



关联数据、BIBFRAME及资源发现

贾君枝

中国人民大学

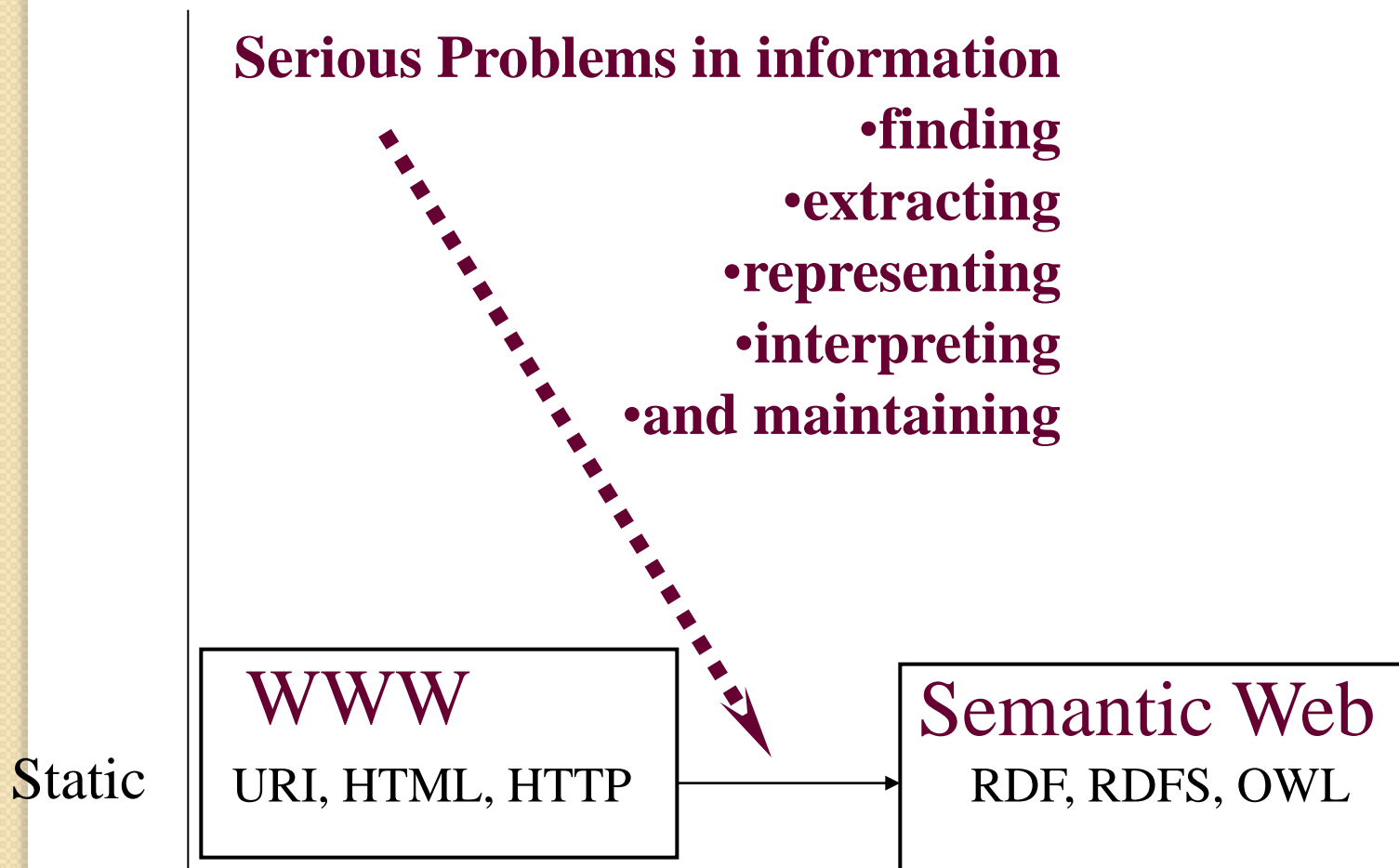
2019.10.17于福州

第六届全国文献编目工作研讨会

内容

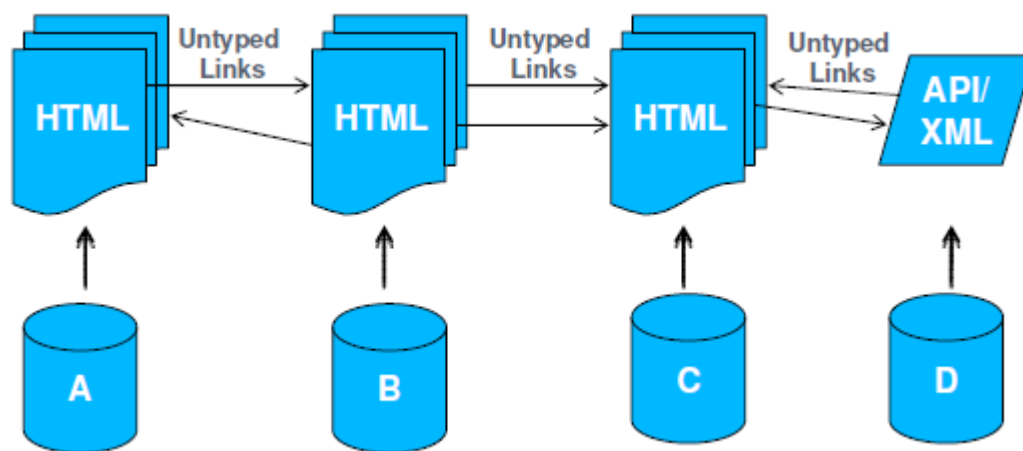
- 语义网环境下图书馆新要求
- 关联数据特点与功能
- BIBFRAME目标、模型及实现
- 基于关联数据的资源发现

万维网发展中的问题



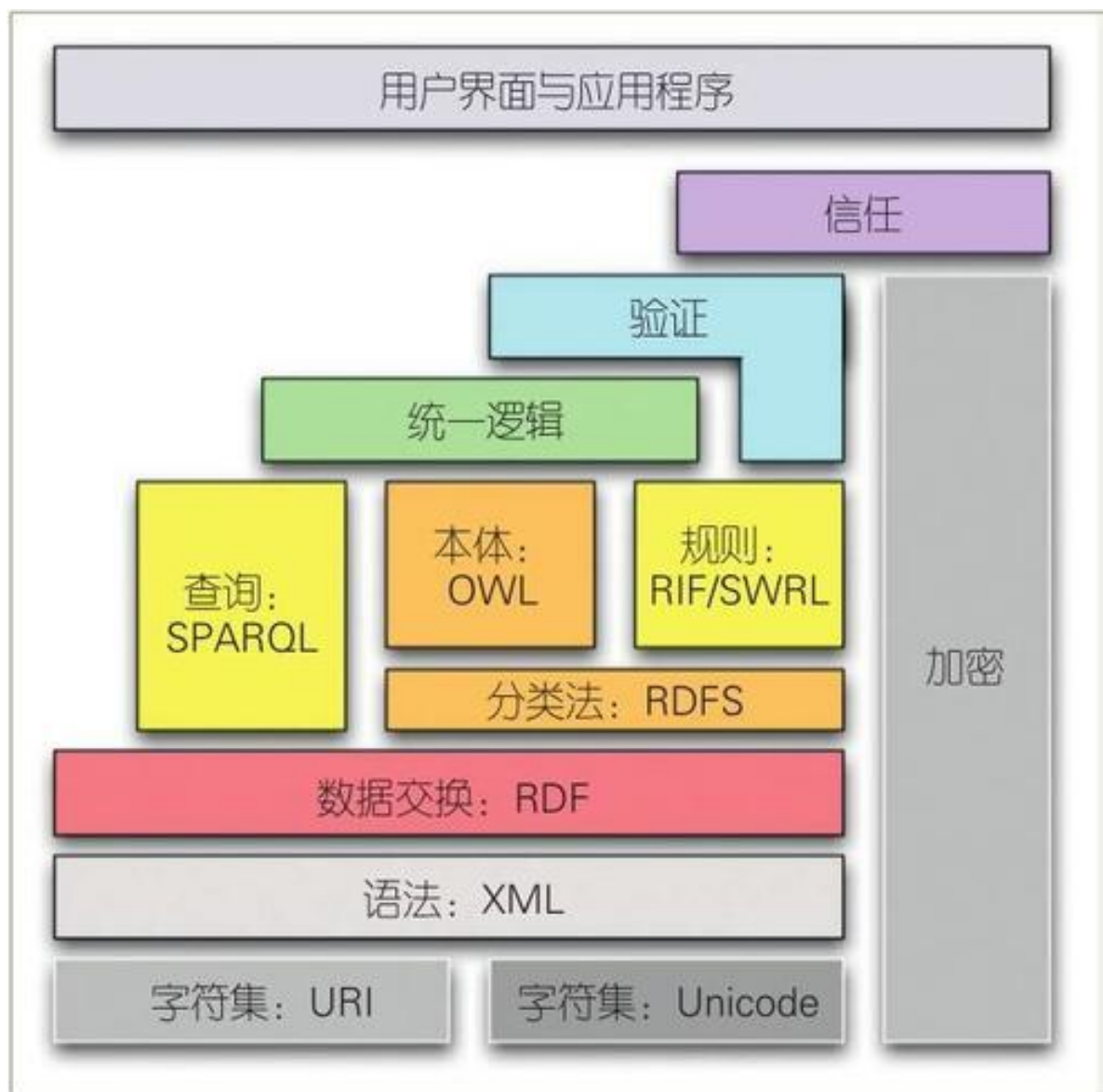
Web of document

- 以文档为中心
- 文档之间建立关系
- 人可读
- 隐含语义内容



语义网

语义网提供了允许数据跨应用、部门的共享及重用，其基于 RDF 基础上构建。



语义网络层

第一层Unicode和URI。该层是整个语义Web的基础，其中Unicode处理资源的编码，URI负责标识资源

第二层XML+NS+XML Schema，用于表示数据的内容和结构

第三层RDF+RDF Schema，用于描述Web上的资源及其类型

第四层Ontology vocabulary，用于描述各种资源之间联系

第五层到第七层是在上面4层的基础上进行的逻辑推理操作，包括公理和推理规则、认证机制及信任机制

语义网环境下图书馆发展要求

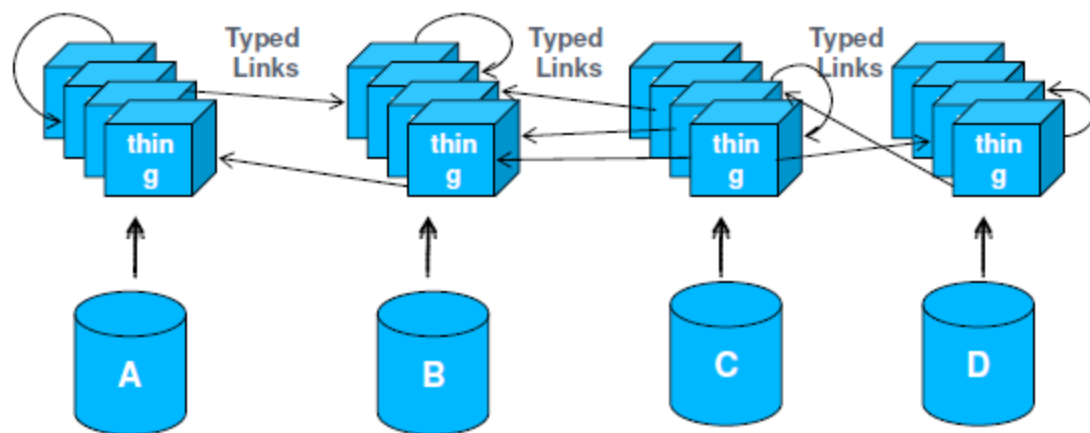
- 图书馆系统支持新的格式、 workflow 更新
- 实现更高层次的互操作
- 构建互联网合作新模式
- 提升用户服务体验
- 促进资源的可发现性

关联数据

- 关联数据是语义网发布数据的一种方式，用来在语义网中使用URI和RDF发布、分享、连接各类数据、信息和知识。
- 优点在于：
- 鼓励数据重用
- 降低冗余
- 扩大数据内外部连接
- 形成数据网络推动数据集成与发现

数据网络

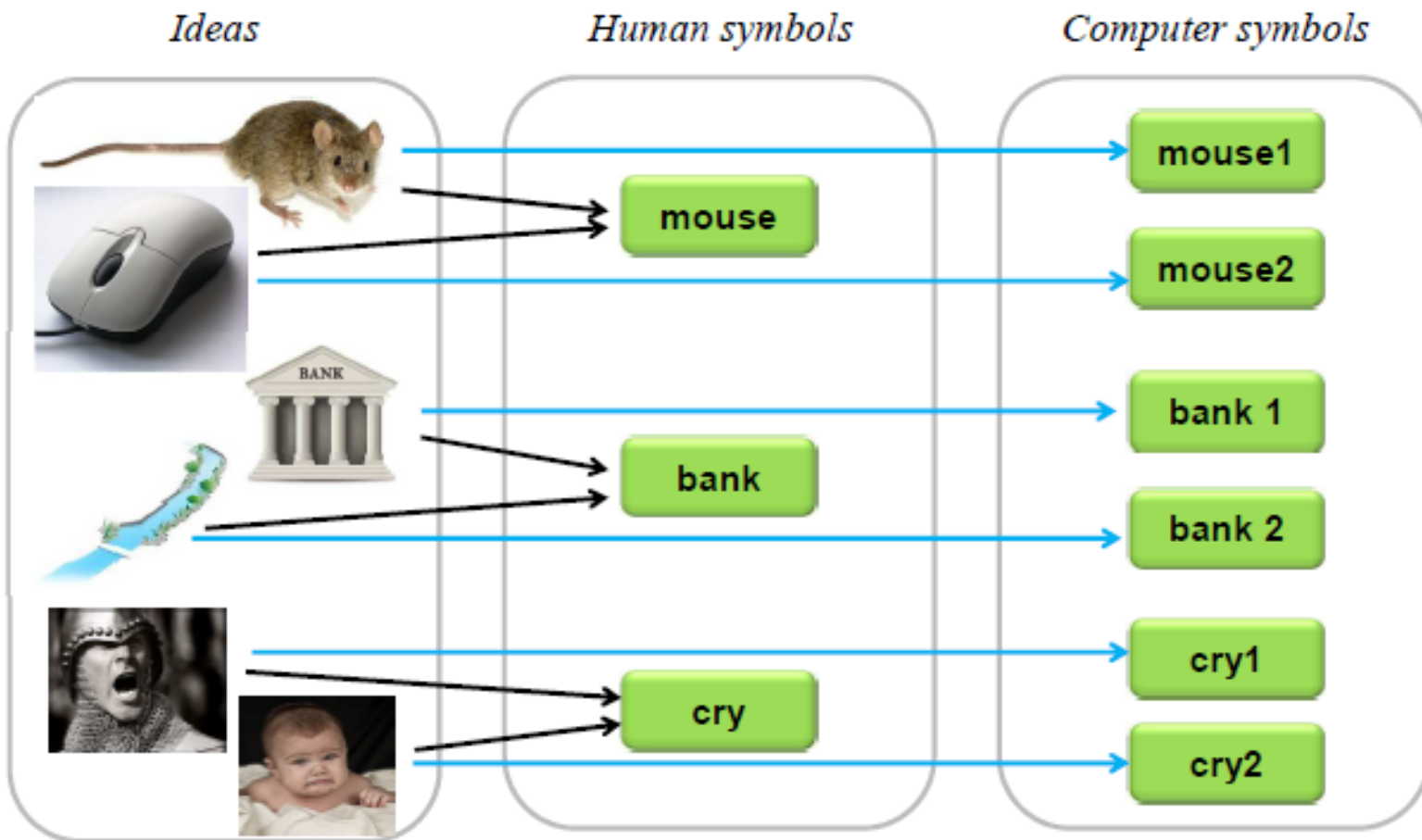
- 事物对象为核心
- 事物对象建立联系
- 语义明确
- 人机可读



发布关联数据的要素

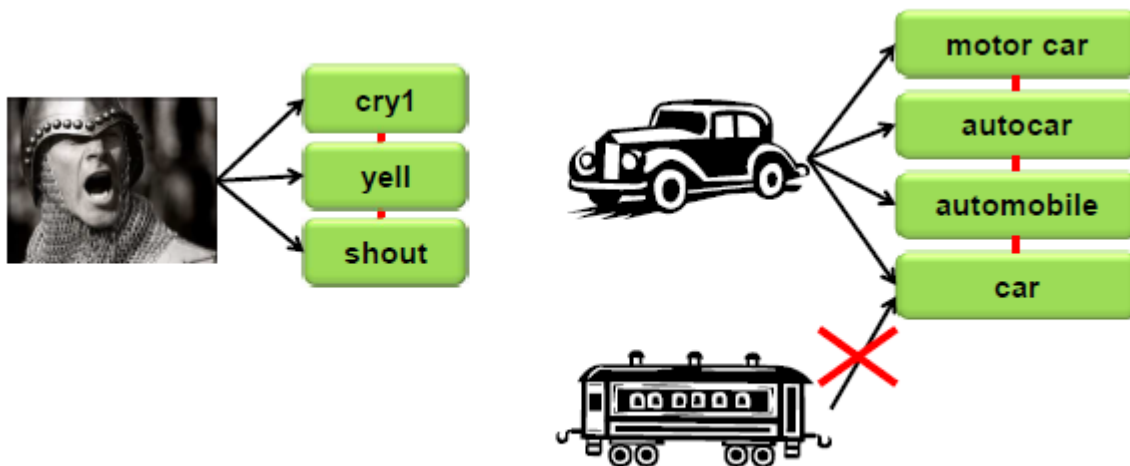
- 资源：任何有意义的，有引用必要的，都可以称之为资源。
- 资源标识：任何一个资源都用一个 HTTP URI (Uniform Resource Identifiers) 来标识。
- 资源描述：数据 Web 主要通过 RDF 格式 (Resource Description Framework) 来表示。RDF 将一个资源描述成一组三元组 (主语，谓语，宾语)。
- 数据关联：数据与数据建立链接。

数据的歧义

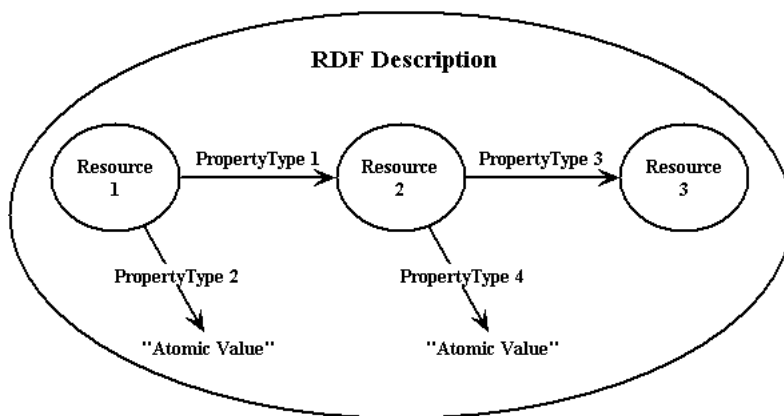


解决方案

- 给定唯一的识别符 (URI)
- 通过本体达到对领域知识的共同理解

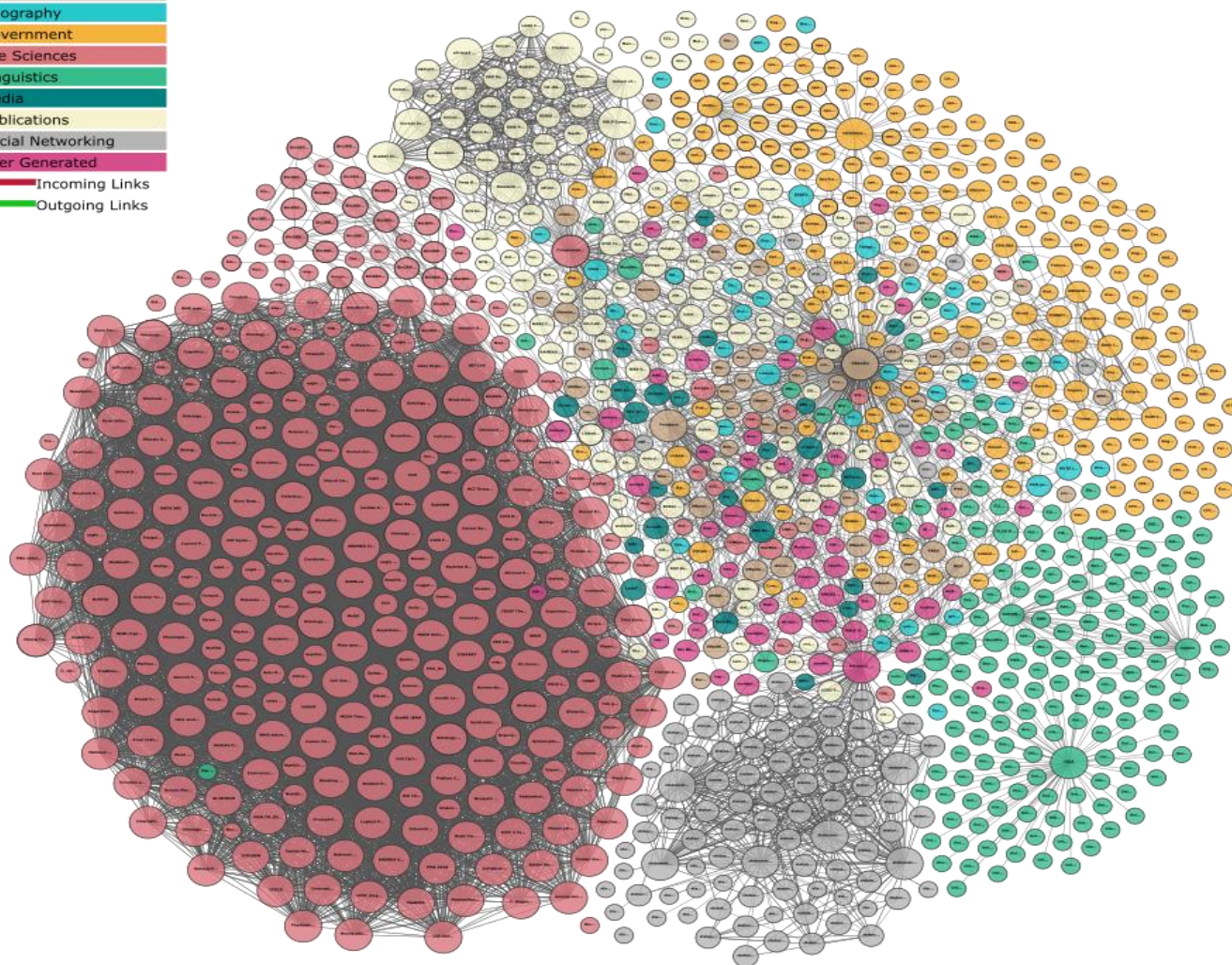


RDF的三元组



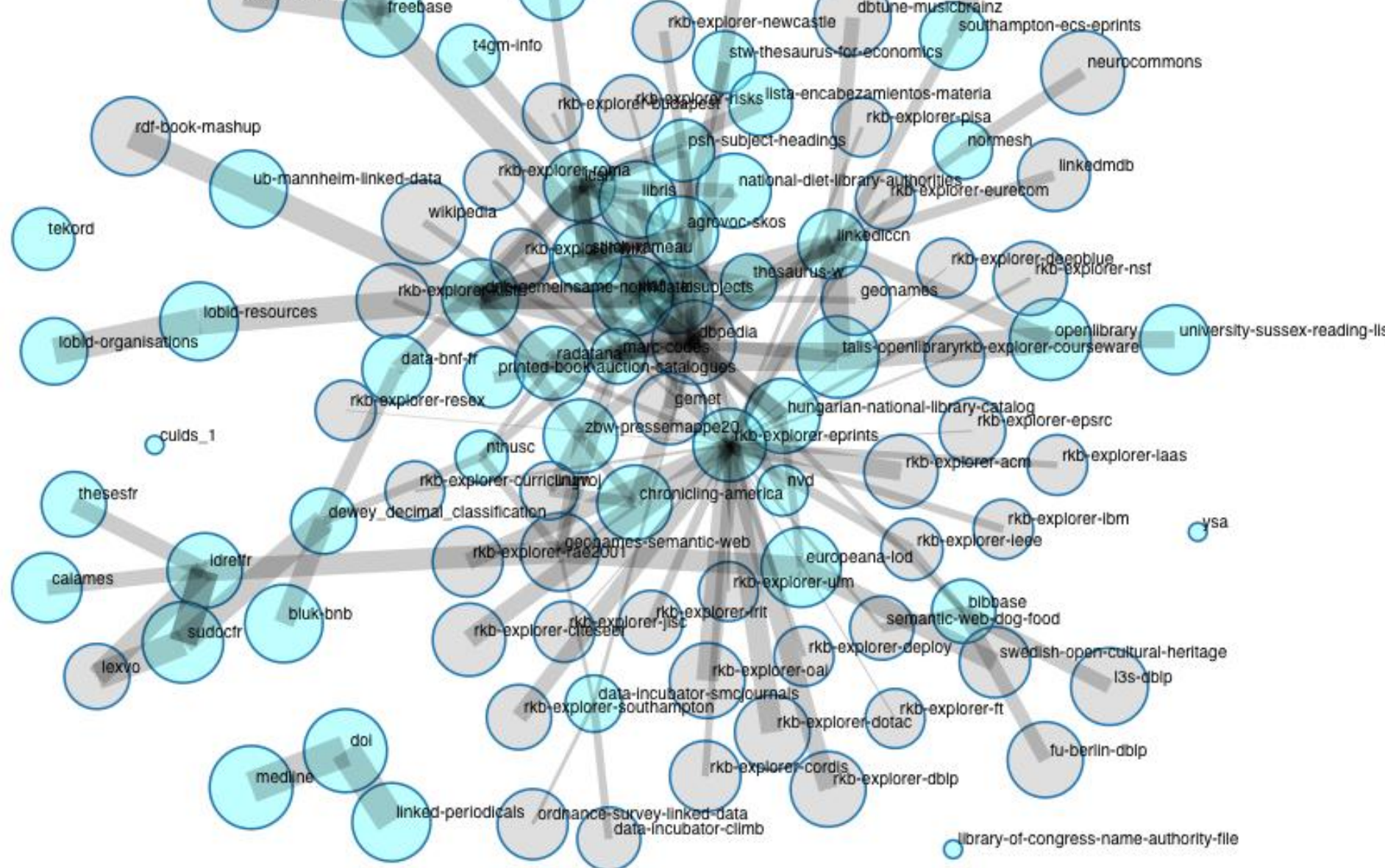
- 一组三元组称为RDF图，一个RDF图也是资源。
- 多个RDF图组成RDF数据集。
- 将RDF图或数据集编码为字符串，成为RDF文档，对应不同的语法格式有Trurtle、XML、RDFa、JSON-LD等。

关联数据云图-LOD



图书馆关联数据

- W3C成立 图书馆关联数据孵化组（[Library Linked Data Incubator Group](#)），致力于图书馆的关联数据发布，旨在提供图书馆数据在网络上的互操作性。
- 图书馆数据包括：
- 书目记录数据集
- 值词表（分类法、主题词表）
- 元数据集（DC/MARC/FRBR/RDA）



图书馆关联数据的优势

- ❖ 元数据的开放与共享
- ❖ 资源的发现
- ❖ 识别用户行为
- ❖ 分面浏览

元数据的开放与共享

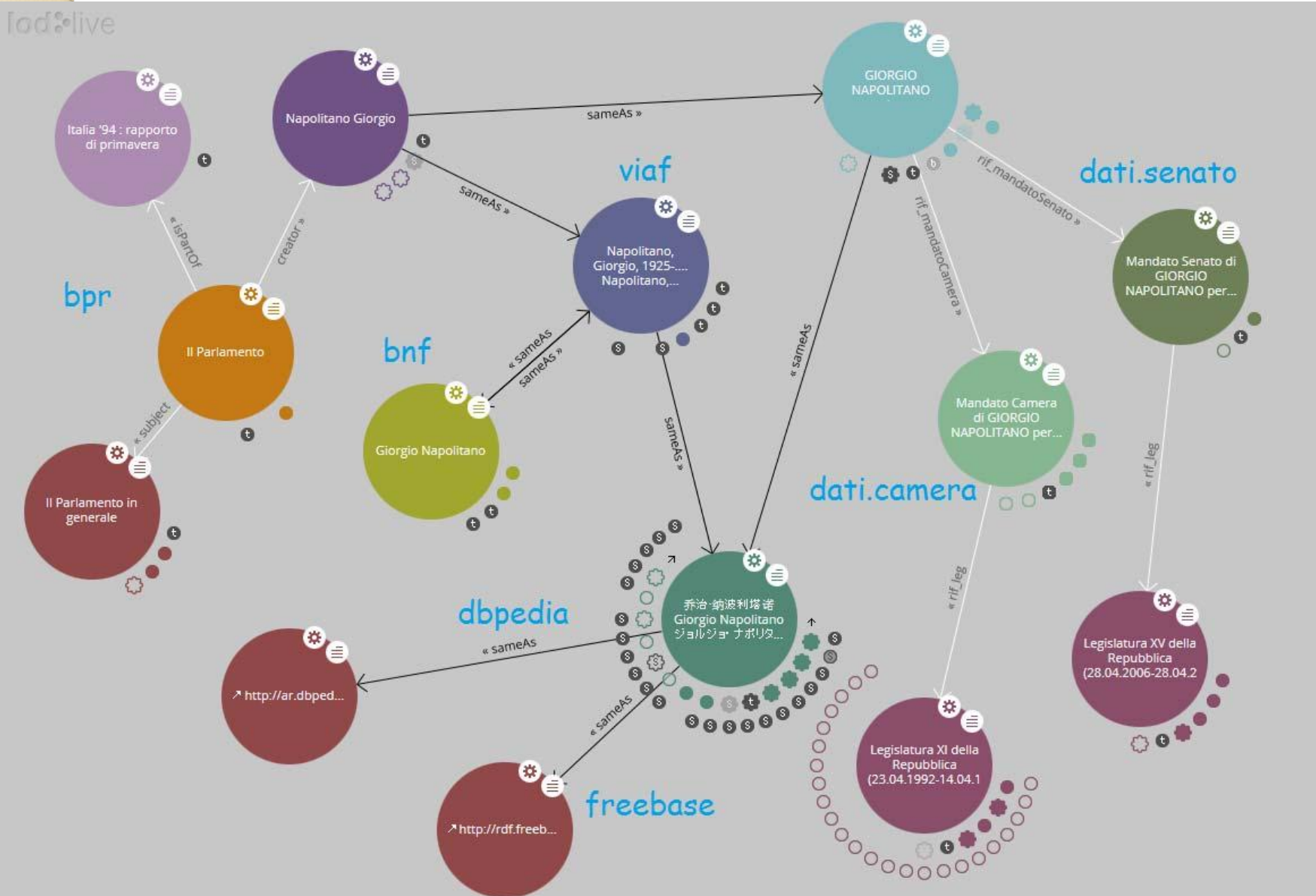
开放书目数据通过使用全球性的唯一识别符以分配给作品、地点、人、事件、主题及其它对象或者感兴趣的概念，以确保被自由地使用、重用及再分配。同时图书馆允许资源被其他广泛的数据源所引用，将使图书馆元数据能够被重用，以消除数据不必要的重复，会使自身的元数据描述更加丰富。

Europeana 项目

- 向欧洲用户提供跨机构的文化遗产数字化资源，用户采用其各自的语言检索获取到相关知识。
- 典型的知识聚合项目，将来自不同环境中的数字对象的元数据进行聚合及利用，定义了统一的数据模型Europeana Data Model (EDM)，将: MARC21, UNIMARC, MODS, METS, Dublin Core等格式的元数据转换为 EDM。

资源的发现

通过开放书目数据的发布，图书馆将连接到无数且相关的数据库、资源，有效地实现了异构数据的聚合，从而允许用户无缝地从一个资源跳转到另外一个资源，实现对数字图书馆的偶然发现与存取。



识别用户行为

图书馆元数据与用户生成的社会类元数据更易于合并，增加了额外的语义信息，既更好地提供丰富完整的元数据描述，又能提供针对用户使用的各种解释说明及术语，以准确地反映用户的信息行为，从而使分析用户的使用资源方式(如点击数、下载量、标签云、评价、推荐性资源)更为客观与便利，以帮助图书馆在获取资源、提供服务时做出准确决策。

分面浏览

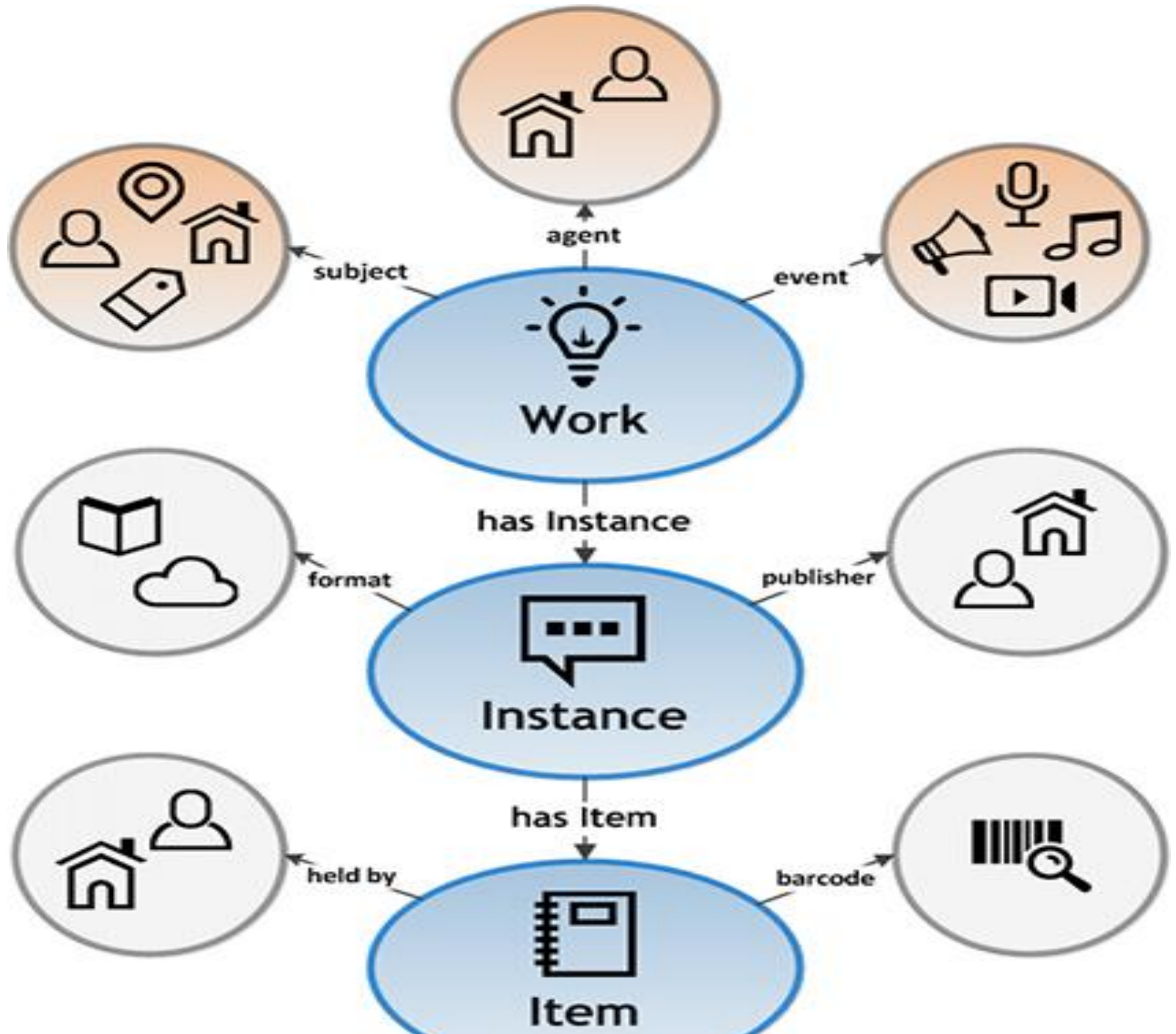
结构化数据允许搜索引擎生成良好结构的面，以区分结果的二义性，降低噪音及选择超载（choice overload）的前提下提高检索的精确率。

BIBFRAME模型

国会图书馆提出以关联数据技术为核心的资源描述模型，以FRBR/RDA为基础，为图书馆适应未来网络世界的书目描述提供基础，提供MARC 21 转换途径，通过数据交换支持资源共享及节省编目成本。

BIBFRAME目标

- 提供给搜索引擎可以使用的描述方式
- 推动FRBR、RDA模型的概念应用
- 为更广泛的文化遗产社区提供资源描述的扩展途径



BIBFRAME模型构成

- 作品层（内容）：不同题名或语种聚类，有施事者、主题、事件属性
- 实例层（载体）：以载体记录为核心，有资源类型、出版者属性
- 可描述的资源类型有：
- Text、Cartography、Audio、NotatedMusic、NotatedMovement、Dataset、StillImage、MovingImage、Object、Multimedia、MixedMaterial
- 单件层（版权）：以馆藏记录为核心，有条码、馆藏机构属性

国会图书馆的关联数据服务之一： BF本体

- <http://id.loc.gov/ontologies/bibframe.html>
- 提供常用本体、受控词表及其他书目描述信息的机器存取服务
- 定义185个类，186个属性
- 分类、列表、RDF方式展示

Class List

AbbreviatedTitle	Generation	PlaybackCharacteristics
AccessPolicy	GenerationProcess	PlayingSpeed
AcquisitionSource	GenreForm	Polarity
+ AdminMetadata	GeographicCoverage	PostalRegistration
+ Agent	GrooveCharacteristic	PresentationForm
Ansi	Gtin14Number	Print
AppliedMaterial	Hdl	Production
Archival	+ Identifier	ProductionMethod
Arrangement	Illustration	Projection
AspectRatio	ImmediateAcquisition	+ ProjectionCharacteristics
Audio	+ Instance	ProjectionSpeed
AudioIssueNumber	IntendedAudience	+ ProvisionActivity
AudioTake	Isan	Publication
Barcode	Isbn	PublisherNumber
BaseMaterial	Ismn	RecordingMedium
BookFormat	Isni	RecordingMethod
BroadcastStandard	Iso	ReductionRatio
Capture	Isrc	RegionalEncoding
Carrier	Issn	ReportNumber
Cartographic	IssnL	Resolution
CartographicDataType	Issuance	RetentionPolicy
CartographicObjectType	Istc	Review
Cartography	Iswc	Role
Chronology	Item	Scale
+ Classification	Jurisdiction	Script
ClassificationDdc	KeyTitle	+ ShelfMark

Property List

acquisitionSource	ensembleType	partNumber
acquisitionTerms	enumerationAndChronology	pattern
adminMetadata	equinox	physicalLocation
agent	exclusionGRing	± place
appliedMaterial	extent	polarity
arrangement	firstIssue	preferredCitation
ascensionAndDeclination	fontSize	productionMetho
aspectRatio	frequency	projection
assigner	generation	projectionCharac
awards	generationProcess	provisionActivity
baseMaterial	genreForm	provisionActivity
bookFormat	geographicCoverage	qualifier
capture	grantingInstitution	reductionRatio
carrier	heldBy	+ relatedTo
cartographicAttributes	hierarchicalLevel	responsibilitySta
classification	historyOfWork	review
classificationPortion	identifiedBy	role
code	identifies	scale
colorContent	illustrativeContent	schedulePart
content	immediateAcquisition	seriesEnumeratio
contentAccessibility	instrument	seriesStatement
contribution	instrumentalType	shelfMark
coordinates	intendedAudience	soundCharacteri
copyrightRegistration	issuance	soundContent
count	itemPortion	source

BF本体注册到LOV平台

检索—中国知网 x LINKED OPEN VOCABULARIES x Linked Open Vocabularies x 安全检查 | 百度云加速 x agdistis.aksw.org/demo/ x

lov.linkeddata.es/dataset/lov/vocabs/bf

应用 淘宝 爱淘宝 聚划算 头条视频 热门游戏 传奇新品 JD 京东 九秀直播 hao123 苏宁 唯品会 1号店 看新闻 天气预报 影视大全 飞猪

BIBFRAME Vocabulary (bf)

Metadata

URI <http://id.loc.gov/ontologies/bibframe/>

Namespace <http://id.loc.gov/ontologies/bibframe/>

isDefinedBy <http://id.loc.gov/ontologies/bibframe.rdf>

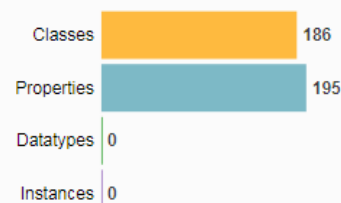
homepage <http://www.loc.gov/bibframe>

Description The BIBFRAME model and vocabulary consider resources that are cataloged as works with corresponding instances (physical and/or electronic). The metadata describing a work is an amalgamation of the some of the data that was formerly associated with the uniform title authority record combined with subject data that was associated with the bibliographic records. Thus each cataloging resource must have a work description, and, if the cataloging resource exists physically or electronically, one or more instance descriptions. @en

Language



Statistics

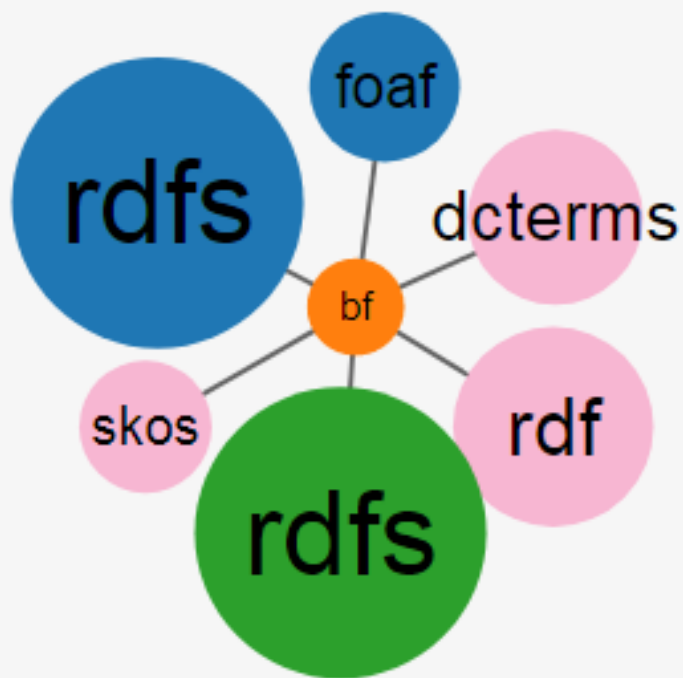


Expressivity

RDF RDFS

BIBFRAME核心类...caj 书目框架_BIBFRA...caj 基于BIBFRAME模...caj 语义联合目录_基...caj BIBFRAME_2.0词...caj 全部显示

BF建立与其他词表的关联



- Foaf:agent
- Dctermns:
- Skos:
- Owl:ObjectProperty
- Owl:DatatypeProperty
- owl:SymmetricProperty
- Rdfs:Literal

基于BF模型的关联数据转换

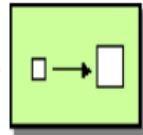
- 应用所设计的用户接口完成内容选择、转换、发布、管理任务。
- 转换：将选择的内容转换成RDF数据集
- 发布：将数据集发布在关联数据云平台上
- 链接：通过共同的实体或概念，将RDF数据集与关联数据云的其他数据集建立链接

BIBFRAME实现

- 书目数据已创建了实例层，需要在数据模型上创建作品层。
- 人物实体聚类：名称规范档、VIAF、ISNI
- 作品聚类：先链接到实例标题，再到OPAC个体的单件层
- 根据映射规则实现MARC到RDF映射

001 27283
300 \$a108 p. ; \$c15 cm.

R300c

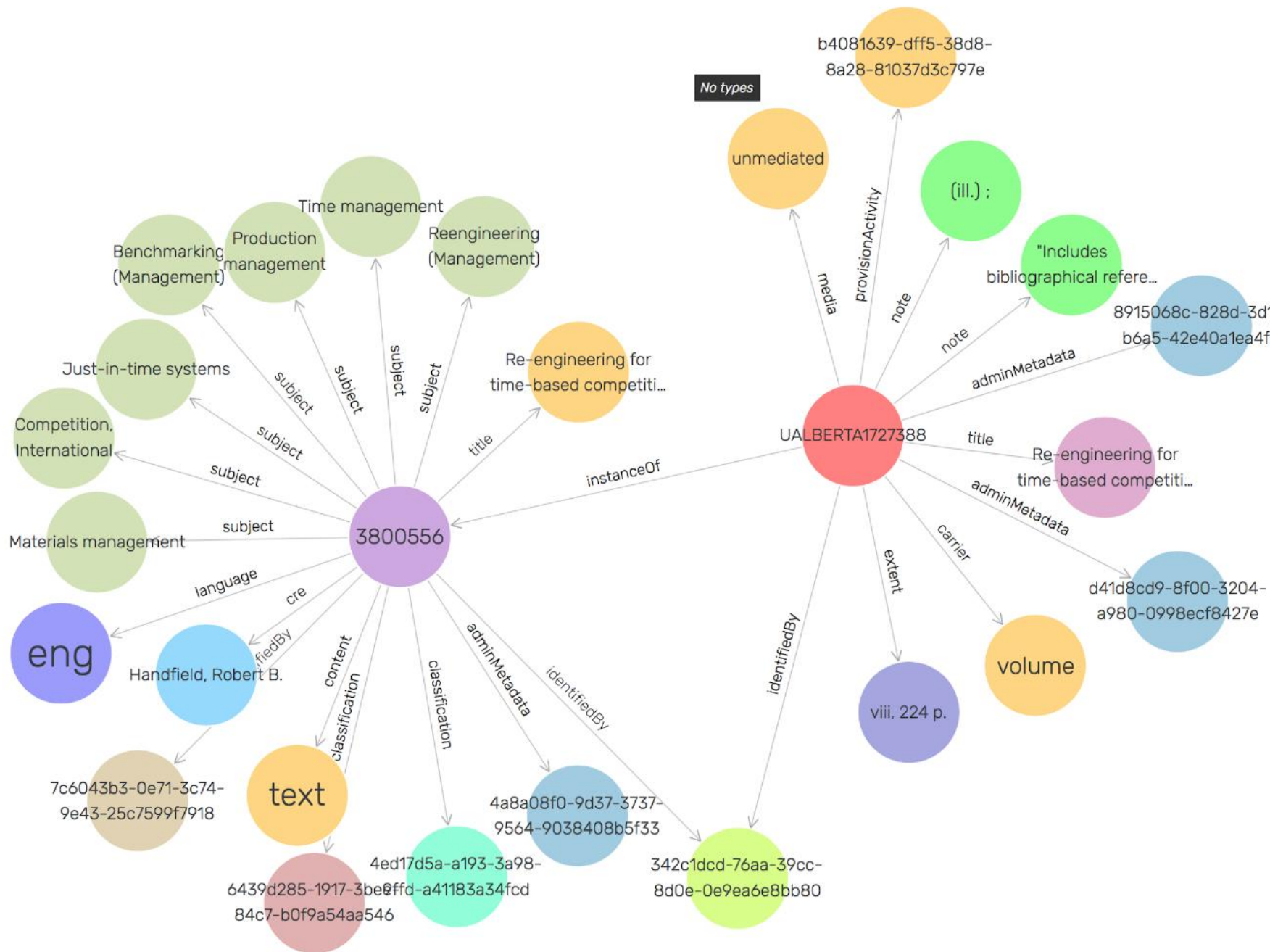


R300a

<<http://share-vde.org/sharevde/rdfBibframe/Instance/27293>> <bf:dimensions> "15 cm."

<<http://share-vde.org/sharevde/rdfBibframe/Instance/27293>> <bf:extent>
<<http://share-vde.org/sharevde/rdfBibframe2/Extent/02e3f96a>>

<<http://share-vde.org/sharevde/rdfBibframe2/Extent/02e3f96a>> <rdfs:label> "108 p. ;"



资源发现

通过统一检索入口快速获取到图书馆内外部资源。通过对各种不同数据库和检索系统的信息资源进行预索引，建立元数据的集中索引仓储，提供“一站式”的检索服务和更高级的发现服务。

集成实现

ILS Vendor

Millenium
Sierra
Polaris
Symphony
Horizon
WorldShare
Voyager
ALEPH500
Alma

Discovery Layer

Encore
EDS
Summon
AquaBrowser
WorldCat Local
Vufind
Primo
BlackLight

Drag to rearrange nodes.

Federated Search

Search:

Search Results

ILS Data

Digital
Collections

ProQuest

EBSCOhost

...

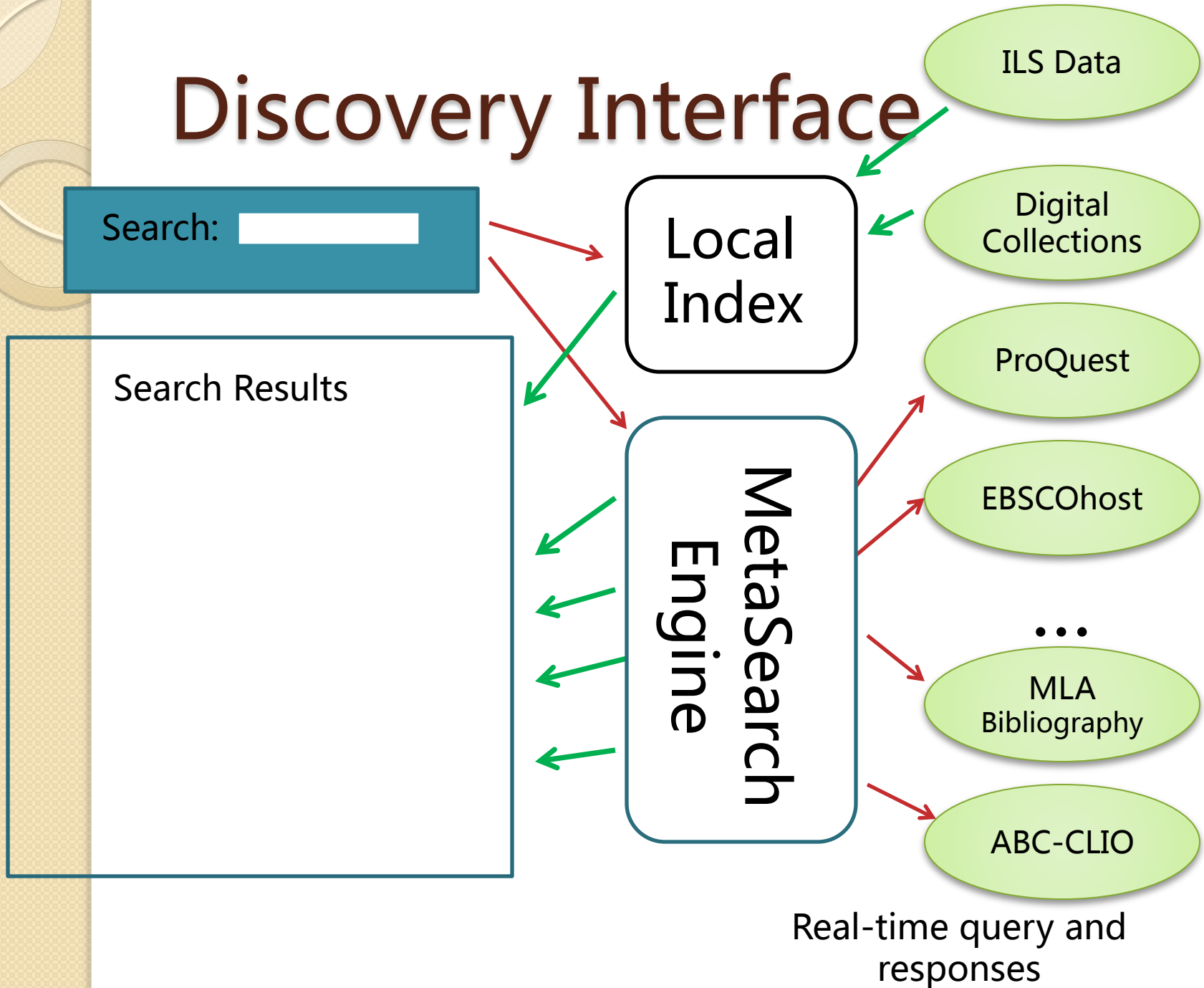
MLA
Bibliography

ABC-CLIO

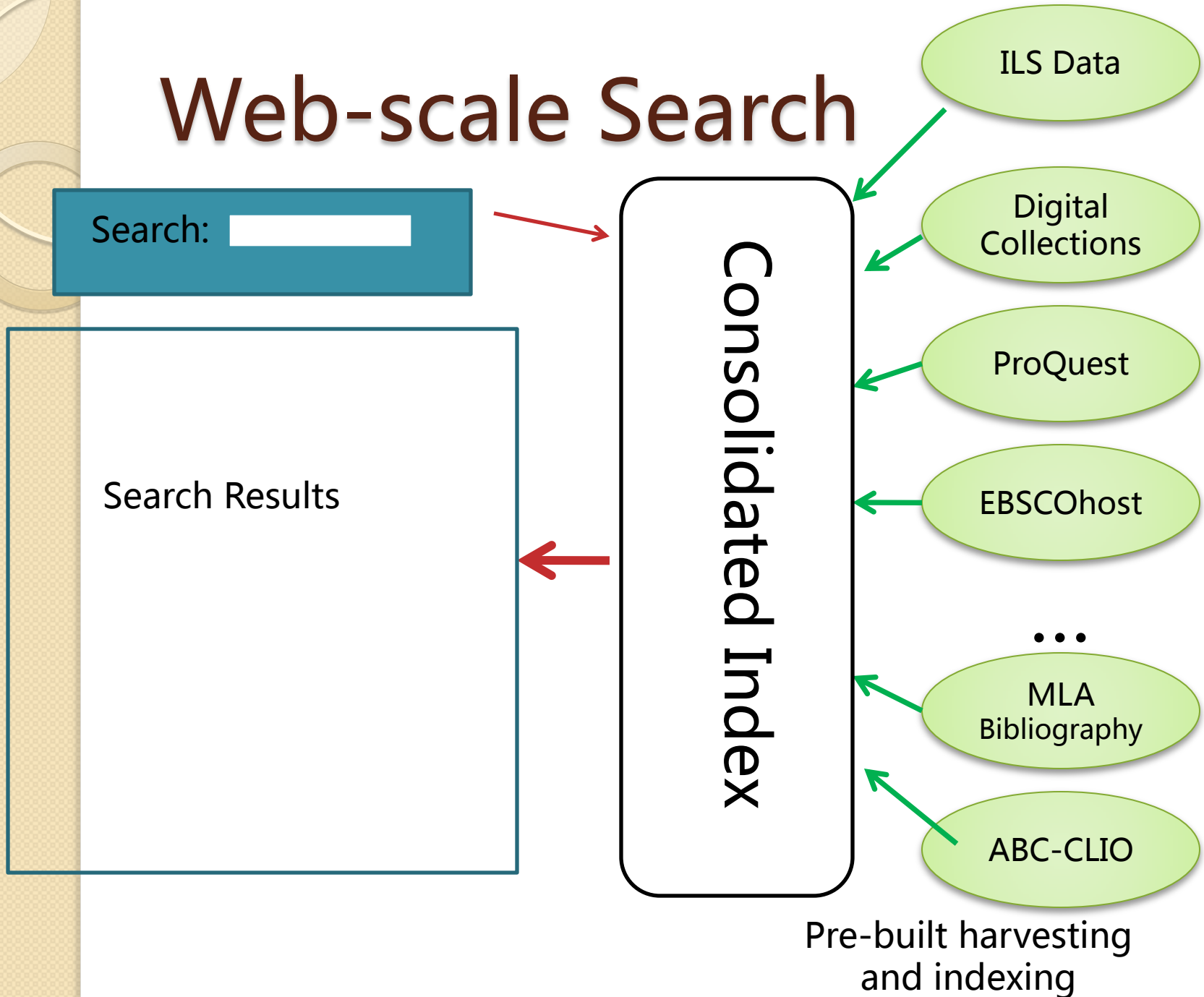
Real-time query and
responses

The diagram illustrates the Federated Search process. A search input box on the left sends red arrows (queries) to a list of data sources on the right. These sources include ILS Data, Digital Collections, ProQuest, EBSCOhost, an ellipsis, MLA Bibliography, and ABC-CLIO. Green arrows (responses) return from each source to a 'Search Results' box on the left. A label at the bottom indicates that these interactions are 'Real-time query and responses'.

Discovery Interface



Web-scale Search



Web-scale检索特点

- 包含图书馆馆藏外部的资源，如网络资源
- 通过概念及概念关系实现语义检索
- 全文索引，实现基于内容的检索

基于关联数据的图书馆资源发现

- 使用基于HTTP的URI识别作品、人物，解决名称冲突问题，充分利用名称规范。
- 传统数据库是封闭世界，关联数据是开放世界，可以获得更多的外部数据，基于索引的资源发现服务将收割、索引关联数据的内容，充分利用关联数据的关系，以提高基于关键词检索的发现服务。
- 资源发现系统使用URI，集成其他来源信息，生成人物的个人简介描述信息。

卡萨里尼图书馆关联书目项目

Browser tabs: linked x, Linked x, LOC BIBFRA x, SHARE x, SHARE x, W3 Library x, The Ro x, LOC BIBFRA x, LOC BIBFRA x, interna x, Precisi x

Address bar: 不安全 | share-vde.org/sharevde/searchNames?n_cluster_id=105909

Navigation bar: 应用, 爱淘宝, 聚划算, 头条视频, 热门游戏, 传奇新品, JD 京东, 九秀直播, hao123, 苏宁, 唯品会, 1号店, 看新闻, 天气预报, 影视大全, 飞猪

Site header: Browse, SHARE, Discovery, Virtual, Environment, Person/Work, Italiano, Info | Contacts, Go to Publication

Search filters: Person, Work, Publishers, Subject

Search results: Search Person/Family/Corporate body

Buttons: EXPAND ALL, CLOSE ALL

Person details: This person in, Wikidata, Library of Congress, WorldCat Identities, data.bnf.fr, isni, VIAF

Person image: Lu, Xun, 1881-1936

Other name forms: Lu, Xun, 1881-1936, Lu, Xun, Lu, Xun, pseud., 1881-1936, 1936-1881, 鲁迅, 1881-1936, 鲁迅, 1881-1936, 鲁迅, 1881-1936, 鲁迅, 鲁迅

Works: Works

个人简历

EntityJS



War Between the States collection

GO TO SEARCH

Lincoln, Abraham, 1809-1865.

About

Abraham Lincoln /ˈɛbrəhæm ˈlɪŋkən/ (February 12, 1809 – April 15, 1865) was the 16th President of the United States, serving from March 1861 until his assassination in April 1865. Lincoln led the United States through its Civil War—its bloodiest war and its greatest moral, constitutional and political crisis. In so doing he preserved the Union, abolished slavery, strengthened the federal government, and modernized the economy. Reared in a poor family on the western frontier, Lincoln was a self-educated lawyer in Illinois, a Whig Party leader, state legislator during the 1830s, and a one-term member of the Congress during the 1840s. He promoted rapid modernization of the economy through banks, canals, railroads and tariffs to encourage the building of factories; he opposed the war with Mexico in 1846. After a series of highly publicized debates in 1858 during which he opposed the expansion of slavery, Lincoln lost the

Related Works

As a Creator

[Abraham Lincoln : miscellaneous papers, 1840-1864.](#)

[Abraham Lincoln campaign leaflet 1864.](#)

[Abraham Lincoln collection, 1824-1865.](#)

[Abraham Lincoln collection, 1847-1864.](#)

[Abraham Lincoln correspondence.](#)

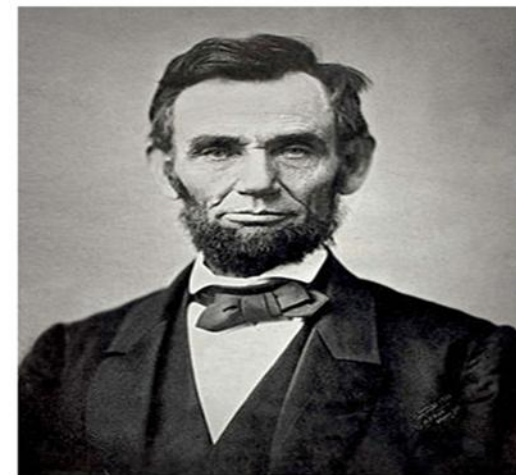
[Abraham Lincoln Document, ca. early 1860s.](#)

[Abraham Lincoln note and photograph, 1863 August 20.,](#)

[Abraham Lincoln papers \(at Chicago History Museum\), 1800-\[ongoing\].](#)

[Abraham Lincoln papers. 1774-1948.](#)

Related Media



[Source: Wikimedia Commons](#)

Related People

[Ulysses Simpson Grant](#)

Related Organizations

[United States. Army.](#)

Related Places

[United States](#)

提升用户体验



Malcolm Gladwell

Born: September 3, 1963
Role: Author

Malcolm T. Gladwell, CM (born September 3, 1963) is a Canadian journalist, bestselling author, and speaker. He has been a staff ...

[Read more »](#)

Works by 13

-  [Outliers: the story of success](#)
-  [The tipping point: how little things can make...](#)
-  [Blink: the power of thinking without th...](#)

Works about 109

-  [Malcom Gladwell, Editor, Good news, ...](#)
-  [Features - How Change Really Happens - It's...](#)
-  [There's A Convincing Theory That ...](#)

Sort by

Relevance Only ▼

Content ▼

☐ Full Text

☐ Peer Reviewed

Format ▼

[David and Goliath : underdogs, misfits, and the art of battling giants](#)

by Malcolm Gladwell

Held by: OCLC Library



View all editions and formats 18

 Book	2005	1st ed	English	 Available 3
 eBook	2005	—	English	 Access Online
 Book	2008	2nd ed	Spanish	 Available 2

- 
- 汇报完毕!
 - 谢谢!