



作为关联开放数据的德国国家书目：应用与机遇

Jürgen Kett

Sarah Beyer

Mathias Manecke

Yvonne Jahns &

Lars G. Svensson

Deutsche Nationalbibliothek
Frankfurt am Main, Germany

中文翻译：赵梦（中国国家图书馆）
Chinese Translator: ZHAO Meng
National Library of China

Session: **215 — *What is a national bibliography today and what are its potential uses?* — Bibliography**

摘要

本文探讨了 21 世纪对国家书目的要求。作者认为，除了传统标准数据所要求的完整性、可靠性，数据的现时性，可参照性和永久保存性外，国家书目还需要增加一项要求，即国家书目需要将其自身整合进万维网，因为在今天万维网才是信息交换的场所。要实现这一点，就需要应用关联数据技术在开放协议下并实现数据公开。作为一个案例，我们研究了德国国家图书馆的做法。其数据库中大约 70% 的内容都已经作为关联数据得以公开。

导言

国家书目可以被定义为某一地理限定区域出版物的完整列表。¹在万维网时代，我们应当进一步拓展这一定义，因为万维网既改变了“出版物”的意义，也改变了我们出版的方式。

传统纸质印刷书是独立的、静止的，而万维网出版物没有很好的范围界定，它是动态的、互动的，内容千变万化的。进一步说，用以记录文本的技术手段已经发生变化：随着全文检索成为可能，元数据目录，叙词表和分类体系的使用正逐渐被边缘化。这并不是因为全文检索优于基于元数据的检索，而是因为它更便于自动化操作：我们可以轻而易举地在网页上找到全文资

¹ 参看 Anderson (1974) p. 12

源，然而，高质量的元数据需要被生成。元数据的创建或编目，要么需要复杂的自动化程序，要么需要人工介入。

但是，现在大多数由人工完成的过程都是在图书馆之外完成的：出版商使用自己的元数据来提高其自身的可见性，在线平台，诸如维基百科或开放图书馆则广邀普通大众发表文章并用元数据描述这些发表的文章，这样就成为全文检索的完美补充。近期大规模数字化和 OCR 的广泛兴起，在先前没有数字化的资料中开放全文检索成为了可能，都使这一情形得到了很大改善。对于像 Google 这样的搜索引擎公司来说，这样显然比使用传统的编目方法更有效率-在他们的搜索引擎中，基于元数据的检索仅仅被视为一种补充方式。²

在这样的背景下，图书馆需要自问，在当下和未来，在传统编目之上构建的国家书目可以传递哪些附加价值。

对国家书目的要求

传统意义上，国家书目是以出版商或图书销售、图书馆和终端用户（特别是研究者和文学学者）这三组不同对象为目标的。这三组对象有着一定的共同点，即，他们都是建立在以下四个国家书目的特性之上的：

1) 数据的完整性和可靠性

对于书商、出版商和研究者来说，国家书目要记录所有的出版物（专业性的书目），不能有任何政治或内容上的偏见，这曾是且仍将是必需的。而且，对于编目规则的遵循扮演了关键角色，特别是在数据积累过程中。

2) 数据的现时性

特别是对于图书馆销售和图书馆来说，国家书目中的数据现时性显现得格外重要。20 世纪后半期，为了应对出版物数量的急剧增长，缩短出版周期，我们在数据加工上花费了大量的精力。

3) 可参考性

具有完整性、可靠性和现时性几个特点的国家书目便可以作为科研目的的参考点：如果一本书在国家书目中列出，那么我们就可以确认它的存在，反之，如果一本书并没有在国家书目中列出，那么很有可能情况就是该书从未出版。

4) 数据的永久保存

为了作为引文参考点，书目记录仅仅是遵从上述质量标准还是不够的，还需具备数据永久保存性。直到 21 世纪初，这还不是个问题，因为书目本身以印刷本的形式出版。只有在积累的书目中修改不合适的条目，但是除此之外没有方法删除一本书已被出版的证据。

未来国家书目要求需要进一步估量其自身在万维网中的应用与复用。

² 参看 <http://books.google.com/intl/en/googlebooks/about.html>

万维网越来越变成信息交换的开放空间：就是所谓的关联数据云。³自 2008 年来，相互关联的数据集的集合在规模上增长巨大，但是人们对于数据质量和数据永久保存的却知之甚少。为了让语义网起作用，我们需要一定水平的信息可靠性，这种可靠性既要考虑信息质量问题又要特别注意信息永久保存的问题。如果知道我们今天所标注的信息在下周就会唾手可得地被引用，那么这一网络便可以持续地发展壮大。⁴这种情形尤其应用于在线出版物及它们的元数据中。为了确保信息的长久可靠性，我们必须保证这部分万维网的稳定性，而这将可能成为图书馆和其他文化遗产保护机构未来工作的一项重要任务。

由此我们可以得出以下结论：

1) 信息的完整性和可靠性

在线出版的兴起事实上使得完整收集与国家书目相关的所有创造性著作变得几乎不可能。而有可能的情况是，国家书目机构甚至不能确实地预测其出版物是多么的不完整。然而，他们却可以自己限定其想要完整收藏的出版物种类（例如：印刷型媒体、高质量博客，以及某种类型的学术成果，独立出版物），并且明确对哪些网页的定期关注，可以满足其信息完整性要求。

一部分元数据将被自动创建，而这些数据并不是那么完美。这并不是说元数据无用，而是说生成的过程应当得到良好的记录，并且应当让数据使用者明了这些数据是机器生成的，哪些情形适合使用这种类型的数据。

对某种规则集合的遵循需要我们少一些关注 AACR⁵或是 RAK⁶编目规则，而多一些关注技术和语义数据的互用性来整合外部服务。

2) 信息的现时性

多数数据使用者期望在线出版物能一出版就被国家书目所收录。这样的做法与数据的可参考性和永久保存的要求相悖：如果国家书目基于的元数据是以出版物早期版本著录的，就会相应地增加书目著录变更的风险。这就要求图书馆信息系统能够存档不同版本的元数据以保存出版物版本变更的不同状态。

3) 信息的可参考性

国家书目在未来将继续保持其出版物参考中心的地位。特别是对于电子出版物，这一角色的重要性将有待进一步提高。为了提供参考服务，我们必须为书目数据提供在电子环境中可被辨认的唯一标识符。

³ 关于关联数据和关联数据云的简介参看 Heath and Bizer (2011)

⁴ 参看 Schuster and Rappold (2006)

⁵ 英美编目条例, cf. <http://www.aacr2.org/>

⁶ 德国字顺编目规则, 德国和奥地利使用的编目规则, 参看

http://www.dnb.de/EN/Standardisierung/Regelwerke/regelwerke_node.html#doc3132bodyText1

4) 信息的永久保存

确保数据的长期可用性和书目著录的完整性是数据信息长期可参考性的先决条件。这不仅是书目数据和规范数据长久可识别的问题，而且还关系到版本控制、处理修改与删除和特定状态的识别等一系列问题，比如通过使用时间戳的方法来记录时间。

德国国家书目在数字时代的网络拓扑结构

考虑到这些要求，国家书目显而易见地不能将其自身隔离于万维网之外。在德国，国家书目通过使用越来越复杂的数据结构，已然从卡片目录集合变身为电子数据库，在这些书目数据库中，与规范数据关联的书目著录经常要与系统之外的资源关联。下一步就是要让国家书目融入万维网中，以便使国家书目在各个方面与网页融为一个整体，无论是网络拓扑结构上、功能上还是技术上、组织上。

通观国家书目的特点，我们便可以理解作为万维网中一个图谱的国家书目。这一图谱不仅与其自身相互关联，而且还与万维网的其他部分相互关联。图书馆信息作为万维网中的一个图谱的这种观点不仅应用于德国国家书目：文化遗产对象、规范文档、叙词表和分类法等其他目录也都应被纳入为网页的一部分。

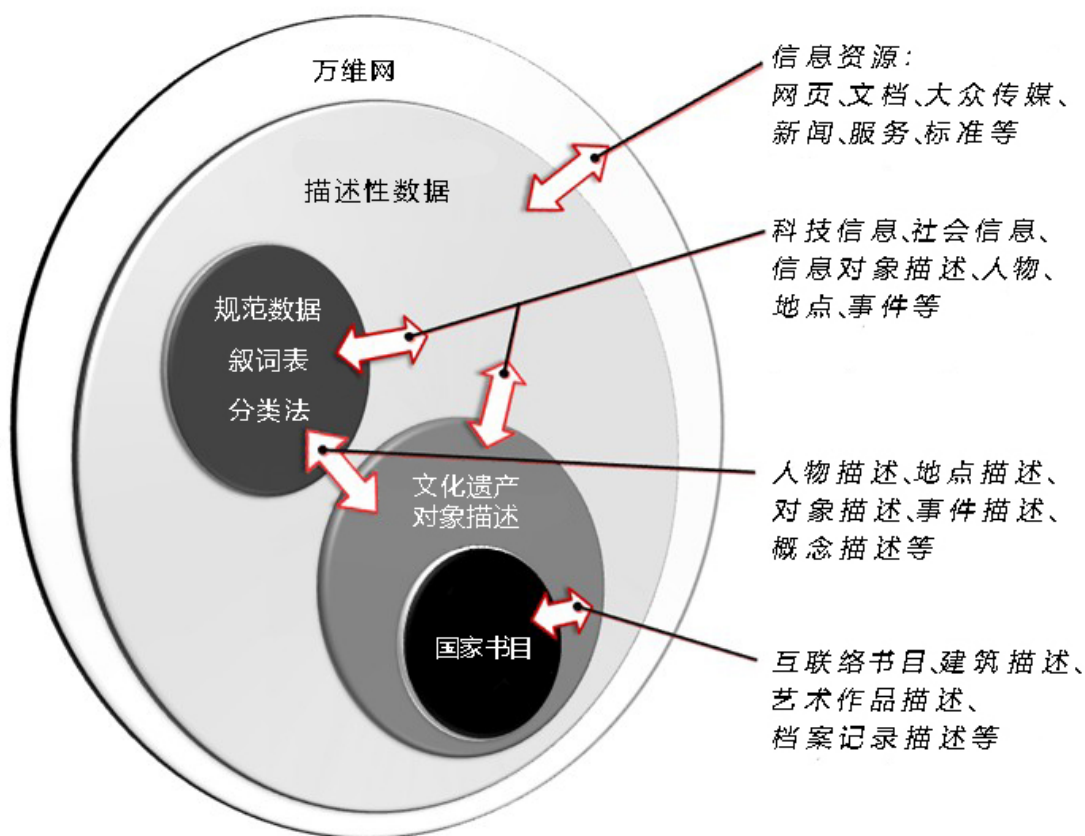


图 1: 未来的国家书目在万维网中的模型图示. 图示中椭圆为子图, 子图间的箭头表明子图中节点间相互关联的可能性。

因此，未来的国家书目不是一个封闭的体系，也不可能完全地独立于万维网之外。纸质印刷版的书目或完整的图书馆数据备份可能仅会变成为人遗忘的衍生品，因为国家书目的完整性依赖于图书馆数据库之外的数据。

在本图中节点代表由智力或艺术而产生的“人工制品”。这些“人工制品”我们可以统称其为“文化资产”，他们可以等同于 FRBR 中的第一组实体⁷。每个节点都集合了他们所代表对象的重要属性并用这些属性去描述它。这些属性大多与我们能在编目卡片上找到的著录有着很多共同点，因为它们都是遵循一系列特定的规则而获取的信息（或事实）。一种出版物的正题名仍是同一字符串，它的 ISBN 号是同一数字序列。大多数其他属性要参考其他实体，即，定向边界连接一个节点与万维网其他节点。其他的节点也可作为元数据节点：文化资产，人物，团体实体，时间，地点，概念等描述。其中一些节点是由图书馆编辑的（例如 GND,⁸ RAMEAU,⁹ and LCSH¹⁰）。从图书馆元数据到其他节点的参考也有可能的，例如，关联到对象本身（如果是一个万维网资源）或是参考由图书馆界之外机构组建数据集中的节点，比如，由名称规范到相应维基百科词条的关联。

从这个观点看来，被称为德国国家书目的信息源的图表中，德国国家书目已经是大量相互关联的数据集。当我们用到“数据源”这个概念时，我们指的是整个图表中的现在已经可获取和未来将能够获取的全部信息。即，数据源不仅包括由德国国家图书馆负责编辑的数据而且还包括与德国国家书目数据相关的所有数据集。由于德国国家图书馆向万维网开放其数据集（通过提供关联开放数据服务的方式），我们的数据源将不可避免地包括极具增加的非图书馆数据。数据源中的关联来源于以下几种机制：通过国际标准号的使用，如 ISBN 或 ISMN，使用标准号与图书馆的馆藏单件发生暗含的关联；由映射受控词表、智力(如，GND, LCSH, and RAMEAU)或机器运算程序(如，VIAF¹¹)的规范文档创建的关联；来源于合作的第三方机构的关联（如，维基百科）。

范围与限度

我们很难界定一个国家书目确切的范围，但是我们可以非常容易的找到包含所有与其创建相关信息的数据超集。处于顶层中的国家书目是万维网中所有描述性数据的一个特有子集。而且也是文化语境中所有描述性信息的子集。这里提到的数据包括与文化遗产语境的相关数据，诸如人物、主题、概念和地点等描述。此外，还应有些限制，如特定的地理地域限制（国家）和文化遗产对象的特定类型的限制（如，图书、地图、明信片、声音记录等等）¹²。

另一方面，国家书目界限是指行政责任。为了满足长期可靠的高质量数据的要求，责任实体必须有足够的政府支持，必须充足装备来满足可预见未来的数据管理和数据加工要求。对于国家图书馆或国家书目机构而言，这是他们存在理由的核心和应担当的一部分责任，经常但并非总是为法定缴存法所支持。¹³这并不能推断所有由国家书目机构组织管理的数据都会自动成为国

⁷ FRBR 实体的一个描述 参看 IFLA (2009)

⁸ The Gemeinsame Normdatei (GND -综合规范文档)涵盖全部类别实体的德国规范文档，作为图书馆书目信息，其他规范文档编目信息的使用者，如档案馆，博物馆，项目，科学文化机构的普遍权威的参考体系。见 http://www.dnb.de/EN/Home/home_node.html

⁹ RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) 由法国国家图书馆、法国大学图书馆以及部分法国公共图书馆所使用的文献语言。见 <http://rameau.bnf.fr/>

¹⁰ 美国国会主题词表(LCSH) 用于标引国会图书馆和世界范围的资料 见

<http://id.loc.gov/authorities/subjects>

¹¹ VIAF -虚拟国际规范文档 (<http://viaf.org/>) “为提供方便存取世界主要名称规范文档而设计的国际服务” 参见 <http://www.oclc.org/viaf/>

¹² 对此进一步讨论，参见 IFLA (2009)

¹³ 参见 Andersen (1974) p. 11.

家书目的一部分：德国国家图书馆管理多个数据集，包括与德国国家书目没有关联的特藏部分。

内容

在德国国家图书馆的一个研讨会上，参与者给出了各自认为的德国国家书目的构成定义。答案从“图书馆组织的所有数据”到“除特藏外的仅仅是题名”都有。但是即使是支持最弱解决方案的组也没有准备即刻完全放弃规范数据。这并不令人惊讶，因为规范文档是编目过程的产物，包含着各种信息，例如：作者姓名，它对于一个正确的 ISBD 显示是必要的，也就是说，没有这一信息，书目记录就没那么有意义。另一方面，并非规范记录中的所有数据对于印刷版国家书目中书目著录的 ISBD 显示都有重要意义。

这便得出如下推论，我们可能不能回答哪些数据在实体层面而不是在个体特征层面组成了在国家书目这样的问题：不是文化遗产实体记录所需的全部特征都是该文化实体本身的特征，而是依附于其上的其他实体。¹⁴ 同样应用于检索环境中。理想化的状态是，联机目录应当使用其全部可获取数据来优化检索过程。这就包括了没有任何参考书目实体而创建的图书馆范围以外的数据。源自像 geonames.org 数据集地理坐标的使用可以使我们对地理位置或地理相邻性进行检索。

在联机目录中 ISBD 的显示和检索仅仅是围绕国家书目众多服务中的两项。并非所有在这两项服务中用到的数据都必然地得被认为是书目的组成部分，这样我们也不能通过观察这些服务（纸质书目，联机检索，国家书目服务），得出究竟是什么组成了国家书目。我们只能在这些服务之上构建国家书目。

当我们基于不限于国家书目的数据来构建国家书目服务时，其后果必将导致服务与数据依赖性被切断。如果我们把注意力集中到国家书目的核心：出版物上，情景会就清晰一致起来。那样的话我们就可以得到文本元素和参考的最小数据集，但他们不能独立起作用。近观 FRBR¹⁵ 和 RDA，我们可以得出更确切的说法：国家书目记录载体表现，就是“作品的内容表达的物质体现”¹⁶。作品、内容表达和单件数据等作为规范记录都是他们自身数据集的组成部分，国家书目中的数据仅包括对其他实体数据集的参考。

创建过程

国家书目应当越来越多地成为由信息消费者和信息生产者组成全球性信息生态系统的一部分。国家书目与越广泛的研究机构所拥有的独立信息的关联性越强，就有更多的应用从它的知识库中受益。图书馆所提供的服务将越来越多地有赖于外部资源信息。与文化遗产相关的实体附加信息的有效性，无论是文化资产，个人物还是概念，使得我们改进自己的书目服务。对于部分信息来说，缩减我们自己组建数据而是利用第三方信息都是有可能的。

¹⁴ 这本质上是对以下陈述的延伸：“书目著录通常应当基于单件作为载体表现的代表，并可能包括属于被体现作品和内容表达的属性” 源自 IFLA (2009) p. 4.

¹⁵ 书目记录的功能性要求. 参见 <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

¹⁶ IFLA (1998) p. 21.

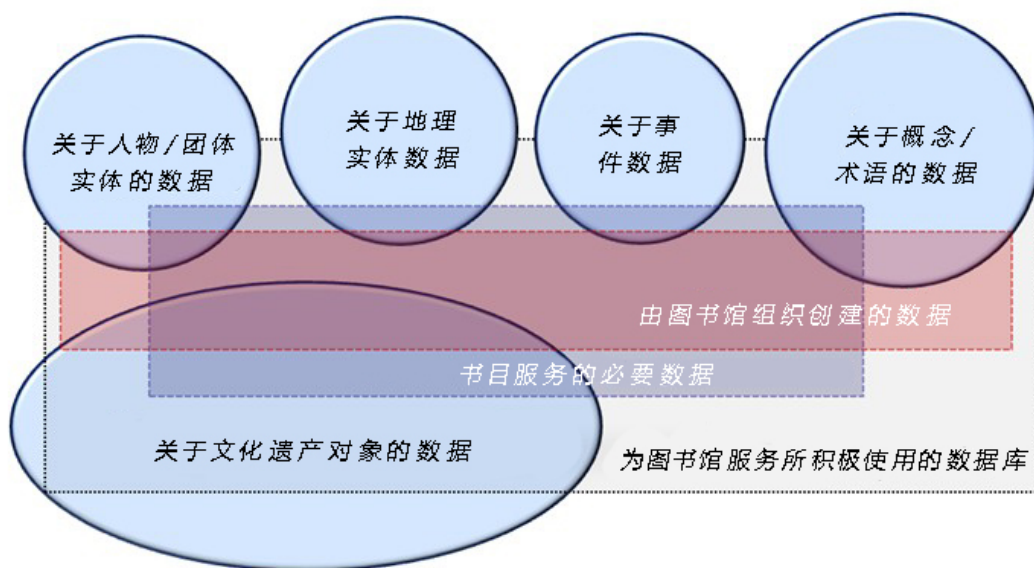


Figure 2: 图书馆信息创建与信息使用的未来可能的分布情况

1. 不是图书馆目录所有信息对于他们的书目服务都是必要的。但他们又可能是不可缺少的，如，对于流通目的或是存档。
2. 并非所有书目服务所需的必要信息都是由图书馆目录提供的。
3. 图书馆使用的数据库基础要远大于那两个数据集之和。

这种方法并非是全新的。在德国国家图书馆，我们至今仍不能人工创建书目著录的所有部分，而是再利用第三方信息，如由出版社提供的元数据，或半自动创建信息如目录，或机器生成主题词。在这种背景下，信息复用的扩展和对选定网络社区的受控开放编目，似乎成为符合逻辑的下一步。

对于图书馆来说，放弃一些工作似乎不是问题，问题在于更多地关注我们的力量，例如：高质量信息的创建，知识组织，规范控制。当共享信息作为普通货品，有效地分工必将使每个信息提供者都能运用其核心竞争力。国家书目中心负责以文本形式出版的文化遗产对象，负责从这些对象到其他实体的参考，负责他们的属性，如果这些属性对于书目服务是必要的且如果没有其他机构能够提供质量相当的信息。其他信息提供者负责其他属性。

由于出版物数量的急剧增加，图书馆要限制手工创建，维护和核实元数据到一定范围内，在这一范围中这些工序相对于自动数据创建更能附加价值。在上述状况中，对于什么附加价值的定义，并不是依赖于对特定的编目条例的遵循，而是应当一方面关注信息选择和过滤有用性的增长，另一方面关注图书馆用户的商业模式。

质量特征

为了确保图书馆服务的质量不受到第三方数据整合的影响，但反而从中受益，图书馆需要认真考虑他们复用的是哪些外部信息资源。数据的可靠性（可参考性，永久保存，标准使用和质量保证）是关键。数据的来源，如果可能的话，数据创建背后的过程应当是透明可见的，应当进行适当地过程质量控制。特别是每个外部数据集都应当以长远的观念得以必要的评估。

在国家书目中复用第三方数据集可能会引发与现行编目条例的冲突。多数、甚至是几乎所有的图书馆界之外的信息提供者都不会去留意 AACR II, RAK, 或是 RDA, 而且我们也不能期待他们在这些领域投以任何关注。相反地，我们应当理清现存的编目条例，带着他们是否能够帮助国家书目实现其意图的问题去评估他们：作为国家出版物全貌的综合记录去服务。只有这些条例可以达到这样的意图，那么我们才能将其作为评价第三方数据集质量的标准。

一个机构使用第三方元数据时必须首先定义什么构成一个可信的数据提供者。为了给出这个定义，我们必须有一份类似于可靠数字档案的标准表。在生成这个标准表时，重要的标准应当用于判断数据提供者所提供的数据是否满足互用性的要求。非源于可信数据提供者所提供的数据必须相应的标出。

这同样应用于机器生成数据。在这里，重要的是良好的存档数据生成过程并确保结果是可复制的。如果我们可以为国家书目子集生成同质数据，用户接纳度将大幅攀升。能达到这样的一种方法是，为可信的自动生成过程协同编辑一份标准表。不言而喻的是，定期监管这些过程是必要的。

技术实现

以上，我们讨论了国家书目可以被建立在融合了规范文档和关联数据集信息的载体表现元数据的最小集合之上。这样，带着国家书目成为万维网图谱中一部分的构想，关联数据技术的应用成为一个自然而然的结果——本质上，是一个为实现实体间可机读关联的以超文本协议为基础的框架——为了书目的技术实现。

下面让我们来看看 Tim Berners-Lee 对于关联数据的要求，¹⁷执行这些要求相当简单：

1) 使用统一资源标识符（URIs）作为对象的名称

我们在国家书目中所用到的每个实体都需要其自己特有的标识符。这个方法至少应用于载体表现，受控词表，团体，个人，理论上也应当用于其他规范数据如出版社。

2) 使用超文本协议（HTTP）URI 以便使人们可以查找他们的名称。

当图书馆将标识符分配给他们的实体时，这些实体应当支撑万维网的基础架构。

3) 当有人查找一个 URI 时，提供有用的信息

为了最大化地复用图书馆数据，当有人不参考标识符时，我们必须提供由 URIs 标识的信息节点的精确描述。由此，我们可以邀请数据提供者使用 URIs 通过关联规范数据或书目信息来复用图书馆数据集。

¹⁷ Berners-Lee (2006)

4) 包括与其他 URI 的关联，以便于使用者发现更多查找内容

当我们复用规范信息或复用我们书目记录中由其他内容提供者所提供的信息时，我们可能会选择显示某些关联实体的属性，如，著者名称或是主题标目的首选形式，与载体表现的题名一起，同时还要提供链接到一个使用者能找到包含更多信息的网页，如作者的出生日期或是职业。

在构建这一基础架构时，信息模型是否使用 RDF 还是任何其他技术手段都并不重要，关联数据本身的格式是不可知的。RDF 技术有很大可能使数据和语义关联以通用的方式变得可被机读（进而被机器理解）。但这将在很多情形下产生高度复杂的数据模型，在使用 RDF 展开应用时发生了很多问题，这些问题引起了很多争议讨论。¹⁸在德国国家图书馆，我们曾经历过很多这样的情形，如，临时关联或图表源，需要在不使用过度复杂的模型的情况下比 RDF 阐释更具表达性——至少在现在。这一情况或将随着 RDF 模型的成熟和诸如命名图表概念的流行而得到改变。只要我们坚持在关联数据原则的基础上构建我们的数据模型，在万维网以稳定的标识符公开我们的数据，参考其他实体来丰富数据的表现形式，我们将为更具表达性的著录语言铺平道路。

使用

在网页上检索信息时，大多数用户使用如像 bing, ask.com 和 Google 这样的商业搜索引擎，或是去社交网络，如 LinkedIn 和 Facebook。¹⁹对于国家图书馆和其他文化遗产机构来说，这将意味着我们必须给那些公司这样的可能性：创造性地将我们的数据纳入他们服务之中，以便使用户从网页搜索中受益于这些信息。例如，欧洲文化遗产门户网站 Europeana 便为藏品中的每个对象都创建了这样一个特殊的登录页面，且这些登录页面都设有唯一的 URI，用以增加该藏品信息的可见度并提高来自搜索引擎的通信量。²⁰

这表明图书馆自己的联机目录在信息格局中将继续占有重要地位。作为一个高度专业的搜索引擎，该标杆应用显示了图书馆的数据可以怎样处理，如何为特定用户组群体提供定制的信息。另一方面，图书馆目录的附加价值和专业网页搜索引擎间的边界日益模糊。我们可以想见未来一些搜索引擎提供商将会关注书目信息。例如，为特定科学社区。高质量书目信息的可用性嵌入了用户检索环境中，这将有益于我们的用户并使他们能够更容易地查找、选择并获取恰当的资源。

为了使图书馆信息在其他服务中更易被复用，必要的是要明确地表述协议术语。特别是对于国家书目来说，附加价值是通过增加可视性和信息复用来创造的，鼓励这种做法的最简便方式是在灵活开放的协议下提供明确允许商业复用的信息。这可能是矛盾的，考虑到像 Google 这样的大参与者在经济上都处于要向图书馆元数据付费的地位上。尽管如此，他们是否愿意还是一个问题，并且如果图书馆信息在主要的搜索引擎上是不可见的，那么对于多数通过网页检索的人们来说，这些信息就是不可见的。开放许可证的使用使得图书馆信息被整合到开放的平台，如维基百科或是中小型网站，如新闻在线门户。通过这种方式，可见性的增加无疑会超过（可能）的收入损失。

¹⁸ 作为众多例子中的一个，见 Miličić (2011) 的相关讨论。

¹⁹ 依据近期 PEW 报告，92% 的在线成人使用搜索引擎查找信息。参见 Purcell (2011)

²⁰ Clark et. al. (2011) p. 15-17

德国国家图书馆目前的工作和德国国家书目的含义

德国国家图书馆也像其他国家图书馆一样，也在实施一些重大的改革。

自 2008 年起，德国国家图书馆陆续将其全部数据库作为关联开放数据公开；目前大约 70% 的题名数据是可获取的。最先公布的数据集是来自 GND 的规范数据，然后是杜威十进制分类法的德文译本。在为期六个月的项目结束时，我们已经可以将国家书目的题名信息的主要部分作为关联数据传递出去。除了在 CC BY-NC-ND 协议下有效的 DDC 数据，所有其他德国国家图书馆的信息也都使用创作共享 0 协议得以公布。

而且，我们越来越多地使用自动化程序创建描述性和主题元数据。一条线是为了丰富在目录中的书目描述，如，有了目录我们就可以增加来源于出版社和大学的元数据内涵，包括主题类别，例如，BISAC 编码——和主题词表。对于国家书目 O 辑的在线出版物，我们全部复用由内容托存者提供的描述性和主题元数据。²¹

这是目前我们评估自动为书目记录添加主题标目的可能性所做的一个项目。其过程是建立在自动标引器之上进行的。我们分析标引器如何处理不同类型的文档，并分析对自动标引的出版物的检索质量。初步成果将于 2013 年公布。

自动标引网页资源所使用的受控规范主题词表与我们智力标引过程所用的相同，这样可以保持信息检索的一致性。从而可以提供可靠的规范信息。而且，人物、地点等关联信息，还有规范信息中的语义关系也会被有效地使用。

今天的用户对于书目信息有着各种各样的期待，特别是主题检索，它包括可获取著作的概览，书目引文，或者直接访问出版物。在德国国家图书馆，我们秉承 IFLA 的一致标引政策的信念，并且使用在提供有序一致信息中仍然有着重要地位的受控检索。²²我们所有的用户都会从组织良好的主题结构中获益——这些结构独立于他们使用国家书目的信息格式、分配渠道、或代表显示之外。分类法体系和主题词表支持用户获得（查找，识别，选择或是探索）他们所需要的信息。大部分用户（终端用户或是专业用户）对于部分数据集感兴趣而不是全部的书目。大多数用户在检索我们的联机目录时可能甚至都不知道他们是在书目中检索。对国家书目需求数量的大幅提高是由于个人层面检索书目或 RSS 订阅请求数的增加。附加关联内容信息作为目次或是摘要范例的来帮助用户理解他们所检索的信息。在线内容关联信息的提供（特别是源自电子书和电子期刊）符合用户的需求。

由于发布信息数量的大量增加，将信息归类成可管理的单元，使其可读、可选并能够做到精确检索的变得越来越重要。理想的是完整细化的标引被应用与所有的文档中，以便帮助用户查找相关检索主题或附加内容信息。面对德国巨大的纸质出版物数量和 .de 域名下网页资源出版数量的增加，以及预算和人员的限制，我们决定依照渐进标引策略，这一策略能使我们在不同细节层面管理全部国家成果。这一模块策略²³允许多种存取点，允许选择不同媒体类型，但这都在质量标准控制之下是透明的、受控的。

²¹ 更多关于德国国家图书馆应对在线出版物的信息参见 Gömpel and Svensson (2011)

²² IFLA (2012)

²³ 德国国家图书馆书目信息的不同分布渠道概况 见 Svensson and Jahns (2010)

最低的标引级别给几乎所有资源提供一个十分广泛的分类号，**Sachgruppe/**主题组相当于一个粗略的分类号。这个结构以杜威分类法（百位类目或第二层级）的上两层级为基础。整合更深层面做出一些改变来满足本地用户的需求。小说和儿童书籍还有学校课本可以被分别选择。

其他一些项目帮助提高与其他数据集的关联性。最佳的起点是我们自有信息的优化。显而易见的是，文本著录到参考文献的转换十分适合于自动操作。其他为命名实体识别和实体名称到参照的转换的过程目前还处在实验阶段。在创建与外部数据集的关联时，德国国家图书馆有效的结合利用了合作，手工和自动程序的方式。通过 **VIAF** 我们可以自动将 **GND** 的人物关联到其他的规范文档，然而 **MACS**²⁴使用可编辑程序从 **GND, LCSH, and RAMEAU** 关联主题词表。一项与德国维基百科的合作就是在维基百科的文章和 **GND** 人物之间不间断的传输连接，还有在 **CONTENTUS** 中我们发展了一项技术来建立自 **GND**、维基百科文章和 **geonames.org** 的地理名称的网络。

为了进一步推动这一发展，**DNB** 与 **HBZ, Northrhine-Westfalian** 图书馆网络一起建立了 **culturegraph.org** 这一平台。基础的想法是提供一个技术的有组织的基础构架用以支持书目信息、叙词表、分类法和其他规范数据之间的相互关联。信息网络成果的也应作为关联信息被公布。这个平台的核心特征是支持可引性。目前，这一平台包括德国图书馆网络中自 1945 年以来的所有出版物的书目著录。初步的中期目标是创建数据簇来识别哪些书目记录描述相同的载体表现，使用唯一普遍的标识符来命名和发布数据簇。进一步的评估将表明，是否可能为（**FRBR** 意义上的）作品创建数据簇，是否那些信息可以用作知识作品的规范文档的种子。

所有的国家行动需要统合于国际化背景中，可能最重要工作将是继续参与致力于未来编目与信息交换的国际项目和标准组织的工作。

结论与未来工作

本文讨论了基于关联数据原则的图书馆数据的生态环境，在开放协议下将图书馆信息发布到万维网。为了更广泛地传播信息，我们应当以与主要的搜索引擎提供商合作为目标，并使他们将我们的书目信息统合入他们的数据库中。

为了服务广泛多样的用户需求，依据数据选择和格式，必要的是要使基础数据库尽可能的灵活。使用标引类推，我们需要从先组数据转变成为一种信息结构体系，在这种体系中我们可以依据满足不断变化的要求后组零碎信息。

目前的讨论是关于编目规则必须关注万维网中书目信息所表现出来的需求。**RDA** 把成为“为数字世界所设计的资源著录与存取的一个新标准”作为目标。²⁵尽管对于传统编目范畴正当的批评仍然很有偏见并且没有足够推进，所提供的增长的灵活性是迈向正确方向的一步。毫无疑问，进一步的修订还是必须的。

可能比图书馆编目规则更重要的是与图书馆生态系统之外信息互动的可能性。大部分可获取的有用信息——如，可能包含用户检索相关词的新出版物的出版社广告——被查找和检索目的所忽略。这要求图书馆拥有手段去储存第三方信息源使得这一信息可见于终端用户。

²⁴ 多语种主题存取 (**MACS**)是使用现存的标引语言，以开发一个允许多语种主题存取图书馆目录的体系为目标的项目。它目前使用 **RAMEAU, LCSH, 和源自 GND** 的论题主题标目 参见 http://www.nb.admin.ch/nb_professionnel/projektarbeit/00729/00733/index.html?lang=en

²⁵ **RDA** 战略计划 2005-2009. 见 <http://rda-jsc.org/stratplan.html>

进一步说，必要的是使引用有一个质量标准。如果一条书目著录被标记为“可引”，它就永不会被删除并且将有可能追踪到它被做出的所有更改。为了达到这个目的，我们必须提高数据创建和维护的过程，并且为自动过程得出的所有数据增加源信息。

最后，我们必须在关联图书馆数据到其他数据集上加大投资。只要有可能，我们就必须使用指向规范数据的参照，而不是指向文本内容的参照。为公共规范文档的国际化合作——如，**VIAF**，将向正确的方向跨出重要的一步，并且将帮助避免重复工作。作品题名的中心仓库将大大的帮助跨国界和语言的书目著录关联。出版社数据库（团体的一个子集）不仅能提供受控存取点，还能推进图书馆与出版社的合作。

国家书目应当是网页图谱中的一个组成部分。数字时代已经引发了信息行业的改变。我们不能止步于现存的结构和实践，图书馆应当接受改变而不是试图去扮演变化的代理。更灵活更开放是一个很好的开始。

参考文献

Anderson (1974) Anderson, Dorothy: Universal bibliographic control : a long term policy; a plan for action. Pullach 1974.

Berners-Lee (2006) Berners-Lee, Tim: Linked Data. Available at <http://www.w3.org/DesignIssues/LinkedData.html>

Clark et. al (2011) Clark, D. J.; Nicholas, D.; Rowlands, I.: D3.1.3 – Publishable report on best practice and how users are using the Europeana service. Available at http://www.europeanaconnect.eu/documents/D3.1.3_eConnect_LogAnalysisReport_v1.0.pdf

Gömpel and Svensson (2011) Gömpel, Renate; Svensson, Lars G.: Managing Legal Deposit for Online Publications in Germany. 2011. urn:nbn:de:101-2011061609. Online available at <http://nbn-resolving.de/urn:nbn:de:101-2011061609>

Heath and Bizer (2011) Heath, Tom; Bizer, Christian: Linked Data: Evolving the Web into a Global Data Space. San Rafael (Calif.) 2011. (Also online at <http://linkeddatabook.com/editions/1.0/>)

IFLA (1998) Functional requirements for bibliographic records : final report. München 1998. All references in this paper are to the 2009 revised online version at http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

IFLA (2009) IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its Glossary. Ed. Barbara Tillett and Ana Lupe Christián. München 2009. For a list of online versions (incl. translations) see <http://www.ifla.org/publications/ifla-series-on-bibliographic-control-37>

IFLA (2012) , Guidelines for Subject Access in National Bibliographies. Ed. Yvonne Jahns. Berlin: 2012.

Miličić (2011) Miličić, Vuk: The Ultimate Problem of RDF and the Semantic Web. Blog post, available at <http://milicicvuk.com/blog/2011/07/19/ultimate-problem-of-rdf-and-semantic-web/>

Purcell (2011) Purcell, Kristen: Search and email still top the list of most popular online activities: Two activities nearly universal among adult internet users. 2011. Available online at http://pewinternet.org/~media/Files/Reports/2011/PIP_Search-and-Email.pdf

Schuster and Rappold (2006) Schuster, Michael; Rappold, Dieter: Social Semantic Software – was soziale Dynamik im Semantic Web auslöst. In: Semantic Web: Wege zur vernetzten Wissensgesellschaft. Ed. Tassilo Peregrini. Berlin 2006.

Svensson and Jahns (2010) Svensson, Lars G.; Jahns, Yvonne: PDF, CSV, RSS and other acronyms: redefining the bibliographic services in the German National Library. 2010. Available online at <http://www.ifla.org/files/hq/papers/ifla76/91-svensson-en.pdf>