

Network of Digital Libraries in Poland as a Model for National and International Cooperation

Cezary Mazurek, Marcin Werla
Poznań Supercomputing and Networking Center, Poland
{mazurek, mwerla}@man.poznan.pl

Abstract. The development of digital libraries in Poland started around 10 years ago. The first Polish digital library was the Digital Library of Wielkopolska made available publicly in October 2002. It is now the largest Polish digital library, giving access to over 130,000 digital objects. One of the keys to the success of this initiative was its organizational model based on regional cooperation. Since 2005 this model has been widely adopted in other regions of Poland, which resulted in a cooperating network of digital libraries. In this paper we provide a brief overview of the development of digital libraries in Poland, followed by a more detailed description of the current state of the digital libraries infrastructure, including the Digital Libraries Federation portal. It is presented with a focus on the cooperation on many levels, the automation of metadata aggregation and processing, and the wide promotion of resources from Polish digital libraries. The paper ends with an outline of future directions of development and possible synergies with the SYNAT national strategic research project.

Keywords: digital libraries infrastructure, metadata aggregation, regional digital libraries, Digital Libraries Federation, Europeana, DART-Europe

1. Introduction

Digital Library Reference Model introduced by the DELOS project [Candela, et al., 2007] describes the digital library as a three-tier construct (see Figure 1). The core component is the Digital Library Management System (DLMS), very often an off-the-shelf component (or set of components), a generic piece of software providing basic functionality required by the particular digital library. On top of such system the second tier is built – the Digital Library System. This system enriches the DLMS with the specific functionality and/or configuration required by the digital library. Finally, the third layer is the Digital Library itself, understood as an organization collecting and preserving digital content and giving access to it.

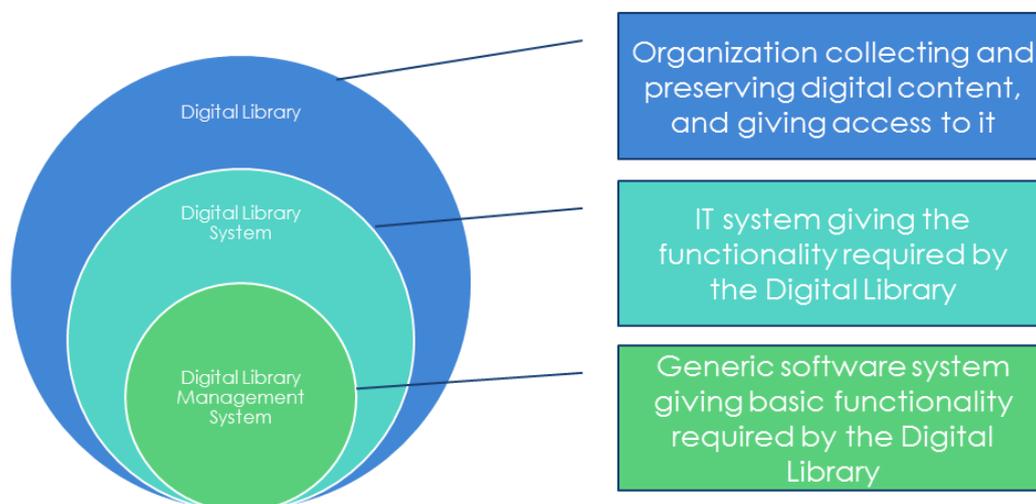


Figure 1. DELOS Digital Library Reference Model

The majority of issues connected with the first two tiers, which may arise while building a digital library, have technical background and are connected with proper choice of the DLMS and its later configuration/extension. A successful creation of the last tier – the organization – is not a purely technical task. It can be associated with significantly more diverse set of issues, including IPR concerns or tasks related to management, logistics or even interpersonal communication.

In the first decade of the 21st century the development of digital libraries in Poland, which is the main topic of this paper, was a bottom-up process led by scientific and public libraries. It was not initiated by any central organizations or government authorities. Instead of following obligatory regulations or policies (because such regulations and policies did not exist), libraries willing to create their digital branches were looking for case studies of analogous activities already performed by similar institutions. Technical and organizational solutions successfully tested in practice were copied and adapted.

The next section of this paper describes those practices, covering both technical and organizational aspects. The third section introduces the Digital Libraries Federation – a network service developed in the PIONIER network in order to facilitate further development of digital libraries in Poland and increase the visibility and use of digital content available in such systems. The fourth section is dedicated to international cooperation of the Digital Libraries Federation with Europeana and DART-Europe. The paper ends with a summary including the directions of future works.

2. From Institutional Repositories to Regional Digital Libraries... and Back

The largest Polish digital library is the Digital Library of Wielkopolska (<http://www.wbc.poznan.pl/>). It is also one of the oldest of such services in Poland. The Digital Library of Wielkopolska was created as a regional initiative of scientific libraries. It was launched publicly in October 2002 [Górny, 2003]. The coordinators of this initiative were the Poznań Foundation of Scientific Libraries and Poznań Supercomputing and Networking Center (PSNC). PSNC was responsible for technical issues including the software platform and network infrastructure. The tasks of Digital Library Management System were in this case performed by the dLibra Digital Library Framework (<http://dlibra.psnc.pl/>), a platform developed by PSNC since 1999. Since the beginning one of important dLibra features was its distributed architecture [Mazurek & Werla, 2005]. This system was designed to allow not only remote access to the digital library resources, but also to facilitate the creation of digital library by a number of scattered content providers. This approach matched very well the needs of the mentioned regional initiative.

Together with the public launch of the Wielkopolska Digital Library, a regional digital library model was created [Mazurek, Parkoła, & Werla, 2006]. In this model a consortium (or simply a group) of institutions from a particular region or city is interested in on-line publishing of their collections. This group designates an institution (sometimes two, rarely more) which is made responsible for the technical infrastructure (or for the Digital Library System from the DELOS model). Such approach has many benefits, including:

- single access point to regional collections – the regional digital library website – which is easier to use and promote/explore than a number of separate institutional repositories, thus is more valuable for end users;
- close cooperation between institutions from the region, which is necessary to maintain the digital library with certain level of quality and may lead for example to new funding possibilities like common projects;
- lowered costs of the technical infrastructure, in most cases shared by digital library members.

This organizational model together with the underlying technical solution was adopted by a similar initiative named Kujawsko-Pomorska Digital Library (<http://kpbc.umk.pl/>), launched in 2005 [Bednarek-Michalska, 2006]. This project was the first Polish digital library co-funded by the European Union. Later on, more Polish regions decided to use the same solution, resulting in around 30 regional digital libraries to date.

An alternative model – institutional repositories – assumes that each institution maintains its own digital library/repository. This model is also quite popular in Poland – the number of institutional repositories is similar to the number of regional digital libraries. The most important premise for creation of such repositories is the need for promotion/identification of the particular institution and its collections. This is especially important in case of research institutions, as their institutional repositories filled with modern research papers show the scientific potential of the institutions' staff.

The increase of importance and popularity of institutional repositories caused slight tensions in the regional consortia. Often regional consortium partners (scientific or university libraries), beside the participation in a regional digital library, wanted to have their institutional repositories. The most obvious decision in such cases was to publish older cultural heritage materials in the regional digital library and to move recent outcomes of the scientific activity of the institution to its institutional repository. An example of such division is the Adam Mickiewicz University from Poznań. The library of this university is one of the main contributors to the Wielkopolska Digital Library. On the other hand the university has also an institutional repository – AMUR (<http://repozytorium.amu.edu.pl/>) [Rychlik & Karwasińska, 2010]. Such approach, although somehow justified from the university's point of view, decreases the positive effect of the regional digital library model, because:

- the university has to maintain the repository and participate in maintenance of the regional digital library;
- end users no longer have access to all university content through one regional portal – they have to visit the regional digital library and/or institutional repository, depending on the type of content they are looking for.

In 2010 PSNC in cooperation with the Poznań University of Technology proposed an extension to the regional digital library model (see Figure 2) and offered support for it in the dLibra software. The extension is based on the assumption that each participant of the regional digital library should have the possibility to set up a separate (institutional, thematic etc.) portal giving access to a selected part of resources available in the regional digital library – in most cases the part contributed by particular institution.

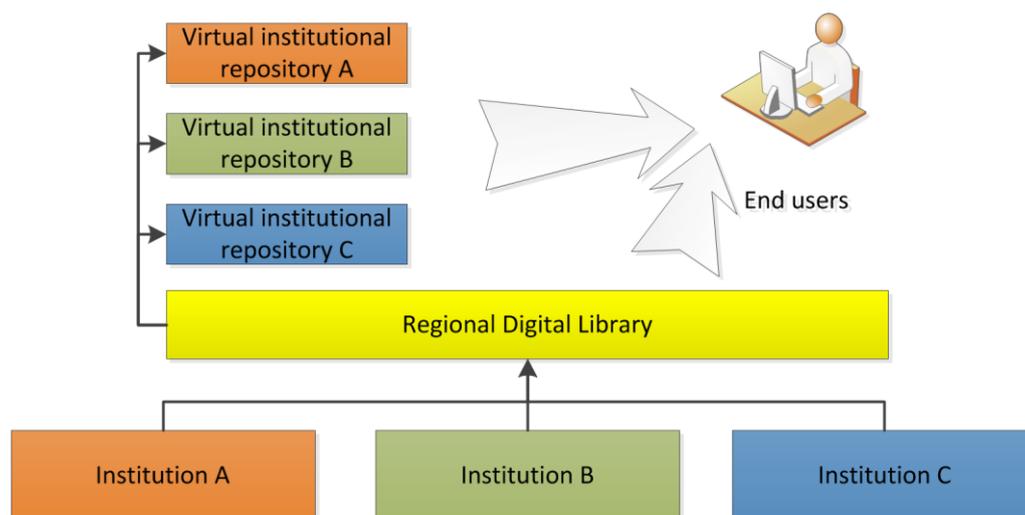


Figure 2. Virtual institutional repositories built on top of regional digital libraries

Thanks to the new features of the dLibra system, the institutional portal has the same functionality as the regional digital library. Users of the regional digital library do not have to use multiple websites and the maintenance issues are still limited to the regional digital library and its consortium. Also, there is no need to publish different types of collections on different websites, with different tools. The described extension and its implementation in the dLibra system will be used in 2011 in at least two deployments:

- to create the Digital Library of the Poznań University of Technology on top of the Wielkopolska Digital Library [Bajer & Ober, 2010];
- to create a shared repository for libraries of over a dozen of research institutes of the Polish Academy of Sciences and a series of virtual repositories for those institutes on top of the shared repository.

The intensive growth of the number of digital libraries and repositories which began in Poland in 2005/2006 was for PSNC the motivation to create a technical solution allowing to integrate existing scattered services into a country-scale virtual infrastructure of digital libraries. The core element of this infrastructure is described in the next section of the paper.

3. Digital Libraries Federation, or How to Cooperate (Automatically)

In June 2007 Poznań Supercomputing and Networking Center made available the Digital Libraries Federation portal (<http://fbc.pionier.net.pl/>), developed as a part of the activities related to the PIONIER Network (the Polish NREN, <http://www.pionier.net.pl/>). One of the main aims of this portal was to create a single point of access to the resources distributed in tens of digital libraries cooperating with the Federation. The Federation mission was defined in the following way:

- To facilitate the use of resources from Polish digital libraries,
- To increase the visibility and popularity of resources from Polish digital libraries in the internet,
- To enable new advanced network services based on the resources from Polish digital libraries to internet users and digital libraries creators.

It was also assumed that cooperation with the Federation will be free of charge and based on open standards, and the digital content will be stored and accessed on the level of digital libraries, while the Federation will be copying and using the metadata describing the content [Mazurek, Parkoła, & Werla, 2008].

Technically, the data aggregation process was implemented with the use of the OAI-PMH protocol. The aggregated metadata is processed, stored and indexed on the Federation servers to allow fast and easy access to it. Presently the main functions of the Federation are the following:

- Basic and advanced search in the aggregated metadata – users can input a query (or a series of queries connected with Boolean operators) and as a result they get the list of objects which metadata matches the conditions expressed by the query. Each search result contains also a link to the on-line version of the found object. A special kind of search is based on persistent identifiers used together with the OAI-PMH protocol (the OAI Id schema). The DLF has a REST service available at <http://fbc.pionier.net.pl/id/> allowing to retrieve current URL the object with particular identifier. For example the URL:

<http://fbc.pionier.net.pl/id/oai:www.wbc.poznan.pl:8711>

resolves identifier oai:www.wbc.poznan.pl:8711 into:

<http://www.wbc.poznan.pl/dlibra/docmetadata?id=8711&showContent=true>

For objects having other identifiers in their metadata (like ISBN numbers) it also works.

- Access to digitisation plans – besides aggregating information about the objects which are available on-line, the Federation also aggregates information about objects which are planned for digitisation, if particular digital library publishes such data. This information is processed in

the same way as the data of digitized object, additionally a PDF report with all digitation plans is generated.

- Support for automated coordination of digitisation – the DLF not only aggregates the metadata, but it also provides a service allowing to perform an automated query to find objects registered in the Federation database (both digitized and planned for digitization). This service takes the metadata of an object which some digital library would like to add to its digitisation plans and checks if there already is an object with very similar metadata. The comparison is based on heuristic algorithm utilizing fuzzy search and the information about title, creator and publication date. The results can be used by a digital library to decide which objects should be digitised. Support for this mechanism is implemented in the dLibra software and is a standard element of digitisation workflow in many Polish cultural heritage institutions. The same algorithm is used to check the DLF database for already existing duplicates. At the moment the number of such automatically detected duplicates is lower than 0.2% of all already digitised objects.
- Database of digital libraries – The DLF is managing a database of Polish digital libraries available at <http://fbc.pionier.net.pl/owoc/list-libs>. Any institution can register its digital library in this database (even if it is not OAI-PMH compatible), and after the new entry is accepted by the DLF administrator, it is published on the DLF website. Each entry contains basic descriptive information (like name, URL and description), but also a list of cooperating institutions with their geographical location and contact address (in case of regional digital libraries), some recommended objects and the technical information about the OAI-PMH protocol.
- Statistics showing the current state and growth of the DLF resources.

Such set of features makes the Federation attractive for end users, but also for digital libraries creators and contributors. At the end of April 2011 the number of digital libraries registered in the Federation database was 67. The number of objects was around 615,000 (plus around 20,000 in the digitisation plans). The overall number of objects to which the Federation portal directed its users in 2010 was close to 8.4 million.

With a well-developed infrastructure of digital libraries on the country level, it was natural to start works focused on the integration of this infrastructure with important European and international initiatives. The first step was the connection with Europeana (<http://europeana.eu/>) and DART-Europe (<http://www.dart-europe.eu/>). This is the subject of the next section.

4. From Chelm to the Hague, from Rzeszów to London

In June 2008 PSNC started cooperation with Europeana as a participant of the EuropeanaLocal project (<http://europeanalocal.eu/>). The main aim of this three year project co-funded by the European Commission under the eContent*Plus* programme was to develop a Europe-wide network of digital libraries and metadata aggregators cooperating together and delivering metadata to Europeana. At the same time, the project was also kind of a large testbed for technical standards, guidelines and tools developed by Europeana, as the members of the project were very often the first to adopt these in the production environment.

One of the outcomes of this project was the establishment of a metadata transfer process between the Federation and Europeana at the end of 2009 [Werla, 2010]. The result of the first data transfer were 257,000 new objects in Europeana, after which the ratio of Polish content reached of 5.5% of all Europeana data. The data transfer is repeated periodically. The most recent operation was performed in April 2011. The exact number of ingested metadata records was 551,176.

The role of the Federation is not only that of a simple data proxy. The aggregated metadata is cleaned, normalized and mapped from the Dublin Core metadata schema into the Europeana

Semantic Elements – a schema obligatory for all Europeana data providers [Mazurek, Stroiński, Werla, & Węglarz, 2011]. Such operations are complex, and require both programming skills and a good knowledge of the Europeana-related standards and guidelines [Europeana Group - Technical Requirements, 2011]. Chełm, a town mentioned in the title of this section, is the home of the most eastward located digital library cooperating with the Polish Federation, available on-line at <http://cyfrowa.chbp.chelm.pl/>. It contains over 4,000 digital objects and is maintained by the Chełm Public Library. It was made publicly available on the 1st of December 2009. During the following eight months their data was transferred first to the Digital Libraries Federation, and then to Europeana. The data transfer process was handled fully automatically and required no additional effort from the digital library administrator. Most probably, without a project like EuropeanaLocal and a digital libraries infrastructure like the one created in Poland, a local initiative managed by small memory institutions would have had a significantly harder and longer way to the Hague – where the headquarters of the Europeana are located.

The metadata aggregation and provisioning mechanism implemented in the Federation and used for connection with Europeana was extended with the support for dynamic sets by the OAI-PMH protocol [Mazurek & Werla, 2008]. This extension allows the OAI-PMH protocol to be used to obtain from the Federation a set of metadata records that matches ad hoc defined criteria. In such case OAI-PMH service providers do not have to rely on sets defined by the data provider, they can simply create a query describing what kind of objects they would like to obtain and then use this query to create and harvest a dynamic set.

This approach was applied in January 2010 to establish cooperation between PSNC and the DART-Europe portal hosted by the University College London and endorsed by the Association of European Research Libraries (LIBER). The aim of the portal is to improve the global access to European research theses [Moyle, 2008]. The technical mechanism for data retrieval is the same as in case of Europeana – the OAI-PMH protocol. However, it is worth noting that Europeana updates the data from the Federation on demand (a few times a year), while DART-Europe is does it in a fully automated way on the nightly basis – as often as the Federation is updates data from its content providers.

Quite an interesting conclusion can be drawn from the analysis of the percentage share of data provided by the Federation in the whole dataset of DART-Europe portal. At the end of April 2010 the DLF provided metadata of almost 2,000 open-access research theses, which constitutes around 1% of the overall amount of metadata records accessible in the portal. On the other hand, the 34 polish universities providing this data are around 10% of the overall number of universities cooperating with DART-Europe. The fact that only 1% of all theses come from Poland shows that there is a lot to do in Poland in the matter of open access publishing of research theses. But the 10% of universities shows that the data aggregation infrastructure is in place and as soon as new e-theses are published, their metadata can be transferred to the European portal. For example the University of Rzeszów is one of partners of the regional Podkarpacka Digital Library. If a librarian from this university publishes some theses, on the next day it is visible in the Federation portal, and a day later it is also visible in the DART-Europe portal. This seems to be quite good a pace for the dissemination of research outcomes. It is implemented as an automated process, requiring minimum maintenance efforts.

The cooperation with Europeana and DART-Europe portals described in this section are good proofs that the network of digital libraries developed in Poland, based on loose cooperation between many diverse institutions, brings benefits both to its participants and end users – also on an international scale.

5. Summary

In this paper we have described how the environment of digital libraries in Poland evolved from few dispersed local initiatives to a consistent network of cooperating regional digital libraries and

institutional repositories with a fully automated metadata exchange and a connection with key European infrastructure. Crucial elements which allowed to achieve the present state were:

- the will of cooperation between different types of institutions, like public and academic libraries, but also archives, museums, NGOs or even private persons – on regional and national level;
- technical solutions providing means to realize such cooperation, based in open standards and developed continuously, in close cooperation with users.

In Poland these solutions (the dLibra software and the Digital Libraries Federation service) are developed by PSNC and are widely used. Besides, the fact that the Federation is based on open standards allows to use any software compliant with these standards to join it.

In August 2010 a Polish national research project named SYNAT (<http://www.synat.pl/>) was started. This project, financed by the National Center for Research and Development, is aimed to conduct a research task titled “Creation of universal, open, repository platform for hosting and communication of networked resources of knowledge for science, education and open society of knowledge”. It is a part of Strategic Research Programme “Interdisciplinary system of interactive scientific and technical information”. The project is coordinated by the Warsaw University – ICM, and Poznań Supercomputing and Networking Center is one of the main partners. This project will have a significant influence on the directions of development of digital libraries in Poland.

Bibliography

- Bajer, J. & Ober, K. (2010). Wykorzystanie regionalnej biblioteki cyfrowej do tworzenia repozytorium instytucjonalnego. *Konferencja "Polskie Biblioteki Cyfrowe 2010"*. Poznań: Poznańskie Centrum Superkomputerowo-Sieciowe.
- Bednarek-Michalska, B. (2006, July). Kujawsko-Pomorska Biblioteka Cyfrowa – pragmatyka tworzenia biblioteki cyfrowej. *Biuletyn EBIB(77)*. <http://hdl.handle.net/10760/8786>
- Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C. et al.. (2007). *The DELOS Digital Library Reference Model*. Retrieved April 30, 2011 from location DELOS - Network of Excellence on Digital Libraries: http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf
- Europeana Group - Technical Requirements*. Retrieved April 30, 2011 from location <https://version1.europeana.eu/web/guest/technical-requirements/>
- Górny, M., Gruszczyński, P., Mazurek, C., Nikisch, J. A., Stroiński, M., & Swędryński, A. (2003). Zastosowanie oprogramowania dLibra do budowy Wielkopolskiej Biblioteki Cyfrowej. *Zeszyty Naukowe Wydziału ETI Politechniki Gdańskiej. Technologie Informacyjne*. (pp. 109-117). Gdańsk: Wydawnictwo Politechniki Gdańskiej. <http://dl.psync.pl/biblioteka/publication/43>
- Mazurek, C., & Werla, M. (2005). Distributed Services Architecture in dLibra Digital Library Framework. Schloss Dagstuhl: DELOS Network of Excellence. http://www.delos.info/files/pdf/Proceedings/Dagstuhl_2903_010405/delos-dagstuhl-handout-all.pdf
- Mazurek, C., & Werla, M. (2008). Extending OAI-PMH Protocol with Dynamic Sets Definitions Using CQL Language. *IADIS International Conference Information Systems 2008*. Algarve, Portugal: IADIS. <http://dl.psync.pl/biblioteka/publication/176>

- Mazurek, C., & Werla, M. (2009). Europeana Local: Rola projektu w europejskiej infrastrukturze bibliotek cyfrowych. *PIONIER Magazine* (nr 3 (04) / 2009).
<http://dl.psnc.pl/biblioteka/publication/218>
- Mazurek, C., Parkoła, T., & Werla, M. (2008). Building Federation of Digital Libraries Basing on Concept of Atomic Services. *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2008*. Pittsburgh, PA, USA: ACM. <http://dx.doi.org/10.1145/1378889.1379003>
- Mazurek, C., Parkoła, T., & Werla, M. (2006). Distributed Digital Libraries Platform in the PIONIER Network. *Lecture Notes in Computer Science vol. 4172* (pp. 488-491). Berlin: Springer.
http://dx.doi.org/10.1007/11863878_50
- Mazurek, C., Stroiński, M., Werla, M., & Węglarz, J. (2011). Distributed Services and Metadata Flow in the Polish Federation of Digital Libraries. *International Conference on Information Society (i-Society 2011)*. Londyn: IEEE.
- Moyle, M. (2008). Improving access to European e-theses: the DART-Europe Programme. *LIBER Quarterly*(18 (3/4)), pp. 413-423. <http://eprints.ucl.ac.uk/14006/>
- Rychlik, M., & Karwasińska, E. (2010). AMUR - Repozytorium Uniwersytetu im. Adama Mickiewicza w Poznaniu. *II Międzynarodowa Konferencja Open Access w Polsce. V Konferencja EBIB Internet w bibliotekach*. Warszawa: Stowarzyszenie Bibliotekarzy Polskich, Komisja Wydawnictw Elektronicznych.
http://www.ebib.info/publikacje/matkonf/mat21/rychlik_karwasinska.php
- Werla, M. (2010). Polskie biblioteki cyfrowe, FBC i Europeana – etapy i bariery w przepływie informacji. *Biuletyn EBIB*(Nr 1/2010 (110)). <http://www.ebib.info/2010/110/a.php?werla>