

Digital Object Lifecycle in dLibra Digital Library Framework

Cezary Mazurek and Marcin Werla

Poznan Supercomputing and Networking Center
Poznan, Poland
{mazurek,mwerla}@man.poznan.pl

Abstract. In this article we describe support for a digital objects lifecycle, which we have created during our digital library related works. Basing on our experiences from WBC digital library, we intend to show how a properly constructed lifecycle can extend the functionality of digital libraries for both library readers and librarians responsible for maintaining content gathered in a digital library.

Introduction

dLibra [1, 2] is the first Polish Digital Library Framework. It has been developed in Poznan Supercomputing and Networking Center (PSNC) since 1999. It is used as a software platform in one of the largest Polish digital libraries – in the Digital Library of the Wielkopolska Region (WBC) [3], which now consists of more than 4000 various publications grouped into four thematic collections: cultural heritage, regional materials, educational materials and music notes. dLibra was also deployed for the Digital Library of Wroclaw University of Technology (BCPWr) [4]. Startup of another seven dLibra-based academic digital libraries is planned for the nearest future.

The dLibra digital object lifecycle is a result of our experiences from the cooperation with multiple libraries associated in Poznan Foundation of Scientific Libraries (PFSL). It is a consortium organized to create and maintain the WBC regional digital library. All of the associated libraries digitalize the most valuable books from their collections and publish them in the WBC library. In this article we shortly introduce the dLibra framework and then we describe the dLibra digital object lifecycle.

dLibra Digital Library Framework

dLibra is a distributed and portable digital library framework. It is created to be a highly configurable software basis for digital libraries. dLibra has a typical multitier architecture. In the center of this architecture there is a dLibra server separated into six services: MetadataServer, ContentServer, UserServer, SearchServer,

DistributedSearchServer and EventServer. Additionally there is the SystemServices server, which allows inter-service communication.

Access to a digital library based on the dLibra software can be performed by two client applications. The first one is the Editor's/Administrator's Application. It allows users to add new objects to the library and manage all the gathered content. It also gives administrators a possibility to manage some core DL parameters like used metadata scheme, metadata dictionaries etc. The second client application is the Reader's Application, based on WWW pages. It gives users read-only access to the library with such functionality as browsing, searching, news bulletins etc. More detailed information about the dLibra architecture can be found in [5].

The main element in the dLibra digital repository structure is the directory. The dLibra directory can be compared to a directory in a file system. It can contain other directories and publications. Information about all DL directories and publications together with their metadata are managed by the MetadataServer. The publication content is managed by the ContentServer. These two services together are responsible for the support of a digital object lifecycle implemented in dLibra.

Digital Object Lifecycle

In the described lifecycle model a digital object can be in four states (see Fig. 1). The first state corresponds to a digital object that is planned to be added to a digital library. This can be useful in a situation when the creation of a digital object is a time-consuming process, but the creator of this object wants to notify other DL users about it. In case of the WBC digital library it is used when the number of libraries cooperates and each library adds publications which it is going to digitalize in the nearest future, as planned publications. Other institutions can check the list of planned publications and avoid a situation in which, for example, two libraries work on the digitalization of the same old manuscript. The planned object can be described with the metadata or assigned to a collection exactly the same as any other digital object in the dLibra-based digital library.

When the digital content is added to an object in the planned state, this object is converted to an unpublished digital object. Such object has both content and metadata. Objects in this state are available for all library editors and administrators, but cannot be accessed by library readers (both anonymous and registered). To access unpublished objects, the user must have special privileges.

The library editor, after preparing the digital object for its readers, can mark this object as published. The published object is a basic type of object visible in the digital library. In the dLibra-based libraries such objects can be grouped within hierarchic structures called "group publications" and can be assigned to several collections.

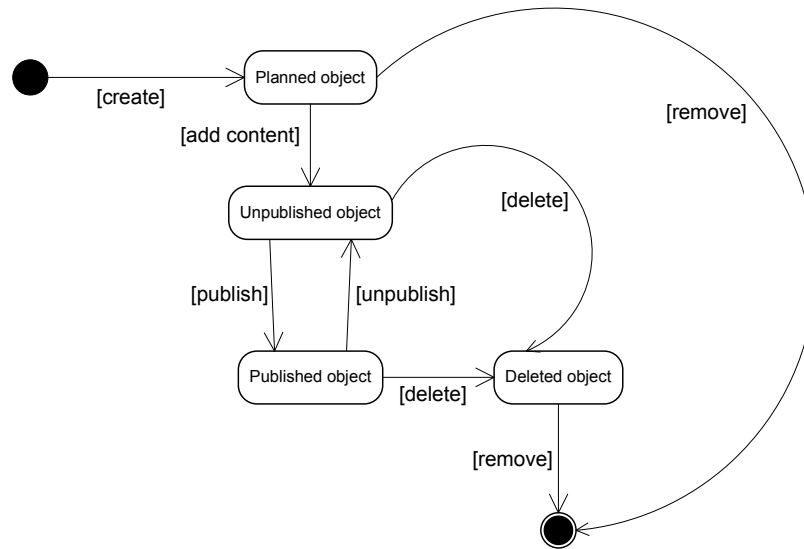


Fig. 1. dLibra digital object lifecycle statechart.

Additionally each object can have multiple editions. Digital object editions can be compared to editions of a book. All digital object editions can be available to the digital library users at the same time. It is achieved by the content versioning system built in the dLibra Framework. This mechanism gives a possibility to show how the content of the digital object evolved during its life. It is especially valuable when the digital library is used not only as a repository for digitalized writing relics but, for example, as a documents repository in SME, where the documents content can often change. We assumed that each digital object must consist of a number of files, and one of those files is the main publication file. It is a file from which users start reading the publication. Each of the publication files can have multiple versions. The edition of a digital object can be seen as a set of versions of those object files – one chosen version of each file. Each edition can be published or not – just as described above for a basic digital object with only one edition.

The last state in which the digital object can exist is a deleted state. Digital libraries, in general, are systems focused on long term preservation of digital assets, therefore the deleted state is perhaps more common for CMS-like repositories. However, we claim that there are two basic situations in which the deletion of an object from a digital library should be supported. The first one is the most common. It is a situation in which there is an error in the object’s content, or the object was added by mistake or only for test purposes and should be entirely removed from the system. In this situation all object content and metadata is removed and the only thing that stays in the system is the information that the given object (seen as persistent and unique id) existed and was deleted on a certain date for a certain reason.

There is also a possibility that object content must be removed from the library because of some legal obligations. A good example of that can be expiration of the license agreement signed between the digital library and the author of a digital

publication. In such cases the object also comes into the deleted state, and its entire content is removed. But, in contraposition to the standard deletion, all object metadata and information about all editions of the object's content stays in the system and becomes read-only. Additionally, information about the cause of the deletion is added.

In both described deleted states, when an external system connects to the digital library, it can obtain data about the deleted publications. Such mechanism is used in dLibra-based digital libraries to improve distributed search implemented with the OAI-PMH protocol [5]. Moreover, when the reader follows, for example, a bibliographic reference and it is a reference to a deleted publication, instead of a typical message like HTTP 404 "Not Found" error, the dLibra system displays information about the deleted digital object with detailed metadata and the deletion cause and date. This is especially useful when the metadata or the deletion cause contains information about the author or the owner of the digital object. Having such information, the digital library reader can try to obtain the searched digital object from its legal owner.

Conclusion

The described digital objects lifecycle is a result of long term cooperation between PSNC and many Polish academic, scientific and public libraries in the field of DL systems. It is designed to extend digital libraries functionality, increase its reliability and preserve the maximum amount of information about gathered digital content. Support for this lifecycle has been entirely implemented in the dLibra Digital Library Framework and is used in all dLibra installations. Our future works will be partially focused on increasing possibilities of a reader – digital library interactions with features like annotations or more sophisticated semantic search possibilities. We hope that next dLibra deployments will allow our digital objects lifecycle to evolve into an even more sophisticated and useful mechanism.

References

- [1] Gruszczyński, P.; Mazurek, C.; Osinski S.; Swedrzynski A.; Szuber S. "dLibra Content Maintenance for Digital Libraries" in *Euromedia'2002*, pages 28–32, 7th Annual Scientific Conference, April 2002.
- [2] Mazurek C.; Stroiński M.; Swędrzyński A.; „dLibra – Integrated Framework for Publishers and Libraries” – poster at 7th European Conference on Digital Libraries, Torndheim, Norway, August 2003.
- [3] Digital Library of Wielkopolska Region. <http://www.wbc.poznan.pl/>.
- [4] Wroclaw University of Technology Digital Library. <http://dlib.bg.pwr.wroc.pl/>.
- [5] Mazurek, C., Werla, M. – “Distributed Services Architecture in dLibra Digital Library Framework”. 8th International Workshop of the DELOS Network of Excellence on Digital Libraries on Future Digital Library Management Systems, 29.03-01.04.2005, Schloss Dagstuhl, Germany. Workshop Proceedings.

- [6] Lagoze, C.; Van de Sompel, H. – “The Open Archives Initiative: Building a low-barrier interoperability framework”, pages 54-62, Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke, VA, USA, June 2001.