

# 全国名称规范共享方案

关联数据的概念  
和编目工作有什么关系？

## 什么是编目业务下一步发展？

- 是不断提高编目的质量？
- 是不断拓展编目的种类？
- 是不断提高时效性？

## 编目工作的变与不变

- 文献编目：一般包括文献著录和目录组织以及文献技术加工 等基本程序。但也有一种 看法，认为广义的文献编目还应包括文献分类和主题标引在内。例如，在美国图书馆界，就把描述文献形式特征的著录称作描述性编目，而把主题标引称为主题编目。不同编目机构进行的文献编目和不同类型文献的编目，在具体过程及详略程度上有所不同。

## 公共图书馆宣言

### 教科文组织

本宣言系与国际图书馆协会(IFLA)共同拟订的。  
1994 年 11 月  
原文：英文

### 知识之门

社会和个人的自由、繁荣与发展是人的基本价值。只有当有文化的公民能够行使其民主权利并能在社会上积极发挥作用时，这些价值才能实现。富有成效的参与和民主的发展有赖于良好的教育和对知识、思想、文化及信息的自由和不受限制的获取。  
公共图书馆是各地通向知识之门，为个人和社会群体的终生学习、独立决策和文化发展提供了基本的条件。

本宣言表明教科文组织深信公共图书馆是开展教育、传播文化和提供信息的有力工具，也是在人民的思想中树立和平观念和丰富人民大众的精神生活的重要工具。因此，教科文组织鼓励各国政府及地方政府支持并积极参与公共图书馆的发展。

## 公共图书馆

公共图书馆是各地的信息中心，用户可以随时得到各种知识与信息。  
公共图书馆应不分年龄、种族、性别、宗教、国籍、语言或社会地位，向所有的人提供平等的服务，还必须向由于种种原因不能利用其正常的服务和资料的人，如语言上处于少数的人、残疾人或在病院及在押犯人等提供特殊的服务和资料。  
不同年龄的人都应该在图书馆中找到适合其需要的资料。藏书及各种服务必须包含各类必要的媒体形式和现代技术以及传统的资料。最主要的是保证质量和适合当地的需要及情况。馆藏资料必须反映当前的各种潮流和社会的演变情况以及有关人类的成就和幻想的记录。  
藏书与服务均不受任何形式的思想、政治或宗教审查，也不受商业的压力。

### 公共图书馆的任务

- 以下信息、扫盲、教育和文化方面的主要任务应当成为公共图书馆服务的核心内容：
1. 从小培养和加强儿童的阅读习惯；
  2. 支持个人自学以及各级正规教育；
  3. 为个人发展创造力提供机会；
  4. 激发儿童与青年的想像力和创造力；
  5. 提高对文化遗产的认识，对艺术的鉴赏力以及对科学成就与发明的了解；
  6. 提供通过各种表演艺术来表现文化的途径；
  7. 促进文化间对话和文化多样性；
  8. 发扬口述传统；

(下转 45 页)

- 公共图书馆是各地通向知识之门，为个人和社会群众的终生学习、独立决策和文化发展提供了基本的条件。
- 不变的是公共图书馆的核心职能，是如何建设“知识之门”，是如何帮助“个人和社会群众的终生学习”

## MARC的特性

- MARC: Machine Readable Cataloging
- 上世纪60年代，美国国会图书馆研制成功，并使用到现在。该协议在图书馆界得到了广泛的应用，提升了整个图书馆信息系统工作的进程。

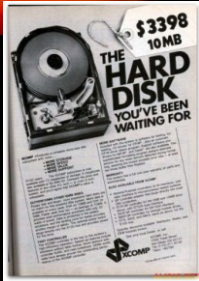
010	a 978-7-208-09279-2  d CNY25.00
035	a (A330000ZJL)012007245257
049	a A330000ZJL  b UCS01004407963
100	a 20100906d2010 em y0chiy50 ea
1010	a chi
102	a CN  b 310000
105	a y z 000ay
106	a r
2001	a 双面胶  9 shuang mian jiao  b 专著  f 六六著
205	a 2版
210	a 上海  c 上海人民出版社  d 2010
215	a 236页  d 24cm
6060	a 长篇小说  y 新加坡  z 现代
701 0	c (新加坡)  a 六六  9 liu liu  c (女)  4 著
801 0	a CN  b A330000ZJL  c 20100906

记录	
010a	978-7-208-09279-2
010d	CNY25.00
200a	双面胶
2009	shuang mian jiao
200b	专著
200f	六六著
210a	上海
210c	上海人民出版社
210d	2010
606a	长篇小说
606y	中国
606z	当代

- MARC的缺点：基于ML (Makeup Language) 的角度审视
  - 可读性差：（什么是577字段？）
  - 可扩展性不足：（最多1000个字段，且不容易建立共识）
- 最重要的是MARC是基于“记录” (Record) 的；

## MARC是基于“记录”

- 换句话说，MARC的所谓“机读”是指让机器读出来给人看的。
- 并不是真正意义上的机读，机器并没有读懂



## MARC的历史局限性

• 01288nam0 2200337 450  
00100100000005001700010010003200027035002600059049004900085100004100134101000800  
175102001500183105001800198106000600216200010100222210003100323215002700354300001  
7003813070035003983300027400433606003900707606001300746406001900759606001300778610  
001000791690001600801701004300817701003100860801002300891910003600914006185176201  
30826101324.0 a978-7-302-32242-9dChy44.50 a(A100000NLC)006755834  
aA100000NLCbUCS01006185178c006755834dNLC01 a20130807d2013 em y0chiy50 ea0  
achi aCnB110000 aa z 000yy arl aUG NX 8.0 中文标准教程9UG NX 8.0 zhong wenban  
biaozhun jiao cheng9 专著张瑞萍, 张玉伟编著 a北京清华大学出版社d2013 a288页c图  
a2dcmel 光盘 a新华电脑学院 a附光盘: ISBN 978-7-89414-510-9 a本书以UG NX 8.0中文版为操作  
平台, 全面介绍使用该软件进行产品设计的方法和技巧。本书共分9章, 重点放在详细讲解UG/CAD模  
块进行产品设计的方法, 主要内容包括绘制草图、创建及编辑曲线、特征建模、创建曲面、装配建模、  
创建工程图和总装模具设计。覆盖了使用UG NX设计各种产品的全部过程。0 a工业产品计算机辅助设  
计x应用软件0 a工业产品0 a计算机辅助设计0 a应用软件0 aUG NX aT847239v5 0a张瑞萍9zhang  
rui pingc[计算机]4编著 0a张玉伟9zhang yu wei4编著 2aCnBOLCCc20130827  
aA100000NLCINLC01n006755834z2

- 磁带机：“磁带是磁带存储系统是所有存储媒体中单位存储信息成本最低、容量最大、标准化程度最高的常用存储介质之一。”——搜狗百科

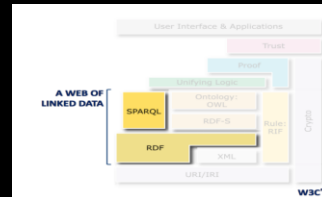
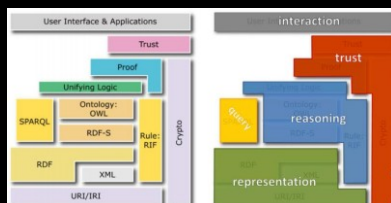
- 磁带机特点：顺序读取



- MARC的缺点并不是因为设计者的能力，而是时代的局限。
- MARC有些理念在现在也并未过时，比如“纯文本格式”
- 要达到记录知识，传递知识的目的，我们需要其他的方法。

## •关联数据

## 关联数据（语义网）



## RDF

- RDF (Resource Description Framework) 从名字就可以看出来, 是一种描述资源的方式。简单来说每一条描述都是一个**主谓宾**三元组构成的短句
- {S,P,O}
- { 苹果, CEO, 库克 }
- { 库克, 是, 男人 }



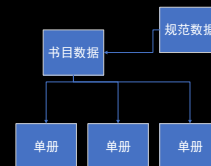
Item	=>title	双面胶
Item	=>ISBN	978-7-208-09279-2
Item	=>Price	CNY25.00
Item	=>GMD	专著
Item	=>author	六六
Item	=>public	上海人民出版社
Item	=>year	2010

## 关联数据

- 知识表达方法上: 一种**严谨的自由**
- 关联数据不是终点, 在更好的知识表达方式问题上, 人类还在不断的探索。

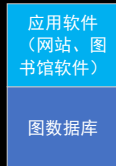
- 为什么要用关联数据?
- =》因为关联数据是目前人类通过实践找到的最理想的知识表示方式。

- FRBR家族 (FRBR、FRAD、FRSAD、PRESSoo) :
- IFLA-LRM (国际图联图书馆参考模型)
- BIBFRAME: Work-Expression-Instance
- FRBR: Work-expression-manifestation-Item



### 什么是关联数据正确的使用方法？

- 底层和中层：使用图数据库Graphstore，作为数据存储的底层（AllegroGraph、Virtuoso等）
- 分析机构数据，尽可能依照已经其他机构已经定义过的元数据方案拆分数据；
- 没有定义的数据，按本机构需求定义并拆分。
- 通过关联数据的算法来实现其他应用的数据排布。  
例如：某些基于关联数据为基础的网站

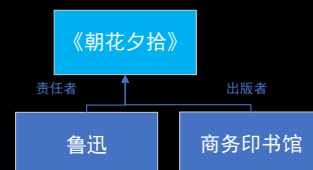


- 图数据库有非常成熟的商业软件可以选择，但目前没有合适的图书馆软件进行应用层支撑。

- 图书馆集成软件
- 如：
  - Aleph: 国家图书馆、首都图书馆等
  - Interlib: 福建图书馆、广州图书馆等
  - LibSys: 吉林省图书馆等
  - ULAS: 深圳图书馆等
  - .....



- 因为没有图数据库，在传统图书馆集成软件中，无法实现真正意义上的关联数据生产。
- 但是传统的图书馆编目结构中，同样存在关系。
- 那就是名称规范数据



## 区别是什么？

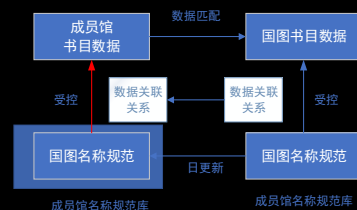
- 这是理念上的区别：
- 在MARC的观念中，以“记录”为核心，各著录项是对记录的补充和完善。
- 在关联数据的理念中，一种书为一个实体对象，其他著录项分别对应不同的实体对象。
- 不同的实体对象间，通过关联关系进行揭示

- 美国国会图书馆的PCC项目组：  
Report for PCC Task Group on the Creation and Function of Name Authorities in a Non-MARC Environment  
PCC任务组在非MARC环境中名称规范的创建和功能报告
- 758字段：资源标识符（2017年新增） \$0和\$1字段
- IFLA的新版UNIMARC: <https://www.ifla.org/node/8415>
- 书目数据中506、507、576、577字段（2016年新增），以及名称规范字段中241和242字段

## 名称规范推广工作的原则

- **方案的可行性**：从实际出发，确保每一阶段进度的可行性，在资源、技术上提出的要求应该与当时已有的条件或在一定时间能争取到的条件相吻合；
- **方案的前瞻性**：注意保持整体工作规划的前瞻性，避免不同阶段制定的标准和接口之间相互矛盾。对元数据；
- **方案的兼容性**：在接口标准和数据规范的制定上，兼顾公共图书馆与高校图书馆的MARC标准。用统一的标准覆盖不同的用户需求；
- **方案的短期价值**：为了保障推广的效果，共享工作成果要确保短期内对图书馆用户在业务和服务层面产生切实有效的提升和帮助。

## 名称规范推广工作的业务模型



## 新的拓展字段：URI

- 全国图书馆联合编目中心唯一标识符：
- 049\$a机构代码
  - \$b唯一标识符：库代码+001字段。
- 例：书目数据中UCS01002578493，名称规范中UCS10000437824
- 名称规范启用049字段
- 书目数据中启用\$3子字段，用于保存联合编目中心唯一标识符，例：701、711字段

- 唯一标识号：通过联编唯一标识号为书目数据和名称规范数据分配，唯一标识号生成规则为：数据库代码+系统号。唯一标识号用于名称规范标识的同时，也作为实体间相互区分的标识的资源名存在。
- 命名空间：英文名Name Space，为不同实体构建的树型元数据体系（命名空间可由我馆建成元数据体系后另行添加）。
- URI：统一资源标识符(Uniform Resource Identifier)由存放资源的主机名+命名空间+唯一标识号组成。未来关联化后，将直接使用现有系统号生成关联数据URI。

- owl:sameAs: OWL语言的内置属性, 用于把两个实例连接在一起。在名称规范工作中, 该属性用于解决名称规范数据合并的异构系统间数据同步问题。
- 数据的删除: 名称规范数据删除时, 仅标识删除, 保留原数据的全部内容。

### 数据的关联回溯

- 由中心提供书目数据的名称规范受控的回溯建议, 用户批量修改所有数据;
- 日常书目数据受控由编目员, 根据中心名称规范数据挂接完成;
- 受控规范数据的修改, 由联编中心提供的日更新文件自动完成。

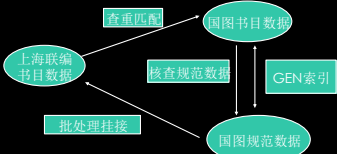
### 名称规范数据共享方案

- 批处理方案
  - 1, 各馆的书目数据
  - 2, 中心的名称规范数据
  - 3, 中心的书目数据

2001 | a鲁迅文学全集 | 9 lu xun wen xue quan ji | h 5 | l 杂文 | b 专  
著  
210 | a北京 | c 群言出版社 | d 2017  
215 | a 360页 | d 24cm  
330 | a 本书收录了《且介亭杂文》《且介亭杂文二集》《且介亭杂文  
未编》三部杂文集。  
6060 | a 鲁迅著作 | j 全集  
6060 | a 鲁迅杂文 | j 杂文集  
701 0 | a 鲁迅 | 9 lu xun | f (1881-1936) | 4 著  
801 0 | a CN | b 浙江省新华书店集团公司 | c 20170906  
801 2 | a CN | b OLCC | c 20170912  
| u  
8564 | u  
http://www.zxhsd.com/kgsm/ts/2017/09/01/3880092.shtml

### 总体技术思路

- 上海联编书目数据与国图书目数据查重匹配。匹配成功的, 挂接国图规范记录。
- 此图由上海图书馆提供



### 索引模式

- 保留原标目控制字段信息, 引入唯一标识符, 启用唯一标识符控制, 以名称规范的唯一标识符 (以 049\$b 作为数据间的关联控制, 原标目内容不进行索引拓展)

- 名称规范数据回溯：
- 在联合编目系统UCS10库中新增049字段，值为\$aA100000NLC,\$bUCS10+名称规范库系统号。其中049\$b值为联编中心唯一标识号。在名称规范库中，作为不同实体相互区分的标识信息。
- 书目数据回溯：
- 为现有书目数据的相应受控7字段增加\$3，值为7字段标目对应名称规范数据的联编中心唯一标识号049\$b。

## 名称规范的实施

- **名称规范查询接口：**
- 支持高级检索，支持不同字段的检索拓展，为成员馆提供包含各种名称规范字段的查询和简要信息的展示；
- **名称规范下载（确认）接口：**
- 提供成员馆确认使用的完整名称规范记录，同时记录该成员馆哪条数据与该条名称规范数据进行了关联。

## 名称规范的实施

- **名称规范日更新接口：**
- 根据成员馆使用的名称规范信息和名称规范库数据当天的更新信息，生成每个成员馆的日更新文件，并通过接口方式发送。

## 编目的未来

- 目前仍然在路上
- 编目的未来属于基于内容的揭示。