

# 元数据仓储详解

国家图书馆 萨蕾





一、元数据标准详解

二、元数据制作

# 元数据标准详解



## ◆ 元数据：

元数据是关于数据的数据。此术语指任何用于帮助网络电子资源的识别、描述和定位的数据。

——IFLA

关于信息资源或数据的一种结构化的数据。

——国家图书馆元数据应用总则

# 元数据标准详解



元数据是对信息包(Information package)的编码描述，其目的在于提供一个中间级别的描述，使得人们据此就可以做出选择，确定孰为其想要浏览或检索的信息包，而无需检索大量不相关的全文文本。

# 元数据标准详解



目录应当是经济有效的工具，它使用户能够：

- ◆ 在利用资源的属性或关系进行检索后，从某一收藏中查找（**find**）书目资源：
- ◆ 识别（ **identify**）一个书目资源或代理
- ◆ 选择（ **select**）一个适合用户需求的书目资源
- ◆ 获取（ **acquire**）或存取（ **obtain access**）所著录的文献，或者检索、获取或存取规范数据或书目数据
- ◆ 浏览（ **navigate**）目录和目录以外的其它领域

——国际编目原则声明

# 元数据标准详解



编目的基本原则：实现元数据功能

- ◆ 强调统一性与规范性
- ◆ 关注资源特色
- ◆ 遵守通用编目规则
- ◆ 保持结构化特点

# 元数据标准详解



## 著录对象:

- ◆ 实体文献与数字对象
- ◆ 整体与部分
- ◆ 内容与形式

# 元数据标准详解



## 著录信息源:

- ◇ 主要信息源: 资源本身
  - 1.资源本身内在的信息源
  - 2.物理载体
- ◇ 规定信息源: 实体文献
  - 参考资源
  - 其他



# 元数据标准详解



## 著录用标识符:

- ◇ 方括号[]
- ◇ 圆括号()
- ◇ 半角分号;
- ◇ 半角冒号:
- ◇ 斜线/
- ◇ 问号?

# 元数据标准详解



## 编目用文字:

- ◆ 题名、责任说明、版本、出版发行、丛编等的信息通常转录自资源，因此，要求要求客观描述，按资源本身的文字著录，外文字母的大小写、标点符号等应遵照其语言文字的书写规则，著录我国各少数民族文献，应遵循其民族文字书写规则
- ◆ 主题、描述、类型、格式等由编目员进行描述或标引的信息均采用中文著录
- ◆ 有规范或规定编码标准的要求按标准著录，如语种

# 元数据标准详解



## 参考书：

- ◇ 国际标准书目著录（统一版）
- ◇ 中国文献编目规则（第二版）
- ◇ UNIMARC Manual: Bibliographic Format (3<sup>rd</sup> Edition)
- ◇ 新版中国机读目录格式使用手册
- ◇ DCMI Metadata Terms
- ◇ DC图书馆应用纲要
- ◇ 国家图书馆元数据应用总则规范汇编

# 元数据标准详解



## 著录细则：

### ◇ 元素及修饰词

名称	责任者	主题	描述	出版者	日期
类型	原始标识符	格式	语种	关联	权限
发布地址	使用对象	标识符	数据提交单位	数据提交日期	其他

# 元数据标准详解



## ◆ 例1:

邓小平—当代中国马克思主义的创立者

邓小平：我是中国人民的儿子

## ◆ 例2:

李大钊：[电视文献记录片]

## ◆ 例3:

责任者名称1 (法)法布尔

责任方式1 著

责任者名称2 浦漫汀;曹文轩

责任方式2 主编

# 元数据标准详解



## ◆关于“内容形式、媒体类型”

### 一般资料类型标识(General Material Designations)

1977年版ISBD(G)中明确指出：“一般资料标识的目的在于用概括性的术语在著录的开始说明文献所属的资料类别，GMD在题名后用方括号表示，属于可选项。”

# 元数据标准详解



IFLA发布的 ISBD（《国际标准书目著录》）统一版中设置了第0项“内容形式和载体类型项”，三部分构成：

- ◆内容形式（Content Form）：表示资源内容的一种或多种基本形式。
- ◆内容限定（Content Qualification）：对内容形式的限定说明，即指出受编资源的类型、感觉性质、维数和/或活动的有无。
- ◆载体类型（Media Type）：用以承载资源内容的一种或多种载体的类型。

# 元数据标准详解



2014年4月中国国家标准GB/T 3469-2013  
《信息资源的内容形式和媒体类型标识》颁布

参考相关国际标准，将信息资源类型按照资源内容和资源载体为内容形式、内容限定与媒体类型三类。

增加了一个在国际标准中没有的限定词表“资料类别说明”。描述对象主要针对传统印刷品及其衍生品（如缩微品和电子图书）。但要求严格控制使用，只有当内容形式为“文本”或“图像”时，才能作为特殊限定词出现。



# 元数据标准详解



内容形式	内容限定				媒体类型
数据集	测绘型	动态	二维	听觉	音频
图像	记谱型	静态	三维	味觉	电子
运动	表演型			嗅觉	缩微
音乐				触觉	显微
实物				视觉	投影
程序					立体
声音					视频
话语					多媒体
文本					其他媒体
多种内容形式					
其他内容形式					

# 元数据标准详解



## 扩充内容限定词（资料类别说明）

会议文献	专利文献
学位论文	期刊
手稿	拓片
专著	标准文献
报纸	技术报告
古籍	

# 元数据标准详解



例：

地图：图像（测绘型）

视频：图像（动态）

图片：图像（静态）

乐谱：音乐（记谱型）

歌剧：音乐（表演型）

评书：话语

# 元数据标准详解



## ◆ 关于综合著录与分析著录：

综合著录：对一部文献或一组（套）文献进行整体描述，形成综合款目的著录方法

分析著录：在综合著录的基础上，对文献的组成部分进行独立描述，形成分析款目的方法。

# 元数据标准详解



## ◆ 关于多载体配套资源

有主次之分：以主要部分为主要著录对象，  
将次要部分著录于附注

无主次之分：可选择综合著录或分析著录

# 元数据制作



◇ 如何充分利用已有数据？

## MARC与DC映射

<1> 一对多或多对一

如：DC的主题元素(subject)对应MARC的多个用于标引的字段，包括主题标引、关键词标引、分类标引等。

<2> 无对应关系

如：DC的日期元素(date)与MARC描述日期的相关字段存在着语义上的差异，无法完全匹配。

# 元数据制作



## <3> 语义内涵的不一致

如：CNMARC中为实现描述与检索两个功能，以多个字段著录同一内容，一部分字段承担描述功能，一部分字段承担检索与规范的功能，例如题名与统一题名，用于著录责任者的2字段与7字段。而在DC中没有承担规范功能的元素。

## <4> 自由词与规范词转换带来的不一致

# 元数据制作

## 如何利用EXCEL制作元数据？

1 左键点击某一栏目

2 点击资源

3 从Excel创建记录

4 保存(S)

名称	责任者名称1	责任方式1	责任者名称2	责任方式2	分类	主题词或关键词
丑末寅初	刘宝全	演唱者	姚士豪			

名称	修改日期	类型
ac48794dd69f54909df5e8679e	2012/2/28 周二 ...	文件夹
alipay	2014/3/24 周一 ...	文件夹
ASUS WebStorage	2011/4/22 周五 ...	文件夹
ASUS.DAT	2010/8/14 周六 ...	文件夹
Config.Msi	2014/9/3 周三 2...	文件夹
Download	2012/7/6 周五 1...	文件夹
eSupport	2010/6/6 周日 2...	文件夹
Intel	2010/6/6 周日 2...	文件夹
log	2012/12/21 周五...	文件夹
MSOCache	2010/9/10 周五 ...	文件夹



# 元数据制作

1 找到“data.xls”文件并打开

2 将做好的元数据拷贝到此文件

The screenshot displays a Microsoft Excel spreadsheet with a table of metadata. The table has columns for name, performer, role, classification, and various attributes. A 'Save As' dialog box is open over the spreadsheet, showing the file name '天津相声曲艺' and the save type 'XML 数据'. The 'Save' button is highlighted.

1	名字	责任者名称1	责任方式1	责任者名称2	责任方式2	分类	主题词或关键词	语种	资源类型	内容类型	语言	访问方式	发布地址	适用对象	音视频时
2	大西厢	刘宝全	演唱者	姚士泉	表演者	J8	京剧大鼓, 大西厢	崔富富白音胡蝶	曲艺	中文	互联网访问	http://www.	普通用户	16分50秒	
3	华容道	刘宝全	演唱者	姚士泉	表演者	J8	京剧大鼓, 华容道	三国时期	曲艺	中文	互联网访问	http://www.	普通用户	5分50秒	
4	游武庙	刘									互联网访问	http://www.	普通用户	6分54秒	
5	别紫鹃	白									互联网访问	http://www.	普通用户	6分45秒	
6	黛玉焚稿	白									互联网访问	http://www.	普通用户	6分09秒	
7	探晴雯	白									互联网访问	http://www.	普通用户	6分23秒	
8	宝剑扑蝶	乔									互联网访问	http://www.	普通用户	17分30秒	
9	凤仪亭	乔									互联网访问	http://www.	普通用户	6分19秒	
10	王二姐捧	乔									互联网访问	http://www.	普通用户	6分55秒	
11	昭君出塞	乔									互联网访问	http://www.	普通用户	6分17秒	
12	花蝴蝶口	马									互联网访问	http://www.	普通用户	7分07秒	
13	玲珑塔	马									互联网访问	http://www.	普通用户	10分39秒	
14	偷年糕	马									互联网访问	http://www.	普通用户	2分20秒	

3 另存为“XML数据”格式的文件

4 保存(S)

# 元数据制作

2 选择“从外部导入记录”

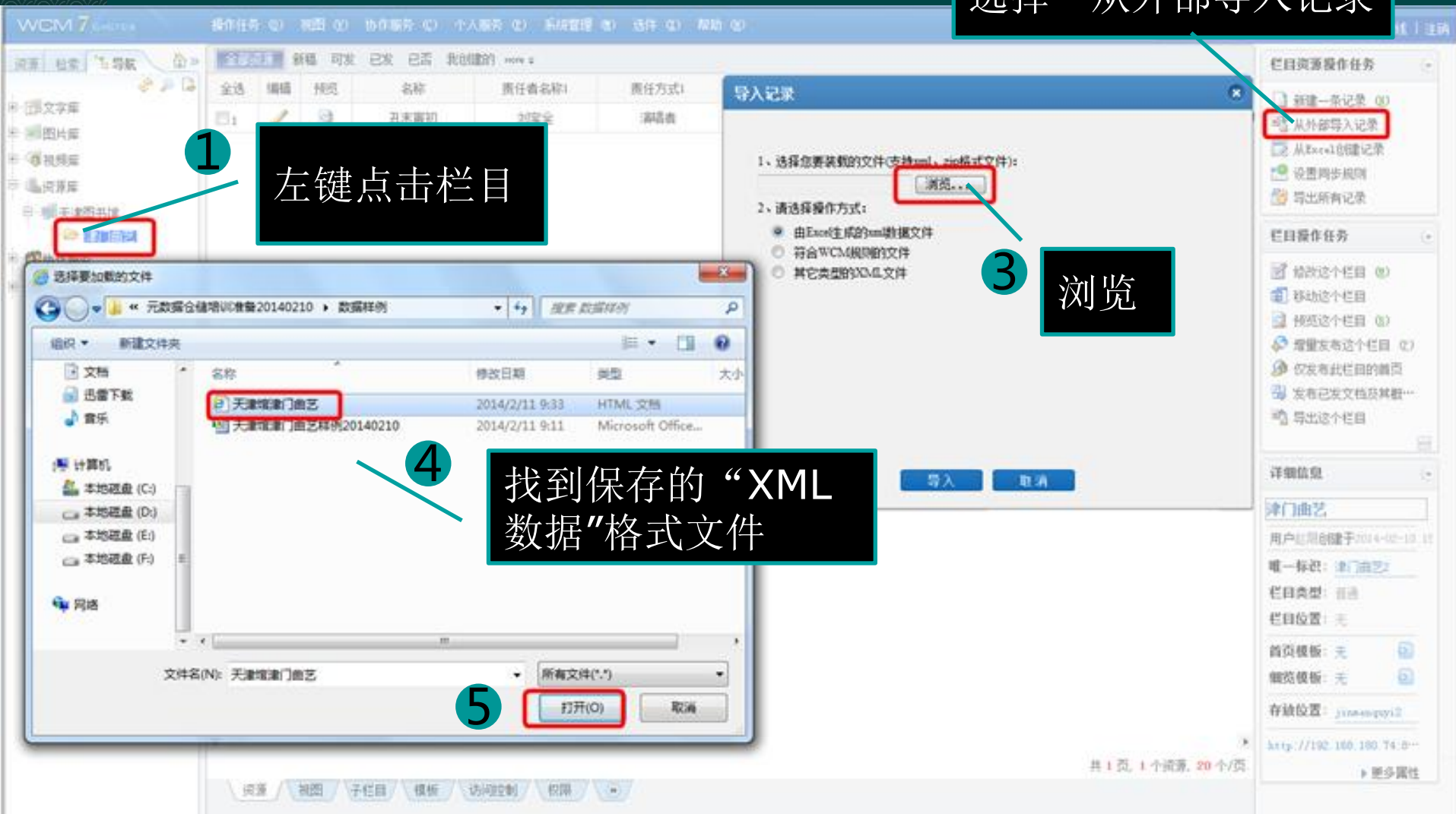
1 左键点击栏目

3 浏览

4 找到保存的“XML数据”格式文件

5

打开(O)



# 元数据制作



1

导入记录

2

导入完成后，弹出导入结果对话框，关闭此框数据导入结束





THANK YOU  
QUESTIONS?