

从数字化到数据化：编目的未来之路

胡小菁

华东师范大学图书馆

CALIS中西文编目员研讨会 2018.11.26 厦门

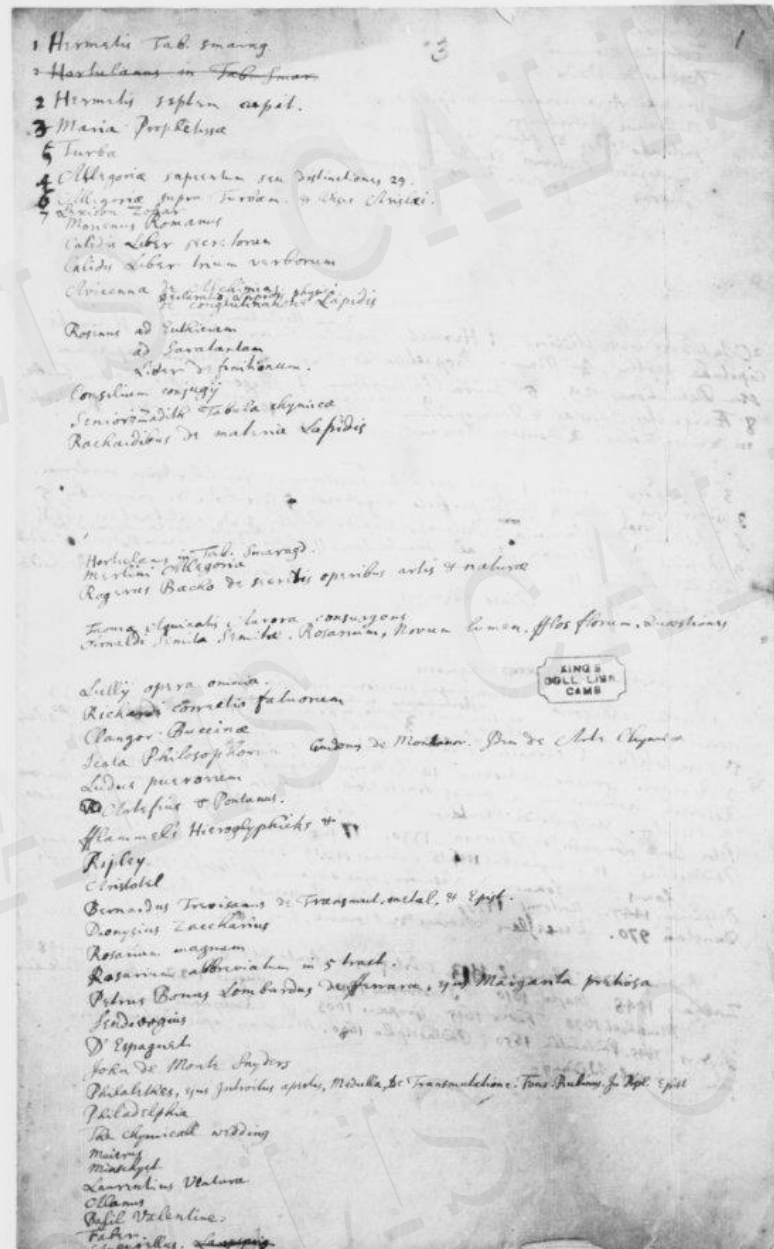
数字化 vs 数据化

- 人读

- 扫描
- 文本/OCR
 - 人类可检索

- 机读

- 标记
 - 机器可处理/操作



1 Hermetis Tabula Smaragdina

2 Hermetis septem capit.

3 Maria Prophetissa

5 Turba

4 Allegoriae sapientum seu distinctiones 29.

6 Allegoriae supra Turbam. & Visio Arislaei.

7 Lexicon Zohar

Morienus Romanus

Calidis Liber secretorum

Calidis Liber trium verborum

Avicenna de Alchimia

Declaratio lapidis physici

de conglutinatione Lapidis

Rosinus ad Euthicium

ad Saratantam

Liber definitionum.

数字化

数据化

Nike, Inc.

Footwear manufacturing company

Nike, Inc. is an American multinational corporation that is engaged in the design, development, manufacturing and worldwide marketing and sales of footwear, apparel, equipment, accessories and services. [Wikipedia](#)



Stock price: [NKE](#) (NYSE) \$103.80 -0.08 (-0.08%)

Jun 12, 7:43 PM EDT - Disclaimer

Customer service: +1 800-806-6453

Founded: January 25, 1964

CEO: [Mark Parker](#)

Headquarters: [Beaverton, OR](#)

Founders: [Bill Bowerman](#), [Phil Knight](#)

Profiles



Facebook



YouTube



Twitter



Instagram



LinkedIn

People also search for

[View 15+ more](#)



Adidas



Foot Locker



Puma SE



New Balance



Under Armour

```
<script type="application/ld+json">
{
  "@context" : "http://schema.org",
  "@type" : "Organization",
  "url" : "http://www.your-company-site.com",
  "contactPoint" : [
    {
      "@type" : "ContactPoint",
      "telephone" : "+1-401-555-1212",
      "contactType" : "customer service"
    }
  ]
}
```

```
<script type="application/ld+json">
{
  "@context" : "http://schema.org",
  "@type" : "Organization",
  "name" : "Your Organization Name",
  "url" : "http://www.your-site.com",
  "sameAs" : [
    "http://www.facebook.com/your-profile",
    "http://www.twitter.com/yourProfile",
    "http://plus.google.com/your_profile"
  ]
}
```

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "Organization",
  "url": "http://www.example.com",
  "logo": "http://www.example.com/images/logo.png"
}
```

从卡片目录到机读目录

图书馆目录：数据化先行者

```
200 1 红楼梦|Ahong lou meng|f(清) 曹雪芹, 高鹗著
210 北京|c中华书局|d2018
215 5册 (13, 1837页)|d21cm
225 2 中华大字经典|Azhong hua da zi jing dian
410 0 |12001 |a中华大字经典
606 0 章回小说|Azhang hui xiao shuo|y中国|z清代
606 0 古典小说|Agu dian xiao shuo|y中国|z清代
606 0 长篇小说|Achang pian xiao shuo|y中国|z清代
690 I242.47|v5
701 0 曹雪芹,|Acao xue qin|f约1715-1763|4著
701 0 高鹗,|Agao e|f约1738-约1815|4著
```

机读 or 人读

遗留的标点符号

- MARC21格式

245 10\$aPour les valeurs bourgeoises /\$cpar Georges
Hourdin. 

- UNIMARC/CNMARC检索点

701 #0\$a曹雪芹,\$f约1715-1763\$4著



MARC21准备弃用ISBD标识符

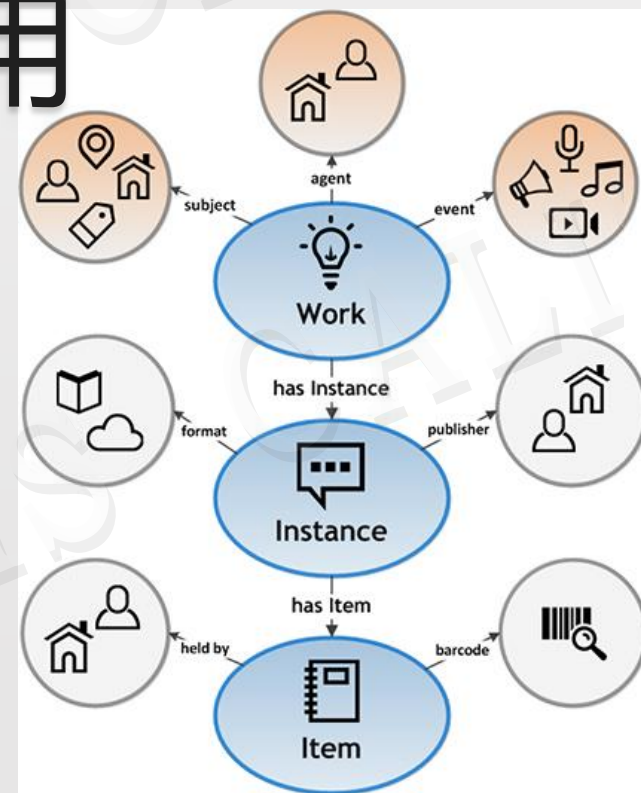
- 《PCC ISBD和MARC任务组（2016年修订最终报告）》
- 测试使用有限ISBD标识符（2018.2.6-7.1）
 - 停止使用[字段末尾句点](#)、停止使用与MARC 21[编码相对应的分隔标点](#)

参见：[MARC21准备弃用ISBD标点](#)（2018-3-3）

取代MARC，美国国会图书馆，2011-

BIBFRAME：面向关联数据应用

BIBFRAME 2.0模型 (2016-4-21)



BIBFRAME进展

2017.5：全部MARC目录转换到BIBFRAME 2.0

- 1700+万**书目记录**转换为BF作品、实例和单件
- 120万**统一题名**规范记录转换为BF作品
- BIBFRAME数据库**每日更新**：由来自MARC书目和规范记录的新作品/实例/单件描述

2017.6- 第2阶段试验

- 60名编目员，BIBFRAME原编

2018.6：BIBFRAME**数据集**

- 作品1900万+实例2400万+单件2260万=43亿三元组
- 作品、实例提供**批量下载**

2018.11：BIBFRAME**编辑器**代码、编辑器演示网站新发布

The new revolution: from record to **entity**

Full Record	MARC Tags	Where to Request
000 02384cam a2200445 a 4500		
001 15789102		
005 20100204124659.0		
008 090824s2009 enka b 000 0 eng		
010 _ _ ja 2009025024		
015 _ _ ja GBA959347 (2 bnb)		
016 7_ _ ja 015291646 (2 Uk)		
020 _ _ ja 9780521519748 (hbk.)		
020 _ _ ja 0521519748 (hbk.)		
020 _ _ ja 9780521732565 (pbk.)		
020 _ _ ja 0521732508 (pbk.)		
035 _ _ ja (DCLC)ecm456716652		
040 _ _ ja DLC (c DLC (d YDXCP (f BWNKJ (g BWNK (h FDA (i UKM (j BTSTA (k CEF (l COX (m DLC		
050 00 ja PR2803.A2 (b HQ5 2009		
062 00 ja 822.3/3 (2 22		
100 1_ _ ja Shakespeare, William, fl 1564-1616.		
245 10 ja <i>As you like it</i> / (c edited by Michael Hattaway.		
250 _ _ ja Updated ed.		
260 _ _ ja Cambridge, UK : (a New York : (b Cambridge University Press, (c 2009.		
300 _ _ ja iv, 240 p. : (b ill. ; (c 24 cm.		
490 1_ _ ja New Cambridge Shakespeare		
504 _ _ ja Includes bibliographical references (p. 240).		
505 0_ _ ja Introduction. Journeys : Plays within the play. Theatrical genres : Pastoral ; Counter-pastoral ; The condition of the country ; Politics ; 'Between you and the women the play must please' ; Gender ; Nuptials ; Sources ; Date and occasion. Stage history ; Recent critical and stage interpretations -- Note on the text -- List of characters -- The play -- Textual analysis -- Appendixes : 1. An early court performance? ; 2. Excerpts from Shakespeare's principal sources, Lodge's Rosalind ; 3. The songs.		
600 10 ja Shakespeare, William, fl 1564-1616. (t <i>As you like it</i>		
650 0_ _ ja Fathers and daughters (v Drama.		
650 0_ _ ja Estates (v Drama.		
655 7_ _ ja Pastoral drama. (2 psald		
655 7_ _ ja Comedies. (2 psald		
700 1_ _ ja Hattaway, Michael.		

As you like it [print]



As you like it [on-line]

Cambridge University Press

Cambridge Press



Publisher

Cambridge Univ. Press

Shakespeare, William, 1564-1616

Шекспир, У. 1564-1616 Уильям



Agent

Saixpēr, Gouilliam, 1564-1616

As you like it



Come ti piace

Comme il vous plaira

Fathers and daughters



Subject

Padri e figlie

Pères et filles

Tiziana Possemato.
From MARC to
BIBFRAME in the
SHARE-VDE project.
BIBFRAME Update
Forum, ALA Annual
Conference 2018

不是发布为关联数据，就可以自动关联了

- 字符串

莎士比亚

Shakespeare, William

皆大欢喜

As you like it

关联的RDF

- URI/IRI

<http://id.loc.gov/rwo/agents/n78095332>

<http://viaf.org/viaf/178316948>



用 URI/IRI 强化MARC记录

.....

100 0#\$aBridget,\$cof Sweden, Saint,\$dapproximately 1303-1373.

\$0<http://id.loc.gov/authorities/names/n50044402>

240 10\$aRevelationes.\$lEnglish\$s(Searby)

\$0<http://id.loc.gov/authorities/names/n84007202>

.....

650 #0\$aVisions.**\$0**<http://id.loc.gov/authorities/subjects/sh85143882>

700 1#\$aSearby, Denis Michael,\$etranslator.

\$0<http://id.loc.gov/authorities/names/nr98021028>

.....

<http://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=13910411>

与 URI/IRI 有关的新增与重定义

MARC21的数据化改造

控制子字段\$0和\$1：捕获 URI/IRI

- \$0：规范记录控制号或标准号（省略前缀(uri)）

2016修改

100 1#\$aTrollope, Anthony,\$d1815-1882.\$0(isni)0000000121358464

- \$1：真实世界对象(RWO, Real-World Object) URI

2017新增

710 2#\$aCalifornia Poets in the Schools (Project),\$eissuing body,
\$epublisher.\$0http://id.loc.gov/authorities/names/n85319780

\$1http://id.loc.gov/rwo/agents/n85319780

2017新增

新增758字段：资源标识符

\$a标签
\$i关系信息
\$4关系（代码或URI）

100 1#\$aBeard, Henry.

245 10\$aBored of the rings :\$ba parody of J.R.R. Tolkien's **The lord of the rings** /\$cby Henry N. Beard and Douglas C. Kenney of the Harvard lampoon.

758 ##\$4<http://rdaregistry.info/Elements/m/P30135>

\$iHas work manifested:\$aBored of the Rings
\$1<http://www.wikidata.org/entity/Q1613936>

书目记录描述的资源

758 ##\$4<http://rdaregistry.info/Elements/w/P10197>

\$iParody of (work):\$aTolkien, J. R. R. (John Ronald Reuel), 1892-1973.
Lord of the rings\$0<http://id.loc.gov/authorities/names/no97079452>

书目记录的相关资源

[758 - Resource Identifier \(R\)](#)

758的作用

- 扩大MARC支持[关联数据](#)标识符.....是图书馆将关联数据构建到元数据生产过程的[迁移路径](#)的一部分
 - 将书目描述与适当的作品或其他资源实体相[关联](#)，是将MARC书目数据转换为关联数据的主要挑战。
 - 记录MARC记录中相关资源的URI，使得当MARC记录转换为关联数据表示时，URI可以作为对象包含在RDF语句中。
 - 在无URI可用的情况下，非URI标识符对于正确的数据调和（实体解析）仍然是有价值的。

[MARC PROPOSAL NO. 2017-09](#)

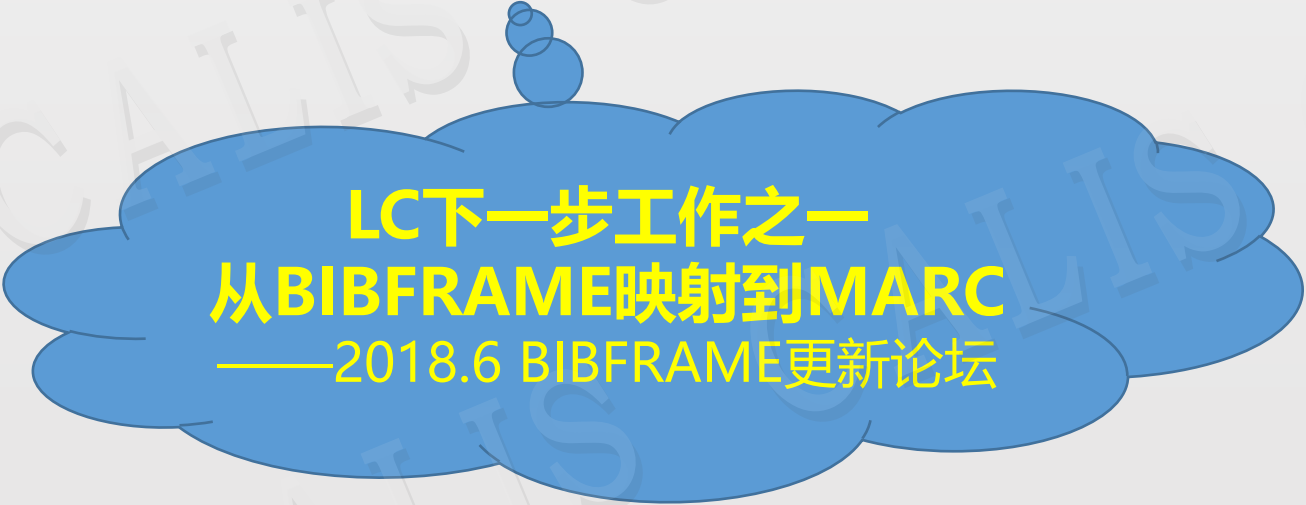
何处获取IRI/URI

- PCC 《创制和获取URI：常用词表和参考源指南》
 - 总结了40个图博档领域常用的词表和参考源的信息
- 上海图书馆开放数据平台
<http://data.library.sh.cn/>

参见：[创制和获取URI的常用词表和参考源指南](#) (2018-3-2)

MARC更新的意义

- 暂时或相当一段时间内MARC会继续使用
- 在BIBFRAME实施前，需要软件在两个格式间来回转换
 - 忽略标点可以简化加上标点、除去标点这样的编程工作
 - 具有URI对转换处理很重要



LC下一步工作之一
从BIBFRAME映射到MARC
——2018.6 BIBFRAME更新论坛

RDA Toolkit beta site, 2018.6

新RDA的变化

无核心元素 条款无编号

title proper

Definition and Scope

A chief name of a manifestation, that is, a title normally used when naming a manifestation.

An alternative title is treated as part of a title proper.

- 新RDA

- 原RDA

2.3.2 Title Proper

CORE ELEMENT

2.3.2.1 Scope

title proper▼: A chief name of a manifestation, that is, a title normally used when naming the manifestation

An alternative title is treated as part of the title proper.

4种记录方法

1. 记录非结构化描述
照录/转录
2. 记录结构化描述
规定格式，如检索点、结构化附注
3. 记录标识符
如标准号码、本地编号
4. 记录IRI（国际资源标识符）

Recording methods

The data value of an RDA element can be recorded using one or more of the following methods:

1. Recording an unstructured description →
2. Recording a structured description →
3. Recording an identifier →
4. Recording an IRI →

- 每个元素适用一种或多种记录方法

一切条款皆可选

可选=原RDA:

- 交替
 - 可选增加
 - 可选省略
 - 例外
-
- 新RDA例
 - 团体首选名称：首冠词

Initial articles

OPTION

When recording a *preferred name of corporate body*, include an initial article, if present.

Example

The Library Association

Der Wehrbeauftragte

OPTION

Omit an initial article (see [Initial articles](#)) unless the name is to be accessed under the article (e.g., a corporate name that begins with an article that is the first part of the name of a person or place).

Example

Library Association

Amis de la terre du Morvan

原RDA例：团体首选名称： 首冠词

- 原RDA
 - 不说明即选用正文条款
- 新RDA
 - 必须说明采用哪个选项

11.2.2.8 Initial Articles 2012/04

When recording a preferred name for corporate body, include an initial article, if present.

EXAMPLE

2012/04

The Library Association

Der Wehrbeauftragte

Alternative LC-PCC PS NLA PS BL PS D-A-CH SKL LAC/BAC-BAnQ KBSP

Omit an initial article (see appendix **C** RDA) unless the name is to be accessed under the article (e.g., a corporate name that begins with an article that is the first part of the name of a person or place). 2012/04

EXAMPLE

2012/04

Amis de la terre du Morvan

not Les Amis de la terre du Morvan

虚拟国际规范档

名称的规范形式
重要性下降

IRI: <http://viaf.org/viaf/96994048>

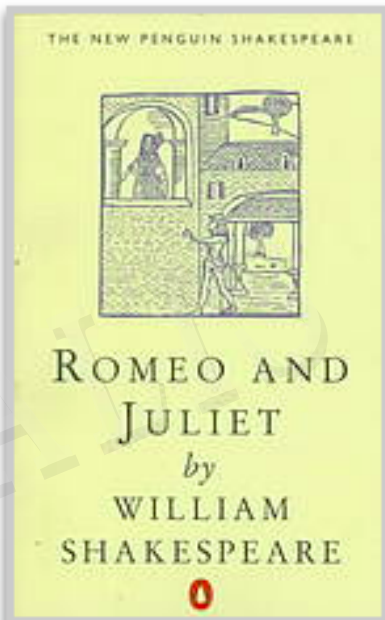
← → ↻ 🔒 https://viaf.org/viaf/96994048/#Shakespeare,_William,_1564-1616

VIAF

Virtual International Authority File

Search

Select Field: ▼ Select Index: ▼ Search Terms:



THE NEW PENGUIN SHAKESPEARE

ROMEO AND JULIET

by WILLIAM SHAKESPEARE

0

Shakespeare, William, 1564-1616. 🇺🇸 🇹🇼 🇷🇺 🇨🇭 🇸🇪 🇪🇸 🇯🇵 🇫🇷

Shakespeare, William 🇩🇰 🇳🇱 🇮🇹 🇪🇺 🇭🇷

William Shakespeare 🇮🇹

Шекспир, У. 1564-1616 Уильям 🇷🇺

Шекспир, Уильям, 1564-1616. 🇷🇺 🇧🇪

شكسپير، وليام 🇷🇺

שיקספיר, ויליאם, 1564-1616 🇷🇺

Shakespeare, W. 1564-1616 William 🇷🇺

شكسپير، وليام, 1616-1564 🇮🇷 🇪🇬

Shakespeare, William (English playwright and poet, 1564-1616)

VIAF ID: 96994048 (Personal)

Permalink: <http://viaf.org/viaf/96994048>

这本书按中文
还是西文做？

新RDA：数据多样性

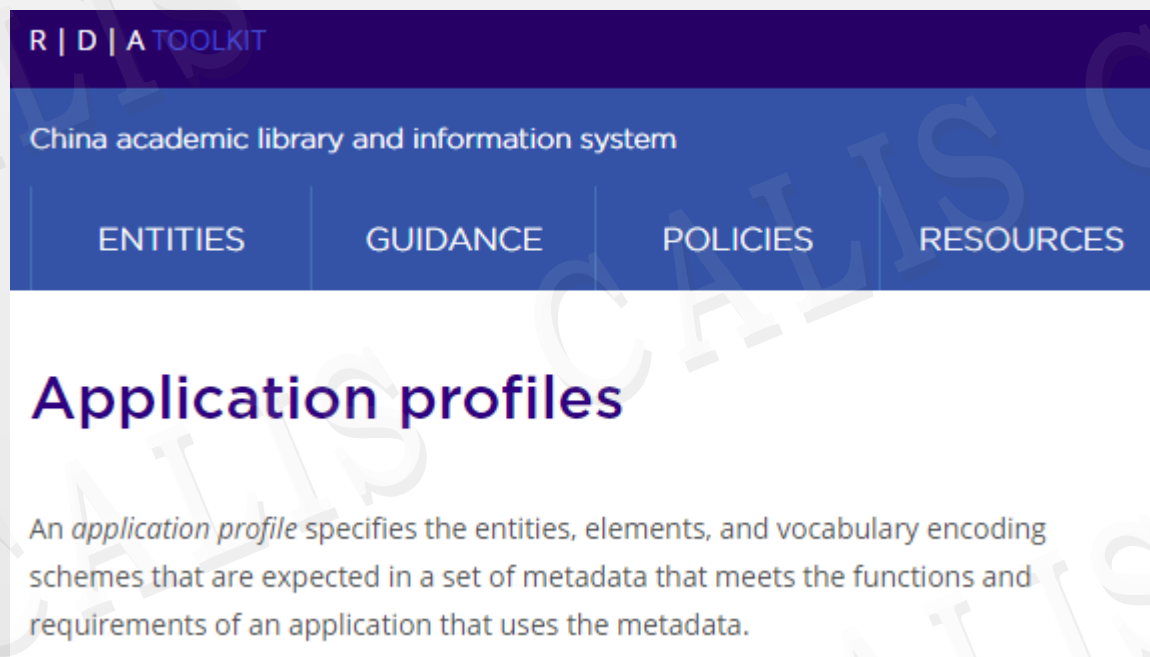
同一个RDA，不同的选项

- 可扩展性
 - 比如用本地控制词表代替工具包中提供的词表
- 选项和配置
 - 没有自上而下、一刀切的“全球书目控制”系统的全球协议
 - 必须与应用纲要一同使用，选择本地应用的元素和记录方法

——RSC现任主席Gordon Dunsire

Gordon Dunsire. Accommodating local cataloguing traditions in a global text.
Diversity of Data: RDA in the international context, National Library of Malaysia, 2018-8-23

应用纲要/应用配置文件 application profiles



- RDA本身基本上成为一本数据字典，指定元素、其定义以及在描述中提供它们的指导原则。使用RDA的机构决定使用哪些元素、遵循哪些可选准则以及用于系统创建完整描述的结构化路径，并记录在应用纲要中

——西班牙国家图书馆Kevin M. Randall

RDA不再是编目规则？

摘自RDA-L邮件组

RDA Registry

元素集

- RDA Registry (Home)
 - Elements (RDA element sets)
 - Classes
 - Agent properties
 - Expression properties
 - Item properties
 - Manifestation properties
 - Nomen properties
 - Place properties
 - Time-span properties
 - Work properties
 - RDA Entity properties
 - Meta-element properties
 - Unconstrained properties
 - RDA/ONIX Framework elements

取值词表

- Values (value vocabularies)
 - RDA values
 - RDA/ONIX Framework values
- Data (Linked data using RDA vocabularies)
 - Curie prefixes (Abbreviations for compact URIs, XML namespaces, etc.)
 - Examples (Single resource)
 - R-Balls (Multiple resources)
 - Datasets (Multiple resources)
- Tools
 - Maps (RDF maps between RDA vocabularies and other namespaces)
 - Alignments (Alignment tables for RDA vocabularies and other namespaces)
 - Profiles (Application profiles using RDA vocabularies)
 - RIMMF (RDA data editor)

注册URI/IRI

RDA Reference value vocabularies

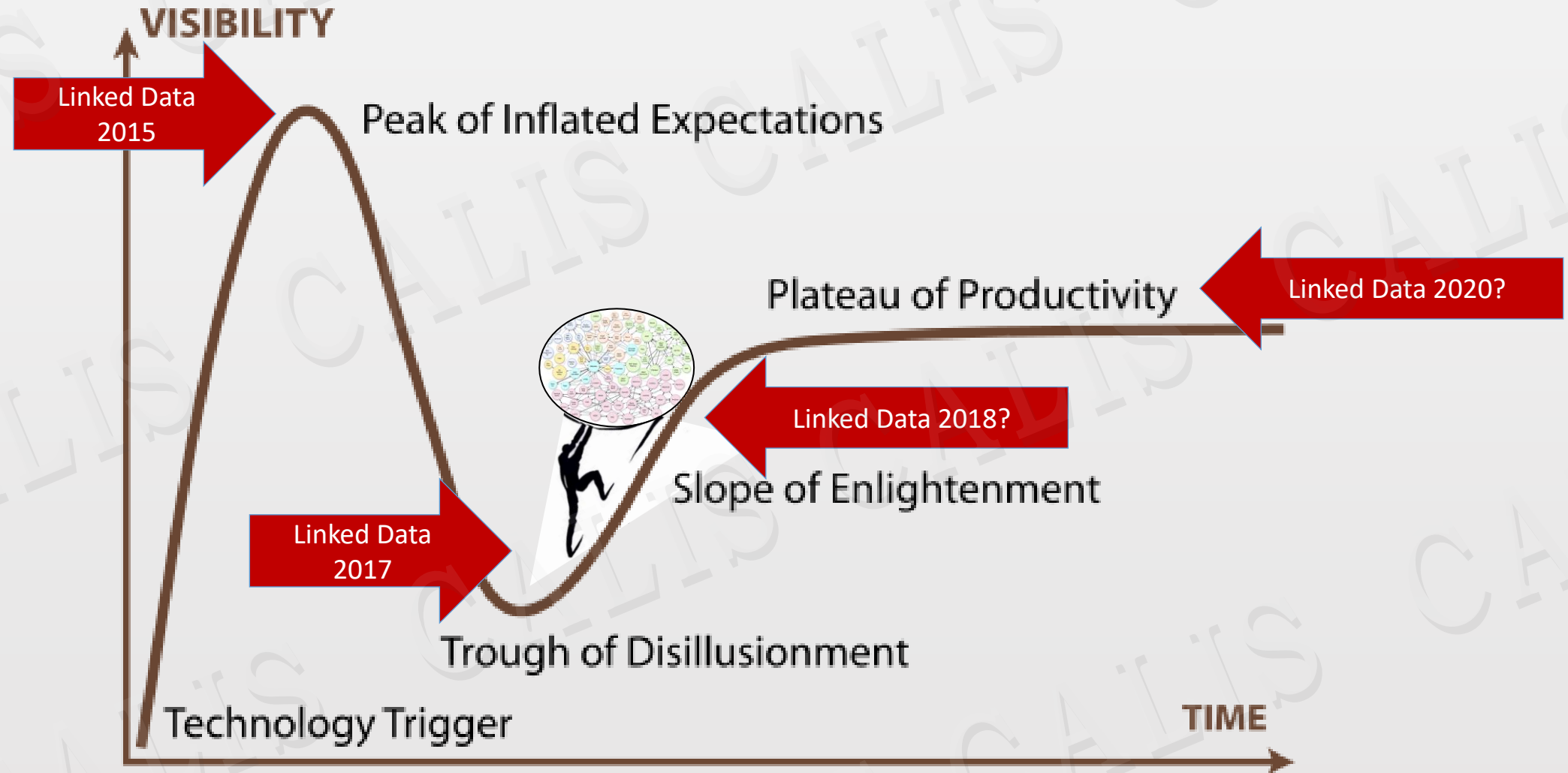
- RDA Aspect Ratio Designation
- RDA Bibliographic Format
- RDA Broadcast Standard
- RDA Carrier Extent Unit
- RDA Carrier Type
- RDA Cartographic Data Type
- RDA Colour Content
- RDA Configuration of Playback Channels
- RDA Content Type
- RDA Conventional Collective Title
- RDA File Type
- RDA Font Size
- RDA Form of Musical Notation
- RDA Form of Notated Movement
- RDA Form of Tactile Notation
- RDA Format of Notated Music
- RDA Frequency
- RDA Generation
- RDA Groove Pitch of an Analog Cylinder
- RDA Groove Width of an Analog Disc
- RDA Illustrative Content
- RDA Layout
- RDA Material
- RDA Media Type
- RDA Mode of Issuance
- RDA Polarity
- RDA Presentation Format
- RDA Production Method

图书馆目录与规范的关联数据发布

2008-

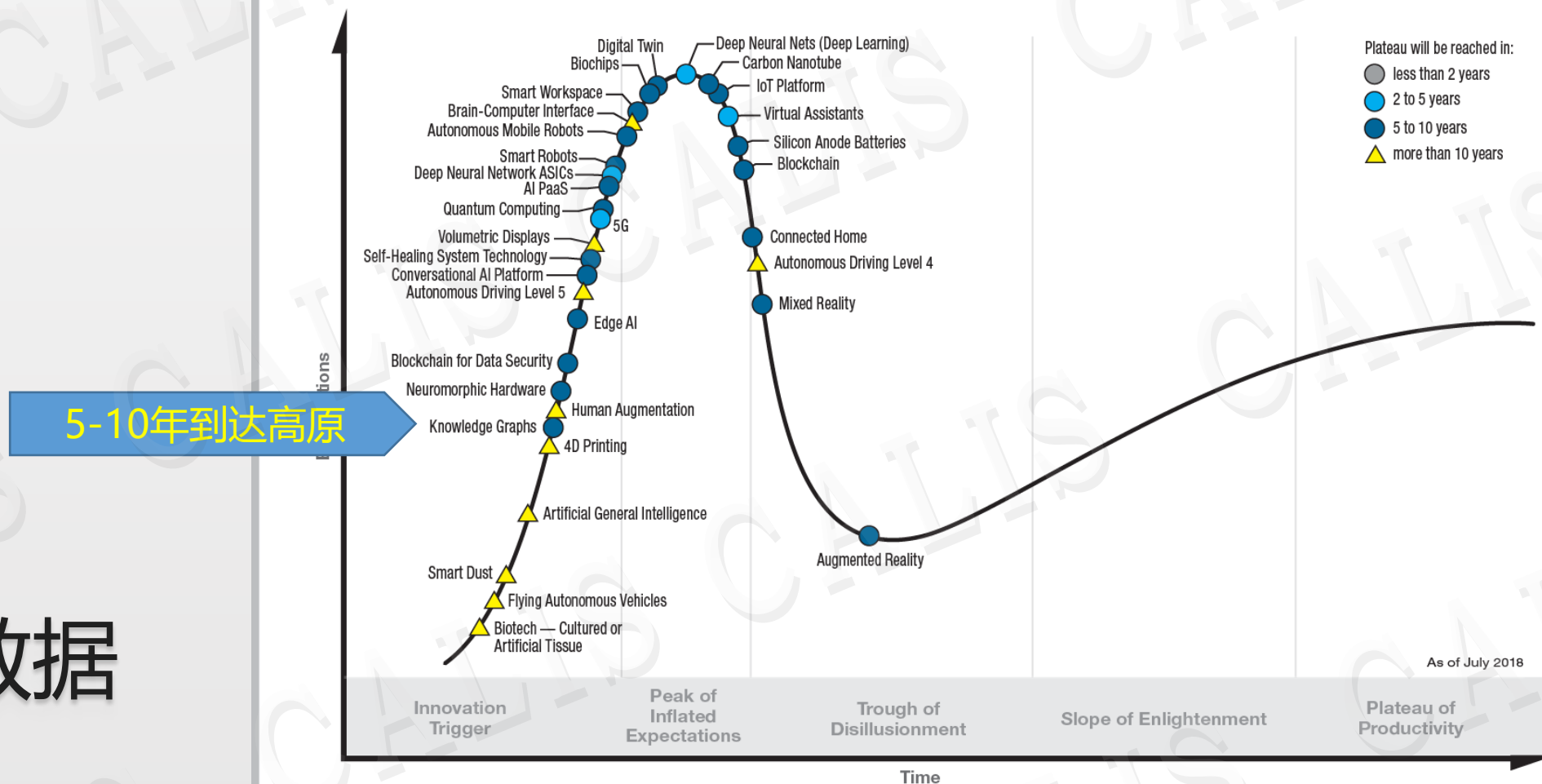
- 瑞典国家图书馆
- 虚拟国际规范档 (VIAF)
- 匈牙利国家图书馆
- 德国国家图书馆
- 不列颠图书馆
-

Gartner Hype Cycle of Emerging Technologies



关联数据 ~ 知识图谱

Hype Cycle for Emerging Technologies, 2018



gartner.com/SmarterWithGartner

Source: Gartner (August 2018)
© 2018 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

[5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018](#) (August 16, 2018)

大型关联数据项目：LD4L系列(梅隆基金资助)

- 2014-2016
 - **LD4L**：创建一个既可以在各个机构内部工作，又可以通过协调、可扩展的关联开放数据网络运行的模型
- 2016-2018
 - **LD4L Labs**：帮助图书馆使用关联数据来改善对学术资源信息的交流和理解
 - **LD4P**：转变技术服务**生产**工作流程
- 2018-2020
 - **LD4P2**：为编目社区建立一条途径，开始转向关联数据来描述图书馆资源

编目员的未来？

数据标注员
了解下？

博文：[RDA不再是编目规则](#)

(2018-7-7)

- 对于国内编目，大家最愿意吐槽的就是国图和CALIS记录的不一致。很多RDA研究（包括我的RDA中文化课题），也把重点放在中国编目规则与RDA的比较上。如果这些差异都不再重要？
- 然后呢？编目员就消失了.....

微信中的神回复

- 编目员现在的工作就是AI的教练啊
[偷笑]真到了AI消灭一切的时候，你还关心编目员有没有工作干嘛？
[发呆]

未来之路：面向关联数据的发展

MARC21

- 准备弃用ISBD标识符
- 新增与重定义URI/IRI相关字段758、子字段\$0\$1

新RDA

- 记录方法：增加IRI
- 应用纲要：无核心元素，一切皆可选，数据多样性

BIBFRAME

- 数据集下载（作品、实例三元组）
- 新编辑器发布

