

《中分表》数据语义化 及关联数据服务

喻菲 国家图书馆

主要内容

- 一、关联数据概述
- 二、国家图书馆关联数据注册与服务平台
- 三、《中分表》数据的语义化描述
- 四、《中分表》注册与服务平台
- 五、现在和未来

一、关联数据概述

- 关联数据（Linked Data）
 - 使用RDF数据模型发布和部署实例和类数据的一组最佳实践的集合。
 - 使用统一资源标识符（**URIs**）来命名数据实体，通过**HTTP**协访问数据，使机器可以理解它所强调的数据之间的联系、相互关系和上下文语境。
- 基本要素:**URI**、**HTTP URI**、**RDF**以及**RDF**链接
- 关联数据开放云图(Linked Open Data Cloud-LOD)
 - 2007年：12个数据集
 - 2018年：1200多个数据集

- 图书馆拥有多种资源类别，包括：
 - 书目记录、馆藏记录、规范记录、主题词表、分类表以及叙词表.....

国外

- 瑞典国家图书馆：
 - 2008年，**LIBRIS**（瑞典联合目录），世界上第一个被整体发布为关联数据的联合目录
 - 约**600万**条书目记录、**2,000万**条馆藏记录及**20万**条规范记录发布为关联数据
- 匈牙利国家图书馆：2010年，书目数据发布为关联数据
- 西班牙图书馆：2012年，**40万**条书目记录，**400万**条个人、团体、统一题名及主题的规范记录
- 德国国家图书馆：2012年，书目数据，**200万**条个人名称规范记录、**18.8万**条主题规范记录、**130万**条团体名称规范记录，**51,478**个类概念
- 法国国家图书馆：书目数据、主题词表**RAMEAU**
- **OCLC: WorldCat**书目记录发布为关联数据
- 美国国会图书馆：**LCSH**

国内

- 上海图书馆
 - 上海图书馆开放数据平台：基础知识库（人、地、时、事、物），文献知识库（家谱、手稿档案、古籍等），本体词表
- 国家科学图书馆
 - 机构知识库的语义扩展
- 台湾“国家图书馆”
 - 中文主题规范系统：《中文主题词表》、《中文图书分类法(2007 年版)》、《中文人名权威控制》
- 叙词表
 - 《中国农业科学叙词表》

二、国家图书馆关联数据注册与服务平台

- 对不同类型的关联数据的统一注册、发布、管理与服务，以及国家图书馆数据之间的内联以及与外界数据的外联，为来自其他机构的资源提供相关链接，并通过链接到这些资源的上下文信息来扩大知识资源的使用范围。
- 《中分表》注册与服务平台
- 国图公开课
- 中日韩数字图书馆

三、《中分表》数据的语义化描述

- 2008年，国图开始研究《中分表》数据的语义化描述问题，采用W3C颁布的SKOS推荐标准中的数据模型表达类目、主题词及其体系的结构和内容，设计了一套通用格式用于中文数字资源组织用的分类号和主题词在互联网上进行数据交换和数据关联，实现数据的获取、共享和重用，并完成了《中分表》数据的语义化描述。

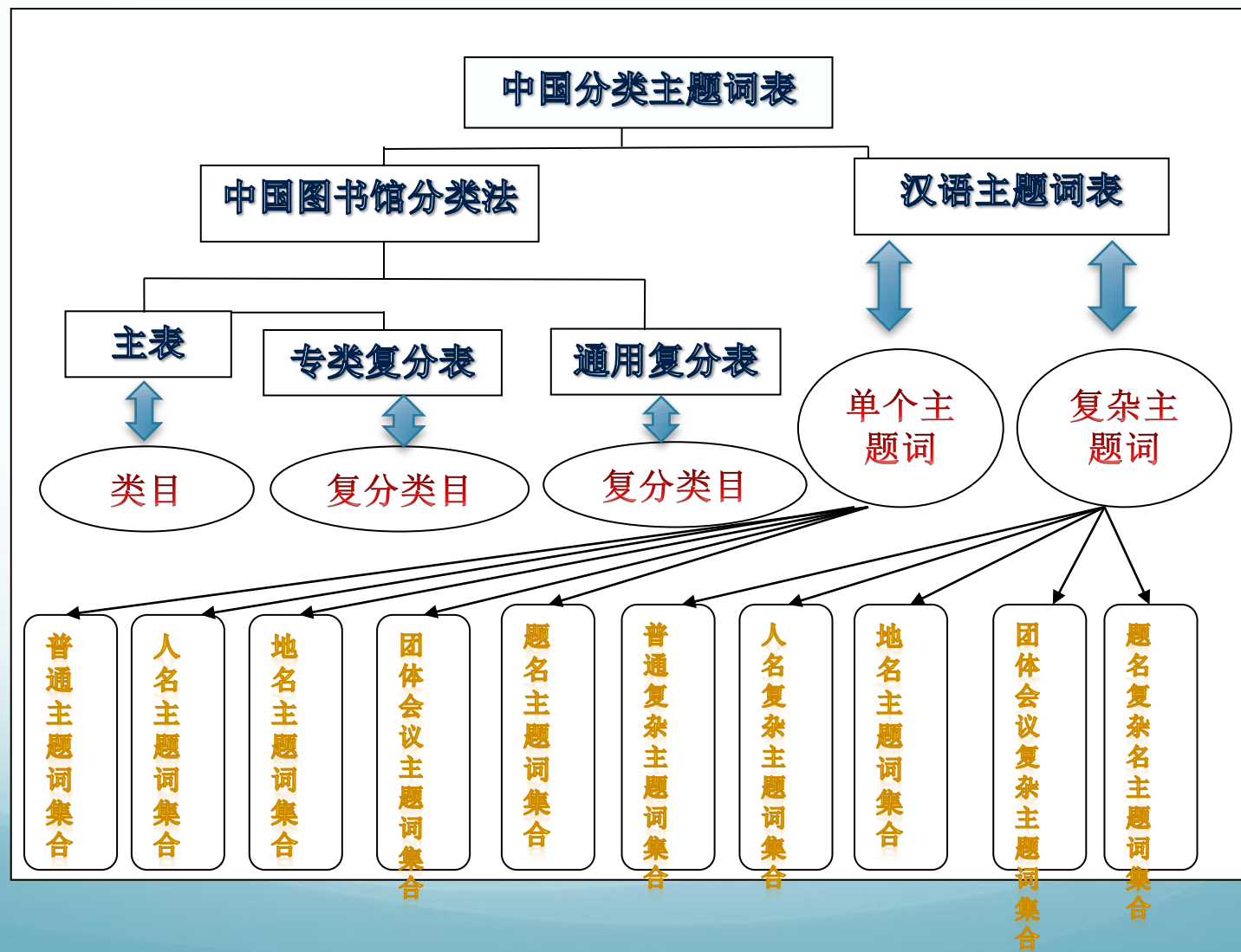
《中分表》数据的语义化描述

- 《中分表》关联数据的总体结构
- 《中分表》**SKOS**数据描述结构
- 《中分表》各种类型数据的**SKOS**数据描述

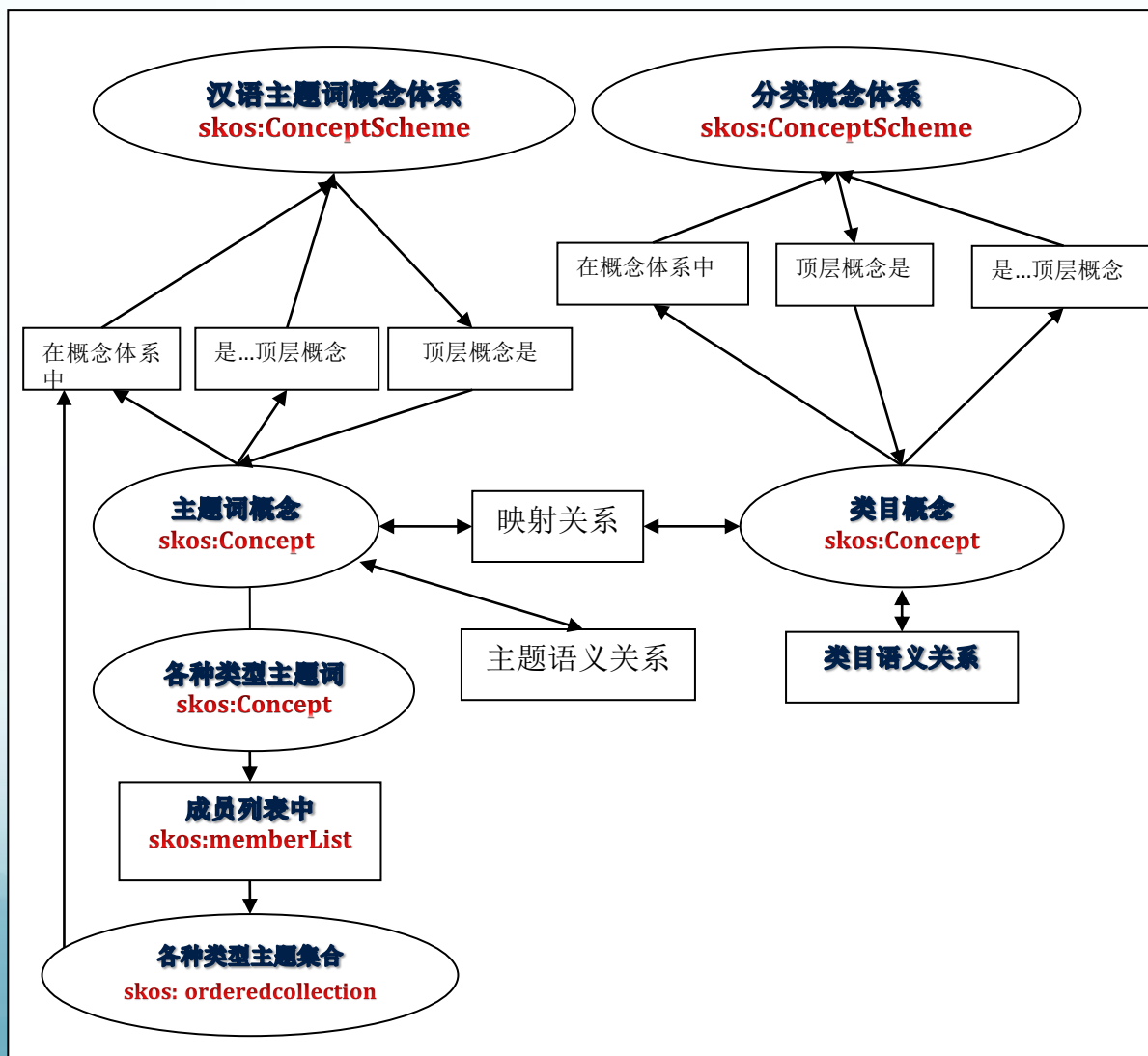
SKOS: Simple Knowledge Organization System

- 2005年，W3C发布《SKOS核心词汇规范》
- 2008年，W3C发布《SKOS参考》（SKOS Reference）、《SKOS初级读本》（SKOS Premier）
- 建立在RDF基础上，提供了表达各种知识组织系统基本结构和内容的通用模型，将概念模式及语义关系表达为机器可理解的方式，还支持概念在万维网上编辑和发布，将概念与网络上的数据相链接，以及将概念集成到其它概念体系中。
- SKOS词表是由一系列RDF定义的类和属性组成，这些类和属性涵盖了各种传统知识组织系统结构中使用到的标签。

《中分表》关联数据的总体结构



《中分表》SKOS数据描述结构



SKOS要素

概念

标签

语义关系

文档注释

概念体系

《中分表》各种类型数据的SKOS数据描述

- 概念体系
- 《主题词一分类号对应表》

- 汉语主题词表

`<http://data.nlc.cn/thesaurus/cct&heading/thesaurus#conceptScheme> rdf:type skos:ConceptScheme`

- 汉语复杂主题词表

`<http://data.nlc.cn/thesaurus/cct&heading/SubjectHeading#conceptScheme> rdf:type skos:ConceptScheme`

- 《分类号—主题词对应表》
- 主表
<http://data.nlc.cn/thesaurus/cct&heading/classification_main#conceptScheme> rdf:type skos:ConceptScheme
- 通用复分表（以“总论复分表”为例）
<http://data.nlc.cn/thesaurus/cct&heading/classification_general_subdivision#conceptScheme> rdf:type skos:ConceptScheme
- 专类复分表（以“A56专题汇编”下专类复分表为例）
<http://data.nlc.cn/thesaurus/cct&heading/classification_A56_subdivision#conceptScheme> rdf:type skos:ConceptScheme

分类法类目

- 款目结构

F74 F741.1 F741.2

F741 政策

互惠政策入此。

国际贸易政策；国际贸易政策\互惠待遇

出口战略；出口政策；非关税壁垒；技术壁垒；|贸易战略|；配额

记录控制号：C004277

```
<skos:Concept rdf:about="http://data.nlc.cn/thesaurus/cct/clc-main/C004277#concept">
<skos:inScheme rdf:resource="http:// data.nlc.cn/thesaurus/cct/clc-main #conceptScheme"/>
<skos:prefLabel xml:lang="zh">政策</skos:prefLabel>
<skos:prefLabel > F741</skos:prefLabel>
<skos:notation rdf:datatype="http://data.nlc.cn/CLCnotation">F741</skos:notation>
<skos: scopeNote xml:lang="zh">互惠政策入此。 </skos: scopeNote>
<skos:broaderTransitive
```

```
rdf:resource="http://data.nlc.cn/thesaurus/cct/clc-main/C002900#concept"/>
```

```
<skos: broader rdf:resource=" http://data.nlc.cn/thesaurus/cct/clc-main/C004266#concept"/>
<skos: narrower rdf:resource="http://data.nlc.cn/thesaurus/cct/clc-main/C004278#concept"/>
  <skos: narrower rdf:resource="http://data.nlc.cn/thesaurus/cct/clc-main/C004279#concept"/>
  <skos:scopeNote xml:lang="zh">互惠政策入此。 </skos:scopeNote>
```


<skos: broadMatch

rdf:resource="http://data.nlc.cn/thesaurus/cct/subheading/H006852#concept"/>

<skos:editorialNote xml:lang="zh">CNLC20051001</skos:editorialNote>

<skos:broadMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S103653#concept"/>

<skos:broadMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S103654#concept"/>

<skos:broadMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S017434#concept"/>

<skos:broadMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S020774#concept"/>

<skos:broadMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S109264#concept"/>

<skos:broadMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S113465#concept"/>

<skos:exactMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S028997#concept"/>

<skos:relatedMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S112465#concept"/>

</skos:Concept>

主题词款目

jiao xiang qu

交响曲

Symphony

J614

D 交响乐曲

S 器乐曲

F 小交响曲

Z 乐曲

C 交响乐

记录控制号: S038576

```
<skos:Concept rdf:about="http://data.nlc.cn/thesaurus/cct/subject/S038576#concept">
  <skos:inScheme rdf:resource="http://data.nlc.cn/thesaurus/cct/subject#conceptscheme"/>
  <skos:prefLabel xml:lang="zh">交响曲</skos:prefLabel>
  <skos:altLabel xml:lang="zh-pinyin">jiao xiang qu</skos:altLabel>
  <skos:altLabel xml:lang="en">Symphony</skos:altLabel>
  <skos:altLabel xml:lang="zh">交响乐曲</skos:altLabel>
  <skos:altLabel xml:lang="zh-pinyin">jiao xiang yue qu</skos:altLabel>
  <skos:broader rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S058989#concept"/>
  <skos:narrower rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S082471#concept"/>
  <skos:related rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S038573#concept"/>
  <skos:broaderTransitive
rdf:resource="http://data.nlc.cn/thesaurus/cct/subject/S092154#concept"/>
  <skos:broadMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/clc-main/C007401#concept"/>
  <skos:editorialNote>CN;NLC;20181127</skos:editorialNote>
</skos:Concept>
```

复杂主题词

- 变文\中国\近代

D 中国近代变文

```
<skos:Concept rdf:about="http://data.nlc.cn/thesaurus/cct/subheading/H010558#concept">  
  <skos:inScheme rdf:resource="http://data.nlc.cn/thesaurus/cct/subheading#conceptscheme"/>  
  <skos:prefLabel xml:lang="zh">变文\中国\近代</skos:prefLabel>  
  <skos:altLabel xml:lang="zh-pinyin">bian wen\zhong guo\jin dai</skos:altLabel>  
  <skos:altLabel xml:lang="zh">中国近代变文</skos:altLabel>  
  <skos:altLabel xml:lang="zh-pinyin">zhong guo jin dai bian wen</skos:altLabel>  
  <skos:exactMatch rdf:resource="http://data.nlc.cn/thesaurus/cct/clc-main/C006934#concept"/>  
  <skos:editorialNote>CN;NLC;20181127</skos:editorialNote>  
</skos:Concept>
```

四、《中分表》注册与服务平台

主要功能：

- 《中分表》关联数据的注册与发布
- 《中分表》关联数据服务
- 后台数据管理
- 平台用户管理
- 系统管理

关联数据注册与服务平台主要功能



《中分表》关联数据的注册与发布

- 平台的主要功能
- 角色用户
 - 系统管理者、数据提交者、数据审核者、数据管理者、一般用户
- 数据状态
 - 数据审核：“待审核”、“不通过”、“合格”、“失效”、“被替代”
 - 数据发布：“未发布”、“预发布”、“已发布”、“屏蔽”

《中分表》关联数据的展示

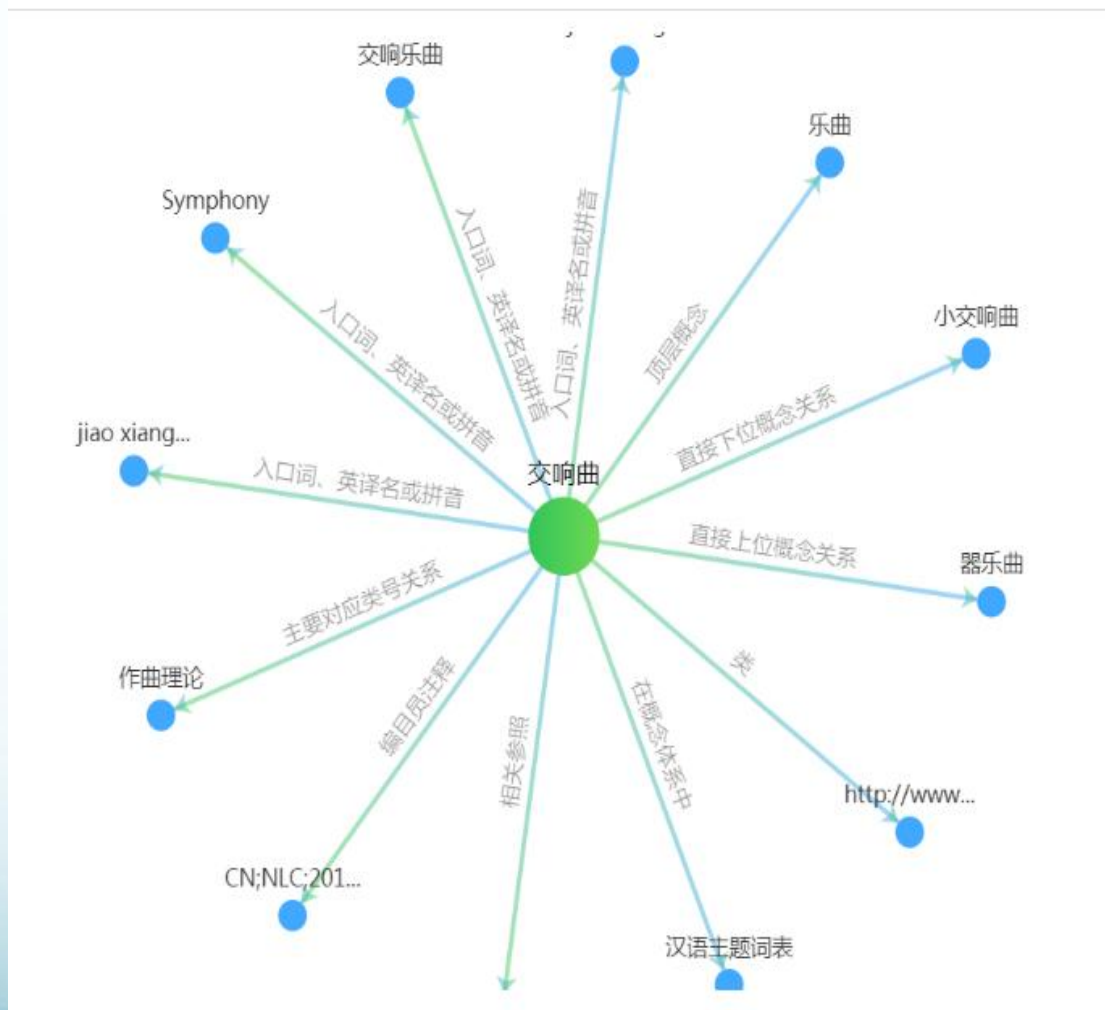
- 数据集展示
 - 树状形式
- 数据内容展示
 - 关联图
 - 三元组
 - 属性列表

《中分表》数据集展示

▼ 主表

- ▶ A 马克思主义、列宁主义、毛泽东思想、
- ▶ B 哲学、宗教
- ▶ C 社会科学总论
- ▶ D 政治、法律
- ▶ E 军事
- ▼ F 经济
 - ▶ F0 经济学
 - ▶ F1 世界各国经济概况、经济史、经济
 - ▶ F2 经济管理
 - ▶ F3 农业经济
 - ▶ F4 工业经济
 - ▶ F49 信息产业经济
 - ▶ F5 交通运输经济
 - ▶ F59 旅游经济
 - ▶ F6 邮电通信经济
 - ▼ F7 贸易经济
 - ▶ F71 国内贸易经济
 - ▶ F72 中国国内贸易经济
 - ▶ F73 世界各国国内贸易经济

《中分表》数据关联图



《中分表》数据三元组

| Subject | Predicate | Object |
|---|---|---|
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://www.w3.org/2004/02/skos/core#Concept |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#narrower | http://data.nlc.cn/thesaurus/cct/subject/S082471#concept |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#broadMatch | http://data.nlc.cn/thesaurus/cct/clc-main/C007401#concept |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#prefLabel | 交响曲 |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#inScheme | http://data.nlc.cn/thesaurus/cct/subject#conceptscheme |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#altLabel | jiao xiang qu |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#altLabel | jiao xiang yue qu |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#altLabel | 交响乐曲 |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#altLabel | Symphony |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#editorialNote | CN/NLC;20181127 |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#related | http://data.nlc.cn/thesaurus/cct/subject/S038573#concept |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#broaderTransitive | http://data.nlc.cn/thesaurus/cct/subject/S092154#concept |
| http://data.nlc.cn/thesaurus/cct/subject/S038576#concept | http://www.w3.org/2004/02/skos/core#broader | http://data.nlc.cn/thesaurus/cct/subject/S058989#concept |

《中分表》数据属性列表

| 元素名 | 中文译名 | 值 |
|------------------------|------------|---|
| rdf:type | 类 | http://www.w3.org/2004/02/skos/core#Concept |
| skos:narrower | 直接下位概念关系 | http://data.nlc.cn/thesaurus/cct/subject/S082471#concept |
| skos:broadMatch | 主要对应类号关系 | http://data.nlc.cn/thesaurus/cct/clc-main/C007401#concept |
| skos:prefLabel | 主题标目 | 交响曲 |
| skos:inScheme | 在概念体系中 | http://data.nlc.cn/thesaurus/cct/subject#conceptscheme |
| skos:altLabel | 入口词、英译名或拼音 | jiao xiang qu |
| skos:altLabel | 入口词、英译名或拼音 | jiao xiang yue qu |
| skos:altLabel | 入口词、英译名或拼音 | 交响乐曲 |
| skos:altLabel | 入口词、英译名或拼音 | Symphony |
| skos:editorialNote | 编目员注释 | CN;NLC;20181127 |
| skos:related | 相关参照 | http://data.nlc.cn/thesaurus/cct/subject/S038573#concept |
| skos:broaderTransitive | 顶层概念 | http://data.nlc.cn/thesaurus/cct/subject/S092154#concept |
| skos:broader | 直接上位概念关系 | http://data.nlc.cn/thesaurus/cct/subject/S058989#concept |

《中分表》关联数据的检索

- 一般搜索
 - 一键式检索，在用户不明确检索词内容的具体归属时选择此检索，系统默认实现对数据集、元数据和数据内容的检索。
- SPARQL查询
 - 利用SPARQL语句对数据进行查询。
 - SPARQL是为RDF开发的一种查询语言和数据获取协议。

《中分表》关联数据的下载

- 提供受权限限制的下载功能
- 提供多种格式的下载功能。包括：RDF/XML（默认）、N3、NTriple、Turtle、RDF/JSON（目前比较常用的关联数据发布格式）
 - RDF/XML：W3C制定的标准推荐格式
 - Turtle家族：包括N3、Ntriple，最接近RDF表达的数据格式，表达更简单
 - RDF/JSON：JSON是由Javascript支持的网络数据交换格式，RDF/JSON是前者的扩展
- 整体下载
 - 该数据集的全部资源进行下载
- 单条资源
 - 下载该条资源
- 批量资源下载
 - 下载已选中的多条资源

REST API

- **Web Service**技术，能使得运行在不同机器上的不同应用无需借助附加的、专门的第三方软件或硬件，就可相互交换数据或集成
- 最主要的两种实现方式是：**SOAP** 和 **REST**
- 标准的基于**HTTP**协议的对外**REST API**服务接口及接口说明，其他平台遵循该接口规范，可实现对平台关联数据的应用。
- 提供接口：
 - （1）通过访问资源 **URI** 获取数据
 - （2）通过**SPARQL**语句获取数据
 - （3）获取所有数据集列表
 - （4）根据数据集**ID**和页码获取数据集的资源
 - （5）根据数据集**ID**或资源**URI**数组下载相应的资源

五、现在

- 《中分表》数据注册与服务平台的基本功能
- 类目树的可视化显示
- 类目树与主题词基本互动显示
- 类目及主题词数据的内容显示
- 类目及主题词数据的关联显示
-

未来...

图书馆的特色数据发布为关联数据



内部数据发布

外部数据发现

语义检索

无界的数据资源网络

全面的资源服务

谢谢！

喻菲

yfei@nlc.cn