**Shopping for current day news in library database.
Is it possible?**

Ilona Dukure
ilona.dukure@lnb.lv

Anita Goldberga
anita.goldberga@lnb.lv

Meeting:	123 *SI* - Cataloguing
Simultaneous Interpretation:	Yes

*Summary. The report reflects the experience of the National Library of Latvia and IT company ''Lursoft'' implementing the co-operation project **Latvia's Model for Periodicals' Content Distribution**. The aim of the project is to facilitate users' access to the full text information of daily periodicals. The authors analyse the potential of the national library to perform bibliographic control of e-resources, and the possibility to exploit the metadata generated automatically from the resources themselves in the library's database. The advantages of the model from the viewpoint of users and co-operation partners, and the problems still to be solved are pointed out.*

I come from Latvia which is one of the Baltic countries situated just in the middle - between Estonia and Lithuania. Our neighbours are also Sweden (in the West) and Russia (in the East).

Ways of organizing the bibliographic control of the web resources on metadata level:

1. To record web resources applying a library information system;
2. To design or to adapt a separate system for the control of digital resources (recording in *Dublin Core)*);
3. To apply the metadata delivered by publishers;
4. To make an attempt to extract and generate metadata from resources automatically;
5. To apply a combination of all the versions mentioned above;
6. To publish the index of the archived and available resources on website;
7. To use more or less structured library metadata for some portal or for the search engines like *Google*, in Latvia – *Siets.*

To my mind there are several ways of organizing the bibliographic control of the web resources on metadata level:

1. To record web resources applying a library information system (usually MARC format);
2. To design or to adapt a separate system for the control of digital resources (recording in *Dublin Core* or some other minimally structured format);
3. To apply the metadata delivered by publishers;
4. To make an attempt to extract and generate metadata from resources automatically;
5. To apply a combination of all the versions mentioned above considering the differing nature of resources;
6. To publish the index of the archived and available resources on website or in other similar way;
7. To use more or less structured library metadata for some portal or for the search engines like *Google*, in Latvia – *Siets.*

Implementations in NLL

1. Traditionally: applying the Library information system Aleph500;
2. Supplying preliminary metadata delivered by publishers about forthcoming publications at the moment of ISBN assignment to the national bibliography database;
3. Passing library metadata for TEL portal (project TEL-ME-MOR) and Latvian full text search portal SIETS

The National Library of Latvia (NLL) has been applying three of the possible ways given on the first slide for the control of the digitally born resources. The first way on that slide is the traditional way. Next, there is ongoing co-operation with publishers concerning the assignment of ISBN numbers (in future also of ISSN numbers) to forthcoming publications in remote access and the supply of preliminary metadata about these publications to the national bibliography databases employing the Internet facilities. We are implementing the last way noted on that slide within the project TEL-ME-MOR together with other national libraries in Europe.

We are not able to *"cover"* the Internet (in the sense of bibliographic control) without respecting the opinion and activities of publishers. There are active e-publishers in Latvia, but there are rather few of them among the book publishers. Therefore for the first part we focussed on serials. Today I would like to discuss more in detail the 4th way in our – Latvian version. We have named it **Latvia's model for periodicals' content distribution**.

Latvia's model for periodicals' content distribution (1)

- Collaboration project among NLL, company Lursoft IT and "Culture Information Systems" (financial support)
- The database of the National analytics contains meta-information on the articles of the newspapers, magazines and journals published in Latvia in 1989 – 2006 (about 1 500 000 records) http://www.lnb.lv/eng/db/nat_bibl.htm

In 2005 the NLL initiated a co-operative project with the aim to provide the Internet access to the full texts of newspapers, magazines and journals while searching in the National Analytics database of NLL. This database contains meta-information on the articles of the newspapers, magazines and journals published in Latvia in 1988 – 2005 (about 1 500 000 records). The system Aleph500 (version 505.14.2) facilitates the display of content via the Internet http://www.lnb.lv/eng/db/nat_bibl.htm.

The National Analytics database was started in 1996. Since 2002 the records are produced in the information system Aleph500 applying MARC 21. The data from the previous local system (666 700 records) have also been converted into Aleph500.

- The National Analytics database:
  - supplies topical information from Latvian periodicals published in Latvian, Russian, English, etc.;
  - covers articles from about 450 periodicals.

The National Analytics database is searchable according to the author, title, and the source of the article (the title of the serial), keywords, names of persons mentioned in the article, and other criteria. This database is widely used in Latvia and abroad as it supplies topical information from Latvian periodicals published in Latvian, Russian, English, etc.

Latvia publishes about 600 serials annually, and two thirds of them are catalogued in the National Analytics database. Comics, prospectuses, crossword puzzles, programmes of TV and radio broadcasts, and the like, are not processed analytically.

The idea of accessing the full texts of articles from the National Analytics database had occurred earlier, but in May, 2005 we could start the implementation of it.

In 2004 the state agency ''Culture Information Systems'' administering the project of the national joint library information system posted a price poll on providing the linkage of bibliographic descriptions with full texts. As a result it made the agreement with Lursoft IT on providing full text access from the National Analytics database.

- Company Lursoft IT (http://www.news.lv)
  – Digital newspapers library
    – offers access to 43 serials and reviews of press agencies;
    – facilitates reading of full texts;
    – provides connection with the linked information in the databases of enterprises and other official registers by the same interface

The company Lursoft IT (http://www.news.lv) – our collaboration partner – is compiling the full-text archive of the electronic newspapers, magazines and journals published in Latvia. The archive contains the digital copies of all the main central and regional press publications acquired legally on the basis of Lursoft's IT agreements with each of the publishers. There is more than 2707000 full text articles available in Lursoft IT newspaper library and this amount is expanded by ~ 1000 every day.

In 2006 Lursoft IT offers access to 43 serials and reviews of press agencies from the Lursoft Digital newspapers' library. It facilitates reading of full texts, though charts, graphs and drawings are not added to the texts. The Digital newspapers' library does not contain photos, prospectuses, advertisements.

In order to exploit the potential of the electronic publications as efficiently as possible it was decided that Lursoft IT would provide not only URLs but also other data that can be acquired from the lay-outs of the serials.

NLL defined the requirements for the descriptions to be delivered – format, layout, coding.

In May 2005 Lursoft IT started the automated delivery of electronic descriptions of the published articles. These **metadata are daily extracted automatically** from the newspaper lay-outs delivered to Lursoft IT by publishers.

Lursoft IT has agreements with publishing houses (list on left side http://www.news.lv) and receives full lay-out sheets of newspapers since 1993. Articles automatically (and partly manually corrected) are extracted from layout sheets and converted in structured XML database entries using specially developed software by Lursoft IT. With specially developed tools, some of the data (Bibliographic data elements) are converted and delivered to NLL.

Technology, including XML database management tools, are developed by Lursoft IT (for more information, see http://www.siets.biz) and used in many local and international news / official documents' databases
(Also see: http://www.siets.biz/solutions/usecases/ and http://www.siets.biz/company/)

The descriptions contain data on the author of the article (if indicated), the title of the article, the title of the magazine, journal or newspaper, its date and number, and a unique identification for each article. The descriptions of appropriate publications of the current day are delivered in pre-defined format (Aleph Sequential – for downloading into library information system) till 3 a.m. daily.

The administrators of the information system Aleph500 in Latvia (the agency ''Culture Information Systems'') provides the import of the daily-received files into the system Aleph500. A separate database was created for this purpose, as not all the received records are included in the National Analytics database. Only about a half of the received descriptions of the central

press publications, and about a fourth to a sixth of the descriptions of the regional publications are subjected to cataloguing. Only significant, valuable articles are selected and catalogued - those concerning research, politics, economy, history and local studies.
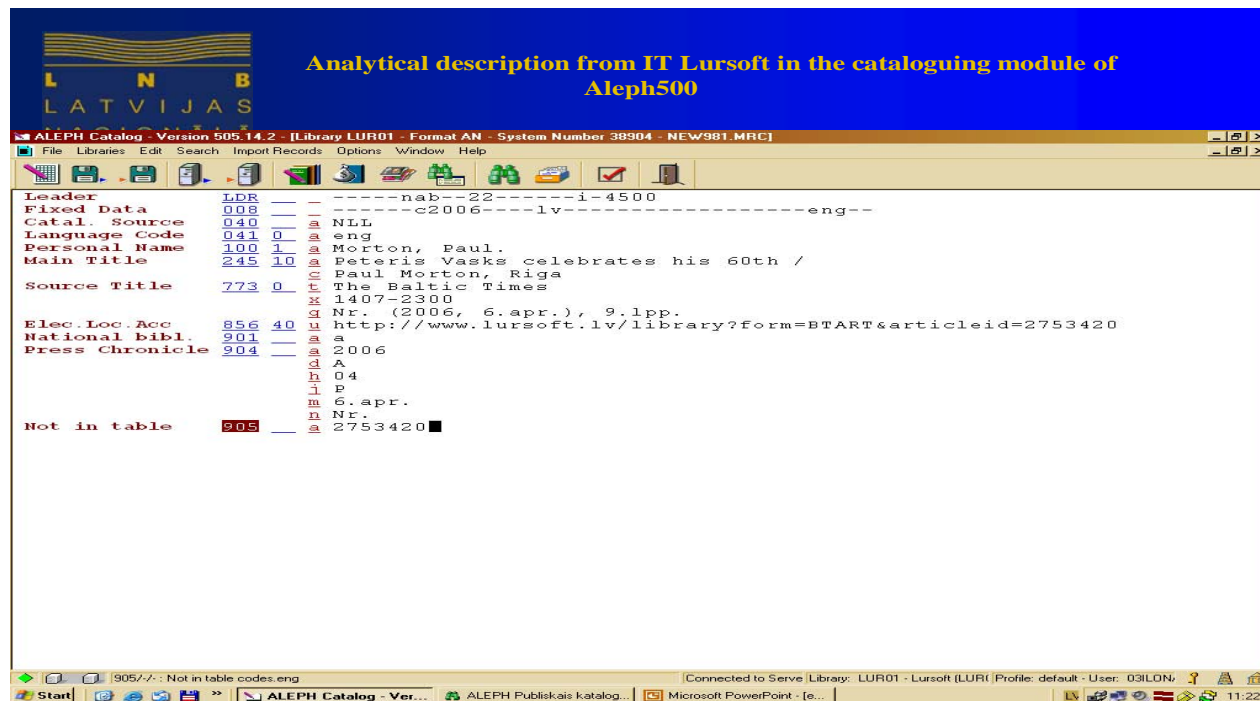
This is what the analytical description looks like when received from Lursoft IT in the cataloguing module of Aleph500.

List of MARC data elements supplied by Lursoft IT

Leader	LDR
Fixed Data	008
Cataloguing Source	040
Language Code	041
Personal Name	100
Main Title	245
Source Title	773
Electronic Location Access (URL)	856
National bibliography designation	901
Local data for national bibliography publishing	904
Lursoft identifier for article	905

There are encoded fields of fixed length to be filled for the description; the data import into the database cannot be performed without them. Additionally we have requested that Lursoft IT include such fields as, for example, the source of cataloguing, language, national bibliography designation, URL. The presence of these fields in metadata format decreases the amount of cataloguers' work, as they do not need to be added manually.

Latvia's model for periodicals' content distribution (5)

- Authorised NLL staff members have the right:
    - to include these descriptions into the National Analytics database;
    - to enrich the received information according to the requirements of MARC21.
- The database of the National Analytics is located on NLL server, but the full texts – on Lursoft's IT servers.

Authorised NLL staff members have the right to include these descriptions in the National Analytics database and enrich the received information according to the requirements of MARC21.

Analytical record processed by the cataloguer in the cataloguing module of Aleph500

In this slide we see the record already processed by the cataloguer – UDC index, annotation of the article, added description fields, key-words, data about the supplement of the publication.

The database of the National Analytics is located on our server; the full texts are on Lursoft's IT servers.

.

How to access the full text of an article?

- Get the internet connection to NLL home-page http://www.lnb.lv/eng/db/nat bibl.htm , find the menu ''Databases'', and select – ''National Analytics database''.
- The linkage of the National Analytics database record with the respective full text is performed using the unique identification of the article (See field 856).

When the user's query *via* the Internet is received, the linkage of the National Analytics database record with the respective full text is performed using the unique identification of the article (See field 856.)   The access to the National Analytics database is free of charge, to full texts – according to the agreement with concrete publisher (authorised access). This way copyright law is respected as the digital archive of Lursoft IT contains legally acquired information (See field 856).

How does a user access the full text of an article? We get the Internet connection to NLL home-page http://www.lnb.lv/eng/db/nat bibl.htm , find the menu "Databases," and select "National Analytics database."

Result list of records in National Analytics database

We select the necessary information option (for example, the criterion ''title''), and search for the articles about hockey. The search has resulted in 62 records.

Records in National Analytics database with URL

We open one of records and see that there is the indication about the location of the article.
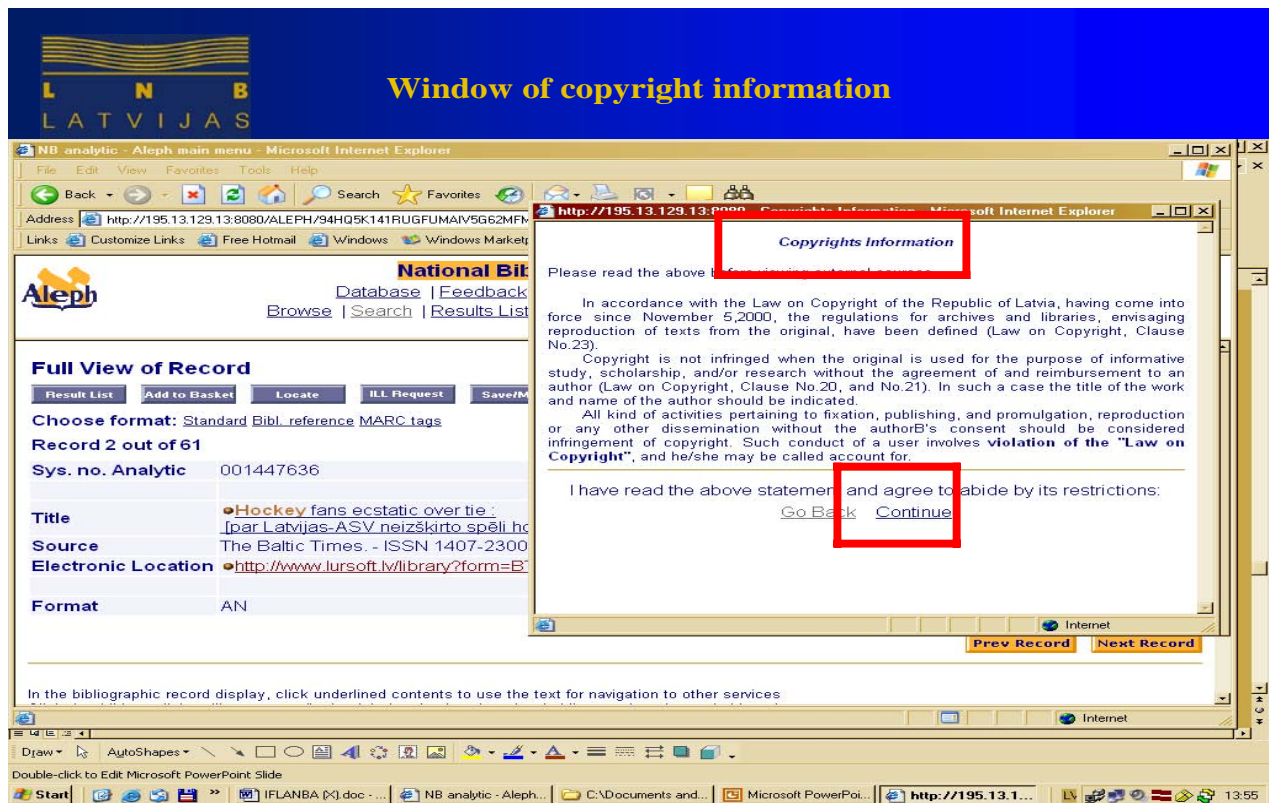
When clicking on this link the window of copyright information opens.

Access to Digital newspapers' library of IT Lursoft
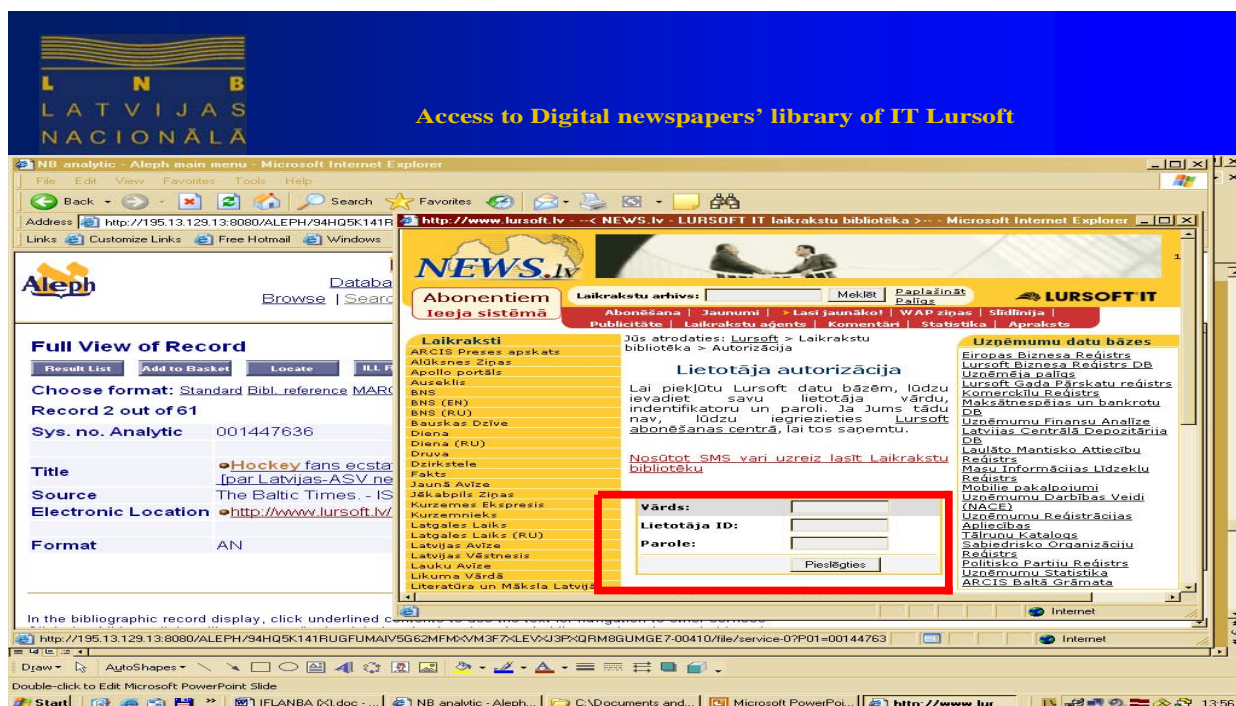
We get to the Digital newspapers' library of Lursoft IT. Authorized users enter their ID and password.

Article in Digital newspapers' library of IT Lursoft
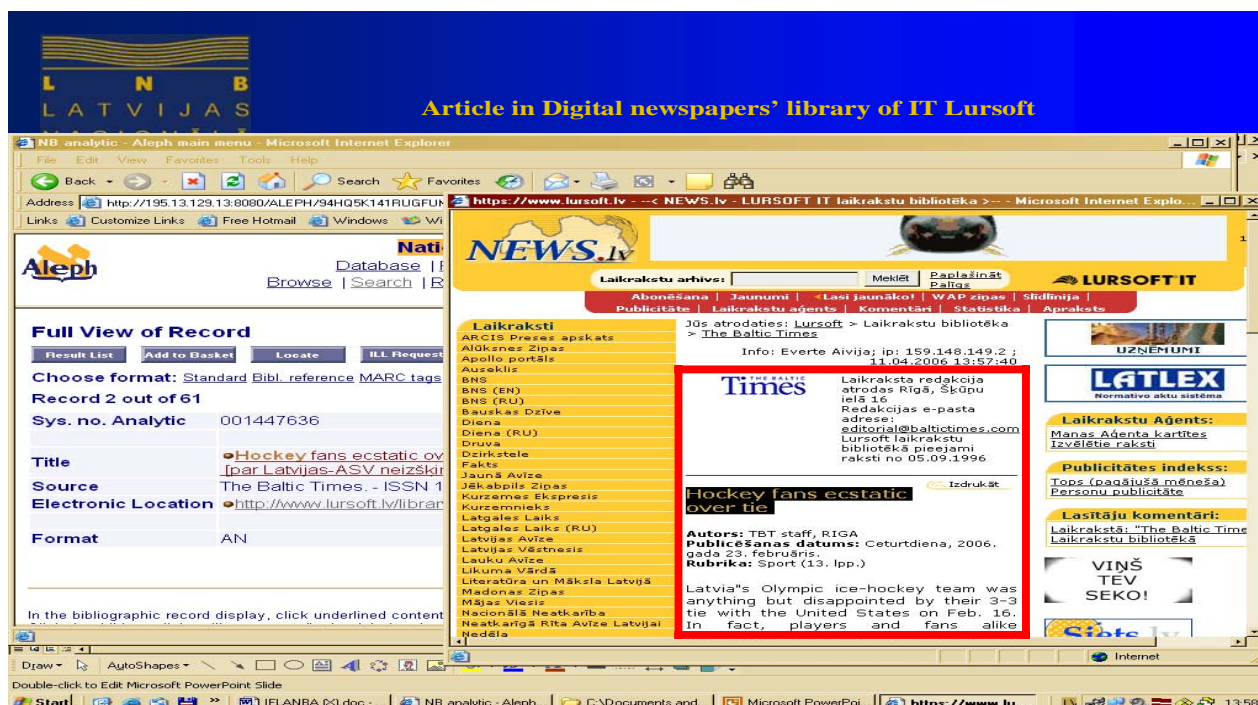
As soon as this is done, the full text of the required article becomes visible.

The cooperative project is at the stage of ongoing development. At present we are continuing to form linkage between the retrospective part of the full text newspapers' archive and the existing records (~ 400.000) of the National Analytics database (1993-2004).
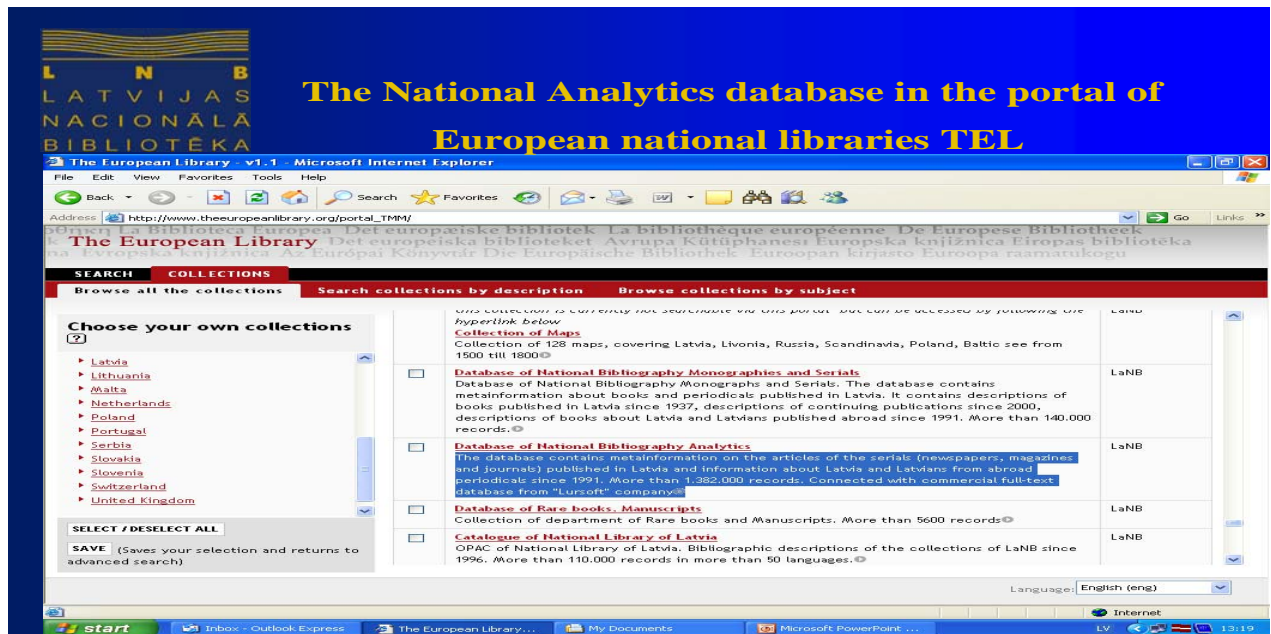
The National Analytics database can be accessed also *via* the portal of European national libraries ''TEL'' within the framework of the project TEL-ME-MOR supported by the European Union in 2005.

Problems:
- Sometimes a daily file is transferred twice;
- The links are not provided to all articles of the serial because the publisher considers the article is not to be offered for public use;
- The users, naturally, consider this information should be completely free of charge.

The mentioned problems do not influence the process substantially.

Benefits:

- **publishing – metadata – access is provided in the same day;**
- NLL
  - receives extracted metadata (short records) of the current day e-articles;
  - the quality of records improves – grammar errors that could occur when rewriting title or author manually are eliminated;

Benefits (2) :

- **publishers** get reliable statistical data on the most frequently accessed articles and authors**;**
- **publishers and Lursoft IT** attract new users;
- **users** can switch easily from the full text of the article to the linked information**;**
- non-stop information shop is open for **users.**

- **publishers** get reliable statistical data on the most frequently accessed articles and authors**;**

- **publishers and Lursoft IT** attract new users;
- **users** can switch easily from the full text of the article to the linked information (e.g., persons cited in the article)**;** connection with the linked information in the databases of enterprises and other official registers is provided by the same interface;
- non-stop information shop is open for **users.**

National Analytics database

- **The database of the National analytics**
  http://www.lnb.lv/eng/db/nat_bibl.htm

- **Lursoft IT** **http://www.news.lv**

Ilona Dukure **ilona.dukure@lnb.lv**
Anita Goldberga **anita.goldberga@lnb.lv**

2006.06.09.

Ilona Dukure **ilona.dukure@lnb.lv**

Anita Goldberga **anita.goldberga@lnb.lv**