



World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

Code Number: 142-E
Meeting: 86. Bibliography
Simultaneous Interpretation: Yes

The perpetuation of the national bibliographies in the new virtual information environment

Marcelle Beaudiquez

Bibliothèque nationale de France
Paris, France

Summary

The author wonders about the possibility of applying the principles of universal bibliographic control to web sites and the recommendations concerning the compilation of the national bibliographies, established in the 70s for traditional documents. She then identifies several scenarios to “create” a national bibliography of the sites and insists on the need to perpetuate into the inventory of web sites the functions of the national bibliographies as elements of national memory. She ends by asking for a strong involvement of IFLA in the establishment of new recommendations.

The question of the future of the national bibliographies¹ in the context of a universal bibliographic control in the 21st century² has already been broached in 2001 in Boston. Today's presentation can thus be considered the 2nd constituent of this thought. In only two years, the development of these different components (“documents” for inventory, depot legal legislations, the Internet environment, ...) has accelerated. The decisions to be taken by the national libraries

¹ The consideration only applies to bibliographies produced by a national library or an official organization benefiting from the depot legal.

²) “ Usages et utilité des bibliographies nationales : quelles perspectives ? “ in IFLA Journal, vol. 28 (2002), n° 1 (<http://www.ifla.org/IV/ifla67/papers/114-119f.pdf>)

are now imminent. This communication contributes to the international thought on this domain and equally applies to certain considerations as to the Bibliotheque nationale of France³.

Already since several years, the contents of the Internet sites have been considered an element belonging to the national memory, worthy of being collected, identified, conserved, communicated in the within the framework of the tasks of the national libraries.

Thus the new depot legal laws will henceforth take into account the deposit of “signs, signals, writings, images, sounds or messages of all sorts which are part of line communications”,⁴ of numerous experimentations done in different countries in order to test the collection and archival of the sites. For the web sites⁵, henceforth included in the depot legal laws, the inventorying becomes necessary, and they thus enter into the universal bibliographic control system, which includes the establishment of a current national bibliography. Starting with this statement, one needs to ask oneself and define what universal bibliographic control are we talking about? Is there advancement in implementing it from the moment where the web sites are concerned in the same capacity as the traditional documents? What national bibliographies should be established? For what use?

In order to define if it is important to continue to establish current national bibliographies in an Internet context and the usage reserved for them, it seemed helpful to me to rationalize in view of the principles and recommendations already established for traditional documents.

1. The principals of the CBU and the web sites

Universal bibliographic control is based on the principle of **national responsibility** in the matter of collection and inventory. This principle of nationality is not automatically transposed in the virtual world like that of the Internet whose field of distribution has no border, but close reference points can be found in the notion of country domain (such as .fr, .uk, .se...) on condition of questioning the fate of that which thus escapes the “national” inventory in the .com, .org, .net... However, let’s not forget that the basis for the national bibliographic inventory of traditional documents can be equally linguistic or documentary (dealing with the country): these same bases can equally be adopted for the inventory of web sites.

The universal bibliographic control is also largely founded on **the principle of exhaustiveness**. When applied to the web sites that are considered the new national memory, this principle multiplies by 5 or 6 the number of “depots” and thus the number of potential “inventory.”

³ in France, the next law on the depot legal of the web sites should be voted by the parliament toward the end of 2003, followed by its application decree. It is already known that an important delay (no doubt about 36 month) in implementation will be agreed to in order to allow the various concerned parties to be prepared, according to their human and financial resources.

⁴ It is the formula of the future French law on the Internet depot legal.

⁵ We will use in the text the simple phrase of “web site” for convenience, knowing fully that in reality the Internet is more complex.

However, resulting from the inquiries that already, for the traditional document, the search for the exhaustiveness has become a race without end, of which one meanwhile talks of “reasoned exhaustiveness” if not “reasonable”, and that one has already introduced the principal of selection of the collection in certain depot legal laws for certain types of document (radio and television broadcasts). It is thus not against the orthodoxy of universal bibliographic control to envisage a selective collection of Internet sites. One must watch out however, not to strongly introduce the notion of documentary interest in the criteria in order not to pass value judgment on the collection, a principle that has generally been respected until now.

But one can already depend on the criteria of various nature: for example, only selecting the web sites which, in whole or in part, ensuring the continuity of the existing collections, because they are or will continue the types of document already collected on other support: or because they are part of the domain in which the national library is already a pole of excellence.

2. The principles of current national bibliographies and the web sites

The current national bibliography, developing with the depot legal laws, has progressively inventoried types of document more and more diversified, edited or produced on equally diversified supports. As to document, the national bibliography has even changed since only a dozen years the “editorial” support, which it has accumulated at first (paper + cederom), then substituted on for the other (only cederom). Today, one already questions the necessity to or not to continue to produce a cederom support, whereas one also proposes a national bibliography online. One must however underline that this putting online of a certain number of current national bibliographies is nothing but the transfer online of the organization of the preexisting bibliography without taking advantage, in the quasi totality of the cases, of the new web research technology or interfacing, although the indexes are however better developed⁶. These online national bibliographies, inventoried almost always documents on the “traditional” support, going from paper to DVD. Thus the national bibliography adapts itself to the editorial changes to which it bears witness without having a veritable breaking point.

The current national bibliography is established on the basis of the depot legal and the recommendation is to **describe in the bibliography all that has been deposited there**. However, one already knows that for traditional documents this principle is not always applied and that the “reasoned” exhaustiveness of the inventory goes together with the reasoned exhaustiveness of the depot. The selection criteria of the web sites could be

⁶ The first important change of the current national bibliography that could arise rapidly is independent of the “virtualisation” of the documents it collects: it would put to action the modelling which always meets with a larger number of supporters among the librarians.

The FRBR approach identifies the object and the expression agrees well with the identification “ad vitam aeternam” of a national edition on an intellectual plan, while the manifestation translates the materiality of the collecting of the supports. The retrospective national bibliographies, all types of mixed document, could thus change in their ergonomics et the “ferberisation” would thus naturally gain the whole of the catalog. The adoption or not of the modelling of the FRBR type in the current bibliographies would more or less be tied, in first place, to the change in cataloging norms (the result of work by the workshops on the international coding of cataloging will tell us more at the proper time; that is why I tried to include this possible change in the general thought).

completed with selection criteria in the bibliographic inventory, if that were deemed necessary.

The current national bibliography of traditional documents rests on another large principle, the unitary description (although the treatment by sets already exists for “ephemera” or pictures) for complete standardized description that serves as an identity card of a document. With the web sites, this idea of “bibliographic unity” seems to be reserved to the whole site, while even the search engines allow for a finer granularity in identifying a page, by example.

With the traditional documents, the current national bibliography immediately guarantees a **perpetual identification** because the physical unit web description is static, all change in content dealing with a new edition, a new depot, a new identification. The web sites are dynamic resources, the “continuing resources”, often modified and updated; they can also be fragmented in an environment where all or part of the site exists only as a display in a search query (web “à la volée” or deep web). However, for the length of its existence, it is the site itself that is identifiable and which is the closest equivalent to the identification of the physical unity

As a support document, the current national bibliography restores a **bibliographic notice** without links with the preservation of the identified document, even if it is evident that, the current national library being edited on the basis of the depot legal, there at least exists a place where one can consult the identified document even long after its publication, it’s the national library of the concerned country. The current national bibliography vouches for the existence of stable physical unit of which at least a copy is kept at the national library, but which have been produce as thousands of copies preserved in other organizations, at the publishers or at the various individuals who bought them. For the web site, the description, the location and the conservation which can be done by the national library correspond to its state for only an “instant”, that of the capture of the site and the trace of which are not necessarily conserved in another institution, nor even by its “creator”, at the same moment, and even less in its entirety. The manner of capture and the depot, the archival methods reserved to restore a dynamic communication of the site with its original links would, for example weigh necessarily on the possibilities and obligations of identification and use of the sites in the duration.

In the matter of **rapid response**, and for a longtime, the current national libraries depending on an effective depot in a beneficial institution, rivaling with the current commercial bibliographies often depending on information from “forthcoming” documents. For all users of a current national bibliography, research leads them to a bibliographic note which will allow them next to consult the document, there, where it is inevitably preserved, the national library. For a modern netsurfer, his search would lead him directly to the site. This immediate access to the site makes all current national bibliographic records of sites useless, as long as those are active (less than 75 days) and free. The simultaneity of the search/ identification and visualization of the site are a criterion showing the documentary use of the patrimony. It implies a repositioning of the role of the current national bibliographies: their rapid response no longer has the same

stake, as there is no longer any added value for the current research. On the contrary, it's in retrospective searching that the national bibliography will find its place if it is accompanied with a perpetual archive.

Thus the other parts of the chain of tasks tied to the depot legal also develop when it concerns the depot legal of web sites;

- the collection becomes a voluntary activity of the national library in the sense that it must put in place the technical means of site searches or the investigation of collection support (FTP files);
- the preservation/archiving is inseparable from the collection which brings about technical choices;
- communication, if one doesn't want it to reduce the communicable "virtual document" to a piece both static and closed of the original web site, should be able to restore access to the site in its original environment (its links), that is to say to a foreseeable capture (thus an effect linked to the collection and conservation).

3 – Particularities of identification of web sites

3.1 which sites?

The actual observed slow evolution is without a doubt normal if one considers that the principle subject of the national bibliographies remains "traditional": the mass of printed and audiovisual documents to be inventoried remain largely greater than the mass of online documents as far as the depot legal of the later concerns only a very limited number of countries. But, as regards legal deposit, a number of signs show that the transformation is speeding up.

3.1.1. Continuity of the types of collected documents

One must, however, rapidly find solutions for the most urgent case, that is to say the type of traditional document collected and inventoried on support and which are nevertheless not issued except online, such as the component part of a web site. The continuity of the collections of the depot legal is a legitimate worry of the librarians and it is important to be able to discover nevertheless among the information on the web those which have replaced earlier traditional support editions (for example e-books or online serial publications)

That these virtual documents, according to country, are already obligatorily submitted or not to the depot legal, their inventory, their description in a library should be studied, if not the editorial traditional bibliographical lists will be suddenly cut off.

One realizes that this separate place aids the continuation of a traditional type of documents already submitted to the depot like books, periodicals, but eventually also pictures, sounds, television programs, films,;... (considering the Internet as the new support of the moment) can make up a strong criteria in the selection of web sites submitted to the depot.

Thus, in France the National Audiovisual Institute of France who will be one of the administrators of the Internet depot legal from the promulgation of the law has identified 5,000

online television and radio sites, to extend the collection of information (selection of television and radio programs) of which it already has charge of.

For its part, the Bibliotheque nationale de France is studying the feasibility of identifying online French periodicals on the web in order to eventually collect them and to treat them as priority among the defined selections in the future law.

3.1.2 Continuity of the basis of inventory

For other countries, the national and patrimonial documentary interest of the web sites was the first criteria of choice to capture or describe. Thus Sweden since 1997 has automatically captured 126,000 Swedish web sites in order to obtain “ a true picture of the Swedish space at the time of archiving” ; Australia, through the PANDORA project has identified and described around 2500 sites “that have research value for studying the history of Australian society” even before the depot became obligatory by law. For Canada, it’s the occasion for the creation of the new organization Bibliotheque and Archives of Canada that defines its area of collection on “the Internet documents which are attributable to a Canadian origin or which present an interest for Canada”.

In these cases of typological or thematic selection and of weak volume, one considers that the volume to be administered authorizes yet a whole bibliographic handling.

But outside of these cases, for which different solutions to bibliographic treatment can be found (cf 3.2) it seems evident that from the moment that the content of the web sites is considered as a component part of the national memory in the same manner as traditional documents, one will always have to be able to answer research questions:

- what sites (addresses and titles) have been deposited/ captured at such a moment or at such periods.
- What “creators” have deposited at a given time period
- What sites have been created in such or such themes?

Those are the functions that fill a national library in the meantime, in relation to a body of national information; but filling the functions does not presuppose the type of production, nor the distribution for this current national bibliography.

3.2 What bibliography?

Thus it remains to predict what method(s) of establishment of a national bibliography of web sites is (are) possible taking into account the specificities of the Internet world.

Three parameters must be addressed in the decision process:

- the percentage of web sites to potentially inventory in relation to the percentage of affected traditional documents and their number, knowing that at least the number of supporting traditional documents will not sensibly diminish and that they will need to be inventoried into the current national bibliographies known to us, even if those are more and more accessible online:
- the specificity of the Internet environment and the frame of references that it already has: a maximum use of the technical network possibilities

and the automation that one can expect in place of the manual interventions in the creation of “notices”, always absorbing human resources;

- legal conditions of communication both local or afar of these sites functioning for the protection of the author’s rights in each country. Certain protectionisms in matter could lead to different measures and give a temporary different perspective in the matter of the nation bibliography; the more the legal and technical difficulties tied to the depot and to the communication of subscription or locked sites would lead to their exclusion from the collection. But must it also be for the national bibliography, which then would be voluntarily and irremediably cut off (which is not the case with the samples).

In this context, the establishment of the national library can come up with variants, and would adopt for the time, “mixed” solutions, in comparison:

- a) if the number of sites to be inventoried is low, or if the selection criteria defined for the collection renders it such, one would chose to include in the structure of the actual current national bibliography, in a single list or separate lists, descriptive notices of the web sites which would follow the rules of the latest ISBD revised to this effect⁷;
- b) whatever the number of sites to inventory, one can chose to treat their bibliographic description with the metadata of the Dublin Core type, recovered from the site itself or manually created and including the utilization of univocal identification numbers developed for the information of everything accessible on the web like l’ISRC (International Standard Record Code), le GRID (global Release identifier) or le DOI (Digital object Identifier) for example.
- c) once the volume of the sites affected by the depot legal law becomes higher and exceeds the possibility of manual identification, it will become necessary to find the maximum of automatic means and take advantage of the frame of references and tools of the Internet world. The virtual document is not only a supplemental reincarnation of changes in production support and the distribution of information; it has its own logic and its own tools. It appears, then, more important to take advantage of the Internet environment, instead of replicating the traditional model of inventorying.

4- Establishment of the national bibliography fitting to the Internet sites

Basis of the inventory:

The country domain (.fr, .uk, .se) supplemented if possible by the .com or .net

⁷ ISBD (CR) ...International Standard Bibliographic Description for serials and other continuing resources. Revised from ISBD (S)....IFLA-UBCIM, 2002

Description of the site:

Reuse of the identifying elements figuring in the data base of the national administration of domains names, which in each country maintains a data base for management allowing to **identify the creator of the site** (organization and/or physical person). A cooperation should be established between this administrator and the national library charged with the depot legal of the Internet sites, so that it may best take advantage the basic elements of administration.

Value added by the national library:

- One can foresee in order to optimize this identification and, if the local legal context allows it, to develop the forms of stating domain names or to depend on, and to create links with, national authority files for corporate bodies and for commercial brands or of authority material (without foreseeing the systematic creations of natural manual links).
- One can as well foresee the development by the library of a search engine proper to the national bibliography arising from the particular subject indexation of our documentary institutions.

Perpetuation of the base data:

The data base created by the national bibliography by the reuse of the domains basis will be perpetuated in such a way to be able to preserve the identification data. It is more than is done today by search engines such as Google or even the administration basis of domain names, which do not take into account temporary perspective.

Intelligent archival of the sites:

The current national bibliography, having become virtual, will have more the characteristic of being able to give direct access to the site itself, restored in a dynamic way with its original links, thanks to an ad hoc archiving.

5 – Conclusion

These thoughts conclude with the change of different parts of a universal bibliographic control from the XXth century that does not deny anything of its origin, but evolves with the information which it contributes in spreading around while leaving each country the possibility to develop according to its own editorial and legislative reality.

Until now, in order to answer the questions “who? what ? when?”, a bibliographic list was established which was itself published and preserved as all other traditional documents. As long as the “traditional” documents continue to be produced, the list will continue to be produced along the same principles, even if it is put online and if the cumulative are computer files.

For web content captured or voluntarily deposited , one will have to abandon this system of lists of whole description comprising of the risk of incorrect addresses towards sites not updated or vanished. The current national bibliography will be administered as an online system, searchable for that duration, and which would provide retrospective products as long as they are needed.

One would also have to guarantee an intelligent archival of sites which would foresee the needs of future researchers and foresee a virtual retrospective national bibliography which would then answer their needs.

IFLA, at the heart of the concept of universal bibliographic control in 1973, could keep a very active and decisive part on the definition of the new universal bibliographic control.

In effect, if one admits the rapprochement of the administrators of the name domains and the need for each national library to cooperate with these in order to take advantage of the elements of identification aiding a perpetual search, the IFLA could work together with the international administrator of domain names, the IRCANN (Internet Corporation of Assigned Names and Numbers) which assures and oversees the tasks linked to the function of the DNS. It would mean assuring that the statute of the delegation of domain names of local administrators mentions the elements of collection identification and the use of the information to the administrative institutions of the Internet depot legal. In the matter of identification, among the proliferation of univocal international numbering, the IFLA could also be prescriber, in drawing closer to international organizations like ISO and its supporting agencies. Also, in the matter of indexation, the work of the IFLA on “subject gateways” should lead to recommendations that the use of the search engines would be a real additional value.

One knows better now the needs and the technical possibilities and the professional requirements’ linked to the Internet legal deposit . A certain number of libraries have already realized this. It will soon be time to think about recommendations in order to insure a global consistency of processing. This is the role of IFLA in the matter of bibliographic description since the beginning , it is also the role of IFLA to set the development of it concept of universal bibliographic control.

Let us not allow a tower of Babel to build up again in an environment already clarified 30 years ago.