

Federating Digital Library Services for Advanced Applications in Science and Education

Michał Kosiedowski, Cezary Mazurek, Maciej Stroiński, Marcin Werla, Marcin Wolski

Poznań Supercomputing and Networking Center, Poznań, Poland
e-mail: {kat/mazurek/stroins/mwerla/maw}@man.poznan.pl

(Rec. December 20, 2007)

Abstract: Over the recent years numerous digital libraries have been deployed and enabled in the Internet with the use of various digital library frameworks. They contain valuable digital resources use of which is important to a large community of users, especially within the scientific and educational communities. The access to distributed digital resources has been a subject of research for a few years now. In this paper we present our own work in this area, discussing our approach at the distributed digital library platform that is widely used within the PIONIER optical network. We present the methodology for building distributed search and access to digital publications that allowed to create a national federation of digital libraries and that is now under investigation for possible use within various application fields including medicine, science and education.

Key words: digital libraries, metadata

I. INTRODUCTION

One of the objectives of the national programme *PIONIER – Polish Optical Internet. Advanced Applications, Services and Technologies for the Information Society* realized in years 2001-2006 [1] was to enrich the content-based services in Polish NREN. To reach this goal several digital library installations have been launched providing an assistance to librarians and university publishers in digital content management and publishing. Continuing this effort and reflecting expectations of the users (e.g. authorized local access to academic scripts, identification of the owner of manuscripts, preservation of regional cultural heritage, etc.), PIONIER introduced the concept of regional digital libraries. In parallel, there were other actions taken towards building advanced services for information society, covering also the area of access to and processing of huge amounts of data. They implied the new approach for building digital library infrastructure in the PIONIER network. After providing solid and robust applications for storage of scientific content in digital libraries and institutional repositories, new requirements appeared, requesting the same level of infrastructure for scientific data as well as data management support for advanced scientific applications. This required strong focus on interoperability issues within heterogenous digital library environments as well as between different services in PIONIER network (e.g. grid computing, telemedicine services, media

delivery, etc.). Therefore, we introduced the concept of atomic services for digital libraries which allowed to build advanced scientific applications on top tier including digital library services and other middleware services interoperating with one another.

In the following sections we present the outcome of our work starting with description of distributed digital library platform in Section II. Then, we present the concept of atomic services followed by discussion of advanced content- and metadata-based services. Finally, we present three examples of advanced applications built with digital library services: applications for human sciences, medical applications and virtual laboratories.

II. DISTRIBUTED DIGITAL LIBRARY PLATFORM IN PIONIER NETWORK

The first digital library installation in Poland started in 2002 as the result of the cooperation between Poznań Supercomputing and Networking Center and Poznań Foundation for Scientific Libraries. Digital Library of Wielkopolska Region, as it was named, was built with dLibra Digital Library Framework which had been developed by PSNC since 1999 [2].

dLibra Digital Library Framework is a distributed and portable digital library framework. It is created to be a highly configurable software basis for digital libraries. dLibra-

based digital library can be used to preserve, manage and access digital objects consisting of content (text, sound, video, etc.) and metadata – object description in the user defined form. Digital objects in the dLibra Framework are organized in a hierarchical way using so called directories and group publications. Additionally, to present publications for the needs of Internet users and distributed search mechanism, dLibra uses collections of objects defined by the digital library administrator.

Thanks to the close cooperation with libraries in following years further installations of digital libraries have been started. By the end of 2005, four regional digital libraries and additionally several institutional installations were deployed. Currently there are 9 regional and 7 institutional dlibra-based digital libraries in the PIONIER network and at least five more are planned to be available by the end of 2007 (Fig. 1). All these libraries create together a distributed platform of digital libraries with over 100 000 digital objects including writing relicts, manuscripts, cultural heritage, regional and educational materials, etc. Additionally, the PIONIER platform of digital libraries provides some value-added services for its users, e.g. searching through distributed resources, unique object identifiers, virtual collections of digital objects.

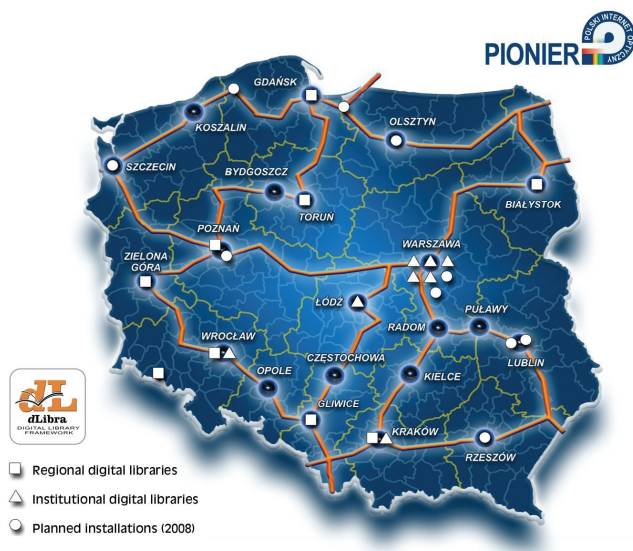


Fig. 1. dLibra installations in the PIONIER network

One of the most important functionalities of dLibra provided for all PIONIER digital libraries is periodic metadata synchronization between all installations, based on the OAI-PMH protocol [4]. This process is based on selective harvesting of metadata about the added, modified and deleted records. Custom collections defined in digital libraries are transformed into OAI-PMH sets and custom metadata

schemas are transformed into DCMES attributes. The functionality applied to all PIONIER digital libraries transformed those libraries into one distributed platform where each digital library has full information about all metadata in the platform. This metadata can be used by each digital library to provide its users with means to access all resources stored in all other digital libraries and to create new advanced content- and metadata-based services.

III. DIGITAL LIBRARY AS A SET OF ATOMIC SERVICES

Digital libraries with their universal functionality [3] may be even more flexible and reusable, if designed in a service-oriented manner [5, 18]. Such design should allow to decrease the effort of the creation of new digital libraries and the effort of maintenance and scaling of currently existing large installations. Standardized interfaces to such services should also allow to reuse them in other systems and to use services from other systems as components of digital libraries.

In [17] we have analyzed and compared the functionality of world most popular digital library management systems like Fedora (<http://fedora.info/>), DSpace (<http://www.dspace.org/>) or OpenDLib (<http://www.opendlib.com/>). This analysis led us to the distinction of four basic groups of digital library functional requirements: storage of digital objects, storage of metadata and annotations on digital objects, searching in metadata and in digital objects content and the composition of complex digital objects and creation of relations between digital objects. Further analysis of those functional groups conducted us to the definition of the following atomic services for digital libraries [16]:

- *Content storage service*, which should allow to store any type of digital object content and to version it. This service can be based on various, already existing, network services like FTP or WebDAV servers or even versioning systems like SVN or CVS. The condition of the service content type independence is a result of an assumption, that the digital object content will be analyzed by the service clients and only the clients need to be aware of the content internal format, structure etc.
- *Metadata storage service*, which should allow to store metadata describing digital objects. There may be many metadata records related to one object and each of those records may have different internal structure. The metadata should be defined and stored in XML. Such approach gives flexibility of the metadata storage and allows to use tools like XQuery and

XPath for metadata processing. The service may be implemented with an XML database, for example Xindice (<http://xml.apache.org/xindice/>). This service may also be responsible for the storage of user annotations on digital objects. Such annotations may be treated just as any other metadata records associated with the digital object, but they have an additional user context.

- *Composition and relation service*, which should allow to define various relations between digital objects and to define multiple groups of related objects. Those relations may be used to create complex digital objects, to show parent/child relationships between objects or to create collections of digital objects. This service may be implemented with one of many existing RDF frameworks (like Jena - <http://jena.sf.net/>).
- *Search service*, which should allow to search in previously defined sets of digital objects. Those may be objects with specific content type or objects described with a particular metadata schema. This service should be able to generate a list of digital object references matching a query passed by the service client. The implementation of this service may be very different, depending on the type of resources that will be indexed and searched and on the supported query language.

Services described above should allow to compose a digital library, giving the core functionality required in such systems. The above description is focused on functional requirements of digital library services. The non-functional aspects like access authorization or performance were not included. It was assumed that such aspects are strongly connected with a particular digital library

deployment and should be analyzed within the context of such deployment.

IV. ADVANCED CONTENT- AND METADATA-BASED SERVICES IN DIGITAL LIBRARIES

Atomic services described above are the basis for building distributed digital libraries. They are used by digital library applications to provide functionality required by the application level services. Application level services are more complex services, which can be used directly by the end user or by other atomic or application level services. An example schema of the service composition and reuse is shown in Fig. 2.

This figure presents two digital libraries. The one placed in the bottom of the figure is a basic digital library composed of four atomic services described in the previous section. On the application level this library can provide additional services for its end users, like private collections or annotations. Private collections are collections of digital objects created directly by end users, as opposed to public collections which are created by the digital library staff and are visible to all library users. Private collections can be easily implemented with the composition and relation service. A similar service is a private documents service which allows its users to create their own digital objects composed of other objects (or their parts) already existing in the digital library. Another example of an application level service may be the annotation service, allowing its users to

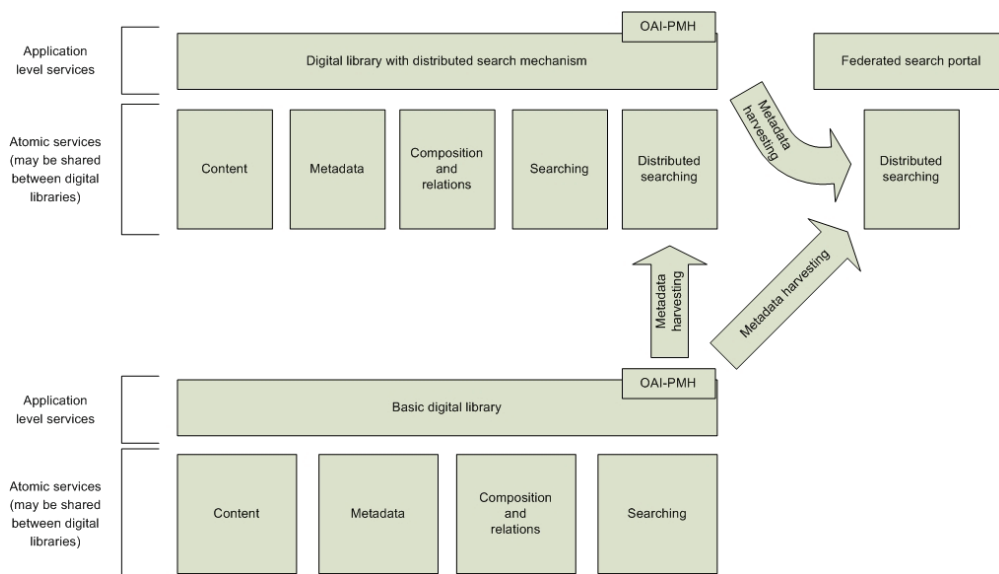


Fig. 2. Example atomic and application level services composition

annotate digital objects with any kind of metadata. Similarly to the private collections, annotations are also visible only to the users who have created them. Annotations can be based on the metadata storage service. All data created with application services described above (private collections, documents and annotations) is, as mentioned, private by default. However, it can be easily shared among users, if the digital library allows it through additional community services.

The other digital library presented in the upper part of the example figure is an advanced digital library composed of four atomic services described in section III, additionally featuring a specialized version of the search service – the distributed search service. This service harvests metadata from various remote digital libraries with the OAI-PMH protocol and exposes searching and browsing in the harvested metadata functionality to its users. Such a specialized atomic service can be used to build new application level services based on the metadata distributed in multiple digital libraries available on the Internet. The following application services based on distributed metadata shared in this model were deployed in the PIONIER network [7]:

- *Distributed search*, which allows to search through resources of all OAI-PMH compatible digital libraries available in the PIONIER network, from the interface of any of those libraries.
- *Virtual dynamic collections*, which allow users to define private collections containing digital objects from several digital libraries. Such collections are defined as sets of conditions for the digital objects metadata (for example: *subject="physics" and type="article"*) and are available as RSS feeds.
- *Unique identifiers resolving*, which allows users to easily identify any digital object and reference to it. It is based on the identifier schema defined in the OAI-PMH protocol specification.

The distributed search service may also be a basis for other application services, like a federated search portal dedicated to a specific domain. Such case is described in the next section.

V. BUILDING ADVANCED APPLICATIONS WITH DIGITAL LIBRARY SERVICES

A. Federation of Digital Libraries for Human Science

Digital Library Federation application (DLF) aims at logical linking of all digital libraries and repositories available in the Polish Internet together with providing new advanced network services realized in this environment.

Additionally, the application endeavors to introduce better usability of resources from the Polish digital libraries and repositories as well as to ensure increased worldwide visibility of those resources.

The DLF application has been designed and developed at PSNC. From the very beginning it was assumed that the communication layer providing access to digital libraries and repositories will be based on open protocols and data formats. For such a reason the OAI-PMH protocol and Dublin Core schema have been chosen as the technical assumption for implementation phase. This assumption lead to another important feature of DLF: it is independent of particular software framework as well as of software developer for any federated digital library or repository. The only requirement against a repository, which is to be federated in DLF is the provision of an open interface for communication through the OAI-PMH protocol.

Formally, the DLF application is based on a free accession of particular repository to the federation of services. There is no need to copy resources or do any implementation work for the repository administrator. What is only required is the URL identifier which enables metadata harvesting through the OAI-PMH protocol for services of DLF. This model is similar to other OAI-PMH-based servers, e.g. OAIster (<http://www.oaister.org/>).

The DLF application automatically and periodically harvests metadata from federated digital libraries collecting in this way information about all digital objects and storing this information in the local directory. Basing on this directory the DLF application provides value-added services for users who can browse and search in distributed platform of digital libraries (Fig. 3).

The functionality provided by the DLF application is based on atomic services implemented by the dLibra framework. However, it is important to note that any other digital library framework could be a basis for deploying a digital library into the discussed federation. Communication within the federation is based on open protocols and thus the DLF application is independent of specific digital library framework functionality.

The functionality provided by the DLF application is based on two instances of the atomic search service. The first instance is responsible for harvesting, indexing and searching the metadata about all digital objects available in digital libraries in the federation. The second instance has similar functionality, but it operates on metadata about resources that are planned for publication in digital libraries in the federation. Such configuration allowed to provide the following functions for the whole federation of digital libraries:

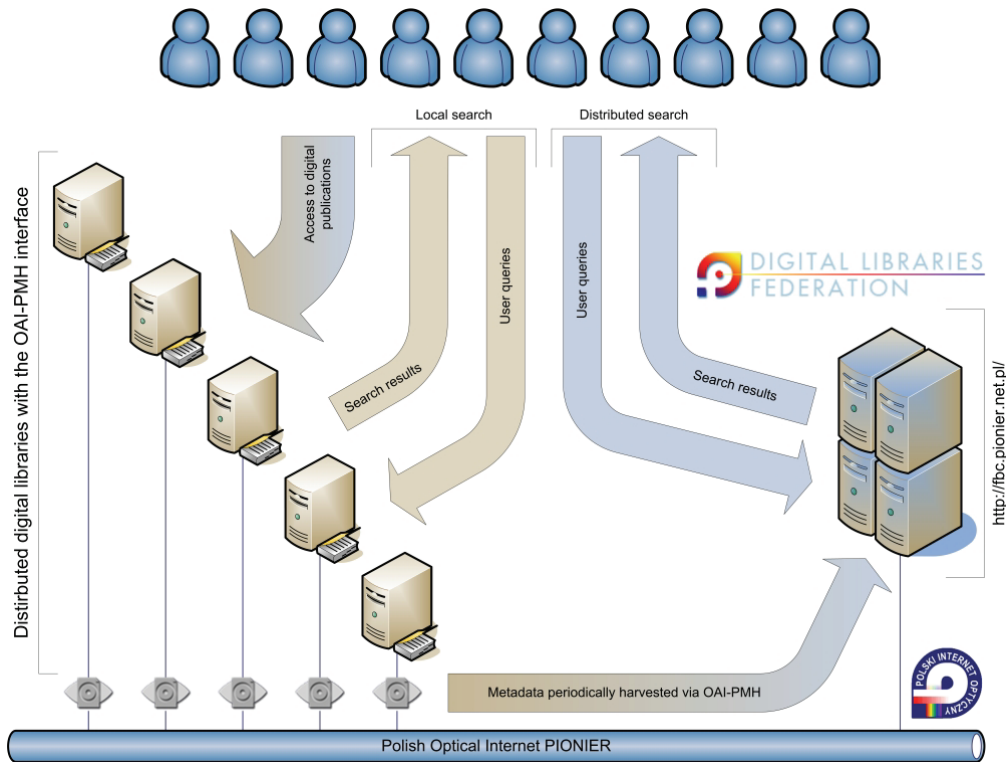


Fig. 3. Digital Library Federation application

- distributed search,
- persistent digital object references,
- coordination of digitalization.

This set of DLF functions supports the development of research infrastructure for human sciences even if resources are not stored in the same place where services are enabled. We will now describe the functions mentioned above in details.

1) *Distributed search*: Searching through distributed resources of digital libraries is possible in one of two ways. The first use-case which was possible also before the DLF application has been launched, allows searching locally in every digital library and collect all results step by step. The second use-case available now, allows the use of distributed searching within the DLF portal. In this case users receive one common list of search results and by choosing one entry they are redirected to the digital library where the selected object is actually stored. The example of search results is presented in Fig. 4.

It is also worth noting that this search function can be added to any webpage. Web administrators interested in providing access to the DLF services from their webpage can embed a DLF search window through simple modifications of the page's HTML code. A plug-in based on OpenSearch format for embedding into the web browser toolbar has also been implemented.

2) *Persistent digital object references*: The service of resolving unique OAI identifiers provided within the DLF portal is based on the functionality of automatic identifier assignment for digital objects provided through OAI-PMH. The identifier format is based on OAI-PMH specification and is constructed in the following way: *oai:<repository domain>:< resource id within repository>*, and its example instance might look as follows: *oai:www.wbc.poznan.pl:8711*.

The DLF application provides the mechanism of resolving these identifiers which allows users or other applications to obtain information and current location of a particular digital object. This mechanism is ready to use also in building persistent references to digital objects e.g. in bibliography. Such a reference is built as follows: *http://fbc.pionier.net.pl/id/<oai_idenfifier>*, and its example instance might be then as the following: *http://fbc.pionier.net.pl/id/oai: www.wbc.poznan.pl:8711*.

Accessing such an address from any web browser will allow to browse basic metadata on the resolved object together with references to full metadata records and to object content (e.g. text, image, etc.).

Through the mechanism of the persistent digital object identifiers the DLF application can provide persistent object identification despite of digital library framework used to manage it as well as despite of current location of the object.

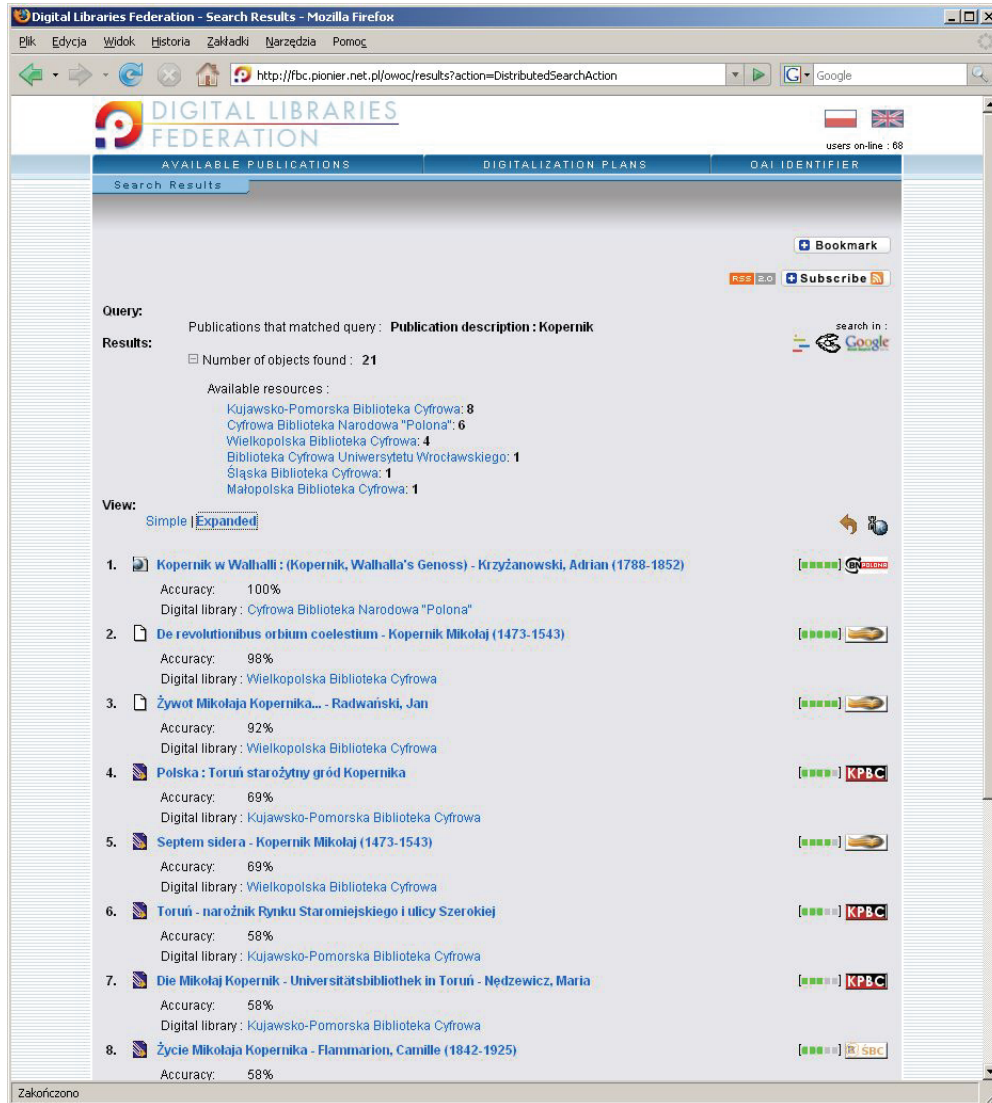


Fig. 4. An example result of a distributed search performed on the PIONIER Federation of Digital Libraries

3) *Coordination of digitalization*: The extraordinary functionality provided within the DLF application is searching through the collection of resources planned for digitalization in any digital library connected to the federation. This function is based on information collected through OAI-PMH from all digital libraries which manage the list of planned publications. Besides of searching it is also possible to browse the full list of planned publications as well as to find out the possible duplications of digitalization efforts.

According to our knowledge, the dLibra Framework is currently the only digital library framework supporting digitalization plans functionality and providing this information through open interfaces. Since 2005 the publication lifecycle in dLibra has been beginning with "planned publication" state, which allows to store metadata about digital

object which is yet neither available nor existing [6]. Such an instance of a digital object might be then transferred to the full shape by appending publication content, namely uploading publication files to the digital library. At the time of writing the current paper there are more than 1 200 planned publications registered by the DLF portal. The future work concerning this function aims at building a social platform for coordination where users (i.e. scientists) will be allowed to apply for digitalization of particular volumes.

B. Medical Digital Library for Interhospital Teleconsultations

Medical Digital Library is planned as the next step in the development of an advanced telemedical system for the

support of the regional healthcare in the region of Wielkopolska. It is developed within the Telemedycyna Wielkopolska project, which is an interdisciplinary collaboration between Poznań Supercomputing and Networking Center, the Institute of Computing Science of Poznań University of Technology, and the Division of Trauma, Burns and Plastic Surgery of Poznań University of Medical Sciences [9]. Although the system and the tools are designed to be used by various medical domains, the pilot use case scenario is based around services for the support of trauma treatment. To this end a system of medical teleconsultations in the area of trauma has been designed and deployed as a pilot for use within several hospitals in Wielkopolska. Further works of the project focus on developing a multimedia Medical Digital Library that will store and enable information and knowledge in the area of trauma. The first service delivered based on this digital library is the Registry of Medical Cases, again with a prototype designed for use in trauma.

The Medical Digital Library is being built using the well-established technologies known from the grid computing domain. This includes solutions in the area of data storage, data processing, security, and data and service access. Especially interesting in the context of digital library federations are two of these areas, that is data storage and data access. The first is done within the Medical Digital Library with the use of the Grid Data Management System (GDMS) [10] which has been designed within the PROGRESS project [11] and further developed within such projects as SGIgrid [12] and ACGT [14]. GDMS provides mechanisms for file virtualization on the grid. It offers its functionality via GSI-based Web Services that allow to manage the stored files, metadata and file access rights. As far as the data access is concerned, the architecture proposed by the Grid Service Provider within the PROGRESS project has been adopted to enable concurrent access from multiple user applications [13]. The Grid Service Provider introduced a flexible architecture of accessing grid services at high-level of granularity to facilitate construction of grid user interfaces. This allows to build user access points on various types of terminals including web browsers, desktop computers and mobile terminals relatively easily. The above-mentioned Registry of Medical Cases is such a high-level service that enables easy creation and management of digital “publications” – the medical cases – and facilitates access to them and any digital objects that compose individual medical cases.

While analyzing the architecture of the Medical Digital Library one can easily notice that three of the four atomic digital library services discussed in section III are directly available within this environment thus enabling possible

federation with other digital libraries in the future. The content storage service is offered within the Medical Digital Library through GDMS’s Data Container services. Data Containers are Web Services that enable storage of data on various types of media: generic file systems, tape archivers or relational databases. The transmission of data to and from the system takes place with the use of various data transfer protocols such as, for example, GridFTP or FTP. The metadata storage service is offered by GDMS’s Metadata Repository service, which allows to assign metadata to digital objects using any metadata scheme, either selected from those already available with the system, for example Dublin Core, or created from scratch. Finally, the search service is offered by the Data Broker service, which allows to search for digital objects depending on any parameter, starting with the file name and ending with any metadata attribute associated with individual digital objects stored within GDMS.

The fourth atomic digital library service identified within the afore-mentioned research project conducted by Poznań Supercomputing and Networking Center, that is the service for composition of digital objects and creation of relations between them, is delivered within the Medical Digital Library through a specific advanced telemedical service such as the Registry of Medical Cases. This means that for each such service there will exist a specific type of “publications” and an instance of the composition and relation service for each of these “publication” types (Fig. 5). The objective of the Registry of Medical Cases is to enable management of medical cases, containing detailed descriptions of individual cases encountered by doctors in their everyday practice, together with their classifications according to medical standards and descriptions of the treatment process. The Registry creates relations between the medical domain descriptions (metadata), the actual medical case description data (textual content) and illustrative material (multimedia content). Since the medical cases can belong to only one medical domain (for example trauma), the Registry automatically creates “collections” of medical cases belonging to a specific domain. Further on, the concept of the Registry of Medical Cases foresees that certain cases stored in the Registry will be assigned the reference case status, thus enabling to create a “collection” of reference cases for specific medical domains. While this functionality is very limited and covers only a small part of the functionality of the discussed atomic digital library service it could be relatively easily encapsulated under an umbrella of a service enabling the full functionality of the atomic digital library service in question to enable Medical Digital Library collections within digital library federations. This is planned for the next version of the Registry.

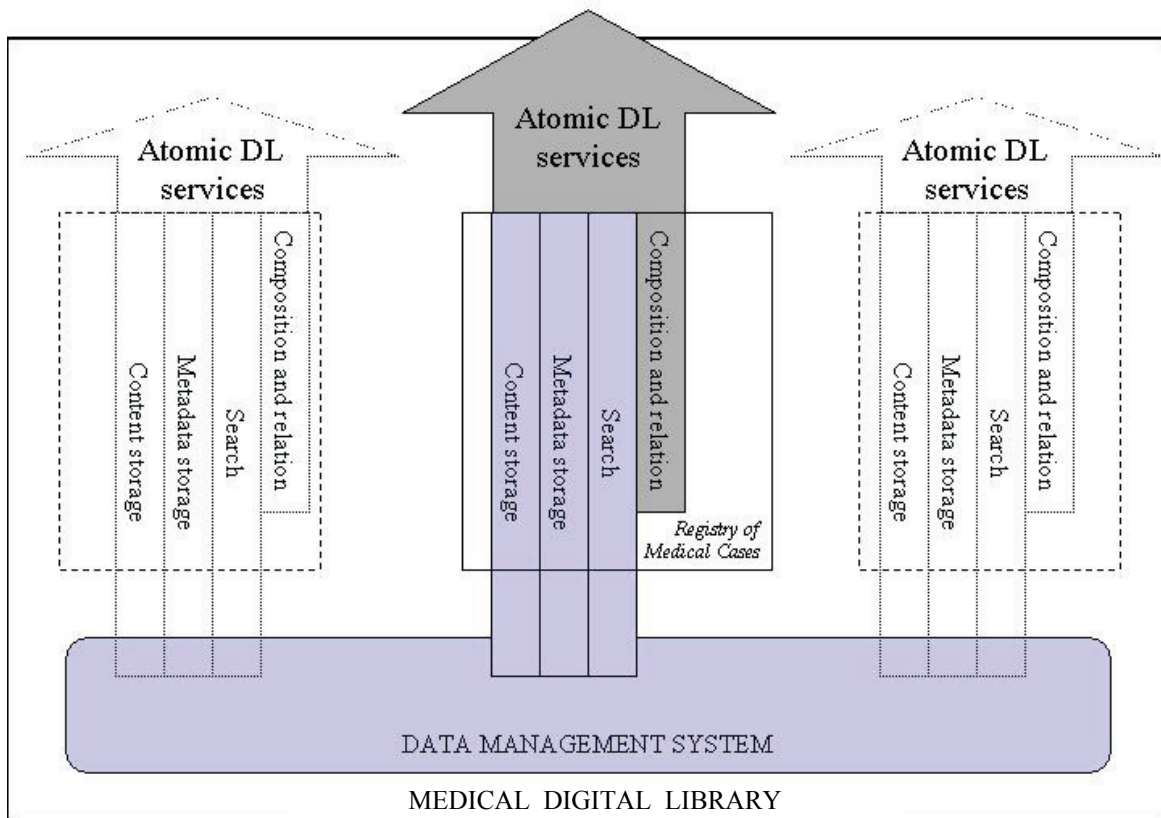


Fig. 5. Atomic digital library services in the Medical Digital Library

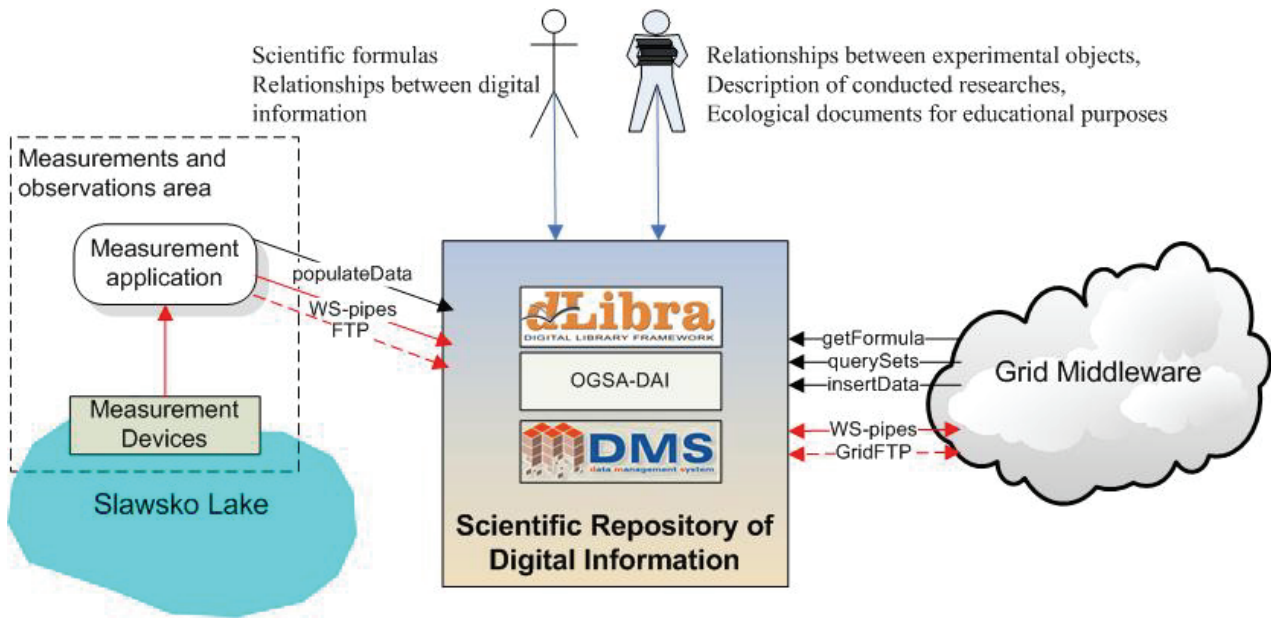
It is also worth mentioning that the Medical Digital Library concept offers some value-added services, utilizing the functionality of the Library. These services include realization of services built on top of the Registry of Medical Cases and construction of some advanced telemedical services. The former concerns two envisaged scenarios: medical scientific research performed on the collected cases by the specialist at Poznań University of Medical Sciences, and healthcare system management support scenarios, in which scenario the collected data may be analyzed to report the current trends in the medical domain in question, thus helping to better estimate needs of the regional health care. The latter concerns two advanced scenarios as follows:

- *Medical Data Archiving*, that is archiving of the image documentation of patient examinations conducted with the use of various types of medical modalities such as, for example, computed tomography, radiography, magnetic resonance imaging and others, and archiving of the digital recordings of events taking place at the hospitals such as, for example, operations performed in the operation rooms

- *Knowledge Driven Decision Support*, that is deep analysis of the medical data stored with the Medical Digital Library, aiming at discovering knowledge that could support clinical diagnosis. This scenario is connected with one of the other elements of the Medical Digital Library, that is with solutions enabling to process data on the computing resources connected to the grid.

C. Scientific Repository in Virtual Laboratory

Another environment where digital library architectures helped us to fulfill requirements imposed by the usage scenarios was the virtual laboratory for the Slawsko Lake. The notion of virtual laboratory indicates a distributed workgroup environment, with the main task of providing a remote access to the various kinds of rare and expensive scientific laboratory equipment and computational resources. The Virtual Laboratory research project [12] has been developed at Poznań Supercomputing and Networking Center (PSNC) since the beginning of the year 2002.



The Slawsko Lake virtual laboratory environment comes with specific requirements to deliver a unified storage of different types of scientific data (measurement data, experimental data, computation results) bound with digital library specific entries, e.g. publications, reports or measurement sheets.

The above-listed requirements of the Slawsko Lake virtual laboratory environment are fulfilled by the concept of the scientific repository of digital information which is based on mature and widely-used software developed in PSNC: the Gridge Data Management System, that was already mentioned in subsection V.2, and the dLibra digital library framework (see Fig. 6). Our main aim was to integrate these two products into a component-based solution, dedicated for storing, preserving and accessing various digital content, such as e.g. text, audio or video, enriched with metadata, and on the other hand being able to adapt to various kinds of application realms, business needs or service scenarios.

1) *Initial assumptions:* As an outcome of the requirements analysis for the Slawsko Lake, a set of primary measurement parameters (meteorological and hydrological) have been distinguished. This data, for example air pressure or maximal daily temperature, is gathered from various devices (e.g. a buoy), and collected by the measurement application. The raw data is transformed into XML documents and transferred to a Gridge Data Management System instance. GDMS is responsible for effective data storage and efficient access methods. It also offers a possibility to describe stored files with metadata which creates an abstract, semantic and explorable layer of resources. This

capability responds perfectly to the Slawsko Lake needs and makes it possible to define new formulas (e.g. free water surface evaporation), verify and modify existing ones, as well as define new relations between the stored data.

2) *Pilot installation:* The initial assumptions presented above do not embrace the main requirements requested for the digital repository for Slawsko Lake. These requirements include, among others, a possibility to model relationships between experimental objects, describe the conducted experiments and collect a broad scope of ecological documents for educational purposes. The scientific repository of digital information, which combines GDMS and dLibra systems, enhanced with selected OGSA-DAI [15] features, makes it possible to group the gathered data into collections, describe the resources with metadata on different levels of abstraction and associate experimental results with publications and reports.

We assume that the scientific repository stores the following three main types of scientific data, which together stand for digital information:

- *Publications and digital objects collections* – typically small files, usually containing only textual information, which may be indexed, searched and displayed using commonly used mechanisms;
- *Generic measurement data and experimental results* – files of different size (typically large ones), usually with binary content like pictures or DNA sequences. In general, it is hard to provide common mechanisms to process this data and usually it is required to deliver specific extensions to handle each file type;

- *Sequential measurement data* – as it was mentioned above, measurement results gathered within the virtual laboratory for Slawsko Lake (e.g. air pressure) are transformed from raw format into XML documents and transferred to Grid Data Management System. However, GDMS does not provide appropriate mechanisms for e.g. full-context search and XML documents manipulation. Thus standard GDMS functionality is enhanced with selected OGSA-DAI features in order to perform all basic operations on XML documents.

Apart from just simple store and retrieve possibilities, the digital repository for scientific information also allows its users to perform many other operations associated with the managed content. Examples of such operations can be: searching through metadata or textual content, subscribing news feeds, browsing through automatically created multi-level metadata indexes or creating persistent references to accessed objects.

3) *Summary*: The scientific repository of digital information integrates three software components – GDMS, dLibra and OGSA-DAI, linking them together in accordance with the service oriented architecture paradigm. The repository constitutes a novel solution dedicated to scientific domains, aiming at acquisition and management of various digital information and equipped with unique capabilities to bind various types of digital content. This novel approach fosters the idea of access, delivery and semantic integration of digital datasets in terms of metadata and ontologies, digital resources (storing and archiving), collections, content profiles, full-content and metadata-based, searching, annotations, assessments (Web 2.0) and collaborative work. Additionally, through utilization of atomic digital library services, it creates an opportunity for further integration of the virtual laboratory digital resources within digital library federations.

VI. CONCLUSIONS

In this paper we presented the results of our work on methodologies for federating digital libraries. On the basis of our long track of successful development of a digital library framework we were able to foster deployment of a significant number of regional and institutional digital libraries connected to the PIONIER national optical network. Further on, basing on open interfaces and thanks to the definition of atomic digital library services we were able to create an open national federation of digital libraries thus providing a powerful tool to academic and educational community. Our research proved that defining atomic digital library services helps to plan digital libraries appropri-

ately, while taking account of the large picture of digital library federating supports visibility and usability of the stored digital objects.

The two above-mentioned major concepts underpinning digital library design and deployment are important also when planning more complex environments supposed to manage and deliver digital information. We used these in such differing environments as a telemedical system and a virtual laboratory. In the former case, a Medical Digital Library has been designed to support telemedical consultations between hospitals and to provide other advanced telemedical services. In the latter case, the capabilities of a classical digital library software were combined with those of grid data management system to create a robust scientific repository of digital information. Both these advanced and complex digital libraries utilize atomic digital library services and thus are capable of efficient storage and provision of content, fulfilling the requirements of the end-users. They both are envisaged to collect interesting scientific and educational content, inclusion of which into the mainstream digital library systems (or federations) would help to distribute the information to the potential users. This is again possible thanks to basing the architectures on atomic digital library services which, with addition of open protocols for communication, creates an opportunity to federate the collected resources with other digital libraries.

The concept of atomic digital library services used in the dLibra framework is an example of SOA applied to digital library systems. Other examples of such approach can be found in [19] and [22]. The problem of searching in distributed resources is also very widely recognized by the digital library community. It was for example described in [20] and [21]. A solution to this problem based on harvesting the metadata from distributed digital libraries, similar to ours, was proposed by F. Simeoni [23].

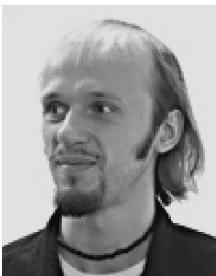
Acknowledgements

Part of the work presented herein has been conducted under the research project no. 3 T11C 02330: Mechanisms of atomic services for distributed digital libraries financed by Polish Ministry of Science and Higher Education.

References

- [1] J. Rychlewski, J. Węglarz, S. Starzak, M. Stroiński and M. Nakonieczny, *PIONIER: Polish Optical Internet*, in: Proceedings of ISThms 2000 Research and Development for the Information Society conference, Poznań, p. 19-28 (2000).

- [2] C. Mazurek, M. Stroiński and S. Szuber, *Digital Library for Multimedia Content Management*, in: Proceedings of 2nd International Workshop of the DELOS Network of Excellence on Digital Libraries, DELOS, p. 29- 34 (1999).
- [3] M. Kosiedowski, C. Mazurek and M. Werla, *Digital Library Grid Scenarios*, in: Knowledge-Based Media Analysis for Self-Adaptive and Agile Multi-Media, Proceedings of the European Workshop for the Integration of Knowledge, Semantics and Digital Media Technology, EWIMT 2004, London, p. 189-196 (2004).
- [4] C. Lagoze and H. Van de Sompel, *The Open Archives Initiative Protocol for Metadata Harvesting*, accessed from <http://www.openarchives.org/OAI/openarchivesprotocol.html>, Open Archives Initiative (2004).
- [5] C. Mazurek and M. Werla, *Distributed Services Architecture in dLibra Digital Library Framework*, in: Proceedings of 8th International Workshop of the DELOS Network of Excellence on Digital Libraries on Future Digital Library Management Systems, Schloss Dagstuhl (2005).
- [6] C. Mazurek and M. Werla, *Digital Object Lifecycle in dLibra Digital Library Framework*, in: Proceedings of the 9th International Workshop of the DELOS Network of Excellence on Digital Libraries on Interoperability and Common Service, Heraklion (2005).
- [7] C. Mazurek, M. Stroiński, M. Werla and J. Węglarz, *Metadata harvesting in regional digital libraries in PIONIER Network*, Campus-Wide Information Systems **23**(4), 241-254 (2006).
- [8] M. Kosiedowski, C. Mazurek, M. Stroiński and J. Węglarz, *Gridsupported Medical Digital Library*, Studies in Health Technology and Informatics **126**, 127-136 (2007).
- [9] J. Błaszczyszki, M. Kosiedowski, C. Mazurek, R. Słowiński, K. Słowiński, M. Stroiński and Sz. Wilk, *Telemedical Portal "Telemedycyna Wielkopolska"*, in: E. Piętka, J. Leski and S. Franiel (Eds.) MIT 2006 Medical Informatics and Technology, Wisla, p. 230-235 (2006).
- [10] J. Pukacki, M. Kosiedowski, R. Mikołajczak, M. Adamski, P. Grabowski, M. Jankowski, M. Kupeczyk, C. Mazurek, N. Meyer, J. Nabrzyski, T. Piontek, M. Russell, M. Stroiński and M. Wolski, *Programming Grid Applications with Gridge*, Computational Methods in Science and Technology **12**, 47-48 (2006).
- [11] M. Kosiedowski, C. Mazurek and M. Stroiński, *PROGRESS – Access Environment to Computational Services Performed by Cluster of Sun Systems*, in: Proceedings of 2nd Cracow Grid Workshop, Kraków, p. 45- 56 (2002).
- [12] M. Lawenda, N. Meyer, T. Rajtar, M. Okoń, D. Stokłosa, M. Stroiński, Ł. Popenda, Z. Gdaniec and R. W. Adamiak, *General Conception of the Virtual Laboratory*, Lecture Notes in Computer Science **3038**, 1013-1016 (2004).
- [13] M. Bogdański, M. Kosiedowski, C. Mazurek and M. Wolniewicz, *GRID SERVICE PROVIDER: How to improve flexibility of grid user interfaces?*, Lecture Notes in Computer Science **2657**, 255-263 (2003).
- [14] ACGT – Advancing Clinico Genomic Trials, <http://eu-acgt.org/>.
- [15] The OGSADAI Project, <http://www.ogsadai.org.uk/>.
- [16] C. Mazurek, T. Parkola and M. Werla, *Atomowe usługi w środowisku rozproszonych bibliotek cyfrowych*, in: Proceedings of II Krajowa Konferencja Naukowa „Technologie przetwarzania danych”, 24-26.09.2007, Poznań. p. 458-467, 2007. ISBN 978-83-7143-349-8. Wydawnictwo Politechniki Poznańskiej.
- [17] A. Dudczak, M. Heliński, C. Mazurek, T. Parkola and M. Werla, *Analiza funkcjonalności wybranych modeli i systemów zarządzania bibliotekami cyfrowymi*, in: Zeszyty Naukowe Wydziału ETI Politechniki Gdańskiej, Seria „Technologie Informacyjne”, **13**, 407-416 (2007). ISBN 978- 83-60779-01-9.
- [18] M. Heliński, C. Mazurek and M. Werla, *Distributed Digital Library Architecture for Business Solutions*, in: Proceeding of 11th Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications – EUROMEDIA'2005, 11-13.04.2005, Toulouse, France. p. 5-10 (2005). ISBN 90-77381-171-1.
- [19] D. Bainbridge, K. Don, G. Buchanan, I. Witten, S. Jones, M. Jones and M. Barr, *Dynamic Digital Library Construction and Configuration*, Lecture Notes in Computer Science **3232**, 1-13 (2004).
- [20] M. Bender, S. Michel, G. Weikum and Ch. Zimmer, *Challenges of Distributed Search Across Digital Libraries*, in: Proceedings of 8th International Workshop of the DELOS Network of Excellence on Digital Libraries on Future Digital Library Management Systems, Schloss Dagstuhl, 2005.
- [21] R. Larson, *Distributed IR for Digital Libraries*, Lecture Notes in Computer Science **2769**, 487-498 (2003).
- [22] G. Pyrounakis, K. Saidis, M. Nikolaidou and I. Lourdi, *Designing an Integrated Digital Library Framework to Support Multiple Heterogeneous Collections*, Lecture Notes in Computer Science **3232**, 26-37 (2004).
- [23] F. Simeoni, *Servicing the Federation: The Case for Metadata Harvesting*, Lecture Notes in Computer Science **3232**, 389-399 (2004).



MICHAŁ KOSIEDOWSKI has been employed by Poznań Supercomputing and Networking Center in the Network Services Department since 1997. His professional interests include service oriented systems, grid systems and web technologies. Michał has been the technical coordinator of the PROGRESS project, which developed access environment to grid services. Currently, he is the project leader of the Telemedycyna Wielkopolska project which develops a telemedical system for the support of the healthcare in the region of Wielkopolska. He is the author or co-author of over 15 publications in professional journals and conference proceedings.



CEZARY MAZUREK is the Head of the Network Services Department at Poznań Supercomputing and Networking Center. He received his PhD in Computer Science from Poznań University of Technology in 2004. His research interests include a wide variety of advanced network services including portal solutions, digital multimedia libraries, streaming technologies, distance learning and access to grid services. He has been conducting numerous research projects in those fields including the Multimedia City Guide, Polish Educational Portal, Digital Library Framework: dLibra, Nabor 2003-2007, PROGRESS and interactive TV platform. He is the author or co-author of over 30 papers in professional journals and conference proceedings.



MACIEJ STROIŃSKI received the Ph. D. degree in Computer Science from the Technical University of Gdańsk in 1987. Currently he is the Technical Director of Poznań Supercomputing and Networking Center. He is also a lecturer in the Institute of Computing Science of the Poznań University of Technology. His research interests concern computer network protocols and management. He is the author or co-author of over 100 papers in major professional journals and conference proceedings.



MARCIN WERLA started working as a software developer in Poznań Supercomputing and Networking Center in November 2002. After he received his master's degree from Poznań University of Technology in 2004 he changed his position at the Center and is now a computing systems analyst. He is the technical leader of the dLibra project and the PIONIER Network Digital Library Federation portal. He is also the author or co-author of several papers in conference proceedings.



MARCIN WOLSKI started working at PSNC in 2001 as a database and network programmer. In January 2002, he joined the Data Management System (DMS) team, where he worked as a system analyst and developer. In 2003, he began to lead the DMS and modified its development direction towards grid environments and large-scale solutions. Consequently, he analyzed the Data Management System software utilization in the context of virtual laboratories. Other research interests of Marcin's include GIS systems, system oriented architectures, autonomic computing and enterprise application frameworks. He is the author or co-author of several papers in professional journals and conference proceedings.