

基于 ISLI 标准的科技文献和科学数据的关联

朱江^{1,2}, 李欣怡^{1,2}, 姜恩波^{1,2}, 刘春江^{1,2}, 向彬¹

(1. 中国科学院成都文献情报中心; 2. 中国科学院大学经济与管理学院)

摘要: 国际标准关联标识符 (ISLI) 国际标准是我国新闻出版界主导制定的第一部国际标准, 用于标识两个实体之间的关联关系, 为科技文献和科学数据的关联提供了新的解决方案。文章在分析科技文献和科学数据关联形式的基础上, 论述了科技文献和科学数据的四种关联类型、ISLI 标准元数据集扩展、关联对象的扩展和规范、ISLI 编码体系设计, 并提供了基于 ISLI 标准的科技文献和科学数据的具体关联实例, 为制定和完善相关标准提供参考。

关键词: 国际标准关联标识符; 科技文献; 科学数据; 关联

中图分类号: G250.7 文献标志码: A 文章编号: 1005-8214(2020)05-0080-04

DOI:10.14064/j.cnki.issn1005-8214.2020.05.015

Linkage between Scientific Literature and Scientific Data Based on ISLI

Zhu Jiang, Li Xin-yi, Jiang En-bo, Liu Chun-jiang, Xiang Bin

Abstract: The International Standard Link Identifier (ISLI), which is used to identify the relationship between two entities, is the first international standard led by the field of press and publication in China. It also provides a new solution for the linkage of scientific literature and scientific data. Based on the analysis of association forms between scientific literature and scientific data, the paper firstly discusses four association types of scientific literature and scientific data, the extension of ISLI standard metadata set, the extension and specification of association objects, and the design of ISLI coding system. Then, the paper provides a specific example of linkage between scientific literature and scientific data based on ISLI, so as to provide a reference for the development and improvement of the relevant standards.

Keywords: International Standard Link Identifier (ISLI); Scientific Literature; Scientific Data; Linkage

1 ISLI 简介与应用现状

ISLI 即国际标准关联标识符 (International Standard Link Identifier), ISLI 国际标准《ISO 17316: 2015 信息与文献——国际标准关联标识符 (ISLI)》^[1] 于 2015 年 5 月 15 日正式发布, 这是我国新闻出版界主导制定的第一部国际标准。与单一对象标识符不同, ISLI 不以标识和识别单一对象为目的, 而是主要用于标识两个实体之间的关联关系, 并不改变实体的各种属性和标识,^[2] 其表现形式是一个带有关联信息 (元数据) 的标志码, 以创建实体之间的关联, 实现多个实体的关联、共同呈现等目的。这些实体可以是图文音像、数据、知识元等具有不同细粒度和表现形式的信息与文献领域内的所有具体实物, 也可以是地理位置

坐标、时间点等更为抽象的事物。^[3] ISLI 关联模型包括源、目标和两者之间的关联三个要素。^[4]

关联关系是文献与信息领域不同类别、属性、细粒度实体之间尚待进一步挖掘与实现的宝贵资源, 有着重要的现实意义。ISLI 的提出、完备与推广对于定义相关实体间的关联关系有重要作用。① 稳定可靠。在 ISLI 标准框架下, 各实体就如同一个个“零部件”, 通过 ISLI 编码实现彼此间共性主题的关联, 同时各实体仍可保持其个体独立性, 因为 ISLI 标准并不改变实体的独立形态及其原本功能。也就是说, 依托 ISLI 标准构建的关联关系不会因为实体所处的仓储系统或其他客观环境的变化而失效。② 原理简洁。不同于以识别和标识单一对象为目的的标识符, ISLI 的关注点

[基金项目] 本文系 2017 年科技部国家重点研发计划重点专项“专业内容知识服务众智平台与应用示范”(项目编号: 2017YFB1402400) 的研究成果之一。

在于使用 ISLI 编码体系创建两个实体之间的关联关系，一旦编码标准的细节制定成熟完备，便可快速、低成本地盘活现存资源，以关联推进新的资源价值产生。^③ 拓展灵活。在 ISLI 标准的编码方案中，关联字段所包含的数字编码的长度和结构均可由注册机构根据应用对象的分类和需要进行自定义，^[4] ISLI 标识符的编码结构采用了定长但不限长的十进制数字编码及相对自由的编码要素语法规则，因此，ISLI 编码体系中的关联字段具有良好的可扩展性。

ISLI 标准源自我国的多媒体印刷读物 (Multimedia Print Reader, MPR) 标准，^[5] ISLI/MPR 被视为 ISLI 标准最初的典型应用，并成功在国内外出版领域得到认可与推广。^[6] 随后，ISLI 标准积极探索与更多领域的融合模式，并在知识关联服务 (Knowledge Linking Service, KLS)、增强现实 (Augmented Reality, AR) 出版、插画等领域逐步形成 ISLI/KLS^[7]、ISLI/AR^[8]、ISLI/WIS^[9] 等应用。国家新闻出版广电总局数字出版司于《新闻出版业“十三五”科技发展规划总体思路》规划中明确表示“要大力推进 ISLI 标准在国内外的产业应用”。^[10] 2018 年 5 月，中国 ISLI 注册中心与中国科学院计算机网络信息中心就新闻出版与物联网标识的元数据共享开展合作，^[11] 这意味着 ISLI 的推广落地将在非新闻出版领域有所突破。

2 科技文献和科学数据的关联

科技文献是学者向社会展现其科研成果的重要途径之一，而科研过程中产生的科学数据实则与科研结果保持着同等重要的位置，其如同“证据”一般详实地记录着科研结果的推导过程，一些科学数据也已成为科研成果的重要表现形式之一。科技文献与科研数据间关系可谓紧密。随着认识的提高和技术的突破，人类进入了大数据时代，科学研究不可避免地也有了新的延伸方向：数据在科研中的地位逐渐重要，数据密集型科学作为科学探索的新的第四范式随之产生。^[12] 加之开放获取浪潮的不断推动，人们对于科技文献中相关科学数据的可获取性呼声日益高涨，诸如 GBIF、DataStaR、OTA 之类的数据仓储平台应运而生。目前，有学者基于数字对象唯一标识符 (DOI)、元数据等角度对科学数据和科技文献的关联展开了研究。^[13,14] ISLI 标准的出现为实现科技文献和科学数据的标准化关联提供了新的解决方案，同时也有利于提升两者间关联关系的稳定性和复用性。

以科学数据原创性和独立性为标准，可将科学数据分为科技文献内科学数据和科技文献外科学数

据。^[15] 前者指在某项科研过程中具体产生的原始科研数据，与科技文献内容完全整合在一起，通常以表格、数字、图像等格式呈现，此类科研数据并未被单独存储，而是依附于科技文献，文献本身即为此类科研数据的天然载体，对这一类科研数据通常以文献引用代替数据引用；后者是指与科技文献分离、独立存在的科学数据，通常来自于特定数据仓储平台的科学数据集和数据记录。目前，还出现了以刊载和发行规范化科学数据为主要目标的数据出版物，如数据期刊和数据论文，这种类型的科学数据可作为直接引用源。

科技文献、科学数据的关联形式多样、类型复杂，主要的关联形式有以下三种。^[16] ① 硬关联。一篇科技文献与它在特定数据仓储平台提交的科学数据集之间的关联，呈现一对一或一对多的关系。这种关联是有意识的、人为的关联，主要实现科学数据与其来源科技文献形式上的关联。② 软关联。一篇科技文献与它引用的科学数据集或主题、内容相关的科学数据集之间的关联，大多数情况下是一对多的关系。这种类型的关联是从某一特定内容出发，将与该内容相关的多方科技文献和科学数据进行集成，可为用户参考提供便利。但这种关联尚不成熟，准确性有待考证。③ 其他关联。一篇科技文献与在出版、传播过程中由编辑或同行科技工作者添加的其他科学数据集或解析工具之间的关联等。上述关联均可利用 ISLI 模型实现。

科技文献和科学数据并非最小的关联单位，可将其解构成细粒度更小的知识单元。如果把独立的单篇科技文献和独立的单个科学数据集看作“资源”，把科技文献和科学数据集中更小的知识单元看作“知识”，则可将科技文献、科学数据集之间的关联划分为四种类型 (见表 1)。根据上述关联类型，可在 ISLI 标准元数据集的基础上对关联类型元数据项、关联对象及取值范围进行扩展，同时对 ISLI 关联编码体系进行定义。

3 科技文献和科学数据关联的 ISLI 元数据集扩展

3.1 ISLI 标准元数据集的扩展

(1) 关联类型的扩展。在 ISLI 标准元数据集的基础上，对科技文献和科学数据集的关联类型元数据项进行扩展，划分为上述的四种类型，并预留进一步扩展的空间。

(2) 关联对象的扩展和规范。科技文献具有诸如论文、图书、研究报告等多样的表现形式，而科学数

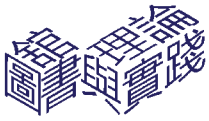


表 1 科技文献与科学数据集的关联类型

资源—资源	资源—知识
科技文献-科技文献	科技文献-科技文献中的知识单元
科技文献-科学数据集	科技文献-科学数据集中的知识单元
科学数据集-科技文献	科学数据集-科技文献中的知识单元
科学数据集-科学数据集	科学数据集-科学数据集中的知识单元
知识—资源	知识—知识
科技文献中的知识单元-科技文献	科技文献中的知识单元-科技文献中的知识单元
科技文献中的知识单元-科学数据集	科技文献中的知识单元-科学数据集中的知识单元
科学数据集中的知识单元-科技文献	科学数据集中的知识单元-科技文献中的知识单元
科学数据集中的知识单元-科学数据集	科学数据集中的知识单元-科学数据集中的知识单元

据分为原始性基础数据与按照不同需求加工后的数据集和相关信息。不同的科技文献和科学数据之间的关联位置、层级、程度不可能千篇一律。因而对于关联对象而言，其取值范围、取值标准和校验方法亟需得以扩展与规范，以便区分与囊括不同类型和细粒度的关联对象，打破载体和介质的屏障，直接进入内容资源层建立一套具有统一规则的、完整的关联关系。

(3) 关联行为元数据项的扩展和作用。实体间存在目的不同的关联行为，可参照 Subject-Action-Object (SAO) 的语义表达结构理念，在 ISLI 关联的“源”和“目标”中间增加一个关联行为元数据项，其作用相当于 SAO 中的“A”，以便对庞大的关联编码进行聚类。关联行为元数据项的取值可根据实际应用不断扩展，以适应不断创新、变化的关联行为。

3.2 命名实体的抽取及 KOI 标识的建立

出于对海量实体及实体种类“身份”管理的需求，需要对“有意义”的实体进行唯一性标识，并通过识别、抽取命名实体和创建实体唯一标识符实现对实体的准确性、系统性管理。

- ① 明确命名实体的命名规则及组织、保存和更新规则，以便对不同类别的实体进行有效归类、检索，明确实体边界，便于机器学习，实现对命名实体的自动抽取、命名、保存和更新。
- ② 建立 KOI (知识对象标识符) 标识。ISLI 关联的是“源”和“目标”两个实体，虽然 ISLI 标准规定“源”实体一般是确定的，“目标”实体可以是确定的也可以事后确定或建立，但“源”和“目标”作为实体，除了实体名称外，一般还应建立一个唯一标识符。独立的单篇科技文献和独立的单个科学数据集可被视为粒度较粗的“资源”层面，该层面现已存在数字对象标识符 (DOI) 作为唯一标识符。但对于科技文献和科学数据集中的知识实体 (或知识单元) 还尚未建立统一且被广泛使用的唯一标识符体系，在此可

借助 KOI [7] 对系统内抽取出来的知识实体 (或知识单元) 进行标识和存储。由于 KOI 尚无公认的标准，因而只能在一定范围内解析使用。

4 科技文献和科学数据关联的 ISLI 编码体系设计

ISLI 编码由十进制数字构成，分为服务字段、关联字段和校验字段。参照相关文件，[4] 一个 ISLI 编码的形式如图 1—图 3 所示 (连字符“-”及“ISLI”并不构成标识符的组成部分，只为便于阅读)。服务字段的编码一般为 6 位，由 ISLI 注册中心 (ISLI RA) 分配；关联字段的编码长度可变，具体长度、结构由 ISLI RA 根据应用需求进行定义，如果分段，一般可细分为前置编码和后置编码两部分；校验字段的数值 (校验码) 由 ISLI 系统依据规则自动计算。[4]

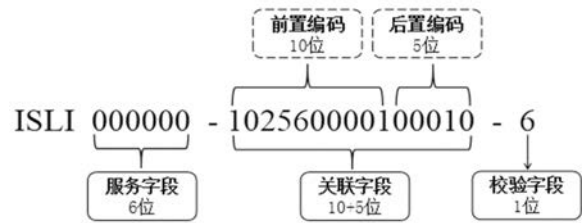


图 1 图书、音像制品 ISLI 编码字段结构

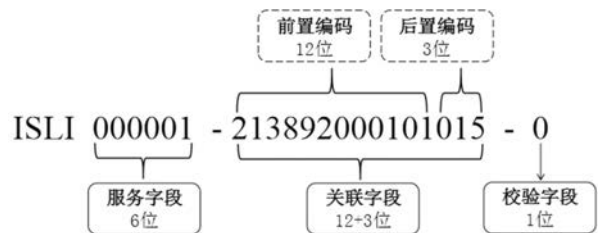


图 2 连续出版物 ISLI 编码字段结构



图 3 以互联网传播的音频内容产品 ISLI 编码字段结构

在科技文献和科学数据关联服务中，假设服务字段为“200000”，关联字段的前置编码为 10 位数字，容有 100 亿个码段，可在不扩容情况下供 100 亿个独立的科技文献或科学数据集使用；后置编码为 5 位数字，容有 10 万个无重复的编码，除去用于标识 1 个科学文献或科学数据集和它的元数据之间关联的“00000”这个特定的 ISLI 标准标志码外，还有

99,999 个 ISLI 标准编码可用于标记该实体与该实体内部的知识单元、外部其他实体（包括外部其他科技文献、科学数据集及其包含的知识单元）的各种关联，并使用经过扩展的 ISLI 标准元数据集来表示各种复杂的关联。

如“10+5”的关联编码字段容量不足，还有根据实际需要增加关联编码字段的长度，变成“10+6”或“12+6”等形式，以容纳更多的关联编码。

为了更加具体地说明科技文献和科学数据的关联是如何基于 ISLI 标准来实现的，特举例如下。论文 A 根据“2018 年成都市日降水量观测记录表”汇总了一张“2018 年成都市月降水量统计表” KOI(a)，并与已获得 DOI(B) 的论文 B 中的“2018 年武汉市月降水量统计表” KOI(b) 进行了比较。论文 A 发表后获得 DOI(A)， “2018 年成都市日降水量观测记录表” 按要求提交到数据仓储平台，获得 DOI(a)；将 DOI(A) 的 ISLI 关联编码的前置编码定为“1000000001”，DOI(a) 的 ISLI 关联编码的前置编码定为“1000000002”，而后置编码按一定的编码段或流水号顺序分配的话，则可形成最基本的关联和 ISLI 关联编码（见表 2）。相应地，对于 DOI(B)、KOI(b) 来说，也可以采取类似的方法，将它与科技文献及其他科学数据集的关联标识出来。

表 2 基于 ISLI 标准的科技文献与科学数据集的关联

关联编码		关联类型	源	目标	关联说明
前置编码	后置编码				
1000000001	10001	资源 - 知识	DOI(A)	KOI(a)	论文 A 与论文 A 中“2018 年成都市月降水量统计表” KOI(a) 的关联
1000000001	30001	资源 - 资源	DOI(A)	DOI(a)	论文 A 与论文 A 提交的科学数据集“2018 年成都市日降水量观测记录表” DOI(a) 的关联
1000000001	30002	资源 - 资源	DOI(A)	DOI(B)	论文 A 与参考文献论文 B 的关联
1000000001	50001	知识 - 知识	KOI(a)	KOI(b)	论文 A 中“2018 年成都市月降水量统计表” KOI(a) 与参考文献论文 B 中“2018 年武汉市月降水量统计表” KOI(b) 的关联
1000000001	70001	知识 - 资源	KOI(a)	DOI(a)	论文 A 中“2018 年成都市月降水量统计表” KOI(a) 与论文 A 提交的科学数据集“2018 年成都市日降水量观测记录表” DOI(a) 的关联
1000000002	10001	资源 - 知识	DOI(a)	KOI(a)	论文 A 提交的科学数据集“2018 年成都市日降水量观测记录表” DOI(a) 与论文 A 中“2018 年成都市月降水量统计表” KOI(a) 的关联
1000000002	30001	资源 - 资源	DOI(a)	DOI(A)	论文 A 提交的科学数据集“2018 年成都市日降水量观测记录表” DOI(a) 与论文 A 的关联

将上述 ISLI 编码及其元数据存储到科技文献和科学数据关联服务系统中，并提交给 ISLI 注册服务中

心，即可在 ISLI 服务系统的支持下实现 ISLI 编码的解析和服务。

5 结论

对于科技文献和科学数据关联中存在的“源”和“目标”细粒度不同、关联类型多样等问题，ISLI 标准以其稳定可靠、原理简洁、可拓展性强等较为独特的优点提供了一种角度新颖的解决途径。在科技文献和科学数据关联的具体实践中，以标识符定义被关联对象之间的关联关系，有利于快速发现并精准定位到所需的高品质内容，同时也使得科技文献和科学数据之间的引用与考证变得更加容易。应全面总结科技文献和科学数据关联的类型，并借鉴 ISLI 标准的其他典型应用案例，完善基于 ISLI 标准的科技文献和科学数据关联标准，以充分展现科技文献和科学数据的关联。

[参考文献]

[1] ISO 17316 : 2015 information and documentation — International standard link identifier (ISLI). [EB/OL]. [2019—08—24]. <https://www.iso.org/standard/59560.html>.

[2] 蔡逊. 关联——ISLI 国际标准的理念与价值 [J]. 出版发行研究, 2015 (7) : 39—41.

[3] ISLI 的适用对象是什么? [EB/OL]. [2019—08—27]. <http://www.cisli.org/isli/resources/dist/#/content/2>.

[4] ISLI 应用指引 (信息内容产业—简体中文). [EB/OL]. [2019—09—04]. https://www.isli-international.org/download/Guide_for_using_ISLI.pdf.

[5] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 27937.1—2011 MPR 出版物第 1 部分: MPR 码编码规则 [S]. 北京: 中国标准出版社, 2011.

[6] 左美丽. ISLI 在出版领域的应用综述 [J]. 出版参考, 2016 (8) : 9—11.

[7] 赵海涛, 等. ISLI 在知识服务中的应用研究 [J]. 出版参考, 2017 (4) : 14—17.

[8] 蔡逊. 国际标准关联标识符 (ISLI) 与新闻出版产业创新 [J]. 出版参考, 2016 (8) : 5—8.

[9] 高珊. 助画方略全力打造 ISLI+ 全球插画资源聚合平台 [J]. 出版参考, 2016 (8) : 14—15.

[10] 新闻出版业“十三五”科技发展规划总体思路 [EB/OL]. [2019—09—02]. <http://data.chinaxwbc.com/epaper2016/epa->

(下转第 91 页)

提供窗口自然也不例外。县级公共图书馆在资源紧张的情况下可以采取如下策略提高读者的满意度和吸引力。① 跟踪社会热点事件。及时在微信公众平台上发布群众关注度较高的消息,尤其是文化类、民生类、教育类等,这样的内容能迅速抓住读者的注意力,提高读者的阅读体验。② 适当转发热门文章。搜集网络爆款文章,筛选符合社会主流价值观和图书馆业务特点的文章加以转发,并注明转载来源,同时,要控制好原创文章与转载文章的比例。③ 适当开展奖励性活动。如,赠书、转发有奖、评论有奖、投票等,在有正向反馈的情况下,读者对文章的阅读兴趣和认可度会大大增加,也会乐于同朋友进行分享,能够极大地提升图书馆微信公众平台的知晓度。

[参考文献]

- [1] 闻波,等. 国内省级公共图书馆微信公众号推送文章研究:基于 Tableau 与 WCI 的分析 [J]. 图书馆杂志, 2018, 37 (4): 58-65.
- [2] 张晨. 全民阅读环境下公共图书馆微信推送服务探析——以天津图书馆为例 [J]. 图书馆工作与研究, 2018 (S1): 158-161.
- [3] 杨思洛, 李小娟. 湖南省地市级公共图书馆微信阅读推广活动调查分析 [J]. 公共图书馆, 2019 (2): 76-82.
- [4] 赵谓炯,等. 少儿图书馆微信公众平台应用研究 [J]. 图书馆工作与研究, 2017 (8): 119-123.
- [5] 倪萍. 少儿图书馆微信阅读推广研究 [J]. 数字图书馆论坛, 2017 (6): 58-64.
- [6] 欧阳红红. 福建省区县级图书馆微信公众平台利用情况调查与分析 [J]. 图书馆学研究, 2017 (14): 17-23.
- [7] 王莉红. 县级图书馆微信公共平台服务发展现状及优化策略 [J]. 神州, 2018 (1): 216-217.
- [8] 秦美丽. 县级图书馆微信公众平台的问题及优化 [J]. 文化产业, 2018 (5): 59-60.
- [9] 毛小敏. 县级公共图书馆微信应用的思考——以平阳县图书馆为例 [J]. 内蒙古科技与经济, 2016 (9): 57-58.
- [10] 中国图书馆学会, 国家图书馆. 中国图书馆年鉴 2017 [J]. 北京: 国家图书馆出版社, 2017: 436.
- [作者简介] 刘阳 (1985—), 男, 硕士, 山东省图书馆馆员, 研究方向: 信息检索, 网络信息服务; 雷辉 (1973—), 女, 山东省图书馆副研究馆员, 研究方向: 公共图书馆新媒体服务。
- [收稿日期] 2019-11-28 [责任编辑] 李梦竹
-
- (上接第 83 页) per/d6193/d7b/201601/64564.html.
- [11] 打通数据壁垒 实现跨界融合: 计算机网络信息中心与中国 ISLI 注册中心签订战略合作协议 [EB/OL]. [2019-09-02]. http://www.cnica.cn/xwdt/ttxw/201805/t20180524_5016767.html.
- [12] Tony Hey, et al. The Fourth Paradigm: Data-Intensive Science Scientific Discovery [M]. Washington: Microsoft Research, 2009: 1334-1337.
- [13] 郭学武. 基于引文的科学数据与科技文献关联研究 [J]. 情报科学, 2014, 32 (4): 59-62, 125.
- [14] 黄筱瑾. 基于内容特征的科学数据与科技文献关联研究 [J]. 现代情报, 2018, 38 (1): 56-59.
- [15] 孙文佳, 常娥. 科学数据与科技文献关联分析 [J]. 图书馆理论与实践, 2017 (3): 49-53.
- [16] 姜恩波, 裴玉香. 科学文献与科学数据的融合方法与实例研究 [J]. 知识管理论坛, 2019 (2): 69-79.
- [作者简介] 朱江 (1968—), 男, 中国科学院成都文献情报中心研究馆员, 中国科学院大学经济与管理学院硕士生导师, 研究方向: 数字资源建设与组织; 李欣怡 (1997—), 女, 中国科学院大学经济与管理学院 2018 级硕士研究生在读, 研究方向: 数字资源建设与组织; 姜恩波 (1972—), 男, 中国科学院成都文献情报中心研究馆员, 中国科学院大学经济与管理学院硕士生导师, 研究方向: 知识资源集成系统研究与设计; 刘春江 (1984—), 男, 中国科学院成都文献情报中心馆员, 中国科学院大学经济与管理学院 2017 级博士研究生在读, 研究方向: 知识资源集成系统研究与开发; 向彬 (1980—), 男, 中国科学院成都文献情报中心助理馆员, 研究方向: 知识资源集成系统开发。
- [收稿日期] 2019-11-26 [责任编辑] 阎秋娟