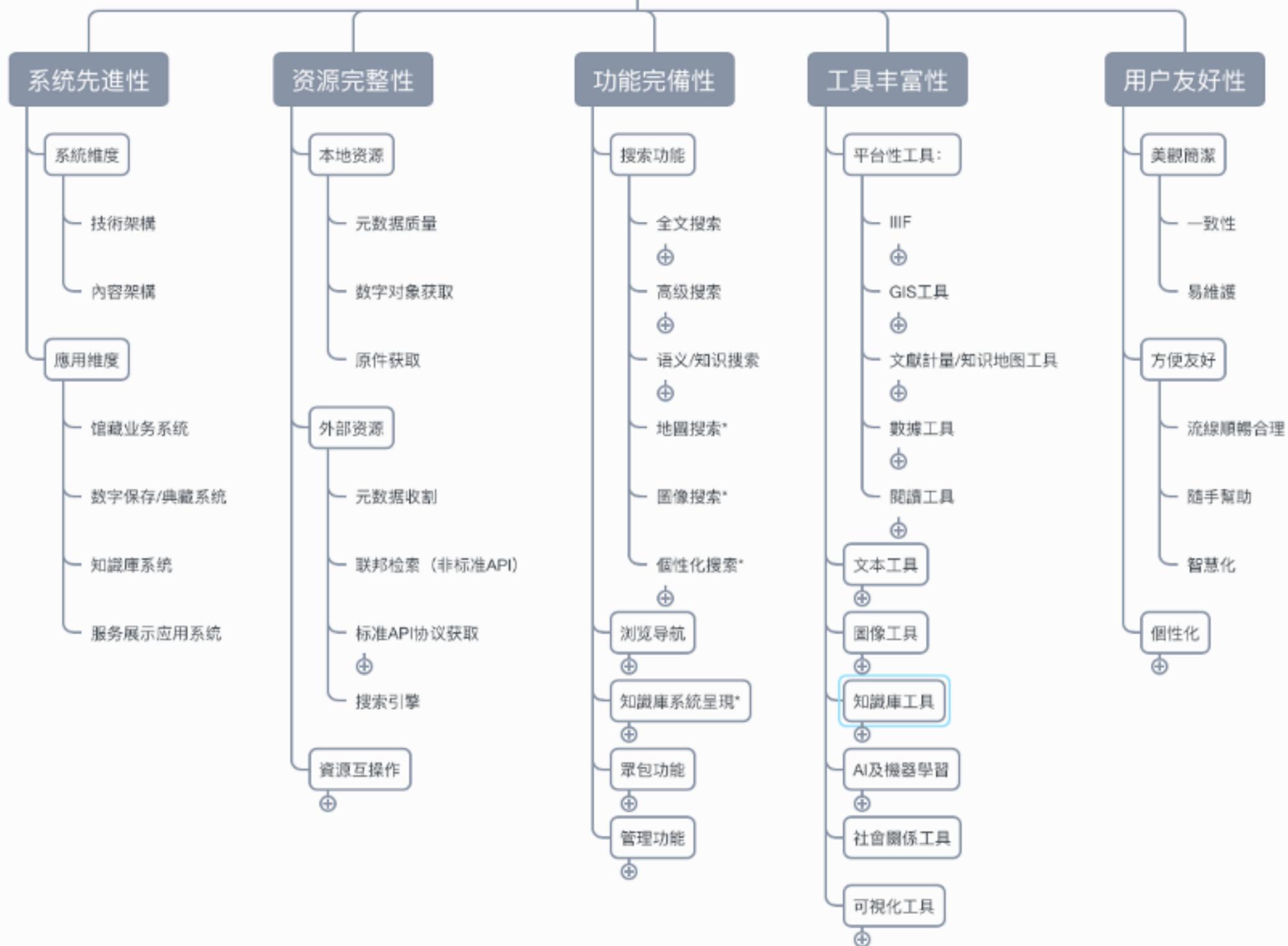
## ③ 微服务的容器及编排规范。

## ④ 多种数据类型的管理，包括底层关系数据库、图数据库（包括三元组语义数据）、对象数据、流媒体的管理，。

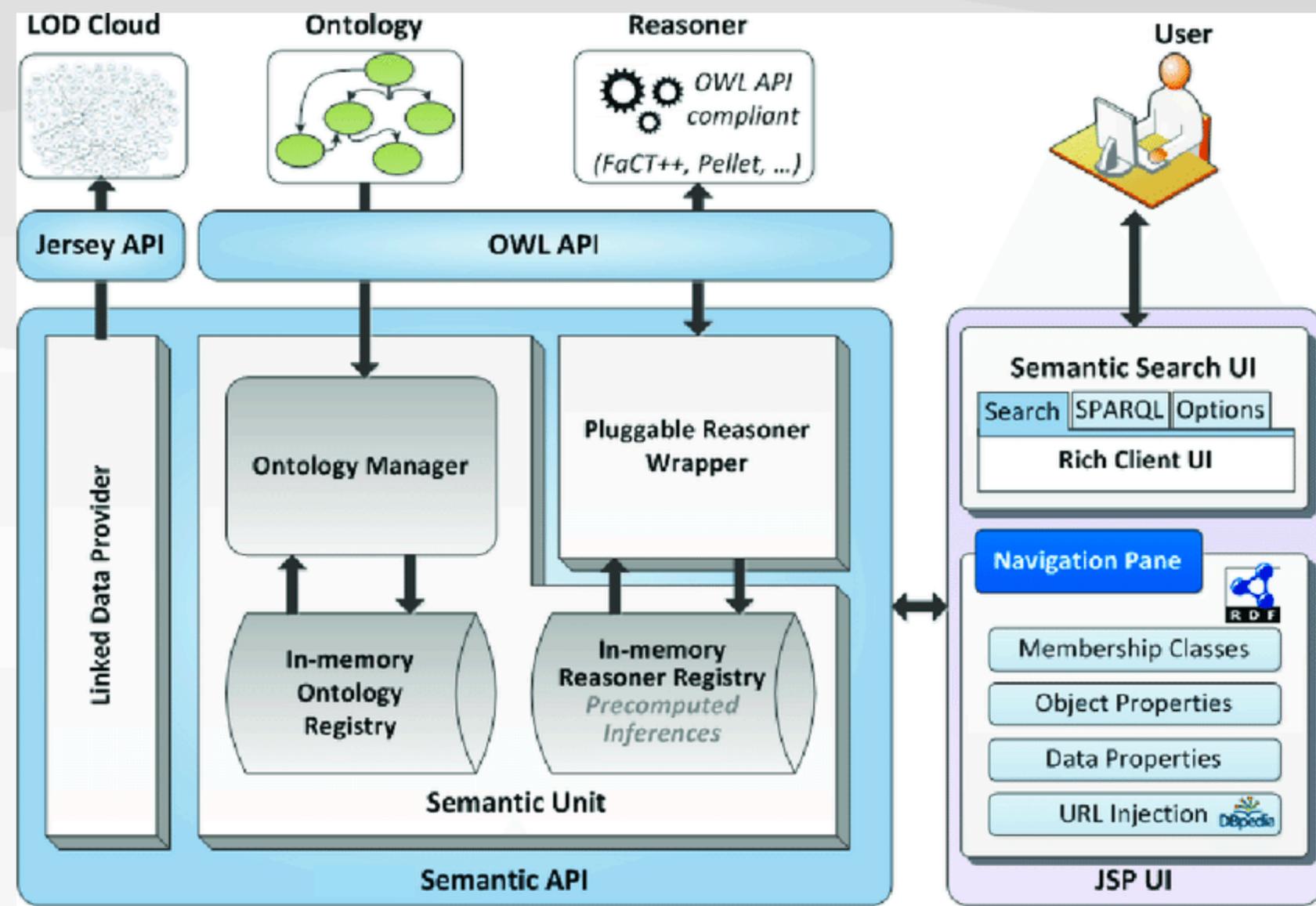
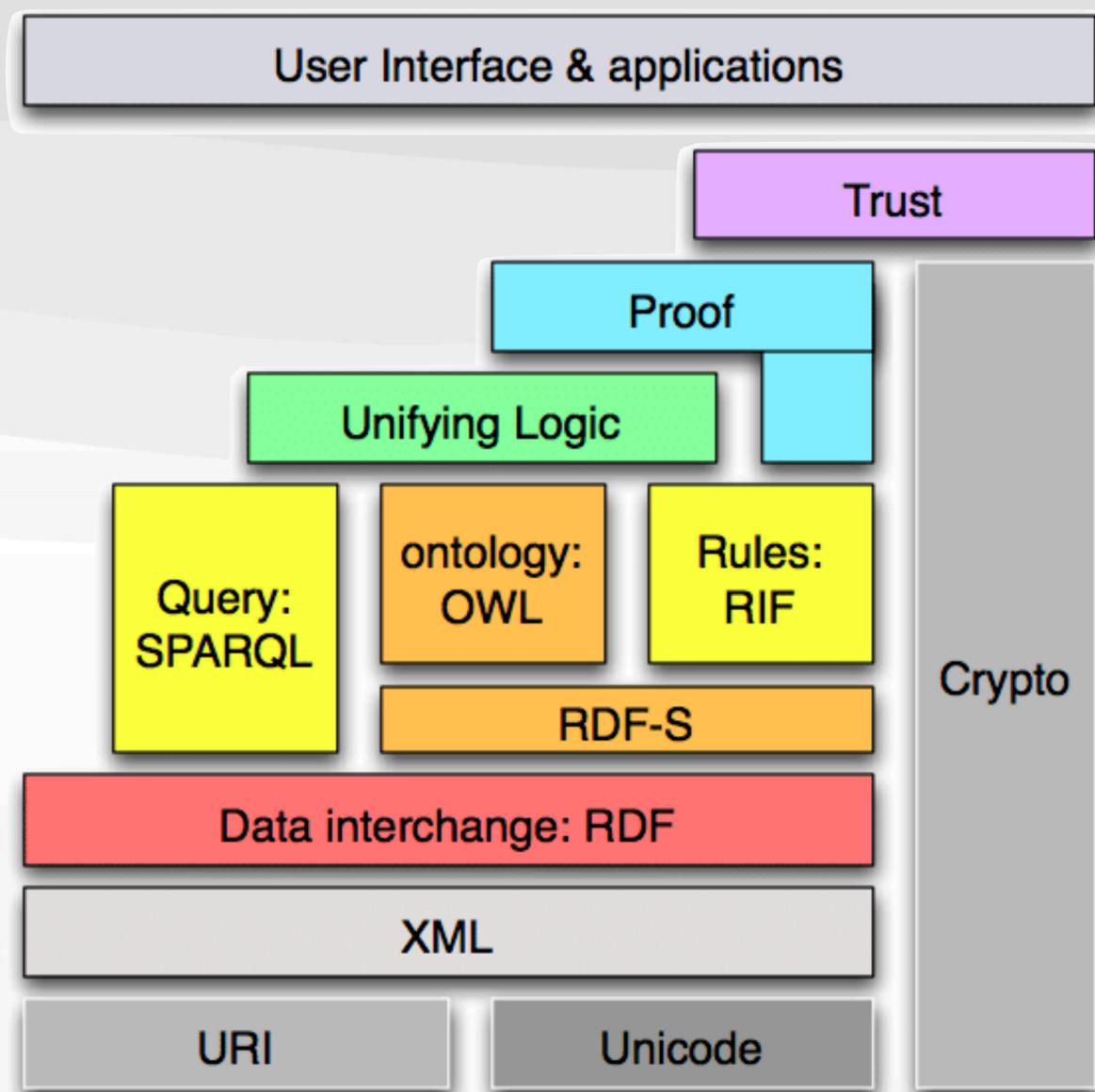
## ⑤ 复杂但统一的用户及授权管理，包括远程访问管理。

---

# 数字人文平台设计



# 从功能到语义：知识服务的新架构



# 内容架构的应用场景

- ① 一致性/相似性计算。
- ② workflow 定义对研究流程的支持。
- ③ 各类图像功能（如图像查询、对比、标注等）的支持。
- ④ 文本与图像关联（可提供加工平台，或研究对比）。
- ⑤. 提供证据链服务（纪录从底层文献到研究结果的整个过程中实体来源及变化，包括引用参考等）。
- ⑥ 海量数据可视化支持（遥读）。
- ⑦ 事实的可信度计算及排序（需建立可迭代的可信度模型）。
- ⑧ 众包数据加工平台的数据管理。
- ⑨ 数据系统迭代进化的支持（数字化、文本化、数据化（实体提取、建立关联等））。

# 数据服务的FAIR原则



## **FINDABLE**

Data has rich metadata and unique identifier



## **ACCESSIBLE**

Data can be easily downloaded or used by using standard protocols



## **INTEROPERABLE**

Metadata use an accessible and standard language



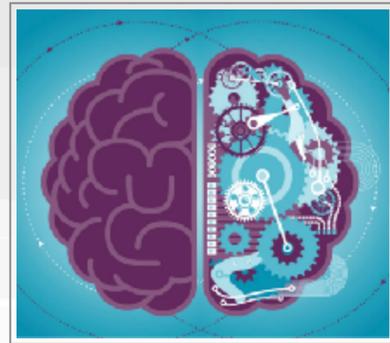
## **REUSABLE**

Data is well-described and provides clear usage of licences



# 从孪生到宇宙：诞生中的崭新世界

A



**Artificial intelligence**

B



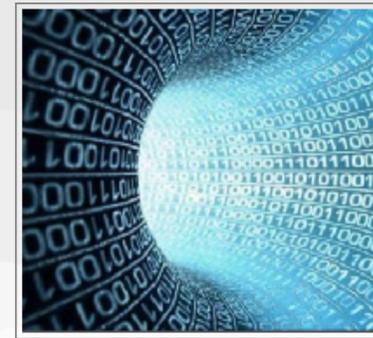
**Blockchain**

C



**Cloud computing**

D

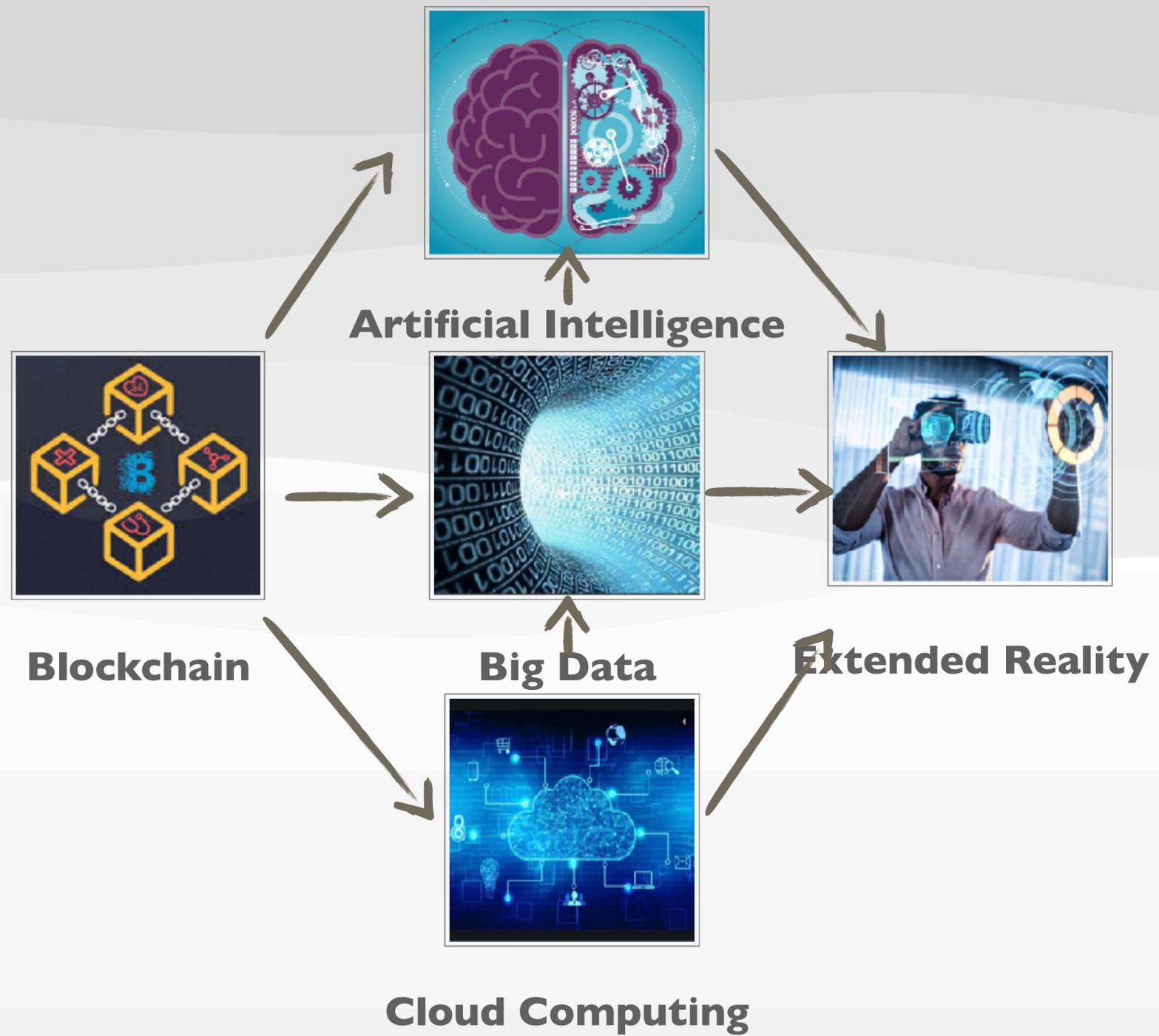


**big Data**

E



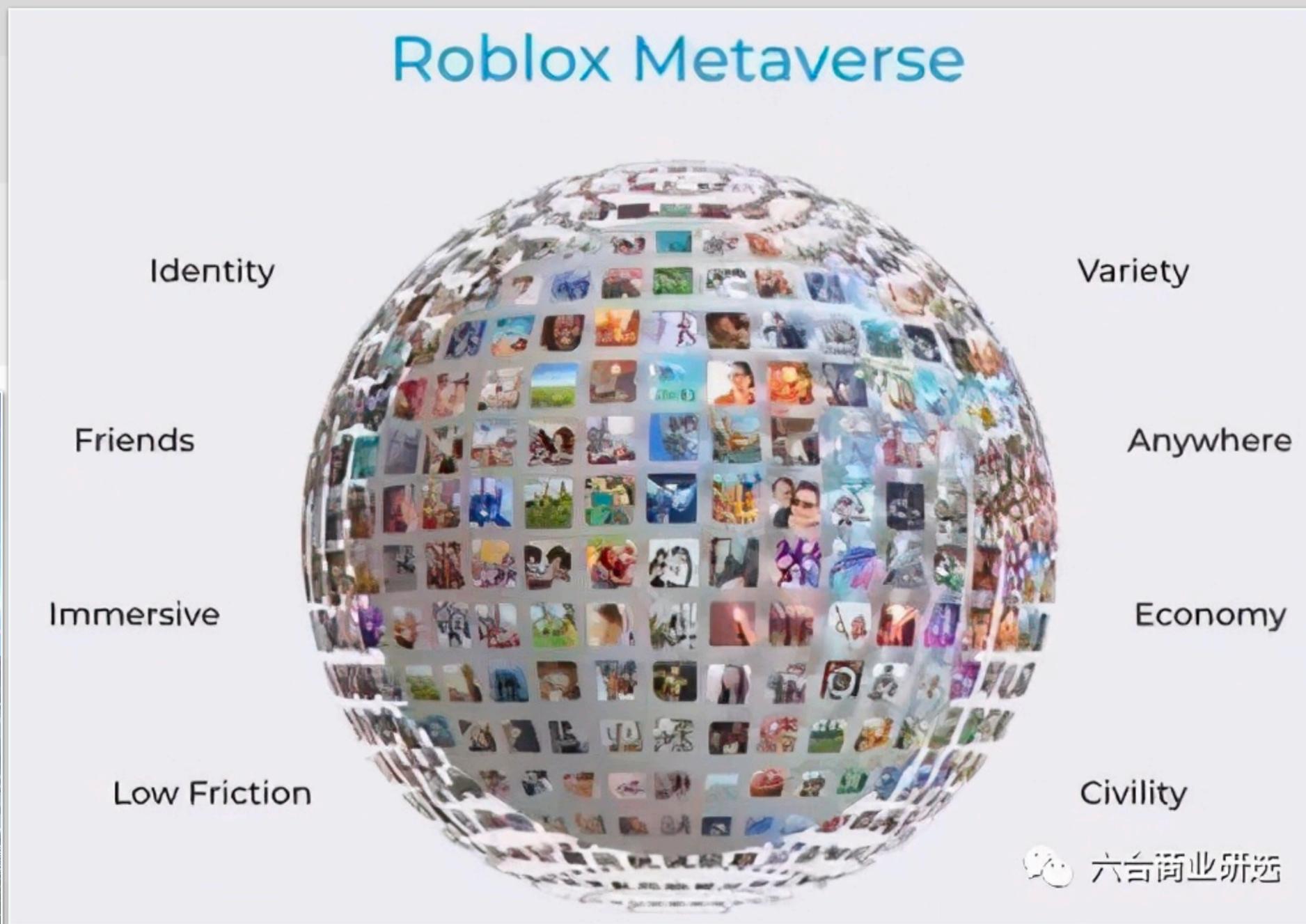
**Extended reality**



2016 Venice 3D



# 从数字孪生到元宇宙





---

# 前景无限

- 趋势：“ABCDE”的技术组合正在改变认知世界的方式
  - 转型：GLAM可以利用MetaVerse，帮助人文学者从文化载体记忆中重建历史场景
  - 颠覆：提供更强大的叙事和阐释能力，成为人文研究知识生态的使能者
-

# 总结

- 1. 着力解决瓶颈问题：基础设施建设
  - 2. 认清技术发展趋势：技术应用演进（三大方向）
  - 3. 改进平台架构设计：技术架构、语义架构和场景架构
-

---

# 数位人文平台建设愿景

让资料唾手可得

让建模随心所欲

让分析无师自通

让结果美轮美奂

**让人文研究变得简单！**

---

谢谢!



刘 炜 [wliu@libnet.sh.cn](mailto:wliu@libnet.sh.cn)

上海图书馆

