

融合开放视角下的文献编目新形态



四川大学图书馆 黄毕惠

E-mail : libdcatl@scu.edu.cn

厦门 2018年11月



编目变革缘起

技术环境扫描

编目标准发展

编目流程重组



前言

文献编目着眼于文献资源的形式特征及内容特征的描述与标引，其工作成果通过生成的书目数据得以呈现。

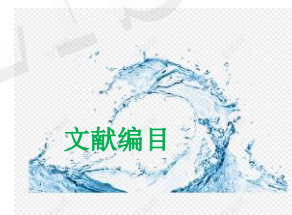
书目数据是揭示、检索、发现、利用文献资源的切入点，也是知识组织与知识利用的重要途径。随着信息技术及数据环境的发展变化，书目数据已经成为全社会大数据的重要组成部分。



前言

书目数据体量的指数级增长趋势和网络化的全文检索环境的形成，促使文献编目工作不断向纵深发展：

融合、开放的数字化网络环境正在形成文献编目的新形态





一、影响文献编目发展变化的因素

一、编目变革缘起

- 用户需求驱动

全域网络环境下，用户希望在万维网中以书目数据做为检索入口，查找、发现、识别、链接图书馆资源，并便捷获取广泛的数字化资源及深度的知识服务。用户的需求是图书馆文献编目工作的发展方向，以用户需求为导向的文献编目工作将不断深化与加强。

- 编目原则规定

IFLA制定的“国际编目原则声明”中，将编目规则的最高目标定义为“用户的便利性”。用户的便利是文献编目工作变革上升的源动力，IFLA编目目标的定义为文献编目工作的发展指明了方向。

一、编目变革缘起

- 社会化应用推动

网络环境下，书目数据的社会化应用呼声日益高涨。冲破当前“信息孤岛”的藩篱，将广泛分布的、自治的、异构的数据和各类书目检索系统进行整合，向用户提供便捷服务并实现用户的主动参与和互操作。书目数据社会化应用的趋势不断推动文献编目变革的步伐。

- 知识组织需求

用户查询书目信息的最终目的，是为了获取特定的、有效的知识信息。当前书目数据粗粒度、无语义、运行环境封闭等现状，与用户获取知识的需求相距甚远。利用现代信息技术和网络语义技术，对信息识别、分析、综合、重组，进而达成智能化的知识组织目标，完成书目组织向知识组织的升华，是对文献编目的新要求。

一、编目变革缘起

上述变革因素催生了文献编目新形态的出现，变革体现在如下方面：





二、信息技术环境扫描

二、新兴技术环境扫描

- SKOS

SKOS (Simple Knowledge Organization System, 简单知识组织系统), 是在语义网框架下, 用机器可理解的语言来表示知识组织系统的一个模型。它是W3C (万维网联盟)在2005年公布的知识组织系统概念框架表示的推荐标准, 主要用于表示各种较为简单的知识组织系统, 如叙词表、分类法、主题词表、术语表等其他类型的概念框架, 针对这些概念框架提供统一的表示标准。

SKOS建立了表示概念框架的知识组织模型, 具有简易、通用、易扩展等特点, 可以以机器可读方式构建知识组织系统, 提供基于RDF的知识组织系统之间的互操作, 最终形成语义环境下机器可以处理的知识组织系统; 并为语义Web环境下的数字信息资源整合提供了描述和转化机制。

二、新兴技术环境扫描

- ontology

在信息科学领域，本体(ontology)是指一种“形式化的，对于共享概念体系的明确而又详细的说明”。其实质就是对特定领域之中某种概念及其相互之间关系的形式化表达（**formal representation**）。本体用于描述由对象类型（概念或者类）、属性以及关系所构成的世界，它依据某种类别体系来表达实体、概念、事件及其属性和相互关系。

知识本体可以将异构性、分布性的数字资源进行有效地整合与集成，可将不同模式、不同语义描述的各类数字资源整合转换为基于本体的资源集成。其主要作用为：领域知识的形式化；提供资源库领域知识的规范描述；提供描述型元数据有关语义描述的知识地图；提供元数据映射方案；提供智能代理与信息环境之间基于语义的理解机制；成为跨平台、跨系统之间的通信中介；实现一定程度的领域知识的重用。

二、新兴技术环境扫描

• 语义技术

语义技术（Semantic Technology）是一系列关于信息资源处理与组织技术的总称，包括自然语言处理技术、信息抽取与挖掘分析技术、语义Web技术等。尤其是语义Web技术，已经在数字图书馆等诸多领域广泛应用。通过使用语义技术，对传统的数字图书文献资源进行语义标注、语义关联等富语义化操作，形成可操作性和互动性强的新型电子资源，具有精确查询、阅读便利等诸多优势。语义技术能有效整合万维网的海量异构数据，实现异构数据的互操作和多源数据的无缝链接，为大数据的有效分析提供一种技术途径。

语义技术标准主要包含：网络资源描述框架(Resource Description Framework, RDF)、网络本体语言(Web Ontology Language, OWL)、RDF查询语言SPARQL、规则交换格式(Rule Interchange Format, RIF)。语义技术标准的基本作用是对网络资源进行描述，用于提供语义唯一标识，同时让数据内容独立于表达形式；使得基于语义技术的应用系统可以非常方便地融合网络大量共享数据。

二、新兴技术环境扫描

- 关联数据技术

2006年，“互联网之父” Tim Berners-Lee提出了“关联数据”的概念。关联数据采用RDF模型，利用 URI命名和标识数据对象，通过 HTTP超文本传输协议揭示并获取数据，从而构建一个计算机能理解的具有结构化和富含语义的数据网络，并在此基础上实现更智能的应用。

以关联数据为基本数据模型的书目数据能够实现以细小粒度的数据为单元，每个数据单元既独立存在，又与其他数据单元建立可被机器理解的关联关系，数据相互关联且富含语义；同时数据单元可以方便地被链接到其他数据集中，与不同来源的数据混搭整合，构建新的服务模式。

书目数据、规范数据等相关数据均可发布为关联数据。关联数据作为实现Web上资源语义互联的关键技术，可以有效解决非结构化、半结构化的数字资源语义互联问题，实现跨领域的知识互联，从而更加有效地在全域网组织信息资源。

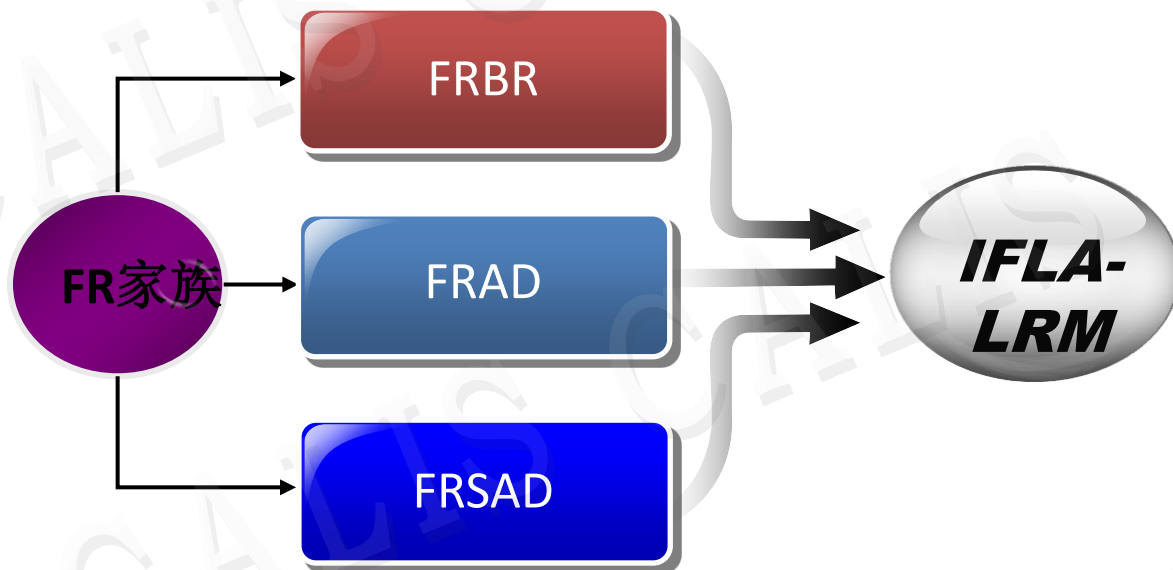


三、编目标准的发展变化

三、编目标准与规则的发展

1. IFLA-LRM

- IFLA-LRM（IFLA-Library Reference Model，国际图联图书馆参考模型），是在实体-关系模型框架下开发的高层概念模型，是对FR家族三个独立的模型整合。



IFLA-LRM对上述三个模型中的实体、属性及关系进行了整合，更强调统一性、通用性和开放性。其主旨是应用关联数据环境，支持并推动关联数据环境下的书目数据使用。

三、编目标准与规则的发展

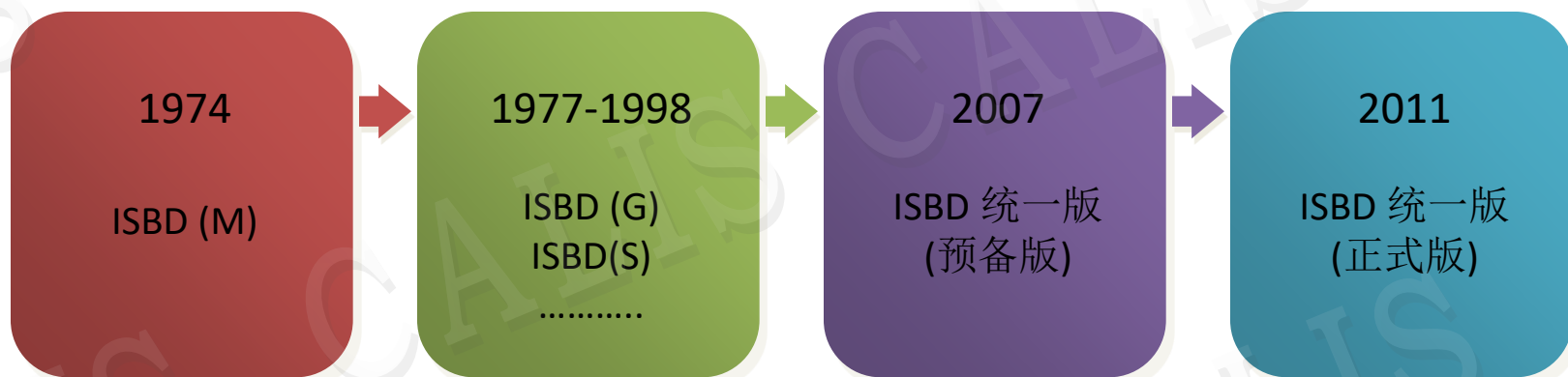
1. IFLA-LRM

- IFLA-LRM是帮助人们认识书目世界复杂事物及其关系的思维工具，为人们认识书目世界提供了一个新视角。
- IFLA-LRM以用户任务作为建模的出发点，其用户任务被定义为：查找、识别、选择、获取、探索。它强调终端用户需求，遵循了ICP的“用户便利性”的原则。
- IFLA-LRM作为书目信息的概念模型，对编目原则(ICP)、著录标准(ISBD)、编目规则(RDA)、书目格式(BIBFRAME)的修订及研究产生了深远的影响，成为构建书目世界国际性编目标准与规范的基础。

三、编目标准与规则的发展

2. ISBD

- ISBD (International Standard Bibliographic Description, 国际标准书目著录), 是现行各类信息资源描述著录的国际标准, 为世界范围内的文献编目标准化做出了巨大贡献。



ISBD发展历程

三、编目标准与规则的发展

2. ISBD

- 1998年FRBR的正式出版，极大地推动了ISBD的变革。为了确保ISBD规则内容与FRBR“基本国家书目记录”的要求相一致，ISBD做出了积极的改进，2004年《ISBD著录单元与FRBR实体属性和关系之间的映射》研究报告发布；2011年ISBD统一版增加第0项文件(内容形式和媒介类型项)，贯彻了FRBR的内容表达和载体表现的理念，为制定基于FRBR的编目条例奠定了基础。
- 2011年IFLA确定了RDF命名空间，并在其开放元数据注册网站对ISBD元数据进行了注册，使ISBD元素成为RDF词表的一部分；2016年，IFLA的《ISBD用作关联数据指南》发布，为如何在关联数据环境中使用ISBD命名空间提供了规则和依据。这些研究，为ISBD元数据的关联应用打下了基础。

三、编目标准与规则的发展

3. RDA

- RDA(Resource Description and Access, 资源描述与检索) 是数字环境下资源著录、描述与检索的新标准。RDA以FRBR为框架, 以“用户编目原则”为指导, 基于实体-关系模型, 服务于用户任务, 帮助用户查找、识别、选择和获取所需信息; 支持书目记录的聚类, 通过聚类展示作品和创作者之间的关系。
- RDA通过关系元素和关系说明语, 将责任者与资源的不同层次进行关联, 有利于查找、识别、显示不同层次的资源及相关的责任者, 满足用户的个性化需求。

三、编目标准与规则的发展

3. RDA

- RDA支持不同内容、不同载体资源的描述，适用于纸质资源和数字资源的著录，能够广泛适应数字化、网络化环境，具有各类型资源编目的普适性。
- RDA与现有的许多国际标准保持兼容，包括业内行业标准和与其他行业相关标准，如与档案馆、博物馆的标准兼容，能在多个行业内广泛应用。

三、编目标准与规则的发展

4. BIBFRAME

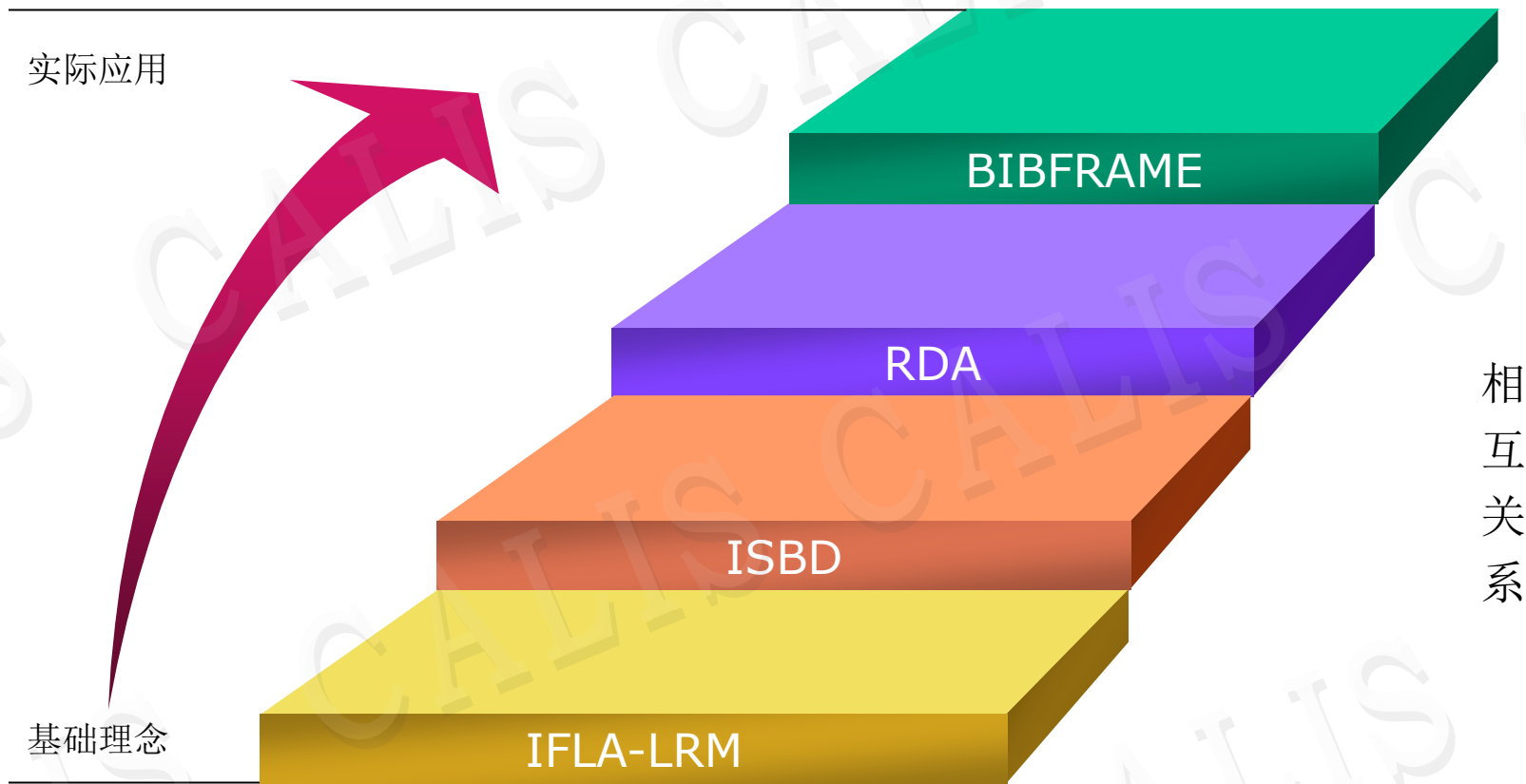
- 书目框架，是取代MARC的新一代书目格式，具有开放性、拓展性、兼容性、细粒度等特点，能揭示语义、强调关联关系。同时支持国际化和本地扩展，为封闭的书目数据集合向开放互联的语义网发布奠定了基础。
- BIBFRAME 采用实体—关系模型，对实体、属性及其关系等进行分析和标识。BIBFRAME 将描述对象简化为创造性作品（Creative Work）和实例（Instance）两个层次，并单独定义了规范实体和注释实体，以满足规范控制和扩展描述的需要。它利用关联数据技术中的URI明确标识不同的实体和属性，揭示资源（实体）之间的关系。

三、编目标准与规则的发展

4. BIBFRAME

- BIBFRAME的抽象模型和技术框架不仅仅提供一种格书目记录格式，还能容纳多种领域本体、元数据方案、数据序列化格式；不仅仅用于描述书目，还能描述博客文章、图像、影像、计算机程序、软件、数据集、实物等等，并为MARC、RDA、Schema.org、DCMI等资源描述模式提供映射方案。因此它不仅适用于图书馆，也适用于博物馆、档案馆等其它类似机构。

三、编目标准与规则的发展





四、编目工作流程重组

四、编目业务流程的重组

• 1. 文献编目现状



四、编目业务流程的重组

- 2. 文献编目业务流程重组

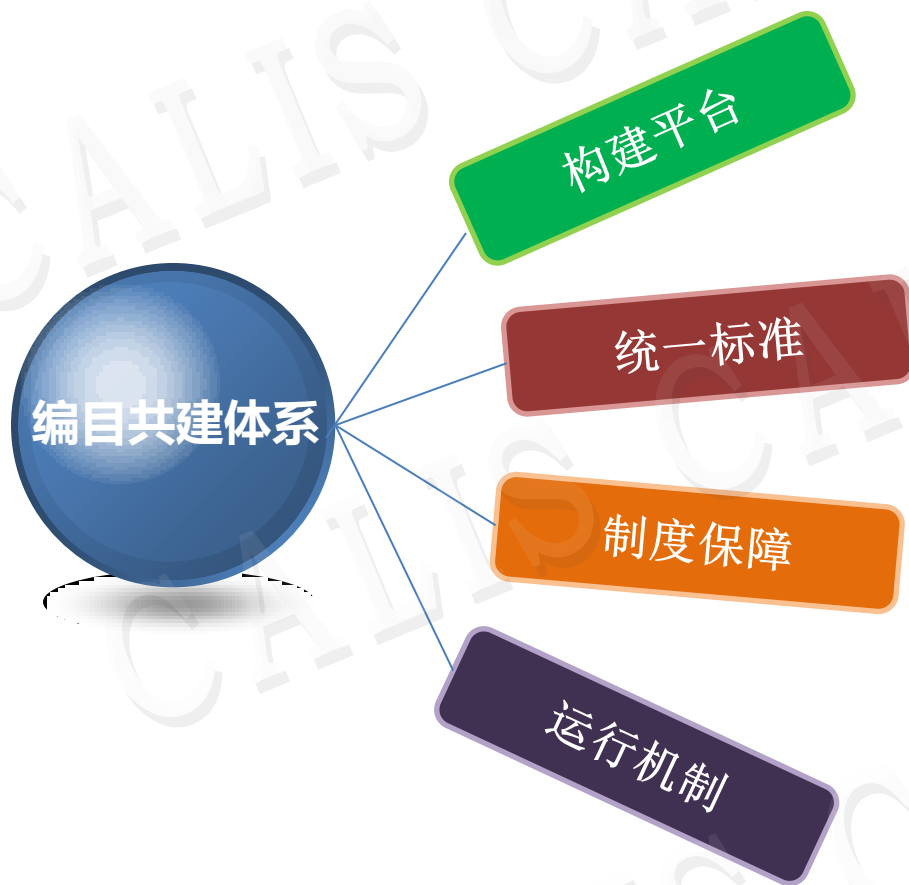


基于CNONIX标准，通过出版发行业、CIP中心、联合目录、图书馆之间的深度融合，实现编目前置，推动书目数据的跨平台应用。

四、编目业务流程的重组

- 3. 重组方式

基于书业生态链构建跨行业文献编目共建体系



结 语

文献编目新形态



以用户需求为导向，以数字资源为主体，积极利用信息技术发展的成果，适应语义网环境下知识组织、发现、关联、获取的需要，文献编目正日益呈现出标准化、一体化、关联化、智能化的发展趋势。

结 语



欢迎指教



谢谢大家