

Dig *Italia*

Numero 2 - **2007**

Rivista del digitale nei beni culturali

ICCU-ROMA

Croatian Digital Web Archive

From project to service of the National and University Library in Zagreb

Tanja Buzina

National and University Library of Zagreb

Mirna Willer

Department for Librarianship of the University of Zadar

Since 1997 online publications are included in the legal deposit. Being a deposit library, NUL has an obligation to collect and preserve online publications. For this reason, NUL began experimenting with cataloguing Web publications as early as 1998. However, the processed publications were not archived, therefore, they could only be accessed at their original address.

In order to set up a tool for capturing and archiving online contents, the National and University Library and the University Computing Center (SRCE) in Zagreb launched a project Design of the System for Harvesting and Archiving Legal Deposit of Croatian Web publication on November 1, 2003. The objective was to establish a tool for harvesting and archiving the deposit copy of Croatian publications on the Internet while preserving, to the largest measure possible, the original contents, formats and functionalities, in order to ensure access and use in the future.

The digital archive is based on the concept of selective collection of static and dynamic Web publications in interaction with the NUL's WebPAC. Each Web resource is fully catalogued and thus retrievable in the NUL's catalogue. The Digital Archive is fully integrated with the library information system and is running as a service since January 2004.

Background: laying the ground

Bibliographic identification, processing and archiving of Web resources is being considered a legal obligation by the Croatian National and University Library in Zagreb (NUL).

The library law of 1997 included “electronic publications” as a legal deposit copy. Electronic publications were specified as: compact discs, magnetic tapes, diskettes, databases and on-line publications¹. In 1998 the cataloguing of Web serial publications freely available on the Internet, or resources that were recognized as such, was started as a research project.

¹ *Zakon o knjižnicama [Library law]*, 9.10.2007, <http://www.nn.hr/clanci/sluzbeno/1997/1616.htm>.

The project was mainly concerned with problems of identification, selection and bibliographic description. Serials were identified on the Web by various methods such as analysing portals and websites, by use of search engines, scanning news from media (radio, TV, newspapers), getting information in printed serials, from publishers' information, reference librarians or users' information, in a way, by word of mouth. Gathering information in such a way was soon recognized as sporadic and unsystematic, and therefore inadequate for a sustainable system for collecting Web legal deposit.

The selection criteria were based on the criteria for the choice of the national legal deposit with some specific criteria added, based primarily on the nature of Web resource itself, like the type of resource (a Web site, but not a mailing list), and appropriateness of the format of the resource for archiving². Identified Web resources were catalogued in the library system, and linked via URL to their original version on the Web due to the fact that the only technical solution that was available was the local archive of word and pdf formats which eventually turned out to be unretrievable. It was therefore necessary to systematically approach the problem of identification of Web resources in general – not only serials but monographic and integrating ones, and to start the design of an archival system that would enable archiving and retrieval of different types and formats of Web resources. Additionally, it was recognized that processing Web resources requires the definition of a completely new working model.

As part of the research project, international work in the field was closely observed, and special attention was paid to types of archiving – harvesting national domains or a wide range of websites, and focused or selective archiving methods³, as well as to the design solutions of a deposit system.

The decision to continue with selective archiving was supported by international experiences which proved that this method supports the following goals:

1. the functionality of each Web resource to be archived can be qualitatively assessed;
2. each Web resource to be archived can be fully catalogued and thus retrievable

² Sofija Klarin – Sonja Pigac, *Hrvatske daljinski dostupne elektroničke serijske publikacije* [Croatian remote electronic serial publications], «Vjesnik bibliotekara Hrvats», 43 (2000.[i.e. 2001]), n. 4, p. 156-167.

³ Sofija Klarin, *Predmet, motivi i metode arhiviranja sadržaja Weba* [Subjects, motives and methods of Web archiving], in: *8. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova*, uredila Tinka Katić, Zagreb: Hrvatsko knjižničarsko društvo, 2005, p. 22-35; Paul Cunnea, *Web arhiviranje Škotskoj i Velikoj Britaniji: stižu li problemi?* [Web archiving in Scotland and UK: flies in the ointment], in: *10. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova*, uredile Ivana Marinković Zenić i Mirna Willer, Zagreb: Hrvatsko knjižničarsko društvo, 2007, p. 27-39.

- in the integrated catalogue and available for inclusion in the national bibliography⁴;
3. each Web resource in the archive can be accessed, depending on the conditions allowed by the publisher;
 4. the specific nature of each Web resource can be analyzed and described, which enables upgrading of the harvesting and archiving system, and defining requirements for preservation.

The EU project NEDLIB's solution of a design of an existing library automation system linked to a deposit system of electronic publications (dSEP) (Figure 1)⁵ particularly appealed to the researchers as the optimal model for the NUL's digital Web archive. It was therefore necessary to link the work already started on selective identification and bibliographic processing of Web resources in the NUL's library information system with the design of an archive that would enable the harvesting of selected resources and integration with the library catalogue.

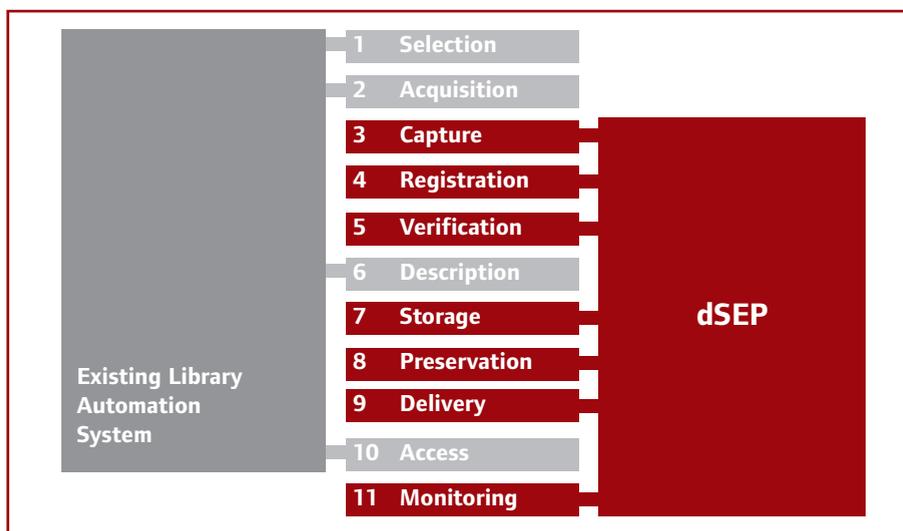


Figure 1. NEDLIB model for deposit system of electronic publications (dSEP).

⁴ Sofija Klarin – Tomislav Murati, *Identifikacija, odabir, obradba i osiguravanje dostupnosti mrežnih publikacija u kontekstu nacionalne bibliografske kontrole [Identification, selection, bibliographic and access management of net publications in the context of national bibliographic control]*, in: *4. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u kontekstu globalne informacijske infrastrukture: zbornik radova*, uredile Mirna Willer i Tinka Katić, Zagreb: Hrvatsko knjižničarsko društvo, 2001, p. 41-56.

⁵ Johan Steenbakkers, *Setting up a Deposit System for Electronic Publications*, The Hague: Koninklijke Bibliotheek, 2000, <http://nedlib.kb.nl/results/NEDLIBguidelines.pdf>; NEDLIB Project, <http://nedlib.kb.nl/>.

From project to full service

In 2003 the National and University Library approached the University Computing Centre at the University of Zagreb (Srce) for their expertise in Web technologies, and their acquaintance with Croatian Web space, in particular with the proposal for co-operation in Web archiving. The requirements were to design a system that would enable harvesting and archiving of selected Web resources, and be integrated with the existing library automation system. The goal of building such a system was enabling preservation of the original Web sites as much as possible – their «look and feel»: content, form and functionality, for future use. A two year project *Design of the System for Harvesting and Archiving Legal Deposit of Croatian Web Publications* was started that same year, followed by two more one year projects, the third one being finished by 1st October 2007. Development plans for the next one year period have been proposed to the NUL's management (November 2007).

Archiving of Web resources started in the beginning of 2004, after the first test version of Digital Archive of Croatian Web Resources – DAMP – was designed. Version 1.0 was released in November 2004. By the 1st of October 2007, the archive contained 1,840 items with 17,680 instances of those items. 147 Web resources had disappeared from the live Web since 2004, while 10 are password protected due to commercial reasons. The total size of the archive was 1TB. The growth of the archive is assessed using the figures for the last year to be at the rate of approximately 300 items per year, with the rise of 9,349 instances. In the survey performed in 2005, the types of resources and types of formats in the archive for the first two year period (2004-2005) were analysed. The result showed the following distribution of types of resources – 40.33% continuing integrating resources (e-zines, websites of institutions, government bodies, events, research projects), 30.99% serials (9.33% newspapers and 21.66% periodicals), and 28.66% monographs (including finite integrating resources); type of formats mainly used by publishers or authors were text/html, image/jpeg, image/gif, application/pdf, and application/x-java script⁶. The results of the 2007 research study showed the change of the distribution of the type of resource. The sample showed that the highest percentage of items were integrating resources – 54.81%, followed by 28.57% of serials and 16.62% of monographic resources. Compared to the analysis done for the first two years of processing Web resources (2004-2005), we can conclude that the percentage of integrating resources has risen (from 40.33% to 54.81%), percentage of serials has slightly fallen (from

⁶ Sonja Pigac – Tanja Buzina, *Selektivno arhiviranje hrvatskog weba: rezultati i otvorena pitanja* [Selective archiving of Croatian Web: results and open questions], in: 9. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova, uredile Mirna Willer i Ivana Zenić, Zagreb: Hrvatsko knjižničarsko društvo, 2006, p. 30. There is a lack of 0.02% in the cited source which can be added to continuing integrating resources.

30.99% to 28.57%), while the percentage of monographs has significantly fallen (from 28.66% to 16.62%)⁷.

Access to the archive is via the WebPAC of the NUL's catalogue, using the Web interface of the archive (Figure 2). Title and URL searches are available, with browse possibilities.

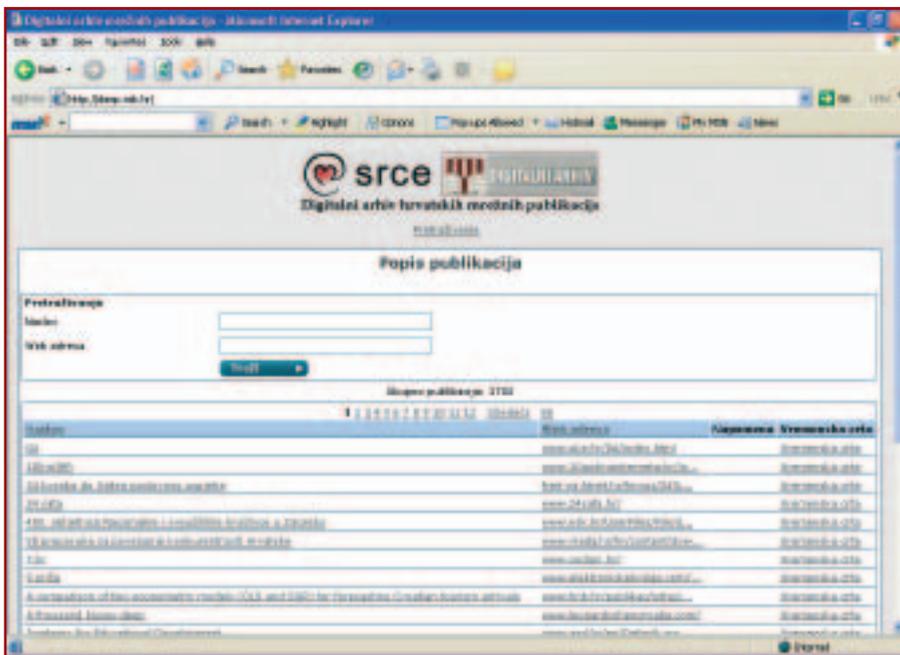


Figure 2. Web interface of the Digital Archive of Croatian Web Resources: <http://damp.nsk.hr>.

The functionality of the first version immediately included linkage of DAMP to the NUL's automation system, i.e. library catalogue, thus forming a functional integrated system. During 2004, when cataloguing and archiving started, the basic workflow for processing Web resources was established. After the end of the first phase of the project, during which the members of the Project team from different processing units took part in the development and setting up the service, the Library selected three members of the staff to work full time in a newly created Unit for Processing Web Resources (UPWR). The UPWR's jobs were broadly defined as identification, selection, cataloguing and archiving Web resources⁸.

⁷ Tanja Buzina – Karolina Holub – Miroslav Milinović – Nebojša Topolščak – Mirna Willer – Jasenka Zajec, *Selective archiving of Croatian Web resources: a study of processing costs at the National and University Library of Croatia*, Dubrovnik-Mljet: LIDA, 2007, http://www.ffos.hr/lida/datoteke/LIDA2007-Buzina____.ppt.

⁸ Sonja Pigac, *Digitalni arhiv mrežnih publikacija: rad u sustavu za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija* [Digital archive of online publications: working within the system for capturing and archiving legal deposit of Croatian online publications], in: *8. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova*, uredila Tinka Katić, Zagreb: Hrvatsko knjižničarsko društvo, 2005, p. 61–67.

The head of the unit was elected from the project team members. The UPWR was supported by the project manager and staff members from different Library departments working part time on Web resources, as well as by the development and technical team, and the project manager from Srce. The UPWR's everyday tasks ran in parallel to the project. The workflow in the present organisation consists of the following tasks being performed in this order:

1. *Identification:*

Sources for identification of a possible candidate for the archive are:

1. Web;
2. communication with publishers related to ISSN, CIP, ISBN and ISMN;
3. assessment of the online registration form on NUL's Website filled in by authors/publishers.

2. *Selection:*

Selection is done according to the criteria for ISSN assignment, and the *Selection Criteria for the Choice of the Legal Deposit Copy of Web Publications*⁹.

3. *Formal and Subject Cataloguing:*

The tasks involve creation of new bibliographic and authority records for serials, continuing and finite integrating resources, and monographs; subject cataloguing and classification; and updating bibliographic records for serials and integrating resources (continuing and finite) on the bases of changes in the resource.

4. *Archiving:*

Metadata for bibliographic identification of the item (i.e., title proper, ISSN, ISBN, ISMN, automated catalogue record identification number), Web URL, frequency of issuance and its regularity are automatically transferred from the catalogue to the Digital Archive each day. Only the UPWR staff is involved in archiving. There are two groups of tasks:

4.1 New items: archiving process:

- 4.1.1 checking the item on its live address on the Web;
- 4.1.2 defining the harvesting parameters; registering the item to harvesting queue;
- 4.1.3 checking the quality of the first harvesting:
 - 4.1.3.1 repeating, when necessary, the harvesting with changed parameters;
 - 4.1.3.2 deleting unsuccessful or poor instances of harvesting;
 - 4.1.3.3 checking the archived item and for display in the catalogue and the Digital Archive's Web interface;

⁹ Nacionalna i sveučilišna knjižnica u Zagreb, *Kriteriji odabira obveznog primjerka mrežne grade za obradu i arhiviranje* [Selection Criteria for the Choice of the Legal Deposit Copy of Web Publications], <http://www.nsk.hr/DigitalLib.aspx?id=83>.

- 4.1.4 defining the frequency of harvesting;
- 4.2 Existing items: quality control of archived instances:
 - 4.2.1 checking the availability of the item at its live Web address according to the monthly automatic report;
 - 4.2.2 changing harvesting parameters if a change in properties/structure has taken place:
 - 4.2.2.1 deactivating harvesting parameters if the Web resource has disappeared from the live Web;
 - 4.2.2.2 control of the multiple harvesting instances;
 - 4.2.2.3 deleting unsuccessful harvesting;
 - 4.2.3 checking automatic daily reports on possible duplicates, and deleting them;
 - 4.2.4 changing frequency of harvesting parameters;
 - 4.2.5 reporting on harvesting problems to Srce technical team.
- 5. *Updating the catalogue:*
After an archived item is checked ready for display in the catalogue and the Digital Archive's Web interface, a programme automatically updates the catalogue over night.

With the fully established workflow of processing Web resources integrated within the organisational structure of the NUL, the project has developed into fully defined service of the library (Figure 3).

In spring of 2007, the project team carried out a two month research of the working process established so far, with two aims. The first one was to assess the costs for processing Web resources, taking into account the time and type of task per item archived and the average time per task per type of resource processed, as well as to assess other costs related to maintenance and development of the service. The second aim was to analyse and assess the present organization and workflow of processing Web resources and propose, if necessary, improvements. The results of the research showed the following:

- the present workflow meets the requirements of efficient processing of Web resources;
- almost the same time is used for original cataloguing and updating existing records due to the changes of resource characteristics: specific to Web resources vs. print publications;
- a high percentage of time is used for archiving, and thus employment of a new staff member with technical skills, and advanced knowledge of Web technology and techniques is needed;
- a high percentage of time of the UPWR staff members, and to a minor extent by ISSN office staff members that is dedicated to development (research, ser-

vices and tools – guidelines for cataloguing) shows that UPWR staff members have taken over these tasks from the Project members as part of their everyday activities¹⁰.

With new organizational changes that are taking part in the NUL at present it is expected that the requirement for a new member staff will be met. The fact that the head of the UPWR is taking the place of the NUL's project co-ordinator in the newly planned project with Srce will bring into the unit an even greater part of the new service design and development.



Figure 3. Web site of the Digital Archive of Croatian Web Resources as part of the NUL's services: <http://www.nsk.hr/digarhiv>.

Architecture of the DAMP system

The DAMP system was developed using open source software. It consists of six main parts: *gatherer* – a subsystem for gathering Web resources, *controller* – a subsystem that controls gathering process and archives gathered content, *scheduler* – a subsystem for scheduling jobs, *data storage* that consists of a database

¹⁰ Sonja Pigac – Tanja Buzina, *Selektivno arhiviranje hrvatskog weba: rezultati i otvorena pitanja* [Selective archiving of Croatian Web: results and open questions], in: 9. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastructure: zbornik radova, uredile Mirna Willer i Ivana Zenić, Zagreb: Hrvatsko knjižničarsko društvo, 2006, p. 28-39.

and specially organised file system, *Web interface* that provides easy access to the archive and allows control of the system, and *data exchange module* that provides exchange between DAMP and NUL's catalogue. The functional scheme of the system is given in Figure 4¹¹. A Web interface to the archive which enables librarian/archivist to monitor results of archiving process of each resource is shown in Figure 5.

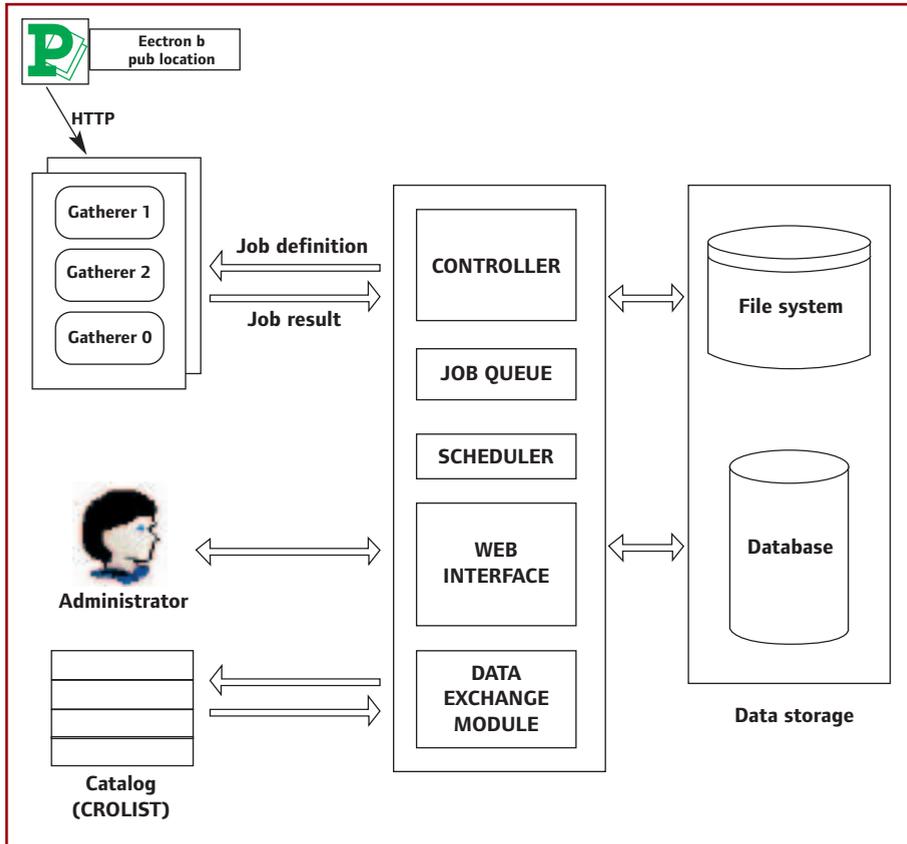


Figure 4. Functional schema of the DAMP system.

¹¹ Miroslav Milinović – Nebojša Topolščak, *The architecture of DAMP*, «Widwisawn», vol. 3, 2005, n. 3, http://widwisawn.cdli.strath.ac.uk/Issues/Vol3/issue3_3_1.html.



Figure 5. Web interface to the working module of the DAMP system: Harvesting details.

The DAMP system has proved to be a useful tool for archiving a limited and well defined set of Web resources; however, it also proved to have some limitations. The specific use of certain Web technologies in the design of Web sites or resources in general is the cause of many difficulties in the archiving process and, in some cases, of unsuccessful archiving. Therefore, it proved to be vital to constantly improve the system, educate the librarians/archivists in Web technologies and use of gathering parameters, but also educate the publishers in order to avoid innovative but non-standard use of Web technologies.

The assessment of the archive, and its comparison to the NEDLIB dSEP model showed that apart from setting up the basic service, further work should be done in the field of direct downloading of instances of Web resources from publishers and harvesting Web resources that require authorization/authentication. The field of long term preservation has not been addressed yet¹², but the NUL is closely following the work of the International Internet Preservation Consortium – IIPC¹³.

¹² Mirna Willer – Miroslav Milinović, *Prema trećoj generaciji knjižnično-informacijskih sustava: hibridna knjižnica za hibridne usluge [Towards the third generation of library information systems: hybrid library for hybrid services]*, in: *8. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova*, uredila Tinka Katić, Zagreb: Hrvatsko knjižničarsko društvo, 2005, p. 36-67.

¹³ International Internet Preservation Consortium – IIPC, <http://netpreserve.org/about/index.php>.

Communication with publishers: essential factor in selective archiving of Web resources

With the aim of improving the archiving results, but also to make the identification process sustainable, the project team devised various ways of raising publishers' awareness of Web resources as part of the legal deposit copies. Apart from promoting the service by publishing on it on the Web, in magazines and journals, presenting at publishers' aimed conferences, etc., various documents and/or services aimed directly at publishers were produced.

In the case of a freely available resource, UPWR staff report to publishers on Web resources being harvested and archived upon each first instance of harvesting. In this report, the publisher is directed to relevant information on the Digital Archive website related to documents on legal deposit, selection criteria and two manuals *Digital archive of Croatian Web publications: guidelines for publishers/authors* and *Recommendations for the creation of Web publications*¹⁴. Publishers are further asked to fill in the online registration form for their publication which is then processed in the identification phase of the workflow. Additionally, the harvesting robot identifies itself upon each instance of harvesting.

In the case of Web resources with restricted and/or password protected access, the UPWR staff consult the publisher on the password and conditions of use. The registration form offers three possibilities for access to be defined by publisher: free access via Internet, access restricted to IP privileged computers within the NUL, and access allowed only to librarian/archivist for processing and archiving a resource. In the case of serials, the UPWR staff establish an agreement with the journal editors/Web masters to report back about posting of a new issue/iteration, thus enabling controlled harvesting as oppose to setting up automatic harvesting (which can harvest the same issue because a new one is not posted yet).

By analogy with their tasks for printed material, ISSN, ISBN, ISMN and CIP units contact publishers, along with UPWR staff, on various issues like giving advice on the presentation and layout of Web resources, sending queries to publishers about updating or lack of posting of new content/issues according to publication pattern, sending information about poor harvesting results, and referring to the *Guidelines*.

¹⁴ Nacionalna i sveučilišna knjižnica u Zagreb – Sveučilište u Zagrebu, *Sveučilišni računski centar (Srce): arhiviranje obveznog primjerka mrežne grade: smjernice za nakladnike/autore* [*Digital archive Of Croatian Web publications: guidelines for publishers/authors*], Verzija 1.0, 2005, http://www.nsk.hr/UserFiles/File/dokumenti/Smjernice%20za%20nakladnike_Verzija%201%200.pdf; Nacionalna i sveučilišna knjižnica u Zagreb, *Preporuke za izradu mrežnih publikacija* [*Recommendations for the creation of Web publications*], <http://www.nsk.hr/DigitalLib.aspx?id=356>.

Future developments

The programme for the fourth phase of the project *Design of the System for Harvesting and Archiving Legal Deposit of Croatian Web Publications* was defined with the following developments:



Figure 6. Time line of an integrating resource *Geografija.hr*.

- finding solutions for harvesting Web resources that show technological idiosyncrasies or new technological solutions;
- updating the Web interface of the archive as a Website of its own with a temporal scale of archived instances implemented;
- develop the Web portal for *Digital archive of Croatian Web publications*; time line is in the testing phase (Figure 6);
- mapping of MARC 21 to Dublin Core data elements, and export of DC meta-data from the NUL's catalogue to DAMP for indexing according to the OAI standard;
- developing the application for indexing the content of the archive for retrieval and harvesting according to the OAI-PMH standard;
- designing a module for monitoring the use of the archive.

The analysis of archived material, representation of certain type of domains and type of data, technical problems or even impossibility of harvesting certain types

of Web resources, and manner of defining the boundaries of a Web resource by use of different harvesting parameters¹⁵ shows that processing of this type of resource requires up-to-date checking of the archived resources, particularly integrated ones, continuing education in technical and technological aspects of the Web as a source for publications and working space, and re-evaluating our understanding of the nature and functionality of the Web content – of the information objects¹⁶!

Le pubblicazioni on-line sono incluse tra le opere soggette al deposito legale dal 1997. In qualità di biblioteca adibita al deposito legale, la NUL è pertanto tenuta a raccogliere e conservare le pubblicazioni on-line, ragion per cui già nel 1998 la biblioteca iniziò a sperimentare la catalogazione delle pubblicazioni Web. Purtroppo, le pubblicazioni catalogate non venivano archiviate, e vi si poteva quindi accedere esclusivamente all'indirizzo originale.

Per rispondere alla necessità di sviluppare uno strumento per la cattura e l'archiviazione di contenuti on-line, il 1 novembre 2003 la National University Library e l'University Computing Centre (SRCE) di Zagabria hanno lanciato il progetto Design of the System for Harvesting and Archiving Legal Deposit of Croatian Web publication. Obiettivo del progetto era l'istituzione di uno strumento per l'harvesting e l'archiviazione della copia di deposito delle opere croate pubblicate su Internet che fosse capace di preservare quanto più possibile i contenuti, i formati e le funzionalità originali dei documenti, così da garantirne l'accessibilità futura.

L'archivio digitale è fondato sul principio della raccolta selettiva di pubblicazioni Web statiche e dinamiche mediante l'interazione con il WebPAC della NUL. Ogni risorsa Web è ampiamente catalogata e quindi reperibile tramite il catalogo della NUL. L'archivio digitale è pienamente integrato ai sistemi informativi della biblioteca ed è operativo come servizio da gennaio 2004.

Les publications en ligne font partie des œuvres sujettes au dépôt légal depuis 1997. En qualité de bibliothèque destinée au dépôt légal, la NUL doit donc collecter et conserver les publications en-ligne, c'est pourquoi, dès 1998, cette bibliothèque a commencé à expérimenter le catalogue des publications Web. Cependant, les publications cataloguées n'étaient pas archivées et on ne pouvait y accéder que par l'intermédiaire de l'adresse originale.

Pour répondre à la nécessité de développer un outil pour capturer et archiver les contenus en ligne, le 1er novembre 2003 la National University Library et

¹⁵ Sonja Pigac – Tanja Buzina, *Selectivno arhiviranje hrvatskog weba: rezultati i otvorena pitanja* [Selective archiving of Croatian Web: results and open questions], in: 9. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastructure: zbornik radova, uredile Mirna Willer i Ivana Zenić, Zagreb: Hrvatsko knjižničarsko društvo, 2006, p. 28-39.

¹⁶ Anne J. Gilliland, *Setting the stage*, in: Tony Gill – Anne J. Gilliland – Mary S. Woodley, *Introduction to metadata: pathways to digital information*, edited by Burtha Baca, online ed., ver. 2.1, J. Paul Getty Trust, 2005, http://www.getty.edu/research/conducting_re

l'University Computing Centre (SRCE) de Zagreb ont lancé le projet Design of the System for Harvesting and Archiving Legal Deposit of Croatian Web publications. Le but du projet était celui d'instituer un instrument, pour l'harvesting et l'archivage de la copie de dépôt des œuvres croates publiées sur Internet, en mesure de préserver au mieux les contenus, formats et fonctionnalités originaux des documents afin d'en assurer l'accessibilité à l'avenir.

L'archive numérique se fonde sur le principe de la collecte sélective des publications Web statiques et dynamiques à travers l'interaction avec le WebPAC de la NUL. Chaque ressource Web est amplement cataloguée et donc repérable à travers le catalogue de la NUL. L'archive numérique est parfaitement intégrée aux systèmes d'information de la bibliothèque et fonctionne en tant que service depuis janvier 2004.

REFERENCES

- Tanja Buzina – Karolina Holub – Miroslav Milinović – Nebojša Topolšćak – Mirna Willer – Jasenka Zajec. *Selective archiving of Croatian Web resources: a study of processing costs at the National and University Library of Croatia*. Dubrovnik-Mljet: LIDA, 2007, http://www.ffos.hr/lida/datoteke/LIDA2007-Buzina____.ppt.
- Paul Cunnea. *Web arhiviranje Škotskoj i Velikoj Britaniji: stižu li problemi?* [Web archiving in Scotland and UK: flies in the ointment]. In: 10. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova, uredile Ivana Marinković Zenić i Mirna Willer. Zagreb: Hrvatsko knjižničarsko društvo, 2007, p. 27-39.
- Sonja Pigac – Tanja Buzina. *Selektivno arhiviranje hrvatskog Weba: rezultati i otvorena pitanja* [Selective archiving of Croatian Web: results and open questions]. In: 9. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova, uredile Mirna Willer i Ivana Zenić. Zagreb: Hrvatsko knjižničarsko društvo, 2006, p. 28-39.
- Anne J. Gilliland. *Setting the stage*. In: Tony Gill – Anne J. Gilliland – Mary S. Woodley. *Introduction to metadata: pathways to digital information*, edited by Burtha Baca, online ed., ver. 2.1. J. Paul Getty Trust, 2005, http://www.getty.edu/research/conducting_research/standards/intrometadata/.
- Sofija Klarin. *Predmet, motivi i metode arhiviranja sadržaja weba* [Subjects, motives and methods of Web archiving]. In: 8. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova, uredila Tinka Katić. Zagreb: Hrvatsko knjižničarsko društvo, 2005, p. 22-35.
- Miroslav Milinović – Nebojša Topolšćak. *The architecture of DAMP*. «Widwisawn», vol. 3, 2005, n. 3, http://widwisawn.cdli.strath.ac.uk/Issues/Vol3/issue3_3_1.html.
- Nacionalna i sveučilišna knjižnica u Zagreb – Sveučilište u Zagrebu. *Sveučilišni računski centar (Srce): arhiviranje obveznog primjerka mrežne grade: smjernice za nakladnike/autore* [Digital archive Of Croatian Web publications: guidelines for publishers/authors]. Verzija 1.0, 2005, http://www.nsk.hr/UserFiles/File/dokumenti/Smjernice%20za%20nakladnike_Verzija%201%200.pdf.
- Sonja Pigac. *Digitalni arhiv mrežnih publikacija: rad u sustavu za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija* [Digital archive of online publications: working within the

- system for capturing and archiving legal deposit of Croatian online publications*]. In: 8. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova, uredila Tinka Katić. Zagreb: Hrvatsko knjižničarsko društvo, 2005, p. 61-67.
- Mirna Willer – Miroslav Milinović. *Prema trećoj generaciji knjižnično-informacijskih sustava: hibridna knjižnica za hibridne usluge [Towards the third generation of library information systems: hybrid library for hybrid services]*. In: 8. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova, uredila Tinka Katić. Zagreb: Hrvatsko knjižničarsko društvo, 2005, p. 36-67.
- Sofija Klarin – Tomislav Murati. *Identifikacija, odabir, obradba i osiguravanje dostupnosti mrežnih publikacija u kontekstu nacionalne bibliografske kontrole [Identification, selection, bibliographic and access management of net publications in the context of national bibliographic control]*. In: 4. Seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u kontekstu globalne informacijske infrastrukture: zbornik radova, uredile Mirna Willer i Tinka Katić. Zagreb: Hrvatsko knjižničarsko društvo, 2001, p. 41-56.
- Sofija Klarin – Sonja Pigac. *Hrvatske daljinski dostupne elektroničke serijske publikacije [The Croatian remotely accessible electronic serial publications]*. «Vjesnik bibliotekara Hrvats», 43 (2000 [i.e. 2001.]), n. 4, p. 156-167.
- Johan Steenbakkers. *Setting up a Deposit System for Electronic Publications*. The Hague: Koninklijke Bibliotheek, 2000, <http://nedlib.kb.nl/results/NEDLIBguidelines.pdf>.
- Nacionalna i sveučilišna knjižnica u Zagreb. *Kriteriji odabira obveznog primjerka mrežne grade za obradu i arhiviranje [Selection Criteria for the Choice of the Legal Deposit Copy of Web Publications]*, <http://www.nsk.hr/DigitalLib.aspx?id=83>.
- Nacionalna i sveučilišna knjižnica u Zagreb. *Preporuke za izradu mrežnih publikacija [Recommendations for the creation of Web publications]*, <http://www.nsk.hr/DigitalLib.aspx?id=356>.
- Zakon o knjižnicama [Library law]*. 9.10.2007, <http://www.nn.hr/clanci/sluzbeno/1997/1616.htm>.