



图书馆中的关联数据：专门服务于图书馆的记录结构转变为基于RDF的数据模型，我们能得到怎样的益处

Getaneh Alemu

Brett Stevens

Penny Ross

&

Jane Chandler

University of Portsmouth
Portsmouth, United Kingdom

中文翻译：李恺

(美国雪城大学信息管理学院)

Chinese Translator: LI Kai

(Information School, Syracuse University, USA)

Meeting:

92 — New futures for bibliographic data formats: reflections and directions — UNIMARC Core Activity

摘要

当前的元数据原则和标准往往导致了以文档为中心的元数据，而不是以数据为中心的元数据；导致了人类可读的元数据，而不是机器可处理的元数据。为了让图书馆能够创造并且利用可共享的、可混搭的、以及可重用的元数据，图书馆要把当前的图书馆模型——比如RDA、以及FRBR——转换到根据关联数据建立的模型，以实现概念上的转换。在技术格式上，图书馆能够实现关联数据技术格式的跳跃性的发展，比如资源描述框架（Resource Description Framework, RDF），而不用让当前图书馆元数据的操作发生断裂。本文对于图书馆和标准机构提出了六个重要的推荐。图书馆要应对关联数据提出的挑战、采用关联数据原则的最小需求、开发本体、决定需要保留哪些需求成为关联数据网络的一部分、以及采用混合的元数据方式（基于标准的以及社会构建的）；最后，本文归纳并且讨论了这种元数据重新概念化的五个主要的益处，作为本文的结尾。这些好处包括元数据的开放性和共享、帮助对信息资源的偶然的发展、关联元数据来发现资源使用的模式、时代精神以及新出现的元数据、基于面的导航、以及使用链接来丰富元数据。

关键词：元数据，元数据标准，MARC，关联数据，RDA，社会构建的元数据，混合元数据的方式

介绍

当前图书馆使用的、用以支持元数据功能的原则、标准和协议有着悠久的历史。（Denton, 2007; Dunsire, 2009; IFLA, 2009; Lubetzky, 1953; Wright, 2007）图书馆编目的历史横亘数千年，可以追溯到公元前3世纪亚历山大图书馆（Library of Alexandria）的时代，当时的图书馆馆长 Callimachus 准备了希腊文学的系统性的书目。（Day, 2005; Wright, 2007）而 Morville 则又把编目的历史进一步上溯到亚述帝国（Assyrian Empire）的时期，国王 Assurbanipal 建立的图书馆在公元前650年收藏了30000块泥板。但是，编目的标准化的尝试在19世纪才真的出现。

（Coyle & Hillmann, 2007; Denton, 2007; Weinberger, 2005, 2007; Wright, 2007）现代编目重要先驱包括 Anthony Panizzi 爵士、Charles Cutter、Melville Dewey、Paul Otlet、S.R. Ranganathan、以及 Seymour Lubetzky（Denton, 2007; Lubetzky, 1953; Wright, 2007）。根据这些先贤的著述，地区和国际性的联盟开展了众多项目，旨在建立严格的编目原则和规则——其结果就是诸如1961年的《巴黎原则》（Paris Principles）、1967年的《英美编目条例》（the Anglo-American Cataloguing Rules, AACR）、1960年代晚期的MARC格式、1971年的《国际标准书目著录：专著》（International Standard Bibliographic Description (ISBD) for Monographic Publications）、1996年的《书目记录的功能需求》（Functional Requirements for Bibliographic Records, FRBR）、以及2010年的《资源的描述和检索》（Resource Description and Access, RDA）的这些标准出版物。（Denton, 2007）

但是，随着数字图书馆的馆藏规模持续增长，当前的元数据原则和标准也面临着严苛的检验。

（Alemu, Stevens, & Ross, 2012; Coyle, 2010; Coyle & Hillmann, 2007; Lagoze, 2010; Mathes, 2004; Shirky, 2005; Veltman, 2001; W3C, 2011; Weinberger, 2005, 2007）批评者认为当前的图书馆标准带来了一些限制，这些限制源自传统的卡片目录系统，它们是数字图书馆在规模上的失败、以及图书馆无法和当前更大的信息环境产生互操作的原因所在。（Coyle, 2010; Coyle & Hillmann, 2007; Weinberger, 2005, 2007）使用当前编目标准和技术格式产生的元数据主要是为了人类的消费，而不是为了机器的处理。这是因为图书馆中使用的标准和技术同时存在着概念和技术上的限制。Coyle（2010）认为：“图书馆目录一直以来都是图书馆数据的唯一情境”，因而图书馆数据无法和外部信息提供者进行互操作。

对于原则、标准和协议的挑战可以用两种视角进行观察：概念的视角和技术的视角。当代元数据标准——比如FRBR和RDA——所蕴含的概念，毫无疑问产生了作为人类消费的文档的元数据记录，而不是用于机器处理的元数据记录。（Coyle & Hillmann, 2007）而且，当这样的元数据标准通过MARC这样的技术格式来推行的时候，所产生的记录又产生了元数据冗余、数据不一致、缺乏粒度、以及复杂性这样的问题。（Coyle, 2010; Coyle & Hillmann, 2007; Day, 2000; Guenther & McCallum, 2003; Tennant, 2002）即便当“让MARC死”这样的号召出现了很长一段时间之后（Tennant, 2002），MARC标准仍然是图书馆中绝大多数人使用的元数据结构。这个现象可能有如下几个原因，比如MARC深深的嵌合在图书馆的系统和功能中，因而任何改变都是太困难而且昂贵的；或者是，归根结底，MARC对于图书馆、以及实现其目的来说是“足够的”了；又或者，替代性的格式，比如可扩展标记语言（eXtensible Markup Language, XML）并没有带来额外的功能，来让改变产生足够的好处和必要性。但是，仍然有人质疑MARC是否够用；另有一些元数据专家断言，这个标准并不适合于机器可操作、可处理的元数据。

（Coyle, 2010; Coyle & Hillmann, 2007; Wallis, 2011a, 2011b）

人们建议的替代性方案包括采用关联数据。作为关联数据的发明者，Berners-Lee等人（Berners-Lee, et al, 2001）提出，最初的网络在概念上是以文档为中心的，链接本身并没有任何语义；而当网络通过超文本标记语言（Hyper-Text Mark Up Language, HTML）这样的技术格

式得到推行的时候，所产生的页面更适合于人类的消费，而不是机器的处理。虽然网页能够非常有效的分享文档和创造协作的机会，但是文档需要人们的干涉，其语义才能被理解，即在电脑屏幕上被呈现出来之后。简单来说，机器不能理解这样的文档。为了解决这个不足，Berners-Lee等人发明了语义网的概念。推行语义网的技术包括RDF、RDFS、SPARQL以及OWL。（Berners-Lee, 1997; Berners-Lee, 1998; Berners-Lee, et al., 2001）根据Allemnag和Hendler的观点（2008），“[关联数据]的大意就是在数据层次，而不是在文档呈现的层次支持分布式网络。它不再让一个网页指向另一个网页，而是让一个数据对象指向另一个数据对象，它使用了全球范围的参引，称之为统一资源标示符（URIs）。”尽管人们对关联数据的兴趣日增，但是有一个讨论仍然在持续的、大范围的进行着，那就是是否应该对传统的以图书馆为中心的概念性原则，比如RDA和FRBR，进行渐进性的改变。另一个问题同样重要，那就是从MARC到RDF的转变是否应该是逐渐进行的，还是我们有足够的理由来抛弃一些或者全部这些传统的格式，继而采用新的方式。（Coyle, 2010; Coyle & Hillmann, 2007; Marcum, 2011; Styles, 2009; Styles, Ayers, & Shabir, 2008; W3C, 2011; Wallis, 2011a, 2011b）

标准机构和图书馆应当考虑的另一个问题，在当前元数据的重新概念化这方面，就是网络2.0方式对于数字图书馆创建和使用元数据的影响。尽管网络2.0在图书馆中的角色已经被广泛接受，但是正如Evans（2009）所说“其意义并没有被完全认识，尤其是对于图书馆学的影响”，其原因或许如Lagoze（2010）所说，传统图书馆模型和当下的网络2.0的方式之间存在着不相容。社会构建元数据（Web 2.0）的方式被人批评为平面、单向并且充满了不一致（Gruber, 2007, 2008）；但是在另一端，以标准为中心的元数据方式则受到了刻板的等级制、以及不能表达用户词汇表这样的批评。（Shirky, 2005; Veltman, 2001; Weinberger, 2007）如Morville（2005）所说，专家控制的元数据和用户生成的元数据不应该被看做是两种截然相反的方式，它们不仅能够共存，而且应该相互影响彼此。从语义网对于元数据优化的角度（Web 3.0），混合的元数据方式——不仅包含对于信息对象的物理描述（比如作品、题名、ISBN、主题、格式等等），也包括了描述其社会文化方面内容的元素（用户标签、评论、书评、链接、评分、推荐）——是至关重要的。换言之，对于元数据社会空间的表现（Web 2.0）和对于数据元素标准化和客观记录（Web 1.0）应该是同等重要的，它们都被用来描述信息对象的实体特征。这对于元数据的丰富性尤其相关，在这方面，我们可以和根据标准的、图书馆员创建的元数据一道，来利用用户生成（社会构建）的元数据。

本文讨论了对当前元数据原则和标准的挑战，并且提出了从以文档为中心的元数据方法转变到以数据为中心的方法能够带来怎样的好处和意义。我们识别了两个重要的大类，也就是元数据原则（比如RDA和FRBR）以及技术记录格式（比如MARC），并且推荐了这些类别中的哪些需要被修正、改变或者维护，以适应关联数据的原则。

关联数据的原则

关联数据的数据模型，正如它的名字所指的，能够识别、描述、链接并且关联结构化的数据元素，这和关系数据库系统的运行方式有些相似，尽管前者是在网络级别上运行的。关联数据的总体目标在于帮助数据的重用、交叉关联、整合以及共享。（Berners-Lee, 2009; Shadbolt, 2010; W3C, 2011）Berners-Lee（2007）指出：“增加一个页面能够提供内容，但是增加一个链接则提供了组织、结构以及对于网络上的内容的支持，这些让内容作为一个总体变得更有价值。”

我们需要指出，关联数据是一种“元模型”，它提供了一个框架来定义、设计、开发并且维护一个给定域名中任何类型、任何大小的编码方案和词汇表。这实际上意味着机构，比如图书馆，并不需要放弃现有的元数据标准、受控词汇表、规范列表以及遗产元数据。但是，为了和

关联数据云这个世界实现互操作，人们期待各个机构能够采用关联数据的结构。

关联数据的原则始于关联数据最基本的组成部分，也就是，使用全球独一无二的统一资源标识符来清楚的指明信息对象、个人、地点和事件这样的事物。（Berners-Lee, 2009）其他重要的原则包括能够参引给定的URI、使用超文本交换协议（Hypertext Transfer Protocol, HTTP）来获取特定的URI所指涉的相关信息、使用RDF/XML这样的数据格式、使用词汇表来定义语言，比如RDFS和OWL、使用称之为SPARQL的检索语言、最后但也很重要重要的是，在数据集的内外整合各个方向的链接，于是丰富数据并且提供在情境上的重要性。

关联数据作为图书馆元数据模型

关联数据及其相关的技术在建立元数据模型、编码、呈现以及共享上扮演着重要的角色。使用URI用于元数据元素的名称、标签和关系减轻了元素使用中命名和发现的冲突。人们已经指出，RDF简单的数据模型能够让我们在信息资源之间创建语义链接。（Coyle, 2010; Coyle & Hillmann, 2007; Day, 2000, 2003a, 2003b; Helen, 2010; Nilsson, 2010; Rothenberg, 2008; Styles, et al., 2008; W3C, 2011; Wallis, 2011a; Wilson 2010）

封闭系统和开放系统的假说

关联数据原则运行于一个开放、动态和互动的系统中，并且在这个系统中成为可能。但是，图书馆标准在大多数情况下运行于封闭和静态的环境中，这些环境往往脱离于一般的网络信息环境。根据Allemnag和Hendler（2008）的观点，关联数据提供了一个分布式的数据模型，在其中，“任何人可以对任何主题说任何事情”会导致实体意义的“多样和冲突”。这两位作者证明，“对于在这种意义上的开放的世界，我们必须假定，在任何时候，新的信息都能够保持透明；而且我们不能根据某一时刻可以得到的信息就是全部信息这个假定，来得出任何结论。”

在数字图书馆的情境下，开放系统的假说让它自己对于来自于不同观点的新的元数据贡献保持开放，因而会我们从我们称之为网络效应的现象中得益甚多。对于URI和本体的使用——这些在一个给定的领域中通过使用词汇表和概念之间的关系，明确的界定了概念（Gruber, 1993）——能够混合、匹配和融合。

当前关联数据在图书馆中的使用

这看起来是一个急切的问题，尤其是在公共部门中，比如国家图书馆，好让它们的数据数据自由和开放的被人获取。（Wilson, 2009）2011年见证了两个重要报告的发表，这两份报告都提到了使用容易获取和可重新使用的开放图书馆数据的格式的重要性。（Library of Congress, 2011; W3C, 2011）正如Haslhofer和Isaac指出的那样（2011），Europeana以及和它相关的国家和地区图书馆已经同意采用关联数据的方式。大英图书馆已经开发了一个关联数据的模型。

（The British Library, 2011）相似的是，Europeana修订了其元数据的模型，从欧洲语义元素指南（Europeana Semantic Elements specification）变成了Europeana数据模型（Europeana Data Model, EDM），后者更加适用于关联数据。（Doerr et al., 2012; Haslhofer & Isaac, 2011）

Wilson（2010）指出，大英图书馆决定自由、免费的发布其书目数据，这个决策是和英国政府透明化和承担责任的承诺相一致的。作者还指出，向关联数据转移的部分可以归因于人们期待的增长，正如图书馆用户所表达的，他们想要看到图书馆随着最新的技术趋势与时俱进。大英图书馆的数据模型包括并且使用了多个当前基于URI的词汇表和本体，包括虚拟国际规范文档（Virtual International Authority File, VIAF）、国会图书馆主题标目（Library of Congress Subject Headings, LCSH）、Lexvo、Geonames、MARC国家和语种代码、Dewey.info、以及

RDF的图书混搭（图书和作者的信息）。（The British Library, 2011）

图书馆元数据缓慢的采用关联数据

W3C图书馆关联数据孵化小组（The W3C Library Linked Data Incubator Group, 2011）的报告认识到图书馆采用关联数据缓慢的速度，但是中报告也提供了图书馆拥抱关联数据原则的强有力的案例和建议。这个报告强调了让图书馆书目数据以“可共享、可扩展、并且可轻易的重新使用”的方式开放和免费获取的重要性。（W3C, 2011）它也指出：“图书馆的标准（MARC、Z39.50）只是为了图书馆社区而设计的”，这让外人很难重新使用这些数据，并且把图书馆的数据和其他数据进行重新的混合。这个报告和Coyle（2010）以及Styles等人（Styles, et al., 2008）进行的评估完全一致，它们都强调了当前图书馆元数据开发和使用上的挑战，并且号召通过拥抱以网络为中心的标准，更加强调可重用性、交叉关联和元数据的共享，来解决这些挑战。

Shadbolt和Hall（Shadbolt and Hall, 2006）评论说，这个吸收是“希望能够到达某个点上，让对数据的随机重用——不管是你自己这样做还是别人这样做——成为可能。”他们后悔人们还没有做很多事情来推动对于语义网范围更广的使用，这部分是因为人们对于“语言、形式主义、标准和语义”的强调。Weinberger（2012）相信，“最初的语义网强调构建本体，作为世界的‘知识呈现’，但是如果我们想要径直发布很多相互关联但是并不完美的数据，以标准化的形式让它们对所有人开放，那么整个互联网就会变成被有很多改进的知识基础设施。”

考虑到各个机构——比如IFLA、美国国会图书馆、OCLC以及许多其他的国家图书馆——已经投入的大量投资和努力，来开发并且维护MARC格式，向关联数据的转换并不是小事。图书馆有大量MARC格式的元数据记录，而且这些记录是有价值的，这件事是毫无疑问的。比如，从1950年代开始，大英图书馆就开始通过订阅获取，向外人提供其国家书目记录，最初使用的是打印的格式，但是后来随着图书馆自动化的推行，就开始使用MARC标准。（The British Library, 2011; Wilson 2010）

Wallis（2011b）观察到尽管现在只有很少的国家图书馆和地方文化遗产机构（比如CENL和Europeana）已经公开宣布会走上开放关联数据之路，重用其数据的基础的努力还是很多的。他认为，尽管把图书馆的遗产记录转化为RDA的工作是重要的，作为其结果的数据集采用的仍然是只属于图书馆领域的语言和术语，这“对于非图书馆员查找有用的数据，仍然会造成困难。”

在图书馆中采用关联数据的挑战

采用关联数据之路并非一路顺畅。尤其关于图书馆，目前有三个主要的挑战。首先，对于MARC标准大规模的使用，作为当前图书馆关系系统和元数据遗产的基础，这种实践已经行之有年。MARC格式，尽管占据了统治地位，却是一种以记录和文档为中心的元数据结构，而不是可操作的以数据为中心的格式。（Coyle, 2010; Coyle & Hillmann, 2007; Styles, 2009; Styles, et al., 2008）目前有几十亿条MARC格式的记录，即便是脱离MARC格式最小的努力，也会对资源产生巨大的影响。即便我们已经意识到MARC标准的局限性，从21世纪初开始，图书馆却也一直在沿用这个格式。（Tennant, 2002）当前的问题并不是图书馆和标准机构未曾意识到MARC的缺陷，而或许是，替代性的格式，比如XML，并不适合作为代替MARC格式的新标准。

第二个挑战，如W3C图书馆关联数据孵化小组曾经指出的那样（W3C, 2011），是图书馆和基

于网络的标准之间存在着术语上的差别。在这方面，Styles（2009）引用了使用FRBR模型可能会造成的歧义。他说：“没有人用‘作品’、‘内容表达’和‘载体表现’这些词汇（虽然这些都是FRBR模型中的核心概念），我们为什么要这么描述数据？”Wallis有着相似的见解，他（2011a）提出了一个更为简单的数据模型，更多地关注对象（单件）——比如一本具体的书，而不是其抽象概念，后者比如作品、内容表达和载体表现。上述两位作者都意识到，这些抽象并不是我们日常使用的概念，也并没有被出版商、编目员和图书馆使用者日常性的使用。Wallis特别推荐，图书馆和关联数据的社区应当携手合作，来弥补这样的差距，以便帮助其他领域的使用者来重用和扩展图书馆数据。Wallis认为，开发图书馆标准的新项目，比如RDA和FRBR，应当以简单为要，但也要充分利用元数据的丰富性——使用关联数据让后者成为了可能。

第三个重要的挑战是，关联数据潜在的采用者可能会面对关联数据技术的复杂性，比如RDF/XML、RDFS、OWL以及SPARQL。在图书馆中创建关联数据明显缺乏工具和应用。Berners-Lee评论道：“[当前的]网络正在增长，因为我们很容易就能写出一个网页，然后把它链接到其他网页。”（Berners-Lee, 2007）有鉴于此，上述技术必须是相对容易学习和使用的，就像在互联网的早期，创建HTML的网页一样容易。而现在，这些技术对于关联数据社区以外的人来说基本都是太复杂了。对于外部社区来说，任何人如果有基本的网页设计技能都应该能够根据关联数据技术来创建一个网页。最后但是并非最不重要的，当前人们采用关联数据原则和出版书目数据的努力关注的是以标准为中心的元数据遗产，但是似乎忽略了社会构建（用户驱动）的元数据方式。

实现概念转变的推荐

接下来，本文提出了6个推荐，帮助图书馆和标准机构来解决上述3个主要的挑战。这些推荐包括：

- 图书馆要应对关联数据提出的挑战；
- 采用关联数据原则的最小需求；
- 开发本体：实体（类别）、元素（属性）和值（个例）；
- 决定需要保留哪些需求；
- 成为关联数据网络的一部分；
- 采用混合的元数据方式；

图书馆需要应对关联数据提出的挑战

受到当前网络2.0和网络3.0范式的启发，并且为了得益于全球统一识别机制、以及关联数据原则带来的交叉链接和可重用性的特点，图书馆、标准机构以及和图书馆有关的机构——比如IFLA——应当应对当前技术趋势所带来的挑战，并且拥抱随之而来的机会。对于这点，Coyle建议说“我们必须应对的变化是网络在越来越多的成为搜索者和研究者的信息源，因而图书馆需要和数据的网络互联。”前文提到的W3C图书馆关联数据孵化小组（W3C, 2011）也意识到了我们需要从特定领域的图书馆标准和记录格式（比如MARC和Z39.50）转换到关联数据的标准和数据格式，比如RDF。图书馆的遗产元数据从OPAC的情境转移到网络的情境是一个巨大的挑战，因而不同的利益相关者——比如图书馆、档案馆、博物馆、出版商以及标准机构之间

的密切合作和协作——是非常重要的。

采用关联数据原则的最小需求

图书馆和标准机构需要采用关联数据原则的最小需求。因而，如果要把图书馆遗产元数据转化为关联数据的结构，使用RDF/XML序列化的格式是必须的。W3C图书馆关联数据小组的报告（W3C, 2011）指出，图书馆、标准机构和厂商将会得益于“扩展其范围，并且和关联数据标准化项目保持协调一致”（W3C, 2011）。使用URI、HTTP URI、RDF以及我们的数据集内部的链接将会让我们满足关联数据原则的最小需求。

开发本体：实体（类别）、元素（属性）和值（个例）

我们也需要实现概念的转换，让图书馆数据从文档或者记录，转化为Coyle（2010）所说的可行动的元数据，这个概念指的是机器可读、可混搭、以及可重新混合的元数据。为了实现这一点，我们必须提前开发元数据的模型（本体）。在这方面，FRBR和RDA应当扮演中心的角色，作为高等级的图书馆数据模型。

根据Horridge、Knublauch、Rector、Stevens以及Wroe的看法（2004），本体通过概念和关系，来捕捉特定领域内的知识。如Noy以及McGuinness（Noy and McGuinness, 2000）所描述的，本体帮助对于信息结构通用的理解在不同个人和软件代理之间共享；通过让领域的假设外显、把领域知识和操作性知识分离，来帮助对于领域知识的重用；并且帮助对领域知识的分析。根据Berners-Lee等人的看法（Berners-Lee, et al., 2001），本体这个词指的并不是“关于存在本质、以及哪些事物是存在的”，相反，这个概念是从虚拟智能社区开始使用它的时候建立的。因而，本体这个概念被定义为“正式定义概念之间的关系的文档或者文件。”（Berners-Lee, et al., 2001）根据Horridge等人（Horridge, et al., 2004）的看法：“本体用于捕捉特定兴趣领域内的知识”。但是本体定义中最广为引用的是Gruber（1993）所下的定义：“对于概念体系明确的详细说明。”（Gruber, 1993）

根据Noy以及McGuinness（Noy and McGuinness, 2000）的看法，开发本体需要我们定义类别、子类、属性以及个例（值）。我们强烈建议，图书馆元数据本体的开发应当利用语义网技术所使用的结构，来指定词汇表高等级的定义。而且，我们还建议本体的开发应当是中心化的（保证权威性），并且最好是由IFLA、RDA开发协调委员会（Joint Steering Committee for Development, JSC）、大英图书馆、国会图书馆以及OCLC这样的机构来进行的，尽管我们也应该包括个人和社区进行的去中心化、协作性的项目。

决定需要保留哪些需求

在图书馆模型和记录格式之间，我们需要区分的一点就是，图书馆能够从拥抱支持图书馆元数据的原则中获得巨大的好处，正如当前的以图书馆为中心的数据模型，比如RDA和FRBR。但是，图书馆也能够通过放弃无人使用的数据格式以及搜索和检索协议（比如MARC格式和Z39.50），来获得好处。有鉴于此，本文的作者认为号召放弃当前的元数据模型（比如RDA和FRBR）是有悖于语义网的特点的。而且，我们建议，IFLA这样的国际机构以及区域性的标准机构和图书馆，在总体上应该领先一步，开始协调性的项目，让当前的图书馆模型和标准兼容于关联数据的原则。为了从这个过程中获得优化的好处，图书馆应当通过受控词汇表，最大化的利用其知识和技能，比如LCSH、医学主题词表（Medical Subject Headings, MeSH）以及规范列表。

以上面的建议作为出发点，本文识别了两个先决条件，即对当前模型（比如RDA和FRBR）的

重新概念化，让它们和关联数据原则兼容的好处；以及图书馆应该把当前和MARC相兼容的元数据记录转换为RDF/XML的序列化格式，以此来逐渐放弃MARC及其衍生格式。换句话说，我们需要概念性的转变，从以记录为中心的元数据格式转变为基于RDF的数据格式。

成为关联数据网络的一部分

当我们根据开放关联数据的原则发布了元素集，这个过程就产生了一个全球相连的语义网，在其中，用户和应用能够发现、选择、改编、使用并且重用数据。这个数据云，就当前来说，看上去是很零碎的，因为数据分属于政府、企业和图书馆，它们被储存在孤井中，也被单独维护。（W3C, 2011）正如W3C（2011）报告所评论的：“今天关联数据云中的许多内容都是特别的、一次性的开放元素集转化为RDF的结果，它们并没有得到定期的准确性检查或者维护的更新。”这意味着，我们需要更多一致的行动，来开发并且维护一致并且可靠的命名空间、词汇表（本体）以及数据集。通过适当的采用关联数据的原则，图书馆不仅可能成为这个数据云的一部分，还能够成为其中最重要的参与者，因为它们掌握了大量的遗产书目数据和规范列表的数据，即便现在，这些数据的绝大多数都还锁在孤井和图书馆专用的格式中。保证数据云的一致和可靠的一种方法是重用现有的本体（词汇表）和URI。一些现有的命名空间包括RDFS、OWL、都柏林核心数据、国际虚拟规范文档、FOAF、Schema.org、大英图书馆术语（British Library Terms, BLT）、Lexvo、Geonames、MARC国家和语种代码、Dewey.info以及RDF图书混搭。

采用混合元数据的方式

如果我们同意这个看法，本体（分类法）和网络2.0的方式将会是相互补充，而不是相互对立的（Gruber, 2007; Morville, 2005; Wright, 2007），那么重要的是，任何元数据模型的重新概念化都将最大化的利用这些不同的范式。比如，在图书馆OPAC中嵌入协作和用户参与的内容将会让前者成为一个网络2.0的服务。与之相似，网络2.0和网络3.0的技术也是互补的，因而我们能够利用这些技术，通过从前者收割大量的数据，而在后者中获得结构（建立模型）和技术的能力，来满足用户的体验（Gruber, 2007, 2008）。Gruber（2007）说，社会构建元数据的方式是单向的，因而受困于不一致以及缺乏组织；而根据标准建立元数据的方式则无法呈现用户的词汇表。相似的是，Morville（2005）也指出，社会网络和与以往能够共存并且相互影响。在对这个看法的详细阐述中，他把Stewart Brand的《建筑是如何学习的》（How Buildings Learn. Brand, 1994; Brand, 1999）一书中提出的“速度分层”（Pace Layering）的理论加以情景化，他据此指出“分类法和本体提供了一个坚实的语义网络，这个网络把界面连接到下层结构”，而民俗分类法则被置于分类法元数据的基础结构之上，为后者提供了快速移动并且灵活的用户词汇表。（参见：Campbell & Fast, 2006; Smith, 2008）在范式转换的过程中，图书馆应当选择性的决定哪些原则、标准和格式应当被延续到未来的范式中，而上述哪些内容应该被抛弃。比如，从由图书馆员控制（根据标准的）向社会构建元数据的方式转变中，我们认为很重要的就是图书馆员降低新用户建立条目的门槛，以便让他们能够参与到创建元数据的工作中，这就意味着放弃死板的元数据质量的控制。

数字图书馆图书馆关联元数据的好处

在图书馆标准中采用关联数据的原则有很多好处，但是我们在下文中将只讨论5个最重要的。这些好处包括：

- 元数据的开放性和共享

- 帮助对信息资源的偶然的发展
- 关联元数据来发现资源使用的模式、时代精神以及新出现的元数据
- 基于面的导航
- 使用链接来丰富元数据

元数据的开放性和共享

网络2.0技术的核心竞争力之一就是其结构能够有助于参与（O'Reilly, 2005），这会降低贡献者参与的门槛。位于这种参与的结构中心的，是根据相互的信任，来拥抱共享和协作的开放性。（Alexander, 2006; Anderson, 2006, 2010; Shirky, 2005; Tapscott & Williams, 2010; Udell, 2004; Weinberger, 2005, 2007）网络2.0更多的是关于态度和文化，而不是技术（Miller, 2005）。换言之，网络2.0并不是一种新的发明，它的特征是参与和协作的文化被移植到网络1.0的技术中，让用户能够成为活跃的内容创建者和消费者。在其畅销书《维基经济学》（Wikinomics）中，Tapscott和Williams（Tapscott and Williams, 2010）断言：“因为技术、人口学、商业和世界的剧烈的变化，我们进入了新的时代，在这个时代中人们对经济的参与前所未有。”他们发现开放性是大规模协作之下隐藏的重要原则，其他的原则还包括同级关系（自组织的协作者同侪网络）、共享以及全球范围内的行动，其中的每个原则都让我们有更大的可能参与到越来越大的智能库中。他们还提到：“开放和真诚、透明、自由、灵活、可扩展性、参与和获取有关。”根据他们的看法，当前的经济、社会和技术的趋势表明，开放性和对知识产权的侵害无关。他们还进一步的表明，开放性的文化以及持续地意识到其潜在的好处调和了“人们传统的看法，这种看法认为公司应该通过把它们最隐秘的资源锁在暗箱里来和别人竞争。”（Tapscott & Williams, 2010）

Alexander（2006）断言，信息在生产者、消费者、不同的领域、服务器、机器的各种方向上流动，这让我们有必要打开信息的孤井，增强分享的服务。作者认为，即便是商业性的网站，比如Amazon.com，也允许其用户“从其列表中收割ISBN号”。他证明：“开放性是这个正在进行的运动中的标志，不管是在意识形态上，还是从技术的角度。”Miller（2005）把这个现象放到了图书馆的领域加以讨论，他认为网络2.0的原则让“数据自由”的观点成为了可能，而后者反过来又让数据能够以多种方式被“揭示、发现和操作”，因而创造出了难以想象的可能性来重新设定数据的目的，并且重用数据。

关联数据并不需要开放就能够使用（Cobden, Black, Gibbins, Carr, & Shadbolt, 2011; Shadbolt, 2010; W3C, 2011），但是就像Berners-Lee（2010）强调的，开放数据对社会有很多好处。在他的Ted演说中，Berners-Lee（2010）重新提到了他关联数据的愿景，并且引用了一组国际、地区和社区层面上的倡议计划和项目，这些项目采用了关联数据的原则，并且通过使用URI和RDF技术让它们的数据开放可得。正如作者所强调的，开放数据可以通过难以想象得多的方式被重用。比如，开放政府数据可以让纳税人查看选举出的政府是如何花掉他们的钱的，在哪里花掉，以及为什么花这些钱。Berners-Lee（2010）还强调，把数据进行开放的动力“才刚刚开始”。

开放关联数据的概念对于元数据在图书馆中被创建、获取、共享和重新组合的方式有着深远的影响。元数据可以“自由、免费的使用、重用和重新传播……通过使用全球范围内独一无二的标示符来命名作品、地点、个人、事件、主题以及其他感兴趣的对象或者概念，图书馆将会让资源能够在广泛的数据源中得到引用，这会让它们的元数据描述被更丰富的获取。”采用关联

数据的原则，并且让图书馆元数据能够得到重用，将会通过可靠的信息源来消除数据不必要的冗余，这种冗余如今已经无所不在。这让一些图书馆在它们的资源之间交叉链接，进而增进合作。最后，开放关联数据通过链接让图书馆元数据变得更加可见，而且还能让它们和非图书馆的信息源相连，比如Google、Wikipedia、LibraryThing、CiteULike以及Amazon。

帮助对信息资源的偶然的发展

很多重要的科学创新都来自于科学家的偶然发现。（Campa, 2008; Stoskopf, 2005; Zuckerman, 2010）这种意料之外的发展可能是科学家惊喜。正如Stoskopf（2005）提到的，这样的发现更多的发生于开放和保持好奇心的观察者，而不是把这样的线索看作是分散注意力的科学家。他断言：“基本的好奇心和观察是科学发现的必要先驱。我们应该意识到，偶然的发现对于科学的发展非常重要，而且经常是重要的智力跳跃发展的基础。”（Stoskopf, 2005）偶然性（serendipity）一词是英国的历史学家Horace Walpole在1754年首次创造出的（Merton & Barber, 2004）。在他写给他的朋友Horace Mann的46封信中，他写到：

“我想要尝试向你解释：偏差比起定义来，会让你获得更好的理解。有一次，我读到了一篇愚蠢的童话故事，故事的名字叫‘锡兰三王子’：他们三个人到处旅游，他们经常有一些发现，偶然的或者通过聪明才智观察到的，发现了之前他们未曾探究过的事情：比如，其中的一个人发现一只右眼失明的驴子最经常同一条路，因为路左边的草比右边的要少的很多。”（Merton & Barber, 2004）

偶然的信息发现在图书馆也和图书馆和信息科学的研究有关；尽管对于这个题目的研究很少。

（Foster & Ford, 2003）Foster和Ford（Foster & Ford, 2003）提出，当前对于用户信息搜索行为的模型（比如：Ingwersen, 1996; Wilson and Walsh, 1996; Wilson, 1997; Kuhlthau, 1993; Saracevic, 1996 & Spink, 1997）并没有考虑到偶然性的因素。他们报告说，最接近这个概念的研究，是Wilson和Walsh以及Wilson的研究，这两个研究把这个现象描述为“消极的注意力”以及“消极的搜索”。Foster和Ford（Foster & Ford, 2003）在对这个问题经过仔细的研究之后，他们得出结论说，对于信息的偶然获取加强了“研究者对于现有问题的认知和解决”，并且帮助研究者勾勒出探索和发现的新的领域。尽管它和完全理解用户的信息搜索行为有关，但是Foster和Ford（Foster & Ford, 2003）还是指出，对偶然性研究在本质上是存在问题的，问题一部分是来自于系统控制的困难，以及对它如何能增强用户体验的预测的困难。随着图书馆的服务进入网络，用户依靠搜索数据库来获取先定的信息源，这样用户会失去对于信息源的偶然的发现，后者是他们在浏览实体图书馆的书架的时候得到的。（Foster & Ford, 2003; Massis, 2011）我们需要指出的是，图书馆数据库具有先定的搜索界面，非常不可能有机会让我们得到偶然的搜索结果，因为这也并非它们设计时预想的方式。

实体图书馆数据的开放获取让人们偶然的发现一本书或者一份期刊成为可能。闭架阅览的图书馆就难以提供这样的偶然的发现。OPAC的设计就是为了回答对于资源先定的检索（比如，作者的姓名，或者题名），它和实体书架上的图书闭架阅览的实践很相似，二者都要依靠对于用户检索和馆藏中记录的匹配，来回复用户对于信息源的检索。于是这就引发了一个问题：“这种偶然性如何能够被整合到数字图书馆的设计中呢？”因为在数字图书馆中对于信息对象的获取依赖于对于元数据的使用，部分的解决方案就取决于对于元数据方式的选择。因而，重要的是探索社会构建元数据的方式如何能够和图书馆的元数据共同使用，来通过偶然的发现来改进用户的体验。相似情况是，关联数据对于促进偶然发现的需求，起到了怎样的作用。

通过发布关联数据，连接到许多不同的、但是相关的数据库和来源，并且让图书馆在这些来源中并置资源将会使图书馆受益。使用URI帮助从不同的来源中聚合这些资源。因而，让用户从一个来源无缝的访问另一个来源就在数字图书馆中重新引入了偶然获取，这和图书馆开架的方式是一样的。

关联数据发现资源的使用模式、时代精神以及新出现的元数据

一个元数据系统越多的反映对象名称系统的变异和组合（比如飓风、捷克共和国是之前的捷克斯洛伐克、或者俄罗斯是之前的苏联、或者Mac是Macintosh的缩写），那么它也就能越好的支持信息对象的可发现性。这是因为“信息职业工作者/图书馆员的词汇表可能和内容创建者和用户的语言不同，这隐藏了分类法应当揭示的信息。”（Barbosa, 2008）因为把元数据和分类法结合起来的首要目标就是保证其发现，因而元数据应当根据是否能够满足这个目标来得到评判，而不是看元数据是否根据客观和符合本体的方式来描述了信息对象。术语与时俱进，因而元数据应当能够意识到并且反映出这些变化。在卡片目录的时代，图书馆尽其所能来更新和维护其卡片目录的即时性。更新记录的意义在于，即便是新的术语和名称系统出现，而且图书馆员注意到了，必要的更新还是会晚很多才会发生。这个问题并不只有图书馆员会面对，它也发生在印刷的百科全书修正事实错误这件事情上。

混合元数据方式（根据标准的元数据以及社会构建的元数据）、以及额外得到关联数据（URI、RDF、OWL、SPARQL）的语义和技术能力支持的数字图书馆系统能够提供更丰富和更完整的对于信息对象的元数据描述，更精确的反映出用户所采用的各种解释和术语。而且，关联数据还能够帮助对于资源使用模式的分析（比如点击量、下载数、经常使用的标签（标签云）、平均打分以及得到推荐的资源）。比如，LibraryThing.com现在就提供了时代精神这样的功能，还能够通过平均打分、标签云和评论最多的图书来分析哪些书最受欢迎。

关联数据作为一个数据模型和一组技术框架，它不能支持自身，而且如果没有底层的数据，它并没有什么意义甚至是完全没有意义。因而，整合图书馆员创建的图书馆遗产数据以及用户创建的元数据对于用户来说是至关重要的，而在其中，二者同等重要。整合后者和用户尤其相关，尤其是我们需要发现使用模式和时代精神。Weinberger（2005）提出：“当作者决定她的作品要写什么的时候，她是一个权威，但是她没办法决定这个作品对别人意味着什么。在搜索中，作品对于搜索者意味着什么远比作者的意图要重要得多。”关联数据提供的建模和技术的优势将能够让发现用户所使用的资源使用模式、时代精神、以及新出现了什么词汇表这样的活动取得极大的成功，这能够帮助图书馆在采购资源的时候（购买图书、订阅期刊）做出符合战略的决定，而这将再一次证明图书馆的投资回报率。

基于面的导航

分面浏览指的是在信息系统导航中使用不同的维度和属性，以增加可发现性。（Morville, 2005）根据Morville（2005）的看法，在图书馆中使用面的想法得到了阮刚那赞（S.R. Ranganathan）的推动而被人关注。使用面和使用分类法不同，在后者，对象在一个单一、有限的导航方式中只处于一个单一的位置。Weinberger（2007）表达了对Ted Nelson的“交织性”（intertwingularity）这个概念的赞同，他反对追求单一、先组以及完美的导航结构的观念。他写到：“人们一直在假装他们好像能把事情变得高度的等级化、可分类并且有秩序，但实际上他们不能。万事万物都是交织的。”“交织性”这个词是由Ted Nelson发明的（他也发明了另一个概念“超文本”（hypertext）），这个词指的是观念（概念）之间的相互关联和相互依赖，他以此表明，任何试图分类的努力都是虚假的（关于Ted Nelson的“交织性”的概念

更多的信息，请参见：Weinberger, 2007; Wright, 2007）。Weinberger认为，数字信息对象可以被挑选、分类，并且根据用户的需求和情境，根据不同的方式被呈现出来。（Weinberger, 2007）下面引用他（Weinberger, 2007）的一大段话：

“在第三种秩序中[在全文的数字图书馆中]，一片叶子[信息对象]可以被挂在很多树枝上，它可以为了不同的人被挂在不同的树枝上，而且一个人也可以让它改变所挂的树枝，如果她决定用不同的方法来对待这个对象。这并不是说，我们知识的世界是一棵树的形状，或者正在变成一些难以想象的四维的树。在第三种秩序中，知识没有形状。有太多有用、强大而且美丽的方法来给我们的世界赋予意义。”

在他的《无法全知》（Too big to know. 2012）一书中，Weinberger提出：“对于信息过载问题的解决方案是创造更多的元数据”，他也提到了根据不同的面过滤相关结果的重要性。在《简单法则》（The Laws of Simplicity）一书中，Maeda（2006）提出：“简单性是减去了明显的事情，增加了有意义的事情。”他写到：“如果把更多拿远了，那么更多看上去也就变少了。因而，如果我们让结果贴近我们、让实际的工作离我们更远，那么体验也就变得更加简单。”

作为原则，数字图书馆并不受到传统图书馆实体书架带来的束缚，因而能够让用户通过多种方式，使用关键词和类别作为面，来过滤信息，发现信息对象。通过关联数据链接结构的力量以及元数据的丰富性——通过同时采用根据标准的元数据方式以及社会构建的元数据方式来获得，我们可以使用多种面来过滤呈现给用户的内容。这样元数据的丰富性也能够利用用户的档案，好让数字图书馆系统能够对搜索结果进行情境化和定制化。当前，Last.fm、LibraryThing以及Amazon.com以及其他服务都在利用分面导航和元数据的过滤。

通过链接丰富元数据

目前，OPAC中指向元数据的链接只能指向特定的点，而且很快就会走入死胡同，如果某个数据元素没有被进一步链接的话。比如，当前的元数据界面允许用户搜索一本特定的书，他们点击作者的姓名来看到他的出版物，或者就是找到一本相关/相似的书。OPAC通常都并不提供指向作者传记页面的无缝链接，或者从这个页面再到另外一个页面，比如Wikipedia、Google图书，然后再从这些页面链接回图书馆的列表。在关联数据环境中被丰富描述的图书馆可以被连接到任何有关的内容，用户来决定是否要停止导航，如果他们的信息需求业已得到满足，或者是由应用来提供情境或相关的信息。换言之，每个元数据都能通过链接得到丰富，除非这个元数据是文字的类型——比如一个人的年龄，因为这种元数据不需要被链接到另一个页面上。

通过情境化和相关链接增强的元数据能够让用户在离散的图书馆数据库和外部信息提供者（比如其他图书馆和搜索引擎）之间得到无缝的导航。通过全球独一无二的方式识别实体（比如作品、个人、地点和事件）、元数据的元素或者属性（作者、题名、主题、关系）以及相应的值（个例），关联数据提供了多种方式通过元数据来丰富信息对象，这会有助于信息的搜索，改进用户在信息图书馆中的体验。

结论

当前图书馆的标准和模型受制于传统卡片目录的一些本质的限制。使用这些模式生产的元数据只要是用于人类的消费而不是机器的操作。关联数据被认为能够生产以数据为中心、机器可操作的元数据。但是，在实际图书馆元数据生产和利用的工作中利用关联数据，这件事才刚刚开始。

我们建议，对当前的图书馆模型，比如RDA和FRBR进行重新概念化，好让它们兼容于关联数据的原则；图书馆应当把现在和MARC兼容的元数据记录转换为RDF/XML的序列化格式，以便逐渐放弃MARC及其衍生的格式。有鉴于此，我们认为，放弃当前的元数据模型（比如RDA和FRBR）的主张是违背关联数据的原则的，因为它允许单独的标准和受控词汇表“欢快的”在同一个领域中同时存在，只要这些资源（对象）使用了全球独一无二的统一资源标志符进行标记，而且采用了高等级的数据模型，比如RDF。和它有关的，图书馆已经掌握了一个世纪之久的开发、维护、保存受控词汇表（比如规范名称和主题标目）的最佳实践和经验教训，对于用户需求的经验，也可以在每个采用了关联数据原则的项目中得到利用。我们认为，在图书馆元数据中适当的采用关联数据将会帮助图书馆开发图书馆的应用和服务，比如分面浏览、偶然性的浏览、元数据的时代精神的数据以及对于新趋势和用户词汇表的发现。

参考文献

- Alemu, G., Stevens, B., & Ross, P. (2012). Towards a Conceptual Framework for User-Driven Semantic Metadata Interoperability in Digital Libraries: A Social Constructivist Approach. *New Library World*, 113(1/2), 38-54.
- Alexander, B. (2006). Web 2.0: A New Wave of Innovation for Teaching and Learning? *EDUCAUSE Review*, 42(2), 32-44.
- Allemnag, D., & Hendler, J. (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Amsterdam: Morgan Kaufmann.
- Anderson, C. (2006). *The Long Tail: How Endless Choice is Creating Unlimited Demand*. London: Random House Business Books.
- Anderson, C. (2010). *Free : How Today's Smartest Businesses Profit by Giving Something for Nothing*. London: Random House Business Books.
- Barbosa, D. (2008). *Taxonomy Folksonomy Cookbook*. Retrieved from http://solutions.dowjones.com/cookbook/ebook_sla2008/cookbookebook.pdf
- Berners-Lee, T. (1997). *Metadata Architecture*. Retrieved from <http://www.w3.org/DesignIssues/Metadata.html>
- Berners-Lee, T. (1998). *What the Semantic Web can Represent*. Retrieved from <http://www.w3.org/DesignIssues/RDFnot.html>
- Berners-Lee, T. (2007). *Testimony of Sir Timothy Berners-Lee Before the United States House of Representatives Committee on Energy and Commerce Subcommittee on Telecommunications and the Internet*. Retrieved from <http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html>
- Berners-Lee, T. (2009). *Linked Data*. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. (2010). *Tim Berners-Lee: The Year Open Data Went Worldwide*. from http://www.ted.com/talks/tim_berniers_lee_the_year_open_data_went_worldwide.html
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The Semantic Web: A New form of Web Content that is Meaningful to Computers will Unleash a Revolution of New Possibilities*. *The Scientific American*, 284(5), 34-43

- Brand, S. (1994). *How Buildings Learn: What Happens after They're Built*. London: Viking.
- Brand, S. (1999). *The Clock of the Long Now: Time and Responsibility* (1st ed.). New York: Basic Books.
- Campa, R. (2008). Making Science by Serendipity. A review of Robert K. Merton and Elinor Barber's *the Travels and Adventures of Serendipity*. *Journal of Evolution and Technology* 17(1), 75-83.
- Campbell, D. G., & Fast, K. V. (2006). From Pace Layering to Resilience Theory: The Complex Implications of Tagging for Information Architecture. Paper presented at the The 7th Information Architecture Summit, Vancouver, British Columbia, Canada.
- Cobden, M., Black, J., Gibbins, N., Carr, L., & Shadbolt, N. (2011). A Research Agenda for Linked Closed Data. Paper presented at the Second International Workshop on Consuming Linked Data. Retrieved from <http://eprints.ecs.soton.ac.uk/22711/3/position.pdf>
- Coyle, K. (2010). Library Data in a Modern Context. *Library Technology Reports* 46(1), 5-13.
- Coyle, K., & Hillmann, D. (2007). Resource Description and Access (RDA): Cataloging Rules for the 20th Century. *D-Lib Magazine*, 13(1/2).
- Day, M. (2000). Resource Discovery, Interoperability and Digital Preservation: Some Aspects of Current Metadata Research and Development. *VINE*, 36(117), 35-48.
- Day, M. (2003a). Integrating Metadata Schema Registries with Digital Preservation Systems to Support Interoperability: A Proposal. Paper presented at the International Conference on Dublin Core and Metadata Applications, Seattle.
- Day, M. (2003b). Preservation Metadata Initiatives: Practicality, Sustainability, and Interoperability. Paper presented at the ERPANET Training Seminar on Metadata in Digital Preservation. Retrieved from <http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf>
- Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., et al. (2000). The Semantic Web: The Roles of XML and RDF. *IEEE Internet Computing*, 15(3), 63-74.
- Denton, W. (2007). FRBR and the History of Cataloging. In A. G. Taylor (Ed.), *Understanding FRBR: What it is and How it Will affect Our Retrieval* (pp. 35-57). Westport, Connecticut: Libraries Unlimited.
- Doerr, M., Gradmann, S., Henniecke, S., Isaac, A., Meghini, C., & Van de Sompel, H. (2012). The Europeana Data Model (EDM). Paper presented at the World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden.
- Dunsire, G. (2009). UNIMARC, RDA and the Semantic Web Paper presented at the World Library and Information Congress: 75th IFLA General Conference and Council Retrieved from <http://www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf>
- Evans, W. (2009). *Building Library 3.0: Issues in Creating A Culture Of Participation*. Oxford: Chandos.
- Foster, A., & Ford, N. (2003). Serendipity and Information Seeking: an Empirical Study. *Journal of Documentation*, 59(3), 321-340.
- Gruber, T. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies*, 43, 907-928.

- Gruber, T. (2007). Ontology of Folsonomy: A Mash-up of Apples and Oranges. *International Journal on Semantic Web & Information Systems*, 3(2).
- Gruber, T. (2008). Collective Knowledge Systems: Where the Social Web Meets the Semantic Web. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1), 4-13.
- Guenther, R., & McCallum, S. (2003). New metadata standards for digital resources: MODS and METS. *Bulletin of the American Society for Information Science and Technology*, 29(2).
- Haslhofer, B., & Isaac, A. (2011). data.europeana.eu - The Europeana Linked Open Data Pilot. Paper presented at the DC-2012 International Conference on Dublin Core and Metadata Applications, The Hague, The Netherlands.
- Helen, W. (2010). Linked Data and Libraries. *Catalogue & Index*(160), 2-5.
- Horridge, M., Knublauch, H., Rector, A., Stevens, R., & Wroe, C. (2004). *A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools Edition 1.0*
Available from <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>
- IFLA. (2009). *Functional Requirements for Bibliographic Records: Final Report: International Federation of Library Associations and Institutions*.
- Lagoze, C. (2010). *Lost Identity: the Assimilation of Digital Libraries into the Web*. Cornell University, School of Information Science.
- Library of Congress. (2011). *A Bibliographic Framework for the Digital Age*. Washington, D.C.: Library of Congress.
- Lubetzky, S. (1953). Development of Cataloging Rules *Library Trends* 2(2), 179-186.
- Marcum, D. (2011). *A Bibliographic Framework for the Digital Age*. Washington, D.C.: Library of Congress.
- Massis, B. E. (2011). "Serendipitous" browsing versus library space. *New Library World*, 112(3/4), 178-182.
- Mathes, A. (2004). *Folksonomies - Cooperative Classification and Communication Through Shared Metadata* Retrieved from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Merton, R. K., & Barber, E. (2004). *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*. Princeton: Princeton University Press.
- Miller, P. (2005). *Web 2.0: Building the New Library Ariadne*, 45.
- Morville, P. (2005). *Ambient findability*. Sebastopol, CA: O'Reilly.
- Nilsson, M. (2010). *From Interoperability to Harmonization in Metadata Standardization: Designing an Evolvable Framework for Metadata Harmonization*. Unpublished PhD, School of Computer Science and Communication, KTH, Stockholm.
- Noy, N. F., & McGuinness, D. L. (2000). *Ontology Development 101: A Guide to Creating Your First Ontology*. Retrieved from http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

- O'Reilly, T. (2005). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Retrieved from <http://oreilly.com/web2/archive/what-is-web-20.html>
- Rothenberg, J. (2008). Interoperability as a Semantic Cross-cutting Concern Interoperabiliteit: Eerlijk zullen we alles delen. Den Haag.
- Shadbolt, N. (2010). data.gov.uk - The Linked Data Revolution. Paper presented at the Innovating Through Information Lecture Series. Retrieved from <http://eprints.ecs.soton.ac.uk/18787/>
- Shirky, C. (2005). Ontology is Overrated: Categories, Links, and Tags. Clay Shirky's Writings About the Internet. Retrieved from http://www.shirky.com/writings/ontology_overrated.html
- Smith, G. (2008). Tagging : People-Powered Metadata for the Social Web. Berkeley, CA: New Riders.
- Stoskopf, M. K. (2005). Observation and cogitation: how serendipity provides the building blocks of scientific discovery. *Institute for Laboratory Animal Research*, 46(4), 332-337.
- Styles, R. (2009). Bringing FRBR Down to Earth. <http://dynamicorange.com/2009/11/11/bringing-frbr-down-to-earth/>
- Styles, R., Ayers, D., & Shabir, N. (2008). Semantic MARC, MARC21 and the Semantic Web. Paper presented at the Linked Data on the Web (LDOW2008). Retrieved from <http://events.linkeddata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf>
- Tapscott, D., & Williams, A. D. (2010). *Wikinomics: How Mass Collaboration Changes Everything* (Expanded ed.). London: Penguin Books.
- Tennant, R. (2002). MARC must die. *Library Journal*. Retrieved from <http://www.libraryjournal.com/article/CA250046.html>
- The British Library. (2011). Free Data Services. Retrieved December 14, 2011, from <http://www.bl.uk/bibliographic/datafree.html>
- Udell, J. (2004). Collaborative Knowledge Gardening with Flickr and del.icio.us, Social Networking goes Beyond Sharing Contacts and Connections. *InfoWorld*. Retrieved from <http://www.infoworld.com/d/developer-world/collaborative-knowledge-gardening-020>
- Veltman, K. H. (2001). Syntactic and Semantic Interoperability: New Approaches to Knowledge and the Semantic Web. *New Review of Information Networking*, 7(1), 159-183.
- W3C. (2004a). OWL Web Ontology Language Overview. from <http://www.w3.org/TR/owl-features/>
- W3C. (2004b). RDF Primer. W3C Recommendation. Retrieved from <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/#basicconcepts>
- W3C. (2011). Library Linked Data Incubator Group Final Report: W3C.
- Wallis, R. (2011a). Library of Congress to Boldly Voyage to Linked Data Worlds. <http://consulting.talis.com/2011/11/library-of-congress-to-boldly-voyage-to-linked-data-worlds/>
- Wallis, R. (2011b). Will Europe's National Libraries Open Data In An Open Way? Retrieved from <http://dataliberate.com/2011/09/will-europes-national-libraries-open-data-in-an-open-way/>
- Weinberger, D. (2005). Tagging and Why it Matters. Retrieved from <http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/07-WhyTaggingMatters.pdf>

Weinberger, D. (2007). *Everything is Miscellaneous*. New York: Times books.

Weinberger, D. (2012). *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. New York: Basic Books

Wilson, N. (2010). *Linked Data Prototyping at the British Library*. Paper presented at the Talis Linked Data and Libraries event. Retrieved from <http://blogs.talis.com/nodalities/neil-wilson-lightning-talk-%E2%80%93-linked-data-and-libraries-2010>

Wright, A. (2007). *Glut: Mastering Information Through the Ages*. Ithaca: Cornell University Press.

Zuckerman, E. (2010). *Desperately Seeking Serendipity*. Retrieved from <http://www.chi2011.org/program/plenaries.html>

About the authors

Getaneh Alemu is a PhD student in the School of Creative Technologies, and Dr Brett Stevens (Principal Lecturer, School of Creative Technologies), Dr Penny Ross (Senior Lecturer, School of Computing) and Jane Chandler (Principal Lecturer, Faculty of Creative and Cultural Industries), all at the University of Portsmouth, Portsmouth, UK. Getaneh Alemu is the corresponding author and can be contacted at: Getaneh.Alemu@port.ac.uk