



Date : 01/07/2006

在图书馆数据库中采集当日信息的可行性

*Translation by Ms. Hong Gao,
Ms. Chong Luo and
Ms. Wei Zhang
National Library of China*

Meeting:	123 Cataloguing
Simultaneous Interpretation:	Yes
<p>WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL 20-24 August 2006, Seoul, Korea http://www.ifla.org/IV/ifla72/index.htm</p>	

[摘要] 本文介绍拉脱维亚国家图书馆和 Lursof 公司的合作项目“期刊内容发布模型”(*Latvia's Model for Periodicals' Content Distribution*) 的经验。目的是帮助用户存取每日报刊的全文信息。作者分析了拉托维亚国家图书馆实现对电子资源书目控制的潜力，并探研从图书馆数据库资源本身自动生成元数据的可能性。报告还从用户和合作者的视角出发，阐述该模型的优点以及今后需要解决的问题。

一、网络资源书目控制的元数据组织方式

我来自拉脱维亚，一个位于爱沙尼亚和立陶宛之间的波罗的海国家，西邻瑞

典、东邻俄罗斯。

我认为，网络资源书目控制的元数据组织方式有以下几种：

- 1、在图书馆信息系统中描述网络资源，通常用 MARC 格式；
- 2、设计或采用一个独立的系统，对数字资源进行控制，如用 DC 描述，或用其他一些最简级的结构化格式描述；
- 3、利用出版商提供的元数据；
- 4、尝试从资源本身自动抽取和生成元数据；
- 5、针对不同种类资源，将上述提及的方发组合应用；
- 6、在网站上发布已存档和可获取资源的索引，或采用其他类似的方法；
- 7、对门户或搜索引擎（如 Google），适当采用结构化的图书馆元数据，如拉托维亚的“Siet”。

二、拉脱维亚国家图书馆的实践

拉脱维亚国家图书馆（NLL）在上述提到的方法中，已经采用了三种方法对原生数字资源进行控制。第 1 种是应用于图书馆集成系统 ALEPH500 的传统方式；第 2 种目前正在进行中，即与出版商合作，以远程访问的方式为即将发行的出版物分配 ISBN 号（未来还可能分配 ISSN 号），以及向开通 Internet 的国家书目数据库提供这些出版物的预编目元数据；第 3 种方法是与欧洲其他国家图书馆联合进行的“TEL-ME-MOR”项目，即在 TEI 门户和拉脱维亚全文搜索引擎 Siets 中使用图书馆元数据。

从书目控制的意义讲，考虑出版商的所思所为，我们就能够“覆盖”因特网。拉脱维亚有许多活跃的电子出版商，但图书出版商却寥寥无几。因此，第一部分

我们要关注连续出版物。今天,我想详细讨论拉脱维亚元数据组织的第4种方式,即“期刊内容发布模型”(Latvia's model for periodicals' content distribution)

三、拉脱维亚的合作项目“期刊内容发布模型”

1、项目介绍

2005年,拉脱维亚国家图书馆启动了一个合作项目,其目的是为用户在检索国家分析数据库时,为用户提供报纸、杂志和期刊的Internet全文检索。该数据库包括1988-2005年拉脱维亚出版的报纸、杂志和期刊的元数据信息(约1,500,000条),Aleph500系统(14.2版)可通过Internet显示文献内容。
[\[http://www.lnb.lv/eng/db/nat_bibl.htm\]](http://www.lnb.lv/eng/db/nat_bibl.htm)

国家分析数据库始建于1996年。自2002年开始,其数据在Aleph500系统中采用MARC21格式建立。先前本地系统中的数据(666,700条)已转入Aleph500系统。

2、国家分析数据库

国家分析数据库可提供著者、题名、文章来源(连续出版物题名)、关键词、文章中提及的个人名称、其他标准等检索途径。该数据库在国、内外应用广泛,因为其数据提供拉脱维亚语、俄语和英语等语种的论题性主题。

拉脱维亚每年出版期刊约600种,其中2/3在国家分析数据库中有编目数据。漫画、内容说明书、填字游戏、电视节目和电台广播节目等类似内容不在分析处理之列。

从国家分析数据库检索全文的想法较早提出,到2005年5月才开始实施。

2004年,“文化信息系统”(负责管理国家联合图书馆信息系统项目的机构),

开始对书目著录与全文链接项目进行招标。最终，与 Lursoft 达成了协议，提供国家分析数据库的全文检索。

3、报纸全文文档

我们的合作伙伴 — Lursoft (<http://www.news.lv>)，正在编辑拉脱维亚出版的电子报纸、电子杂志和电子期刊的全文文档。文档包括所有中央和地方出版的、由 Lursoft 和各出版商达成协议的、合法获得的出版物的电子复本。Lursoft 报纸图书馆提供的全文超过 2,707,000 篇，并以每天 1000 篇左右的速度增长。

2006 年，Lursoft 提供 43 种期刊的检索，以及来自 Lursoft 数字报纸图书馆提供的出版机构的评论。该数字图书馆可提供全文阅读，但图表、曲线图、草图等尚未添加至文本中。数字报纸图书馆不包含照片、内容说明书和广告。

为了尽可能高效地开发电子出版物的潜力，Lursoft 决定不仅要提供统一资源地址 (URLs)，还要提供从期刊中获取的其他信息。拉脱维亚国家图书馆负责定义书目著录描述的要求：格式、版式、代码。

4、元数据的自动传送

2005 年 5 月，Lursoft 开始进行出版物电子描述信息的自动传送，其元数据由出版商每天从报纸的版面中自动抽取并传送给 Lursoft。

Lursoft 和出版商之间早有协议，从 1993 年开始它就接收出版商提供的报纸的版面表单。文章从表单中自动抽取 (部分进行人工更正)，并使用 Lursoft 研发的专门软件将其转换为结构化的 XML 数据库记录。通过专门研发的工具，一些数据 (书目数据元素) 被转换并传送给拉脱维亚图书馆。

相关技术包括由 Lursoft 研发的 XML 数据库管理工具 (详细信息见 <http://www.siets.biz>), 已广泛应用于国、内外的新闻和官方文献数据库中 (参见 <http://www.siets.biz/solutions/usecases/>和 <http://www.siets.biz/company/>)。

书目著录包括文章的著者 (若著者已知)、文章篇名、连续出版物 (杂志、期刊或报纸) 题名、卷期编号、以及每篇文章的唯一标识符。书目信息在预定格式 (下载到图书馆信息系统中的 Aleph Sequential) 中描述, 当天的信息在凌晨三点之前传递到拉脱维亚国家图书馆。

拉脱维亚国家图书馆的系统管理员负责将每天收到的重要文档转入 Aleph500 系统。为此, 需要建立一个单独的数据库, 因为收到的记录并不需要全部包含在国家分析数据库中, 而是有所选择。仅有约 1/2 的中央出版社出版物和 1/6-1/4 的区域性出版物需要编目。此外, 还要看文章的意义和价值, 如那些关于研究、政治、经济、历史和本地研究的文章需要编目。

书目著录首先需要著录定长字段, 否则数据无法转入系统。其次, 我们还要求 Lursoft 提供诸如编目源、语种、国家书目标识、URL 等的信息, 这些信息以元数据格式显示, 不需人工录入, 大大降低了编目员的工作量。

权威馆员有权限将以上描述纳入国家分析数据库中, 并依据 MARC21 标准对收到的信息进行扩充。

下面这张幻灯片是一条编目员处理过的记录, 包括 UDC 标引、主要题名、变异题名、丛编信息、文章注释、主题词、附加款目及关于出版物补编的数据。

四、如何检索全文?

国家分析数据库装在我们的服务器上, 全文数据库装在 Lursoft 服务器上。

通过互联网收到用户请求后，国家分析数据库的记录就可以通过文章的唯一标识符与全文信息进行链接。访问国家分析数据库是免费的，全文则依据与特定出版商的协议来获取权限。版权法是 Lursoft 数字文档的一部分，是其法定获得信息的途径。

用户如何获取一篇文章的全文呢？

首先，链接拉脱维亚国家图书馆的主页 http://www.lnb.lv/eng/db/nat_bibl.htm，找到“数据库”菜单，并选择“国家分析数据库”。

其次，选择必要信息选项（如“题名”），并检索关于“曲棍球”的文章，检索结果有 62 条记录。

再次，打开其中一条记录，可看到文章的电子资源定位与访问地址。点击此链接时，会弹出版权信息窗口。

最后，可以看到 Lursoft 数字报纸数据库。授权的用户通过他们的 ID 和密码登录后，即可阅读全文。

该合作项目正处在发展阶段，目前所做的工作是将国家分析数据库中 1993-2004 年的记录（约 400,000）与报纸全文文档的回溯部分进行连接。

国家分析数据库也可通过欧洲国家图书馆 TEL 门户进行访问，该门户被纳入 2005 年由欧盟支持的 TEL-ME-MOR 项目的框架内。

五、问题

目前还存在以下几个问题：

- 有时每日的文档需传输两次；
- 并非所有连续出版物的文章都提供链接，因为出版商认为这些文章不应向公众开放；

- 用户希望这些信息应全部免费获取。

但上述问题并不影响以后的工作进程。项目在以下方面使图书馆、出版商和用户获得受益：

其一：图书馆

- 出版、元数据提供、检索可在同一天进行；
- 拉脱维亚国家图书馆可获得从当日电子文章中抽取的元数据（简单记录）；
- 记录的质量提高，可避免手工录入中题名、著者的输入错误等。

其二：出版商、用户

- 出版商可得到关于文章、著者访问频率的可靠性统计数据；
- 出版商和 Lursoft 不断地吸引新的用户；
- 用户可以很容易地从文章的全文转移到被链接的信息（如文章中被引用的个人名称）。在同一界面还提供商业数据库和其他正式注册者数据库的连接。
- 为用户提供不间断的信息服务。

作者信息

Ilona Dukure ilona.dukure@lnb.lv

Anita Goldberga anita.goldberga@lnb.lv

中文翻译：高红、罗翀、张玮

中国国家图书馆

Chinese translation by Hong Gao, Chong Luo, Zhang Wei
National Library of China