

PDF Watermarking for Content Management Systems

A Datalogics, Inc. White Paper

Introduction

This document outlines an approach for implementing PDF watermarking for documents exported from Content Management Systems as a mechanism for extending information rights management capabilities outside the repository.

Background

Content Management Systems (CMSes) are traditionally implemented as document/information ‘repositories’ of information – generally a content storage location powered by a database technology, with rigorous security controls on the repository to control and audit access to the information stored within. These controls encompass things like:

- user authentication (login);
- role-based access (checkin/checkout, edit, make new versions, create/store new content);
- workflow (initiate, review, approve, comment); and
- auditing

The Challenge

CM vendors have been successful in controlling information (documents/files) while it is under the control of the CMS; it becomes more challenging to control these documents once they’ve left the repository:

- how do you control forwarding of documents to unauthorized parties?
 - how do you control unauthorized editing of a document (and subsequent dissemination of altered information) once it has left the repository?
 - how can you determine accountability in these situations?
-

Clearly, there are two distinct business challenges: first, how can you prevent the “leak” of privileged information; and second, how can you determine accountability when it happens?

Heavyweight Solutions

A number of ‘heavyweight’ solutions have been developed to address these needs: a recent entrant in this market is Adobe’s LiveCycle suite of document services (in particular, their Policy Server service). These solutions rely on embedding a “security token” with a document, once it’s left the repository, which causes the document to “phone home” to an authentication server to determine the access policy in real time before allowing it to be opened.

These types of solutions do allow the CM vendor to extend their information rights management (IRM) capabilities outside the repository; however, they also require additional overhead to:

- implement a service to receive and process authentication requests when documents outside the repository are opened; and
- define and maintain a set of policies to support the various permutations of access rights for each role, for each document type managed in the repository.

Watermarking: a Lighter-weight Solution

In some instances a lighter-weight mechanism may be more desirable, something which is: easy to implement; easy to maintain; does not place undue burden on the system; and does not intrude on the day-to-day activities of the user.

Adding a watermark to PDFs exported from the repository can be an effective way to provide a subset of IRM capabilities which, in many instances, might be sufficient and most appropriate. Indeed, some CMS vendors (such as IXOS/Open Text and d.velop AG in Germany) have already implemented such functionality.

A watermarking solution could be implemented as follows:

1. A request to retrieve a PDF from the repository is initiated (for onscreen presentation to the user, via “export” functionality, etc.).
2. The request is intercepted by a watermarking “module” (either be a physically distinct module, or a functional component of the CMS engine itself.).
3. The module edits the PDF document, inserting a watermark into the document. The watermark might include, for example, username, date/time information, and other security information as relevant.
4. The document request is then processed, and the (watermarked) document is presented to the user.

Functional Details

When developing such a module, the following features might be worth taking into consideration:

- Contents of the watermark should be configurable (e.g., specifying date/time, and date/time format; format of username (full name or userID); ability to include graphics).

- Appearance of the watermark should be configurable (e.g., ability to make watermark visible or not visible¹; font/style used; location of watermark - on the page, on all pages or just the first, etc.)
- Selective watermarking of PDFs – support the idea that perhaps some customers may have certain types of documents that they don't want watermarked, e.g., maybe fully public information like product collateral where this type of security/accountability isn't a concern.

Results

Implementing a watermarking module as described could have several benefits:

- it provides a basic level of accountability on exported PDFs;
- it mitigates the risk of leaking information (since the user's name will appear as part of the watermark, they will be more likely to be more careful with it); and
- it does it in a way that does not place unreasonable burdens on the system (performance), and does not impose restrictions or changes to the way that users interact with the system currently.

Summary

Implementing a “lightweight IRM solution” via watermarking could have wide applicability in CMS environments. Clearly such a system would not provide “end-to-end foolproof security” for documents outside the repository. Rather, it would provide a simple and flexible level of accountability that might be sufficient in many scenarios; and provides real added-value to customers.

Content Management vendors have already adopted such a mechanism within their systems. With the continued proliferation of PDF within the enterprise, it may be only a matter of time before this becomes a critical feature of any Content Management System.

*Author: Greg Manuel
January 15, 2007*

For questions or more information, please contact us at:

Datalogics, Inc.
101 N. Wacker Drive #1800
Chicago IL 60606 USA
+1 312 853 8200
<http://www.datalogics.com>

Copyright (c) 2007, Datalogics, Inc. All Rights Reserved.

Datalogics, the Datalogics Logo and all Datalogics product names are either trademarks or registered trademarks of Datalogics, Inc. All other trademarks are the property of their respective owners. Reproduction of this document in whole or in part without the express written consent of Datalogics, Inc. is prohibited.

This document and related materials and information are provided “as is” with no warranties, express or implied, including but not limited to any implied warranty of merchantability, fitness for a particular purpose, non-infringement of intellectual property rights, or any warranty otherwise arising out of any proposal, specification, or sample. Datalogics, Inc. assumes no responsibility for any errors contained in this document and has no liabilities or obligations for any damages arising from or in connection with the use of this document.

¹ Some customers may want to use the watermarking feature as a DRM-type “forensic” feature, with a hidden stamp that could be used for accountability purposes.