

# OCLC Registry of Digital Masters – Opportunities for European Cooperation

by JANET LEES

## INTRODUCTION

I would like to thank the Preservation Division for the invitation to participate in the programme today. I would also like to thank the LIBER Board for their interest in the early stages of this project culminating in the announcement LIBER President Erland Kolding Nielsen made last night about [LIBER](#) support and collaboration with [OCLC PICA](#) with respect to the Digital Registry.

This initiative reflects my new role within OCLC PICA, which is to identify opportunities for joint development with our major shareholder OCLC, and sometimes 3<sup>rd</sup> parties such as LIBER.

My main purpose today is to provide a descriptive paper to inform LIBER members about the OCLC Digital Registry and to outline the possible role for LIBER and its membership in a model for European participation.

My presentation will be in three phases: the need for a registry, a description of the Registry and finally, and perhaps most importantly, to outline opportunities for LIBER and its members in establishing a model for European contribution.

We are all familiar with the concept of a registry through our participation in union catalogues and perhaps also from our experience with microfilms through the European Register of Microform Masters ([EROMM](#)) hosted by our friends in Göttingen.

## THE NEED FOR A REGISTRY

Today we are all engaged somewhere along the road of digitisation. We are converting some of our p-collections to e-collections and we are licensing or purchasing born digital materials. The totally digital collection is still a long way in the distance and we are all conscious that funding for this journey is limited. We do not want to waste available resources duplicating the efforts of others, but we want to be assured that the standards used in digitisation by others meet our standards and access to the resulting materials is guaranteed as much as they can be. Finally as in any new process we also have a professional need to share our experiences as we move up the learning curve so that we can develop and follow “best practice” as these activities become part of the library workflow.

In short we want “more access at less cost” which is to say that we are willing to *collaboratively* build a larger mass of digitised material than we could ever hope to achieve *individually*. We are also conscious that this effort like many other activities we embark on these days has a global dimension as scholars and serious reference users become used to the idea that the resources they need can be obtained from way beyond their local area.

As with many digital initiatives the Digital Library Federation ([DLF](#)) was the catalyst for the registry. DLF identified a Registry as being a key component of the infrastructure that would be required by Digital libraries. In 2001 DLF brought together an initial working group comprising representatives from CLIR, Columbia, Harvard, Michigan and OCLC to begin to define the functional requirements for a Registry and to set up more specific working groups that would look at the definition of the bibliographic and digitisation standards that would be adopted.

OCLC was asked to work with the initial group to develop a prototype that would embed registry records in OCLC WorldCat. This initiative took place at the point where OCLC was at the beginning of its migration to its new technological platform that would allow the Registry to be both a part WorldCat and also exist as a defined and OAI harvestable subset.

DLF also decided that the creation of the Registry should act as an encouragement for libraries to adopt high or preservation standards in their digitisation projects. For this reason the working group decided that the Registry records should contain information about the standards followed by the digitisation agency.

Initially the Registry was envisaged as containing descriptions of reproductions of print based materials that are digitised for preservation or to extend access to the original materials. This was later extended to include born digital materials including serials.

To summarise this part of the presentation I can do no better than to quote the summary of the Registry Purpose as declared by the DLF. "By recording materials in the Registry, institutions are signalling the intent to preserve and maintain the accessibility of the described materials over an extended timeframe (decades or centuries, not years). This implies the materials are digitised carefully, complying with established standards and best practices and are stored in professionally managed systems" (Registry, 2004).

## DESCRIPTION OF THE REGISTRY AND ITS FUNCTIONALITY

There are essentially three components: metadata standards, digitisation standards and access functionality standards. Links with other registries is anticipated and I will say more about that in the final section.

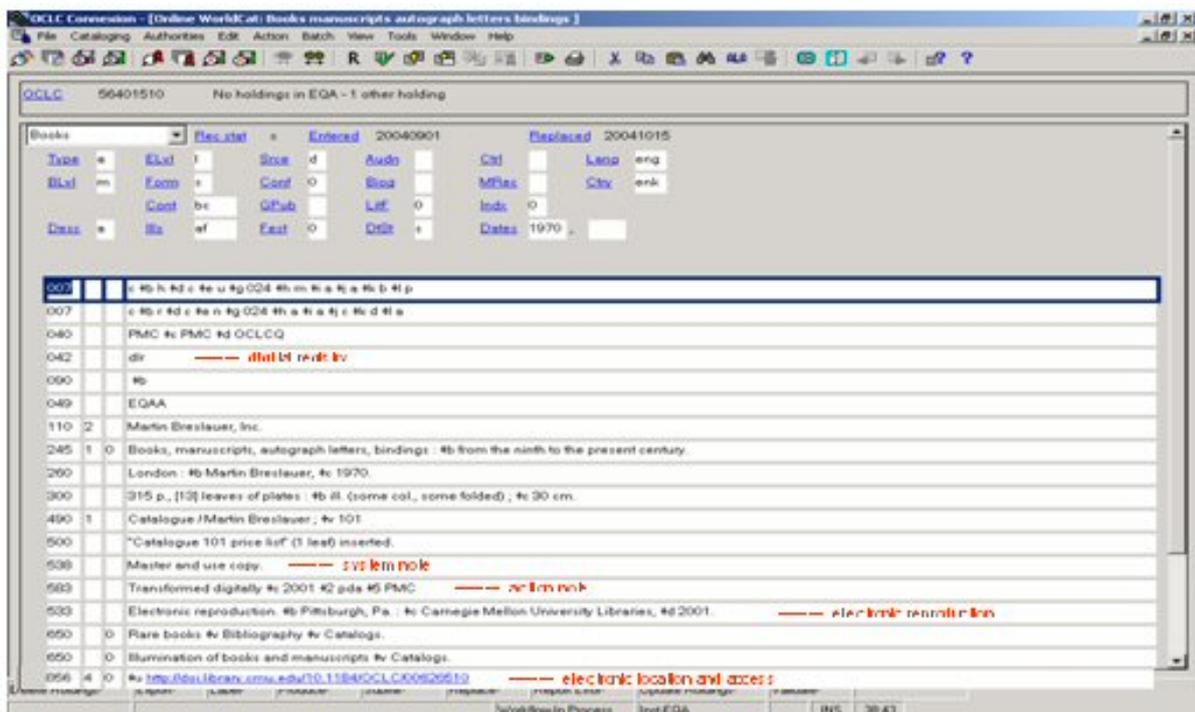
Let's start with the metadata. The records are based on existing bibliographic standards with additional fields for digital actions and access. The Registry is hosted by OCLC and the format used by OCLC is MARC21 but this does not preclude conversion of records from other standards eg UniMARC or Dublin Core. In terms of the prototype records can be entered through the OCLC online cataloguing interface, by ftp or via OCLC batch loading - allowing items to be entered one by one or as large collections. There is no requirement for OCLC membership for submitting institutions - the Registry is open to all.

Persistent links - URL or URN - are used to link to the use copy and also to descriptions of the terms and conditions of access. Equally URLs or other persistent identifiers are used to find information about the master archive copy to describe the technical standards used in creating the archival master and the repository practices being followed in its storage and maintenance.

The statement of intent to digitise is an encouragement for libraries to input records to the registry as soon as a definite decision to digitise has been made or funding obtained to help in reducing unnecessary duplication of effort. A projected date is included so that libraries can be reminded to update the information in case of any delays or changes to projects. URLs are also used to link the user from the metadata record to the item.

Here is a sample registry record created by Carnegie Mellon University with the registry fields highlighted:

Picture 1



042 – Authentication code: indicates that this record has been contributed to the registry

538 – System details: indicates master and use copies. Has the ability for a subfield for an URL

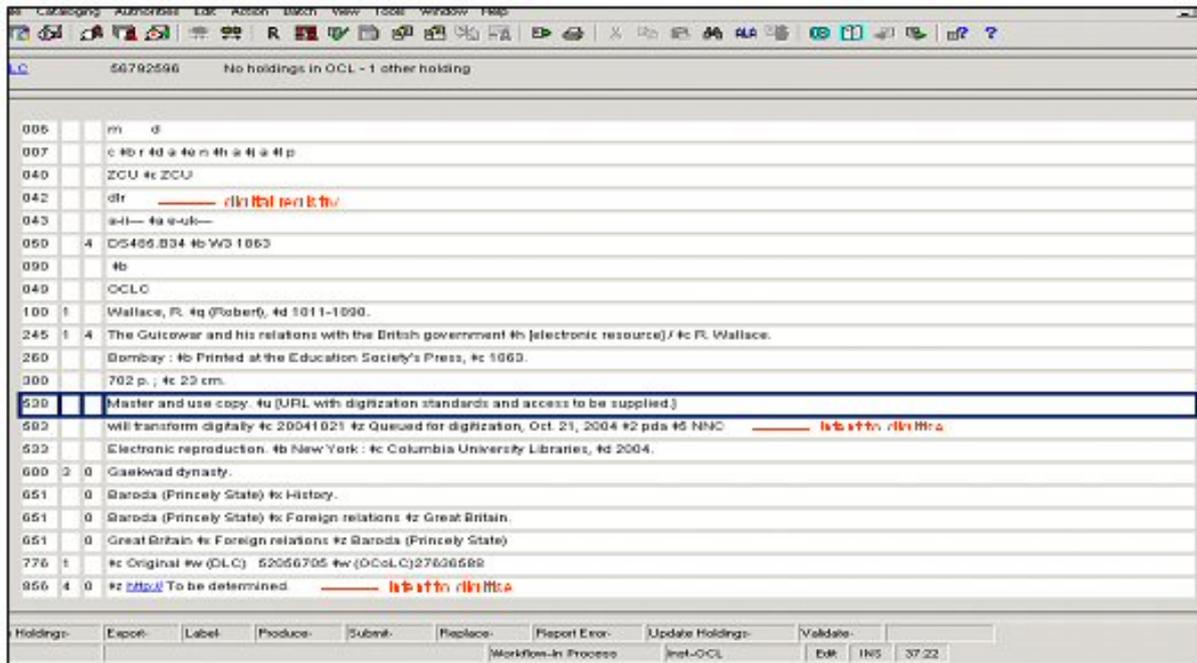
583 – Action note: indicates institution who and when the item was digitised

533 – Electronic reproduction note

856 – Electronic location and access information

This sample registry record is an example of the intent to digitise

Picture 2



583 field contains “will transform digitally” and gives the date when the item was queued for digitisation

856 indicates the URL is “to be determined”

In both cases I have used these fields for illustration purposes. A detailed document [Registry of Digital Masters Record Creation Guidelines](#) developed by the DLF Registry Working Group is available from the DLF web site (Registry, 2004).

The working group weighed up the pros and cons of establishing specific or minimum standards for digitisation for contributed items. On the one hand they wanted registered items to be representative of high quality digitisation but on the other hand did not want to reduce the usefulness of the registry by excluding large collections that might be valuable additions.

In the end the group decided to work on a “minimum benchmark” of what they called a “faithful reproduction” based on a desire to build consensus in what was and is still understood to be an ever changing technological environment in which standards would be continuously improved (Benchmark, 2002).

Given that at the research library level communication between a relatively small number of practitioners in libraries and other organisations is fairly well established this should not be an issue and should go some way in increasing user confidence that the quality will support the persistence of access and interoperability between collections - however the scalability of such an approach across a large cross section of institutions could be questionable.

I can only stress that at this stage the DLF has prepared and endorsed these benchmarks as minimum standards for digitisation, not for the registry per se. However in terms of meeting the objective of the registry to reduce duplication of effort the creation of baseline standards should help.

*DLF Minimum Benchmarks*

Printed Texts	Illustrated Texts (black & white)	Illustrated Texts (colour)	Rare and Early Printed Books
600 dpi 1-bit, TIFF	400 dpi 8-bit, TIFF	400 dpi 24-bit, TIFF	400 dpi 9- or 24-bit, TIFF

The minimum benchmark I have shown here is for digital page images and is expected to be the first in a series of benchmarks for other types of digital reproductions. This is very much a work in progress and I feel it is important that we have key European library input to the debate as soon as possible.

The objectives for access and discovery is to allow users - librarians and end users - to be able to globally identify where digitised masters and use copies are located and under what terms and conditions they may be accessed. In cases where multiple digitised copies exist the Registry should be able to provide the user with comparable information about all existing copies. The use of persistent identifiers to achieve this has already been mentioned but should be underlined.

The DLF benchmark recommends that digitised materials should be “faithful” reproductions - preservation context - retaining the same sequential navigation as the original including the component parts such as title page, table of contents, illustrations and index. This requires the digitisation agency to include blank pages where they exist in the original and to provide placeholders for any data that is missing in the original for whatever reason.

Let’s look at discovery of items from the prototype. For the purposes of the initial phase of the Digital Registry project the records are displayed in the OCLC FirstSearch interface. As they are part of WorldCat they can also be searched via the OCLC PICA end user service PiCarta. Ultimately the DLR records will also be available as a distinct database and available from the DLF and potentially other hosts. Records from organisations participating in OCLC Open WorldCat will also have their records visible through the Open WorldCat Search engine partners - currently Google, Yahoo and Ask Jeeves.

Here is the initial set of records ... at present there are not many. Three institutions - Cornell, Harvard and Library of Congress - are preparing the first live files for inclusion and we expect these some several thousand records to be ready to seed the Registry very shortly. The following sequence of screen displays shows the initial display set, an individual record, transfer through to an OPAC and navigation to the item itself.

Picture 3

**FirstSearch** Test Account **OCLC** **PICA**

WorldCat Detailed Record | My QuestionPoint | Ask a Librarian | Library Resources | Google

Click on a checkbox to mark a record to be e-mailed or printed in Marked Records.

Home Databases Searching Results Resource Sharing Staff View My Account Options Comments Exit Hide tips

List of Records Detailed Record Marked Records Go to page

WorldCat results for: ac= "dlr" and ((dt= "bks") or (dt= "url")). Record 1 of 4.

Subjects Libraries E-mail Print Export Help

WorldCat

WorldCat results for: ac= "dlr" and ((dt= "bks") or (dt= "url")). Record 1 of 4.

Prev 1 Next Mark:

**Books, manuscripts, autograph letters, bindings :**  
from the ninth to the present century.

Martin Breslauer, Inc.

1970

**English** Internet Resource 315 p., [13] leaves of plates : ill. (some col., some folded) ; 30 cm.  
London : Martin Breslauer,

**GET THIS ITEM**

Access: <http://doi.library.cmu.edu/10.1184/OCLC/00626510>

Availability: **Check the catalogs in your library.**

- Libraries worldwide that own item: 1
- Connect to the catalog at University College Cork

External Resources:

- Click here to check availability from other services
- BIBSYS Extended Services
- Find this item at Barnes & Noble.com (Bookseller)

**FIND RELATED**

More Like This: [Advanced options...](#)

Find Items About: [Martin Breslauer, Inc.](#) (10)

**Title:** **Books, manuscripts, autograph letters, bindings :**  
**from the ninth to the present century.**

Corp Author(s): [Martin Breslauer, Inc.](#)

Publication: London ; Martin Breslauer,  
Year: 1970

Description: 315 p., [13] leaves of plates : ill. (some col., some folded) ; 30 cm.

Language: English

Series: Catalogue / Martin Breslauer ; 101; **Variation:** Catalogue (Martin Breslauer, Inc.) ; 101.

Access: <http://doi.library.cmu.edu/10.1184/OCLC/00626510>

**SUBJECT(S)**

Descriptor: [Rare books -- Bibliography -- Catalogs.](#)

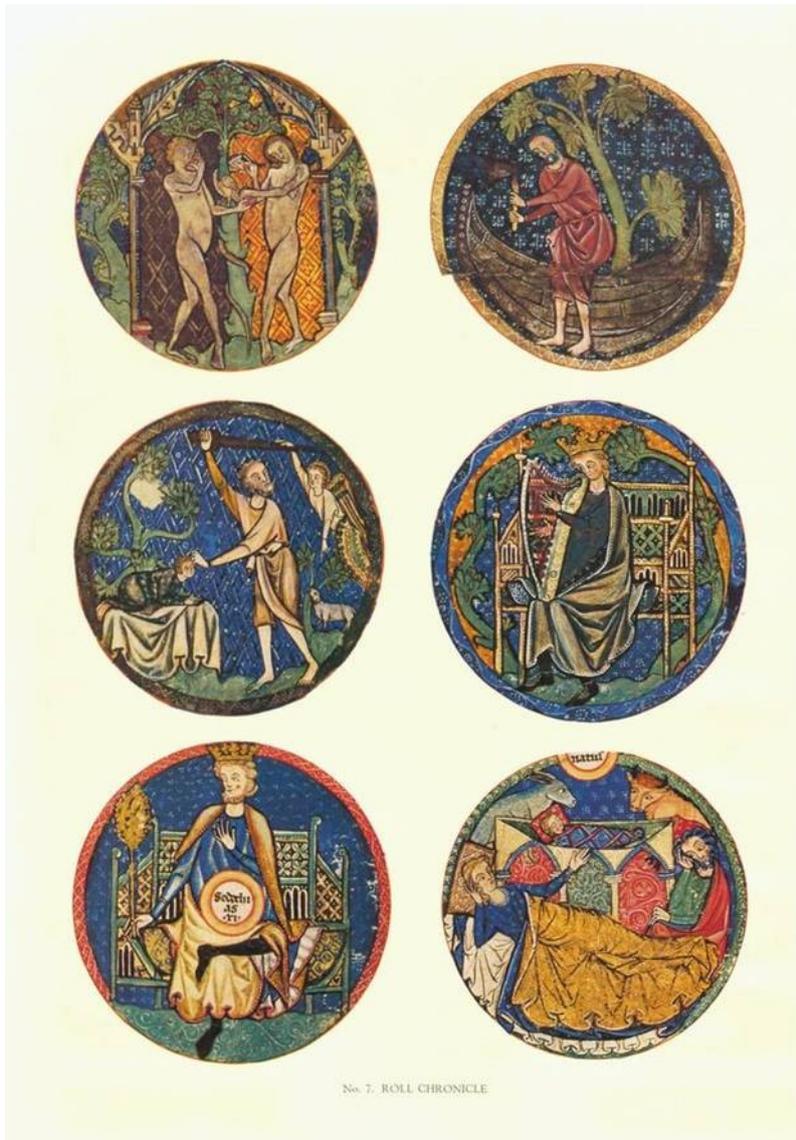
Picture 4

The screenshot shows the FirstSearch WorldCat interface. At the top, there's a navigation bar with 'Home', 'Databases', 'Searching', 'Results', and 'Resource Sharing'. Below that, there are tabs for 'List of Records', 'Detailed Record', and 'Marked Records'. The main content area displays the title 'Books, manuscripts, autograph letters, bindings : from the ninth to the present century.' by Martin Breslauer, Inc., published in 1970. It includes an 'Access' link, 'Availability' information, and 'External Resources'. A 'GET THIS ITEM' section is also present. The bottom part of the page shows 'Find Items About' and 'FIND RELATED' sections.

Picture 5

The screenshot shows the Carnegie Mellon University Libraries Library Catalog page. It features a book cover image on the left and the title '016.09 B84B 1970 "Books, manuscripts, autograph letters, bindings : from the ninth to the present century."' on the right. Below the title, there are links for 'Full Text' and 'Library Catalog'. The page also includes a 'Search' and 'Browse' section, and a footer with navigation links like 'Search', 'Browse', 'About Collection', 'Home', 'Help', and 'Note'.

Picture 6



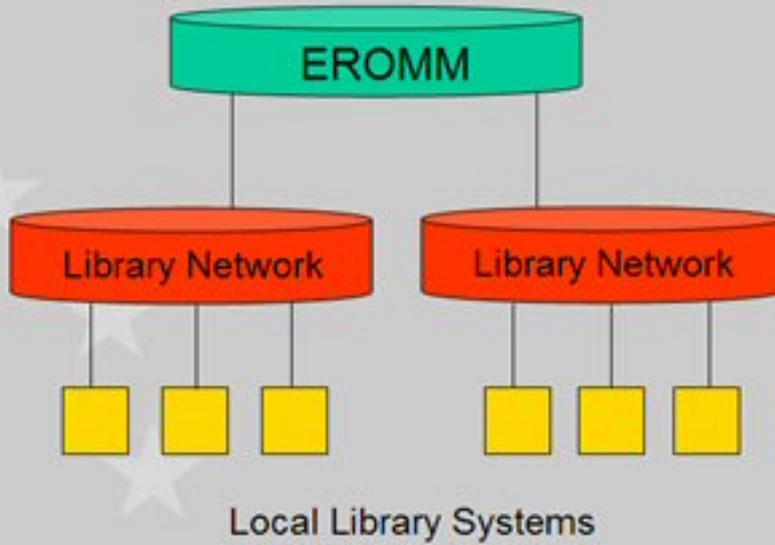
## A MODEL FOR EUROPEAN CONTRIBUTION

Now that you have an understanding of the content and functionality of the Registry, let's turn our thoughts to a model for European participation. For whilst the foundations for the Registry are US orientated it makes no sense for the Registry to be anything but global in scope. Multiple registries would add confusion and dilute the possibility of achieving the generic registry goals. The omission of the vast cultural riches of European library collections dilutes or conversely a European Registry without US participation would not meet user expectations for comprehensiveness.

With these thoughts in mind and with the [Future of LIBER](#) document signalling a possible role for LIBER in coordinating this activity OCLC PICA met last year with representatives from the LIBER Board who were very positive in their support for the Registry concept (Future, 2004). The initiative was discussed at the January LIBER Executive Board meeting in Uppsala which set up a working group chaired by Paul Ayris which was tasked with recommending possible models for collaboration. This group met with OCLC PICA again in the spring and discussed a possible working model for European cooperation in a global Registry.

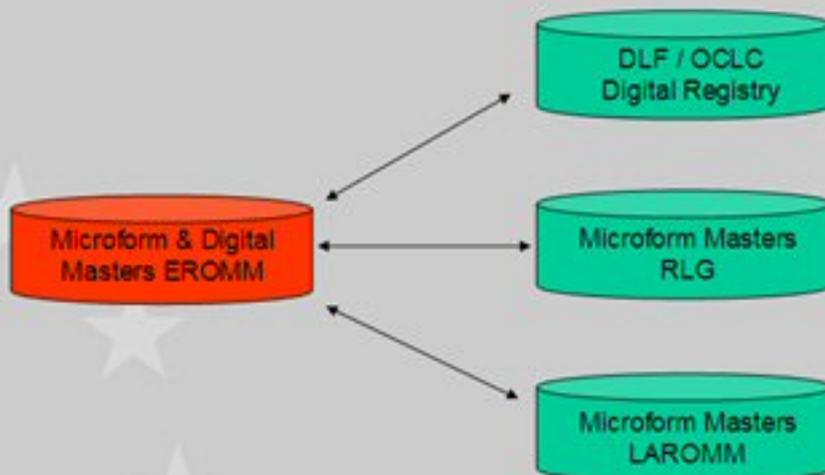
The LIBER/OCLC PICA proposal is based on developing the existing role of [EROMM](#) which has been in existence since 1990 and which since 1994 has included records of digital surrogates.

## European Contribution Model



EROMM would act as a coordinating centre for the registration of digital masters contributed from European libraries working within their own national library networks. These records would then be exchanged with the DLF/OCLC Registry and possibly others to create global coverage.

## Proposed Global Model



This model builds on the current workflows of libraries that catalogue initially in their regional or national union catalogues and the existing experience at EROMM in developing and promoting cataloguing and digitisation standards across Europe.

LIBER support for this initiative is very welcome and we believe LIBER can play a very proactive role in promoting the coordination of digitisation initiatives and standards. The initial registry contributors and their colleagues from ARL would also welcome LIBER support.

So far so good but as most of you know the exchange of bibliographic records at any level poses a number of challenges and requires good understanding and communication by all parties.

At this point EROMM has provided a sample file that has been analysed by OCLC. The review of this file has been shared with EROMM. There is already an appreciation that there is unevenness in content and quality and possibly purpose from records contributed by different libraries. I should add that there is nothing that we have found to date that would preclude an exchange of records.

The next step is to provide EROMM with sample records from the initial seeding North American libraries, which we expect will identify more variations and stimulate further discussion on standards. We hope that this will lead to a joint meeting later this year.

The goal will be to try and harmonise the metadata, digitisation and access standards initially between North America and Europe and ultimately elsewhere in support of building a global registry that will meet the initial objectives of the DLF. Your reaction, comments and ultimately participation, if not now to me, Paul Ayris, Raf de Keyser or any member of the LIBER Board will be very welcome as we begin the process in this early start up phase of the Registry.

## **REFERENCES**

Benchmark for Faithful Digital Reproductions of Monographs and Serials. Digital Library Federation Benchmark Working Group (2001-2002). <http://purl.oclc.org/DLF/benchrepro0212>

The Future of LIBER – a consultation paper. Prepared by MacDougall Consulting Ltd., January 2004. <http://www.kb.dk/liber/future/index.htm>

Registry of Digital Masters Record Creation Guidelines. Version 1.0 July 2004 prepared by DLF Digital Masters Working Group. <http://purl.oclc.org/DLF/rdm0704>

## **WEB SITES REFERRED TO IN THE TEXT**

DLF - Digital Library Federation. <http://www.diglib.org/dlfhomepage.htm>

EROMM. European Register of Microform Masters. <http://www.eromm.org/>

LIBER. <http://www.kb.dk/guests/intl/liber/>

OCLC PICA. <http://oclc-pica.org/>