

2013 年第 1 期（总第 16 期）

# 数字图书馆标准规范

## 跟踪扫描

主办单位：中国科学院国家科学图书馆

2013 年 2 月

为传播科学知识，促进业界交流，  
特编译《标准规范跟踪扫描》，仅供个人  
学习、研究使用。

## 目 录

【标准规范推介】 .....	1
一、科研数据的元数据与标准 .....	1
1、开放档案信息系统(OAIS)参考模型 .....	1
2、ISO 14721:2003 空间数据与信息转移系统——开放档案信息系统(OAIS)——参 考模型 .....	2
3、CERA-2 .....	2
4、欧洲社会科学数据档案理事会(GESSDA)目录 .....	3
5、CF元数据与天气和预测元数据公约 .....	3
6、Gesis——莱布尼兹社会科学研究所以 .....	3
7、数据批准印章(DSA) .....	4
8、数据批准印章 .....	4
9、用任意查询收割DataCite元数据 .....	5
10、杜威十进制图书分类法 .....	6
11、DDI——数据文档计划 .....	6
12、DOI系统 .....	7
13、ISO 26324:2012 信息和文献——数字对象标识符系统 .....	8
14、基于风险评估的数字存储库审计方法(DRAMBORA) .....	8
15、都柏林核心——都柏林核心元数据计划 .....	8
16、欧洲语言社会科学叙词表(ELSSST) .....	9
17、Genericode——定义代码列表的一种标准格式 .....	9
18、地球科学的地面欧洲网络互操作——数字存储库(GENESI-DR) .....	10
19、句柄系统与CNRI——美国国家研究创新联合会 .....	10
20、ICPSR——美国校际社会科学数据共享联盟 .....	11
21、ICPSR数字保存政策框架 .....	13
22、ISCO——国际标准职业分类 .....	13
23、ISO 639-1:2002 语种名称表示代码——第一部分: α-2 代码 .....	14
24、国家代码——ISO 3166 .....	14
25、ISO 19115:2003 地理信息——元数据 .....	15

26、ISO 20252:2006 市场、观点和社会研究——词汇和服务需求 .....	16
27、ISO/IEC 11179, 信息技术——元数据注册 (MDR) .....	16
28、数字格式可持续性 .....	17
29、电子资源长时期保存元数据 (LMER) .....	18
30、机读目录 (MARC) .....	18
31、气候模拟数字存储库的通用元数据 .....	18
32、海洋元数据互操作 .....	19
33、元数据编码和传输标准 .....	19
34、元数据对象描述框架 .....	19
35、NESTOR——德国数字资源长期保存专业技术网 .....	20
36、NetCDF (网络通用数据格式) .....	21
37、OASIS编码列表表达技术委员会 .....	21
38、地理和环境数据发布网络 .....	22
39、保存元数据: 实施战略 .....	22
40、持续的统一资源定位符 .....	22
41、观看标准: 元数据世界的可视化 .....	23
42、元数据标准的词汇表 .....	23
43、欧洲航天局地球观察: 欧洲的标准档案格式 .....	23
44、ISO 17369:2013 统计数据与元数据交换 (SDMX) .....	24
45、SKOS简单知识组织系统 .....	24
46、跨文化调查中最好实践的指导路线 .....	25
47、十原则——数字保存仓储的十项基本特征 .....	26
48、可信赖的仓储审核和认证: 标准和清单 .....	26
49、Triple S. 调查交换标准 .....	27
50、统一资源名称 .....	27
51、因特网工程任务组 .....	27
52、世界数据中心: 气候 .....	28
53、Z39.50: 信息检索协议 .....	28
二、数据文档计划 (DDI) 技术规范第一部分: 综述 版本 3.1 .....	29

## 【标准规范推介】

### 一、科研数据的元数据与标准

#### 1、开放档案信息系统(OAIS)参考模型

本文档为开放档案信息系统(OAIS)定义了国际标准组织(ISO)参考模型。OAIS 是一种档案馆,由负责保存信息并为指定团体服务的组织(可能是更大的组织的一部分)、人和系统组成。它需履行本国际标准中规定的责任,这使得 OAIS 档案集与其他“档案馆”区分开来。OAIS 中的“开放”一词用于暗示该标准以及未来的相关国际标准可在开放式讨论会上发展改进,但并不意味着无限制地使用它。

所维护的信息被视为需要长期保存,即使 OAIS 本身并不是永久的。长期对于关心技术变化的影响来说是足够长的,包括对新的媒体和数据格式的支持或不断变化的用户社区。长期可无限延长。在此参考模型中,人们尤其关注数字信息(既作为持有信息的原形,也作为数字和物理档案材料的支持信息)。因此,该模型兼容固有非数字信息(如,一个物理样本),但是这些信息的建模和保存并未详细涉及。本参考模型:

- 提供了长期数字信息保存和使用所必需的档案概念的理解与增强意识的框架;
- 提供了非档案机构有效参与保存过程所需的概念;
- 提供了描述和比较现存与未来档案的建立与操作的框架(包括术语和概念);
- 提供了描述和比较不同长期保存的策略和技术的框架;
- 提供了比较由档案保存的数字信息的数据模型以及探讨数据模型和潜在信息的时移变化的基础;
- 提供了通过其他努力可扩展至覆盖非数字化形式信息(如物理媒体和物理样本)的长期保存框架;
- 延伸了对长期数字信息保存与使用的共识,并扩大了厂商可支持的市场范围;
- 指导了与 OAIS 相关标准的识别与产生。

该参考模型涉及了全套档案信息保存功能,包括摄取、档案存储、数据管理、访问和传播。它还涉及数字信息到新媒体和格式的迁移、用于表示信息的数据模型、信息保存中的软件作用以及档案之间数字信息的交换。它能够识别归档功能的内部和外部接口,并且它可识别这些接口中的一系列高层次服务。它提供了各种说明性例子以及一些“最佳实践”建议。它定义了称为 OAIS 的档案的责任极小集,还定义了提供一系列有用术语和概念的最大档案。(编译自:Reference Model for an Open Archival Information System (OAIS).

<http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf>. [2009-08])

(岳增慧编译, 吴贝贝校对)

## 2、ISO 14721:2003 空间数据与信息转移系统——开放档案信息系统(OAIS)——参考模型

ISO 14721:2003 为开放档案信息系统(OAIS)定义了参考模型。ISO 14721:2003 旨在建立一个档案信息系统(数字的及物理的), 拥有由负责保存信息并为指定团体服务的人组成的组织方案。

该参考模型涉及了全套档案信息保存功能, 包括摄取、档案存储、数据管理、访问和传播。它还涉及数字信息到新媒体和格式的迁移、用于表示信息的数据模型、信息保存中的软件作用以及档案之间数字信息的交换。它能够识别归档功能的内部和外部接口, 并且它可识别这些接口中的一系列高层次服务。它提供了各种说明性例子以及一些“最佳实践”建议。它定义了称为 OAIS 的档案的责任极小集, 还定义了提供一系列有用术语和概念的最大档案。

ISO 14721:2003 中描述的 OAIS 模型可适用于任何档案。它特别适用于负责信息长期可用的组织。这包括了履行其他职责的组织, 如与规划需要相对应的处理和传播。

(编译自: ISO 14721:2003 Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model.

[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=24683](http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683). [2012-08-21])

(岳增慧编译, 吴贝贝校对)

## 3、CERA-2

CERA-2(气候与环境检索与档案-2)标准的目标是使气候和环境数据能够存储在不同的地理位置, 该机构有能力维护这些数据, 能保障数据的质量并为数据的应用提供科学的技术性建议, 确保在线数据传播和长期保存。CERA-2 还致力于构建生动的用户界面, 使用户能够使用一个虚拟的数据库系统(该系统由不同的、交互操作的数据库系统构成)。同时 CERA-2 还发展共性较高的数据模型来解决数据映射问题。遵循 IEEE 的参考模型, 元数据应该支持四种不同的科研数据管理的界面: 浏览、搜寻和检索, 吸取、质量保证和再加工, 应用到应用的转化, 存储和归档。

(编译自: CERA-2.

[http://www.pik-potsdam.de/cera/Descriptions/Publications/Papers/9807\\_DKRZ\\_TechRep15/cera2.pdf](http://www.pik-potsdam.de/cera/Descriptions/Publications/Papers/9807_DKRZ_TechRep15/cera2.pdf). [1998-07])

(吴贝贝编译, 王妍校对)

#### 4、欧洲社会科学数据档案理事会 (CESSDA) 目录

- 给来自欧洲社会科学数据档案的数据集提供一个无缝接口
- 可以通过顶部搜索框进行自由文本搜索
- 可以通过左侧菜单中的选项进行浏览
- 可以用九种语言中的任何一种进行查看, 默认语言依赖于计算机用户的区域设置
- 可以从上面的下拉列表中通过选择相应的语言在任何时候切换语言。请注意: 改变语言会将用户带回这个 CESSDA 目录主页
- 在自由文本搜索中:
  1. 支持\*截断
  2. 支持 ‘?’ 通配符

(编译自: CESSDA Data Catalogue.

Council of European Social Science Data Archives.

<http://www.cessda.org/accessing/catalogue/>. [2013-01-16])

(王妍编译, 岳增慧校对)

#### 5、CF 元数据与天气和预测元数据公约

天气和预测 (CF) 元数据公约旨在促进文件的处理与共享, 该文件是与 NetCDF API 创建的。CF 公约逐渐地得到认可, 并已被许多项目和组织采用作为首选标准。公约定义了元数据, 它提供了在每种变量中数据表示的是什么以及数据的时空属性的明确描述。这促使不同来源的数据的用户能够决定哪些数量是可比的, 并且利用强大的提取、重画网格和展示功能来构建应用程序。

(编译自: CF Metadata.

Climate and Forecast Metadata Convention.

<http://cf-pcmdi.llnl.gov/>. [2013-01-16])

(王妍编译, 岳增慧校对)

#### 6、Gesis——莱布尼兹社会科学研究所

GESIS 是德国提升科学研究服务的最大基础研究所。其服务可根据研究数据周期从用户的视角进行分类。这一周期由分析至发布阶段的数据和信息研究以及研究计划构成, 包括实证研究数据收集和存储、准备、存档以及研究数据的检索。

(编译自: GESIS – Leibniz-Institute for the Social Sciences.

<http://datacite.org/GESIS>. [2013-02-12])

(岳增慧编译, 吴贝贝校对)

## 7、数据批准印章 (DSA)

数据批准印章确保在未来没有必须的新起点、规定或高成本, 研究数据仍然可以以一种高质量和可靠的方式被处理。数据批准印章和它的质量方针也许是研究机构、归档数据的组织以及数据的用户感兴趣的。它可以通过评估过程被授予给任何申请它的库。

任何人把他或她的数据存档, 都可以在将来查找、识别和使用它。所有的硬件、软件都一直在改变, 不能将其理所当然地视为电子数据。使数据永不过时可以通过确保数据集和元数据符合某些请求来实现。

(编译自: Data Seal of Approval (DSA)).

<http://www.datasealofapproval.org/?q=about>. [2013-01-16])

(王妍编译, 岳增慧校对)

## 8、数据批准印章

### 数字研究数据的质量指南

#### 引言

此文件包含共有 16 个质量方面的应用和验证的指南, 该质量方面与社会科学和人文中数字研究数据的创建、存储以及再利用有关。这些指南为被数据批准印章委员会授予“数据批准印章”提供基础。

#### 目标

在此文件中制定的质量指南对于研究者、创建数字研究文件的机构、将研究文件归档的组织以及研究数据的用户来说是感兴趣的。

数据批准印章的目标是保护数据, 在将来无需新标准的实现、规范或高成本, 而确保高质量、可靠地指导研究数据的管理。

#### 批准印章

- 给予研究者保证, 即他们的研究结果会以一种可靠的方式被储存并且可以被再利用
- 提供研究资助, 保证研究结果能够再利用
- 确保研究者以一种可靠的方式来评估保存有研究数据的库
- 允许数据库来有效地归档和分发研究数据

#### 关于指南



下列指南的基础是五个标准，它们一起决定数字研究数据是否能够成为可持续存档：

- 研究数据可以在网上找到
- 研究数据是可获取的，然而考虑到与个人信息或数据的知识产权相关的立法
- 研究数据在一种可用的格式中是有效的
- 研究数据是可靠的
- 研究数据可以被引用

这些标准会在指南的以下部分被论述。

相关指南与这些标准的实现有关，并且聚焦于 3 个利益相关者：数据生产者、数据库和数据使用者。

1. 数据生产者对数字研究数据的质量负责
2. 数据库为储存的质量和数据的有效性：数据管理负责
3. 数据使用者为数字研究数据的使用质量负责

(编译自：Data Seal of Approval.

Quality guidelines for digital research data.

[http://www.datasealofapproval.org/sites/default/files/DSA%20booklet\\_2-0\\_engels\\_mei2010.pdf](http://www.datasealofapproval.org/sites/default/files/DSA%20booklet_2-0_engels_mei2010.pdf).

[2013-01-16])

(王妍编译，岳增慧校对)

## 9、用任意查询收割 DataCite 元数据

大约一年以前我们发布了 OAI—PMH 服务器，选择要收割的子集的方法是通过每个 DataCite 成员以及他们的每个数据中心的个体集。例如，仅仅收割大英图书馆（设置“BL”）或大陆（设置“TIB. PANGAEA”）是可能的。

我们很高兴地宣布，现在您几乎可以通过任何自定义查询进行选择性地收割。例如，它允许你仅收割数据集、按语言、出版年或在我们搜索服务中任何可用过滤器组合进行过滤。在 DataCite 元数据搜索中建立查询。在结果的底部有 OAI—PMH 链接。

能够尽可能地成为完全 OAI—PMH 兼容，我们给予了极大关注。我们相信每个 OAI 客户可以使用这个新特性。我们编码搜索参数并把它用作为 setname。

(编译自：Harvesting DataCite metadata with arbitrary queries.

<http://www.datacite.org/node/68>. [2012-08-06])

(王妍编译，岳增慧校对)

## 10、杜威十进制图书分类法

杜威十进制图书分类法 (Dewey Decimal Classification) 是由美国图书馆专家麦尔威·杜威发明的, 对世界图书馆分类学有相当大的影响, 已翻译成西班牙文、中文、法文、挪威文、土耳其文、日文、僧伽罗文、葡萄牙文、泰文等出版, 并被许多英语国家的大多数图书馆、以及使用其它相应译文之国家的部分图书馆采用。在美国, 几乎所有公共图书馆和学校图书馆都采用这种分类法。

杜威十进制图书分类法于 1876 年首次发表, 历经 22 次的大改版后, 内容已有相当程度的修改与扩充。最新的版本为 2004 年版。该分类法以三位数字代表分类码, 共可分为 10 个大分类、100 个中分类及 1000 个小分类。除了三位数分类外, 一般会有两位数字的附加码, 以代表不同的地区、时间、材料或其他特性的论述, 分类码与附加码之间则以小数点“.”隔开。

杜威十进制图书分类法的 10 个大分类:

- 000 - 计算机科学、资讯与总类
- 100 - 哲学与心理学
- 200 - 宗教
- 300 - 社会科学
- 400 - 语言
- 500 - 科学 (指自然科学)
- 600 - 技术应用科学
- 700 - 艺术与休闲
- 800 - 文学
- 900 - 历史、地理与传记

(编译自: Dewey Decimal Classification.

[http://en.wikipedia.org/wiki/Dewey\\_Decimal\\_Classification](http://en.wikipedia.org/wiki/Dewey_Decimal_Classification). [2013-02-12])

(岳增慧编译, 吴贝贝校对)

## 11、DDI——数据文档计划

数据文档计划 (DDI) 致力于创建一个描述社会、行为以及经济科学数据的国际标准。以 XML 形式表达, DDI 元数据规范现在支持整个研究数据生命周期。DDI 元数据伴随并使得数据能够概念化、收集、加工、传播、发现、分析、更新和归档。其主要特点是: 围绕数据生命周期记录数据, 可以与其它用户互操作, 可以生成智能数据。



当前 DDI 的实现以及其相对不同应用的效益，见 DDI at Work。

为确保标准的持续支持与不断发展，DDI 分支为两个独立发展线。这两个分支：DDI-Codebook（以前的 DDI2）和 DDI-Lifecycle（以前的 DDI3）；在规范页进行了更深度地解释。想要获得更多 DDI 入门信息，详见网站中大量的资源（包括常见问题和邮件列表）。（编译自：DDI -- The Data Documentation Initiative.

<http://www.ddialliance.org/what>. [2012-08-19])

（岳增慧编译，吴贝贝校对）

## 12、DOI 系统

这是国际 DOI 基金会（IDF）的网站，它在 DOI（数字对象标识符）系统和活动中提供信息。DOI 系统为注册和永久互操作标识符在数字网络上的使用提供了一个技术性和社会性基础设施。DOI 系统实现了处理系统和 indecs 框架。

IDF 是提供 DOI 服务和注册的登记机构联合会的治理和管理主体，是 DOI 系统的 ISO 标准（ISO 26324）注册机构。

（编译自：DOI System.

<http://www.doi.org/index.html>. [2013-01-08])

（王妍编译，岳增慧校对）

### 13、ISO 26324: 2012 信息和文献——数字对象标识符系统

ISO 26324:2012 详细说明了数字对象标识符系统的句法、描述和辨识功能组件, DOI 名称的创建、注册以及管理的普遍准则。

ISO 26324:2012 定义了 DOI 名称的句法, 它是被用来识别任何物质形式的对象或有需求将它与其他对象区分的抽象物。

DOI 名称不能代替, 也不能成为在另一个模式中被使用的标识符的备选, 例如被 ISO/TC 46/SC 9 定义的模式。ISO 26324:2012 描述了如何被用在与另一个标识符模式的结合中, 以及通过 DOI 元数据记录和/或 DOI 句法将其他模式的字符串被集成到 DOI 系统。

ISO 26324:2012 并没有详细说明实现数字对象标识符系统的句法、描述和辨识功能组件的特定技术。

(编译自: ISO 26324:2012 Information and documentation -- Digital object identifier system.

[http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=43506](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43506). [2012-04-23])

(王妍编译, 岳增慧校对)

### 14、基于风险评估的数字存储库审计方法 (DRAMBORA)

数字内容管理中心 (DCC) 和欧洲数字保存 (DPE) 很高兴地宣布基于风险评估数字存储库审计方法工具包的发布。此工具包的目的是通过为存储库管理者提供一种评估他们能力、辨别缺点, 认识优点的方法来促进内部审计。数字存储库还处于初级阶段, 此模型能够适应快速发展。该工具包的发展遵循一个存储库试点审计的集中段, 该试点审计由 DCC 承办, 被各种各样的组织引导, 包括国家图书馆、科学数据中心以及文化遗产数据档案。

(编译自: Digital Repository Audit Method Based on Risk Assessment (DRAMBORA).

<http://www.repositoryaudit.eu/>. [2008-02-01])

(王妍编译, 岳增慧校对)

### 15、都柏林核心——都柏林核心元数据计划

都柏林核心元数据计划 (DCMI) 是一个开放组织, 支持元数据设计与元数据生态最佳实践的创新。DCMI 的活动包括架构和建模工作; DCMI 社区与 DCMI 任务组、全球会议与研讨会的讨论与协作工作; 以及促进元数据标准和最佳实践获得普遍认可的教育努力。DCMI 维持了许多与标准机构及其他元数据组织的正式与非正式联络与关系。

都柏林核心元数据计划 (DCMI) 支持元数据设计与拥有广泛用途和商业模式下的元数据最佳实践的共享创新。

DCMI 通过以下方式实现上述目的:

- DCMI 规范和元数据术语名称空间的长期管理与发展;
- 现有全部 DCMI 工作主题的持续讨论;
- 建立和管理国际和地区性的事件;
- 可用会议资产 (包括会议记录、项目报告和会议纪要) 的管理和公开;
- 元数据最佳实践中训练资源 (包括教程、在线研讨会和讲习班) 的创建与传递;
- 协调全球 DCMI 志愿者社区。

DCMI 的工作原则是:

- 开放共识基础
- 国际视野与参与
- 目的用途与商业模式的中立
- 技术中立
- 交叉学科聚焦

(编译自: Dublin Core -- The Dublin Core Metadata Initiative.

<http://dublincore.org/>. [2013-02-12])

(岳增慧编译, 吴贝贝校对)

## 16、欧洲语言社会科学叙词表 (ELSST)

ELSST 是一个广泛的针对社会科学的多语言叙词表。最初 ELSST 是基于英国数据档案的 HASSET 叙词表, 多年来它已被 CESSDA 成员开发。它的主要目标是促进整个欧洲数据资源的获取途径, 独立于域、资源、语言和词汇。ELSST 用于在 CESSDA 目录中帮助检索。ELSST 的源版本是英语叙词表。此叙词表目前被翻译成 8 种其他语言: 丹麦语、芬兰语、法语、德语、希腊语、挪威语、西班牙语和瑞典语。

(编译自: European Language Social Science Thesaurus (ELSST).

<http://www.cessda.org/accessing/index.html>. [2013-01-16])

(王妍编译, 岳增慧校对)

## 17、Genericode——定义代码列表的一种标准格式

Genericode 开始作为一个项目来定义代码列表 (也称为枚举或受控词表) 的一种标准格式。代码列表贯穿 IT (信息技术) 的所有部分, 但没有一个标准的方式来管理它们。这

意味着每一个新的标准都致力于创建其自己的新解析方案, 这导致了代码列表管理的割据。

Genericode 旨在提供如下内容:

- 代码列表内容的一个标准模型以及 XML 表示;
- 与代码列表项相关的数据的标准模型以及 XML 表示;
- 现有代码列表如何产生新代码列表的一个标准模型以及 XML 表示。

最后一点将 genericode 与仅提供一个代码列表特定版本内容的静态快照的解析方案区别开来。Genericode 不仅展现了代码列表项, 还提供了代码列表如何与其以前版本或其他代码列表相关联的跟踪索引。这简化了对一个新的代码列表版本与以前版本之间区别的理解, 也简化了此变化对现有系统及进程的影响的计算。

(编译自: Genericode -- A standard format for defining code lists.

<http://www.genericode.org/>. [2008-02-16])

(岳增慧编译, 吴贝贝校对)

## 18、地球科学的地面欧洲网络互操作——数字存储库 (GENESI-DR)

GENESI-DR 要为欧洲和全球科学用户无缝地访问和共享来自太空、空中和现场传感器中的所有数据、信息、产品和知识建立开放的地球科学数字存储库的访问途径。GENESI-DR 应该操作、验证和优化集成的访问途径, 并使用有效的数字数据存储库来展示欧洲如何能够最好地应对与地球状况相关的新兴全球需求, 一个目前尚未满足的需求。

(编译自: Ground European Network for Earth Science Interoperations-Digital Repositories (GENESI-DR).

[http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ\\_RCN=10925676](http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=10925676). [2010-04-30])

(王妍编译, 岳增慧校对)

## 19、句柄系统与 CNRI——美国国家研究创新联合会

### 句柄系统

句柄系统为唯一、持久的数字对象标识提供了高效的、可扩展的、安全的解析服务, 且它是 CNRI 的数字对象框架的组成部分。数字对象框架提供了一种在网络环境中管理数字信息的方法。一个数字对象拥有可被识别、访问和保护的身份以及独立平台结构。一个数字对象不仅可包含信息元素 (如论文的电子版、电影或录音), 也包含数字对象及其他有关数字对象元数据的唯一标识符。元数据可包括数字对象访问限制、版权声明、许可协议标识符。

句柄系统包括一组开放协议、一个命名空间以及一个参考实现协议。该协议使得分布式

计算机系统能够存储被称为句柄的任意资源标识符, 并能将句柄解析成为定位、访问、接触、认证或其他利用资源所需的信息。该信息可根据需要而改变, 以在不改变标识符的情况下反映被识别资源的当前状态, 因此允许项名在位置及其他相关状态信息的变化时而保持不变。一些将 HDL 标识符和解析服务作为基础设施的应用实例有: 权限管理应用程序、数字对象注册中心和存储库、机构数据保存和归档。

### **CNRI——美国国家研究创新联合会**

美国国家研究创新联合会 (CNRI) 是一个非营利组织, 旨在从事、培养和促进公共利益研究。基于网络的信息技术战略发展中心提供了信息基础设施研究与发展的领导和资助。

CNRI 从事系统与技术示范项目, 使选定的 (新的基于计算和通信应用) 基础设施组件的进一步设计和实现以开发新的基于计算机和交流的应用程序。目前的项目包括设定、管理和解析数字对象及其他网络资源的持久标识符的句柄系统, 还包括 MEMS 与纳米技术交流、口语对话系统研究以及支持数字图书馆和网络信息技术研究的其他活动。

(编译自: Handle System.

CNRI -- Corporation for National Research Initiatives.

<http://handle.net/>. [2013-02-12]

<http://www.cnri.reston.va.us/>. [2013-02-12])

(岳增慧编译, 吴贝贝校对)

## **20、ICPSR——美国校际社会科学数据共享联盟**

ICPSR (大学间政治与社会研究联盟) 成立于 1962 年, 为数据访问、管理和针对大范围多样性的社会科学研究社区方法分析提供向导和训练。ICPSR 数据档案在其广度和深度上是无可比拟的, 其数据保存涵盖了一系列学科, 包括政治学、社会学、人口学、经济学、历史、教育、老年学、刑事审判、公共卫生、外交政策、健康和医疗保健、教育、儿童保健研究、法律、药品滥用等。ICPSR 也举办一些专注于特定学科或主题的赞助项目, 鼓励各领域的社会科学家在 ICPSR 归档数据。

为方便研究人员收集社会科学方面的数据, ICPSR 为其提供一系列专业化服务, 具体如下:

- 为便于保存和传播, ICPSR 不仅接受定量数据, 同时也接受定性研究数据 (包括成绩单和视听媒体); ICPSR 专注于数字保存, 鼓励研究人员将其数据存放于新兴格式, 如网站、地理空间数据、生物医学数据和数字视频。关于如何保存你项目相关的任何类型的数据内容或格式请联系 ICPSR。
- ICPSR 电子保存形式有利于文件在网页上的安全上传, 使存件人能够描述所存数据

收集。存件人可以指定他人访问文件，并添加信息，如若需要，可以再多个时间段完成。

- ICPSR 为每个数据集合分配一个持久的标识符。
- 数据集文档的 ICPSR 标准运用变量级 DDI XML 标记，这使得搜索更加精确。
- ICPSR 可以协助传播那些为保护人类主体而需特殊处理及一定限制性的数据。受限制性的数据集需要特殊的应用过程（如数据保护计划，IRB 批准等）。
- ICPSR 拥有一个安全的数据飞地以存储并保护那些具有最高机密标准的数据。这些存储于数据飞地中的数据可以在我们现场监控的计算机设备中得以分析。同时，我们正使用一个虚拟的数据飞地，允许对仍存在于 ICPSR 服务器中的限制性数据进行远程的安全访问。
- ICPSR 可以便于存储延迟传播的数据，但前提是数据的释放会对保密构成风险，或者同时有其他的数据传播方案。
- ICPSR 要求所有处理数据的员工都得接受特殊培训，以获得处理涉及保密问题的敏感数据方面的证书。
- 2011 年 6 月，ICPSR 成为首批六个获得数据批准印章的数据存储库。数据批准印章是一个国际标准，可证明档案采取适当措施保证其所含数据的质量及长期可得性。
- 研究人员可以通过在线数据分析系统、调查文档和分析（SDA）进行简单或复杂的分析及重新编码，计算新变量、子集变量或下载情况。
- ICPSR 可以创建专门网页提升数据传播。对于复杂研究，ICPSR 通过创建用户指南来帮助新用户使用数据。这可以通过网站联通其他在线功能实现，如常见问题列表（FAQs）或电子邮件列表。
- 文档准备培训和数据文档计划（DDI）可由 ICPSR 员工提供。学习如何创建 DDI 兼容的文档，包括 XML 码簿，可以提高数据再项目阶段的可用性，也可以在项目完成阶段更好地存储数据。
- ICPSR 员工可以为其他学者提供培训，在社会研究的定量研究方法的暑期项目期间，通过提供专业的数据分析讲座使其学会如何分析数据。当然也可以采用网络讲座的形式。
- ICPSR 为正开展保存项目的管理者提供数字保存培训。
- ICPSR 已获得相应资金，提高其获取、存档、并安全地流式传输视频研究数据的能力。这还需要新的基础设施和程序，目前也正在开发之中。

(编译自 :ICPSR -- Inter-university Consortium for Political and Social Research.  
<http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>. [2012])

(岳增慧编译，吴贝贝校对)



## 21、ICPSR 数字保存政策框架

通过为研究者获取过去和当今社会科学研究成果，ICPSR 很好地发挥了其应有的作用，保存了社会科学的优秀成果。数字保存政策框架有利于更好地发挥其这一作用，代表了 ICPSR 层面数字保存政策文档的最高水平，并使 ICPSR 在通过可理解的数字保存项目的发展和演化来保存数字资产的集合形式方面的作用更加明确。数字保存政策框架反映了 ICPSR 战略计划所定义的目标，同时也包含与其他相关 ICPSR 政策和程序的引注。这一框架的潜在用户包括 ICPSR 成员、员工、数字内容储户、投资者以及用户等。

此外，数字保存政策框架还涉及一个数字保存社区的实际标准：《可信数字存储库：属性和责任》(PDF)。这一框架的组织反映了一个可信数字存储库的七个属性：

- OAIS（开放档案信息系统）依从性
- 行政负责性
- 组织可行性
- 财务可持续性
- 技术和程序的适合性
- 系统安全性
- 程序问责性

(编译自：ICPSR Digital Preservation Policy Framework.

<http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/dpp-framework.html>. [2013-02-12])

(岳增慧编译，吴贝贝校对)

## 22、ISCO——国际标准职业分类

ISCO（国际标准职业分类）是国际劳工组织负责的一个主要的国际分类，隶属于国际经济和社会分类。

ISCO 是根据当前工作中所承担的任务和责任将工作组织为一个定义清晰的工具，其主要目标是提供以下服务：

- 为关于职业的国际报告、比较、统计和管理数据的交流提供基础；
- 为国家和汲取职业分类的发展创立模型；
- 为尚无国家职业分类的国家提供直接可用的职业分类体系。

ISCO 主要用于统计应用及各种面向客户的应用中。面向客户的应用包括：求职者与空缺职位的匹配；国家间长期或短期移民劳工的管理；职业培训项目开发和指导。

首版 ISCO 于 1957 年由第九届国际劳动统计学家会议批准通过，也称作 ISCO-58。此版 ISCO 此后被 1966 年第十一届国际劳动统计学家会议通过的 ISCO-68 所取代。第三版 ISCO-88

于 1987 年由第十四届国际劳动统计学家会议通过。目前很多国家职业分类均是基于以上三版 ISCO。

最近, ISCO 已经更新, 囊括了 1988 年以来世界范围内工作的发展, 并根据采纳 ISCO-88 以来的经验进行了改进。但这种更新并未改变 ISCO-88 的基本原则和结构, 但有一些重要的结构调整。更新后的国家职业分类为 ISCO-08, 于 2007 年通过, 当前很多国家正是基于 ISCO-08 更新其国家职业分类, 或采用新的国际统计标准改进调整。

采用 ISCO-08, 而分类结构和对应表采用 ISCO-88 的方案在 ISCO 网站上有英语、法语和西班牙语几种版本。对 ISCO-08 群组的最终定义目前在 ISCO 官网上仅有英语一种语言。

结构、定义、对应表以及简介基本概括了更新过程、所用的方法和概念模型, 描述了 ISCO-88 和 ISCO-08 的主要差异, 最终将以书卷 ISCO-08 卷一的形式发布。

英语版最终定稿后, 法语和西班牙语版本也将尽快发布。

(编译自: ISCO -- International Standard Classification of Occupations.

<http://www.ilo.org/public/english/bureau/stat/isco/index.htm>. [2010-06-09])

(岳增慧编译, 吴贝贝校对)

### 23、ISO 639-1:2002 语种名称表示代码——第一部分: α-2 代码

本部分 ISO 639 提供了包含语种代码元素的代码, 该单元由表示语种名称的双字母标识符组成。本部分 ISO 639 的语种标识符最初被设计用于术语、词典编纂和语言学, 也可被用于要求以双字母码形式进行语种表示的应用中, 特别是计算机系统。α-2 代码被设计实际使用于世界大多数主要语言, 不仅包括全世界文学作品中最常用的语言, 也包含了大量专业语言和术语。当存在大量以专业语言和术语编写的文档时, 将创建额外的语种标识符。专为机器使用所设计的语言, 如计算机编程语言, 并不包含在此代码中。

(编译自: ISO 639-1:2002 Codes for the representation of names of languages -- Part 1: Alpha-2 code.

[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=22109](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22109). [2013-01-15])

(岳增慧编译, 吴贝贝校对)

### 24、国家代码——ISO 3166

#### 何为 ISO 3166?

ISO 3166 是国家代码以及其分支代码的国际标准, 其目的是为国家、地区、重要的地理区域及内部区域名称的表示建立国际认可的代码。然而, ISO 3166 并不是为这些国家命

名,而是规定代码来代表国家。

ISO 3166 中的国家名称均取自联合国成员。当联合国在《术语通报国家名称》或在《统计用国家和地区代码》(联合国统计部门持有)发布新的名称时这些信名称和代码便会自动更新添加。内部区域名称取自对应国家的相关官方信息。

ISO 3166 首次出版于 1974 年,是建立国家代码的唯一标准。1977 年扩展为三部分,还包括内部区域代码和国家名称代码(当前已停止使用)。这三个部分中,第一部分,ISO 3166-1 最为常用。

#### 如何维护 ISO 3166?

ISO 3166 的三个部分是由专门的维护机构(ISO 3166/MA)进行维护,ISO 3166/MA 的构成反映了两个利益相关群体:国家标准组织,ISO 成员和联合国机构。这两个利益群体主要参与了 20 世纪 70 年代早期 ISO 3166 的发展。

对 ISO 3166/MA 拥有投票权的十名专家中有五人以下是国家标准组织的代表:

- 法国标准化协会(AFNOR)
- 美国国家标准协会(ANSI)
- 英国标准学会(BSI)
- 德国标准化学会(DIN)
- 瑞典标准协会(SIS)

联合国或其他国际组织的其他主要的五个代表(同时也是 ISO 3166-1 用户)为:

- 国际原子能组织(IAEA)
- 国际电信联盟(ITU)
- 互联网名称和编号分配公司(ICANN)
- 万国邮政联盟(UPU)
- 联合国经济和社会委员会(UNECE)

此外,ISO 3166/MA 还有一些成员虽不参加投票,但通过自己的专业技能,对维护机构的程序决策有重大影响。

各成员可通过 ISO 3166/MA 的秘书处进行联系。

(编译自:Country Codes -- ISO 3166.

[http://www.iso.org/iso/country\\_codes.htm](http://www.iso.org/iso/country_codes.htm). [2013-02-12])

(岳增慧编译,吴贝贝校对)

## 25、ISO 19115:2003 地理信息——元数据

ISO 19115:2003 定义了描述地理信息和服务所需的模式架构。它提供了数字地理数据的识别、程度、质量、空间和时间模式、空间参照与分布的信息。

ISO 19115:2003 适用于:

- 数据集编目、交流活动以及数据集完整描述;
- 地理数据集、数据集系列以及个体地理特征与功能属性。

ISO 19115:2003 定义了:

- 强制性和条件性元数据部分、元数据实体和元数据元素;
- 服务于元数据全部应用(数据发现、确定数据适用性、数据访问、数据传输以及数字数据的使用)所需的元数据最小集;
- 可选元数据元素——以便考虑更广泛的地理数据标准描述, 如果需要的话;
- 一种扩展元数据以适应专门需求的方法。

尽管 ISO 19115:2003 适用于数字数据, 其准则可扩展至地理数据的许多其他格式, 如地图、图表以及非地理数据文本文档。

注意, 某些强制性元数据元素可能不适用于这些其他数据格式。

(编译自: ISO 19115:2003 Geographic information -- Metadata.

[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=26020](http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020). [2009-02-28])

(岳增慧编译, 吴贝贝校对)

## 26、ISO 20252:2006 市场、观点和社会研究——词汇和服务需求

ISO 20252: 2006 为从事市场、观点和社会研究的组织和专业人员建立术语、定义以及服务需求。此标准被修订为 ISO: 20252:2012。

(编译自: ISO 20252:2006 Market, opinion and social research -- Vocabulary and service requirements.

[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=39339](http://www.iso.org/iso/catalogue_detail.htm?csnumber=39339). [2012-05-22])

(王妍编译, 岳增慧校对)

## 27、ISO/IEC 11179, 信息技术——元数据注册 (MDR)

ISO/IEC 11179 标准是一个复合标准, 包括以下部分:

- 第一部分: 框架
- 第二部分: 分类
- 第三部分: 注册表元模型与基本属性
- 第四部分: 数据定义规则
- 第五部分: 命名与标识原则
- 第六部分: 注册

### 11179-1: 框架

本部分 ISO/IEC 11179 介绍和讨论了理解该标准集所必不可少的数据元素、价值域、数据元素概念、概念域和分类表的基本思想, 并提供了关联 ISO/IEC 11179 各部分的环境。

### 11179-2: 分类

本部分 ISO/IEC 11179 提供了一个管理分类表的概念模型。有许多组织分类表的结构, 也有许多分类表描述的主题领域。因此, 该部分还提供了一个分类表本身的双面分类。

### 11179-3: 注册表元模型与基本属性

本部分 ISO/IEC 11179 指定了元数据注册的概念模型, 以及当不需要一个完整的注册解析方案时的一套元数据使用基本属性。

### 11179-4: 数据定义规则

本部分 ISO/IEC 11179 提供了如何开发明确的数据定义的指导。ISO/IEC 11179-4 确切介绍了数据定义应该如何形成的一些特定的规则和指南。一个精确的、格式良好的定义是管理项共同理解的最关键需求之一; 格式良好的定义是对信息交换至关重要。只有当每个用户对数据项都有一个共同且准确的理解时, 数据才可进行可靠交换。

### 11179-5: 命名与标识原则

本部分 ISO/IEC 11179 提供了标识管理项的指导。标识是指定或识别一个特定数据项的一个宽泛的术语。标识可以不同的方式完成, 取决于标识符的使用。标识包括对人类无固有意义的数值标识符的赋值; 图标(含义已被分配的图形符号); 通常为了人类理解, 与数据项的定义和值域相关联的嵌入意义的名称。

### 11179-6: 注册

本部分 ISO/IEC 11179 提供了注册申请人如何在一个中央注册机关注册数据项以及如何为每个数据项分配唯一标识符的指示说明。本文档中也指定了已注册的管理项的维护。

(编译自: ISO/IEC 11179, Information Technology -- Metadata registries (MDR).

<http://metadata-standards.org/11179/>. [2012-10-06])

(岳增慧编译, 吴贝贝校对)

## 28、数字格式可持续性

数字格式网站提供关于数字内容格式的信息, 它与后续的网站将会持续记录数字格式的进展。它的目的是支持关于数字内容格式的战略计划, 确保美国国会图书馆数字内容的长期保存; 提供现存的和正在形成的格式信息的详细目录, 包括工具的识别和详细的文件, 该文件保证国会图书馆能管理在内容生命周期中由数字格式创建或接收的内容; 鉴别和描述适合长期可持续的格式, 为维护这些格式(包括附属于工具的推荐和管理所需的文件)制定战略; 鉴别和描述不适合长期可持续的格式, 为维护这些格式所容纳的内容制定战略。该网站

对一些数字内容格式进行了描述, 提供了适度的详细信息和引用。

(编译自: Sustainability of Digital Formats.

<http://www.digitalpreservation.gov/formats/>. [2013-02-08])

(吴贝贝编译, 王妍校对)

## 29、电子资源长时期保存元数据 (LMER)

对于电子文件的长期保存来说, 编制合适的技术元数据是至关重要的。到目前为止, 不幸的是, 并没有建立一个相关的元数据 XML 模式, 尤其是涉及长期保存。与 LMER, 德国国家图书馆提供了一个基于新西兰国家图书馆模型的模式。

(编译自: Long-term preservation Metadata for Electronic Resources (LMER).

[http://www.dnb.de/EN/Standardisierung/LMER/lmer\\_node.html](http://www.dnb.de/EN/Standardisierung/LMER/lmer_node.html). [2012-02-21])

(王妍编译, 岳增慧校对)

## 30、机读目录 (MARC)

MARC 21 格式是书目和机读形式的相关信息表示和通信的标准。MARC 记录包括三个要素: 记录结构、内容指示器以及记录的数据内容。MARC 21 是用来定义编码机读记录的一组代码和内容指示器。定义五种数据类型的格式有: 书目、持有、权限、分类和社区信息。MARC 21 格式由与不同用户社区的国会图书馆维护。

(编译自: Machine-Readable Cataloguing (MARC).

<http://www.loc.gov/marc/96principi.html>. [2013-02-12])

(王妍编译, 岳增慧校对)

## 31、气候模拟数字存储库的通用元数据

METAFOR 的主要目标是定义一个公共信息模型 (CIM), 它用一种标准方法描述气候数据以及产生数据的模型。METAFOR 将建立在现有元数据基础上, 目前用于现有的数据存储库, 并解决问题, 例如元数据片段、缺口或重复。在国际层面上与相关举措的密切互动中, METAFOR 将提出这些存储库中识别、获取和使用气候信息的解决方案。为了支持 CIM 的采用, METAFOR 将建立基于 CIM 的工具和服务, 允许特定的数据和模型在合作网站中被发现和比较。因此, METAFOR 将会优化用来存储知识的气候数据基础设施的方式, 从而为日益广泛的利益相关者增加主要研究数据值。

(编译自: Common Metadata for Climate Modelling Digital Repositories.  
[http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ\\_RCN=10049248](http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=10049248). [2011-02-28])

(王妍编译, 岳增慧校对)

### 32、海洋元数据互操作

MMI 通过加强数据的出版、发现、文档和获取来提升海洋数据的交换、整合和使用。海洋元数据互操作可以帮助用户更容易地找到、获取和使用其它数据集。如果用户创造了海洋科学数据集, MMI 可以更好地宣传、分配和复用数据以及与其他数据集结合。MMI 通过把复杂的元数据领域简化为具体的、易懂的指导来促进海洋科学领域的合作研究。MMI 正在开发网页应用程序和独立的工具使海洋数据系统实现复杂的交互。

(编译自: Marine Metadata Interoperability.  
<https://marinemetadata.org/aboutmmi>. [2013-02-18])

(吴贝贝编译, 王妍校对)

### 33、元数据编码和传输标准

元数据编码和传输标准 (METS) 是针对数字图书馆范围内的对象, 对于描述性的、管理性的和结构性的元数据进行编码的一个标准, 使用 W3C 的 XML 模式语言表达。这个标准由美国国会图书馆的网络发展和 MARC 标准办公室进行维护, 正在被当作数字图书馆联盟的行动来发展。

(编译自: Metadata Encoding & Transmission Standard.  
<http://www.loc.gov/standards/mets/>. [2012-11-29])

(吴贝贝编译, 王妍校对)

### 34、元数据对象描述框架

元数据对象描述框架 (MODS) 是美国国会图书馆网络发展和 MARC 标准办公室联合感兴趣的专家于 2002 年发展的。它有一套可用于多种目的尤其是用于图书馆应用程序的书目元素集。作为一个 XML 框架, MODS 目的是能够传播存在于 MARC21 记录中的筛选的数据, 同时能够创建原始资源的描述记录。MODS 包含 MARC 字段的一个子集, 使用基于语言的标签而不是数字标签, 在一些情况下重新组合来自 MARC21 书目格式的元素。MODS 由 W3C 的 XML 框架

语言进行表达。MODS 可以作为元数据编码和传输标准 (METS) 的一个扩展使用。

(编译自: Metadata Object Description Schema.

<http://www.loc.gov/standards/mods/mods-overview.html>. [2013-02-14])

(吴贝贝编译, 王妍校对)

### 35、NESTOR——德国数字资源长期保存专业技术网

NESTOR 是德国数字资源长期保存与使用专业技术网。该技术网络是所有关于长期数字保存事物的中心接点。其建设经过了两个项目阶段, 大部分项目成员以及一小部分新合作伙伴决定以合作联盟的形式继续该项工作。项目基础设施的建立一直在维护和扩大。

该合作联盟中的伙伴有:

- 德国国家图书馆
- 巴伐利亚州立图书馆
- 下萨克森州和大学图书馆
- 柏林洪堡大学
- 哈根公开大学
- 博物馆研究所/普鲁士文化遗产基金会
- 巴登-符腾堡州档案馆
- 巴登-符腾堡州图书馆服务中心
- 德国语言研究所
- 柏林电脑游戏博物馆
- Goportis
- PDF/A 技术中心

NESTOR 的行政管理办公室在德国国家图书馆。

该技术网络的建立源于图书馆、档案馆和博物馆日益面临数字出版物和其他数字资源需要安全、长期地存档。然而, 数字数据本质上是不稳定的和濒临灭绝的。长期储存和供应(长期保存)在技术上是可能的——但昂贵的。良好的规划和持续的监测对于确保数字数据保持永久访问来说是必要的。

如果个体机构能够利用积累的知识, 已被证实有效的策略和成功测试的工具, 那么它们可以组织长期保存。NESTOR 生成和收集这方面的知识, 并使其可被用于有关机构。

来自各种不同机构的代表就 NESTOR 工作组中数字保存的特定方面进行协商。五个工作组目前致力于以下主题:

- 媒体 (视听材料和多媒体)
- 合作长期保存 (协作的角度)



- 法律 (数字保存的法律方面)
- 数字库存保存 (数字档案组织、保存规划)
- 仿真
- 成本
- 认证

(编译自 : NESTOR -- Competence Network For Digital Preservation.

[http://www.dnb.de/EN/Wir/Kooperation/nestor/nestor\\_node.html](http://www.dnb.de/EN/Wir/Kooperation/nestor/nestor_node.html). [2013-01-30])

(岳增慧编译, 吴贝贝校对)

### 36、NetCDF (网络通用数据格式)

NetCDF (网络通用数据格式) 是一组 面向阵列的数据访问接口, 以及 C、Fortran、C++、Java 和其他语言数据访问库的自由分布集。NetCDF 库支持表示科学数据的与机器无关的格式。同时, 该接口、库和格式支持科学数据的创建、访问和共享。

NetCDF 数据是:

- 自描述的。一个 NetCDF 文件包含了其所含数据的信息。
- 便携的。一个 NetCDF 文件可被计算机以存储整数、字符和浮点数等不同方式进行访问。
- 可伸缩的。一个大数据集的一个小子集可以被有效地访问。
- 可附加的。数据可被附加到一个结构合理的 NetCDF 文件, 而不需要复制数据集或重新定义其结构。
- 可共享的。一个作者与多个读者可同时访问同一 NetCDF 文件。
- 可归档的。该软件的当前和未来版本可支持访问所有 NetCDF 数据的早期格式。

(编译自 : NetCDF(Network Common Data Form).

<http://www.unidata.ucar.edu/software/netcdf/docs/faq.html#whatisit>. [2013-02-12])

(岳增慧编译, 吴贝贝校对)

### 37、OASIS 编码列表表达技术委员会

OASIS 编码列表表达技术委员会的目的是为编码列表创建一个 XML 的相互交换格式。这个格式将支持实际的编码 (包括可选择的编码和混合的多方编码) 和适合文档与管理的每个编码的任意相关信息。格式通过对一套输入编码列表如何转换成一套输出编码列表进行不明确的规范来支持派生编码列表的自动生成。技术委员会提出了一套要求, 一个 W3C 的 XML

框架, 一个 UML 模型, 文件和测试例子 (支持编码列表的衍生功能)。

(编译自: OASIS Code List Representation TC.

<https://www.oasis-open.org/committees/codelist/charter.php>. [2013-02-18])

(吴贝贝编译, 王妍校对)

### 38、地理和环境数据发布网络

PANGAEA 信息系统是一个开放获取图书馆, 目标是归档、出版和分配来自地球系统研究的地理信息数据。大部分数据可以免费获取并且可以在数据集描述的权限内使用。每个数据集的描述是可视的并包含主要负责人。通过使用数字对象唯一标识符可以鉴别, 分享, 发行和引用每个数据集。数据作为对出版物的补充或可引用的数据集进行存档。海洋环境科学世界数据中心 (WDC—MARE) 使用 PANGAEA 作为数据档案馆和分布系统, 同时还是地球系统科学数据 (ESSD) 期刊指定的档案保存处。

(编译自: PANGAEA Publishing Network for Geoscientific & Environmental Data.

<http://www.pangaea.de/about/>. [2013-02-18])

(吴贝贝编译, 王妍校对)

### 39、保存元数据: 实施战略

PREMIS 保存元数据数据词典是一个元数据国际标准, 它支持数字对象的保存, 确保他们的长期可用性。PREMIS 是由专家组成的国际团队开发, 是在世界范围内的数字保存项目中实现, 对 PREMIS 的支持具体表现为大量的商业性和开源的数字保存工具和系统。PREMIS 编辑委员会协调标准的修订和完成, 包含数据词典、一个 XML 框架和支持文档。

(编译自: Preservation Metadata: Implementation Strategies.

<http://www.loc.gov/standards/premis/>. [2013-01-17])

(吴贝贝编译, 王妍校对)

### 40、持续的统一资源定位符

持续的统一资源定位符 (PURLs) 是面临动态变化的网络基础实施时充当永久标识符的网络地址。PURL 工具包受 OCLC 研究办公室和 Zepheira 的积极参与影响, 为识别和管理分散的资源定义一种网络结构, 该分散资源用于满足各种商业和信息管理需求。PURLs 不是直接解析成网络资源, 它提供了一种间接的方式, 允许资源的潜在网络地址可以随着时间变化而不对其所依赖的系统产生消极影响。这种能力为网络资源的参考提供了持续性, 能够因为

商业、社会或技术原因由一个机器移动到另一个机器。

(编译自: Persistent Uniform Resource Locators. Träger OCLC (Online Computer Library Center) und Zepheira.

<http://purl.org/>. [2013-02-18])

(吴贝贝编译, 王妍校对)

## 41、观看标准：元数据世界的可视化

元数据景观的可视化地图的目的是帮助计划者挑选和执行元数据标准。地图上列出的105个标准中的每一个根据应用的强度(在四个轴中定义分类: 团体、范围、功能和目的)来评价。在给定的分类中一个标准的强度由在分类中被采纳的混合指标来决定, 包括标准的设计意图、在分类中总体的使用适合度。地图中的标准都是在文化遗产群落中被密集使用或宣传的。在可视化中标绘的一小部分元数据标准也在绘图中作为显著部分呈现。这些代表文化遗产元数据中被人熟知或最常被探讨的标准。

(编译自: Seeing Standards: A Visualization of the Metadata Universe.

<http://www.dlib.indiana.edu/~jenlrile/metadatamap/>. [2013-02-18])

(吴贝贝编译, 王妍校对)

## 42、元数据标准的词汇表

元数据标准的词汇表, 由印第安纳大学白色专业发展奖金支持, 汇集了众多的元数据标准, 按字母顺序进行排列, 对每条元数据标准给出链接和简介。

(编译自: Glossary of Metadata Standards.

[http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards\\_glossary\\_pamphlet.pdf](http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf). [2013-02-18])

(吴贝贝编译, 王妍校对)

## 43、欧洲航天局地球观察：欧洲的标准档案格式

SAFE 是欧洲航天局地球观察存档设施内保存和传输数据的通用格式。SAFE 标准目的是解决来自地球观察数据的包装和长期保存方面的主要挑战。SAFE 致力于与 ISO 14721:2003 OAIS: (开放档案信息系统) 参考模型和相关的标准例如 CCSDS/ISO XFDU (XML 格式化的数据单元) 包装格式相符合。

(编译自: Standard Archive Format for Europe within ESA Earth Observation.

<http://earth.esa.int/SAFE/>. [2011-07-11])

(吴贝贝编译, 王妍校对)

#### 44、ISO 17369:2013 统计数据与元数据交换(SDMX)

ISO 17369:2013 提供了一种集成的方法来促进统计数据与元数据交换(SDMX), 包括系统内部与系统之间的互操作实现, 涉及统计数据与相关元数据的交换、报告和传播。

ISO 17369:2013 适用于任何需要管理其统计数据与相关元数据的交换、报告和传播的组织机构。ISO 17369:2013 核心的信息模型已发展为支持政府和跨国统计机构的数据收集与使用, 并且该模型也适用于其他涉及统计数据与相关元数据的组织环境。

ISO 17369:2013 是 ISO/TS 17369:2005 的修订版。ISO/TS 17369:2005 主要涉及统计数据与元数据交换的六个主题: SDMX 标准的框架; SDMX 信息模型; SDMX-ML 模式与文档; SDMX-EDI 语法和文档; SDMX 实施指南; 网络服务使用 SDMX 指南。

(编译自: ISO 17369:2013 Statistical data and metadata exchange (SDMX).

ISO/TS 17369:2005 Statistical data and metadata exchange (SDMX).

[http://www.iso.org/iso/home/store/catalogue\\_ics/catalogue\\_detail\\_ics.htm?csnumber=52500](http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=52500). [2013-01-08]

[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=40555](http://www.iso.org/iso/catalogue_detail.htm?csnumber=40555). [2013-01-08])

(岳增慧编译, 吴贝贝校对)

#### 45、SKOS 简单知识组织系统

SKOS (简单知识组织系统) 是一个为诸如索引库、分类表、主题标题系统和分类等知识组织系统而建立的一般数据模型。通过使用 SKOS, 可将知识组织系统看作一种机器可读数据, 然后通过计算机应用进行数据交换, 并通过机器识别模式以网页形式显示出来。

SKOS 数据模型在本规范中被正式地定义为一个 OWL Full 本体 (OWL 语义), 表示为 RDF 三元组 (RDF 概念), 而且可以通过运用任何具体的 RDF 语句 (如 RDF/XML [RDF-XML] 或 Turtle [TURTLE]) 进行编码。

SKOS 数据模型将知识组织系统看作一个包含一系列概念的概念图示, 这些 SKOS 概念以及概念图示通过 URI 进行标识, 可以使任何人都能通过任何情境明确地引证, 并使之成为万维网的一部分。

SKOS 概念可用任意语言的词项 (UNICODE) 表示, 例如 “romantic love” 或 “れんあい” 等, 任何一种给定的自然语言, 可以是英语亦或是日语 (此处用的是日语的平假名)。

任一给定语言的其中一种表示可以成为最佳表示,其他表示则可以成为替代表示。另外,这些表示亦可能进行“隐藏”,这对于那些通过文本索引进行查询的知识组织系统而言是至关重要的。

SKOS 概念可以是一种或多种字符,或者说词汇编码,用这种词汇编码唯一标识某一给定概念图示范围内的概念。URL 是确定计算机系统内 SKOS 概念的首选方案,而字符为确定其他系统内已在应用的 SKOS 概念提供了桥梁作用,例如图书馆目录中应用的分类代码。

SKOS 概念可以用各种类型的笔记进行归档。SKOS 数据模型提供了一套基本的文档属性,支持范围注释、定义和编辑笔记等。这套属性并非面面俱到,而是提供一个框架,可以由第三方进行扩展,为更多特定类型的笔记提供支持。

SKOS 概念可通过语义关系属性关联到其他 SKOS 概念。SKOS 数据模型为 SKOS 概念间的层次连接和关联连接提供支持。同样,通过第三方扩展,这些概念可与 SKOS 数据模型的任何部分作用,为更多特定需求提供支持。

SKOS 概念可集结为集合,然后进行标记和/或命令。SKOS 数据模型这一特性的目的是为同义词典内的节点标签以及一个拥有一组有意义概念的的命令的情境提供支持,亦或是提供一些有用信息。

SKOS 概念可映射到不同概念图示的其他 SKOS 概念。SKOS 数据模型为四个基本的映射连接类型提供支持:层次映射连接、关联映射连接、封闭等效映射连接和完全对等映射连接。

此外,SKOS-XL 还为词汇实体的识别、描述以及连接提供了支持。

(编译自:SKOS Simple Knowledge Organization System.

<http://www.w3.org/TR/skos-reference/>. [2009-08-18])

(岳增慧编译,吴贝贝校对)

## 46、跨文化调查中最好实践的指导路线

比较调查设计和实施(CSDI)指导行动的目的是发展和促进国际认可的指导路线,这些路线强调跨文化和跨国度的比较调查研究实施的最好实践。目标用户为计划或从事跨文化或跨国度研究的研究者和调查从业者。CSDI 目前正发展成为跨文化调查指导路线(CCSG),CCSG 将覆盖调查周期的各个方面。CCSG 目前有 13 个章节,另外两个章节是关于调查质量和调查中的伦理问题的,两者与调查周期中的所有进程都相关。

(编译自:Guidelines for Best Practice in Cross-Cultural Surveys.

<http://ccsg.isr.umich.edu/pdf/00FullGuidelines3.pdf>. [2010-08])

(吴贝贝编译,王妍校对)

## 47、十原则——数字保存仓储的十项基本特征

2007年1月,在芝加哥研究图书馆中心召集了四个保存组织代表,寻求数字保存仓储的核心标准共识,指导审计和认证仓储的进一步国际努力。该组织是:

- 数字内容管理中心(英国)
- 欧洲数字保存
- 内斯特(德国)
- 研究图书馆中心(北美)

与会者确认了数字保存仓储的十项基本特征:

1. 致力于继续维护认证的社区/社区的数字对象的仓储。
2. 展示履行其承诺的组织适应性(包括财政、人事与进程)。
3. 获得和维护必要的合同与法律权利,并履行职责。
4. 拥有一个有效的和高效的政策框架。
5. 根据符合其承诺和能力的规定标准,获取和摄入数字对象。
6. 维护/确保数字对象的完整性、真实性和可用性。
7. 创建并维护必要的、在保存以及保存之前的相关生产、访问支持和使用流程环境下,对数字对象所采取的行动的元数据。
8. 履行必要的传播需求。
9. 拥有一个保存规划和行动的战略计划。
10. 拥有足以继续其数字对象维护与安全的技术基础设施。

核心需求的基本关键前提是所有类型与规模的保存活动的仓储必须与所定义的社区或社区组的需求成比例。

(编译自: Ten Principles -- Ten basic characteristics of digital preservation repositories.

<http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re.> [2007-01])

(岳增慧编译,吴贝贝校对)

## 48、可信赖的仓储审核和认证:标准和清单

TRAC由RLG和OCLC共同发布,进一步明确表达了可信、可靠、可持续的数字仓储的属性 and 责任的框架,这些数字仓储有能力处理由大的和小的文化遗产研究机构掌握的材料。这个框架可以通过为一个可信赖的仓储提供基础来适应于不同的环境、技术体系结构和机构职责。在应对文化遗产资源的长期保存中,这个文件被证明是有用的,并且已经与OAIS联合

发展成为一个数字保存计划工具。作为一个框架, 这个文件主要关注高度组织化和技术性的属性并为数字仓储的认证讨论潜在模型。

(编译自 :Trustworthy Repositories Audit & Certification: Criteria and Checklist. [http://www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf). [2007-02])

(吴贝贝编译, 王妍校对)

## 49、Triple S. 调查交换标准

Triple-s 标准定义了一种方式, 通过该方式调查数据和变量可以在运行于不同的软件和硬件平台上的不同的调查程序之间转化。这个标准的初始版本在 1994 年由 Keith Hughes, Stephen Jenkins and Geoff Wright 设计。Triple-S 的 XML 版本于 1999 年发布。

(编译自 :Triple S. Survey Interchange Standard. <http://www.triple-s.org/ssbackgrnd.htm>. [2013-01])

(吴贝贝编译, 王妍校对)

## 50、统一资源名称

URN 策略主要以希望分配和使用 URNs 的机构作为目标, 例如图书馆联盟、出版社、信息提供者和大学图书馆。德国国家图书馆在命名空间“nbn:de”中为所有 URNs 用户提供了基本的组织化和技术条件来分配、管理和解析 URNs。URNs 保证对象的永久获取。URNs 为德国国内的和国际的参考系统例如参考书目、目录和搜索引擎做参考, 并且可以通过书目交换格式进行传输。EPICUR 项目可以为管理和解析统一资源名称 (URNs) 创造协作的可利用的约定。

(编译自 :Uniform Resource Name (URN).

<http://www.persistent-identifier.de/?link=3352&lang=en>. [2008-02-01])

(吴贝贝编译, 王妍校对)

## 51、因特网工程任务组

IETF 是一个大型的开放的国际化团体, 汇聚了与因特网体系结构和因特网流畅运行发展相关的网络设计者、操作者、供应商和研究者。IETF 实际的技术工作由按主题划分的小组完成。IETF 的任务是通过形成高质量的有关技术性文件来使得互联网工作变得更好, 该

27

文件影响人们设计、使用和管理因特网的方式。IETF 标准进程的基本定义存在于 RFC 2026 (BCP 9)。知识产权规则分散在 RFC 5378 (BCP 78)和 RFC 3979 (BCP 79)中。

(编译自: Internet Engineering Task Force.

<http://www.ietf.org/> [2013-02-08])

(吴贝贝编译, 王妍校对)

## 52、世界数据中心: 气候

WDCC 的任务与 M&D 的基本任务相一致。被收集、存储和传播的气候研究数据为科学团体服务。WDCC 仅发展气候数据产品。WDCC 不会计划进行原始数据存储 (例如卫星或气候模型), 因为全球范围内的原始数据存储没有可以使用的设备。WDCC 正与在主题上相一致的数据中心进行紧密的合作, 例如地球观察、气象学、海洋学、地区气候和环境学, 目的是建立一个气候数据的完整网络。WDCC 包含在 CERA 数据库系统中, 提供的服务有: 用户服务、科学数据管理、项目支持。

(编译自: World Data Center for Climate.

<http://www.mad.zmaw.de/wdc-for-climate/>. [2008-02-19])

(吴贝贝编译, 王妍校对)

## 53、Z39.50: 信息检索协议

Z39.50 是严格基于 ISO 的 OSI (开放系统互联) 参考模型的应用层协议, 是一个美国国家标准。Z39.50 指的是国际标准 ISO23950: “信息检索 (Z39.50): 应用服务定义和协议规范”。Z39.50 的目的是为了信息系统的开放互联, 由于各信息系统分别采用各自的数据库软件, 数据的描述格式、访问方式等都各不相同, 必须为各自数据库系统建立一个抽象、通用的用户视图, 将各个系统的具体实现映射到抽象模型上, 才能使不同的系统在一个相互理解的、标准的通信平台上进行交互, 满足互操作的需要。允许计算机搜索远程系统上馆藏信息的一种协议, 产生可进一步处理的检索结果和检索信息, 主要用于检索书目信息。

(编译自: Z39.50. Information Retrieval Protocol.

<http://www.loc.gov/z3950/agency/>. [2012-09])

(吴贝贝编译, 王妍校对)



## 二、数据文档计划 (DDI) 技术规范第一部分：综述 版本 3.1

### 1.0 介绍

DDI 技术规范第一章：概述部分为数据文档计划 (DDI) 第 3 版的概念模型进行概述并作了技术描述。与之前版本不同，本版 DDI 标准还将包含两个部分：概念模型和 XML 模式及源自 XML 模式的文件类型定义 (DTDs)。这是 XML 词汇表标准化的一个常用方法，可为用户带来巨大便利：词汇表本身可变得更协调，更好理解；概念模型可为程序开发人员提供支持标准的有用信息，因为很多工具允许 XML 直接绑定由通用建模语言 (UML) 或其衍生品表达的模型。这一概念模型是在 DDI 技术规范中建立起来的。第三章：概念模型。

DDI 3 反映了修订后对 DDI 的有意覆盖及 XML 技术发展的整体展望。完成 DDI 覆盖、设计及结构意义转向的描述后，本文将详细阐述 DDI 3 中所用的结构及机械化。DDI 的技术规范，第二章：用户指南介绍 DDI 3 的各种不同应用。

### 1.1 数据生命周期的元数据

尽管最初 DDI 从码簿获得其模型，但当时很多人将这一概念延伸，使其涵盖面要比传统的硬拷贝码簿更广、更复杂。利用 DDI 3.1 版本，我们已经有能力记录社会科学数据整个生命周期的复杂性，参照图一：合并的生命周期模型。

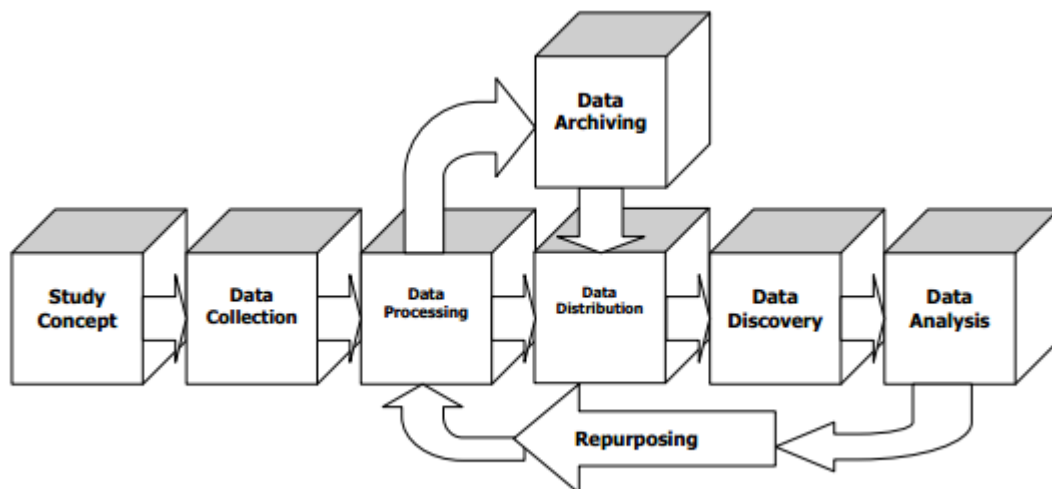


Figure 1: Combined Life Cycle Model

从历史进程来看，从未有过任何作为研究对象的 DDI 实例概念曾被设计、管理或存档。正如下图所示，完成数据的生命周期有诸多步骤，可通过 DDI 实例进行记录。可想而知，对于 DDI-3，一个研究的概念设计可以在 DDI 中得以标记。而且，因为研究贯穿整个生命周期，记录这一过程的 DDI 实例将会在接下来的版本序列中持续更新：通常，生命周期的每一阶段都会有一个。

合并的生命周期模型或者将直接传播整合给用户，或者将通过数据档案的传播整合给用户，知道数据可以在其生命周期后期进行重新加工，创建一个迭代过程。这意味着生命周期已不再是线性而是圆形过程。我们将数据用途的更改看作研究数据的次要利用。这不是从同

一个数据集,如机密数据文件、公用文件和综合数据文件等来创建多个产品。因为多个产品可以在初始概念化、收集和数据处理中进行规划,用途的更改可能是因为数据集的次要利用或者一项真实或虚拟的统一数据集,可以反映一项新的概念框架。这一观点的含义包括:定义概念过程中获得的数据产品间的关系,以及定义数据收集阶段的主要数据源和次要数据源的能力。

随着时间的发展,这种模型的模块化设计逐渐发展起来。这并不是一个突然的转向而是慢慢形成的一种共识。我们需要有处理专门数据文档和数据集的灵活性,还需要创造当前我们工作或开发中所需的各种技术环境。

## 1.2 范围变化

DDI 3 代表了不同于前版本的一个主要变化:范围大大扩大了。一直以来,DDI 都是专注于数据归档,如今这仍然是 DDI 3 的一个主要专注点,而且 DDI 3 支持数据生命周期的各个方面。因此,随着数据收集过程的进行,从概念设计到重新使用,描述这一活动进程的元数据的增长更新可以在 DDI 中得以收集和表达。

范围方面的这一转变在 DDI 的整体设计中具有多方面的影响。这意味着需要更大的实例来适应扩大的元数据集,同时也意味着描述单个数据文档的简单示例,不再广泛应用。与以往 DDI 版本相比,在当前 DDI 3 中,很容易在许多文件中找到用来“研究”的数据,形式往往也更加灵活。同时,这些文件也代表了广泛的需要进行描述的物理数据结构。

支持整个生命周期意味着一个研究与基于这一研究所作的系列研究的关系需要进行记录。因此,这一系列的研究也就需要进行描述,例如,一系列的纵向研究或比较研究等。这一变化的自然结果是提供了一种表达研究可比性的方法,这些研究往往设计的目的是进行比较研究或为了以后进行比较研究。

此外,档案要能记录更多与数据相关的自身活动方面的信息。而信息注意内部处理、收集管理以及组织结构都要求有更有力的支持。

如果不能比纸质收集仪器以更丰富的形式获取调查本身的信息,那么元数据对生命周期的描述是不完整的。如今很多系统都支持重新利用问题,因此仪器元数据是生命周期支撑的一个必要部分。

在 DDI 3 中还有其他一些变化,包括:支持 HTML 标记集合在一些含有较长的、人类可读文本的领域选择使用;可处理重新使用的大类,如笔记和引文已变得更加统一一致;增加了结构的一致性和参阅外部和内部材料的灵活性。为了校准或使用其他产品,其他元数据标准的重要性在这项设计中也得到了充分认识。

尽管 DDI 3 中在范围上体现了很大的变化,但一个重要的设计目标是尽量减少照搬 DDI 2。DDI 归档目的的这一简单应用并没有使版本产生根本性差异。而且 DDI 还会提供当前所有应用领域的映射,以及一些简单实用的免费工具来帮助用户。

## 1.3 技术更新

DDI 3 的许多变化是由 XML 技术的发展带来的。因为 W3C XML 模式定义 (XSD) 的应用已成为主流, DDI DTD 将不再是标准的权威表达。相反, 它将成为这一构架的一个姐妹产品, 尽管仍然描述 XML 实例, 但表达比一个 DTD 更多的验证参数。

XML 名称空间的应用是另一个典型的 XML 实践, DDI 3 将进行介绍。这可以使现已扩大的词汇表模块化, 使之长期可以更容易管理和维护。

应该说, DDI 3 意在提高其所包含的元数据支持计算机处理的程度, 也就是说, 它将超越“人类可读”, 争取实现“机器可读”这一目标。这是一个长期目标, 在早期的 DDI 三个版本中不会得到长足进展。但这在保持当前基于 XML 技术的整体应用, 如网络服务方面, 意义深远。

### 3.0 模式、方案和主要可重用类

DDI 3 包含 22 个 DDI 模式, 数套 DC 集和 XHTML 模式, 14 个方案和许多扩展的类, 所有的模式都是由 XSD 文件表示。DDI 模式共有四种类型:

- Packaging 模块: 是一种可维护的 DDI 模式, 可以建构元数据项而不包含独特的元数据项。
- Scheme-Based 模块: 是一种可维护的 DDI 模式, 其内容包含可维护的方案。
- Non-Scheme-Based 模块: 是可维护的 DDI 模式, 包含元数据项但不包含方案。
- 子模块: 只适用于抽象元数据类的替代品 (并不能独立维护)。
- 共享内容: 包含可供其他 DDI 模式应用的元数据, 不可维护。

方案是可维护的元数据元素列表, 组织可以分别发布的信息并由大量研究重新使用。它们是一些诸如问题银行, 概念银行及变量银行等资源的基础。方案的建构还考虑到其可为他人重用的潜力。很多被提出的资源集合已经得到 DDI 成员的注意, 包括提供标准编码项目的代码方案如 NAICS 代码等的代码方案, 包含 NHGIS 地理信息的地理结构方案, 以及主要档案所含的问题和变量银行。这种材料在统一结构格式的可用性可以支持重新使用和为比较进行的映射。

主要的可重用类是那些在模式中的 reusable.xsd 类, 以及广泛地应用于建构常见特性的类, 如识别、参考、引用、覆盖率、其他材料以及笔记等。因为所有的模式导入可重用, 这一模式中的元数据类在整个 DDI 实例中是可用的。

#### 3.1 XML 模式

模式是由类型列出的, 而且每个描述都包括模式.xsd 名称, 在模式中用作元素名称前缀的缩略形式, 以及模式的官方名称空间。描述之后是在根元素或元素中的元素列表。

##### 3.1.1 Packaging 模块

Schema Name: instance.xsd

[none]

Namespace: ddi:insance:3\_1

DDI 实例模块为包含所有类型的 DDI 实例提供了一个单一的根元素。这至关重要, 因为

处理程序可以处理很多 XML 类型, 而且这些程序需要一个已知起点来处理 DDI XML 实例。

应该指出的是, DDI 实例 (以及通用 DDI XML) 设计的初衷便是既可用作程序间传输的一个持久格式, 也可用作一个临时格式。因此, 也就不能假设一个已知的元数据集可以两次以同样的方式通过一个实例表达出来。在 DDI 3 中所编订、维护和引用的是元数据本身, 而不是表达元数据的 XML。这可能就像一个细微的差别, 而关于如何开发应用程序则有重大意义。

根元素中所含的元素 [最小……最大]:

```
r:Citation [0..1]
r:Coverage [0..1]
g:Group [0..n]
g:ResourcePackage [0..n]
g:LocalHoldingPackage [0..1]
s:StudyUnit [0..n]
r:OtherMaterial [0..n]
r:Note [0..n]
TranslationInformation [0..1]
```

**Schema Name:** group.xsd

**g**

**Namespace:** ddi:group:3\_1

这一模块为可提供其他模块生存所需的 XML 结构, 共有三个最高层元素。一个高层元素是群组, 描述构成群组一部分的子群和研究单位, 以及提供群组内继承和元数据共享信息的其他元素。基本的关系结构是由一组描述特定群组组织原则的属性提供的。

第二个高层元素是资源包, 用于描述可维护的模块或方案, 这些模块和方案可以被一个群组结构之外的多个研究单位使用。第三个高层元素是地方控股包, 可包括存储项 (研究单位或群组), 地方档案的添加以及无需复原存储项的其他信息。

根元素中所含的元素 [最小……最大]:

Group

```
r:Citation [0..1]
Abstract [0..n]
Purpose [1..n]
r:SeriesStatement [0..1]
r:FundingInformation [0..n]
r:Coverage [0..1]
r: UniverseReference [0..1]
r:OtherMaterial [0..n]
```

a:Archive [0..1]  
r:Note [0..n]  
Concepts [0..n]  
DataCollection [0..n]  
LogicalProduct [0..n]  
PhysicalDataProduct [0..n]  
StudyUnit [0..n]  
SubGroup [0..n]  
cm:Comparison [0..1]  
pr:DDIProfile [0..n]  
DDIProfileReference [0..n]

#### ResourcePackage

r:Citation [0..1]  
Abstract [0..n]  
Purpose [1..n]  
r:FundingInformation [0..n]  
r:Coverage [0..1]  
r:UniverseReference [0..1]  
r:OtherMaterial [0..n]  
a:Archive [0..1]  
r:Note [0..n]  
Concepts [0..n]  
DataCollection [0..n]  
LogicalProduct [0..n]  
PhysicalDataProduct [0..n]  
pi:PhysicalInstance [0..1]  
cm:Comparison [0..1]  
pr:DDIProfile [0..n]  
DDIProfileReference [0..n]  
a:OrganizationScheme [0..n]  
c:ConceptScheme [0..n]  
c:GeographicLocationScheme [0..n]  
c:GeographicStructureScheme [0..n]  
c:UniverseScheme [0..n]

d:ControlConstructScheme [0..n]  
 d:InterviewerInstructionScheme [0..n]  
 d:QuestionScheme [0..n]  
 l:CategoryScheme [0..n]  
 l:CodeScheme [0..n]  
 l:NCubeScheme [0..n]  
 l:VariableScheme [0..n]  
 p:PhysicalStructureScheme [0..n]  
 p:RecordLayoutScheme [0..n]

#### LocalHoldingPackage

g:DepositoryStudyUnitReference  
 g:DepositoryGroupReference  
 g:LocalAddedContent

**Schema Name:** studyunit.xsd

**s**

**Namespace:** ddi:studyunit:3\_1

这一模块包含适合单个研究单位的元数据,因此也就在很多方面对应一个 DDI 2.0 实例。应该指出的是, DDI 3 中, 研究单位总可以为其所属的群组或子群中所继承的元数据提供酬金, 也使表达所有将一个特定的研究单位看作一个单独、简单的 DDI 3 实例的元数据成为可能。

根元素中所含的元素[最小……最大]:

r:Citation [1..1]  
 Abstract [1..n]  
 r:UniverseReference [1..n]  
 r:SeriesStatement [0..1]  
 r:FundingInformation [0..n]  
 Purpose [1..n]  
 r:Coverage [0..1]  
 r:AnalysisUnit [0..n]  
 AnalysisUnitsCovered [0..n]  
 KindOfData [0..n]  
 r:OtherMaterial [0..n]  
 r>Note [0..n]  
 r:Embargo [0..n]  
 c:ConceptualComponent [0..n]

d:DataCollection [0..n]  
 l:BaseLogicalProduct [0..n]  
 p:PhysicalDataProduct [0..n]  
 pi:PhysicalInstance [0..n]  
 a:Archive [0..1]  
 pr:DDIProfile [0..n]  
 DDIProfileReference [0..n]

### 3.1.2 Scheme-Based Modules

**Schema Name:** archive.xsd

**a**

**Namespace:** ddi:archive:3\_1

这一模块提供档案特定信息的元数据,如电话号码和地方处理,数据或元数据的生命周期事件,以及与使用组织方案的实例内容相关的所有组织或个体方面的信息。要注意对 DDI 而言,一个“档案”是指任何作为 DDI 内容维护者的任何个体或组织。从这个意义上说,它可以描述最初的研究人员,数据生产机构,图书馆或档案,可以直接包含于任何 3.1.1 中所列出的包装模式中。

根元素中所含的元素[最小……最大]:

ArchiveModuleName [0..n]  
 r:Label [0..n]  
 r:Description [0..n]  
 ArchiveSpecific [1..1]  
 OrganizationScheme [1..1]  
 r:LifecycleInformation [0..1]  
 r:OtherMaterial [0..n]  
 r:Note [0..n]

**Schema Name:** conceptualcomponent.xsd

**c**

**Namespace:** ddi:conceptualcomponent:3\_1

这一模块允许元数据概念部分的归档,包括使用哪个概念以及如何定义、组群并组织成方案。另外,这一模块还含有一个全域方案来描述研究领域的覆盖面和结构,以及两个地理方案。地理结构方案用以获取研究中所涉及到的高层结构类型。地理定位方案为所描述的结构提供具体方位识别。它可依附于任何 DDI 实例的类型中(群组、研究单位、资源等)。

根元素中所含的元素[最小……最大]:

ConceptualComponentModuleName [0..n]  
 r:Label [0..n]  
 r:Description [0..n]

r:Coverage [0..1]  
 r:OtherMaterial [0..n]  
 r:Note [0..n]  
 ConceptScheme [0..n]  
 ConceptSchemeReference [0..n]  
 UniverseScheme [0..n]  
 UniverseSchemeReference [0..n]  
 GeographicStructureScheme [0..n]  
 GeographicStructureSchemeReference [0..n]  
 GeographicLocation Scheme [0..n]  
 GeographicLocationSchemeReference [0..n]

**Schema Name: datacollection.xsd**

**d**

**Namespace: ddi:datacollection:3\_1**

这一模块对数据收集过程进行描述, 可依附于 DDI 实例的任何一种类型, 包括方法论、收集事件、问题方案、组织问题的控制结构和特定顺序的文本、仪器、访谈者指令, 以及与数据收集相关的处理过程。

根元素中所含的元素[最小……最大]:

DataCollectionModuleName [0..n]  
 r:Label [0..n]  
 r:Description [0..n]  
 r:Coverage [0..1]  
 r:OtherMaterial [0..n]  
 r:Note [0..n]  
 Methodology [0..1]  
 CollectionEvent [0..n]  
 QuestionScheme [0..n]  
 ControlConstructScheme [0..n]  
 InterviewerInstructionScheme [0..n]  
 Instrument [0..n]  
 ProcessingEvent [0..n]

**Schema Name: logicalproduct.xsd**

**1**

**Namespace: ddi:logicalproduct :3\_1**

这一模块描述研究单位的逻辑产品, 或者一个群组或子群中的一个共享逻辑产品, 或者资源, 包括变量、范畴、范畴方案、代码方案、Ncubes、以及数据关系方面信息的描述。这



些数据关系信息有逻辑记录内容、独特的记录标识符、记录链接的复杂部分等。此模块通常由很多不同的 DDI 实例共享, 可存在于很多 DDI 实例类型中。

根元素中所含的元素[最小……最大]:

LogicalProductName [0..n]

r:Label [0..n]

r:Description [0..n]

r:Coverage [0..1]

DataRelationship [0..n]

r:OtherMaterial [0..n]

r:Note [0..n]

CategoryScheme [0..n]

CategorySchemeReference [0..n]

CodeScheme [0..n]

VariableScheme [0..n]

VariableSchemeReference [0..n]

NCubeScheme [0..n]

**Schema Name:** physicialdataproduct.xsd

p

**Namespace:** ddi:physicaldataproduct:3\_1

这一模块描述用于一个数据文件中的物理布局。注意, 在 DDI 3 中, 单个数据集可能分布于多个文件中。因为物理数据结构可能在一次研究的很多实例中重新利用, 甚至还会用于多个不同研究中, 这一模块可能出现在大多数 DDI 实例类型中(研究单位、群组或者资源包)。这就使数据文件及相关元数据的收集在管理方面更加灵活。物理结构方案包含了逻辑记录及其物理存储结构的基本物理特征方面的描述。

记录布局方案包含存储在一个特定结构中的记录布局的细节信息。很多记录布局的置换组都支持各种文件格式的描述。

根元素中所含的元素[最小……最大]:

PhysicalDataProductName [0..n]

r:Label [0..n]

r:Description [0..n]

r:OtherMaterial [0..n]

r:Note [0..n]

PhysicalStructureScheme [0..n]

RecordLayoutScheme [0..n]

### 3.1.3 Non-Scheme - Based 模块

**Schema Name:** comparative.xsd

cm

**Namespace:** ddi:comparative:3\_1

Comparative 提供了研究实体与一个群组或子群的比较、与外部标准的比较或一个资源包中两个或更多模式之间比较的元数据。它描述了这些研究实体之间在其领域、概念、问题、变量、类别和编码方案方面如何相互关联。

根元素中所含的元素 [最小……最大]:

ComparisonName [0..n]

r:Label [0..n]

r:Description [0..n]

ComparisonDescription [0..n]

ConceptMap [0..n]

VariableMap [0..n]

QuestionMap [0..n]

CategoryMap [0..n]

CodeMap [0..n]

UniverseMap [0..n]

r>Note [0..n]

**Schema Name:** ddiprofile.xsd

pr

**Namespace:** ddi:ddiprofile:3\_1

该模块允许 DDI 实例描述其所使用的 DDI 的元素和属性。声明哪些元素被使用或未被使用以及将可选元素改为所需的元素是可能的。用作该模块模型的 DDI 核心配置文件，不能在 DDI 2.0 XML 中进行表达。配置文件可以以 ResourcePackage 元素的方式进行描述、通过引用进行重用或被嵌入 Group 和 StudyUnit 模块中。

包含在根元素中的元素 [最小……最大]:

DDIProfileName [0..n]

r:Label [0..n]

r:Description [0..n]

XPathVersion [1..1]

DDINamespace [0..1]

XMLPreixMap [1..n]

Instructions [0..n]

Used [0..n]

NotUsed [0..n]

**Schema Name:** physicalinstance.xsd

pi

**Namespace: ddi:physicalinstance :3\_1**

此模块描述了与数据集的物理实例有关的位置和其他元数据。该模块依赖物理产品模块, 并且其总是针对于特定研究实体。它可直接或通过引用另一物理实例中的那些或通过物理实例所代表的数据文件来包含概要统计和类别统计。

包含在根元素中的元素[最小……最大]:

r:Citation [0..1]  
 Fingerprint [0..n]  
 r:Coverage [0..1]  
 r:OtherMaterial [0..n]  
 r:Note [0.. n]  
 RecordLayoutReference [1..n]  
 DataFileIdentification [1..n]  
 GrossFileStructure [0..1]  
 r:ProprietaryInfo [0..1]  
 Statistics [0..1]

**3.1.4 子模块**

**Schema Name: dataset.xsd** **ds**

**Namespace: ddi:dataset:3\_1**

该模块是一个 BaseRecordLayout 替代结构。它提供了一个将数据标记为物理数据产品的子模块的简单方法。它最适合于非 NCube 数据 (可在其他 DDI 模块中获得)。尽管标记名称不能反映表格式布局, 其含义是中立的, 数据可按行或按列分组。数据也可以随机的顺序与识别其变量名或用例标识的每一项进行输入。

包含在根元素中的元素[最小……最大]:

PhysicalStructureReference [1..1]  
 ArrayBase [0..1]  
 Name [0..n]  
 IdentifyingVariableReference [0..1]  
 DefaultVariableSchemeReference [0..1]  
 CHOICE: [1..1]  
     RecordSet  
     ItemSet  
     VariableSet

**Schema Name: physicaldataprodut\_ncube\_inline.xsd** **m3**

**Namespace: ddi:physicaldataprodut\_ncube\_inline:3\_1**

该模块是一个 BaseRecordLayout 替代结构。该模块允许逻辑产品中描述为 Ncubes 的多维数据的内联描述。

包含在根元素中的元素[最小……最大]:

PhysicalStructureReference [1..1]

ArrayBase [0 ..1]

NCubeInstance [1..n]

**Schema Name:** physicaldataprodut\_ncube\_normal.xsd m1

**Namespace:** ddi:physicaldataprodut\_ncube\_normal:3\_1

该模块是一个 BaseRecordLayout 替代结构。此模块包含了描述多维 Ncubes 的“普通”方法，重点在于将 Ncube 作为数据结构，而不是表象布局。

包含在根元素中的元素[最小……最大]:

PhysicalStructureReference [1..1]

CharacterSet [1..1]

ArrayBase [1..1]

NCubeInstance [1..n]

**Schema Name:** physicaldataprodut\_ncube\_tabular.xsd m2

**Namespace:** ddi:physicaldataprodut\_ncube\_tabular:3\_1

该模块是一个 BaseRecordLayout 替代结构。此模块按其呈现的形式描述多维数据——也就是说，根据一个特定表格式（二维）布局，当记录多维数据或存储于电子表格中数据的历史表格时特别有用。

包含在根元素中的元素[最小……最大]:

PhysicalStructureReference [1..1]

CharacterSet [1..1]

ArrayBase [1..1]

NCubeInstance [1..n]

TopLeftTableAnchor [1..1]

**Schema Name:** physicaldataprodut\_proprietary.xsd m4

**Namespace:** ddi:physicaldataprodut\_proprietary:3\_1

该模块是一个 BaseRecordLayout 替代结构。此模块描述了专有软件中的数据，如 SAS、SPSS 以及 Stata 等统计软件包。

包含在根元素中的元素[最小……最大]:

PhysicalStructureReference [1..1]

CharacterSet [0 ..1]

ArrayBase [0 ..1]

```

r:Software [1..1]
DataItemAddress [0..1]
DefaultNumericDataType [0..1]
DefaultTextDataType [0..1]
DefaultDateTimeDataType [0..1]
CHOICE: [0..1]
  CodedDataAsNumeric
  CodedDataAsText
DefaultVariableSchemeReference
r:ProprietaryInfo [0..1]
DataItem [0..n]

```

### 3.1.5 共享内容

**Schema Name:** reusable.xsd r

**Namespace:** ddi:reusable:3\_1

此模块描述了在整个 DDI 3 模式的不同模块中重用的 XML 类。它没有涉及可重用元数据（如在基于资源或群组的 DDI 实例中所提供的）。

**Schema Name:** dcelements.xsd dc

**Namespace:** ddi:dcelements:3\_1

此模块允许获取和表达本地都柏林核心元素，可作为参考文献或一个特定元数据集的描述来使用。在 DDI 中，都柏林核心并不用作主要引证机制——该模块支持理解柏林核心 XML 但是不理解 DDI 的应用。该模块在 DDI 3 中任何允许引用的地方都可使用。

**Schema Name:** ddi-xhtml11.xsd xhtml

ddi-xhtml11 -model -1.xsd

ddi-xhtml11 -modules-1.xsd

**Namespace:** http://www.w3.org/1999/xhtml

DDI 3 中的 XHTML 支持实例中文本描述的格式化。因为 XHTML 的独特性以及大多数开发环境下的相应支持，可以知道相比于 DDI 特定的格式化标记组，XHTML 为格式化提供了更好的方法。这一模块可用于 DDI 3 中任何有文本描述（可能要求格式化）的地方。只有那些指定元素才支持 XHTML 标记，而且这些元素通常旨在实现与机读相对的人类可读这一目标，且为了传达目标信息，其内容可能需要一定的结构。DDI 3 模式使用以下 XHTML 文件版本：

```

XHTML Modularization 1.1
W3C Working Draft 5 July 2006
http://www.w3.org/TR/xhtml-modularization/

```

**Schema Name:** xml.xsd XS

**Namespace:** <http://www.w3.org/XML/1998/namespace>

该模式用于 DDI 3 中, 允许使用通用 xml 类, 如语言格式的 `xs:lang`、内容域的 `xs:string` 等。

(编译自: Data Documentation Initiative (DDI) Technical Specification Part I: Overview Version 3.1.

<http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/>. [2009-10-18])

(岳增慧编译, 吴贝贝校对)