

Date : 17/07/2006



## Se procurer les nouvelles du jour dans la base de données de la bibliothèque : une possibilité ?

**Ilona Dukure**

[ilona.dukure@lnb.lv](mailto:ilona.dukure@lnb.lv)

**Anita Goldberga**

[anita.goldberga@lnb.lv](mailto:anita.goldberga@lnb.lv)

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

### **Abstract**

*Cette communication relate l'expérience acquise par la Bibliothèque nationale de Lettonie et la société Lursoft au cours du projet coopératif appelé « Le modèle letton pour la diffusion du contenu des périodiques ». Le but de ce projet est de faciliter l'accès des utilisateurs à l'information en texte intégral de la presse quotidienne. Les auteurs analysent le potentiel de la bibliothèque nationale en ce qui concerne le contrôle bibliographique des ressources électroniques, et la possibilité d'exploiter dans la base de données de la bibliothèque les métadonnées générées automatiquement à partir des ressources elles-mêmes. On fera ressortir les avantages du modèle, du point de vue des utilisateurs et des partenaires du projet, ainsi que les problèmes qui restent à résoudre.*

Je viens de Lettonie, un pays balte pris en sandwich entre l'Estonie et la Lituanie. Notre pays est également voisin de la Suède à l'ouest et de la Russie à l'est.

À mon avis il existe plusieurs manières d'organiser le contrôle bibliographique des ressources du Web au niveau des métadonnées :

1. Cataloguer les ressources du Web au moyen d'un système d'information bibliothéconomique (généralement, les formats MARC) ;
2. Concevoir ou adapter un système à part pour le contrôle des ressources numériques (catalogage en Dublin Core ou autre format à structure minimale) ;
3. Reprendre les métadonnées fournies par les éditeurs ;
4. Tenter d'extraire et de générer automatiquement des métadonnées à partir des ressources ;

5. Faire appel à une combinaison des solutions énumérées ci-dessus, en fonction de la nature des ressources ;
6. Publier l'index des ressources archivées et disponibles sur le site Web ou d'une manière analogue ;
7. Faire appel à des métadonnées bibliothéconomiques plus ou moins structurées pour un portail ou pour des moteurs de recherche tels que Google – ou, en Lettonie, Siets.

La Bibliothèque nationale de Lettonie (BNL) a mis en œuvre trois des possibilités énumérées ci-dessus pour le contrôle des ressources numériques natives. La première est la manière traditionnelle. Ensuite, il y a la coopération continue avec les éditeurs pour ce qui est de l'attribution des ISBN (et aussi, dans l'avenir, des ISSN) aux publications à paraître sous forme de ressources accessibles à distance, ainsi que pour ce qui est de la fourniture de métadonnées préliminaires relatives à ces publications dans les bases de données bibliographiques nationales qui font appel à Internet. Nous sommes en train d'implémenter la dernière de ces possibilités au sein du projet TEL-ME-MOR, en partenariat avec d'autres bibliothèques nationales d'Europe.

Nous ne saurions « couvrir » l'Internet (au sens du contrôle bibliographique) sans demander leur avis aux éditeurs. Il y a en Lettonie des éditeurs spécialisés dans les ressources électroniques, mais peu d'entre eux se consacrent aux monographies. C'est pourquoi nous nous sommes tout d'abord concentrés sur les publications en série. Aujourd'hui je voudrais aborder plus en détail la 4<sup>e</sup> possibilité, telle que nous l'avons mise en œuvre en Lettonie. Nous l'avons appelée : « Le modèle letton pour la diffusion du contenu des périodiques ».

En 2005 la BNL a lancé un projet coopératif visant à rendre les journaux, magazines et revues accessibles sur Internet en texte intégral à partir d'une recherche sur la base de données de Dépouillement National de la BNL. Cette base de données contient de la méta-information sur les articles de journaux, magazines et revues publiés en Lettonie sur la période 1988-2005 (soit environ 1 500 000 notices). Le système Aleph500 (version 505.14.2) facilite l'affichage du contenu sur Internet : [http://www.lnb.lv/eng/db/nat\\_bibl.htm](http://www.lnb.lv/eng/db/nat_bibl.htm).

La base de données de Dépouillement National a démarré en 1996. Depuis 2002, les notices sont produites dans le système d'information Aleph500 au format MARC21. Les données provenant du système local antérieur (666 700 notices) ont également été converties en Aleph500.

Sur la base de données de Dépouillement National, il est possible de lancer des requêtes par auteur, titre, et source de l'article (titre de la publication en série), par mots-clés, par noms de personnes mentionnés dans l'article, et sur d'autres critères encore. Cette base de données est largement utilisée en Lettonie et à l'étranger, dans la mesure où elle fournit des informations provenant de périodiques lettons publiés en letton, en russe, en anglais, etc.

La Lettonie publie environ 600 publications en série par an, dont les deux tiers sont catalogués dans la base de données de Dépouillement National. Les bandes dessinées, les prospectus, les mots croisés, les programmes de télé et de radio, etc., ne sont pas dépouillés.

L'idée de donner un accès au texte intégral des articles à partir de la base de données de Dépouillement National s'est fait jour très tôt, mais nous n'avons pu réellement la mettre en œuvre qu'en mai 2005.

En 2004, l'Agence nationale "Systèmes d'information culturelle", qui administre le projet de système d'information de bibliothèques commun au niveau national, a lancé un appel d'offre pour un procédé qui permettrait de relier les descriptions bibliographiques aux textes mêmes. Elle a ensuite passé un accord avec la société Lursoft IT pour construire le système permettant l'accès au texte intégral à partir de la base de données de Dépouillement National.

La société Lursoft IT (<http://www.news.lv>) – notre partenaire – compile les archives en texte intégral des journaux, magazines et revues électroniques qui sont publiés en Lettonie. Ces archives contiennent des copies numériques de toutes les principales publications de la presse centrale et régionale acquises légalement ou sur la base d'accords passés entre Lursoft et chacun des éditeurs. Il y a plus de 2 707 000 articles disponibles en texte intégral dans la bibliothèque de journaux de Lursoft et ce chiffre s'accroît d'environ 1 000 unités par jour.

En 2006, Lursoft fournit l'accès à 43 publications en série et comptes rendus d'agences de presse provenant de la Bibliothèque numérique de journaux de Lursoft. La lecture en texte intégral est facilitée, bien que les tableaux, graphiques et dessins n'y figurent pas. La Bibliothèque numérique de journaux ne contient ni photos, ni prospectus, ni publicités.

Afin d'exploiter tout le potentiel des publications électroniques avec le maximum d'efficacité, il a été décidé que Lursoft fournirait non seulement les URL mais aussi d'autres données qu'il est possible d'extraire de la maquette des périodiques.

La BNL a défini ses exigences concernant les descriptions à fournir – format, présentation, encodage.

En mai 2005 Lursoft a démarré la fourniture automatique des descriptions électroniques des articles publiés. Ces métadonnées sont extraites automatiquement, sur une base quotidienne, des maquettes de journaux fournies à Lursoft par les éditeurs.

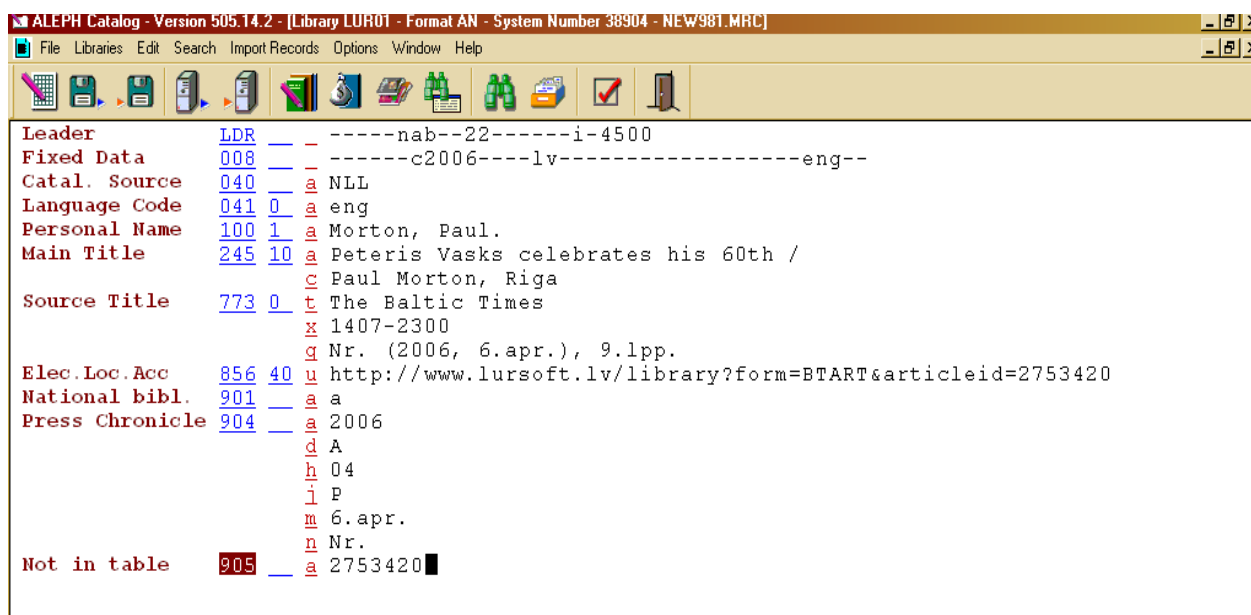
Lursoft a passé des accords avec les maisons d'édition (voir liste à l'adresse <http://www.news.lv>) et reçoit des maquettes de journaux depuis 1993. Les articles sont extraits automatiquement (et partiellement corrigés à la main) des maquettes et convertis en éléments de bases de données structurés en XML à l'aide d'un logiciel qui a été spécialement développé par Lursoft. À l'aide d'outils dédiés, certaines données (les éléments de données bibliographiques) sont converties et fournies à la BNL.

Ces techniques, y compris les outils de gestion de la base de données en XML, sont développées par Lursoft (pour plus de détails, voir <http://www.siets.biz>) et sont utilisées dans de nombreuses bases de données locales et internationales consacrées à l'information et aux publications officielles. (Voir aussi <http://www.siets.biz/solutions/usecases/> et <http://www.siets.biz/company/>).

Les descriptions contiennent des données relatives à l'auteur de l'article (s'il est mentionné), au titre de l'article, au titre de la revue ou du journal, à sa date et son numéro, et comportent un identifiant unique pour chaque article. Les descriptions des publications appropriées du jour sont fournies dans un format prédéfini (Aleph Sequential – pour permettre le déchargement dans le système d'information de la bibliothèque) quotidiennement jusqu'à 3 heures du matin.

Les administrateurs du système d'information Aleph500 en Lettonie (l'agence « Systèmes d'information culturelle ») importent dans le système Aleph500 les fichiers reçus quotidiennement. Une base de données distincte a été créée dans ce but, car toutes les notices reçues ne sont pas versées dans la base de données Dépouillement National. Il n'y a qu'environ la moitié des descriptions reçues de la presse publiée dans la capitale, et un quart à un sixième des descriptions de la presse régionale qui fassent l'objet d'un catalogage. Seuls les articles importants et valables sont sélectionnés pour être catalogués – c'est-à-dire, ceux qui concernent la recherche, la politique, l'économie, l'histoire et les études locales.

Voici à quoi ressemble le dépouillement une fois versé par Lursoft dans le module de catalogage d'Aleph500 :



Il reste des champs de longueur fixe à remplir pour la description ; sans quoi les données importées dans la base ne seraient pas valides. En outre, nous avons demandé à Lursoft de remplir des champs tels que, par exemple, la source de catalogage, la langue, la mention de bibliographie nationale, l'URL. La présence de ces champs dans le format de métadonnées réduit la charge de travail des catalogueurs, qui n'ont plus à les remplir à la main.

Les membres autorisés du personnel de la BNL ont le droit d'intégrer ces descriptions à la base de données Dépouillement National et d'enrichir les informations reçues selon les exigences de MARC21.

Ici nous voyons la notice une fois traitée par le catalogueur – elle comporte un indice CDU, une annotation de l'article, des champs de description ajoutés, des mots-clés, des données relatives au supplément de la publication :

ALEPH Catalog - Version 505.14.2 - [Library NBA01 - Format AN - System Number 1495981 - NEW982.MRC]	
File Libraries Edit Search Import Records Options Window Help	
Leader	LDR --- - ----nab--22-----i-4500
Control No.	001 --- - 001495981
Date and Time	005 --- - 20060411095846.0
Fixed Data	008 --- - 060406c2006----lv-----eng--
Catal. Source	040 --- a NLL
Language Code	041 0 a eng
Local UDC	081 --- a 78
Personal Name	100 1 a Vasks, Pēteris.
Main Title	245 10 a Peteris Vasks celebrates his 60th : b [par radošo darbību : stāsta komp.] / c Pēteris Vasks ; pierakst. Paul Morton
Series/Add.Tit.	440 6 a Iel. "Out & about"
Subject-P.Name	600 1 a Vasks, Pēteris
Indx Term Unco	653 --- a Latviešu komponisti
Indx Term Unco	653 --- a Mūzika
A.E. Pers. Name	700 1 a Morton, Paul, e pierakst.
Source Title	773 0 t The Baltic Times x 1407-2300 q Vol.10, Nr.502 (2006, 6./12.apr.), 9.lpp.
Elec.Loc.Acc	856 40 u http://www.lursoft.lv/library?form=BTART&articleid=2753420
National bibl.	901 --- a a
Press Chronicle	904 --- a 2006 b 04 d A h 04 i P m 6./12.apr. n Nr.502
Not in table	905 --- a 2753420

La base de données Dépouillement National se trouve sur notre serveur ; les textes intégraux se trouvent sur les serveurs de Lursoft.

Lorsqu'une requête d'un utilisateur est reçue via Internet, le lien entre la notice de la base Dépouillement National et le texte intégral est établi sur la base de l'identifiant unique de l'article (champ 856). L'accès à la base Dépouillement National est gratuit, ainsi que l'accès au texte intégral, conformément à l'accord passé avec l'éditeur (accès autorisé). Ainsi, la loi sur le copyright est respectée puisque les archives numériques de Lursoft contiennent de l'information légalement acquise (champ 856).

Comment un utilisateur a-t-il accès au texte intégral d'un article ? Il faut se connecter à la page d'accueil de la BNL ([http://www.lnb.lv/eng/db/nat\\_bibl.htm](http://www.lnb.lv/eng/db/nat_bibl.htm)), trouver le menu « Bases de données », et sélectionner l'item « Base de données Dépouillement National ».

On choisit un critère de recherche (par exemple « titre »), et on lance la requête, par exemple sur le mot « hockey ». Ici, on a 62 réponses :

NB analytic - Aleph main menu - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://195.13.129.13:8080/ALEPH/69RMHULLXEGUF82CAJYYKKM42BGGPEG2BJ4P6INBLBG1PD8H4-01121/con-Ing/eng

Links Customize Links Free Hotmail Windows Windows Marketplace Windows Media

**National Bibliography Analytic**  
 Database | Feedback | Preferences | Help | Ex Libris  
 Browse | Search | Results List | Previous Searches | Basket | Latvian

View Selected Select All Deselect Save/Mail Create Subset Add to Basket Refine Filter Print Page Back

Results for Words= ( hockey ); Sakārtots pēc Year, then Author

Records 1 - 20 of 62 Jump to # Up Forward

#	Author	Title	Year	Source
1	Aykroyd, Lucas.	Skrastins embodies Latvian hockey pride /	2006	Laiks - Ņujorka. - Nr.8 (2006, 18./24.febr.), 4.lpp.
2	Gluhiha, Aleksandra.	Oficiālie hokeja čempionāta suvenīri pārsteidz ar garozību /	2006	Diena. - ISSN 1407-1290. - Nr.97 (2006, 26.apr.), 8.lpp.
3		Hockey fans ecstatic over tie :	2006	The Baltic Times. - ISSN 1407-2300. - Vol.10, Nr.496 (2006, 23.febr./1.mart)
4	Zaviņš, Kaspars.	Izdevjot citu reālitāti.	2006	Neatkarīga Rīta Avīze Latvijā. - ISSN 1407-3403. - Nr.17 (2006, 20.jan)
5	Balandina, Julia.	Riga hosts Olympic ice-hockey qualifiers :	2005	The Baltic Times. - ISSN 1407-2300. - Vol.10, Nr.443 (2005, 3./9.febr.), 13
6		Cost of Ice-Hockey World Championships could hit 7 million :	2005	The Baltic Times. - ISSN 1407-2300. - Vol.10, Nr.473 (2005, 8./14.sept.), 1
7		Hokejs Latvijā -	2005	Hokejs 2006. - Nr.11 (2005), 22, 23.lpp.

On ouvre l'une des notices et on trouve mention de la localisation de l'article :

NB analytic - Aleph main menu - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://195.13.129.13:8080/ALEPH/94HQ5K14TRUGFUMAIV5G62MFMXVM3F7XLEVXJ3PXQRM8GUMGE7-01471/con-Ing/eng

Links Customize Links Free Hotmail Windows Windows Marketplace Windows Media

**National Bibliography Analytic**  
 Database | Feedback | Preferences | Help | Ex Libris  
 Browse | Search | Results List | Previous Searches | Basket | Latvian

Full View of Record Help

Result List Add to Basket Locate ILL Request Save/Mail Print Page Back

Choose format: Standard Bibl. reference MARC tags

Record 2 out of 61 Prev Record Next Record

Sys. no. Analytic 001447636

Title **Hockey fans ecstatic over tie :**  
 [nar Latvijas ASV neizšķirto spēli hokejā Ziemas olimpiskās spēlēs Turīnā (Itālija)]

Source The Baltic Times. - ISSN 1407-2300. - Vol.10, Nr.496 (2006, 23.febr./1.marts), 13.lpp.

Electronic Location **<http://www.lursoft.lv/library?form=BTART&articleid=2710319>**

Format AN

Prev Record Next Record

In the bibliographic record display, click underlined contents to use the text for navigation to other services

En cliquant sur ce lien, la fenêtre d'information sur le copyright apparaît :

The screenshot shows a Microsoft Internet Explorer browser window. The main window displays a bibliographic record from the 'National Bit' database. The record details include: Sys. no. Analytic: 001447636; Title: Hockey fans ecstatic over tie: [par Latvijas-ASV neizšķirto spēli hc]; Source: The Baltic Times. - ISSN 1407-2300; Electronic Location: http://www.lursoft.lv/library?form=B; Format: AN. A copyright information dialog box is overlaid on the right side of the browser window. The dialog box title is 'Copyrights Information'. It contains the following text: 'Please read the above before viewing external sources.'; 'In accordance with the Law on Copyright of the Republic of Latvia, having come into force since November 5,2000, the regulations for archives and libraries, envisaging reproduction of texts from the original, have been defined (Law on Copyright, Clause No.23).'; 'Copyright is not infringed when the original is used for the purpose of informative study, scholarship, and/or research without the agreement of and reimbursement to an author (Law on Copyright, Clause No.20, and No.21). In such a case the title of the work and name of the author should be indicated.'; 'All kind of activities pertaining to fixation, publishing, and promulgation, reproduction or any other dissemination without the author's consent should be considered infringement of copyright. Such conduct of a user involves violation of the "Law on Copyright", and he/she may be called account for.' Below the text, there is a statement: 'I have read the above statement and agree to abide by its restrictions:' followed by two buttons: 'Go Back' and 'Continue'. The 'Continue' button is highlighted with a red box. At the bottom of the dialog box, there are 'Prev Record' and 'Next Record' buttons. The main browser window also has a 'Prev Record' and 'Next Record' button at the bottom right.

**Copyrights Information**

Please read the above before viewing external sources.

In accordance with the Law on Copyright of the Republic of Latvia, having come into force since November 5,2000, the regulations for archives and libraries, envisaging reproduction of texts from the original, have been defined (Law on Copyright, Clause No.23).

Copyright is not infringed when the original is used for the purpose of informative study, scholarship, and/or research without the agreement of and reimbursement to an author (Law on Copyright, Clause No.20, and No.21). In such a case the title of the work and name of the author should be indicated.

All kind of activities pertaining to fixation, publishing, and promulgation, reproduction or any other dissemination without the author's consent should be considered infringement of copyright. Such conduct of a user involves violation of the "Law on Copyright", and he/she may be called account for.

I have read the above statement and agree to abide by its restrictions:

[Go Back](#) [Continue](#)

Prev Record Next Record

On accède alors à la bibliothèque numérique de journaux de Lursoft. Les utilisateurs habilités saisissent leur numéro d'identification et leur mot de passe :

The screenshot displays a web browser window with the following elements:

- Browser Title Bar:** NB analytic - Aleph main menu - Microsoft Internet Explorer
- Address Bar:** http://www.lursoft.lv -< NEWS.lv - LURSOFT IT laikrakstu bibliotēka >- - Microsoft Internet Explorer
- Page Header:** NEWS.lv logo, navigation links (Abonēšana, Jaunumi, Lasi jaunāko!, WAP ziņas, Slīdīnija, Publicitāte, Laikrakstu aģents, Komentāri, Statistika, Apraksts), and a search bar.
- Left Sidebar:**
  - Full View of Record:** Record 2 out of 61. System number: 001447636. Title: Hockey fans ecsta [par Latvijas-ASV ne]. Source: The Baltic Times. - IS. Electronic Location: http://www.lursoft.lv/. Format: AN.
- Main Content Area:**
  - Lietotāja autorizācija (User Authorization):** A login form with fields for 'Vārds', 'Lietotāja ID', and 'Parole', and a 'Pieslēgties' button.
  - Uzņēmumu datu bāzes (Company Databases):** A list of various databases including 'Eiropas Biznesa Reģistrs', 'Lursoft Biznesa Reģistrs DB', 'Uzņēmēja palīgs', etc.

Immédiatement après cette opération, le texte intégral de l'article recherché apparaît sur l'écran :



The screenshot shows a web browser window with two tabs. The active tab is 'NEWS.lv - LURSOFT IT laikrakstu bibliotēka'. The main content area displays a bibliographic record for a newspaper article. The record is titled 'Full View of Record' and includes the following fields:

- Record 2 out of 61**
- Sys. no. Analytic:** 001447636
- Title:** Hockey fans ecstatic over tie [par Latvijas-ASV neizšķirtu]
- Source:** The Baltic Times. - ISSN 1
- Electronic Location:** http://www.lursoft.lv/librar
- Format:** AN

The article text is displayed in a red-bordered box:

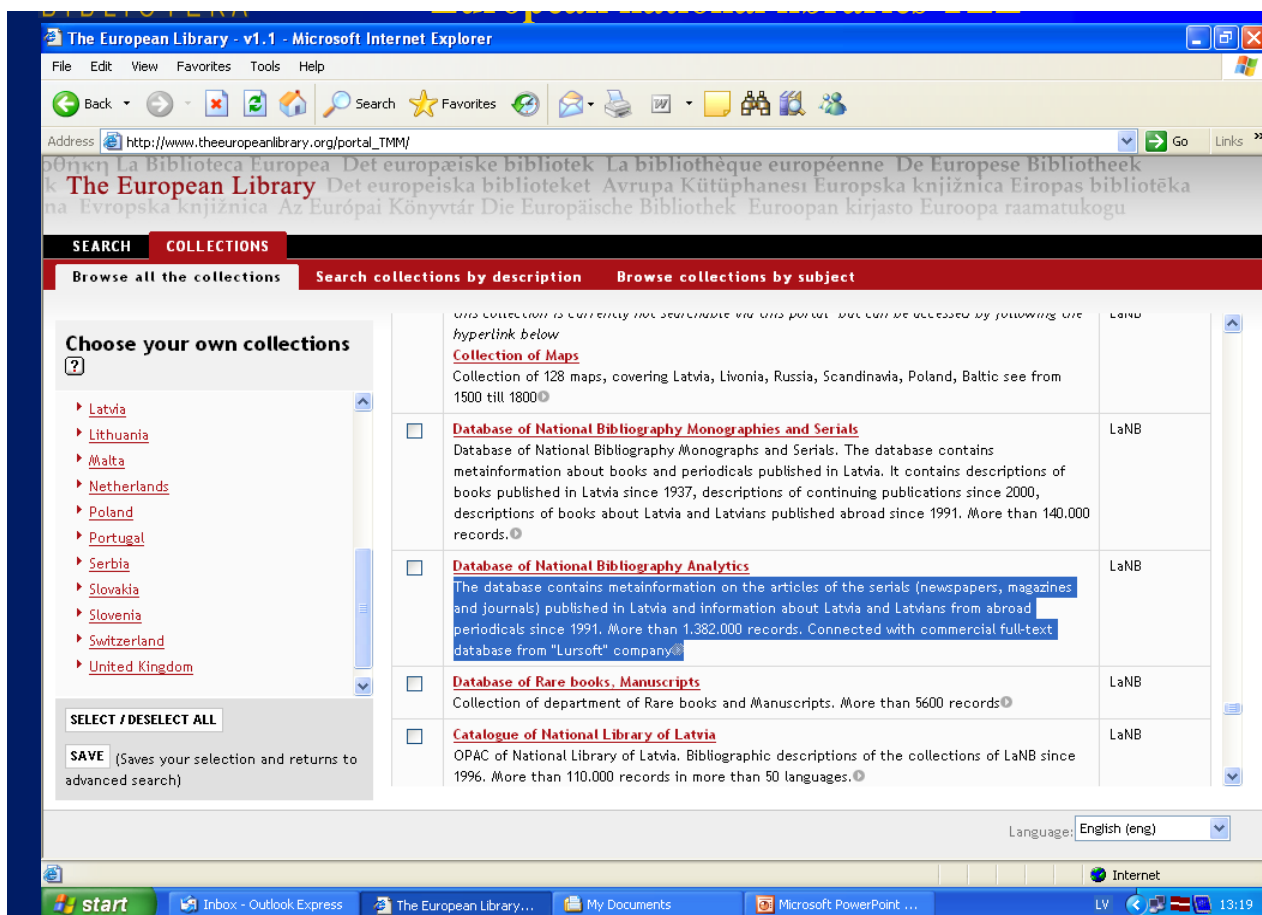
**Hockey fans ecstatic over tie**

**Autors:** TBT staff, RIGA  
**Publicēšanas datums:** Ceturtdiena, 2006. gada 23. februāris.  
**Rubrika:** Sport (13. lpp.)

Latvia's Olympic ice-hockey team was anything but disappointed by their 3-3 tie with the United States on Feb. 16. In fact, players and fans alike

Ce projet coopératif en est à un stade de développement. À l'heure actuelle nous continuons à établir des liens entre la partie rétrospective des archives de journaux en texte intégral et les notices existantes (environ 400 000) de la base de données Dépouillement National (1993-2004).

La base de données Dépouillement National est également accessible via le portail TEL des bibliothèques nationales européennes dans le cadre du projet TEL-ME-MOR soutenu par l'Union européenne en 2005 :



Les problèmes que nous avons rencontrés sont les suivants :

- Parfois un fichier quotidien est transféré à deux reprises ;
- Les liens ne sont pas fournis pour tous les articles d'un périodique parce que l'éditeur considère que certains articles ne sont pas destinés à un usage public ;
- Les utilisateurs, bien entendu, considèrent que l'information devrait être entièrement gratuite. Tous ces problèmes n'affectent pas l'ensemble du processus en profondeur.

Avantages :

- l'accès aux publications et à leurs métadonnées est fourni de la même manière ;
- la BNL reçoit les métadonnées extraites (notices brèves) des articles électroniques du jour ;
- la qualité des notices est améliorée (par exemple, les erreurs qui pourraient survenir lorsque l'on recopie à la main un titre ou un nom d'auteur sont éliminées) ;
- les éditeurs reçoivent des données statistiques fiables sur les articles et les auteurs auxquels les utilisateurs ont le plus fréquemment cherché à accéder ;
- les éditeurs et Lursoft attirent de nouveaux utilisateurs ;
- les utilisateurs peuvent facilement passer du texte intégral de l'article à l'information liée (par exemple, aux personnes citées dans l'article) ; la connexion avec l'information liée dans les bases de données des entreprises et autres sources officielles est fournie par la même interface ;
- un kiosque d'information en continu est donc rendu disponible aux utilisateurs.