

**Koltay Klára**  
**Electronic documents in libraries: OPACs or gateways.**

**The most labor intensive document type**

Both information professionals and the general public agree that electronic documents are not only a convenient means of storing information but they offer fast retrieval of it, they have enhanced search capabilities that multiply the possibilities of creating links among pieces of information that were hitherto far away from each other in the universe of printed documents. On the other hand, however, users and institutions serving them must put strong emphases on similarly enhancing their capabilities of locating and presenting the relevant range of electronic information. If relevance is understood to mean both thematic relevance and direct accessibility of the document in any given situation the task is not easy. While once catalogued books are likely to remain in the shelf area allocated for them when cataloguers added their shelf numbers to the bibliographic records till they are withdrawn from the collection very often several decades later, electronic documents are not so predictable. One may sit on the self embodied in a CD-ROM, which is installed only on request, others can be accessed only at dedicated workstations or at any number of terminals within the local area network. Some others are online databases that require password identification or provide access by the IP address of the users' terminals. Electronic journals are even more difficult. Some are published on CD-ROMs, others are ordered as part of part of the a subscription package of mixed subject area, some are "free" as long as you have a paper subscription to the same title but they require user identification when accessed.

The electronic holdings of a library with all the intricacies of varying and quickly altering access information may be challenging for the librarians to keep pace with and it is even more difficult to lead the patrons to the required bit of information without making him feel himself in a jungle grown over its own keepers heads.

The method, that may give him the experience of a user friendly environment of well-structured information is as old as libraries: documents must be listed in one way or another with data of their accessibility.

In the case of networked documents there emerged two variant methods. Document descriptions are entered into the bibliographic databases, often into the library OPACS themselves, or they appear in alphabetic or thematic lists on library homepages. Both solutions have their advantages. Networked electronic documents in the library catalogue become an integral part of the library's entire collection, they can be retrieved with exactly the same well-known methods by the same indexes as the more traditional documents. A patron searching for a topic in the OPAC will find the electronic documents among the books and printed journals even if it never occurred to him that the library may provide him anything but books. It is especially important in the case of journals that the printed and electronic titles and holdings appear together as two complementary parts of the same collection. Descriptive and subject cataloguers follow their everyday routine and cataloguing practice when dealing with this type of material.

Listing online access electronic documents on homepages or thematic portals have other advantages. It is more plausible for the patron surfing the Internet to select and access information from Web-lists than trying to locate it first by a catalogue search even if the catalogue itself is on the Web and its interface provides a direct link to the networked document from the retrieved description of it. Lists, on the other hand can easily be edited specially to meet certain demands: they can be selected according to the interests of various user groups.

Whichever method we chose to present the library's electronic holdings we have to bear in mind that their maintenance requires permanent attention. It is not only the Internet documents that often change location when the institutional homepage of its publisher is redesigned, if its author changes his affiliation or he simply moves to another Internet provider. Quite a lot of them - as many as 20 % of Web-sites<sup>1</sup> – disappear completely within a year. Even the long-term services a library is subscribed to may change the method how they can be accessed. When a database installed locally on a workstation becomes popular, it may enter the library's LAN CD-ROM service or the library may subscribe to the online service of it; later as for a member of a wider library consortia, the same database may be provided by another member institution.

Patrons, however, are interested only in that they can get access for the whole subscription period and do not want to hear about and remember all service providers, library consortia and changing passwords. It is the task of the librarian not only to register the existence of the document but to follow all the altering access methods and lead the patron to the presently active access point without causing him extra problems.

### **A possible solution**

The Library of Debrecen University (Hungary) taking all the above problems experienced in everyday library practice into consideration has set up a system, which aims at providing some solution to them without putting too much burden on library staff. The model consists of the following elements:

1. Various types of electronic documents are catalogued in the library's ILS receiving UDC strings and subject headings just as other document types in the OPAC and instructions on how to access them together with their URL when relevant.
2. Records are saved into the online catalogue thus they appear integrated into the general holdings of the library for patrons searching the catalogue using familiar indexes and search techniques. The OPAC interface provides a direct link to the URLs recorded in the document's bibliographic records.
3. Based on the bibliographic and access information in the records there created Web lists of various types at certain intervals:
  - an alphabetic list by title
  - a thematic list based on UDC strings
  - a short list containing only the 10 most recent documents

The lists are located at the library's homepage.

4. A mechanism checks the URLs in the records periodically and warns librarians when a URL cannot be accessed.
5. Librarians administer all changes, corrections and deletions in the bibliographic records. The corrections are visible in the OPAC at once and the weblists are corrected when they are regenerated next time.

We hope that the procedure employs librarians' effort to create standard, exchangeable bibliographic records of lasting value and to enrich the library's catalogue an aim which can still be considered a widely accepted one<sup>2</sup> even in at the era when the overwhelming amount of Internet documents make information professionals experiment with more efficient,

---

<sup>1</sup> The figure is quoted by John Byrum in his lecture State of the art and unsolved problems at the Electronic Resources International conference (Rome, 2001) at [http://w3.uniroma1.it/ssab/er/relazioni/byrum\\_eng.pdf](http://w3.uniroma1.it/ssab/er/relazioni/byrum_eng.pdf)

<sup>2</sup> cf. Conference lectures at Bicentennial Conference on Bibliographic Control for the New Millennium (2000) at <http://www.loc.gov/catdir/bibcontrol/> and Electronic Resources International conference (Rome, 2001) at <http://w3.uniroma1.it/ssab/ER/en/program.htm>

automated methods of registration. An alternative presentation of the same data, which can serve slightly different needs is offered without very much additional human effort.

### **The bibliographic record**

All electronic documents that are part of the library's service is to be described in the catalogue: electronic periodicals with full text access, online databases available for the patrons, databases available in the LAN CD-ROM service of the library or at standalone workstations. Some "free" Internet resources that are considered to be important for the academic community can also be subject to cataloguing. The criteria of their selection are laid down in a set of in-house rules. The descriptions are created in MARC format, following the ISBD(ER) standard (Figure 1.) and receive UDC strings and subject headings. The records are saved into the local catalogue and as in the case of other documents into the location database of the National Document Delivery System. (<http://odr.lib.klte.hu>). They are very popular there: they can either be accessed from the location database interface or we receive ILL requests for articles in them, which are fulfilled in a print format, when licenses permit. (Figure 2.)

The bibliographic records have a few mandatory elements that serve the automated processes:

1. Access information can be of two different types. In the case of an on-line document it is the URL (865 \$u) providing direct link between the record and the document itself, which is the most important. In most cases we add a short statement on which part of the document the URL links to and on the group of users who are entitled to access it, if relevant. (856 \$3). If the document cannot be directly accessed a note (590 \$a) describes the exact place and method of access. The data from both fields are displayed in the Web-list entries, as well.
2. While we intend the catalogue to present all electronic documents integrated into all the holdings of the library, we want to present them selected and grouped in the Web-lists. At the moment we maintain lists of "databases", "electronic journals" and "Internet documents" on our homepage. The cataloger of the document decides if the document is to be entered into one of the Web-lists and adds a code to the record (699 \$a) to mark it for the list generating mechanism to select them and write them into one of the lists. In the case of electronic journals the same code carries the information whether it is only the table of contents or the full text of the journal that can be displayed. (Figure 3.)
3. Within the three document groups there created two web-lists. One is in alphabetic order of the document title (245 \$a), the other is a thematic list the categories of which are created and filled with the appropriate items based on the UDC strings (080 \$a) in the bibliographic records. So these fields are compulsory, as well.
4. Our bibliographic records traditionally contain the initials of catalogers who created and updated them. (583 \$x). One of them is selected with an add "Q" character as the person responsible for the validity of the document URL, the person who is warned by the URL checking mechanism in an electronic message when it finds problems accessing the documents.

### **The Web-lists**

Presently nine lists are generated from the bibliographic records in the on-line catalog that contain a code in its 699 field: alphabetic (Figure 4.) and thematic lists (Figure 5.) of

databases, electronic journals and Internet documents and shorter alphabetic lists of the most recent documents call the patrons' attention to the novelties we offer them.

The lists are generated daily so that they receive the new items and updates entered into the bibliographic records the previous day and remain updated.

The display format of list items comply with the conventions of web-list: they are relatively brief as far as bibliographic information is concerned but usually have short summaries or contents and access information displayed. (Figure 6.)

### **Thematic lists**

When designing a thematic list the main problem is to decide the number of subject categories we intend the list items grouped into. If we have a thousand items in the database it is equally useless to divide them into three categories as into as many as 600. The optimal depth of categorization seems to be one in which we have 10 to 30 items in subject category. If we design the system of subject categories to match the initial number of records in the database it can soon be outgrown especially in the case of a document type that increase in number so dynamically. A solution to the problem can be a category system which making use of the hierarchical structure of the UDC strings widens itself as the number of documents grows.

The framework of our subject list is a set of subject categories, which is not predefined but always depends on the UDC strings that are present in the bibliographic records belonging to the given thematic list. The program that generates the lists reads the UDC strings from left to right and works with so many digits as the number of items in a given subject area justifies. If there are only a few items with UDC string starting with 3, they will constitute one group with the heading Social sciences. As the number of items becomes larger than a given number the program steps to the next level, reads the first two digit of the same strings, counts the number of records belonging to them and introduces the subject categories on the next level (sociology, law etc.) If a category on the second level does not have enough items in it, it is not created and its items are still listed under the category on the higher level.

As the structure of UDC strings is not so strict that we can always identify the levels of subject hierarchy with the digits of the UDC strings, we introduced a file which defines the levels of our subject framework, the digits of the UDC strings that has to be taken into consideration at that level and the subject terms to be displayed on the Web-list when the subject level is introduced. (Figure 7.)

### **Checking the URLs**

The present aim of the project is to introduce a mechanism that automatically tries to access the URLs given in the records. If the program finds problems (the page cannot be accessed, its title is not the one accepted) a message is sent to the catalogue responsible for the document and marked in field 583 \$x with a Q. The URL can be corrected in the bibliographic records and the corrections will appear in the Web-lists by next day.

---

000 00424nas 2200289 4504  
 001 bibKLT00304027  
 005 20020118165851.0  
 008 020115c20029999xxu p 0uuu 0eng d  
 040 \$aHuDeKLEK  
**080 \$a63\$a57\$a504\$a3\$a61\$a53\$a54\$a62\$a7\$a8**  
 110 2 \$aInstitute for Scientific Information  
 245 10 \$a**Current contents connect**\$h[elektronikus dokumentum] /\$cInstitute for Scientific Information  
 250 \$aVersion 1.6.  
 260 \$a[Philadelphia, PA] :\$bInstitute for Scientific Information,\$ccop. 2002  
 500 \$aCím és leírás a nyitóképernyőről  
 505 \$tAgriculture, Biology and Environmental Sciences (ABES) -\$tSocial and Behavioral Sciences (SBS) -\$tClinical Medicine (CM) -\$tLife Sciences (LS) -\$tPhysical, Chemical and Earth Sciences (PCES) -\$tEngineering, Computing and Technology (ECT) -\$tArts and Humanities (AH)  
 520 \$a Szemlézett tudományterületek cikkadatbázisa. Az egyes témakörökhöz Interneten elérhető további források gyűjteménye is kapcsolódik.  
**583 \$xQKK**  
**699 \$aA**  
 650 4 \$aMezőgazdaság\$xadatbázisok  
 650 4 \$aBiológia\$xadatbázisok  
 650 4 \$aKörnyezettudományok\$xadatbázisok  
 650 4 \$aTársadalomtudományok\$xadatbázisok  
 650 4 \$aOrvostudomány\$xadatbázisok  
 650 4 \$aHumán tudományok\$xadatbázisok  
 650 0 \$aAgriculture\$xDatabases.  
 650 0 \$aBiology\$xDatabases.  
 650 0 \$aEnvironmental sciences\$xDatabases.  
 650 0 \$aSocial Sciences\$xDatabases.  
 650 0 \$aMedicine\$xDatabases.  
 650 0 \$aHumanities\$xDatabases.  
**856 41 \$3IP alapú elérés a DE számára \$uhttp://www.clib.dote.hu/hun/current\_uj\_hun.html**  
 949 \$ID1/IR  
 999 \$aDEBR\$bD1/IR

Figure 1. A bibliographic record.