

# 从关系数据库到关联数据：W3C标准应用探析\*

夏翠娟 金家琴(上海图书馆)

**摘要** 针对RDB2RDF两种标准规范Direct Mapping和R2RML在实际应用中的作用和用法，基于文献调研的基础上，分析了几大关系数据库和RDB2RDF工具在支持RDB2RDF映射方面的功能原理和实现方式，并对各个产品是否支持两种标准规范得出了研究结论。以MusicBrainz关联数据项目为案例，详细介绍了R2RML在实际项目中的使用方法和实施方案。

**关键词** RDB2RDF, Direct mapping, R2RML, 关联数据

DOI: 10.13663/j.cnki.lj.2015.05.014

## On the Application of W3C's RDB2RDF Standards

Xia Cuijuan, Jin Jiaqin (Shanghai Library)

**Abstract** As two standards of W3C's RDB2RDF Working Group, DM and R2RML have been implemented by many RDB2RDF tools, in the latest version of Oracle, which is one of the most famous RDB providers, and by Music Brainz Linked Data project and Open Semantic Cloud for Brussels Project. This paper analyzes the RDB2RDF tools, the RDB products and the projects implemented with the two standards.

**Key words** RDB2RDF, Direct mapping, R2RML, Linked data

### 0 引言

RDF (Resource Description Framework, 资源描述框架) 是 W3C 为促进语义万维网 (Semantic Web, 简称语义网) 的应用而推出的一系列标准规范, 包括 RDF 抽象模型和一组 RDF 编码格式规范如 Turtle、N-Triples、JSON、N-Quod 等<sup>[1]</sup>。RDF 有利于数据的共享、重用和语义互操作, 越来越多的系统和应用开始利用 RDF 改造底层数据的数据结构。而大量遗留系统中的数据存储存储在 RDB (Relational Database, 关系数据库) 中, 如何方便、快捷且准确地实现“关系数据库中的数据 (RDB 格式) 转换成 RDF 数据格式 (这个过程简称为 RDB2RDF)”, 是实际应用中经常被关注的问题。W3C 因而成立了 RDB2RDF 工作小组, 制定了 DM (Direct Mapping, 直接映射) 和 R2RML (RDB2RDF Mapping Language) 这两个标准, 来规范 RDB2RDF 的实现, 已于 2012 年 6 月成为 W3C 的推荐标准。两年来, 这两个标准已在一些关

系数据库产品和许多 RDB2RDF 工具中得到应用, 也有了一些实际应用案例。本文从关系数据库、RDB2RDF 工具、实际应用案例这三个方面对 DM 和 R2RML 这两个标准规范的应用进行调研, 并对各自的应用效果进行比较和分析, 为基于关系数据库的语义应用提供参考。

### 1 DM 和 R2RML: 作用与用法

W3C 的 RDB2RDF 工作组推荐了两种 RDB2RDF 映射语言 DM 和 R2RML, 用于定义关系数据库中的数据如何转换为 RDF 数据的各种规则, 包括 URI 的生成、RDF 类和属性的定义、空节点的处理、数据间关联关系的表达等。<sup>[2]</sup>

\* 本文系国家社科基金青年项目“W3C 的 RDB2RDF 标准规范在关联数据服务构建中的应用”(项目编号:13CTQ008) 和上海市哲学社会科学项目“基于书目控制的网络信息资源的规范控制方法”(批准号:2009JG502-BTQ040) 的研究成果之一。

DM是直接映射的方式,它将关系数据库表结构和数据直接输出为RDF图(RDF Graph, 可看做是多个三元组的集合),该RDF图完全是关系数据库数据结构的反映,关系数据库中的一个表转换为一个RDF类(Class),一个字段转换为一个RDF属性(Property),用于表示类和属性的术语与关系数据库中的表名和字段名保持一致。DM映射一般由程序自动生成。<sup>[3]</sup>

与DM不同的是,R2RML有高度的可定制性和灵活性,R2RML为RDB2RDF映射定义了系统性的逻辑框架,提出“逻辑表”(Logical Table)的概念,将关系数据库中的一个表、一个视图,甚至是一个有效的SQL查询定义为“逻辑表”,这就突破了关系数据库表的物理结构的限制,在生成RDF数据之前,就可以对RDB中的数据进行计算处理、筛选、清洗和整合,为不改变数据库原有的结构而灵活地按需生成RDF数据奠定了基础。R2RML映射一般作为一个可人工编辑的文本文件存在,以Turtle句法和格式编码,可将关系数据库的数据结构与已有的本体词表映射。通过R2RML映射,一个关系数据库可输出为一个RDF图,该RDF图中所用到的类名和属性名可来自已有的本体词表,如FOAF、DC、SKOS等,可灵活地根据具体需求选取现有本体词表中的术语。如需改变映射规则,R2RML文件可方便地由人工编辑。<sup>[4]</sup>

RDB2RDF映射在实际应用系统中的实现一般有两种模式,Juan F. Sequeda在2013年的国际语义网大会(ISWC2013)的培训<sup>[5]</sup>中介绍了这两种模式,如图1所示。在有的应用系统架构设计中,需要将RDB中转换而来的RDF数据集导出到本地后,再导入专业的RDF存储库(RDF Store,也叫TripleStore),这种方式即也叫“抽取—转换—装载”(Extract-Transform-Load, ETL),简称为ETL模式。由于原有系统的数据仍在不断更新,这种方式往往无法实时提供最新的数据。在另一种应用系统架构设计中,只需提供一个虚拟的RDF数据视图和SPARQL查询接口,并返回RDF数据查询结果,即在原有服务器上增加一个RDF数据封装层(即图中的Wrapper System);而在数据

查询时,大多会经历将前台SPARQL查询请求根据事先定义好的映射规则转换成SQL查询语言,再将查询结果转换为RDF数据的过程,本文称为“实时转换”模式,这种模式往往会对系统的性能提出更多的考验。

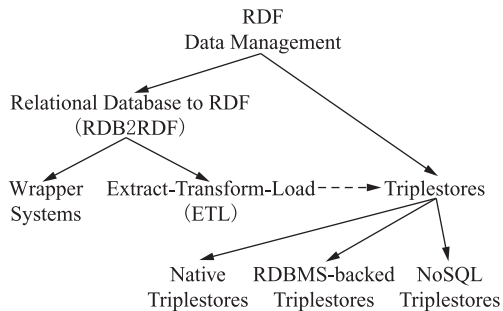


图1 RDB2RDF的两种实现模式<sup>[5]</sup>

以上两种模式各有利弊,视具体需求而定。由于在实际应用中,如何设计RDB2RDF的实现方案、如何选择RDF数据存储和管理工具、如何提供数据消费服务,与RDB2RDF的实现模式有较为密切的关系,因而在考察几种关系数据库产品和各种RDB2RDF工具平台对DM和R2RML实现的方式时,对这两种模式的支持将作为一个重要考察点。

## 2 在关系数据库中的应用

关系数据库作为数据的存储和管理系统,其成熟稳定的功能和性能,仍被大量应用系统采用。虽然它并不是最适合RDF数据的数据管理系统,但仍有大量的数据存于其中,对于那些仍在生长和使用中的应用系统来说,最好能在不改变原有系统业务流程和数据结构的情况下,提供RDF数据和基于RDF数据的语义应用,这就要求从原关系数据库中适时生成或导出RDF数据,会涉及RDB数据格式到RDF数据格式的转换,即RDB2RDF。对于大多数关系数据库,均可通过第三方插件或工具实现RDF数据的管理和RDB2RDF的转换,本小节主要调研Oracle、IBM DB2、SQL Server、MySQL、PostgreSQL目前流行的几个关系数据库本身内置的对RDF数据管理的支持,尤其是对RDB2RDF的支持,以及是否支持W3C的两个推荐标准DM和R2RML,

而通过第三方插件或工具实现 RDF 数据管理和 RDB2RDF 将在第 4 节中详细分析。

本小节的调研分三个层次：(1) 数据库产品本身是否支持 RDF 数据的存储和 SPARQL 检索和更新？(2) 数据库产品本身是否支持以 RDB2RDF 的方式生成或导出 RDF 数据？(3) 如果支持 RDB2RDF，是否支持 W3C 的 DM 或 R2RML 标准？

### 2.1 Oracle

在几个商用数据库产品中，Oracle 对 RDF 的支持最为突出，目前 Oracle 有两种方案来存储和检索 RDF 数据，一种基于 RDB，另一种基于 NoSQL。前者的代表产品是 Oracle Spatial and Graph，是本文调研的主要对象。该产品除了支持地理空间数据外，还支持面向语义网和关联数据应用的 RDF 数据存储管理和 SPARQL 检索，支持原生 RDF 数据的创建、已有 RDF 数据的导入，也支持 RDB2RDF 方式，即存储在关系数据库表中已有的数据，根据事先定义的映射规则，实时转换为 RDF 数据，在转换规则的定义上，支持 W3C 的 DM 和 R2RML 标准。

根据 Oracle 技术文档<sup>[6]</sup>第 10 节中的说明，Oracle 提供一组 API——SEM\_APIS 来支持语义网应用，其中 SEM\_APIS.CREATE\_RDFVIEW\_MODEL 这个 API 可以支持 RDB2RDF 的实现，该 API 可根据默认的映射规则 DM，或根据用 R2RML 自定义的映射规则，生成虚拟的 RDF 视图并通过一个名为 SEM\_MATCH 的 API 对 RDF 视图执行 SPARQL 查询。当使用 DM 时，只需要指定需要转换成 RDF 数据的表名以及 RDF 图的前缀和命名空间，即可根据默认的 DM 规则生成 RDF 视图。当使用 R2RML 时，需定义一个基于 R2RML 语法的映射文件，一般用 Turtle 格式编写，SEM\_APIS.CREATE\_RDFVIEW\_MODEL 这个 API 首先将该文件转换为 N3 格式并存储在数据库的一张表中，然后读取映射规则生成 RDF 视图。

这些虚拟的 RDF 视图并不存储真实的 RDF 数据，如果需要物理地存储 RDF 视图中的 RDF 数据，可以通过 API 从 RDF 视图中导出 RDF 数据并存储在 Oracle 的数据库表中。SEM\_MATCH 支持对 RDF 视图中的 RDF 数据

进行 SPARQL 查询，还支持将 RDF 视图中的数据与原生的 RDF 数据源整合在一次 SPARQL 查询中，进行联邦 SPARQL 查询。

### 2.2 IBM DB2

IBM 于 2012 年宣布在 DB2 第 10 版的更新 (DB2 LUW 10.1) 中嵌入支持 RDF 存储和检索的功能，基于 DB2 的 RDF Store (下文简称 DB2-RDF Store) 被 IBM 官方称为 IBM DB2 NoSQL Graph Store，虽然被称为 NoSQL，但其底层仍然是 DB2 的关系数据库结构，它用几个特殊的表来存储 RDF 数据，并通过一些内置的 RDF 命令和 API 来实现 RDF 数据的支持，这些 API 是对 Apache Jena API 的扩展。DB2-RDF Store 分为两种：一种是基本型，适用于从零开始逐步创建 RDF 数据，另一种是优化型，适用于导入已有的 RDF 数据。创建、编辑或维护 RDF Store 需要用到 DB2 的 RDF 命令或 API，主要利用 Jena API 来实现对 RDF 数据的更新和 SPARQL 数据查询。在 DB2 Version 10.1 Fix Pack 2 之后的版本中，除了支持 SPARQL 查询外，已可以支持 SPARQL 1.1 UPDATE。但未有资料显示 DB2 支持 RDB2RDF 的方式，当然也就不支持 DM 和 R2RML 这两个标准。<sup>[7]</sup>

### 2.3 MySQL 和 PostgreSQL

通过使用 RDFLib 插件，MySQL 和 PostgreSQL 数据库支持 RDF 的存储和检索功能。RDFLib 存储接口提供了一个 MySQLMassLoader 模块实现 RDF 数据的导入和存储，对 RDF 数据的检索和更新是由 RDFLib-SPARQL 处理器来完成的。RDFLib-SPARQL 处理器的原理是在后台将 SPARQL 查询语句翻译成相同的 SQL 查询后再执行，可以支持 SPARQL 1.1 检索和更新。MySQL 和 PostgreSQL 目前仅支持通过第三方插件或者 API 来实现 RDF 数据的存储、检索和更新，不支持 DM 和 R2RML 这两个标准。

### 2.4 Microsoft SQLServer

Microsoft SQLServer 自嵌的配置文件管理器 (Profile Manager) 组件可以支持把 RDF 格式的配置信息存储到 SQL Server 数据库中。配置文件管理器只能存储特定的用户、服务或应用程序所需的属性。在把配置文件信息存储到数据

库之前,连接服务框架(CSF)配置文件管理器会解析传入的数据,并转换成RDF三元组。配置文件管理器支持使用简单协议和SPARQL查询来检索。通过使用分面功能,配置文件管理器可优化检索性能。

Semantics.Server 2.0是基于Microsoft SQL Server存储RDF的解决方案,它利用SQL Server的搜索引擎和内存管理来支持RDF数据存储和SPARQL检索。目前不支持RDB2RDF的方式及DM和R2RML这两个标准。

综上所述,关于第一个问题:Oracle、SQL Server、IBM DB2这三种商业数据库产品,均已公开发布了支持RDF数据管理的产品及相应的技术文档。而两个开源的数据库MySQL和PostgreSQL可通过第三方API支持RDF数据的存储和管理。关于第二个问题:只有甲骨文公司的产品Oracle Spatial and Graph可以RDB2RDF的方式生成和导出RDF数据。关于第三个问题:只有甲骨文公司的产品Oracle Spatial and Graph在RDB2RDF的过程中支持W3C的DM和R2RML标准。见表1。

### 3 在RDB2RDF工具中的应用

除了数据库本身内置了支持RDB2RDF的功能外,利用第三方RDB2RDF工具是一个更为灵活的选择。在W3C的RDB2RDF孵化小组的调研报告<sup>[9]</sup>中,分析了15种工具,这种工具较多,大部分为开源,且每种常见的数据库都能找到相应的工具。在DM和R2RML成为W3C的推荐标准后,RDB2RDF工作小组发布了实施报告<sup>[9]</sup>,列举了参与过用例测试的8种

工具,也有一些支持这两种标准的RDB2RDF工具没有参与过测试,如R2RML Parser。其中有的同时支持DM和R2RML两种标准,有的只支持其中一种,大部分都支持实时转换模式,支持ETL模式的有5种。除ultrawrap是capsenta公司开发的商用产品外,其他都是开源的。各种工具的总体情况见下表2。

以下选取几个典型的RDB2RDF工具作简要介绍和分析。

D2RQ是DM和R2RML出现之前,最为人们熟知的RDB2RDF开源工具平台,由著名的语义技术研究开发机构DERI和惠普实验室等机构开发维护。它包括D2R Server和一种私有的RDB2RDF映射语言D2Rq。D2R Server是一个HTTP Server,用于接收系统前端的SPARQL请求并作出响应。D2Rq提供一种可定制的映射语言,将RDB数据映射成RDF数据模型。D2R Server的SPARQL接口基于SPARQL协议,从前端传过来的SPARQL查询请求被D2R Server转换成关系数据库的SQL查询请求,返回的结果被D2R Server转换成RDF数据。它并没有将RDB发布成真实的RDF数据,而是使用D2RQ映射文件将其映射成虚拟的RDF视图。在2012年发布的v0.8.1版本中,开始支持DM,但仍不支持R2RML。除了实时生成RDF之外,它有一个dump-rdf插件,可以根据默认或定制的映射将整个RDB数据库导出为一个单独的RDF文件,因而也支持ETL模式。<sup>[10]</sup>

Virtuoso利用自有的“元数据方案映射语言”,来实现RDB数据与RDF数据的映射,生成关联数据视图(Linked Data Views),得到

表1 关系数据库产品应用DM和R2RML的情况

RDB产品	是否支持RDF数据存储和SPARQL	是否支持RDB2RDF	是否支持DM和R2RML
Oracle	支持	支持	支持
IBM DB2	支持	第三方RDB2RDF工具	第三方RDB2RDF工具
SQL Server	支持	第三方RDB2RDF工具	第三方RDB2RDF工具
MySQL	通过第三方插件或者API可以支持	第三方RDB2RDF工具	第三方RDB2RDF工具
PostgreSQL	通过第三方插件或者API可以支持	第三方RDB2RDF工具	第三方RDB2RDF工具

表2 RDB2RDF 工具平台对 DM 和 R2RML 的支持

工具名称	DM		R2RML		转换方式	支持的数据库
	是否支持	测试用例通过情况	是否支持	测试用例通过情况		
D2RQ	是	HSQLDB:22/24	否		实时 ETL	Oracle, IBM DB2, MS SQL Server, MySQL, PostgreSQL
RDF-RDB2RDF	是	PostgreSQL 20/24; SQLite:14/24	是	PostgreSQL : 51/62; SQLite : 55/62	ETL	MySQL, PostgreSQL
XSPARQL	是	全部通过	是	全部通过	实时	MySQL, PostgreSQL
SWObjectsdm-materialize	是	全部通过	否			MySQL, PostgreSQL, Oracle
ultrawrap	是	Postgresql 全部通过	是	OraclePostgresql 全部通过	实时 ETL	Oracle, IBM DB2, MS SQL Server, PostgreSQL
db2triples	是	全部通过	是	MySQL : 61/62; PostgreSQL : 全部通过	实时 ETL	MySQL, PostgreSQL
OpenLink Virtuoso	否		是	33/62	实时	Virtuoso
morph	否		是	HSQLDB:54/62	实时 ETL	MySQL, PostgreSQL, MonetDB

RDF 数据，这种功能与 R2RML 异曲同工，是 R2RML 诞生之前各种私有映射语言的一种。由于认识到 R2RML 已经成为 W3C 的推荐标准，基于 R2RML 来实现 RDB2RDF 将成为各种 RDB2RDF 工具的一种趋势，因而 Virtuoso 也开始支持 R2RML。它采用间接的方式，通过安装一个名为 R2RML VAD 的插件，将 R2RML 转换成它自己的关联数据视图的语法，最终还是基于私有的 RDB2RDF 映射语言和内置的功能来实现 RDB2RDF 的转换。数据库管理员可在 Virtuoso 后台把已定义好的 R2RML 脚本导入到 Virtuoso，系统执行映射规则，便可在其关联数据视图中查阅 RDF 数据。<sup>[11]</sup>从 W3C 的 RDB2RDF 实施报告中看出，Virtuoso 的测试用例通过率是较低的，62 个测试用例中，只有 33 个通过，其余 29 个为“cannotTell”。另外，它不支持 R2RML 的 rr:sqlQuery，该句法用于对基于 SQL 查询的逻辑表进行 RDB2RDF 映射定义，是 R2RML 的一大功能特色，不支持此功能意味着放弃了 R2RML 在数据转换之前进行数据筛选和整合的灵活性。在 RDF 数据的生成上，Virtuoso

只支持实时转换模式，不支持 ETL 模式。

Ultrawrap 是 Capsenta 公司开发的一款专用于 RDB2RDF 的商业软件，目的是在语义网环境下，最大限度地开发和利用 RDB 的潜力，将 RDB 中的数据发布到语义网中。Ultrawrap 是全面支持 DM 和 R2RML 这两个标准的平台之一，在 W3C 的 RDB2RDF 用例测试中，全部为“通过”状态。Ultrawrap 包括两个主要组件：RDB2RDF 映射编译器和服务器。前者负责编译 RDB2RDF 映射并生成 RDF 或 OWL 数据；后者负责接收 SPARQL 查询请求并返回数据。在 RDF 数据的生成上，它支持实时转换和 ETL 这两种模式。在性能上，Capsenta 公司声称 Ultrawrap 能够充分地利用现有的 SQL 基础架构，特别是数据库自带的元数据和 SQL 查询优化器，可使 SPARQL 检索达到 SQL 检索同样的性能和速度。<sup>[12]</sup>在 Berlin SPARQL Benchmark 的测试中，分别基于 Ultrawrap 和原生 RDB 对包含 1 亿条三元组的数据进行查询的结果表明，Ultrawrap 执行 SPARQL 查询的速度，与 RDB 执行 SQL 查询的速度相当<sup>[13]</sup>。

Morph 是一组语义技术工具套件,由“本体工程工作组”开发和维护,包括 RDB2RDF 组件 morph-RDB,支持 R2RML,不支持 DM。在 RDF 数据的生成上,它支持实时转换模式和 ETL 模式。在实时转换模式中,它需要根据已定义好的 R2RML 映射将前端的每一次 SPARQL 查询请求转换为相应的 SQL 查询请求;在 ETL 模式中,它需要根据已定义好的 R2RML 映射一次性地将 RDB 中的数据转换为 RDF 数据。Moreph-LDP 是 morph 家族的另一个组件,可以看作是 morph-RDB 的一个扩展<sup>[14]</sup>。它试图实现 W3C 的另一个标准——关联数据平台 (Linked Data Platform, 简称 LDP)<sup>[15]</sup>,LDP 是对关联数据四原则的进一步明确和扩展,规定了如何在通过资源的 URI 获取更多关于资源的信息之外,还能对这些信息进行更新,包括更新的方式和权限控制等。Morph-LDP 的目的是基于 R2RML 来实现 LDP,使得 RDB 中的数据能无缝地集成到 LDP 环境中。它根据 R2RML 映射从 RDB 中生成的 RDF 数据以 LDP 的标准封装,呈现给前台,并接收和处理客户端发送的按照 LDP 标准封装的 SPARQL 查询或更新请求。

## 4 在实际项目中的应用

### 4.1 音乐大脑关联数据项目 (LinkedBrainz)

#### 4.1.1 项目背景和需求

MusicBrainz 是一个开放的音乐维基百科,允许任何人贡献数据和内容并在网站上公开发布,系统基于 RDB 构建,可批量下载 MusicBrainz 数据库的数据及应用软件,数据为 SQL 格式,可导入 PostgreSQL 数据库中,MusicBrainz 不仅公开发布了数据,还公开发布

了数据结构 (Next Generation Schema, NGS) 和数据间的关系定义 (Advanced Relationships, ARs)。在 MusicBrainz 数据库中,艺术家、专辑、作品等已经被赋予了唯一 ID 并拥有丰富的关联关系,已经被关联数据社区广泛利用,但并未直接提供关联数据服务。

LinkedBrainz 项目的目的是将 MusicBrainz 数据库以关联数据的形式发布,其任务包括:为 NGS 和 ARs 转换为 RDF 数据建立映射;为 MusicBrainz 服务器增加内容协商的功能以为资源的 URI 提供“解引 (dereference)”服务;提供一个 SPARQL 端点 (Sparql Endpoint) 以响应 SPARQL 查询请求。由于 MusicBrainz 是一个数据不断更新并持续提供服务的系统,因而 LinkedBrainz 项目面临的一个挑战是在将 MusicBrainz 发布为关联数据时,不能影响系统已有的功能。项目最终选择了简便易行的 RDB2RDF 的解决方案,而非直接修改系统原有代码或利用现成的软件将 PostgreSQL 数据库转换为 RDF。

#### 4.1.2 利用 R2RML

为了更明确地表达 MusicBrainz 数据库中各类数据实体间的关系,NGS 和 ARs 需要与包括“音乐本体 (Music Ontology, MO)”在内的一系列本体词表建立映射。在 2013 年举行的第一届“语义音乐媒体”国际论坛上,Peter Haase 的报告<sup>[16]</sup>介绍了如何利用 R2RML 为 NGS 和 ARs 与 MO 定义映射。R2RML 正好满足了 LinkedBrainz 实现 RDB2RDF 的需求,它允许将 RDB 数据结构映射到一个或多个本体词表。如下图 2 所示,MusicBrainz 数据库中的表“Recording”映射为音乐本体 (其命名空间前

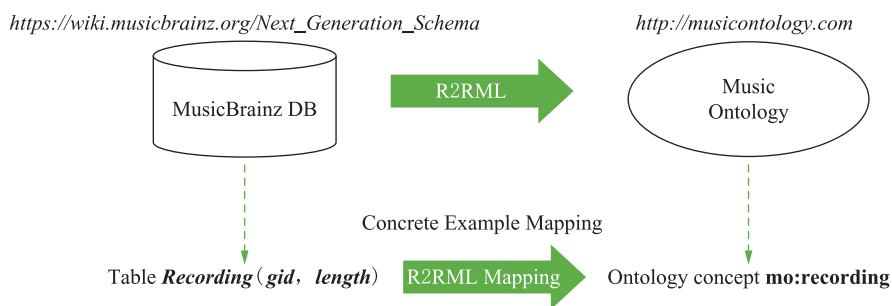


图 2 利用 R2RML 实现 MusicBrainz 数据和音乐本体的映射<sup>[16]</sup>



缀为 mo) 的一个类：mo:recording。

用 R2RML 将表 artist 映射到音乐本体的 mo:MusicArtist 类的代码如下：

其中 rr 为 R2RML 命名空间的前缀，rr:tableName 定义了需要映射的 RDB 表，rr:class

定义了该表映射到哪个类，rr:template 定义了一个 mo:MusicArtist 实体的 URI 生成规则。

用 R2RML 将字段映射到音乐本体的属性的代码如下：

这里用到了 rr:sqlQuery 来作为一个逻辑表

```

lb:Artist a rr:TriplesMap;
  rr:logicalTable [rr:tableName "artist"];
  rr:subjectMap
    [rr:class mo:MusicArtist ;
     rr:template "http://musicbrainz.org/artist/{gid}#_"];
    
```

```

lb:artist_name a rr:TriplesMap ;
  rr:logicalTable [rr:sqlQuery ""SELECT artist.gid,
  artist_name.name FROM artist INNER JOIN
  artist_nameON artist.name = artist_name.id""];
  rr:subjectMap [rr:template
"http://musicbrainz.org/artist/{gid}#_"];
  rr:predicateObjectMap [rr:predicate foaf:name;
  rr:objectMap[rr:column "name"]].
    
```

(rr:logicalTable)，用一个 SQL 语句来获取 RDB 中不同表的字段，而不是如上例中直接用数据库中已有的表。正如上文所述，R2RML 的这种功能为不改变 RDB 的原有结构，而灵活地生成 RDF 数据提供了便利。rr:predicate 定义了 artist\_name 字段映射到 FOAF 本体中的 foaf:name 属性。

### 4.1.3 系统实现

Peter Haase 的报告<sup>[16]</sup>和 Barry Norton 在 EUCLID 项目的培训课程<sup>[17]</sup>中介绍了 LinkedBrainz 的技术架构，如图 3 所示。

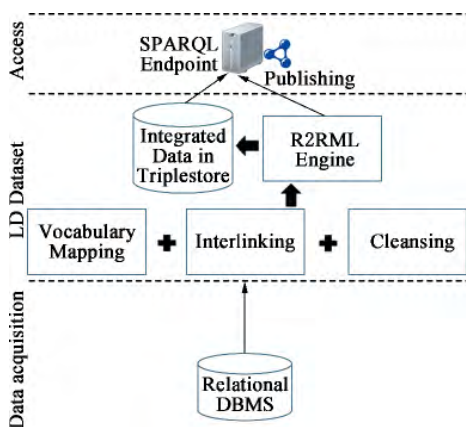


图 3 LinkedBrainz 的技术架构<sup>[17]</sup>

在转换模式上，采用了典型的 ETL 模式，首先利用 R2RML 映射规则将 RDB 数据结构与领域本体建立好映射，再利用 R2RML 映射引擎（也即上文提到的 RDB2RDF 工具），根据 R2RML 映射文件生成的 RDF 数据，导入到专用的 RDF 存储库（也叫三元组存储库，即 Triplestore）中，基于 RDF 存储库提供 SPARQL 查询服务，在 RDB2RDF 工具的选择上，采用了商业平台 ultrawrap。

在 EUCLID 培训课件介绍的 MusicBrainz 关联数据演示系统上，基于向 RDF 存储库发送 SPARQL 查询请求并返回 RDF 数据的模式，实现了各种数据可视化效果，如不同的视图浏览数据、分面检索、可视化标签云图功能等。

## 4.2 布鲁塞尔语义云项目 (OSCB)

### 4.2.1 项目背景和需求

布鲁塞尔语义云 (Open Semantic Cloud for Brussels, OSCB) 项目始于 2011 年 2 月 1 日，其目标是为比利时的布鲁塞尔地区所发生的大事、旅游线路、公交站点等数据建立一个关联数据发布和消费平台，以便不同的数据提供者能够方便地发布他们的数据，最终用户或应用开发者也能以一种通用直接而简单的方式来利

用这些数据。数据提供者一般以关系数据库表或XML格式提供原始数据,该项目采用了基于R2RML将这些数据转换为RDF三元组的方式。因为数据包含了大量的时间和空间信息,需要提供一个可检索时空数据的SPARQL端点,并需采用基于地图的可视化方式来展示数据,因而需要将原关系数据库或XML文档中的数据实体映射到空间或时间本体中的相应概念,R2RML也正好可以满足这一需求。<sup>[18]</sup>

#### 4.2.2 利用R2RML

文献[18]举了三个不同数据源的例子,来说明如何利用R2RML将关系数据库、XML、地理空间数据库转换为RDF三元组。

以下R2RML映射的作用是将XML数据映射为RDF三元组。该XML数据中保存了布鲁塞尔的一些文化事件以及与事件相关的机构。首先利用R2RML的逻辑表(rr:logicTable)的功能,将XPath查询XML的结果当做一个逻辑表,语句rr:class gospl:Address将一条XPath查询结果记录映射为地理空间本体(前缀为gospl)中的Address类,作为三元组的主体(Subject),语句rr:predicate geo:geometry将geo本体中的geometry作为谓词(Predict),语句rr:objectMap定义了该三元组的客体(Object),其取值为XPath查询结果中的" Institution\_

Full\_Address" 字段值。

以下映射将关系数据库中的某一站点信息转换为RDF数据。它将一条SQL查询结果当做一个逻辑表,这条SQL查询获取表stop中的字段值并将其中两个字段值合并成一个字符串,起到了生成RDF数据之前的数据处理作用。在三元组映射定义中,将一条结果记录映射为地理空间本体中的Location类,这里用了来自两个不同本体中的术语当做谓词gospl:Location\_with\_Latitude和rr:predicate geo:geometry,将生成两个三元组。

#### 4.2.3 系统实现

该项目基于R2RML映射以ETL的方式生成了包含空间和时间信息的RDF数据,以stRDF格式(spatiotemporalRDF)存储于名为Strabon的专用时空数据存储库中。Strabon也提供SPARQL端点,允许以stSPARQL(spatiotemporal SPARQL)语言查询存储库中的RDF数据.stSPARQL语言是对SPARQL语言的扩展,允许在SELECT, FILTER, 和 HAVING子句中以空间术语为参数来辅助查询。

由于有的数据提供者所提供的原始数据中的地址没有包含经纬度等地理空间信息,该项目还整合了各种外部关联数据集中的数据,如DBPedia、GeoNames和LinkedGeoData。系统先

```
rr:logicalTable [
  rr:xpathQuery""//Institution_Instance ~
  ./InstitutionID ~
  ./Institution_Street_FR ~
  ...
  ./Institution_City_FR ~
  ./Institution_Full_Address"";];
rr:subjectMap [
  rr:class gospl:Address ;
  rr:template "http://www.agenda.be/db/Address_of_Institution/{
  InstitutionID}";];
...
rr:predicateObjectMap [
  rr:predicate geo:geometry;
rr:objectMap [
  rr:column "Institution_Full_Address";
  rr:termType rr:Literal ;
  rr:datatype virtrdf:Geometry;];];.
```



```
<#TM_location >
a rr:TriplesMap ;
rr:logicalTable [
  rr:sqlQuery ""SELECT stp_identifier, stp_longitude, stp_latitude,
CONCAT('POINT(',stp_longitude,',',stp_latitude,');http://www.opengis.net/def/crs/
EPSG/0/4326') aspointFROM stop"";];
  rr:subjectMap [
    rr:template "http://www.stib.be/location/{stp_identifier}";
    rr:class gospl:Location;];
  ...
  rr:predicateObjectMap [
    rr:predicate gospl:Location_with_Latitude;
    rr:objectMap[
      rr:column "stp_latitude";
      rr:termType rr:Literal;
      rr:datatype xsd:double;];];
    rr:predicateObjectMap[
      rr:predicate geo: geometry;
      rr:objectMap [
        rr:column "point";
        rr:termType rr:Literal ;
        rr:datatype virtrdf:Geometry;];]; .
```

利用这些数据集所提供的数据消费接口获取所需数据，并转换为 stRDF 格式，存储于 Strabon 中。数据检索的返回结果可在地图上可视化地呈现。

## 5 结语

W3C 的两个 RDB2RDF 标准自推出以来，得到了该领域研究者和应用者的关注，其开发者在 ISWC 等国际会议上提供详尽的培训，在多种 RDB2RDF 工具上得到应用，甲骨文公司这样资深的关系数据库提供商也在 Oracle 数据库的新产品中植入了 DM 和 R2RML 的支持，还有了 LinkedBrainz 和 OSCB 这样的实际应用案例，但其应用范围仍然较为狭窄，局限在特

定的需求和环境下，而且作为一种新兴的标准规范，其推广和应用仍需时间和市场的考验。比如像 D2R 这样应用广泛 RDB2RDF 平台，已有自己成熟的映射语言（D2Rq），除了只是象征性地支持了 DM 外，两年多来并没有支持 R2RML 的动作。随着原生的 RDF 存储库和各种 NoSQL 数据库技术的进一步成熟，并得到越来越广泛的应用，基于 RDB2RDF 的语义应用架构可能将作为一种过渡性的解决方案而退居一隅。但对于目前大量的数据仍保存在关系数据库中这样的现实，若要方便快捷地从 RDB 中获取 RDF 数据，采用 RDB2RDF 方法及利用 DM 和 R2RML 这两个标准仍不失为一种可行的选择。

### 参考文献

- [1] W3C RDF Working Group. RDF 1.1 Concepts and Abstract Syntax[EB/OL].[2014-12-03].<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
- [2] 夏翠娟. RDB2RDF标准及应用研究[J]. 现代图书情报技术, 2013(4):10-17.
- [3] Marcelo Arenas, etc. A Direct Mapping of Relational Data to RDF[EB/OL]. [2012-12-08].<http://www.w3.org/TR/rdb-direct-mapping/>.
- [4] Souripriya Das, Seema Sundara, Richard Cyganiak. R2RML: RDB to RDF Mapping Language[EB/OL]. (2012-09-27).[2012-12-12]. <http://www.w3.org/TR/r2rml/>.
- [5] Juan F Sequeda, Barry Norton, Daniel P Miranker.

- (2013) Relational Database to RDF (RDB2RDF) Tutorial: Theory and Practice[EB/OL]. [2014-12-10]. <http://www.slideshare.net/juansequeda/rdb2-rdf-tutorial-iswc2013>.
- [6] Chuck Murray. Spatial and Graph RDF Semantic Graph Developer's Guide[EB/OL]. [2014-12-06]. <http://docs.oracle.com/database/121/RDFRM/title.htm>.
- [7] Mario Briggs, etc. Resource description framework application development in DB2 10 for Linux, UNIX, and Windows, Part 1: RDF store creation and maintenance[EB/OL]. [2014-12-21]. <http://www.ibm.com/developerworks/data/tutorials/dm-1205rdfsdb210/dm-1205rdfsdb210-pdf.pdf>.
- [8] Satya S Sahoo, etc. A Survey of Current Approaches for Mapping of Relational Databases to RDF(R/OL). [2012-12-10]. [http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF\\_SurveyReport.pdf](http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf).
- [9] Boris Villazón-Terrazas, Michael Hausenblas. RDB2RDF Implementation Report[R/OL]. [2014-12-12]. <http://www.w3.org/TR/2012/NOTE-rdb2rdf-implementations-20120814/>.
- [10] Richard Cyganiak, etc. Accessing Relational Databases as Virtual RDF Graphs[EB/OL]. [2014-12-26]. <http://d2rq.org/>.
- [11] OpenLink Software Documentation Team. OpenLink Virtuoso Universal Server: Documentation[EB/OL]. [2014-12-26]. <http://docs.openlinksw.com/pdf/virtdocs.pdf>.
- [12] Juan F Sequeda, Daniel P Miranker. Ultrawrap: SPARQL Execution on Relational Data[R/OL]. [2014-12-24]. [http://apps.cs.utexas.edu/tech\\_reports/reports/tr/TR-2078.pdf](http://apps.cs.utexas.edu/tech_reports/reports/tr/TR-2078.pdf).
- [13] Juan F Sequeda, Daniel P Miranker. SPARQL Execution as Fast as SQL Execution on Relational Data[EB/OL]. [2014-12-26]. [http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/PostersDemos/iswc11pd\\_submission\\_94.pdf](http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/PostersDemos/iswc11pd_submission_94.pdf).
- [14] Nandana Mihindukulasooriya, etc.. morph-LDP: An R2RML-based Linked Data Platform implementation[EB/OL]. [2014-12-24]. [http://2014.eswc-conferences.org/sites/default/files/eswc2014pd\\_submission\\_85.pdf](http://2014.eswc-conferences.org/sites/default/files/eswc2014pd_submission_85.pdf).
- [15] Steve Speicher, John Arwe, Ashok Malhotra. Linked Data Platform 1.0[EB/OL]. [2014-12-16]. <http://www.w3.org/TR/ldp/>.
- [16] Peter Haase. Mapping, Interlinking and Exposing MusicBrainz as Linked Data[R/OL]. [2014-12-24]. 1<sup>st</sup> International Workshop on Semantic Music and Media (SMAM2013) Sydney, Oct 21, 2013.
- [17] Barry Norton, Juan Sequeda. Mapping Relational Databases to Linked Data[EB/OL]. [2014-12-24]. <http://www.slideshare.net/EUCLIDproject/r2-rml-londonsemweb201304>.
- [18] Kevin Chentout, Alejandro Vaisman. Adding Spatial Support to R2RML Mappings[C]. OTM 2013 Workshops, LNCS 8186, pp.398-407, 2013.
- 夏翠娟 女, 上海图书馆(上海科学技术情报研究所), 高级工程师。从事数字图书馆研究开发工作。E-mail: cjxia@libnet.sh.cn 上海 200031
- 金家琴 女, 上海图书馆(上海科学技术情报研究所), 副研究馆员。从事文献提供工作。上海 200031
- (收稿日期: 2015-02-03)