



题目: 利用关联开放数据¹实现规范记录的数据聚合与发布

作者²:

Xavier Agenjo, Project Manager Fundación Ignacio Larramendi and director of Polymath Virtual Library [xavier.agenjo@larramendi.es]

Francisca Hernández. Metadata Librarian. Consultant. DIGIBÍS, Producciones Digitales [francisca.hernandez@digibis.com]

Viedma, Andrés. Software architect. DIGIBÍS, Producciones Digitales [andres.viedma@digibis.com]

中文翻译 / Chinese Translators:

黄亮、刘丹（中国国家图书馆）

HUANG Liang & LIU Dan, National Library of China

Meeting:

80 — Cataloguing: breaking barriers — Cataloguing Section

摘要:

在对博学家虚拟图书馆的分析中，介绍了利用关联开放数据实现规范记录的数据聚合与发布。目的是源源不断地收集关于西班牙、美国西班牙裔、巴西和葡萄牙博学家信息、数据、数字化文本以及网址。照此，它聚合来自于西班牙、美国西班牙裔、葡萄牙和巴西的关于思想、哲学、政治、科学等各文种（拉丁语、阿拉伯语、希伯来语、西班牙语、葡萄牙语...）各时期（自公元前一世纪塞内加至今）的信息。

整个系统的中枢是著者。每个著者的 MARC21/RDA 格式的规范记录都是根据传记资料进行建立并完善的。将特定属性加以分类以加强各个面（专业、职业、性别、会员资料、生卒年、生卒地以及使用的语言）的关联度和导航性。这样，每条规范记录都从多种来源和词汇表中聚合信息。并且根据 MARC/RDA（12 期更新），每个著者作品的数字化版本也在记录中有所描述。

¹ 所有数据和 URL 都已在 2011 年 4 月 30 日核实并更新。

² 感谢 DIGIBIS 研发部门的 César Juanes 对此文的审阅。

类似地，每个著者都与其他责任者（译者、出版者、注释者等）相关联，作为传送他们作品各文本的一种方式而使用著者-题名规范记录，该记录将按照 MARBI 分类为 WEMI。

规范记录的建立参考了不同的可靠来源（规范文档、百科、传记辞典等）。2010 年之前，著者与所找到的相关数据之间的关联都是由人工建立的。2011 年，伊格纳西奥拉腊门迪基金会（Fundación Ignacio Larramendi）正致力于开发从关联开放数据源中半自动地获取数据的不同工具。此文详述了获取 LOD 资源的 URIs 的过程，URIs 利用 SRU/OpenSearch 服务器和 SPARQL 终端产生自动查询，或通过 LOD 上可利用的文档充实语义。在根据 LCSH、VIAF、GeoLinkedData（来自于国家地理研究所）及 GeoNames 词表充实语义之后，数据以 MARC21、Europeana 数据模型 5.2.1、SKOS 和 VIAF 格式导出。

1. 引言：博学家虚拟图书馆

以下关于关联开放数据中规范记录的聚合与发布的思考并没有停留在理论层面，而是基于正在进行中的图书馆实践。这些编目工作是在作为伊格纳西奥拉腊门迪基金会³虚拟图书馆一个组成部分的博学家虚拟图书馆⁴中进行的。博学家虚拟图书馆建设的基本目标是赋予西班牙、葡萄牙、巴西和拉丁美洲博学家作品以特殊的关联性和重要性，使之与其他拥有相似特征的作品相关联，并使其在国际上被更广泛利用，这正是书目工作的一个主要特点。

尽管看似显而易见，本综述还是从一开始就强调了前期书目设计的重要性，书目设计应该优先于技术问题，特别是在像关联开放数据之类的新事物上。实际上，这些书目的目标是成为博学家虚拟图书馆编目实践和具体开发的功能性需求的基础。此项目旨在收集和连接书目信息，或者仅仅是一些信息，这些信息是从历史的角度关于伊比利亚美洲思想，利用汤因比（Toynbee）⁵的概念建立与其他“文明”的关联。

³ 关于该项目的最新进展，参见如下文章：Agenjo Bullón, Xavier y Hernández Carrascal, Francisca. *La Biblioteca Virtual Larramendi: fuente de información bibliográfica para el pensamiento iberoamericano en la Web 3.0*. En: *Jornadas Virtuales Iberoamericanas de Bibliotecología*. [<http://goo.gl/VbwtN>]. 该文章写于 2010 年秋季，项目从那个时候起有一些变化，可以参见：Agenjo, Xavier y Hernández, Francisca. *La Biblioteca Virtual Ignacio Larramendi desde la perspectiva LOD y EDM that was presented at the I Seminario Internacional de la Biblioteca de Galicia* [<http://goo.gl/uFrXx>].

⁴ <http://goo.gl/3kn00>

⁵ 我们可以回忆汤因比在其《历史研究》（*A Study of History*）中可以确定多达 21 种不同的文明，而且尽管这个信息可能看上去过时，甚至亨廷顿在其名著（从我们的观点出发是其随后的著作）中所用的文明的概念也有点过时，事实上各种文化之间存在孤立区域无疑也有信息交流，特别参见 Tim Berners-Lee 的 *Design Issues: Linked Data* [<http://www.w3.org/DesignIssues/LinkedData>].

规范记录的信息源

很明显，对于书目项目的目标有用的多数信息源都没有数字格式，或者缺少能够被数字化的合适结构。在其他情况下，信息源可在网络上获得，而界面设计却使信息容量和效用大打折扣。例如，《皇家历史学院传记辞典》（*Diccionario Biográfico Español de la Real Academia de la Historia*）提供了大量与西班牙历史和西班牙语国家文化有关的人物的主要信息，但却没有以技术手段提供其数字化信息。另一信息源是出版于1980年至2003年的Gonzalo Díaz y Díaz的《西班牙哲学人物和文件》（*Hombres y documentos de la filosofía española*），共7卷，内容庞杂而缺少启发式方法，但却包含了极大的信息量⁶。同样费雷塔莫拉（Ferrater Mora）的《哲学大辞典》（*Diccionario de filosofía*）⁷也提供了极大的信息量，从智识的角度更加严格，但信息本身却很缺乏。

很遗憾，纸质发行的文献会有时效性，并且受版权保护，因此在网络上是不能获得的，尤其是在关联开放数据中。矛盾的地方是，根据各国的相关立法，大型回溯数字化项目在著者离世50年，70年，甚至是80年之后就停止了，因此完全过时的内容很容易获取，而其他更有价值以及最新的内容却不能获取。我们经常忽略开放存取的一个根本问题，这个问题不是数字信息是否可以获得，而仅仅是这样的数字信息是否存在。

这个问题应在图书馆界的会议上引起注意，对于图书馆员来说，书目更应该是一个基础学科，而我们却忘了通常带有各种手稿的最好的评述版以及初步研究等在网络上找不到资源。当然，一些项目显然应用了历史文献的数字化技术，但这始终是少数。鉴于其特殊性，我们引用马克·吐温项目（*Mark Twain Papers & Project*）⁸作为例子来说明相关作者的数字版本。

可喜的是，博学家虚拟图书馆至少已经开始尝试与学者探讨，目的在于介绍具体到每个著者（博学者）的书目研究状态。此外，还确定了在任何情况下都需要参考的起码的信息源：《皇家历史学院传记辞典》⁹，《西班牙国家图书馆规范文档》（*Fichero de Autoridades de la Biblioteca Nacional*）¹⁰，《虚拟国际规范

⁶ *Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico* [<http://goo.gl/NHiVW>]. 文字不很全面.

⁷ Ferrater, Mora J. *Diccionario de Filosofía*. Madrid: Alianza, 1979. 这是 Ferrater Mora 的最终版，此后一个扩大版出版，以 Ferrater Mora 为首，由 Jose Maria Terricabras 教授主持，却被附加了一个星号。

⁸ <http://bancroft.berkeley.edu/MTP/>

⁹ <http://www.rah.es/diccBiografico.htm>

¹⁰ <http://catalogo.bne.es/uhtbin/authoritybrowse.cgi>

文档》(Virtual International Authority File)¹¹ (包括西班牙国家图书馆和葡萄牙国家图书馆的规范文档), 美国国会图书馆规范库 (Library of Congress Authorities)¹², 以及上文提到的《西班牙哲学人物和文件》、《哲学大辞典》。其他常用信息源包括《斯坦福哲学百科全书》(Stanford Encyclopedia of Philosophy)¹³, 《天主教百科全书》(The Catholic Encyclopedia)¹⁴, 《天主教百科全书》(Enciclopedia católica)¹⁵, 犹太虚拟图书馆 (Jewish virtual Library)¹⁶, 伊斯兰哲学在线 (Islamic philosophy Online)¹⁷, 《西班牙工程哲学》(Proyecto Filosofía en Español)¹⁸, 米拉莱斯图书馆 (Biblioteca Miralles)¹⁹, 萨维德拉法哈多图书馆 (Biblioteca Saavedra y Fajardo)²⁰, 塞万提斯虚拟图书馆 (Biblioteca Virtual Miguel de Cervantes)²¹, 萨米恩托项目 (Proyecto Sarmiento)²²。

从规范记录到书目记录

VIAF 是优先被纳入到规范记录信息源中的。使用 VIAF, 可以在很多规范信息源中查找到著者, 例如国会图书馆, 大英图书馆以及德国国家图书馆。VIAF 最大的优势在于从规范记录群可以导航到 VIAF 不同规范文档的每个标目。这就使根据信息源优先选取标目成为了可能: 西班牙国家图书馆编辑西班牙著者的规范标目, 葡萄牙国家图书馆编辑葡萄牙著者的规范标目, 巴西国家图书馆²³ 编辑巴西著者的规范标目等。并且为了实现共享, 这些著者规范标目都可以在国会图书馆中查找到。有一点是很必要的, 通过查询, 可以建立重要的广泛的新特征,

¹¹ <http://viaf.org/>

¹² <http://authorities.loc.gov/>

¹³ <http://plato.stanford.edu/>

¹⁴ <http://www.newadvent.org/cathen/>

¹⁵ <http://ec.aciprensa.com/>

¹⁶ <http://www.jewishvirtuallibrary.org/>

¹⁷ <http://www.muslimphilosophy.com/>

¹⁸ <http://www.filosofia.org>

¹⁹ <http://www.bibliotecamiralles.org/escritores.html>

²⁰ <http://saavedrafajardo.um.es/Biblioteca/IndicesW.nsf/Inicio?OpenForm&m=2>

²¹ www.cervantesvirtual.com

²² <http://www.proyectosarmiento.com.ar/>

²³ 仍未包含在 VIAF 内。

存在于将书目记录关联到规范记录的数字化聚合。这样，著者姓名及姓名的变异形式就可以被检索到，利用关联的书目记录丰富并完善了规范记录，多数情况下这项工作都是由国家书目机构来完成的。

VIAF 目前可以连接近 20 个规范文档，通过这些规范文档，可以获取到大量图书馆目录。有一点必须指出，只有在某些情况下，100 字段才与书目记录的 6XX 字段相关联。如果这一功能性一直有效，著者的相关作品就可以获取²⁴，VIAF 就可以被用作关于一个著者或者同一著者的作品的图书馆资源中心。

书目记录的信息源

正如我们定义了著者的信息源，我们使用同样的机制确立了书目信息源，其中包括：文献遗产联合目录（*Catálogo Colectivo de Patrimonio Bibliográfico*）²⁵，欧洲研究图书馆联盟（CERL）²⁶的传统印刷图书数据库（*Heritage of the Printed Book Database, HPB*），拉丁美洲、西班牙和葡萄牙图书馆 15 世纪到 19 世纪印刷版图书联合目录（*Novum Regestrum*）²⁷ 以及 WorldCat。这里要特别提到梅南德斯·佩拉尤虚拟图书馆²⁸，该虚拟图书馆继续着书目编制工作，该工作是由梅南德斯·佩拉尤（Menéndez Pelayo）在他的作品《西班牙科学》（*La ciencia española*）²⁹中发起的，这部作品扩大了西班牙思想家的影响力，并回答了著名的 M. Masson³⁰问题：在西班牙人们得到了什么？在过去的两个世纪、四个世纪、十个世纪，西班牙为欧洲做了什么？

²⁴ 通过 VIAF 访问加拿大图书馆和档案馆 [<http://www.collectionscanada.gc.ca/>]特别有意思，虽然博学家虚拟图书馆的目的和目标暂时还不明确，没有一个著者形成项目的部分，但是如果方法得到扩展，就会成为可能。

²⁵ <http://www.mcu.es/bibliotecas/MC/CCPB/index.html>

²⁶ Heritage of the Printed Book Database (HPB) del CERL.
<http://www.cerl.org/web/en/resources/hpb/main>

²⁷ <http://goo.gl/fqtHA>. See also *Novum Regestrum: el Catálogo Colectivo del Patrimonio Bibliográfico Iberoamericano* de Xavier Agenjo Bullón, Francisca Hernández Carrascal, *Boletín de la ANABAD*, ISSN 0210-4164, Tomo 44, N° 4, 1994, 127-142
<http://dialnet.unirioja.es/servlet/articulo?codigo=50938>

²⁸ <http://www.larramendi.es/i18n/bvmpelayo/inicio.cmd>

²⁹ Menéndez Pelayo, Marcelino. *La ciencia española*. In: *Biblioteca Virtual Menéndez Pelayo*.
<http://goo.gl/iB5m>

³⁰ Nicolas Masson de Morvilliers (1740-1789)

此外，我们必须提一下两个正在同时进行的大型项目，Hispana³¹ 和 Europeana³²，它们在不同的维度上影响着博学家虚拟图书馆，不仅影响它的内容，而且还影响它的功能需求。尽管有许多关于 Hispana 的描述³³，却很少有资料详述其的创建和进展，但这些有限的资料却出乎意料地介绍了 Hispana 的特征和范围。截止写作本文[2011年4月30日]，Hispana 已经从146个西班牙语资料库中收集了3,181,786种数字资源。关于 Europeana 的文件相当多，而鉴于我们所讨论的问题，这其中值得关注的是2011年3月7日更新的 Europeana 数据模型元素定义 5.2.1 (Definition of the Europeana Data Model Elements)³⁴，以及2010年8月31日发布的 Europeana 多瑙河发布版的功能说明(Functional specification for Europeana the Danube release)³⁵。

博学家虚拟图书馆和伊格纳西奥拉腊门迪基金会的发展与 Europeana 的功能紧密相关，这不是偶然的。原因在于：绝大多数的西班牙人都是通过访问 Hispana 网站（拥有1,367,808的数字对象，远多于拥有19,062数字对象的塞万提斯虚拟图书馆³⁶，尽管后者比前者更为公众所熟知）链接到 Europeana。Hispana 和大多数 Hispana 的内容提供者³⁷都是在 ILS DIGIBIB 目前 6.0 版本环境下运行的，不久会升级到 7.0 版本。

³¹ <http://hispana.mcu.es>

³² <http://europeana.eu>

³³ *Hispana y las iniciativas del Ministerio de Cultura* María Antonio Carrato Mena *Jornada de Difusión de EuropeanaLocal*, 17 nov. 2010 [<http://hdl.handle.net/10421/4765>]. *La aplicación del Modelo de Datos de Europeana a la Biblioteca Virtual de Patrimonio Bibliográfico : bvpb.mcu.es*, María Luisa Martínez-Conde *Jornada de Difusión de EuropeanaLocal*. 17 nov 2010 URI: <http://hdl.handle.net/10421/4783>. More recent: Carrato, María Antonia. *Hispana. I Seminario Internacional de la Biblioteca de Galicia* [<http://goo.gl/6xIqq>]

³⁴ <http://goo.gl/ojIL>

³⁵ <http://goo.gl/P6jme>

³⁶ <http://www.cervantesvirtual.com/>

³⁷ 历史报刊虚拟图书馆，加利西亚语：加利西亚虚拟图书馆，安达卢西亚虚拟图书馆，格雷多斯（萨拉曼卡大学，西班牙），马德里数字图书馆，卡斯蒂利亚-拉曼恰数字图书馆，曼弗雷基金会文档中心，卡斯蒂利亚-莱昂数字图书馆，文献遗产虚拟图书馆，阿拉贡内斯法律虚拟图书馆，皇家历史学院虚拟图书馆，国家档案局图书馆联合目录，阿斯图里亚斯虚拟图书馆，卡斯蒂利亚-拉曼恰图片档案，瓦伦西亚数字图书馆，阿拉贡虚拟图书馆，智者桑丘基金会，穆尔西亚图书馆，阿兰胡埃斯数字图书馆，拉古纳大学，萨拉戈萨虚拟图书馆，拉里奥哈虚拟图书馆，伊格纳西奥拉腊门迪基金会，皇家药学院虚拟图书馆。

有一点不能忽视，发行了 ILS DIGIBIB, DIGIARCH 以及基于 OAI 的资料库 OAsIs 的是一家隶属于伊格纳西奥拉腊门迪基金会的 DIGIBIS³⁸公司，而博学家虚拟图书馆正是其开发项目的检验平台。因此（如果这正是客户所希望的），这里所描述的许多特征都显著地影响着大约 30 个主要的西班牙数字图书馆，如果将文献遗产虚拟图书馆（Biblioteca Virtual del Patrimonio Bibliográfico）与历史报刊虚拟图书馆（Biblioteca Virtual de Prensa Histórica）³⁹的数字馆藏计算在内，那么就会影响影响 100 多家图书馆。

计划在 2011 年 5 月 1 日发布的 Europeana 多瑙河阶段版在 IFLA 波多黎各会议期间应该已经开始实施。虽然计划并不能保证如期实施，然而其延后一定与博学家虚拟图书馆所计划的事情无关，因为 Europeana 新的语义特征已经对开发数据新的语义结构起到了决定性作用。因为拥有这些特征，Europeana 和 Hispana 成为博学家虚拟图书馆的两个主要信息源，以及两个方法性的实例，就像 Hispana 的情况一样，可以反作用于虚拟图书馆。

目前，Europeana 已经发布了基于 OpenSearch 的 API⁴⁰，OpenSearch 可以嵌入到网站搜索中并显示提取自 Europeana 的记录。发挥了其重要作用并且已经嵌入到博学家虚拟图书馆的网络界面中的 API，自动从 Europeana 数据库中的初始查询推出搜索。也就是说，在一个有几千条数据的数据库中进行的查询同样也运行于拥有大约两千万数字对象（17,901,019）的 Europeana 数据库中。

2. 数字聚合体

根据博学家虚拟图书馆既定的研究方法，信息系统的核心内容是“数字聚合体”，其风格和设计与传统的百科全书条目很接近，并且力图达到维基百科（或者如下文中讨论的 DBpedia）中一些条目的复杂性和完整性程度。

术语“聚合体”和“聚合”在不同的使用环境中有不同的用法，可能会让读者产生歧义。正如 Europeana，在收割元数据的过程中，亦可被收割的收割器就可称之为聚合器。然而，博学家虚拟图书馆所使用的数字聚合体的概念与开放文档先导计划之对象重用和交换（OAI-ORE）⁴¹所定义的十分一致，其将聚合体定义为包括一组相互关联的资源。一个数字聚合体就是围绕一个特定作者收集的一组数据和数字信息资源，其核心是一条规范记录。

³⁸ <http://www.digibis.com>

³⁹ 文献遗产虚拟图书馆的子域 *Iberoamérica en las colecciones de la BVPB* 能更好的说明这一点。
[<http://goo.gl/8Jyfa>]历史报刊虚拟图书馆提供的全文搜索也非常有用。
[<http://prensahistorica.mcu.es>]

⁴⁰ <http://www.version1.europeana.eu/web/api>

⁴¹ <http://www.openarchives.org/ore/>

伊格纳西奥拉腊门迪基金会执行的规范控制比大型图书馆或书目机构的更广泛。我们以 FRAD⁴²定义的方式坚持语境化任务而不是实体个人的标识功能，这样，不仅规范文档中不平常的数据被记录了下来，而且记录与其他规范数据源以及其他在网络上可获得的资源和数据也建立了连接。从这点上说，参考数据来源（著录在 670 字段）不仅仅是姓名及其变异形式的选择判断，还是关于著者的书目信息的跟踪来源。

正如 VIAF，每条规范记录都以 MARC21 格式包括其标目（1XX），单纯参照根查（4XX），相关参照根查（5XX）以及标目连接款目（7XX）。在博学家虚拟图书馆中标目连接款目是很重要的，因为所有著者的选取并不单纯因为其在伊比利亚美洲文化中的重要性，更是因为伊比利亚美洲文化对全球文化的积极影响，这也正是必须有其他语言等同信息的原因。这样就有可能实现世界范围的文本转换，以及思想的传输和影响。西班牙也存在这样的问题，在西班牙，除卡斯蒂利亚语和西班牙语以外，还有加泰罗尼亚语、瓦伦西亚语、加利西亚语和巴斯克语。

多语种标目

在 MARC21 格式中很有必要增加一个定义标目语言的特殊元素。很遗憾，论文（2001-DP05）MARC21 规范格式的多语种规范记录⁴³目前为止并没有达到解决这一问题的目的。7XX 字段可以用不同语言著录等同标目，但却无法标明标目语言，而这在多语种的规范文档中极其必要。其他机制也同样存在局限性：008/08 位的语言代码仅能表示英语和/或法语为编目语言，040 字段的 \$b 也仅仅著录编目语言，而编目语言往往与标目语言是不同的。

MADS 和被提议的 MADS/RDF⁴⁴格式虽然合理地克服了这种局限性，但大多数的规范记录都是 MARC 格式的，因此很难将这一类型的数据立即转化到关联开放数据环境中。VIAF 也存在这样的局限性，尽管可以确定记录的机构和来源国，但标目语言却不得而知，除非通过推断。当标目产生于相同语言区，极其相似很难区分时（例如，西班牙语的 *Geografía* 和意大利语以及加泰罗尼亚语的 *Geografia*[无重音]），标目语言的确定显得极为重要。

个人名称的变异形式

在其他规范文档中检索信息的目的不仅在于建立标目间的语言等同关系，还在于丰富名称变异形式的数量以及与其他名称的关系。ILS DIGIBIB 特别适合采取这一做法，允许导入与已建标目适当关联的书目记录，并且控制 4XX 字段中非规范款目的产生。在建立规范记录的初期，这一过程有些慢，但在中期阶段，

⁴² http://www.ifla.org/files/cataloguing/frad/frad_2009-es.pdf

⁴³ <http://www.loc.gov/marc/marbi/2001/2001-dp05.html>

⁴⁴ MADS/RDF 入门。最终公开评测文件。[<http://www.loc.gov/standards/mads/rdf/>]

其愈发显现成效和质量。需要注意的是，如果资源从规范记录的核心开始分析，那么起初的书目项目会更好的完成。

这一过程尤为重要，尤其是当我们将之视为从 Seneca（科尔多瓦, 4 BC-罗马, 65 AD）到 Martín de Riquer（巴塞罗那, 1914 -）的一个著者群，而这其中许多著者在创作不同语言的作品时使用了不同的名称，更不用说他们的名称也会随着著作一起被翻译成其他语言（例如，Averroes 是 Ibn Rus 的拉丁语名称）。要想能够导入具有足够变异名称的资源，就必须全部获取几乎未根据规范记录或文档作出改变的资源。

3. 规范记录

2010 年 10 月出版的 MARC21 第 12 期更新遵循 RDA 的理念，增强了许多对于个人信息聚合十分必要的字段的使用，即 3XX 字段。从历史⁴⁵的角度看，这些字段起初目的在于区分同形同音异义词，而渐渐发展为不可或缺的一部分。因此，100 字段的不同子字段原本只是选择性使用以便区分同名的著者，最后却演变成为必备的，最明显的就是子字段\$d，它从年代时间的角度迅速将一个著者归入特定语境，多数时候能便于参考作品避免归因错误。MARC21⁴⁶为适应 RDA 作出调整之前，并没有像出生地这样起作用的单个属性的具体条目。正是这些属性，指的不是个人名称，而是个人本身，并且有助于以必要的百科全书式的视角看待 MARC 记录，或许我们可以用一个新的词语“dbpedic”来表示。

RDA 的使用强化了博学家虚拟图书馆的作品体系。首先，在著者信息检索中，RDA 同样重视由 FRAD 定义为阐明关系和搜索的用户任务。其次，以 RDA 的作品-内容表达-载体表现-单册⁴⁷的模型构建信息源。为此，DIGIBIB 随后的新版本立即适应了 MARC21 的更新升级，使得 ILS DIGIBIB 也适应了更新了的 MARC21。

除名称标准形式外，著者信息还应尽可能包括名称变异形式和其他语言的等同形式，著者的传记资料（678），生卒年（046），生卒地（370）和其他活动地（370），职业（372 和 374），使用语言（377）或性别（375）。这些个人属性最终目的在于支持数据间的语义搜索和导航。它们使博学家虚拟图书馆网站可

⁴⁵ Agenjo Bullón, Xavier y Palá Gasós, Pilar. El fichero de autoridades del Catálogo Colectivo del Patrimonio Bibliográfico. En: Boletín de la ANABAD, ISSN 0210-4164, Tomo 37, N° 4, 1987, págs. 593-606. [<http://dialnet.unirioja.es/servlet/articulo?codigo=801041>]

⁴⁶ <http://www.loc.gov/marc/RDAinMARC29.html>

⁴⁷ MARC Discussion Paper No. 2011-DP03: Identifying Work, Expression, and Manifestation records in the MARC 21 Bibliographic, Authority, and Holdings Formats
[<http://www.loc.gov/marc/marbi/2011/2011-dp03.html>]

以提供以活动时间、生平地、专业和职业为参照的著者检索，而不仅仅是在 Web OPACs 上已经相对普遍的通过作品进行的检索。

其他元数据方案中的规范记录

博学家虚拟图书馆同样关注档案中团体、个人及家族的信息。档案馆、图书馆和博物馆的数据整合问题很受重视，并且已经有成功案例。它们通常以统一查询界面出现，这就加强了此类数据的信息容量。近期，DIGIBIS 开展了一个新项目卡斯蒂利亚拉曼恰政府（Junta de Castilla-La Mancha），致力于 Archivo de la Imagen⁴⁸（图片）和 Archivo de la palabra（录音制品）⁴⁹。这样，MARC 21 就系统地映射⁵⁰到 AC-CPF、ISAD(G)和 ISAAR(CPF)了。

同样，博学家虚拟图书馆已经实现了将书目资料和档案材料相关联。例如，关于 Agustín de Montiano y Luyando (1697-1794)的资料，他是历史学家，评论家及剧作家，因曾担任建于 1735 年的皇家历史学院的第一任院长而为人所熟知，拉腊门迪虚拟图书馆⁵¹收录了他的论文。

4. 关联数据

博学家虚拟图书馆是 Europeana 的内容提供者，由 Europeana 体系中的国家级聚集器 Hispana 进行元数据收割，进而使用 Hispana 的数据库作为其信息源。因此，它已经建立了采用 Europeana 数据模型（EDM）⁵²和关联开放数据技术的战略性目标。EDM 和关联开放数据（LOD）正迅速引领新变革，并且提供了大量的数据，VIAF⁵³或 DBpedia⁵⁴就是很好的例证。从 Tim Berners-Lee (蒂姆·伯

⁴⁸ http://clip.jccm.es/archivo_de_la_imagen/es/micrositios/inicio.cmd

⁴⁹ http://clip.jccm.es/archivo_de_la_palabra/es/micrositios/inicio.cmd

⁵⁰ http://www.sedic.es/p_boletin_novedades_mensual.asp#Digibis

⁵¹ 事实上，它是由 Ignacio de Larramendi y Montiano（1921-2001）建立的拉腊门迪虚拟图书馆或 DIGIBIS 的前身。参见：Hernando de Larramendi, Ignacio. *Mecenazgo cultural de Ignacio Hernando de Larramendi y Montiano : crónica y testimonios*. [<http://www.larramendi.es/i18n/consulta/registro.cmd?id=954>]

⁵² Doerr, Martin, Gradmann, Stefan, Hennicke, Steffen, ISAAC, Antoine, Meghini, Carlo, and SOMPLE, Herbert van de. *The Europeana Data Model (EDM)*. En: 76th IFLA General Conference and Assembly, 10-15 August, 2010, Gothenburg, Sweden. [<http://goo.gl/7eLs>]

⁵³ <http://ckan.net/package/viaf>

⁵⁴ <http://dbpedia.org/About>

纳斯-李) 2006 年首次提出关联数据概念至今, 除图书馆 MARC 书目记录外, 语义结构化数据和可重复使用的数据以之前在网络上从未出现的速度公布。

值得关注的是, 在 W3C 内部, 图书馆关联数据孵化小组 (LLD)⁵⁵ 已经于 2010 年成立了。这个小组负责准备一些关于档案馆、图书馆和博物馆本体状态的报告和交付成果; 相关的词表和数据集以及图书馆界的相关用例。孵化小组和关联开放数据项目⁵⁶的动态都会适宜地向感兴趣的个人公布。

对于博学家虚拟图书馆来说, 关联开放数据⁵⁷以及附属技术, 提供了一个充分认识“数字化聚合”作为关联数据的节点的方法。事实上, 博学家虚拟图书馆就是一个由 LLD 评定的用例。它不仅捕捉信息, 也以各尽其能的原则将信息连接到 LOD 合适的节点及其来源上。LOD 数据集的资源都由 URI 标识出特殊特征⁵⁸。URI 根据采用的模式 (本体或 RDF 模式) 以及指定功能块的实体和属性架构了界定其他资源属性的关系网络。重要的是, URI 允许标识一种环境下的实体与另一种环境下定义的实体相同或十分相似。例如, Maimonides 在 DBpedia 中代指一个哲学家, 而在 VIAF 中代指个人名称。这一过程包括使用 URIs 连接此类数据, 开放这些数据以方便重用, 利用合适的工具⁵⁹分析数据并根据关系和推论规则推导出其他数据, 更新原始数据以便在修改和扩展之后再次提供服务, 如此反复。

DBpedia 和 VIAF

2010 年博学家虚拟图书馆开始将规范记录与 LOD 中的同类既有资源关联起来。显然, 关于刚刚所描述的方法的信息源是用于获得连接的候选资源。因此, VIAF 资源的 URIs 著录在 MARC21 记录的 024 字段。RDF 的资源 Maimonides (rdf: about = "viaf/100185495"), 集合了来自于 7 个不同的规范文档包含 255 个变异名称的关于此著者的参照标目, 并且还和 DBpedia 的内容 (<http://dbpedia.org/resource/Maimonides>) 连接。经验丰富并且具有专业技术知识的图书馆员, 例如一个高级编目员, 能根据 VIAF 关系网推断出多达 12 个来自于不同国家和语言的关于 Maimonides⁶⁰的标目。显然, 人工获取数据很精准但速度却很慢, 因此在权衡其优势和劣势之后, 数据搜索的自动化很快就显现出其适宜性。构成博学家虚拟图书馆的大约 1000 个著者无疑需要人工操作, 而体现在

⁵⁵ <http://www.w3.org/2005/Incubator/llid/>

⁵⁶ <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

⁵⁷ http://www.w3.org/2005/Incubator/llid/wiki/Use_Case_Polymath_Virtual_Library

⁵⁸ <http://www.w3.org/TR/2007/WD-cooluris-20071217/>

⁵⁹ <http://www.w3.org/wiki/SemanticWebTools>

⁶⁰ Ibn Maymun, Musà or, the much more familiar to the orthodox Jews, Rambam

时间和数据上的效率都会因为采取自动化手段而大大提高。此外，一旦工具被设计并开发，它们将成为 DIGIBIB 的特色。据估计，在此论文发表前后，即 2011 年 8 月 13-18 左右，这些工具将构成 DIGIBIB 7.0 及以上版本的一个模块。

SKOS: 实现关联开放数据的另一种方式

通过采用 SKOS（简单知识组织系统）编码，这些关联方式都得以实现。结果显示，某个概念与其他词表中的其他概念相同或者十分相似。通过书目记录的主题标目，并将概念与其他可在 SKOS 中获得的词表关联，同样可以搜索到关于某个著者的大量作品信息。此外，通过掌握不同词表中概念的关系网及拥有检索与这些概念关联的信息源的能力，可实现自动查找信息源。

目前，博学家虚拟图书馆正尝试研究由图书、档案及图书馆司（Dirección General del Libro y Bibliotecas）开发的西班牙公共图书馆标题表，此标题表通过与西班牙国家图书馆⁶¹的规范文档的融合，以及与美国国会图书馆标题表⁶²的对应而完善。这一过程可以类比为将 RAMEAU⁶³的 SKOS 概念与 LCSH 和德国主题规范文档（SWD）⁶⁴相连接，因此建立不具备高粒度的概念关系是可行的，而这种概念关系融合了可以提供大数据集之间清晰的连接网络的固定数量的著者。

时间和空间

年代数据和地理数据对于某个著者的数据聚合和关联是至关重要的。著者都有具体的生卒年月，与某一时期的同时代人具有社会关系，其作品都有具体的出版、翻译和编辑日期⁶⁵。用这种新的年代分类方式关联著者数据，而不仅仅通过名称和主题，更精确了著者的作品集，尽管他们的作品信息相当分散并且不一致。规范记录和书目记录许多元素中的时间编码通常需要一些属性来表示经常用于人，资源和事件的描述的可能性、不准确性、估计性等等。此类数据对于建立时间线，根据日期执行运算和推导至关重要，WorldCat Identities⁶⁶恰恰说明了这一点。

⁶¹ *Autoridades de la Biblioteca Nacional* [Recurso electrónico]/ Biblioteca Nacional ; software, Chadwyck Healey -- Número 1 (nov. 1996). - Madrid : Biblioteca Nacional : Chadwyck-Healey España, 1996- CD-ROM.

⁶² <http://id.loc.gov>

⁶³ <http://www.cs.vu.nl/STITCH/rameau/>

⁶⁴ <http://www.d-nb.de/standardisierung/normdateien/swd.htm>

⁶⁵ 现在已经有新的格式将不同的年表甚至表达年代的数字的记录方式进行规范化，如果不是按照字母表、拉丁字母表、阿拉伯字母表等。扩大版时间/日期格式 ETDF（Extended Time/Date Format）[<http://www.loc.gov/standards/datetime/>]

⁶⁶ <http://www.worldcat.org/identities/>

关于地理数据的聚合和关联正在进行着同样的工作。事实上，在关联开放数据中存在一个名为 GeoNames⁶⁷的主要方案，它也是迄今已知的唯一通过转换国家地理研究所⁶⁸的地理编码复制成西班牙语，并发布成 RDF 并成为 LOD 的来源之一。这样一来，数据聚合过程就再一次被扩展了。著者出生并逝世于某一地点，与某一地点的同时代人存在社会关系，这些地点可能是学院、大学、宗教机构、设在或关于某一地点的军事、科学和文化团体或政治团体。因此，利用这些地理信息就有可能建立新的恰当的数据聚合和关联。

通过将所选数据转换成 DIGIBIB 可以容纳的有效结构，就可以自动连接到 GeoLinkedData 和 GeoNames 的数据。这些数据已经映射到 MARC 格式，以便自动更新 752 和 151 的所有子字段。这样一来，我们不但能从这些词表中获得名称信息，而且还能找到地理坐标信息。有可能还需要进一步区分管辖权、地理位置和旧的管辖权对应的地理位置的问题。下文还会提到管理个人和其他个人名称的问题。显然，管理管辖权，地理名称和地理位置应该具有不同的数据结构。

5. 使用 LOD

VIAF 网站可以提供查询服务，查询结果以 MARC XML 或 RDF 格式进行显示。借助其 API⁶⁹，VIAF 可以通过 SRU 和 OpenSearch 方式提供查询，查询结果能以 MARC XML、RDF 及其他多种格式和模式进行下载。VIAF⁷⁰也是关联开放数据的信息源之一。在响应号召将数据发布到 LOD⁷¹时，通常数据集是利用可以查询 RDF 文档的 SPARQL 端点⁷² 或 SPARQL 客户端⁷³进行存取和查询的。也就

⁶⁷ <http://www.geonames.org/ontology/documentation.html>

⁶⁸ GeoLinkedData 已经发布来自于国家地理研究所的多种信息源，根据关联数据原则，使其能够以 RDF 格式被获取

[<http://geo.linkeddata.es/web/guest;jsessionid=6E6A9C1E1DAF40F81005E8F4DA3A27E1>]

⁶⁹ <http://www.oclc.org/developer/documentation/virtual-international-authority-file-viaf/using-api>

⁷⁰ <http://ckan.net/package/viaf>

⁷¹ *How to Publish Linked Data on the Web.*

<http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

Best Practice Recipes for Publishing RDF Vocabularies. W3C Working Group Note 28 August 2008.

<http://www.w3.org/TR/swbp-vocab-pub/>

⁷² <http://www.w3.org/wiki/SparqlEndpoints>

⁷³ <http://www.w3.org/wiki/SparqlImplementations>

是说, DBpedia⁷⁴或 VIAF 可以利用 SPARQL⁷⁵语言提供查询服务, 查询可以促使对数据进行逻辑上的普遍对比搜索, 并检索出比使用上文介绍的方式更相关的信息。利用 SPARQL 可以定义更复杂的查询, 例如在 DBpedia 而不是 VIAF 中查询“philosophers influenced by Maimonides”。因此, 博学家虚拟图书馆根据书目计划再次采取措施, 利用 LOD 数据集⁷⁶形式的特定书目信息。并且这样做是从用户和生产者的角度出发的。

所以, 第一步是为适应 Europeana 因多瑙河阶段计划作出的改变而启用 Europeana 数据模型 (EDM)。Europeana 或多或少地自动处理数据, 以将从内容提供者处接收来的信息转化为新的 EDM 模式, 这是可能的。但考虑到这些提供者的数量已经达到几千, 并且使用超过 25 种语言和许多词表, 因此我们相信, 采用自动化手段只能处理不确定的数据量, 结果无法核对, 并且还有可能产生错误的等同和关联。不管这一过程是否切实可行, 是否有能力解决数据不同部分间的连接, 它都不能被普及, 更不能在没有内容提供者本身干预的情况下被完善。一个典型的例子, 地理名称通常是与当前的地理坐标不一致的行政管辖区名称, 因此 15 世纪的卡斯提尔与现在的卡斯提尔只能是近似的地方。至少, 博学家虚拟图书馆已经在国内开始发布连接到 LOD 的数据。

将 MARC 21 格式的书目记录与 Europeana 的语义元素 3.3.1 和 Europeana 数据模型相互映射是比较困难的。但是, 如果仅仅实现了这一过程, 而并没有将数据与外部资源或词表连接, 那么 Europeana 将很难实现准确的数据聚合 (从 OAI-ORE 意义上说)。将 VIAF 的 URIs 合并到规范记录的方法被证明是非常有用的, 因为 EDM 数据的产生可以提供给 Agent 外部词表一个参考。第二个关联过程是通过 *id.loc.gov* 上可获得的 LCSH 实现的。因此, 博学家虚拟图书馆能使用 SKOS 发布主题标目, 然后连接到 LCSH。这样, 属性 *dc: subject* 与 *skos: Concept* 相联系, 并至少与 LCSH、间接与 RAMEAU⁷⁷和 SWD⁷⁸相连接。换言之, 主题标目具有语义结构, 并且与用英语、法语和德语表述的相同标目相联系。这就显示了将不同数据集和一个或多个具体词表连接起来的益处。

⁷⁴ <http://ckan.net/package/dbpedia>

⁷⁵ SPARQL 是一种 RDF 数据查询语言 [<http://www.w3.org/TR/rdf-sparql-query/>]。目前 W3C 正在研究 SPARQL 1.1 联合查询。 *This specification defines the syntax and semantics of a SPARQL 1.1 Federated Query extension for executing queries distributed over different SPARQL endpoints.* [<http://www.w3.org/2009/sparql/docs/fed/service>].

⁷⁶ <http://ckan.net/package>. 见 LLD 之编汇 <http://ckan.net/group/lld>、关联数据 DOIs <http://inkdroid.org/journal/2011/04/25/does-as-linked-data/>, 或关联期刊数据 <http://periodicals.dataincubator.org/html>

⁷⁷ <http://ckan.net/package/stitch-rameau>

⁷⁸ <http://ckan.net/package/dnb-gemeinsame-normdatei>

除此之外，同样的过程也适用于互连国家级数据。使用 LCSH 作为西班牙图书馆主题标目的主要来源之一的结果是，利用这一关系（记录在 670 字段）可以关联不同的西班牙主题标目列表，并有可能是不同语言的。LOD 还是间接产生了一个相当有益的附加作用，它使仍保持相对独立的西班牙图书馆目录和规范文档得以关联，效果高于期望并且更全面。

这里必须指出 Europeana 在完善国家信息系统方面的另一个优势。负责维护 Hispana 的西班牙文化部也已经开始为 EDM 作出调整，正如上文提到的，首先采取的行动之一就是将其公共图书馆主题词表（LEM）转换成 SKOS 格式，可能的话将其款目与 LCSH 相关联。从逻辑上说，博学家虚拟图书馆已经完成了将其主题标目与 LEM 连接的过程。希望 LEM 数据集，以及其他加利西亚语、加泰罗尼亚语、瓦伦西亚语及巴斯克语的西班牙主题标目能够关联起来。2011 年 4 月 30 日，西班牙文化部（Spanish Ministry of Cultura）下属的文献遗产虚拟图书馆启用了 ILS DIGIBIB 7.0，这一版本能够根据 EDM 5.2.1 导出记录。

无论是 EDM 的全面实施还是博学家虚拟图书馆的数据集发布成 LOD，这些过程都有共通点。在这点上说，数据发布为 LOD 应该符合书目计划的进程，因此比较慢，只有建好的、拥有合理的责任度、并且充分关联⁷⁹的数据才能被发布。正如博学家虚拟图书馆的书目项目对资源进行严格挑选，图书馆项目将其资源发布为 LOD 时同样对数据集进行严格挑选。IFLA 会议是一个最好的平台，让大家记住元数据方案，本体和本体调整不能使我们忽视整个图书馆系统，和即将发布在 LOD 中的数据集的书目客体。

6. 自动化过程

作为信息用户，博学家虚拟图书馆正在研发各种用于数据采集，整合及传播的应用程序。本文已介绍了利用 GeoNames 或 LEM 实现数据重用。此外，在写此文之际，新的应用程序也正在设计用于通过半自动化过程采集 LOD 数据集。这一过程可能包括采集与一个或多个著者相关的部分或全部三元组，或者只获取资源的 URI。数据被储存在可以实现搜索，选取和升级操作的中间仓储里。一旦得以核实，数据就被书目数据库或规范文档所容纳，就有可能选择一个资源的所有属性或仅选择其 URI，并更新原记录。

同样的过程将利用书目资源和信息资源的描述得以实现。这一过程将很快启动，因为存在正确标识作品的更多问题。如果规范文档中个人名称的标识都延续了诸如缺乏足够的背景信息等诸多问题，则作品缺少数据的情况更多。因此缺少统一题名和著者-统一题名款目或作品及其译版的题名关系，会产生无效的自动或半自动过程。

⁷⁹ Weibel, Stuart. *Principles of Linked Data Recast*. Weibel Lines [http://goo.gl/sYSDe]

此外，不是所有规范数据或书目数据的资源都可通过 LOD 获得，因此博学家虚拟图书馆使用可以用来启动特殊搜索档案夹的 SRU 服务器。这些档案夹可以根据搜索到的元素来调整查询。这一设计的优势在于检索结果可以被存储到一个可以从 SRU 服务器执行常规元数据收割的 OAI 仓储。这个系统很有可能建立一个国家级甚至国际级的收割器。Europeana 至今取得巨大的进展是基于 OAI 元数据收割和一个可被称作 *Americanae* 的相似系统。这一过程的主要优势是选择性的元数据收割，以从西班牙和国际资源中获得好的信息资源。

7. 规范的百科

此模式从个人名称的规范记录转为个人的记录，不再是名称问题，范围更广。事实上，近期有待解决的一些编目问题是如何将个人属性和个人名称属性联合起来。而关于 VIAF 或 DBpedia 的一些本体和数据集的研究阐述了建立个人名称的类和属性的不同定义。

如果关联开放数据云图的中心是 DBpedia，即维基百科，那么百科式的规范数据就具有与维基百科相似的结构，而只是使用 MARC 作为基础格式，这一说法似乎就理所当然了。即使 MARC 21 的多用性和良好粒度可轻易被转换成 XML 结构，忽略世界上众多（数以十亿计）MARC 记录是非常荒谬的。当然，任何系统都可以从 LOD 任何数据集中获得数据，但这不仅仅是关联数据的问题，还包括数据的维护和升级。这样说来，DBpedia、VIAF、MADS、MARC 21 和 EAC-CPF 的结构可用于融合信息，但这却是不够的，因为仅仅关联数据只会导致无法控制的错误、不准确性或冗余。也就是说，我们需要一个新的关于个人和个人名称的数据结构，可以支持数据管理和维护。

规范记录的结构可能是基于 MARC 格式的方案集合，并参考 MADS / RDF、EAC-CPF⁸⁰，包括 DBpedia⁸¹的一些属性。DBpedia 产生的影响很有可能对于增加定义个人的属性的数量很有必要。关于 DBpedia 本体的研究显示，表示人的类别的一些属性切实地关系到博学家虚拟图书馆的实际建设，这其中包括“影响”、“被影响”或“哲学学派”。

博学家虚拟图书馆的子项目（或称之为子集更恰当）萨拉曼卡学派（Escuela de Salamanca）能更好地体现这一点。萨拉曼卡学派是一个在 16 世纪和 17 世纪初文艺复兴时期，发现新大陆之后，由西班牙和葡萄牙学者（主要为神学家、法学家和经济学家）所发展集合而成的学术流派，以来自萨拉曼卡大学的神学家弗朗西斯科·维多利亚（Francisco de Vitoria）的著作为主要根基。此项目最终目

⁸⁰ 尽管此论文通篇未提到，这里仍需引用近期出版的 CIDOC-ICOM Linked Open Data Recommendation for Museums. [http://www.cidoc-crm.org/URIs_and_Linked_Open_Data.html]

⁸¹ <http://mappings.dbpedia.org/server/ontology/classes/Person>

标在于延续这些在法律、政治、尤其经济领域的先驱者的影响力⁸²。Modern Age 是社会人这个概念的重大变革，而正是萨拉曼卡学派使用新的方法解决了这些问题。弗朗西斯科·维多利亚（Francisco de Vitoria），多明戈·索托（Domingo de Soto），马丁·阿斯皮利奎塔（Martín de Azpilcueta），托马斯·德·梅尔卡多（Tomás de Mercado）和弗朗西斯科·苏亚雷斯（Francisco Suárez）试图将托马斯·阿奎纳的学说（Thomistic doctrine）与新出现的社会秩序和经济秩序相融合。这样，通过这一项目，我们发现了为经济科学⁸³奠基，关注征服的道德合法性问题，并发展创新理论⁸⁴的一群著者。

总之，我们发现博学家虚拟图书馆包含的著者以及分布于许多图书馆的成千作品的著者采用通过 VIAF 将作品与其他著者相连接的过程，而不久这一过程就被扩展，进而包括利用词表或标题表，如 LEM、LCSH、RAMEU 或 SWD，建立图书主题间的虚拟导航，这些著者在空间和时间上相互连接并聚合起来。利用这些数据建立了一个广泛的连接关系，尤其是通过更多指代个人本身而不是个人名称的元数据，团体无疑可被视作为个人。我们应该增加作品特征，因为在绝大多数情况下，得益于印刷商、编辑者和销售商（以及收藏或将文献数字化的图书馆），作品的内容表达才能获得清楚的载体表现。而这些实体同样具有关于空间和时间的说明。这样，根据萨拉曼卡学派的观点，我们就能了解圣埃斯特万修道院（40° 57' 38"N - 5° 39' 47" O）对教师和学生来说起到多么关键的作用，甚至做了一项编辑和印刷的工作，延续至今。图书馆学校，一所看不见的大学，那个时代的术语变得清晰可见。

重新定义 ILS 功能

提供关联数据的可能性将引起对 ILS 功能，即 DIGIBIB 功能的完整的重新界定。首先，如上文所述，允许 MARC 格式原本不包含的数据元素或属性编入其中，以方便更新，管理并导出为 LOD。其次，允许容纳全部或选定的数据集；即具有更新记录的全部或部分的功能。例如，在博学家虚拟图书馆中，有可能实现仅选择或容纳资源的 URI。第三，修改查询界面是非常必要的，使通过 SRU / OpenSearch 或结合 SPARQL、利用 APIs 从其他资源处进行查找和在线糅合成为可能。最后，尽管不是非常详尽，必定可以将没有广泛用作 LOD 数据集的描述的信息种类视为一个整体而管理，这其中必须提到档案检索工具和博物馆记录。

⁸² Agenjo Bullón, Xavier; Hernández Carrascal, Francisca y Juez García, Patricia: *La Escuela de Salamanca considerada desde el punto de vista de la Web semántica y la información en la red*. Paper given on 2011, April, 14th in *X Jornadas de la Asociación de Hispanismo Filosófico: Crisis de la modernidad y filosofías ibéricas*, held in Santiago de Compostela, 2011, April, 13-15. [<http://goo.gl/uFrXx>]

⁸³ Schumpeter, Joseph A. *History of Economic Analysis*. New York: Oxford University Press, 1954.

⁸⁴ Grice-Hutchinson, Marjorie. *The School of Salamanca: Readings in Spanish Monetary Theory, 1544-1605*. Oxford: Clarendon Press, 1952.

8. 数据集的获取

应该注意的是，除数据集的获取问题之外，对这些数据集的严格挑选也是很重要的，而最终形成的是具有新的结构的书目信息资源。其根本问题是从智识点和理论角度对数据质量进行严格的分析。除需要根据 CKAN⁸⁵或 VoID⁸⁶注册数据集外，显然还需要不仅关于本体或数据结构，而且关于款目数量和关系、更新频率、更新的管理方式、以及用于任何数据融合的资源的信息。可以说，图书馆员和目录工作者都在从外部资源关联数据的过程中直接体会到，对于同一件事物的所有判断有时不正确，有时错误，有时冗余，甚至有些时候不信赖数据来源。

语义网和 LOD 计划提高导航的准确度，但可见，尽管 LOD 只拥有极少的可用资源，还是很容易查找到以不同属性不同方式定义的相同事物，而且关联本身并不足以提高导航的准确度。必须指出，目前需要做的就是 LOD 数据集中转换大量高品质的未被公众使用的纸本书目资源和非结构化书目资源。尽管 Europeana 和 Hispana 实施的众多数字化项目都取得了良好的进展，但这仍然是图书馆和图书馆员不可推卸的责任。正如 19 世纪德国的唯心主义理论所阐述的物自体理论，物体本身所固有的属性，它的存在不受主体认识形式的限制。

9. 结论

这里所描述的所有工作都是围绕这两个项目展开的，这两个项目在 2011 年 4 月 30 日结束最后一个评估时，仍然是悬而未决的。一方面，Europeana 多瑙河阶段才刚刚开始启用。另一方面，预计 RDA 在美国图书馆界的广泛的试行的成果或许在下一届 ALA 会议上将被推广。如果我们说，W3C 的图书馆关联数据孵化小组还未得出结论，那么显然我们所做的就只是与这个新范例同时进行的一系列工作。或许我们做出的努力和贡献具有最大的利益，因为博学家虚拟图书馆已经采取类似于 RDA 实施的措施，积极参与 Europeana 的建设，并且是 LLD 的一个用例。我们的实际工作在网络上开放给那些感兴趣的人，包括获取信息（不只书目信息），实现最大可获取性，最重要的是，实现蒂姆·伯纳斯-李阐述的关于关联开放数据的目标。这就是此文的最终意义。

⁸⁵ Comprehensive Knowledge Archive Network. [<http://ckan.net/package/new>]

⁸⁶ Describing Linked Datasets with the VoID Vocabulary [<http://www.w3.org/TR/void/>]