

第三代图书馆服务平台的元数据管理

——以 FOLIO 的 Codex 方案为例*

许磊 夏翠娟

摘要 在数据化的网络时代,用户需求与图书馆服务形式向深度专业的知识化和智慧化发展。面对这种新常态,第三代图书馆服务平台的元数据管理不仅需要解决图书馆集成管理系统甚至是“下一代”图书馆服务平台受困于 MARC 的现状,满足资源的统一管理 with 业务整合,更需要考虑统一的数据建模与编码方式,实现跨类型的数据流通和知识的组织与融合。这就为元数据管理提出了新的需求,包括:完整的生命周期管理,资源类型兼容,标准规范兼容,互联网环境下的书目与规范控制,知识组织与融合。这种运用平台化思维、改变图书馆运行生态的可能性可以在 FOLIO 项目中看到希望。FOLIO 作为以微服务架构设计的图书馆服务平台,在设计之初就定义了一个核心的微服务域——Codex。一方面,Codex 作为数据注册中心和数据链接中介位于 FOLIO 架构的顶层,通过 Codex 消除不同域中数据的编码、格式和存储位置的差异,以减少域之间的交互与耦合。另一方面,Codex 以 BIBFRAME 2 模型为基础设计了“作品—实例—单件/馆藏”的抽象数据模型,并参考 DC 元素设计了一套最小核心元数据元素集。Codex 记录不描述资源的详细信息,只起到揭示与定位的作用。专有的业务信息或更详细的记录信息都可以通过 Codex 链接到相应的记录层获取。Codex 作为核心的元数据管理域,将被设计成为 FOLIO 平台的资源链接中心、规范数据中心、跨域的全媒体资源管理入口,并通过模型中的抽象实体揭示资源间的关系。虽然现阶段仅实现了不存储自身数据的 Codex Search,对关联数据的支持也不够完善,但无论是其抽象的数据模型,还是最小化的元数据方案,以及“统一记录—业务记录—正式记录”的分层描述原则,Codex 都有足够的灵活性支持面向实体对象的描述与组织。图 6。表 6。参考文献 28。

关键词 图书馆服务平台 元数据管理 FOLIO Codex

分类号 G250.7

Metadata Management of the Third Generation Library Service Platform: A Case Study of FOLIO Codex

XU Lei & XIA Cuijuan

ABSTRACT

In the era of digitalized network, user needs and the forms of library services have developed towards the direction of specialized knowledge and smartness. Facing this new norm, the metadata management of the third generation library service platform not only needs to solve the present situation of the library integrated

* 本文系国家社会科学基金一般项目“面向数字人文研究的图书馆开放数据体系构建与服务模式设计研究”(编号:18BTQ027)的研究成果之一。(This article is an outcome of the project “Research on Library Open Data System Construction and Service Model Design for Digital Humanities Research”(No. 18BTQ027) supported by National Social Science Foundation of China.)

通信作者:夏翠娟,Email:cjxia@libnet.sh.cn,ORCID:0000-0002-1859-6979(Correspondence should be addressed to XIA Cuijuan, Email:cjxia@libnet.sh.cn, ORCID:0000-0002-1859-6979)

management system or even the next generation library service platform beset by MARC to meet the requirements of integrated resource management and business integration but also considers a unified method of data modeling and coding to achieve cross-typed data stream and knowledge organization and fusion. This poses new requirements for metadata management, including: complete life circle management, compatibility of resource types and standard specification, bibliographic control and authority control, and knowledge organization and fusion. The possibility of using platform thinking to change the ecology of library operation shows promise in the FOLIO project. As a library service platform based on micro service architecture, FOLIO at the start of its design has defined a core micro domain—Codex. On the one hand, as the data registry center and data intermediary, Codex is positioned at the top of the FOLIO architecture and can remove differences in encoding and format in different domains and span storage location to reduce the data exchange and system coupling between different domains. On the other hand, based on the BIBFRAME 2 model, Codex designs an abstract data model “Work—Instance—Item” and a set of minimum core metadata elements by drawing from the DC elements. Codex does not describe the detailed information of resources and only play a navigate and positioning role. Exclusive business information or more detailed records can be assessed by getting to the recording level through Codex. As the core metadata management domain, Codex will be designed as the FOLIO platform’s resource link center, authority data center and entry point for resource management, and establish relationships between resources through the abstract entities in the model. Although currently only the Codex Search, which does not store its own data, is realized and the support for linked data still needs to be improved, Codex is flexible enough to support the descriptions and organization of entities, whether it is the abstract data model, or the minimized metadata proposal, or the hierarchical description principle of formal records—working records—unifying records. 6 figs. 6 tabs. 28 refs.

KEY WORDS

Library services platform. Metadata management. FOLIO. Codex.

0 引言

自 1980 年代开始,图书馆就依赖于图书馆集成管理系统(Integrated Library Systems, ILS)对馆藏文献进行管理。ILS 的应用不仅提高了图书馆的自动化水平,优化了业务流程,同时也提升了文献服务的质量,在此过程中, MARC 作为全球图书馆界一致认同和遵循的元数据标准规范,对资源的规范化管理、数据的跨系统传输和交换,起到了积极而关键的作用。随着技术的变革和图书馆上下游环境的改变,以纸质资源管理为核心的 ILS 已不能适应当前数字化、云计算、互联网的时代。为应对这种挑战,一方面,

图书馆界在元数据标准规范的变革上进行了一系列的努力,设计 DC 元数据标准规范体系以适应数字化资源、电子资源和网络资源描述的需求,并逐渐从语义网(Semantic Web)和关联数据(Linked Data)中汲取营养,在内容描述方面推出 RDA 以取代 AACR2,在编码格式上以 BIBFRAME、SchemaBibExt、EDA 等更灵活、更易于扩展、更能与现代数据技术和环境兼容的标准来取代 MARC。另一方面,图书馆软件厂商开始利用最新技术开发“下一代”图书馆系统,也就是“图书馆服务平台”(Library Services Platform, LSP)^[1]。LSP 是采用 SOA 架构的分布式网络系统,利用“软件即服务”(Software as a service, SaaS)的模式运作,试图实现包括纸质资源、数

字化资源和电子资源的统一管理和发现,以适应数字时代图书馆对多样化资源管理和需求的需求。LSP 开发商们还希望能满足图书馆系统功能整合和平台化转型的需求,取代传统分散隔离的 ILS、电子资源管理系统、数字资产管理系统、知识库等^[2]。然而,在元数据管理方面,这些下一代 LSP 仍然难以摆脱 MARC 的束缚,无法很好地支持图书馆界的新型元数据标准规范,究其原因,主要在于缺乏网络时代的平台化思维和相对应的商业模式。

当把平台化思维引入图书馆时,可望催生前所未有的创新和变革。想象一个与 IOS APP 商店类似的图书馆应用市场,图书馆可以自主从中择优选择不同开发商甚至其他图书馆发布的各种应用,然后部署到本地或云端。这种平台化的生态将完全改写现在的图书馆软件开发方式和运营模式,也必将对元数据管理带来新的挑战 and 机遇。这种彻底改变图书馆行业的可能性,可以从 FOLIO (The Future of Libraries is Open) 项目中看到希望。FOLIO 不仅致力于开发一个创新的开源图书馆服务平台,更是一个由图书馆、开发商、供应商及其他利益机构共同组成的协作社区。从元数据管理的角度来考察, FOLIO 要解决的不仅是 ILS 和下一代 LSP 遗留的问题,如何将 MARC 留在过去,支持新的元数据标准规范;同时还要处理新的平台化生态带来的数据互操作问题,当图书馆的业务被分割成块,平台上分布式地部署着不同功能、不同供应商开发、基于不同技术框架的 App,仍然要保证图书馆的各种数据在其中畅通无阻,达到语义上的一致性。FOLIO 以 Codex(元典)为核心的元数据管理模式,试图解决这两方面的问题。

虽然围绕着新一代图书馆系统,国内学者已开展了一系列研究。有对国外 LSP 的介绍、对比与经验总结^[3-6],有对 LSP 电子资源管理的深度评析^[7-8],有从背景、功能需求、特征、知识服务等理论方面的研究^[9-11],也有国内 LSP 的研发经验^[12-14]。针对 FOLIO 平台,肖铮等最先全面地介绍了基于微服务架构的 FOLIO 项目^[15-16],周

义刚等基于北京大学图书馆的需求提出对 FOLIO 的期待^[17],谢蓉等则评价其为真正的第三代图书馆服务平台,代表了未来的发展方向^[18]。但目前尚无对新一代图书馆系统的元数据管理做专门研究。因此,本文从元数据管理的角度,分析第三代图书馆服务平台的元数据管理需求,并介绍上海图书馆“FOLIO Codex & MM”小组对 FOLIO 以 Codex 为核心的元数据管理方案进行分析、测试与评估的方法、过程和阶段性结论。

1 第三代图书馆服务平台元数据管理的需求分析

在这个“数据为王”的时代,第三代图书馆正在经历从信息环境到数据环境、知识环境的转变,数据服务、知识服务、智慧服务成为图书馆服务的新方向^[19]。如阿克夫 DIKW 金字塔 (Ackoff's Pyramid) 所描述的,从底层的数据到信息到知识乃至到理解与智慧,元数据是每个层次上升的粘合剂^[20]。面对这种新常态,第三代图书馆服务平台不仅仅要满足多样化资源的统一管理、业务流程和系统功能的整合,更需要考虑从上而下的统一数据建模和自下而上的标准数据编码,以达到数据在不同应用之间顺畅交换、跨越资源类型的藩篱、在知识组织层面实现融合的目的^[11],这就对元数据管理提出了新的需求。

(1) 完整的生命周期管理。首先,支持元数据从采集、编目、加工、服务到增值、交换、统计分析、评估的完整生命周期是元数据管理最迫切的需求,ILS 和下一代 LSP 由于难以打通不同资源类型、不同系统平台之间的界限,使得元数据在不同的生命周期中处于不同的系统和分散的业务流程中,为资源的管理和服务带来了大量的问题。

(2) 资源类型兼顾。一方面,图书馆资源类型的多样化要求元数据能灵活地描述纸质资源、数字化资源、电子资源、网络资源,不受文本、图像、音视频等媒体格式不同的牵制。另一方面,读者也希望得到基于内容需求而无需区分

资源类型和媒体格式的一体化服务。因此,全媒体的资源管理也成为元数据管理的必备需求。

(3) 标准规范兼容。开放的、分布式的信息环境决定了 MARC 作为唯一的元数据标准规范的时代已经一去不复返,基于 DCMI 抽象模型 (DCAM) 和应用纲要 (DCAP) 的方法,大量适应不同资源类型和不同需求环境的元数据标准规范被制订出来,如中国国家数字图书馆标准规范体系就为家谱、古籍、图像、网络资源制订了相同核心集加不同扩展集的元数据规范。而 BIBFRAME 可以看作是一个框架,在此框架内还可以不断扩展出更多的应用纲要。第三代图书馆系统需要支持灵活可扩展的、适合各种资源类型的元数据标准规范。

(4) 互联网环境下的书目与规范控制。书目控制和规范控制是图书馆的核心职能,也是 ILS 和下一代 LSP,以及第三代图书馆服务平台要实现的核心功能,尤其是互联网时代,要求全网域范围内实现书目控制与规范控制。上海图书馆在数字人文服务平台的设计与开发上,已利用关联数据技术探索了家谱、古籍等文献在互联网上的书目控制,以及人物、地名、收藏机构等规范数据在互联网上的规范控制功能^[21]。对于 ILS 和下一代 LSP 来说,只是在图书馆系统内或者图书馆领域内实现了部分书目控制和规范控制功能,而互联网环境下的书目与规范控制则是对第三代图书馆服务平台提出的需求。

(5) 基于内容的知识组织与知识融合。传统图书馆元数据更多地是对文献外部特征的描述,导致为了管理的方便而因资源的形式划分种类,又因资源种类的不同造成系统和数据的割裂。单纯的 MARC 或 DC 都是以文献的描述记录为最小单位,难以很好地揭示资源在内容层面各种对象的特征和关系,无法应对内容数据化和知识服务的挑战。作为“元”元数据的知识本体和关联数据技术被证明是一种有效的解决方案^[21]。因此,应用语义技术实现内容相关而资源类型无关的知识组织与跨领域的知识融通,是对第三代图书馆服务平台提出的更高需求。

2 FOLIO 的元数据管理

在第二代图书馆服务平台不彻底的技术革新后,FOLIO 显示了第三代图书馆服务平台开放的、社区协同的生态环境。作为起源于第二代,但面向第三代的图书馆服务平台,FOLIO 元数据管理方案既有第二代的特点,又有着鲜明的微服务特征^[22]。

为确保每个微服务之间的松散耦合,FOLIO 提出了“域(Domain)”的概念和秉承最小化的原则。FOLIO 依照图书馆的业务需求将整个平台分为几个不同的域,除 Codex 域作为抽象的元数据管理层外,还有其他具体业务域:知识库域(KB Domain)、馆藏域(Holdings Domain)、采购域(Acquisition Domain)、典藏域(Inventory Domain)、流通域(Circulation Domain)。每个域由多个应用(App)组成,共同完成特定的功能。按照最小化的原则,每个域可以按需设计仅满足本身功能需求的元数据方案^[23]。例如,流通中借出与归还应用不需要理解 MARC 题名字段指示符和子字段的确切含义,它只需要一个显示在屏幕上的题名信息,以确认手头的复本是否被借出即可。同样,一个统计论文使用情况的应用程序不需要区分论文导师和学位候选人等元素。而在不同的域之间起到数据整合和桥梁作用的域称为“Codex”。

2.1 Codex 及其功能与定位

在 FOLIO 基于“域”的松散耦合的资源管理框架中,Codex 被定义为规范的虚拟层,通过抽象的数据模型集成来自不同业务域的元数据,而不受格式、编码或存储位置的制约。同时,Codex 作为一种元数据方案,定义了一套最小的元素集来描述和链接来自不同业务域的不同种类的资源^[24]。它建立了一种机制,允许业务域以尽可能简单的元数据方案来描述本域业务流中的数据流,高效地完成工作,而将与其他域进行数据整合和链接的工作交给 Codex。也就是说,各个域只需关心本域的需求,不用考虑与其他

所有域之间单独地建立数据交换和传输机制,而是通过 Codex 作为数据注册中心和数据链接中介,来消除不同业务域中数据的编码、格式和存储位置的差异。如图 1 所示,其他业务域只需

要与 Codex 进行交互即可,而不用与其他业务域产生依赖关系,这样就在很大程度上减少了业务之间的纠缠,减少了整个平台中数据交互的次数,减轻了系统的压力。

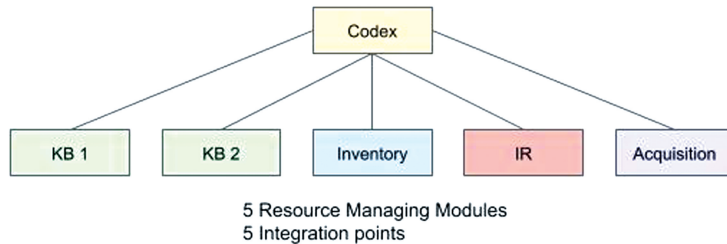


图 1 Codex 的中介分层结构^①

Codex 由一个抽象数据模型和一套最小核心元数据元素集组成。抽象数据模型定义了资源管理所涉及的基本数据对象,而最小核心元数据元素集不描述资源的详细信息,只起到揭

示与定位的作用。当前台用户界面需要调取资源的详细信息时,只需通过 Codex 的定位,通过业务域提供的数据接口来获取详细信息。如图 2 所示,FOLIO 的其他域将其元数据记录通过映射

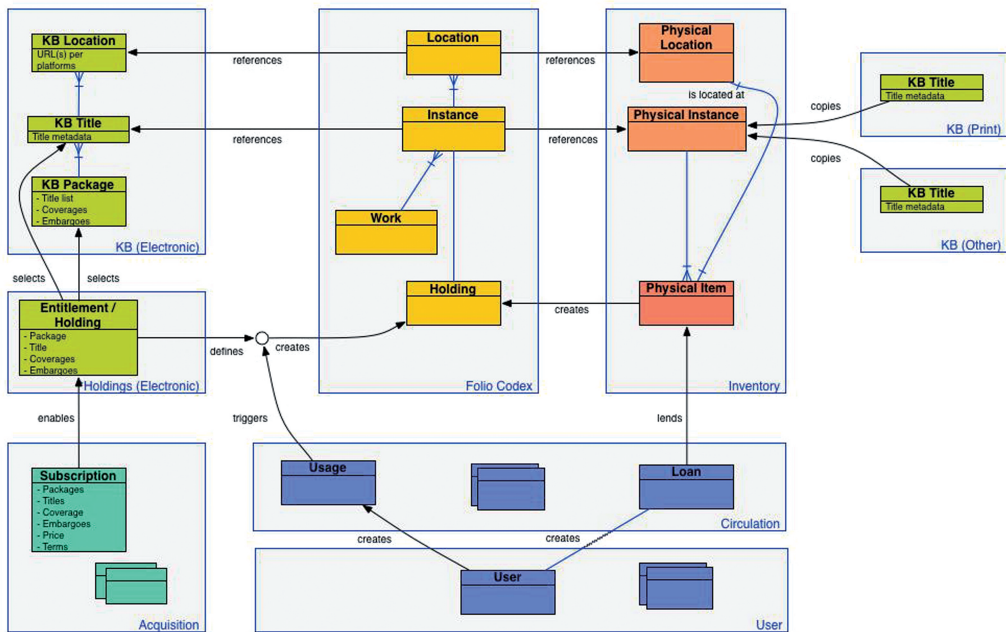


图 2 FOLIO 资源管理域间的数据关系^②

① Bareau V. The Codex vision[EB/OL]. [2019-05-06]. <https://wiki.folio.org/display/PLATFORM/The+Codex+Vision>.

② Resource management data domains[EB/OL]. [2019-05-06]. <https://wiki.folio.org/display/PLATFORM/Resource+Management+Data+Domains>.

转换后在 Codex 域中生成相应的记录。同时, Codex 域也会集中管理规范数据和取值词表。因为此类数据将应用于其他多个模块,如典藏模块、编目模块、数字馆藏模块等。通过 Codex, 其他模块就可以平等便利地获得所需的规范数据服务。另一方面, Codex 的数据记录来源于各业务域, 实施时将从各业务域中提取 Codex 最小核心集所需的部分数据作为 Codex 的数据记录。通过这样的方式, Codex 作为抽象的基本语义层, 以保持足够的灵活性来满足不同业务域中元数据记录间的语义互操作。

FOLIO 对元数据记录的详略应用了分层描述原则, 如图 3 所示。根据目的的不同分成三层^[24]。从下到上依次是:

正式记录层 (Formal Records)。正式记录层作为最详细的元数据或业务数据记录, 保有最原始的、最详细的数据。特定的应用程序可以创建、编辑、验证各种格式的原始编目数据或业

务数据。或者通过数据导入程序将外部数据批量导入到 FOLIO 中, 包括旧系统中的 MARC 记录、BIBFRAME 记录、订单记录等。

业务记录层 (Working Records)。业务记录层作为特定业务域中的元数据记录, 为特定的功能需求服务, 包含有应用程序实现其功能所必需的元素, 并链接到正式记录层的原始编码数据记录。业务记录不需要复制原始编目记录的所有元素。只需保留满足该域业务功能的字段即可。当然在业务记录中也可根据需要新增正式记录所没有的元素。

统一记录层 (Unifying Records), 即 Codex 记录, 目的是提供一个统一和抽象的视图, 对所有域开放。这并不是说 Codex 包含了资源的所有元数据, 而是通过 Codex 记录与各种业务记录相链接, 再通过业务记录链接到正式记录, 从而形成一条从抽象数据到业务数据的最细粒度描述性元数据的链接路径。

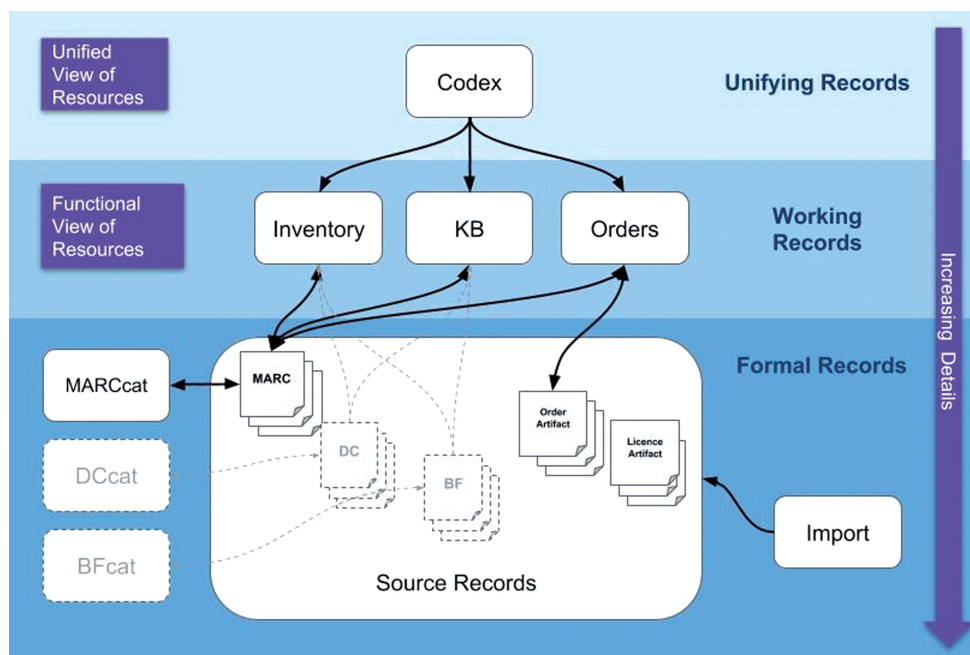


图 3 FOLIO 系统的数据分层^①

① The Codex vision [EB/OL]. [2019-05-06]. <https://wiki.folio.org/display/PLATFORM/The+Codex+Vision>.

2.2 Codex 的数据模型

目前的 Codex 定义了抽象数据模型和元数据方案,数据模型以 BIBFRAME 2.0 为基础。出于 MARC 在图书馆领域可预见的持久应用,以及 BIBFRAME 模型未来潜力的考虑^[25],Codex 基于 BIBFRAME 2.0 设计了“作品—实例—单件/馆藏”的三层核心模型(见图 4)。初始版本中包含了 5 个对象(如图 4 橙色表示)。后续版本可能会加入更多的对象,其中作品(Work)和主题(Subject)是确认会更新的两个对象。Codex 大致遵循“作品—实例—单件”的 BIBFRAME 相似结构,但并不严格。两者不仅对同一个实体对象的属性元素定义不同,Codex 模型还包含了 BIBFRAME 没有的对象:资源包(Package),最新版本已更名为 Container^[26](下文使用 Container)。在数据模型图中,馆藏地(Location)和覆盖范围(Coverage)也作为数据对象被分离出来。这是为了满足对

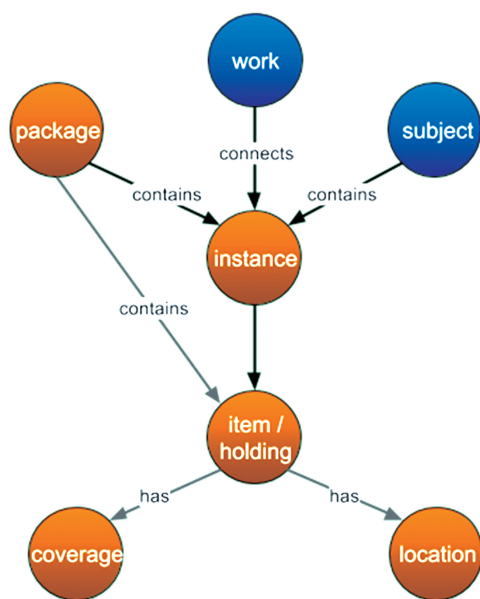


图 4 Codex 的数据模型^①

电子资源的管理,便于被 FOLIO 的其他域重用。但在实践中它们是被嵌套在更高级别的对象中,如单件(Item)或资源包(Container)。

2.3 Codex 的元数据方案

Codex 数据模型中主要数据对象的元数据方案设计则参考了 DC。虽然 Codex 的元数据方案一直处在更新维护中,但最终会类似于 DC,将只保留最小的核心元素集。在最新版中,实例对象有 15 个元素,其中只有 2 个是必备元素。单件/馆藏对象有 23 个元素,其中 12 个元素从实例继承而来,有 3 个是必备元素。资源包有 13 个元素,其中 2 个为必备元素。馆藏地数据有 7 个元素,只有 1 个是必备元素。覆盖范围数据有 4 个元素,其中 1 个是必备元素。

2.3.1 实例(Instance)数据对象的元素集

实例表示一个资源的主要对象,也是一个聚合电子和物理资源的抽象对象。它可能包含一些 BIBFRAME 2.0 归在作品中的属性。实例数据元素见表 1。

2.3.2 单件/馆藏(Item/Holding)数据对象的元素集

单件/馆藏有两重含义:①单件用来表示实例中特定复本意义上的单件;②馆藏描述图书馆或机构与该单件的关系。从数据建模的角度来看,描述它们的数据元素有很多重叠,只用一个对象来表示两个概念是合理的。本文中所述的“单件”可理解为“单件/馆藏”。另外,本文中的馆藏不仅仅是实际拥有,也包括机构可访问的资源。

单件表示一个实例的某种特定材料的复本。换句话说,单件只有一个实例作为其上位类对象。因此,单件可以从其父实例继承多个元素,这些元素将直接出现在单件对象记录中。具体的单件/馆藏数据元素见表 2,不包含继承自实例的 12 个元素。

^① The Codex metadata model. [EB/OL]. [2019-05-06]. <https://wiki.folio.org/pages/viewpage.action?pageId=1415393>.

表 1 实例数据元素

元素	子元素	值类型	必备性	可重复性	说明
id		字符串	Y		实例记录的内部 id(UUID)
题名		字符串	Y		资源的题名信息(或标签)
交替题名		字符串			资源的任何可替代正式题名的其他题名
丛编		字符串			资源的丛编题名信息
责任者		字符串		Y	和资源相关的一组第二作者或创建者
出版者		字符串			资源的出版、发行、制作者
日期		字符串			与资源相关的日期(如出版日期)
类型		字符串			资源类型(如连续性资源、专著、音频等)
格式		字符串			资源的载体材料类型或发行格式(如精装、盒式磁带、dvd 等)
标识符		对象		Y	资源标识符,由标识符名称一值构成
	类型	字符串			标识符类型,如 ISSN
	值	字符串			标识符值
来源		字符串			元数据来源
语种		字符串		Y	资源的语种
权限		字符串			资源的访问权限
版本		字符串			专著的版本
最新修改日期		字符串			管理元素,Codex 中资源信息修改的最新时间

表 2 单件/馆藏数据元素

元素	值类型	必备性	可重复性	说明
id	字符串	Y		单件记录的内部 id(UUID)
实例对象 ID	字符串	Y		父实例对象的 ID
描述	字符串			关于资源的自由文本描述
复本数量	数字			复本的数量
条码号	字符串			单件的馆藏条码号
供应商包	字符串			与单件有关的资源供应商的资源包 (不要与 Codex 的资源包对象混淆)
覆盖范围	对象			详见覆盖范围对象
馆藏地信息	对象		Y	参见馆藏信息对象,一个单件可以有多个馆藏信息,如永久馆藏和临时馆藏
单件状态	字符串			用于说明单件的馆藏状态,如订购中
被选定	布尔值			“holding”的属性,单件的选购状态,一般用于说明电子资源是否有馆藏或购买
被定制	布尔值			管理信息标示,用于说明继承自父实例的元素内容是否被覆盖

2.3.3 资源包(Container)数据对象的元素集 的资源包中,因为这些包表面上是独立的分组,资源包表示实例和/或单件的容器。因此,并且可以依据多种可能的方式对资源进行分组。它具有将这些对象按任意需求组合在一起的能力。资源包数据元素见表3。一个给定的实例或单件可以放在任意数量

表3 资源包数据元素

元素	值类型	必备性	可重复性	说明
id	字符串	Y		资源包记录的内部 id(UUID)
标识符	字符串			用于追溯资源包的外部标识符(如条码号)
名称	字符串			资源包的名称或标签
描述	字符串			关于资源包的自由文本描述
类型	字符串			资源包的类型,表示是什么样的资源集合。如 DVD 套装,合并装订,档案箱等
资源供应商 id	字符串			连接到资源的供应商对象(也可以在另外一个域)
资源供应商	字符串			用于展示目的的供应商名称
平台	字符串			电子资源的使用平台
单件列表	数组	Y		资源包内所含实例或单件的列表
单件数量	数字			资源包中所含单件的数量
馆藏单件数量	数字			一般用于电子资源,用于说明资源包中馆藏资源的数量
覆盖范围	对象			用于资源包本身的覆盖范围,而不是资源包中内容的覆盖范围(参见 Content 对象)
已选购	布尔值			资源包自身的购买状态,而不是资源包中的内容,一般用于电子资源

Codex 资源包不应与电子资源知识库中的资源包对象相混淆。后者是资源供应商用于商业目的的捆绑机制;前者是一个更广泛的概念,可以用来代表各种类型的资源分组。当然,这些商业资源包会在 Codex 资源包描述下用特定的类型表示。不管怎么样,它只是众多的 Codex 资源包类型中的一种。

除了作为其他对象的容器之外,资源包本身也是一个对象。它包含许多用于描述它的元数据元素,这些元素不在容器的有效载荷之内。资源包对象可以作为一个整体,以非常强大的方式操作资源。

2.3.4 馆藏地(Location)数据对象的元素集

尽管馆藏地对象通常被嵌套在单件或资源

包对象中。但是,为确保在 FOLIO 其他部分的潜在重用价值,馆藏地对象还有自己的元素,便于对复杂的馆藏地信息进行细粒度的结构化描述,如特定资源的位置或服务地址。为了描述实体资源的物理地址,它提供了多层结构的地址组件。最顶层的是 Institution-Campus-Library 结构,它力求在 FOLIO 的不同设施和不同租户之间提供一致性。在这三个层次下是强大的“Parking”元素,它是多个“元素—值”对的集合。它可以包含一些标准的条目(例如 LC 索书号)以及机构或者图书馆特定的地址组件(例如排架号或部门)。“平台”元素和“URI”元素则是为了满足电子资源揭示的需求。馆藏地数据元素见表4。

表 4 馆藏地数据元素

元素	子元素	值类型	必备性	可重复性	说明
id		字符串	Y		馆藏信息记录的内部 id (UUID)
机构		字符串			馆藏地的最高一级管理机构(例如大学)
校园		字符串			地理层级的馆藏信息(例如城市)
图书馆		字符串			建筑级别(例如单个图书馆)
典藏位置		对象		Y	扩展的“名称—值”资源集合。可以表述一般的馆藏信息属性(如索书号)或图书馆具体的馆藏信息的细节(例如排架号或部门)
	名称				典藏位置的名称(如 LC 索书号、排架号)
	值				具体的典藏位置名称的值
平台		字符串			电子资源发行平台
URI		字符串			电子资源的访问链接

2.3.5 覆盖范围(Coverage)数据对象的元素集

覆盖范围对象允许定义资源内容的收录范围,这些范围可以是日期,也可以是其他单元,如卷和发行刊号。范围也可以用来表示与专著

相关的单个日期。此外,覆盖范围对象还支持表示资源使用的禁止和开放状态。覆盖范围数据元素见表 5。

表 5 覆盖范围数据元素

元素	子元素	值类型	必备性	可重复性	说明
id		字符串	Y	Y	覆盖范围记录的内部 id (UUID)
范围		对象		Y	描述资源覆盖的范围 起始值与终止值构成一个范围说明 终止值可选(如具体的一个单独的日期) 范围可以是具体的时间(日期)也可以是发行的卷期单元
	开始	字符串			范围的起始值
	结束	字符串			范围的终止值
说明		字符串			资源覆盖范围的自由文本说明
禁止范围		对象		Y	描述资源禁止使用的一组范围
	开始	字符串			禁止的开始
	结束	字符串			禁止的结束

2.4 Codex 与 MARC、DC、BIBFRAME 等其他元数据标准的关系

Codex 的数据模型和元数据方案的设计既未完全遵循 BIBFRAME,也不是单纯的 MARC

或者 DC,而是实用主义地借鉴了这些元数据标准的优势,同时秉承最小化的原则,以简单实用为首要考虑。

MARC 作为图书馆领域历史最悠久且仍占

据主流位置的元数据格式,在 FOLIO 中以特别的方式进行管理。在 FOLIO 微服务架构中, MARC 数据可以通过“数据导入”(Data import)应用或其他工具导入到典藏域中。当 MARC 数据导入系统后, FOLIO 将从 MARC 记录中抽取部分相关字段,将其映射到 Codex 有限的核心元素集。原 MARC 数据将作为附件链接到新的 Codex 记录中。

Codex 的元数据方案设计参考了 DC 的应用纲要设计思想。一方面, Codex 各数据对象的元素集是该类数据对象主要特征的基本要素,对于其他业务域的元数据方案来说,是一个最小的核心集,各业务域可以根据需要在此核心集的基础上自行扩展。另一方面, Codex 数据对象的每个元素定义沿用了最简单的 DC 标准,对元素的描述对象、取值范围、必备性和可重复性做了明确的规定。

Codex 数据模型直接受到 BIBFRAME 的启发,以面向对象的思维将元数据管理中涉及的不同实体类型作为不同的数据对象分别对待,理清其中的关系,定义每个数据对象的属性。也沿用了 BIBFRAME 2.0 “作品—实例—单件”的三层架构,同时还做了特定的扩展。但 FOLIO 避免了像 BIBFRAME 2.0 那样逐步陷入 MARC 的复杂和细节,放弃了直接实施 BIBFRAME 2.0,只是考虑在两者之间建立映射,而一致的面向对象思想和相似的框架将使这种映射具有令人信服的可行性(见表 6)。

表 6 Codex 与 BIBFRAME 核心对象间映射

BIBFRAME 2	Codex
Work	Work
Instance	Instance
Item	Holding
physical Location electronic Locator	Location

2.5 Codex 的现状与愿景

Codex 作为 FOLIO 架构最顶层的微服务域,

试图通过定义一个统一的抽象的数据模型和最小核心集的元数据方案,成为整个系统的资源链接中心、规范数据中心、跨域的全媒体资源管理入口,并通过模型中的抽象实体揭示资源间关系。除了当前 demo 系统可见的跨域检索 Codex Search 应用外,还包括:编目数据导出(Catalog Exports)、规范控制(Authority Control)、资源关系管理(Resource Relationships)等。

关于与遗留系统中元数据的关系,通过 Codex, FOLIO 声称其元数据管理能够实现“将 MARC 抛在后面、支持关联数据、与遗留系统中各种不同的元数据格式无关”这样几个重要目标。FOLIO 系统外的元数据记录,无论是什么格式,都将与 FOLIO 的元数据规范映射后抽取 FOLIO 的 Codex 和各业务域中仅需要的元素导入到 FOLIO 系统中,而将源记录另行存储,不进入 FOLIO 的业务流,仅用作与遗留系统间的数据交换。

上海图书馆的 Codex & MM 小组通过文献调研和测试,发现上述 Codex 的愿景在当前版本的 FOLIO 系统中尚未完全实现。当前版本的 FOLIO 的 Codex 域只有 Codex Search 的功能,属于一个初始版本(见图 5)。在初始版本的 Codex 中, Codex Search 通过分发查询条件到各个资源管理域,最后在 Codex-Mux 模块中进行结果合并与排序,并非通过生成和存储 Codex 数据记录来实现^[27]。但在后续版本中, Codex 域会根据所定义的抽象数据模型和元数据方案生成 Codex 数据记录,实现对跨域资源之间关系的管理与揭示,以满足除 Codex Search 之外其他应用程序的功能需求。另外, Codex Search 也会作为一个嵌入式组件,组合在其他模块中,以便在 FOLIO 的任何地方对资源进行检索与定位(见图 6)。

对于遗留系统中的数据导入问题,上海图书馆的 Codex & MM 小组也进行了测试,将 1 000 种上海联编中心的 BIBFRAME 格式书目数据和 Horizon 系统中的 300 种 CNMARC 格式书目数据、馆藏数据、读者数据、流通数据导入到 FOLIO

Codex: initially

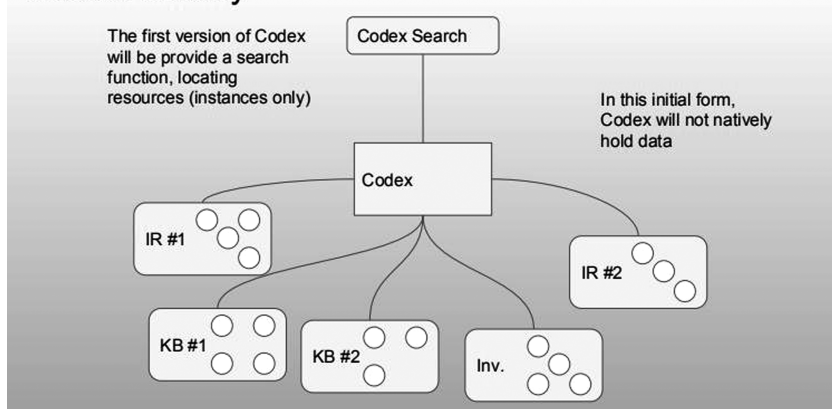


图5 FOLIO 当前版本缺乏 Codex 数据记录的 Codex 域^①

Codex: later

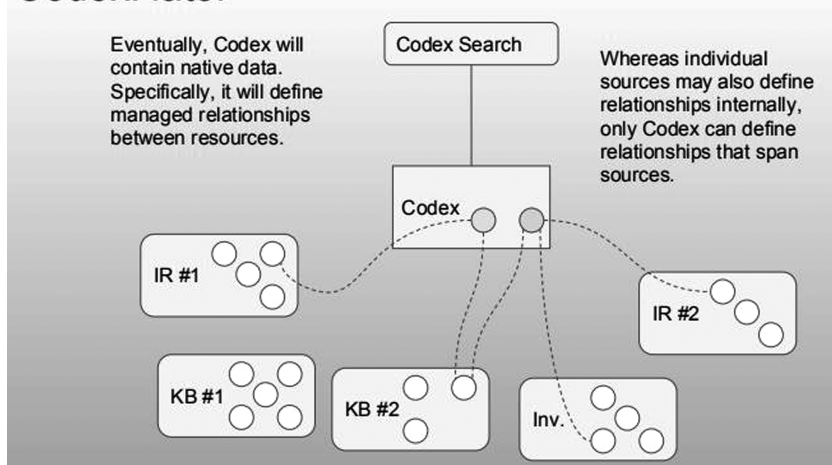


图6 FOLIO 未来版本中存储 Codex 数据记录的 Codex 域^②

的典藏域中。在数据导入的方式上, FOLIO 当前版本的数据导入(Data Import)应用通过在前端设置映射批量导入数据的功能尚未进入可用阶段。项目组针对测试数据, 设计了 BIB-FRAME2.0 和 CNMARC 与 FOLIO 典藏域元数据方案之间的映射, 利用 FOLIO 提供的 API, 进

行了一定的定制开发, 以非侵入式而非直接操作 FOLIO 数据库表的方式导入了上述测试数据, 实现了在 Codex Search 中可检索, 在前端界面可浏览。在 FOLIO 的实际实施中, 针对其现有元数据方案, 必然需要考虑到如何适应中文领域具体需求而进行扩展的问题, 通过测试数据

① Bareau V. Folio Architecture Update 2018-03-20[EB/OL]. [2019-05-07]. <https://wiki.folio.org>.

② Bareau V. Folio Architecture Update 2018-03-20[EB/OL]. [2019-05-07]. <https://wiki.folio.org>.

导入,项目组也对此问题进行了探索,发现当前版本的 FOLIO Inventory 域在数据导入时,支持元数据元素编码体系(取值词表)的扩展,但尚不支持元数据元素的扩展。这需要 FOLIO 项目在接下来的实施过程中统一考虑,制订元数据方案扩展的原则和规范。

3 结语

从纸质文献描述元数据标准 MARC 开始,到数字图书馆时代的 DC,再到 MARC 与 DC 的折中方案 MODS,以及最新的概念模型与本体词表,如 BIBFRAME、Schema.org 等,图书馆元数据标准正在经历从印本到多媒体,从单一标准到应用纲要,从文献描述到知识组织,从机器可读到语义互操作,从专业领域到开放互联的变革。因此,第三代图书馆服务平台的元数据管理也有了不同以往的功能需求,不仅需要支持资源的整合与一体化,更需要支持标准规范的灵活性和可扩展性,以及数据的开放性与语义互操作。

FOLIO 的 Codex 在元数据管理方案设计的方法上,为了既满足新的需求,又要兼顾实用性和可行性,有诸多创新,颇有启发性。如 Codex 元素集在设计之时,坚持最小化的思想,根据核心职能判断元素的取舍。即如果数据元素在检索(包括限定与过滤)中有用,则必须通过 Codex 获取;如果元素在交叉检索结果的显示中有用,则必须通过 Codex 获取。如果元素只与特定的

域有关(如采访),则不应该映射到 Codex^[28]。另外,FOLIO 社区通过描述关键用户案例,分析实现每个业务域功能所需的元素,然后按照开发优先级进行元素设计。设计时优先考虑了与描述印本资源的 MARC 标准之间的映射,其次是电子资源元数据标准,最后才是 BIBFRAME、EAD、DC 等其他标准。

FOLIO 平台的以社区驱动基于微服务架构的创新开发模式,赋予了图书馆更大的自主权,使其能对新一代图书馆系统进行个性化定制。更重要的是,数据作为一种战略资源,第三代图书馆服务平台真正实现了数据的自管。FOLIO 的 Codex 元数据方案,融合了 BIBFRAME 模型特征和 DC 的最小化原则,不仅可以满足未来面向实体对象的元数据环境,而且也降低了元数据互操作的难度。在 FOLIO 微服务架构顶层,Codex 作为一种抽象数据模型,在不影响现有元数据格式与工作流程的前提下定义了资源间的关系,在业务流中建立了数据的链接。同时在底层的格式设计中,Codex 也是一种规范化的数据编码格式,超脱于原生数据格式,实现了跨 APP 的语义互操作。虽然现阶段的 Codex 并没有很好地支持关联数据技术下的实体描述与关系揭示,但无论是其“作品—实例—单件”的核心模型,还是“统一记录—业务记录—正式记录”的层次化元数据记录模型,Codex 都有足够的灵活性支持深度的知识组织与融合,实现图书馆的智慧服务。

参考文献

- [1] Breeding M. Smarter libraries through technology: the beginning of the end of the ILS in academic libraries[J/OL]. Smart Libraries Newsletter,2011,31(8)[2019-06-05]. <https://journals.ala.org/index.php/sln/issue/viewIssue/302/64>.
- [2] Breeding M. Beyond the ILS: a new generation of library services platforms[M]//Robots in academic libraries: advancements in library automation. IGI Global, 2013: 13-36.
- [3] 陈武,王平,周虹.下一代图书馆服务平台初探[J].大学图书馆学报,2013,31(6):82-87.(Chen Wu, Wang Ping, Zhou Hong. Comparative study on next generation library services platforms[J]. Journal of Academic Libraries,2013,31(6):82-87.)

- [4] 包凌,赵以安. 国外下一代图书馆自动化系统的实践与发展趋势研究[J]. 图书馆学研究,2013(9):58-65. (Bao Lin, Zhao Yi'an. Research on practice and development trend of the next generation library automation system in foreign countries[J]. Research on Library Science,2013(9):58-65.)
- [5] 李娟,张雪蕾,杨峰. 基于实证分析的下一代图书馆服务平台选择策略——以 ALAM、Kuali OLE、OCLC WorldShare 和 Sierra 为例[J]. 图书与情报,2017(3):84-92. (Li Juan, Zhang Xuelei, Yang Feng. Empirical analysis on the choice strategy of next generation library services platforms; taking ALAM, Kuali OLE, OCLC WorldShare and Sierra for example [J]. Library & Information,2017(3):84-92.)
- [6] 刘斌,黄婧. 下一代图书馆服务平台 Alma 迁移实践——以北京师范大学图书馆为例[J]. 图书情报工作,2019,63(4):79-85. (Liu Bin, Huang Jing. The migration practice of the next generation library service platform Alma: taking Beijing Normal University Library as an example[J]. Library and Information Service,2019,63(4):79-85.)
- [7] 田晓迪,孙博阳. 下一代图书馆服务平台的电子资源全流程管理功能——以 Alma 为例[J]. 图书情报工作,2016,60(17):65-69. (Tian Xiaodi, Sun Boyang. Electronic resources management based on the library service platform Alma[J]. Library and Information Service,2016,60(17):65-69.)
- [8] 刘素清. 从电子资源管理视角分析我国高校图书馆服务平台的发展[J]. 大学图书馆学报,2018,36(4):11-17. (Liu Suqing. Exploring library services platform for Chinese academic libraries from the perspective of electronic resources management[J]. Journal of Academic Libraries,2018,36(4):11-17.)
- [9] 殷红,刘炜. 新一代图书馆服务系统:功能评价与愿景展望[J]. 中国图书馆学报,2013,39(5):26-33. (Yin Hong, Liu Wei. On the new generation of library service system: functions, visions and prospects[J]. Journal of Library Science in China,2013,39(5):26-33.)
- [10] 刘炜. 关于“下一代图书馆系统”的思考[J]. 国家图书馆学刊,2015,24(5):7-10. (Liu Wei. Thoughts on “the next generation library system”[J]. Journal of the National Library of China,2015,24(5):7-10.)
- [11] 张磊,贺晨芝,赵亮. 面向数据与知识服务的第三代图书馆服务平台[J]. 国家图书馆学刊,2018,27(6):40-47. (Zhang Lei, He Chenzhi, Zhao Liang. The third generation library service platform oriented to data and knowledge services[J]. Journal of the National Library of China,2018,27(6):40-47.)
- [12] 杨新涯,魏群义,许天才,等. 论新一代图书馆系统的特征[J]. 图书馆论坛,2017,37(7):2-8. (Yang Xinyaya, Wei Qunyi, Xu Tiancai, et al. The characteristics of new generation library system[J]. Library Tribune,2017,37(7):2-8.)
- [13] 邵波,张文竹. 下一代图书馆系统平台的实践与思考[J]. 图书情报工作,2019,63(1):98-104. (Shao Bo, Zhang Wenzhu. Practical research and thinking of the next generation library system platform[J]. Library and Information Service,2019,63(1):98-104.)
- [14] 吴元业. 下一代图书馆服务平台 CLSP[J/OL]. 图书馆论坛,2019[2019-08-08]. <http://kns.cnki.net/kcms/detail/44.1306.G2.20190808.1619.002.html>. (Wu Yuanye. Study on next generation library service platform CLSP[J/OL]. Library Tribune,2019[2019-08-08]. <http://kns.cnki.net/kcms/detail/44.1306.G2.20190808.1619.002.html>.)
- [15] 肖铮,林俊伟,陈丽娟. 下一代图书馆开放服务平台 FOLIO 初探[J]. 图书馆学研究,2018(15):34-38,63. (Xiao Zheng, Lin Junwei, Chen Lijuan. The research of FOLIO as a next generation library service platform [J]. Research on Library Science,2018(15):34-38,63.)
- [16] 肖铮,林俊伟. 用微服务构架下一代图书馆服务平台——以 FOLIO 为例[J]. 图书馆杂志,2018,37(11):63-69. (Xiao Zheng, Lin Junwei. Building the next generation library service platform with microservices archi-

- ecture: taking FOLIO as an example[J]. *Library Journal*,2018,37(11):63-69.)
- [17] 周义刚, 聂华. 新一代图书馆服务平台调研及思考——基于北京大学图书馆的需求[J]. *图书馆杂志*, 2019,38(2):69-78. (Zhou Yigang, Nie Hua. Research and thoughts on the new generation of library services platforms: based on the needs of Peking University library[J]. *Library Journal*,2019,38(2):69-78.)
- [18] 谢蓉, 刘炜, 朱雯晶. 第三代图书馆服务平台: 新需求与新突破[J]. *中国图书馆学报*,2019,45(3):25-37. (Xie Rong, Liu Wei, Zhu Wenjing. The third generation library services platforms: new requirements and new breakthroughs[J]. *Journal of Library Science in China*,2019,45(3):25-37.)
- [19] 初景利. 新时代图书馆与新型服务能力建设[EB/OL]. [2019-08-14]. http://202.198.25.38/tgw/editor/attached/file/20180525/20180525123248_95480.pptx. (Chu Jingli. New service capability construction of library in the new era[EB/OL]. [2019-08-14]. http://202.198.25.38/tgw/editor/attached/file/20180525/20180525123248_95480.pptx.)
- [20] Gartner R. Metadata: shaping knowledge from antiquity to the semantic web[M/OL]. Switzerland: Springer, 2016:9-11[2019-06-04]. <https://www.springer.com/us/book/9783319408910>.
- [21] 夏翠娟, 张磊, 贺晨芝. 面向知识服务的图书馆数字人文项目建设: 方法、流程与技术[J]. *图书馆论坛*, 2018(1):1-9. (Xia Cuijuan, Zhang Lei, He Chenzhi. Construction of library digital humanities projects for knowledge services: method, process and technology[J]. *Library Tribune*,2018(1):1-9.)
- [22] Bateau V. Resource management data domains[EB/OL]. [2019-05-16]. <https://wiki.folio.org/display/PLATFORM/Resource+Management+Data+Domains>.
- [23] Behara S. Breaking the monolithic database in your microservices architecture[EB/OL]. [2019-05-16]. <https://dzone.com/articles/breaking-the-monolithic-database-in-your-microserv>.
- [24] Bateau V. The codex vision[EB/OL]. [2019-05-06]. <https://wiki.folio.org/display/PLATFORM/The+Codex+Vision>.
- [25] 2017-07-07 FOLIO data model/codex working group meeting notes[EB/OL]. [2019-05-06]. <https://wiki.folio.org/pages/viewpage.action?pageId=1414350>.
- [26] 2018-07-24 metadata management meeting notes[EB/OL]. [2019-05-06]. <https://wiki.folio.org/display/MM/2018-07-24+Meeting+note>.
- [27] Bateau V. FOLIO architecture update 2018-03-20[EB/OL]. [2019-05-07]. <https://wiki.folio.org>.
- [28] Harnish K. Codex metadata analysis[EB/OL]. [2019-05-06]. https://docs.google.com/presentation/d/1AZxRICVO3If85xv78oUdazE8kBoa11jq4548Ms6EiW4/edit#slide=id.gc6f980f91_0_0.

许磊 上海图书馆助理馆员。上海 200031。

夏翠娟 上海图书馆系统网络中心高级工程师。上海 200031。

(收稿日期:2019-10-12)