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

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 Wielkopolska Region (WBC) [3], which now consists of more than 3300 various publications grouped into four thematic collections: cultural heritage, regional materials, educational materials and music notes. dLibra was also deployed for Wroclaw University of Technology Digital Library (BCPWr) [4]. Startup of seven another dLibra based academic digital libraries is planned for the nearest future.

In our article we want to describe support for digital objects lifecycle, which we have created during our digital library related works. We want to show how properly constructed lifecycle can extend functionality of digital libraries and how it can be used to improve distributed communication mechanisms. Below we briefly point up all stages of dLibra digital object lifecycle.

In described lifecycle model digital object can be in four states. First state corresponds to digital object that is planned to be added to digital library. This state is very useful when number of institutions cooperates with one digital library. Each institution can add objects, which it is going to add in the nearest future, as planned objects. Other institutions can check list of planned objects and avoid situation in which for example two libraries works on digitalization of the same old manuscript. Planned object can be described with metadata or assigned to collection exactly the same as any other digital object in dLibra based digital library.

When digital content is added to object in planned state, it is converted to unpublished digital object.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 To access unpublished objects, user must have special privileges.

Library editor, after preparing digital object for its readers, can publish such object. Published object is a basic type of object visible in digital library. In dLibra based libraries such objects can be grouped within structures called "group publications" and can be assigned to several collections.

Additionally each object can have multiple editions. Digital object editions can be compared to editions of a book. This mechanism gives possibility to show, how content of digital object evolved during its life. It is especially valuable when digital

library is used not only as a repository for digitalized writing relicts, but for example as a publishing platform in SME. We assumed that each digital object must consist of a number of files. Each of those files can have multiple versions. Edition of a digital object can be seen as set of versions of that object files – one chosen version of each file. Each edition can be published or not – just as described above for basic digital object with only one edition.

Last state, in which digital object can be, is a deleted state. There are two basic situations in which object can be deleted. First of them is most common one. It is a situation, in which there is an error in object content or object was added only for test purposes and it should be entirely removed from the system. In this situation all object content and metadata is deleted and the only thing that stays in the system is information that given object existed and was deleted on certain date.

There is also a possibility, that object content must be removed from library because of some legal obligations. In such cases object also comes into deleted state, and its entire content is removed. But, in contraposition to standard deletion, all object metadata, information about all editions and object files stays in the system and becomes read-only. Additionally information about deletion cause is added.

In both described deleted states, when an external system connects to digital library, it can obtain data about deleted publications. Such mechanism is used in dLibra based digital libraries to improve distributed search implemented with OAI-PMH protocol [5]. Moreover, when reader follows for example bibliographic reference and it is a reference to deleted publication, instead of typical HTTP 404 "Not Found" error message, dLibra system displays information about deleted digital object.

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 distribution possibilities and reliability and to preserve maximum amount of information about gathered digital content. Support for this lifecycle has been entirely implemented in dLibra Digital Library Framework and is used in all dLibra installations. We hope that next dLibra deployments will allow our digital objects lifecycle to evolve into 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, 7*th* 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] 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.