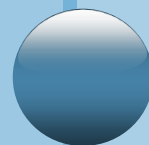




关联数据应用研究进展

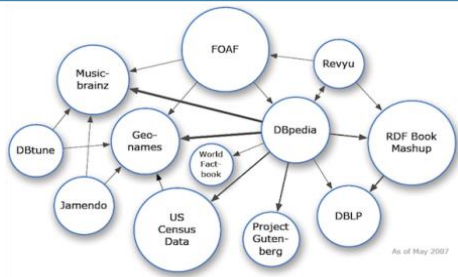


刘晓娟

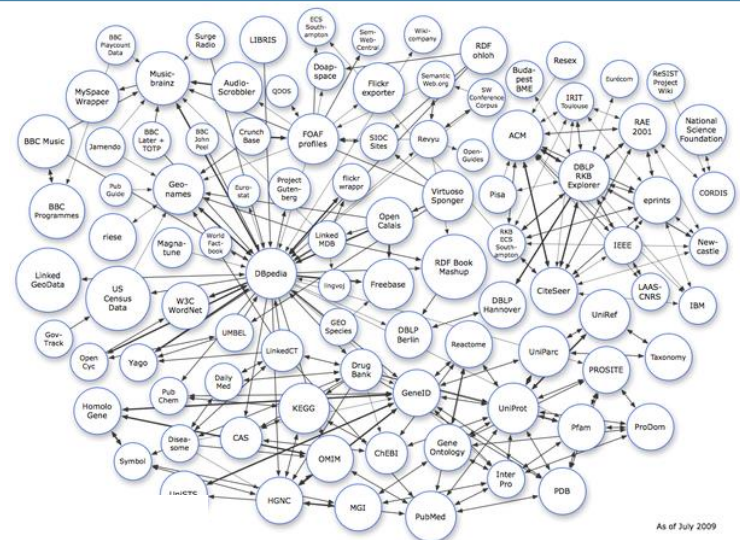




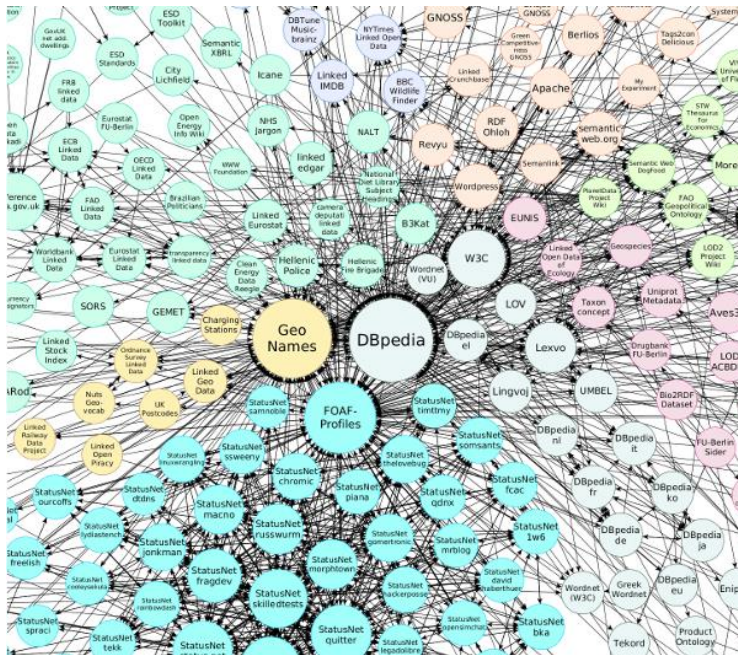
提起关联数据



2007



2009



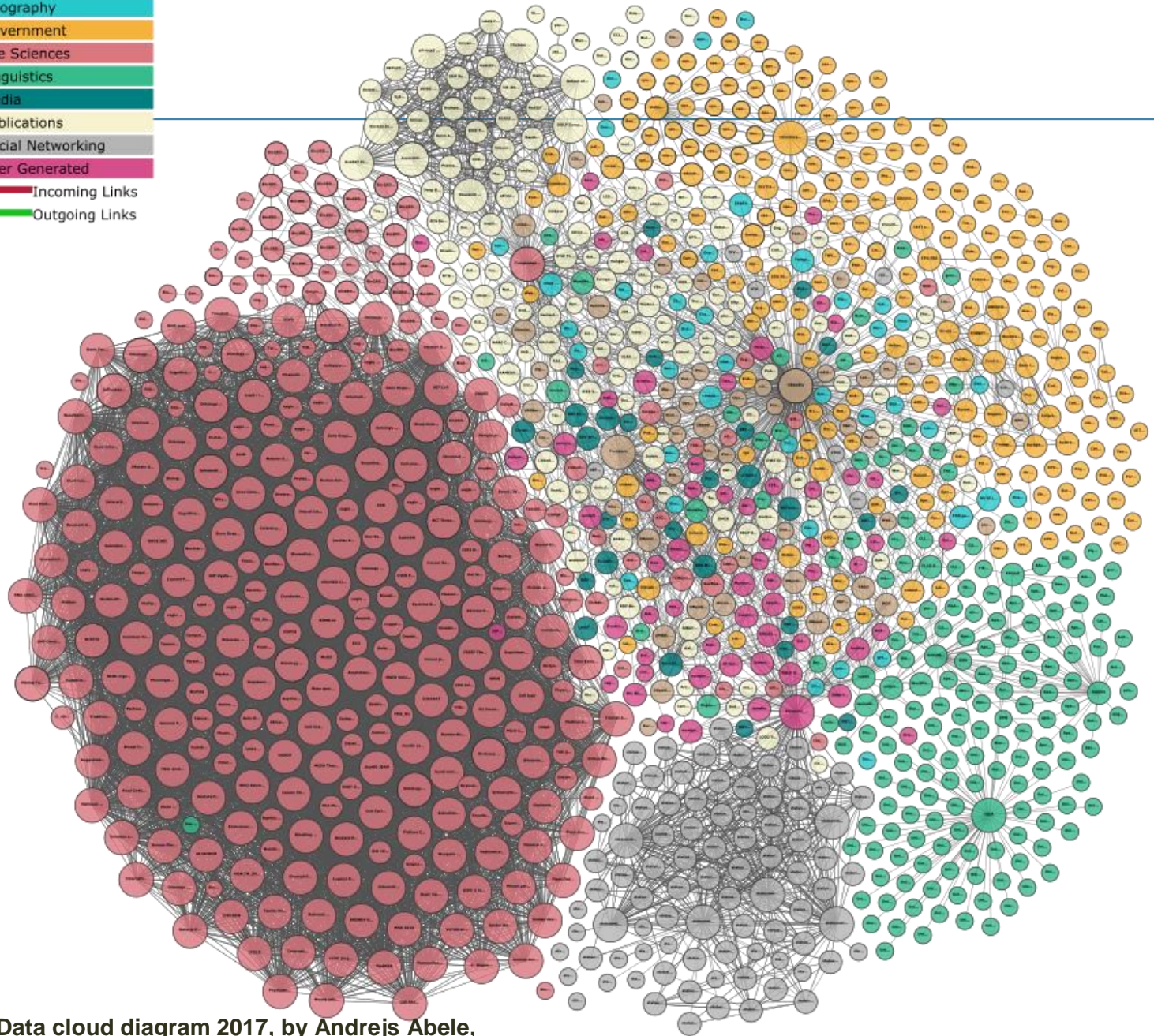
- Media
- Geographic
- Publications
- User-Generated Content
- Government
- Cross-Domain
- Life Sciences
- Social Networking

2014



Legend

- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated
- Incoming Links
- Outgoing Links



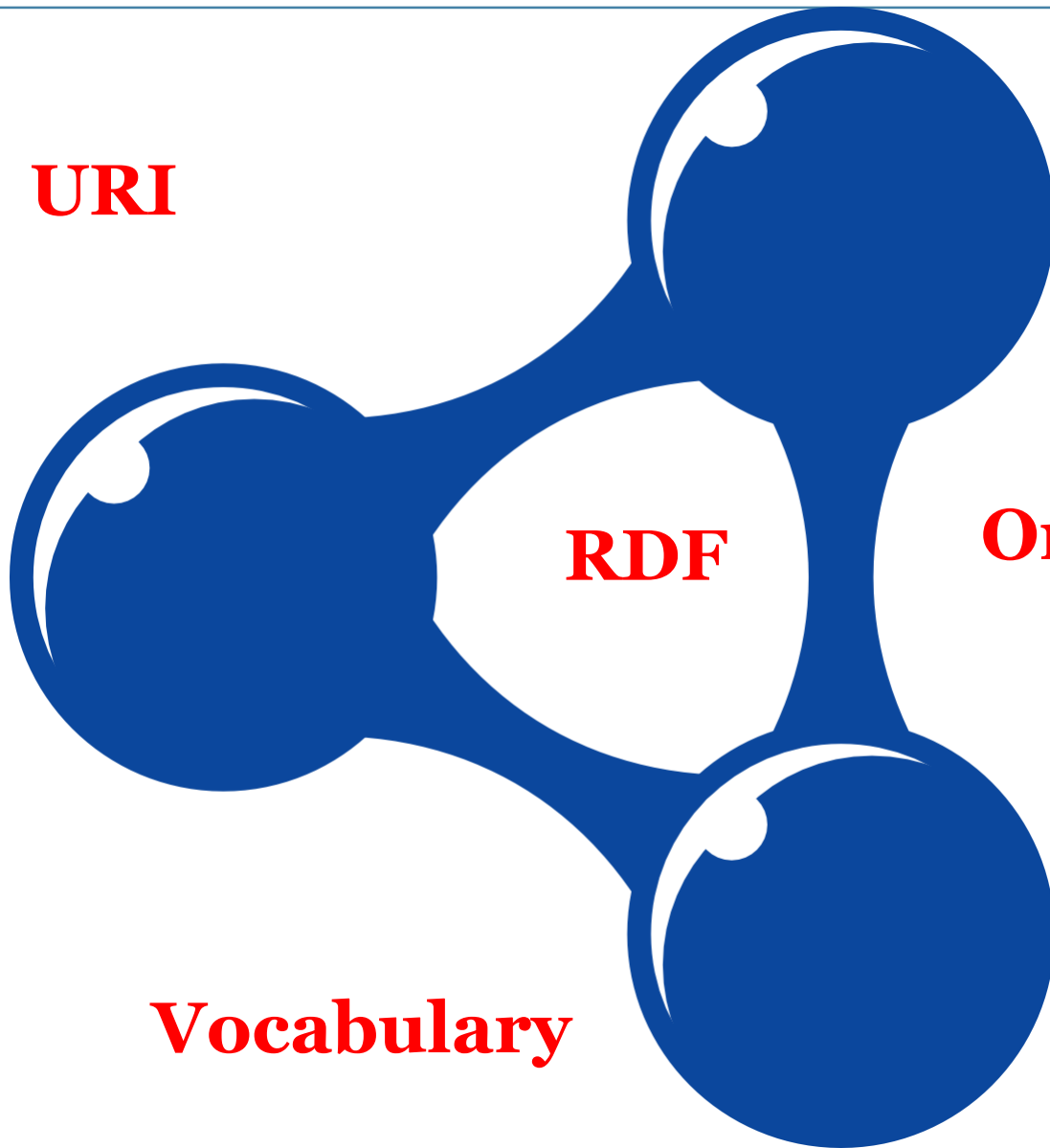
Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cvdaniak. <http://lod-cloud.net/>



提起关联数据



URI



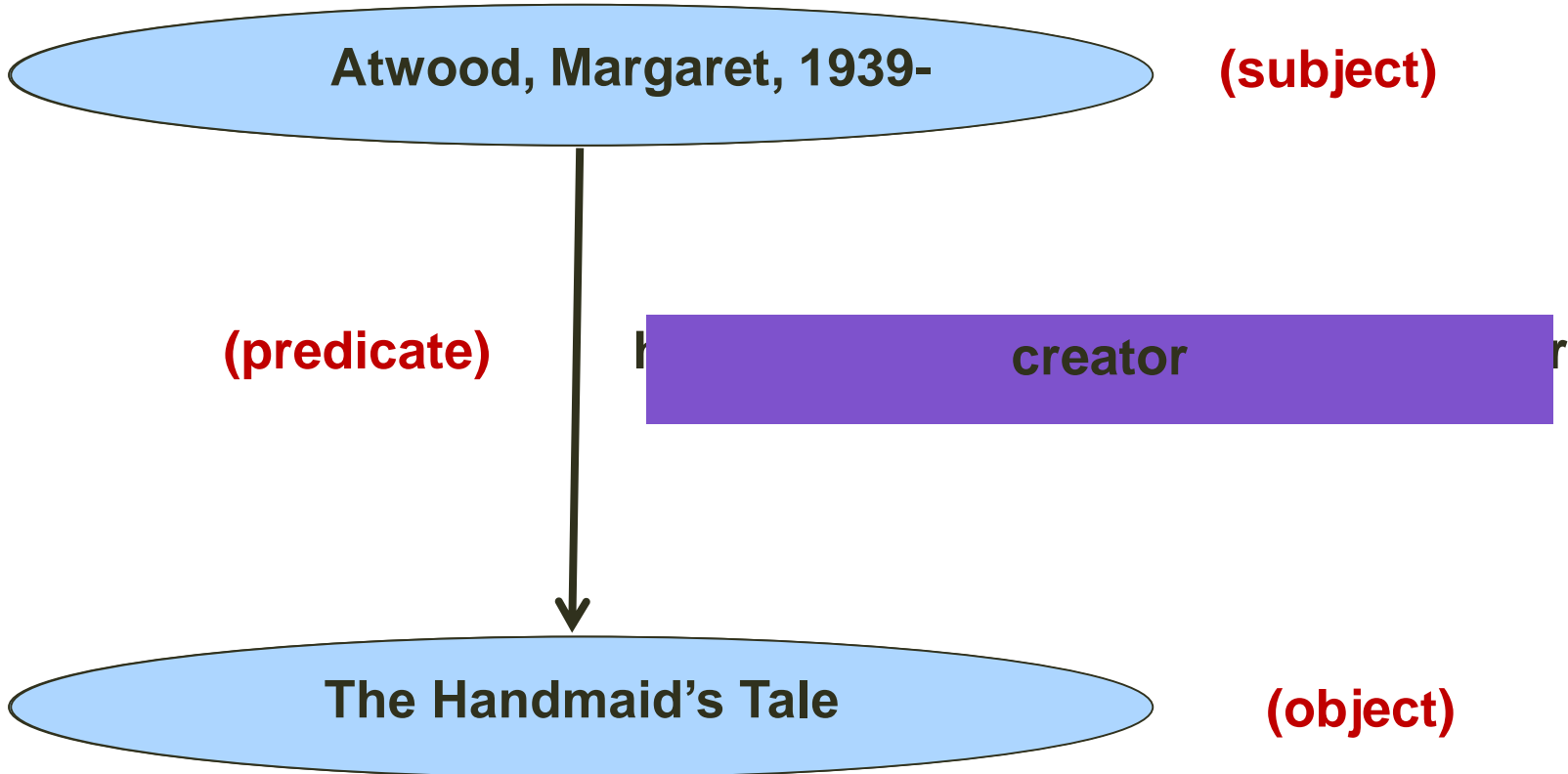
Ontology

Triple

Vocabulary



特别是RDF

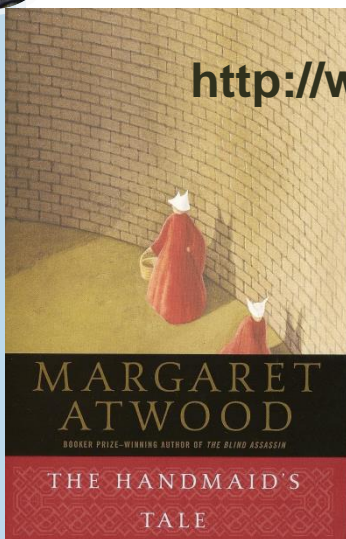




关联数据的四项基本原则



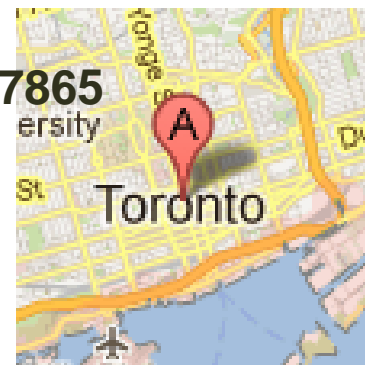
- (1) 使用URI作为任何事物的标识名称，不仅是标识文档
 - Use URIs as names for things



<http://www.worldcat.org/oclc/12558693>

<http://xmlns.com/foaf/0.1/knows>

<http://sws.geonames.org/6167865>



<http://viaf.org/viaf/109322990>



- (2) 使用HTTP URI , 使任何人都可以参引 (dereference)这一全局唯一的名称
 - Use HTTP URIs so that people can look up those names



- (3) 当有人访问名称时 , 以规范的标准(RDF, SPARQL) 提供有用的信息
 - When someone looks up a name, provide useful information



- (4) 尽可能提供链接，指向其它的URI，以使人们发现更多的相关信息

- Include links to other URIs so that they can discover more things

<http://www.utoronto.ca/person/Margaret-Atwood>

http://dbpedia.org/resource/Margaret_Atwood

http://dbpedia.org/resource/Women_novelists

http://dbpedia.org/resource/Booker_Prize_winners

Links to more things

Links to more things



接下来



- 什么是关联数据的应用？
- 谁在应用关联数据？
- 为什么要应用关联数据？
- 如何应用关联数据？
- 关联数据应用中有哪些技术和工具？
- 已有的应用案例有哪些？
- 应用中存在哪些障碍？
-



图书馆、信息服务机构与关联数据



● 作为数据提供者

- 书目资源
- 受控名称规范文档
- 词表类表
- 本体

发布/Publish

● 作为数据消费者

- 利用这些数据为用户服务

消费/Consume



从用户角度来看



- 更完善、丰富的数据
- 跨资源、跨专业的利用与分享
 - 关于人名的规范数据+ dbpedia或VIAF
 - 地名文档+ dbpedia
- 这与人工做的‘链接’是不同的，这些是数据，是无限的数据支持的而且是自动联结的。



什么是关联数据的应用



- Software system that makes use of data on the web from **multiple datasets** and that benefits from links between the datasets



关联数据应用的特点



- 根据关联数据原则对发布在Web上的数据进行**消费**：应用程序应该能够请求、检索和处理访问的数据
- 通过跟踪不同数据源之间的链接来**发现**更多信息
- 将所消费的关联数据与数据源（不一定是关联数据）**结合起来**
- 根据关联数据原则将合并后的数据**公布**在Web上
- 为最终用户提供**价值**



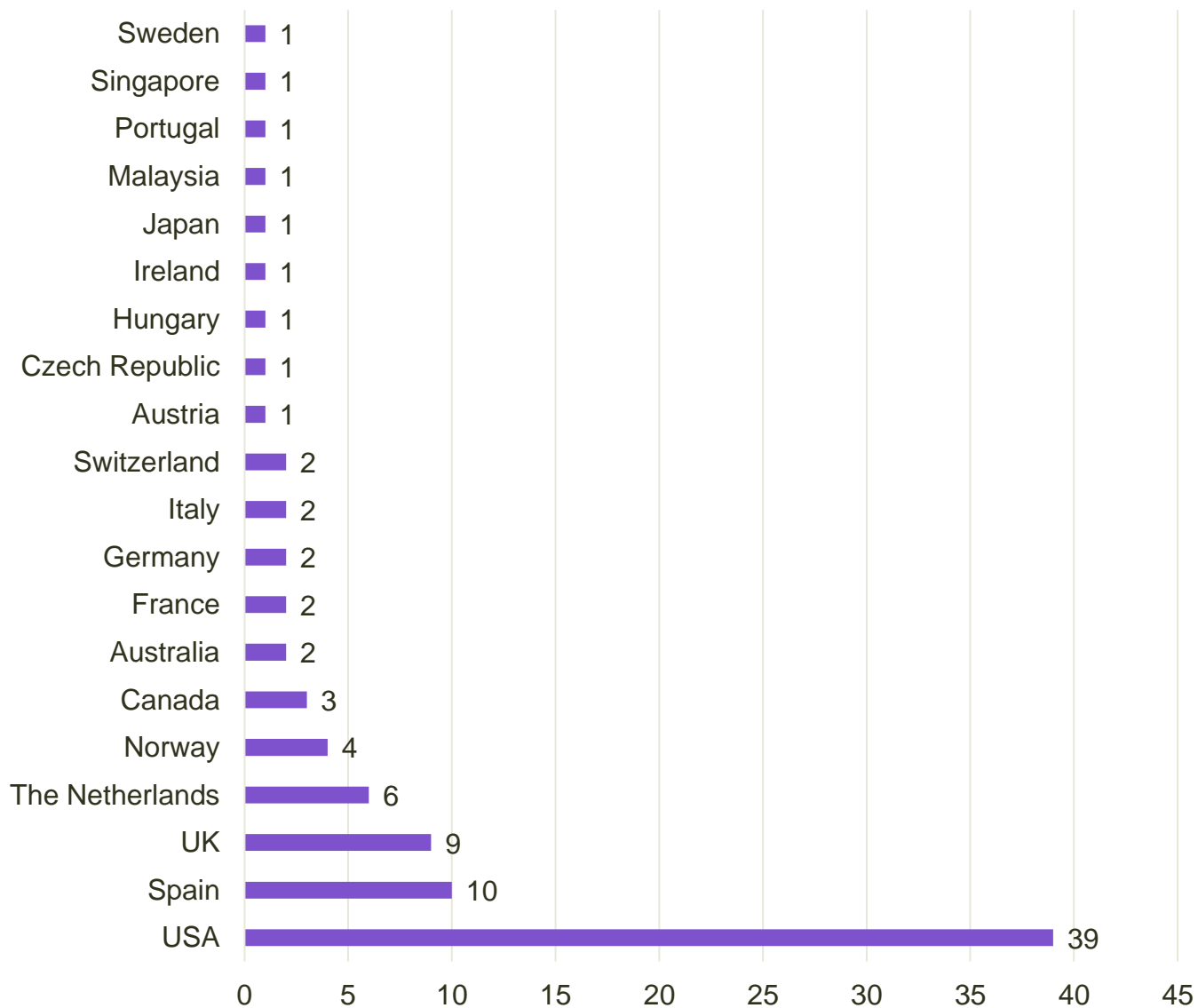
关于关联数据的国际性调查

——OCLC Research

调查轮次	第一次	第二次
调查时间	2014.7.5-8.15	2015.6.1-2015.7.31
#关联数据项目或服务	76	112
#国家 (20)	15	20
#机构 (90)	48	71
调查目的	向那些曾经实施过或者正在实施关联数据项目或服务的人学习经验	增加调查参与度，特别是国家图书馆
		识别关联数据环境中的变化
调查内容	Who's doing it	
	Examples in production	
	Why and what institutions are consuming	
	Why and what institutions are publishing	
	Technical details	
	Advice from the implementers	



调查对象机构的国家分布





项目合作、人员、资助



- 项目/服务所涉及的外部团体：大部分（69%）没有涉及到任何层次的外部合作。合作者包括：其他大学或研究机构；其他图书馆或档案馆；其他联盟成员；系统供应商；公司等。
- 人员配置：几乎所有（98个）已经或正在实施关联数据项目/服务都将关联数据加入现有工作人员的职责中。其中20个有专门负责关联数据项目的工作人员；4个正在或已经增加了关联数据专家为新员工；13个正在或已经增加了具有关联数据专长的临时员工；17个正在或已经雇佣了具有数据专长的外部顾问。
- 资金资助：25个项目获得基金资助；82个由图书馆/档案馆和/或附属机构；5个来自于合作机构的资金支持；1个来自于公司；1个由私人资助。



关联数据是如何利用的？



如何利用	2015	2014
消费	38	25
发布	10	4
消费和发布	64	47



关于关联数据的发布



- 哪些类型的数据被发布?
- 所发布的关联数据规模如何?
- 关联数据发布的动机
- 哪些关联数据集比较常用?
- 哪些RDF词表和本体用的比较多?
- 数据许可标准
- 可访问性
- 所采用的技术
- 关联数据发布中存在哪些障碍?



哪些类型的数据被发布?



- bibliographic
- authority data
- descriptive metadata
- Ontologies/vocabularies
- digital collections
- geographic data
- datasets
- data about museum objects
- encoded archival descriptions
- organizational data
- data about researchers or library staff



所发布的关联数据规模如何？



- 大部分关联数据的数据集都很小。67个反馈中报告了他们的数据集的大小
 - 39个低于1000万个三元组
 - 19个超过了1亿
 - 3个超过10亿
 - North-Rhine-Westphalian Library Service Center (1-5 billion)
 - Norwegian University of Science and Technology's "various" linked data projects totaling 15 billion triples
 - OCLC's WorldCat Linked Data (15 billion triples)



关联数据发布的动机



- 向Web上的更广泛受众公开数据 (91%)
- 论证将数据集作为关联数据可以做什么？ (80%)
- 听说过关联数据，希望通过公开一些本地数据为关联数据进行尝试 (58%)
- 探索将数据发布为关联数据是否可以为本地资源提高搜索引擎优化 (SEO) (38%)
- 将关联数据发布后用于未来项目中的消费和重用
- 最大限度提高数据的互操作性和可重用性
- 测试BIBFRAME3和schema.org
- 项目需求
- 提供整个机构科研活动的稳定、集成、标准化的数据



哪些关联数据集比较常用?



- 已实施的大多数项目/服务在过去六个月中平均每天收到少于1000个请求。
- over 100,000 requests per day
 - Europeana
 - The Getty Vocabularies
 - Library of Congress' Linked Data Service with over 50 vocabularies
 - National Diet Library's NDL Search
 - North Rhine-Westphalian Library Service Center's Linked Open Data service
 - OCLC's WorldCat Linked Data
 - OCLC's Virtual International Authority File (VIAF)



哪些RDF词表和本体用的比较多?



● Top 5

- Simple Knowledge Organization System (skos)
- Friend of a Friend (foaf)
- DCMI Metadata Terms (dcterms)
- Dublin Core Metadata Element Set (dce)
- Schema.org vocabulary (schema)





数据许可标准



● Top 5

- **CC0 1.0 Universal**
- Open Data Commons Attribution (ODC-BY)
- Open Data Commons Open Database License (ODC-ODbI)
- Public Domain Dedication and License or PDDL
- Creative Commons Attribution-NonCommercial-NoDerivatives (BY-NC-ND)



可访问性



- 大部分提供了多种方式呈现数据，其中网页是最常见的，其次是内容协商、file dumps, SPARQL endpoint, SPARQL 编辑器和应用。
- 最常见的关联数据序列化方式是 RDF/XML，其次是Turtle, JSON-LD, N-Triples, RDFa Core, RDF/JSON, Notation3 和N-Quads。



所采用的技术



- 用于发布关联数据的技术多样化，并且大多项目采用多种技术

项目数量	技术 (根据频率排序)
10 或更多	SPARQL, Java, XSLT, Zorba
2-9	Solr, Virtuoso Universal Server (provides SPARQL endpoint), Google Refine, Jena Applications, RDF Store, Drupal7, Python, Apache Fuseki, ElasticSearch, Perl, Metafactory, DIGIBIB, MongoDB, 4store, Apache Marmotta, BlazeGraph, GraphDB (formerly OWLIM by Ontotext Software), Hydra, Numishare (built on Orbeon, eXist and Solr), Rails
1	Apache Tomcat, Cubicweb, Django, dotNetRDF, Fedora Commons, Hbase/Hadoop, JAX-RS, Joomla, LibHub, Mapping Memory Mapper (3M), MARC Report and MARC Global (from The MARC of Quality), MySQL, Node.js, Oracle, PoolParty, Protège, Pubby, r2rml-parsers, Ruby, Ruby Virtuoso triplestore, Sesame, skosmos, Wordpress



关联数据发布中存在的障碍



- 工作人员面临陡峭的学习曲线
- 遗留数据存在不一致性
- 选择适当的本体表示数据
- 建立链接
- 缺乏如何构建系统的文档或建议
- 缺乏工具
- 软件并不成熟
- 确定数据的拥有者
- 限制性许可、资源不足、数据集太大难以作为整体发布、机构支持不足等等



关于关联数据的消费



- 哪些数据被消费的频率较高？



- 关联数据消费的动机

- 关联数据消费中存在的障碍



哪些数据被消费的频率较高？



- Virtual International Authority File (VIAF)
- DBpedia
- GeoNames
- id.loc.gov
- Resources the respondents convert to linked data themselves
- Getty's Art and Architecture Thesaurus
- FAST Linked Data
- WorldCat.org
- data.bnf.fr
- Deutsche National Bibliothek Linked Data Services



关联数据消费的动机



- 为本地用户提供更丰富的体验 (76%)
- 通过使用其他来源的关联数据来增强本地数据 (74%)
- 更有效的内部元数据管理 (48%)
- 提升本地搜索的准确度和范围 (40%)
- 探索从外部资源进行关联数据消费是否可以为本地资源提高搜索引擎优化 (SEO) (28%)
- 实验将不同类型数据合并为单个三元组库 (25%)
- 听说过关联数据，希望通过利用关联数据进行尝试 (25%)



关联数据消费中存在的障碍



- 来源数据与关联数据资源的匹配、消歧和校准
- 词汇映射
- 发布为关联数据的内容往往不可重用或缺乏URI
- 缺乏权威控制
- 数据集未更新
- RDF dumps的尺寸；利用数据之前需要理解数据的结构
- dumps数据格式的波动性
- 缺乏工具；endpoints不稳定
- 很难让其他机构在对象和概念之间进行统一；服务的可靠性
- 跨语言的术语消歧存在困难



Examples in Production



● 国家图书馆



● Networks

● 学术图书馆

● 公共图书馆

● 博物馆

● 学术项目

● 出版商



国家图书馆



- 16个国家图书馆关联数据项目或服务中的大部分将他们的书目或权威记录开放为关联数据，其中有3个国家图书馆的关联数据集进入消费排名的Top10，包括
 - id.loc.gov (Library of Congress)
 - data.bnf.fr (Bibliothèque nationale de France)
 - DNB Linked Data Service (German National Library)

Home

Getting

Welcome

The BNB Linked
open data and
SPARQL editor
to enter a plain

The British Na
Republic of Ire
and has recent
books, already
as well as mul

Flint SPARQL Editor 1.0.4

New

Edit

View

Help



Dataset

<http://bnb.data.bl.uk/>

Mode

SPARQL 1.0+ (5st)

Output

SPARQL-XML

Submit

Query 1

```

14 PREFIX rdfs: <http://www.w3.org/2000/01/
15 PREFIX skos: <http://www.w3.org/2004/02/
16 PREFIX xsd: <http://www.w3.org/2001/XMLSchema
17 PREFIX void: <http://rdfs.org/ns/void#>
18 PREFIX umbel: <http://umbel.org/umbel#>

```

```

21 SELECT ?author ?name WHERE {

```

```

23   ?event a bio:Birth;

```

```

25     bio:date "1945"^^<http://www.w3.org/
26
27

```

SPARQL

Properties

Classes

Prefixes

```
PREFIX bibo: <http://purl.org/ontology/bibo/>
```

```
PREFIX bio: <http://purl.org/vocab/bio/0.1/>
```

```
PREFIX blt: <http://www.bl.uk/schemas/bibliographic/bltern
```

```
PREFIX dct: <http://purl.org/dc/terms/>
```

```
PREFIX event: <http://purl.org/NET/c4dm/event.owl#>
```

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos
```

```
PREFIX isbd: <http://iflastandards.info/ns/isbd/elements/>
```

```
PREFIX org: <http://www.w3.org/ns/org#>
```

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
```

```
PREFIX rdau: <http://rdaregistry.info/Elements/u/>
```

Line: 1; Position: 1; Query is valid

Query Results

Visual Results Mode

author

name

<http://bnb.data.bl.uk/id/person/LudwigStephen1945->

Stephen Ludwig

<http://bnb.data.bl.uk/id/person/LudwigStephen1945->

Stephen Ludwig



Johann Wolfgang von Goethe

Schriftsteller, Publizist, Politiker, Jurist, Naturwissenschaftler, Theaterintendant, Maler, Zeichner
Born: 28. August 1749, [Frankfurt am Main](#)
Died: 22. März 1832, [Weimar](#)

Objects:



[All Pictures \(1756\)](#) | [All Videos \(6\)](#) | [All Audios \(86\)](#)

Involved in:

<http://www.dnb.de/EN/entityfacts>

- **Goethe's sämtliche Werke**
Goethe, Johann Wolfgang(1749-08-28 - 1832-03-22). - Stuttgart, 1858
- **Goethe's sämtliche Werke**
Goethe, Johann Wolfgang(1749-08-28 - 1832-03-22). - Stuttgart, 1858
- **Goethe's sämtliche Werke**
Goethe, Johann Wolfgang(1749-08-28 - 1832-03-22). - Stuttgart, 1858
- **Schriften. 6**
Goethe, Johann Wolfgang von. - Leipzig : Göschen, (1790)

[All Objects \(2322\)](#)

Subject of:

- **Johann Wolfgang von Goethe**
Bis 1907, Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, N 1/78 T 1 Nachlass Robert Arnaud (1885-1945)
- **Sämtliche Werke : in vierzig Bänden. 38, Der Farbenlehre polemischer Theil**
Goethe, Johann Wolfgang von. - Stuttgart [u.a.] : Cotta, 1840
- **Goethe, der Pazifist : zwischen Kriegsfurcht und Friedenshoffnung**
Rothe, Wolfgang. - Göttingen : Vandenhoeck & Ruprecht, 1998
- **Goethes Lyrik in russischer Übersetzung : V. A. Žukovskij und F. I. Tjutčev als bedeutendste Goethe-Übersetzer der russischen Romantik**
Kahlenborn, Ulrike. - München : Sagner, 1985

[All Objects \(208\)](#)



Johann Wolfgang von Goethe
Source: [Wikimedia Commons](#)

Information regarding the license status of embedded media files (e.g., images or videos) may be retrieved by clicking the respective file.

Search:

"Johann Wolfgang von Goethe" in the DDB

External links:

- [Wikipedia \(Deutsch\)](#)
- [Bibliothèque nationale de France](#)
- [NACO Authority File](#)
- [Virtual International Authority File \(VIAF\)](#)
- [Neue Deutsche Biographie \(NDB\)](#)
- [Allgemeine Deutsche Biographie \(ADB\)](#)
- [Historisches Lexikon der Schweiz \(HLS\)](#)
- [Digitaler Portraitindex der druckgraphischen Bildnisse der Frühen Neuzeit](#)
- [Wikisource](#)
- [Wikipedia \(English\)](#)



Networks



- Digital Public Library of America



- Europeana

- North Rhine-Westphalian Library
Service Center

- OCLC



europeana

Add a search term



Explore 52,307,392 artworks, artefacts, books, videos and sounds from across Europe.

Winter Landscape with Ice Skaters, Hendrick Avercamp
Rijksmuseum
© Public Domain

THEMATIC COLLECTIONS
MUSIC COLLECTIONS

THEMATIC COLLECTIONS
ART HISTORY COLLECTIONS

EXHIBITION
CAKE? CAKE!

WHAT'S NEW
HELP US IMPROVE EUROPEANA - SURVEY

FEATURED SITE
EUROPEANA 1914-1918 - UNTOLD STORIES & OFFICIAL HISTORIES OF WORLD WAR I



WHAT'S NEW
WORLD WAR I: A BATTLE OF PERSPECTIVES.



学术图书馆



- 大多数学术图书馆受访者的关联数据项目本质上是实验性的，比如北卡罗来纳州立大学的组织名称关联数据库。



公共图书馆



- 在调查中，很少有公共图书馆对调查作出回应，只有两个图书馆有项目或服务进行中。
- Arapahoe Library District
- Oslo Public Library



博物馆



- 很少有博物馆对调查作出回应



- British Museum's Semantic Web Collection



学术项目



- Dalhousie University's Institute for Big Data Analytics
- Pratt Institute's Linked Jazz project
- Nomisma



Linked Jazz



- 一项探索关联数据在数字人文领域中应用的实验项目,主要关注如何利用关联数据提升文化遗产文献的可视性,即通过爵士历史的数字档案揭示音乐家之间的关系以及他们在爵士领域的社交网络。
 - Analyzer
 - Name Mapping and Curator Tool
 - Ecco!
 - Linked Jazz 52nd Street
 - Linked Jazz Network Visualization Tool

LINKED JAZZ





工作思路



- 从大量包含爵士信息的爵士档案中抽取访谈脚本,识别脚本中包含的爵士艺术家名称并构建爵士名称目录,再与名称规范文档进行映射,对脚本中所提及的名称进行标注,然后让用户利用众包工具分析、标注艺术家之间的关系。

摘自：关联数据在数字人文领域中的应用剖析——以关联爵士项目为例



52大街众包工具

LINKED JAZZ



*Revealing the relationships
of the jazz community*

Let's get started!

选择音乐家、挖掘对话、揭示关系、生成网络

efined



choose a musician



explore the
conversation



uncover relationships



grow our network

Buddy DeFranco



Sam Rivers



Louie Bellson



John Levy



Danny Barker



We need your feedback, tell u



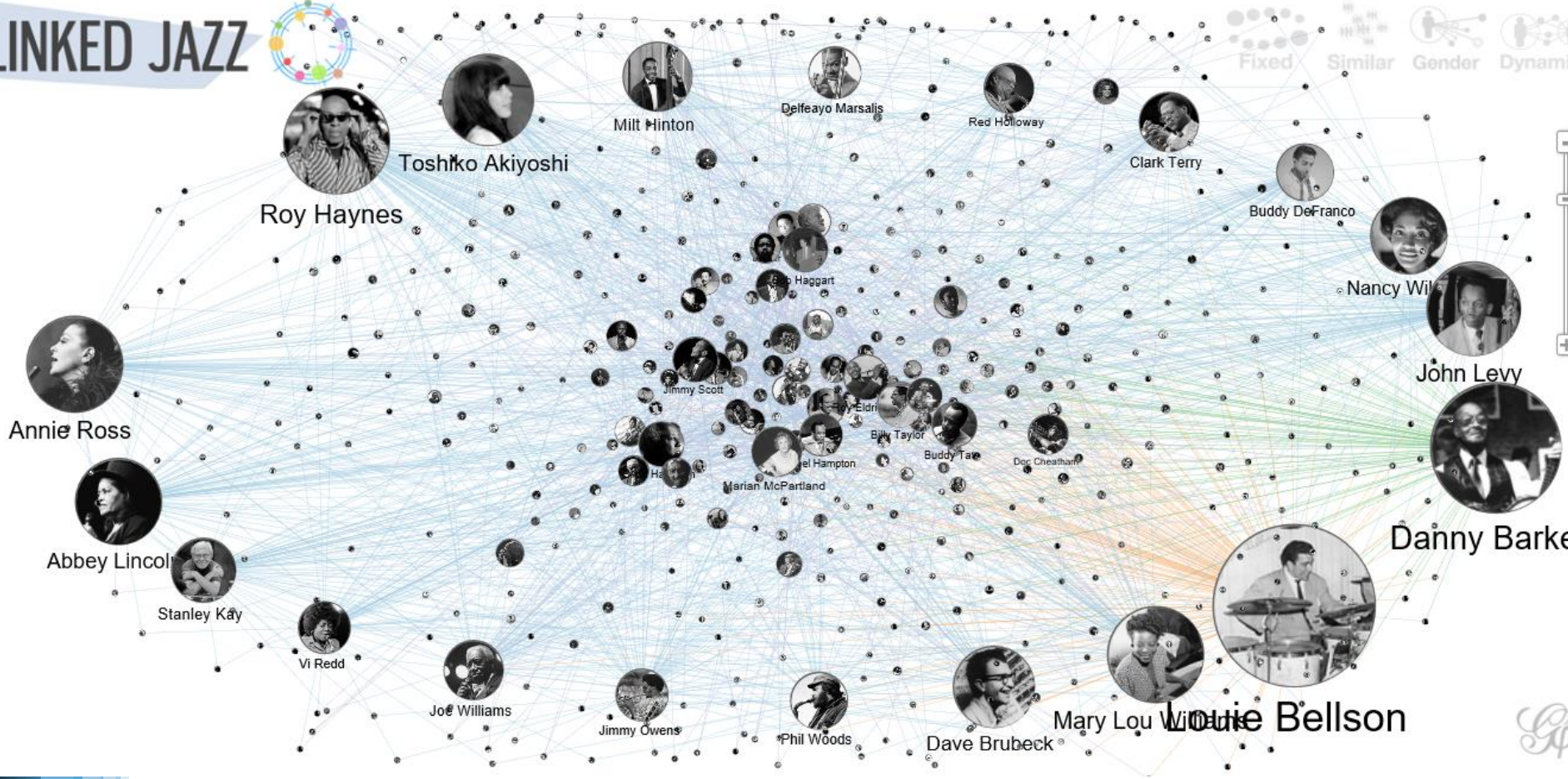
社交网络可视化



LINKED JAZZ



Fixed Similar Gender Dynamic



Nomisma

Nomisma.org is a collaborative project to provide stable digital representations of numismatic concepts according to the principles of [Linked Open Data](#). These take the form of http URIs that also provide access to reusable information about those concepts, along with links to other

Data Export

Nomisma Linked Data



[Quantitative Analysis](#)

ephesus (nmo:Mint)

skos:prefLabel Efese (*af*), Ἔφεσος (*am*), أفسس (*ar*), Efes (*az*), Горад Эфес (*be*), Εφεσ (*bg*), Efesos (*br*), Efes (*bs*), Efes (*ca*), Efez (*cs*), Effesus (*cy*), Efesos (*da*), Ephesos (*de*), Ἐφεσος (*el*), Ephesus (*en*), Efeso (*eo*), Éfeso (*es*), Efeso (*eu*), أفسس (*fa*), Efesos (*fi*), Efesus (*fo*), Ephèse (*fr*), Éfeso (*gl*), οἶοςκ (*he*), Efez (*hr*), Ephesosz (*hu*), Եփեսոս (*hy*), Ephesos (*id*), Efesos (*is*), Efeso

Export

Linked Data [GitHub File](#) [RDF/XML](#) [RDF/TTL](#) [JSON-LD](#)

Geographic Data [KML](#) [geoJSON \(mints\)](#) [geoJSON \(hoards\)](#) [geoJSON \(finds\)](#)



Type

Type Series

Example

Price 1877

Price (1991)

Mint [Ephesus](#)
Denomination [Drachma](#)
Date 300 B.C. - 300 B.C.



Price 1876

Price (1991)

Mint [Ephesus](#)
Denomination [Tetradrachm](#)
Date 300 B.C. - 300 B.C.



Price 1885

Price (1991)

Mint [Ephesus](#)
Denomination [Tetradrachm](#)
Date 200 B.C. - 190 B.C.



Price 1875

Price (1991)

Mint [Ephesus](#)
Denomination [Stater](#)
Date 300 B.C. - 300 B.C.





出版商



- Springer是唯一在调查中予以回应的出版商。
- 它正在将有关学术会议的数据制作为关联数据，使出版物、作者、专题和会议的信息更易于探索，并便于分析作者、研究机构和会议的产出和影响力。



华人家谱总目——上海图书馆家谱知识服务平台Beta版



首页 时空 登录 注册

- A
- B
- C
- D
- E
- F
- G
- H
- J
- K
- L
- M
- N
- O
- P
- Q
- R
- S
- T
- W
- X
- Y
- Z

A
艾
ai
1863
先祖名人 20
家谱文献 14

安
an
1659
先祖名人 18
家谱文献 14

阿
a
1982
先祖名人 5
家谱文献 3

敖
ao
1780
先祖名人 6
家谱文献 9

爱新觉罗
aixinjueluo
1660
先祖名人 94
家谱文献 792

B
卜
bu
1799
先祖名人
家谱文献

卞
bian
1847
先祖名人
家谱文献

巴
ba
1820
先祖名人
家谱文献

布
bu
先祖名人
家谱文献

先祖名人

- 敖奇勤
- 敖彦直
- 敖利
- 敖敏
- 敖陶孫

敖

ao

东汉《风俗通义》收。上古时期敖氏名人罕见，中世以后逐渐出现。宋代有敖陶孙，福清人，为庆元进士，元代有敖继公，长乐人。《万姓统谱》作敖继翁，此从《宋元学案》。宋《百家姓》收为第375姓。《中国人名大辞典》有敖氏9例。

家谱

- 吴西勒氏谱书不分卷（遼寧省丹東市鳳城市）

敖奇勤 先祖

⊕ 吴西勒氏谱书 ⊙ 凤城

- 责任者：吴融
- 摘要：吴西勒氏又译敖奇勤，系巴爾虎蒙古族。始迁祖達力，清代人。1930年据铅印本複印，书名据谱序题。
- 撰修时间：1930年
- 版本：复印本
- 数量：一册
- 馆藏信息



When is Margaret Thatcher born?

Who developed Skype?

What is the largest city in Australia?



Answers

1925-10-13

Skype Limited

Sydney

1 Typed question generator

When → Date/Time
Who → Person/Organization
What → Thing

Margaret Thatcher → Person
Skype → Company
Australia → Place

Expected Answer Type identification

NE recognition and identification

2 Property identifier

[Date] born [Person] →
[Agent] developed [Organization] →
[Thing] largest city [Place] →

[birthDate] relation ←
[developer] relation ←
[largestCity] relation ←

WikiFramework

Pattern matcher Pattern repository

3 Query selector

```
select distinct * where {
  <http://dbpedia.org/resource/Margaret_Thatcher> <http://dbpedia.org/ontology/birthDate> ?v .
} limit 20
```

```
select distinct * where {
  ?v <http://dbpedia.org/ontology/product> <http://dbpedia.org/resource/Skype> .
  ?v rdfs:type <http://dbpedia.org/ontology/Organisation> .
  OPTIONAL { ?v <http://www.w3.org/2000/01/rdf-schema#label> ?l filter (lang(?l)="en") }
} limit 20
```

```
select distinct * where {
  <http://dbpedia.org/resource/Australia> <http://dbpedia.org/ontology/largestCity> ?v .
  OPTIONAL { ?v <http://www.w3.org/2000/01/rdf-schema#label> ?l filter (lang(?l)="en") }
} limit 20
```

Query Manager

DBpedia

DBpedia en français

DBpedia Deutsch

WikiFramework

DBpedia relations

[birthDate]	1955-02-24	birthDate	Steve Jobs
	1685-03-31	birthDate	J. S. Bach
	1961-08-04	birthDate	Barack Obama

[largestCity]	Rome	largestCity	Italy
	São Paulo	largestCity	Brasil
	Paris	largestCity	France

Wikipedia pages

Early life and education

Steven Paul Jobs was born in San Francisco on February 24, 1955 to two university students, Joanna Carole Schieble, of Swiss Catholic descent, and Syrian-born Abdulfattah "John" Jandali (Arabic: جندلي), who were both unmarried at the time.^[32] Jandali, who was teaching in Wisconsin when Steve was born, said he had no choice but to put the baby up for adoption because his girlfriend's family objected to their relationship.^[33]

Relation matching

Steven Paul Jobs	→	Steve Jobs	Steve	Jobs
		X	X	✓
February 24, 1955	→	24th of February, 1955	February 24, 1955	
		X	✓	

Pattern extraction

[birthDate]
[Person] was born in San Francisco on [Date]
[Person] was born in Eisenach, Saxe-Eisenach, on [Date]
[Person] was born on [Date]

Pattern matching

[Date] born [Person] → [birthDate]



调查总结



- 这两次调查可以被认为是一个仍在发展中的关联数据环境的快照。由于调查分析只限于哪些有反馈的受访对象，而它们又主要来自图书馆领域，因此通过调查可以对关联数据的发展有一个局部看法。所描述的许多项目本质上是实验性质的。
- 关联数据项目侧重于数据。与当前的体系结构相比，开发能够提供新的或更有价值的关联数据服务将需要开发利用多个关联数据源和集成来自不同领域的数据集的应用程序。一些人指出，关联数据使他们能够克服自身机构内部的数据孤岛。
- 所描述的许多项目和服务是多机构协作的。这种合作有助于缓解机构层面资源不足的问题，同时提供了一种在本地知识匮乏的情况下扩大学习的方法。



建议



- Focus on what you want to achieve, not technical stuff.
 - Model data that solves your use cases.
 - Strive for long-term data reconciliation and consolidation.
- Add distinctive value: Build on what you have that others don't.
 - Pick a problem you can solve.
 - Involve your institution/community.
- Have a good understanding of linked data structure, available ontologies and your own data.
 - Consume your own published data.
 - Consider legal issues from the beginning.
 - Read as widely as possible and consult community experts.



如何成为Linked Data Librarian



- 第1步：阅读和理解RDF数据
- 第2步：验证RDF数据，转换格式
- 第3步：学习和探索SPARQL查询
- 第4步：探索可链接数据接口和API
- 第5步：数据再处理/可视化

**曾蕾、林夏. 动手体验关联数据与开放数据,
数字图书馆前沿问题高级研讨班, 2015.7**



构建自己的FOAF文件

FOAF-a-Matic

[[Japanese](#)] [[French](#)] [[Spanish](#)] [[Danish](#)] [[Swedish](#)] [[Greek](#)] [[German](#)] [[Italian](#)] [[Korean](#)] [[Trad. Chinese](#)]

作者：[Leigh Dodds](#) 譯者：[李士傑](#)

Introduction

FOAF-a-matic 是一個簡單的 Javascript 應用程式。它讓你能夠建立關於你自己的 FOAF ("Friend-of-A-Friend") 檔案。你可以參考 Edd Dumbill 的文章：[「XML 观察：使用 XML 和 RDF 找到朋友」](#)，拜訪[RDFWeb 的 FOAF 首頁](#)，以及[FOAF 的詞彙描述 \(FOAF vocabulary description \)](#)。

簡單地說，FOAF 是一種用 XML 跟 RDF 描述你自己的方式：你的名字、電子郵件信箱，以及你的朋友們。這種方式使其可以被軟體處理，舉例來說可為搜尋引擎所利用，自動化地發現社群與你的資訊。FOAF 具有帶動線上社群許多有趣發展的潛力。Ben Hammersely 的英國衛報文章 "[Click to the Clique](#)" 提供 關於這些想法的更進一步探討。

FOAF-a-Matic 被用來提供一個簡單、快速的方式，產生你自己的 FOAF 檔案。最基本你至少需要提供你的 名字與 email 地址，並且你可以增加任何你想加入的朋友。把朋友加入 FOAF 檔案中是很重要的（當然想 寫下多少朋友看你高興啦），因為當 FOAF 彙整器索引你的 FOAF 描述時，他們將會自動把你們一起綁成 一群個體之間的網絡關係。

備註：在本頁你所輸入的所有資訊，都不會被存下來或被使用。全部的處理過程全部都是在客戶端進行，以確保你的隱私權。

如果你對這個應用有任何的意見，或者關於 FOAF 的進一步問題，為什麼不參加[RDFWeb-dev 的通訊論壇](#)？

最新消息：我目前正在撰寫一套建立與管理個人 FOAF 資料的桌面應用程式，[FOAF-a-Matic Mark 2 a](#)

表單

個人資訊



個人資料

某些你個人的資訊，以及其他人如何和你聯絡。

稱謂 (Mr, Mrs, Dr, etc)	<input type="text" value="Dr"/>
名字	<input type="text" value="曉娟"/>
姓 (Family/Given)	<input type="text" value="刘"/>
暱稱	<input type="text"/>
電子郵件地址	<input type="text" value="lxj_2007@bnu.edu.cn"/>
首頁	<input type="text"/>
你的照片	<input type="text"/>
電話號碼	<input type="text" value="010-58805387"/>

工作資訊

關於你工作所在的資訊。

工作所在首頁	<input type="text" value="http://www.sg.bnu.edu.cn"/>
在你的工作脈絡中的介紹網頁	<input type="text"/>

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:admin="http://webns.net/mvcb/">
<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="#me"/>
  <foaf:primaryTopic rdf:resource="#me"/>
  <admin:generatorAgent rdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
  <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>
<foaf:Person rdf:ID="me">
<foaf:name>曉娟 刘</foaf:name>
<foaf:title>Dr</foaf:title>
<foaf:givenname>曉娟</foaf:givenname>
<foaf:family_name>刘</foaf:family_name>
<foaf:mbox_sha1sum>685f7296d156e40459dcaecd693baabf0bbd9a84</foaf:mbox_sha1sum>
<foaf:phone rdf:resource="tel:010-58805387"/>
<foaf:workplaceHomepage rdf:resource="http://www.sg.bnu.edu.cn"/>
<foaf:schoolHomepage rdf:resource="http://www.bnu.edu.cn"/>
```

學校

你是哪裡畢業的？

學校首頁	<input type="text" value="http://www.bnu.edu.cn"/>
------	--

你所認識的朋友

告訴 FOAF-a-matic 你所認識的朋友。按下「加入更多朋友」，可以增加空白欄位，輸入更多你所認識朋友的資訊。如果你的朋友已經有了 FOAF 檔案，請把該檔案連結的資訊 (URL) 輸入在「SeeAlso」的欄位中。

Friend-- Name	<input type="text" value="张三"/>	Email	<input type="text" value="bnulxj@163.com"/>	See Also	<input type="text"/>
Friend-- Name	<input type="text"/>	Email	<input type="text"/>	See Also	<input type="text"/>
Friend-- Name	<input type="text"/>	Email	<input type="text"/>	See Also	<input type="text"/>

加入朋友



本网站是上海图书馆的开放数据平台，将陆续以关联开放数据（Linked Open Data）的方式向互联网公开发布各种术语词表、规范档、书目数据以及各种数据清洗和转换工具等。提供各种数据消费接口供开发人员调用，和重用。调用接口前，请注册API Key。

人名规范库

More...

华人姓氏列表

More...

中国历史纪年表

More...

地理名词表

More...

RDB2RDF

More...

ETL

More...

家谱本体

More...

手稿及档案本体

More...

古籍本体

More...

任务

数据采集 >

数据加工 >

数据转换 >

模板

过滤模板

抽取模板

RDF映射模板

资源

Iri列表



Virtuoso SPARQL Query Editor

出生在西安的名人有哪些？但是这些人不在1970年后出生的人当中

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#) | [RDF views](#) | [iSPARQL](#)

Default Data Set Name (Graph IRI)

Query Text

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?name
WHERE {
  ?a <http://dbpedia.org/property/name> ?name.
  ?a <http://dbpedia.org/property/birthPlace> <http://dbpedia.org/resource/Xi'an>.
  filter not exists{
    ?a <http://dbpedia.org/ontology/birthYear> ?date.
    filter(?date>"1970-01-01"^^xsd:date )
  }
}
```

name
"Li Yang"^^<http://www.w3.org/1999/02/22-rdf-syntax-ns#langString>
"You Yong"^^<http://www.w3.org/1999/02/22-rdf-syntax-ns#langString>
"Zhao Fei"^^<http://www.w3.org/1999/02/22-rdf-syntax-ns#langString>
"Huang Jianxin"^^<http://www.w3.org/1999/02/22-rdf-syntax-ns#langString>
"Gu Changwei"^^<http://www.w3.org/1999/02/22-rdf-syntax-ns#langString>

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout:

milliseconds (values less than 1000 are ignored)

Options:

- Strict checking of void variables
- Log debug info at the end of output (has no effect on some queries and output formats)
- Generate SPARQL compilation report (instead of executing the query)

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query

Reset



Link

36 results

acm - ACM Classification Ontology

<http://www.rkbexplorer.com/ontologies/acm>

This ontology is a representation of The ACM Computing Classification System [1998] @en

arch - Archival collections ontology

<http://purl.org/archival/vocab/arch>

An RDF vocabulary for describing archival collections and the names associated with them @en

bf - BIBFRAME Vocabulary

<http://bibframe.org/>

The BIBFRAME n (electronic). The m title authority reco must have a work

@en

Metadata

URI

<http://bibframe.org/vocab>

1 Incoming Links

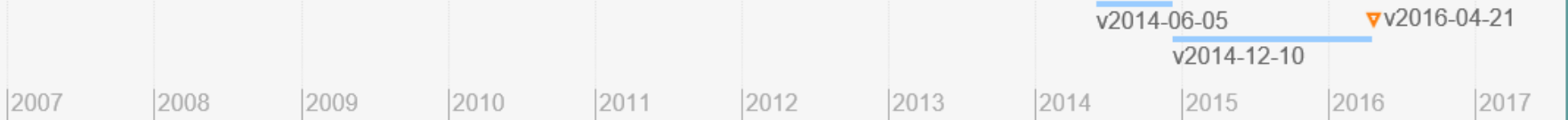
2 Outgoing Links



Metadata Extends Specializes Generalizes Has Equivalences with Has Disjunction with Imports

Vocabulary Version History

Timeline © SIMILE



simple search?

CHOOSE A DATASET

INSERT A KEYWORD

start >>

insert URI?

PASTE A RESOURCE ADDRESS

start >>

browse your data?

PASTE SOME RDF DATA

PASTE AN ADDRESS

start >>



Live on LodLive

CHOOSE AN ENDPOINT

start >>

www.wikidata.org ?

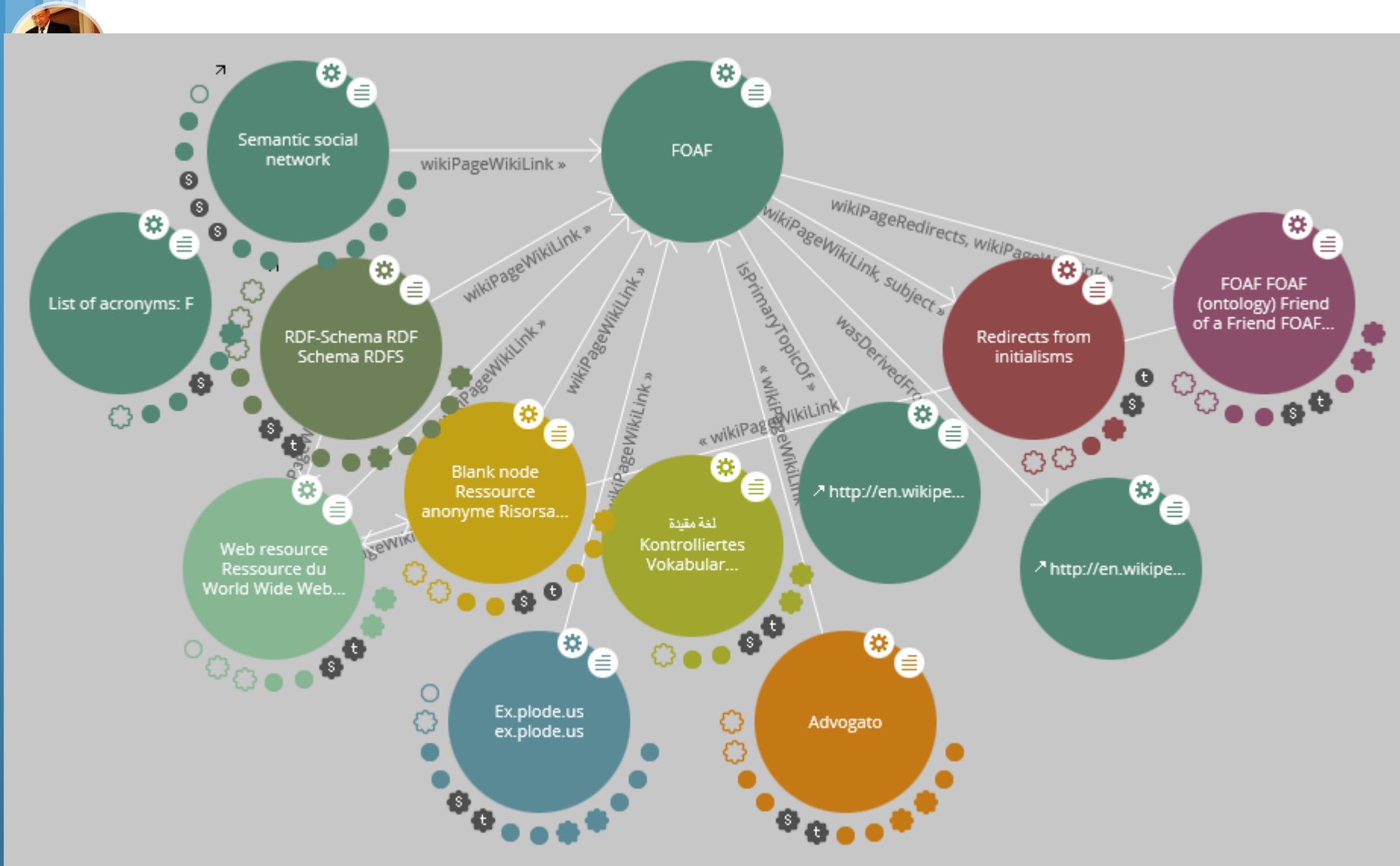
RESOURCE

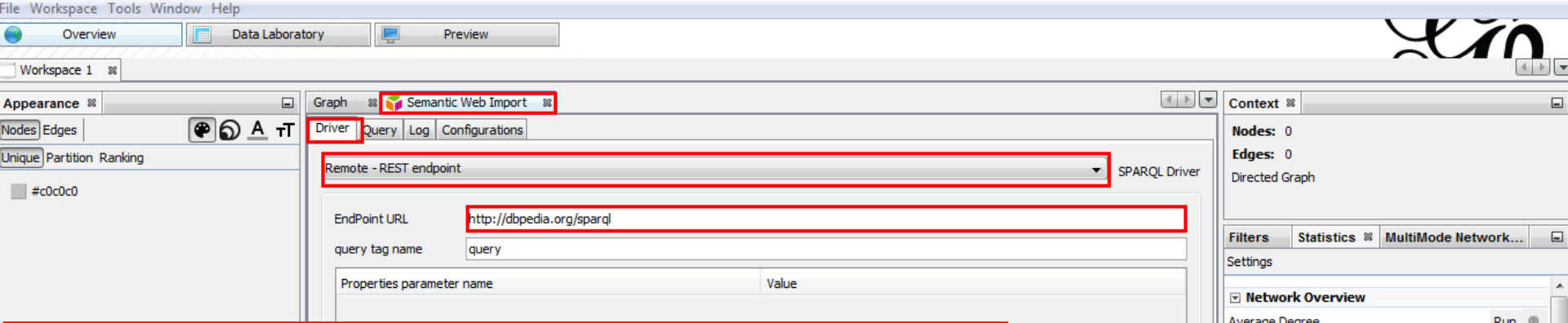
data.ontotext.com/resource/

RESOURCE

start >>

LodLive: FOAF expand related resources

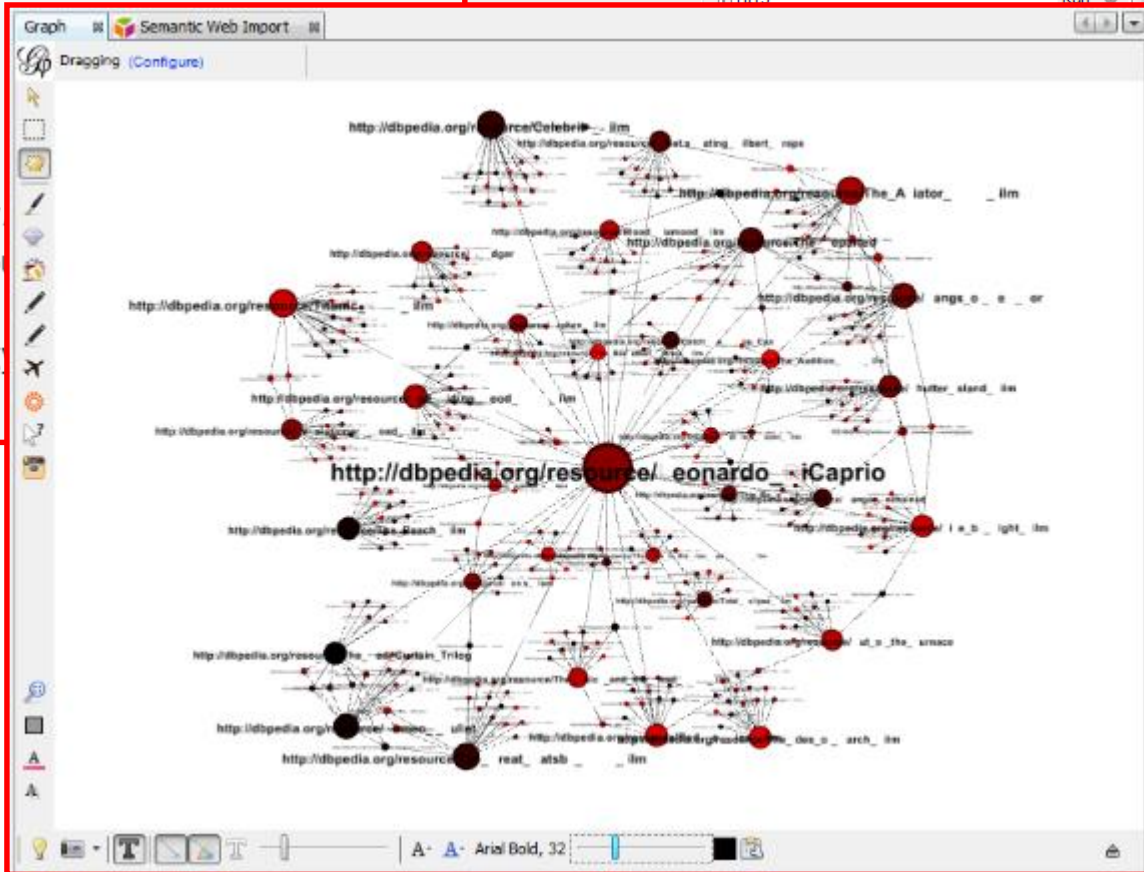




```

CONSTRUCT{
  ?person <http://gephi.org/type> "person".
  ?movie <http://gephi.org/type> "movie".
  ?movie <http://toto/link> ?person .
}
WHERE
{
  ?movie a <http://dbpedia.org/ontology
  ?movie ?rel1 <http://dbpedia.org/reso
  ?movie ?rel2 ?person.
  ?person a <http://dbpedia.org/ontolog
  }

```



Gephi semantic web import



相关文献

ALA American Library Association

LIBRARY LINKED DATA

EARLY ACTIVITY & DEVELOPMENT

Erik T. Mitchell

Library Technology Reports
Expert Guides to Library Systems and Services

JAN 2016
Vol. 52 / No. 1
ISSN 0044-2580

Prepared 14 March, 2017

MacKenzie Smith



Article

Current state of Linked Data in digital libraries

María Hallo
National Polytechnic School, Ecuador

Sergio Luján-Mora
University of Alicante, Spain

Alejandro Maté
University of Alicante, Spain

Juan Trujillo
University of Alicante, Spain

Journal of Information Science
2016, Vol. 42(2) 117–127
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0165551515594729
jis.sagepub.com

谢谢！