Vesna Aleksandrović
Audiovisual and Printed Music Department
National Library of Serbia
Ivan Pešić
Yugoscan Ltd.

DATA MINING – REVEALING THE SOUND RECORDINGS METADATA MEANING

Abstract. What brings together bibliographic record (its format and content), labels, names, gramophone 78 rpm record, matrix numbers, roles, subjects? In the case of a digital library and information accessibility, that would be certainly – metadata. Every librarian knows the meaning of information. Things work right when you do have information to offer, and get. The problem arises when information is limited and almost does not exist.

Now and here, in our hands are so valuable and precious, words, tones, tunes, information, voices of the past, written on one of the earliest sound carrier – 78 rpm gramophone record, representing cultural heritage of the mankind. Besides audio data, every single gramophone record has its story *behind the scene,* which we can hardly find even in old and dusty catalogues of gramophone records publisher houses, archives or in the memories of unique but informal, *The 78 rpm Gramophone Records Collectors and Fans* union.

Our obligation is to find, explore and represent these data, such as recording and publishing date, location, matrix numbers, to resolve pseudonyms, initials, find out who composed or wrote lyrics or libretto and many other things which vary from record to record. Furthermore, these discoveries have to be put in some user friendly form and system, and publicly presented.

The topic of this paper is information extraction for the purpose of wider digital object denotation and presentation. The present COBISS2 platform (used in National Library of Serbia) does not have an appropriate data export format suitable for this type of material. Therefore, we devised a process which extracts all necessary fields and subfields from a record, apply further processing of data, and store it in an XML file. We also developed an XML scheme for internal purposes which is used in MapForce mapping to represent the metadata in the final XML format. Also, there is a plan to create an additional mapping into DUBLIN CORE compatible format.

Keywords., 78 rpm, Gramophone records, Bibliographic record, Mapping, Data transformation, XML Scheme

What are data? Could we be able to signify data without any interchange with larger systems that include and exclude it, at the same time? Could it be just data? In the age of acceleration on every level of being, we are so occupied with finding, gathering and information organizing. What doest it mean, in fact? It means that humankind longs to find fresh and new meaning by reorganizing known fact and data. This is the path of change.

Nowadays, being librarian puts you into a role of information creator and presenter, but user, as well. That means that holding information in your hand does not approve it if you do not share it and pass it on. What we can do to present information, gather it and reorganize into new, much more adoptable way? First, being aware of information we do have. However, it is only information, without any meaning if we do not give it a meaning. The second step is to gather similar types of meanings and create a higher system, which could be beneficial to one who seeks the information. Therefore, we have to pay close attention to foresee every single chance to help the seeker to find what he is looking for.

Here is one example of not thinking in forth. Look at information we get on 78 rpm record. Usually you can find on record label title of recorded work, mostly present is the name of composer and performer, publisher and publishing place. Oftentimes, there is no sign of

date of pressing or recording, who made the recording and where. Some would say that this is not important. Maybe it is not, but maybe it is for someone in this world. Our goal is to keep mind on that starting point – if you do have information, pass it on, and if you do not, your task and duty is to find it. If there is no possibility to find much more (at the time), just accept it and gather all data you have to something useful.

Sometimes, being on this adventurous tour of exploring something which happened times ago, finding additional information and giving it a new meaning, sounds like being and Egyptologist, found oneself in front of splutter signs which one just can not gather to achieve recognition of any known pattern or knowledge. Nevertheless, we do like challenges, don't we? Yes, we do. It moves us towards a closer connection to unknown and unfamiliar, giving us an opportunity to shape it and read it. That is the path of every single dedicated early sound recordings explorer and collector.

Historical notices and their influence at present

Early sound recording industry started nearly 130 years ago. Great invention, the gramophone, indeed. Medium capable of recording and reproducing sounds, voices, music and spoken words. The moment frozen in time, captured and stored, for good. As long as the record exist on physical level.

In the light of mentioned not thinking in forth, today we have so little information concerning the concrete record on its label. However, we do want to find more. Much people do not know the secret world of signs that are stored on each gramophone record, beyond the sound itself, which reveals additional meaning and information, just as guidelines to finding out its history.

Catalogue and Matrix numbers. Early recordists engraved their own initials on the records. Some put additional, obscure signs, numbers, letters, and characters. On the level of major publishing houses, in the beginning of sound recording industry, there were secret codes, some have been revealed, some not. For example, what does it mean - WCG.3974-R? Alternatively, 054199? Is it some random line of letters and numbers? In fact, it has a meaning. WCG are initials of Will Gaisberg, one of the sound recording pioneers, and numbers represent series of his recordings, allocating concrete recording. Recordist such as Fred and Will Gaisberg, Franz and Max Hampe, William Sinkler Darby and many others, concluded that every record has to be given a closer prefix, to gather much more information concerning recording.

Catalogue number allocates concrete gramophone record in order as they were pressed and its place in publishing house catalogue. Along with it, it also represents many other and valuable *information:*

„The Gramophone Company used a rather complex system for assigning catalog numbers to their disks. This involved the use of a 5- or 6-digit number. Counting from the right, a zero in the sixth position indicated a 12-inch record, while its absence indicated a 10-inch record. The fifth position was used to indicate the language of the recording, i.e., 2=Russian, 3=French, 4=German, 5=Italian, 6=Spanish, etc. The digit fourth from the right indicated the voice or instrument, as 2=male, 3=female, 4=concerted, 5=piano, 7=violin, and so forth. This system appears to have been in effect from the first recording of Syria Lamonte in November 1898, and was used consistently throughout the period of single-sided discs."[1]

---

[1] Matrix and Catalog Numbers in G&S Discography - *Gramophone Company Matrix and Catalog Numbers*, by Howard S. Friedman.

Matrix number is somewhat different data and it has close connection to recording. It is well know that a recording has to be shaped through at least several takes, let just take it as some kind of rehearsal.

„Commercial recording began in London on August 2, 1898, almost immediately after the Gaisbergs arrived. Between that date and November 1, no matrix or serial numbers of any kind were assigned to recordings, which were listed by catalog number only. Matrix numbers were introduced on or about November 1, 1898 (although they may have been used earlier). Each matrix plate was identified by a single number, which sometimes was modified by a prefix containing both numbers and letters, or by a suffix of similar makeup, or by both. The matrix number alone provided information as to who, what, and generally when, while the prefixes and/or suffixes, if any, indicated where and how. Either the prefix or the suffix or both could be separated from the main number by a hyphen. The suffix could contain either normal or superscript Roman or Arabic numerals..”[2]

This explains the meaning of numbers on His Master's Voice or sisterly publishing houses. Every publishing house developed its own and unique system. There are so numerous examples of signs, but we keep forward, to expose other severe problems, which we find on the path of gramophone record identification.

Recording speed. 78 revolutions (or rotations) per minute describes frequency of plate rotation during recording and reproducing. It is mostly unfamiliar that rotation speed goes from 60 to 90 rpm. It depends on several issues: publishing house and its technical equipage, period of recording, used stylus and many other things. What do we have as information on the label? Mostly it is 78 rpm. Only Columbia record had a mark 80 rpm.

Initials and unresolved pseudonyms. When we see on the label Mozart, Caruso, Rich. Strauss, or any similar well known name, or last name, we do not have a problem to resolve it and put it in predominant name form into bibliographic entry.

What happens in case like this – E. Mullot? Or, perhaps, the label is illegible, and half of title or information on author/s and performer/s you just can not see and get? On the scene is step-by-step technique: first of all, accept what you have and place it in the appropriate field in bibliographic record. Then you should explore other sources to find valuable and wanted information. If you can not find any kind of information, it is time to leave it behind, the information will come, definitely.

Recording/record date. Open a book, newspaper, CD, DVD near you. Notice that all of them has information on year when it is published, even more about first issuing, copyright date et. Dealing with aged materials such as old, rare books, maps, gramophone records, postcards or photographs, sometimes can be so frustrating to a librarian since such information does not exist in the publication. The same technology marches again, investigation within books, anthologies, periodical publications, advertisements, Internet, catalogues, consultations with other researchers, scholars etc.

Title and other information translations. For the purpose of full information offered to user, it is highly suggested to enter all titles: multi-language, multi-script, head or inside title, and alternative, inserted, original, collective, translated, parallel, variant and additional title information, as well. This requires data investigation, sometimes it goes easy, sometimes is not. Nevertheless, a librarian aware of bringing full information importance will lightly find the way to do so.

---

[2] Matrix and Catalog Numbers in G&S Discography - *Gramophone Company Matrix and Catalog Numbers*, by Howard S. Friedman.

Subject and topical heading. On some labels one could find complete description of the work recorded, such as Beethoven – III Symphonie, Es-dur, Eroica op. 55, 2. Satz, Marcia funebre / IIIrd Symphony E flat Major "Eroica", 2nd mov.ement, or, in case of Serbian music something like this – Mokranjac, II rukovet, peva Akademsko pevačko društvo „Obilić", dirigent Lovro Matačić. That is easy work for a librarian, who has precise title, music genre, and many precious information concerning subject and topic defining, which could help user to find what he is searching for.

Problem arises when there is no precise information but title and performer, or composer. If the record is not digitized, things start to be more complicated. Defining more than one subject and topic heading will be helpful.

Bibliographic issues

Being librarian involves many various skills, knowledge, and of course, different standards which have to be followed. And it is good, as far as following these standards does not deny its purpose. Let's take a look at this full bibliographic record:

001 [a]c - ispravljeni zapis [b]j - zvučni snimci, muzički [c]d - izvedeno delo [d]0 - nema hierarhijskog odnosa [7]vv - više pisama

07121[a]A 192503

100 [b]f - publikacija s procenjenom godinom izdavanja [c]19?? [h]scc - srpski [i]b1 - transliteracija COBISS za ćirilicu [l]cb - ćirilica - srpska

1010 [a]scc - srpski

102 [a]deu - Nemačka

126 [a]a - gramofonska ploča [b]d - 78 o/m [e]d - 10 in (25,4 cm) [i]a - akustična

128 [a]ar - arije (vokalne ili instrumentalne kompozicije koje nisu deo veće celine) [b]oa - simfonijski orkestar [c]vd - tenor

2001 [a]Kavalerija rustikana [b]Zvučni snimak [i]Vino se u čaši peni [e]Vinska pesma [f][Pietro] Maskanji [c]Pajaci [i]O Kolombina--- [e]Harlekinova serenada [f][Ruđero] Leonkavalo [g][izvodi] Živojin Tomić, tenor [g]prati orkestar Državne Opere u Berlinu

210 [a]<LAT>Germany [c]<LAT>Odeon [d][19--?] [e]<LAT>Germany [g]<LAT>Odeon

215 [a]1 gramofonska ploča [d]25 <LAT>cm

3000 [a]Ploča je bez originalnog omota

3000 [a]Na etiketi ploče oznake <LAT>Vse 555 i <LAT>Vse 538

3000 [a]Crvena etiketa.

423 0[1]2000 [a]O Kolombina--- [1]700 1 [a]Leonkavalo [b]Ruđero [4]230 - kompozitor

6063 [a]Arije [x]Tenor, orkestar [w]Gramofonske ploče

675 [a]784.2.087.613.1:785.11(086.72) [s]78 [b]78 [c]784 - Vokalna muzika. Narodne pesme. Pevački horovi

70211[a]Maskanji [b]Pjetro [4]230 - kompozitor [6]01

70211[a]Tomić [b]Živojin [4]590 - izvođač

71202[a]Orkestar Državne Opere [c]Berlin [4]590 - izvođač

830 [a]78rpm

90213[a]<LAT>Mascagni [b]<LAT>Pietro [6]01

992 [b]9505N343, oz, 9505K107, 0501c107,0501c107,78rpm,0909nd107,lukovic1,

99607[d]lD\f2\n304\sSt [v]e - stari fond [f]600901594 [p]7 - potpuna nedostupnost (arhivski primerak ...) [o]20090928

99607[d]lD\f2\n304\sSt\dm [v]e - stari fond [f]600500283 [o]20050124 [p]7 - potpuna nedostupnost (arhivski primerak ...)

996 7[g]tra\oji [d]lCD\n3473 [f]800400705 [v]f - sopstveno izdanje [o]20090928 [p]4 - ograničena dostupnost - čitaonica

or, this one

001 [a]c - ispravljeni zapis [b]j - zvučni snimci, muzički [c]m - monografska publikacija [d]0 - nema hierarhijskog odnosa [7]vv - više pisama

07121[a]Z. 1007

100 [b]f - publikacija s procenjenom godinom izdavanja [c]19?? [h]scc - srpski [i]b1 - transliteracija COBISS za ćirilicu [l]ba - latinica

1012 [a]scr - hrvatski [a]ger - nemački

102 [a]hrv - Hrvatska

126 [a]a - gramofonska ploča [b]d - 78 o/m [d]x - [e]d - 10 in (25,4 cm) [i]a - akustična

128 [a]wz - valceri [a]df - plesni oblici (pojedini plesovi, osim mazurke, menueta, pavane, polke i valcera) [b]of - duvački orkestar

2001 [a]Rosen aus dem süden [b]Zvučni snimak [e]valcer [f]J.[ohann] Strauss [c]Hrvatski plesovi [f]F.[ranjo] S.[erafin] Vilhar [g][izvodi] Muzika Savske diviz.[ione] oblasti [g]kapelnik I.[vo] Muhvić

210 [a]Zagreb [c]Edison Bell Electron [d][19??] [e]Zagreb [g]Edison Bell Electron

215 [a]1 gramofonska ploča [d]25 <LAT>cm

3000 [a]Ploča je bez originalnog omota

3000 [a]Na etiketi ploče oznake Z. 85 i Z. 86.

3000 [a]Crna etiketa.

423 [1]2000 [a]Hrvatski plesovi [1]700 1 [a]Vilhar [b]Franjo Serafin [4]070 - autor

6063 [a]Valceri [x]Orkestar, vojni [w]Gramofonske ploče

675 [a]785.12.085.2(086.72) [s]78 [b]78 [c]785 - Instrumentalna muzika. Orkestarska muzika. Kamerna muzika. Džez

70211[a]Štraus [b]Johan [4]230 - kompozitor [6]01

70211[a]Muhvić [b]Ivo [4]250 - dirigent, horski dirigent

71202[a]Muzika Savske divizione oblasti [4]590 - izvođač

90213[a]Strauss [b]Johann [6]01

992 [b]9505N343,oz,9505K107,0505c107,78rpm,0909nd107,lukovic1,

99607[d]lD\f2\n173\sSt [v]e - stari fond [f]600502928 [o]20050517 [p]7 - potpuna nedostupnost (arhivski primerak ...)

996 7[g]tra\oji [d]lCD\n3315 [f]800400563 [v]f - sopstveno izdanje [p]4 - ograničena dostupnost - čitaonica [o]20090907

It is so obvious that we put so many information concerning this very record, music, composer, performers, but is it readable and clear to user? Therefore, our intention was to reduce all things which are unnecessary and to focus upon all wanted data, gather them and present in user friendly, readable form. Bibliographic standards prescribes so many rules quite appropriate for that very purpose, entering various data… Resolving all names, initials, titles, put in a record in strictly defined format. For example, in the case of resolving initials and names, there is a rule to put it in one field of record along with square brackets, and it looks like this: [f]J.**[ohann]** Strauss. The *f* is subfield in 200 field, and on the record we have J. Strauss, but for we know that this one is the work of Johann Straus, we resolve it in square brackets.

There are so many examples like this concerning titles, names, publishers, publishing years, some really annoying signs put in the record like **LAT** in angle brackets, which tells the system that data we have entered is written in Latin: **<LAT>**Germany [c]**<LAT>**Odeon

We came to conclusion that bibliographic and other standards do not cooperate on so many levels, and from some kind of frustration which aroused in observing that there is no way to translate records from our COBISS database in something useful, we decided to take a journey called *data extraction.* We have decided which data are wanted, and what had to be eliminated. Doing so, we have considered so many things like Dublin Core standard, some really helpful portals like DISMARC, and of course, wide experience with users, their needs, and the point of view of collectors, scientists etc. And then, when all things became clear, we could start to do our data extraction work.

Data transformation

The idea for this project came from the fact that the present COBISS2 platform does not allow an convenient way to extract and map all data relevant for describing a 78 rpm gramophone record. The final goal was to create an XML file with appropriate structure that will contain all necessary data. It may seem a pretty straightforward task, but there were some quite difficult problems that we needed to solve. The software chosen for this task is the MapForce from Altova. Also, the XML scheme was created using their XMLSpy editor.

The first step was to retrieve records from the COBISS database. The only way to get full records is using a full export format. The system sends to user an email with UTF-8 encoded plain text file containing full records in an attachment. Here is a brief description of the file format.

```
    1. ID=15371788
    ¶
    001 [a]c - ispravljeni zapis [b]j - zvučni snimci, muzički [c]m - ¶
     monografska publikacija [d]0 - nema hierarhijskog odnosa [7]vv - više ¶
     pisama¶
    07121[a]<LAT>AA 176033¶
    100 [b]f - publikacija s procenjenom godinom izdavanja [c]19?? [h]scc - ¶
     srpski [i]b1 - transliteracija COBISS za ćirilicu [l]cb - ćirilica - ¶
     srpska¶
```

Each record consists of a header and several fields. The header is separated from the fields with *one* empty line (line breaks are shown here with a pilcrow sign which is not present in the actual file). Records are separated with *two* empty lines. Each field starts with 3 digits, followed by two optional flag digits. Subfields are marked with lower case letter or digit enclosed in square brackets, and separated with *two* horizontal spaces. You can see from the example above that a field can span several lines. These lines begin with 5 horizontal spaces. In order to make parsing as easy as possible, we have decided to remove line wraps, and reformat the file, so that every field stays in just one line, no matter how long the line is. In fact, the line wrapping is necessary because the file is sent by e-mail. One can use any decent text editor to do this. Here, we have used Vim. After carefully analyzing the file, we have come up with these 3 substitute commands which need to be executed in displayed order:

```
  %s/\n \(\[[A-Z]\)/\1/g
  %s/\n \[/ [/g
  %s/\n \([A-zšđčćžŠĐŽČĆ0-9()#\-"]\)/\1/g
```

The result is a file which is suitable for the next step – parsing:

```
1. ID=15371788
¶
001 [a]c - ispravljeni zapis [b]j - zvučni snimci, muzički [c]m - monografska publikacija [d]0
- nema hierarhijskog odnosa [7]vv - više pisama¶
```

07121[a]<LAT>AA 176033¶

100 [b]f - publikacija s procenjenom godinom izdavanja [c]19?? [h]scc - srpski [i]b1 - transliteracija COBISS za ćirilicu [l]cb - ćirilica – srpska¶

     MapForce is a graphical data mapping, conversion, and integration tool. In Enterprise Edition, it comes together with a FlexText, a module that allows advanced processing and mapping of legacy text files. FlexText is used to make a configuration for input or output component in a MapForce mapping. After a careful inspection of the structure of the file, there comes a simple task of applying the appropriate actions. But in this case, there were some major obstacles that we needed to overcome. For example, a subfield can have different meanings. Subfield [f] of the field 200 can denote a performer or an author. Also, the subfield [g] of the same field denotes a performer. So there isn't a one-to-one relationship between subfields and outputs of the component. The solution is to inspect the beginning of the information in the field and make one output for every case. For example, if the above mentioned subfield [f] begins with any of fourteen strings like "[izvodi] ", "[izvode] ", "s", "n", "g" and so on, the contents of the subfield appears on the appropriate output that denotes a performer: "izvode", "izvodi", "izvodi1", "izvodi2", "izvodi3" and so on. In case that the information do not match any of the rules present in the switch, it appears in the output which denotes an author. The important thing to note here is that the choices in this switch are exclusive. That means that the information in the subfield can match only one of the choices and appear on the determined output. Multiple outputs were necessary, because FlexText does not allow mapping of outputs from several actions to one output of a configuration.

     The next step was to create a XML shema for the final XML file. XMLSpy is the editor used to make this scheme: Many elements in this scheme are optional. One of the reasons for this is that there are gramophone records that still do not have all the information discovered. The metadata set is organised as a collection of records. Every record must have at least one track (element 'naslov') with it's title written in the attribute 'ime'. There are also several optional elements that store some more information about that particular track, such as subtitle, additional title, performer, author. There are also elements for publisher, publishing date and place, dimension, catalogue number, useful web link and several bibliografical data.
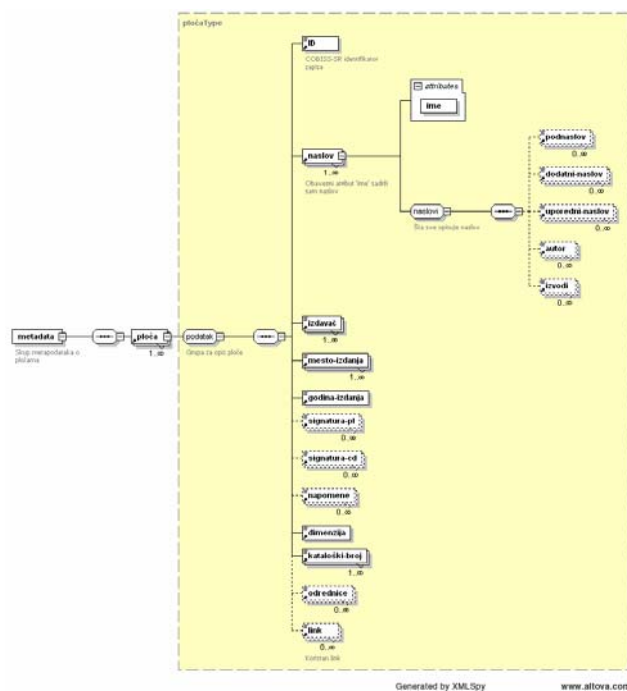
Now we have all that is necessary to make a MapForce mapping between input component defined by FlexText configuration, and the output XML file, defined by the presented schema. This included various substitutions and string formatting, as well as concatenation of strings from several ports of the input component. We had to remove some characters that were added to subfield data for bibliographic requirements, like resolving of names, shorthands, unknown places and years and so on. For example, the signature is entered as "lD\f2\n63\sSt". But we need to write "D II 63 St" to the XML file. All this was accomplished using MapForce built-in functions, and custom created user functions.

A sample of final XML file with metadata for two gramophone records looks like this:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="style1.xsl"?>
<nb:metadata xsi:schemaLocation="http://nb.rs/ns m.xsd"
xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:nb="http://nb.rs/ns"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
   <nb:ploča>
      <nb:ID>15371788</nb:ID>
      <nb:naslov ime="Mijatovke">
         <nb:dodatni-naslov>Razbole se belo Done</nb:dodatni-naslov>
         <nb:autor>Stanislav Binički</nb:autor>
      </nb:naslov>
      <nb:naslov ime="Behar">
         <nb:dodatni-naslov>Bumbul pjeva okolo Mostara</nb:dodatni-naslov>
         <nb:autor>Stevan Hristić</nb:autor>
         <nb:izvodi>Živojin Tomić, tenor uz pratnju orkestra Berlinske Državne
Opere</nb:izvodi>
      </nb:naslov>
      <nb:izdavač>Odeon</nb:izdavač>
      <nb:mesto-izdanja>Germany</nb:mesto-izdanja>
      <nb:godina-izdanja>19--?</nb:godina-izdanja>
      <nb:signatura-pl>D II 63 St</nb:signatura-pl>
      <nb:signatura-cd>CD 1800</nb:signatura-cd>
      <nb:napomene>Ploča je bez originalnog omota</nb:napomene>
      <nb:napomene>Na etiketi ploče oznake xxVse545 i xxVse 563</nb:napomene>
      <nb:napomene>Crvena etiketa</nb:napomene>
      <nb:dimenzija>30 cm</nb:dimenzija>
      <nb:kataloški-broj>AA 176033</nb:kataloški-broj>
      <nb:odrednice>pesme / glas, orkestar</nb:odrednice>
   </nb:ploča>
   <nb:ploča>
      <nb:ID>15380748</nb:ID>
      <nb:naslov ime="Mijatovke">
         <nb:podnaslov>srpske sevdalinke</nb:podnaslov>
         <nb:autor>Stanislav Binički</nb:autor>
         <nb:izvodi>svira Muzika Kraljeve Garde, Beograd</nb:izvodi>
      </nb:naslov>
      <nb:izdavač>Odeon</nb:izdavač>
      <nb:mesto-izdanja>Germany</nb:mesto-izdanja>
      <nb:godina-izdanja>19--?</nb:godina-izdanja>
```

```
        <nb:signatura-pl>D II 48 St</nb:signatura-pl>
        <nb:signatura-cd>CD 1801</nb:signatura-cd>
        <nb:napomene>Na etiketi ploče oznake Vse 444 i Vse 445</nb:napomene>
        <nb:napomene>Teget etiketa</nb:napomene>
        <nb:dimenzija>25 cm</nb:dimenzija>
        <nb:kataloški-broj>A 300379</nb:kataloški-broj>
        <nb:kataloški-broj>A 300380</nb:kataloški-broj>
        <nb:odrednice>sevdalinke / Srbija</nb:odrednice>
    </nb:ploča>
</nb:metadata>
```

MapForce can generate the code in Java and C++, so finally you get an application that performs the required mapping, and which can run on any machine without the need to install the MapForce on it.

For the purpose of using this metadata on the web presentation, the final XML file was transformed using a XSL stylesheet that demonstrates just one of many possibilities for presentation of metadata.

The XML sheme used in this project is made for internal purposes. There is a plan to develop it and make a Dublin Core application.

Literature

[1] http://gasdisc.oakapplepress.com/matrix1.htm
[2] http://ifla.queenslibrary.org/VII/s35/pubs/avm-guidelines04.htm
[3] *The Collector's Guide to Gramophone Company Record Labels 1898 – 1925*. Howard Friedman
[4] *Encyclopedia of Recorded Sound*. Frank Hoffman, Routledge, USA
[5] http://www.altova.com
[6] http://www.vim.org
[7] *Međunarodni standardni bibliografski opis (ISBD), uvodno objedinjeno izdanje*. Po preporukama Radne grupe za ISBD odobrio Stalni odbor Sekcije IFLA-e za katalogizaciju, prevela sa engleskog Ana Savić. Beograd, Narodna biblioteka Srbije, 2009.
[8] *Now That We've Found the 'Hidden Web,' What Can We Do With It? The Illinois Open Archives Initiative Metadata Harvesting Experience*. Timothy W. Cole, Joanne Kaczmarek, Paul F. Marty, Christopher J. Prom, Beth Sandore, and Sarah Shreeves, University of Illinois at Urbana-Champaign, USA
[9] http://www.archimuse.com/mw2002/papers/cole/cole.html

vesna.aleksandrovic@gmail.com
ivan.pesic@gmail.com