

EMC Centera and Oracle Records Database 10g Integration

Applied Technology

Abstract

This paper describes the joint solution of EMC Centera™ content-addressed storage and the Oracle Records Database platform. It provides the foundation for managing, securing, accessing, archiving, and integrating critical business information into your business processes. This joint solution offered by Oracle and EMC provides an easy way for users to leverage Oracle Records DB using a robust records management interface that takes advantage of Centera's retention capabilities.

August 2006

Copyright © 2006 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

All other trademarks used herein are the property of their respective owners.

Part number H2329

Table of Contents

Executive summary	4
Introduction to EMC Centera	4
Audience	4
Centera terminology.....	5
Centera terminology.....	5
Oracle Records Database 10g overview	5
Value proposition	6
Content Services architecture.....	6
How EMC Centera works with Oracle Records Database 10g	7
Declaring records in Oracle Records DB.....	8
Configuring the Content DB/Centera interface.....	8
Installing the EMC Centera software.....	8
Setting domain properties for EMC Centera in Oracle Enterprise Manager.....	8
Specifying credential information for EMC Centera	9
Creating a file plan in Oracle Records DB	10
Centera integration details	13
Conclusion	13
References	13

Executive summary

EMC and Oracle have partnered to provide a new, affordable, easy-to-use content management solution with long-term retention and assured integrity for content compliance. This is the first solution to offer online availability of records and content lifecycle while remaining compliant with the most stringent regulated environments – including SEC 17a-3 and 17a-4. This solution is ideal for all types of users to manage all types of information. This joint solution, which includes EMC Centera™ content-addressed storage and Oracle Records Database, provides the foundation for managing, securing, accessing, archiving, and integrating critical business information into your business processes.

Introduction to EMC Centera

Centera is a networked storage system specifically designed to store and provide fast, easy access to fixed content (information in its final form). It is WORM storage from the perspective that the content, once committed, cannot be changed yet once available for disposition, content can be deleted and the space reclaimed.

Centera is the first solution to offer online availability with long-term retention and assured integrity for content that is compliant with the most stringent regulated environments – including SEC 17a-3 and 17a-4.

Centera greatly simplifies the task of managing, sharing, and protecting large-scale fixed content repositories. Centera enables businesses to use this traditionally offline information to cost-effectively support new sources of revenue generation, expanded business models, and increase service levels to users and customers. A comprehensive SDK is provided for applications to interface to Centera using well-defined access methods.

Centera defines a new space within the EMC storage system portfolio — content-addressable storage (CAS). It is a category of automated networked storage established to store large volumes of fixed content. Each data object gets a unique identifier, or content address (CA), derived from its digital content. This ID is then typically stored in a database such as Oracle

- CAS eliminates the need for applications to understand and manage the physical location of information on storage media.
- Referencing data based on its actual content presents an additional storage benefit. If multiple clients store the same data object on a cluster, the system needs to store only one copy; each client, however, will receive a unique identifier to retrieve it.
- CAS ensures that stored data cannot be changed or tampered with because the identifier is calculated from the binary content of the data. If the data changes, the identifier also changes. Centera is the only solution in this market that meets the most stringent regulatory requirements for protecting sensitive data

Audience

EMC/Oracle customers and partners who are interested in understanding (a) the business value of EMC Centera with Oracle Records DB solution and (b) technical details of how to set up and configure the solution.

Centera terminology

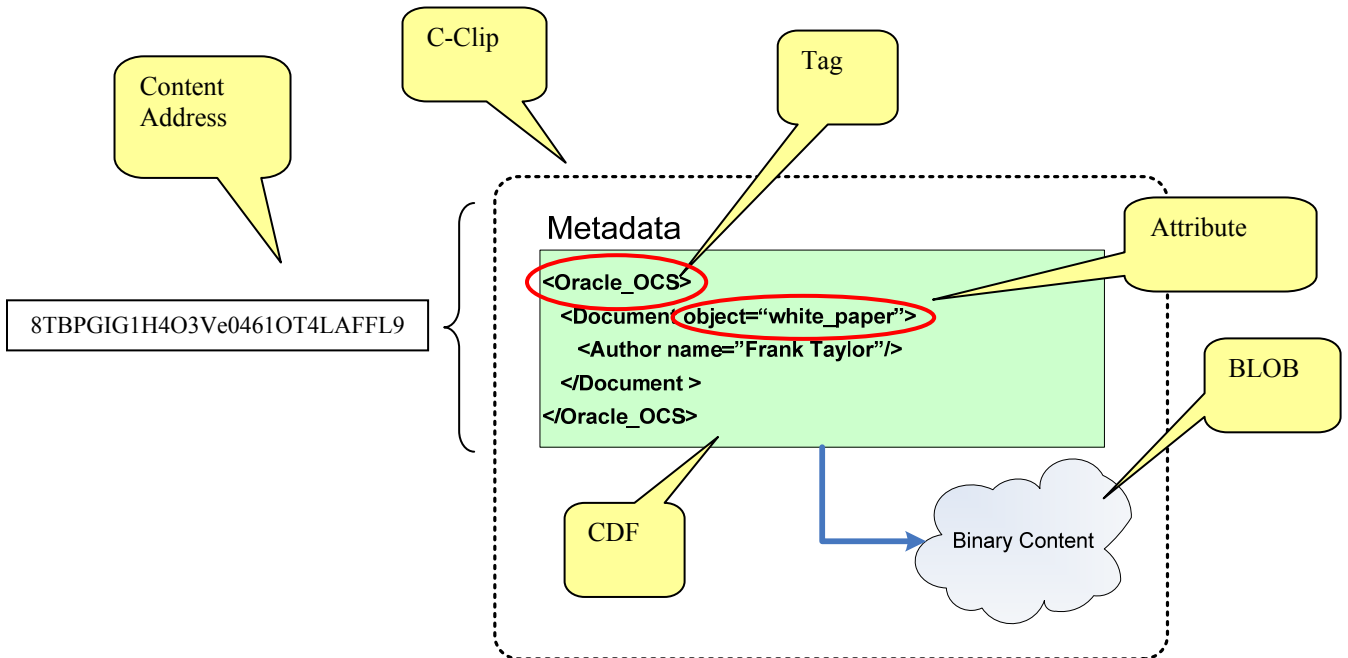


Figure 1. Elements of Centera

C-Clip™: A C-Clip is a package containing the user's data and associated metadata. When a user presents a file or a set of files to a Centera cluster, the system calculates a unique content address (CA) for the data and then stores this address in a newly created XML file, the C-Clip Descriptor File (CDF) together with application-specific metadata. The system then calculates another CA for the CDF, and stores the CDF plus the user's file in the complete C-Clip package on the cluster. The CA of the CDF is a handle for the C-Clip that the system uses to retrieve the user's data.

Tag: A C-Clip contains both user data and associated metadata in XML *tags*. Each tag has a name and can have one or more attributes. The tags representing actual data are called *blob tags*.

Blob: A Centera blob is the Distinct Bit Sequence (DBS) of the user's data. Do NOT confuse it with the term binary large object. Every file consists of a unique sequence of bits and bytes. The DBS represents the actual content of a file and is independent of the filename and physical location. A blob is represented in the C-Clip by a blob tag.

Content address: A content address is a unique identifier created for an object stored on Centera. The content address is generated based on the bit sequence of the object, and is stored in OCS, in order to retrieve the content later.

Oracle Records Database 10g overview

With the introduction of Oracle Records Database 10g, organizations can now manage large amounts of electronic records for regulatory compliance in a scalable, cost-effective, and secure manner. Oracle Records DB provides the foundation for the comprehensive declaration, classification, storage, retrieval, disposition control, and lifecycle management capabilities for electronic records. The records management capabilities are exposed in a simple and non-intrusive fashion focusing on minimizing end-user disruption to ensure successful adoption by all employees throughout the enterprise. Together with Oracle Content Database 10g, it is the first enterprise-wide solution offering a unified records and policy repository as part of the database infrastructure.

Oracle's strengths as an enterprise content management (ECM) solution include:

- Low cost to implement
- High database scalability
- Easily navigated hierarchy
- Fine-grained control over security
- Records management capabilities
- Strong search across all content

Value proposition

Leveraging web services, event-driven workflows, and business process management capabilities, Oracle Records DB allows you to records-enable any business process, portal, or collaborative environment. Using these records management features, companies can manage the records management process by:

- *Declaring records:* Allow content to be declared and classified as records either manually, automatically based on placement in Oracle Content DB, or programmatically leveraging the comprehensive web services support. Once a record has been declared, Oracle Records DB controls the retention and disposition of the document.
- *Managing policy:* Implement an enterprise-wide file plan with hierarchical record policy creation and management, including cutoff, retention periods, disposition actions, record freezes or legal holds, customized workflow notifications, and record searches to control the lifecycle of all your records.
- *Integrating storage management:* Transparently integrate WORM, near-line, and database storage with your record policies for a truly integrated information lifecycle strategy.
- *Disposing of records:* Allow lifecycle management and control over final disposition of the records. Administrators can configure Oracle Content DB 10g to conform to a corporate records retention policy by automatically retaining or disposing of recorded content.
- *Leveraging industry standards:* Web services and Business Process Execution Language (BPEL) support for integration with any legacy, business application, or business process. Oracle eBusiness Suite, PeopleSoft, and SAP users can leverage a standards-based integration to contextually capture records.

Content Services architecture

The logical architecture of Oracle Content DB 10g is shown in Figure 2.

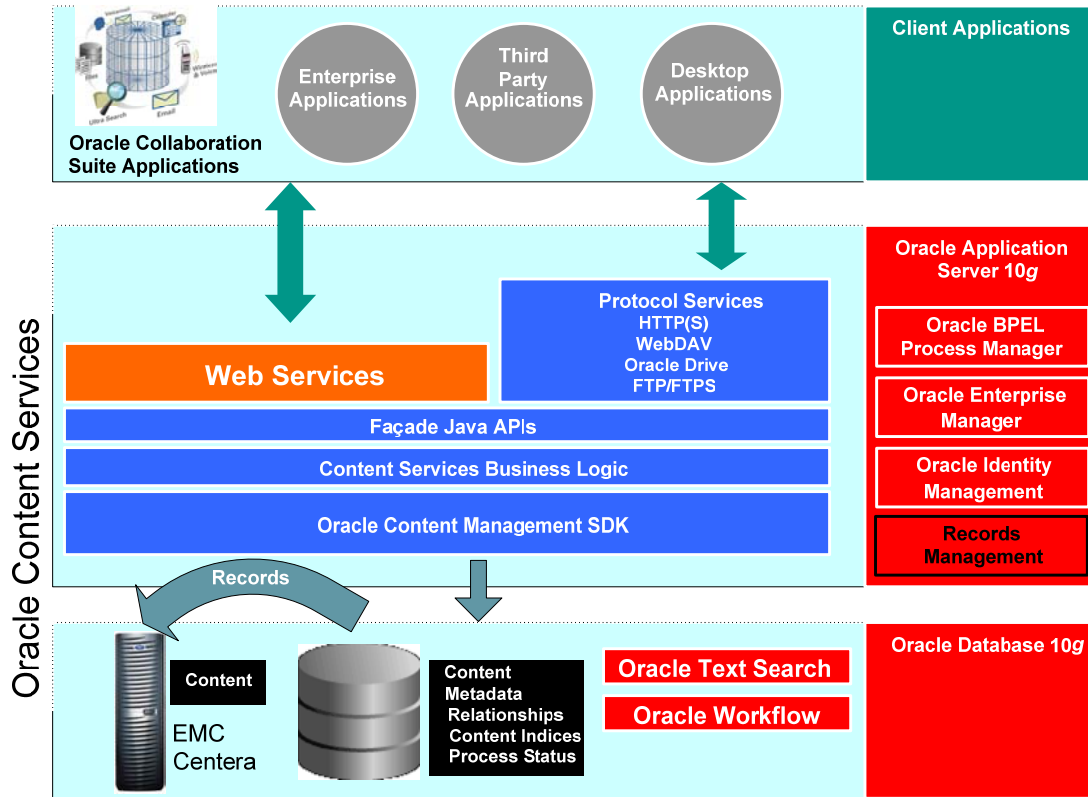
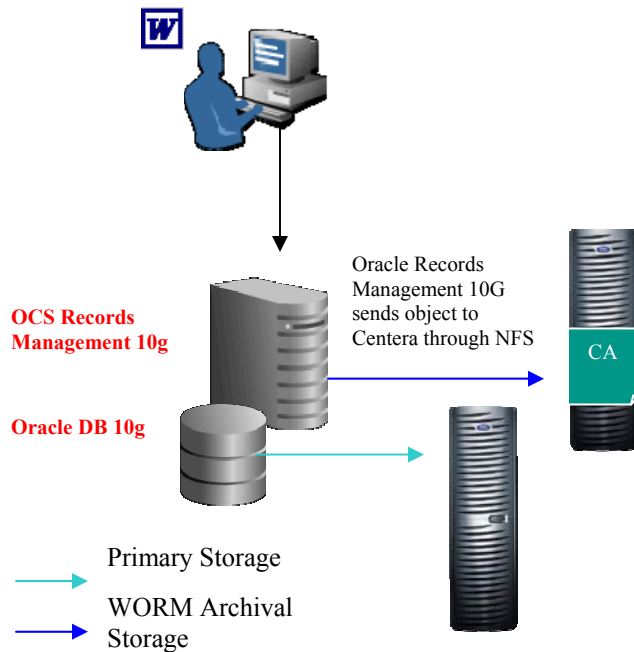


Figure 2. Oracle Content Database architecture

How EMC Centera works with Oracle Records Database 10g



- Configure the Centera interface in Oracle Content DB Enterprise Manager (see Figure 3)
- Create the file plan (disposition lifecycle) and associate the content with a file plan category
- Records DB lifecycle agent sets each file's retention period
- The content object (BLOB) moves to Centera
- The Oracle Database stores the Content Address that will allow later retrieval
- Centera enforces retention/disposition as configured in the OCS file plan category
- Centera prevents alteration, overwrite, or deletion of the file, before expiration of retention.

Figure 3 Centera working with Oracle Records DB 10g

Declaring records in Oracle Records DB

Oracle Records DB enables the declaration of records which can be moved to Centera. This can be accomplished in three ways – automatically using a policy associated with each folder, manually using a web interface, or using a workflow.

Configuring the Content DB/Centera interface

Follow the instructions in this section to configure Oracle Content DB for EMC Centera integration. Before you begin, make sure that Oracle Records DB has been configured and that the Records Management Lifecycle Agent has been activated.

You should perform the steps in the following sections on the Applications tier that is running the Records Management Lifecycle Agent. To determine which Applications tier is running this agent, use the Oracle Collaboration Suite Control to view each node to see where the agent is running

Installing the EMC Centera software

You must install the EMC Centera software on the Applications tier that is running the Records Management Lifecycle Agent. To do this, follow these steps:

1. Download EMC Centera 3.1 Patch 1 SDK or later from OracleMetaLink at <http://metalink.oracle.com>. To find the SDK on OracleMetaLink, search for patch number 5072277.

This SDK contains the necessary .jar and library files. Follow the instructions in the readme file to install the software.

2. Modify the **classpath** in the following XML files to include the FPLibrary.jar file:
 - ORACLE_HOME/opmn/conf/opmn.xml
 - ORACLE_HOME/j2ee/OC4J_Content/application-deployments/content/content/orion-web.xml
 - ORACLE_HOME/j2ee/OC4J_RM/application-deployments/rm/rm/orion-web.xml

3. In ORACLE_HOME/opmn/conf/opmn.xml, include the directory that contains the libraries in the appropriate path variable for the node, OC4J_Content, and OC4J_RM. This variable is called LD_LIBRARY_PATH on Linux and Solaris, SHLIB_PATH on HP-UX, and LIBPATH on AIX. For example, on Linux, include the following entry:

```
<environment>
<variable id="PATH" value "$LD_LIBRARY_PATH:absolute_path_to_Centera_lib_
directory"/>
</environment>
```

Note: If opmn.xml does not include an <environment> entry for the variable for your platform, you must create one.

Setting domain properties for EMC Centera in Oracle Enterprise Manager

Use the Oracle Collaboration Suite Control to set domain properties related to EMC Centera. To do this, follow these steps:

1. On the Content Services home page, click **Domain Properties**.
2. Click **IFS.DOMAIN.RETENTION.StorageDevice**. You may need to move to the next page to find this property, or you can use the **Search** field.
3. In the **Value** field, select **CENTERA** and click **OK**.
4. Click **IFS.DOMAIN.RETENTION.CENTERA.Configuration**.

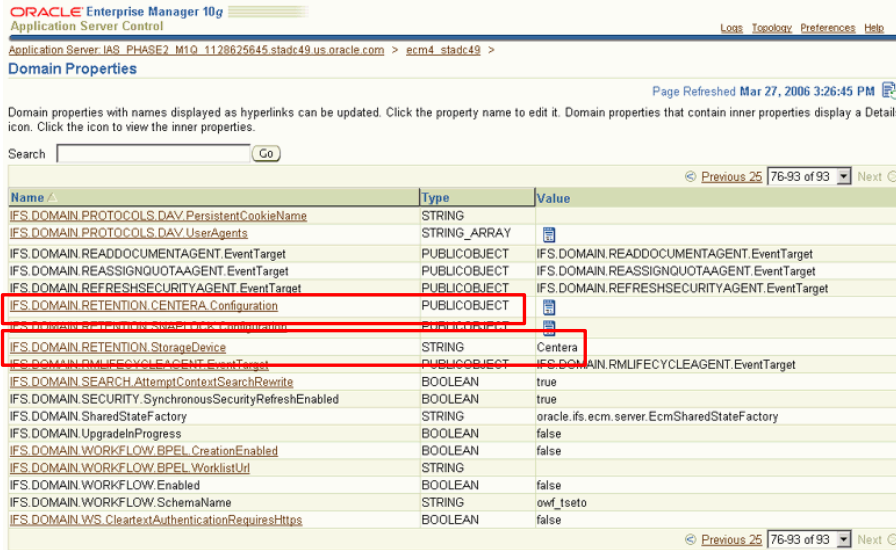


Figure 4. The Centera IP addresses are entered from the Domain Properties interface in Oracle Enterprise Manager.

5. Click **ADDRESSLIST**.
6. In the **Value** field, enter the comma-separated IP addresses of the EMC retention clusters and click **OK**. Which IP addresses for which Centera Access Node will depend on the customer's particular Centera configuration and failover requirements

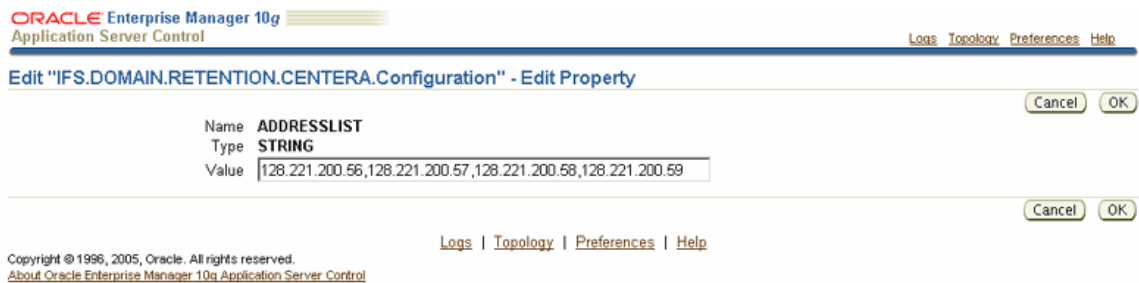


Figure 5. Enter a comma separated list of IP address for the Centera

7. On the Edit Domain Property page, click **OK**.

Specifying credential information for EMC Centera

Use the Oracle Collaboration Suite Control to specify credential information for EMC Centera. To do this, follow these steps:

1. From the Content Services home page, click **Retention Hardware**.

Retention Hardware

Choose a retention device to manage its connection credentials. Cancel OK

Retention Device Type:

Username:

Password:

Confirm Password:

Cancel OK

Figure 6. Enter username/password from the Retention Hardware page

2. In the Retention Device Type field, select EMC Centera.
3. Enter a username for EMC Centera. You must provide a username created by EMC Centera; do not provide an Oracle Content Services username.
4. Enter a corresponding password for EMC Centera, and confirm it in the Password field.
5. Click **OK**.
6. Return to the Content Services home page and click **Restart Domain**.

Creating a file plan in Oracle Records DB

A file plan or hierarchical set of record categories is created by the records administrator. Once logged in as an appropriately privileged user, the administrator is able to create a plan based on business need.

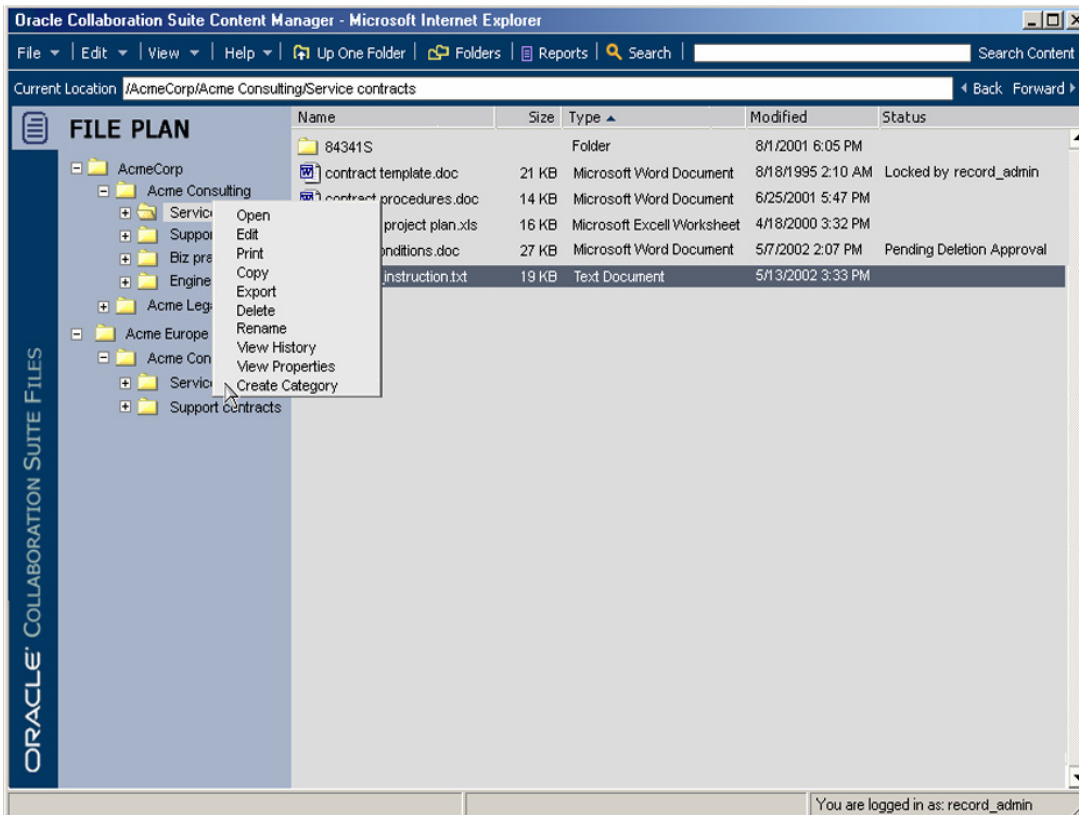


Figure 7. From the web interface, it is straightforward for a privileged user to create a new records category.

Once the records administrator has determined the appropriate part of the file plan to insert the new category, they are prompted to complete a 3-step process:

Step 1 – Name the record category

Enter the name of the category and a description. The name will appear in the list of categories that are made available to users when they choose which type of record they are submitting.

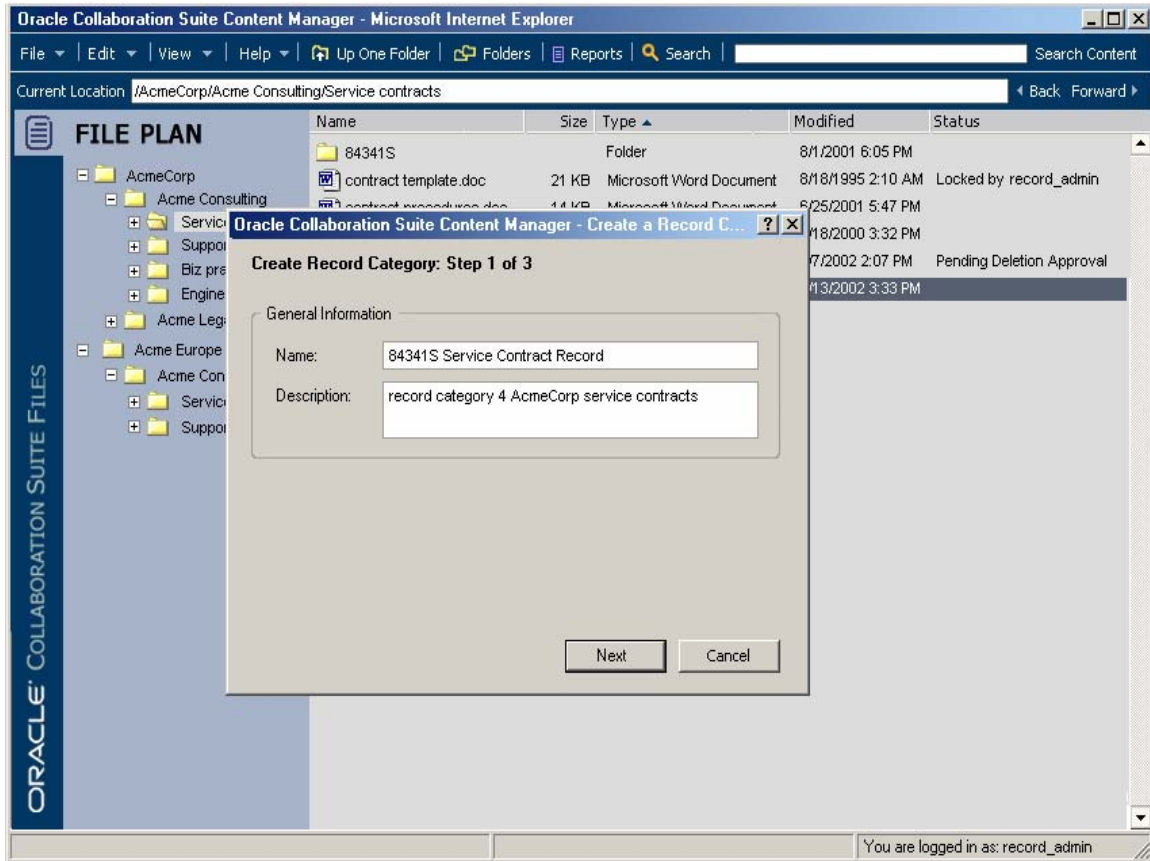


Figure 8. Entering a records category name

Step 2 – Adding category attributes

Category attributes are defined by the records administrator as part of the category. They are used to capture metadata that is pertinent to the category. Category attributes can be mandatory or optional and default values can be used. In Figure 9, the records administrator has defined three category attributes – all of which are required and have default values.

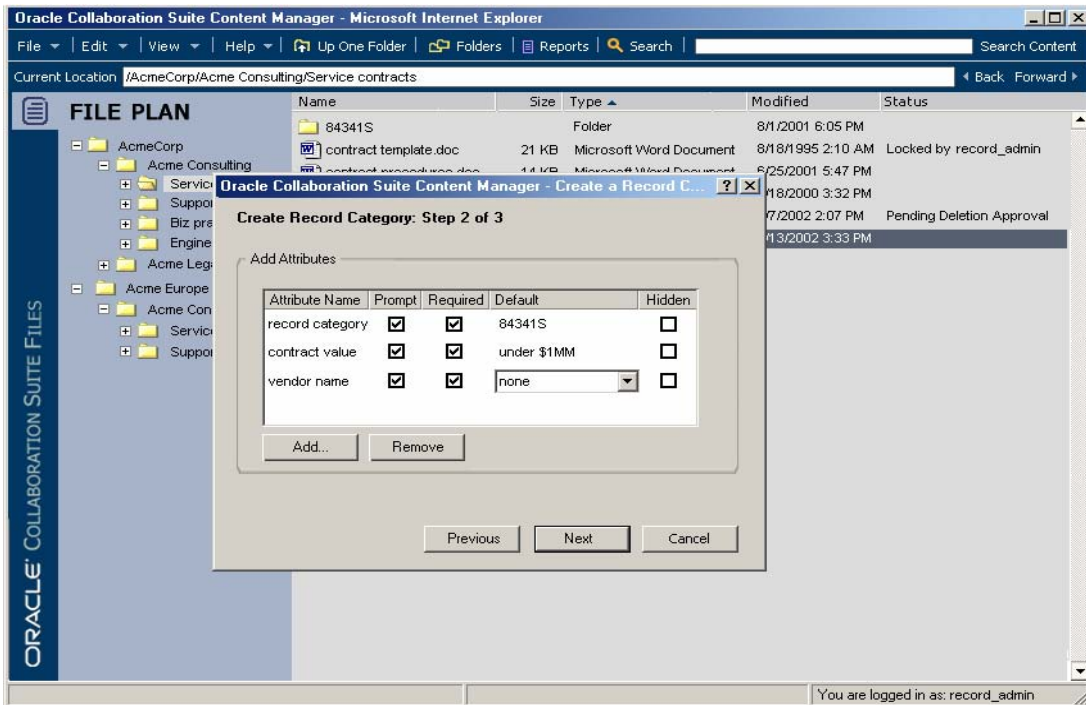


Figure 9. Defining the metadata that is to be acquired during records declaration

Step 3 – Defining the disposition policy

The disposition policy is where the retention and expiration instructions are defined, along with the storage destination for the record.

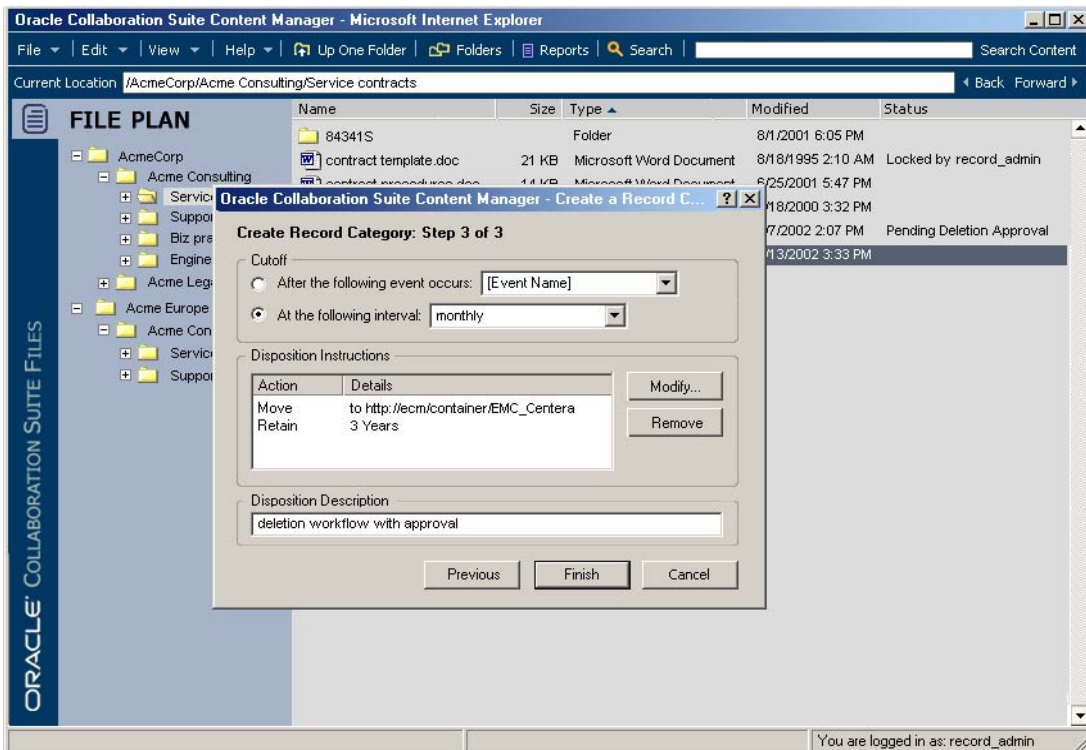


Figure 10. Defining disposition policy

In Figure 10, the disposition instructions indicate that the destination for content placed in this category is Centera and that it should be retained for a period of three years. Once that content has expired, the intent is for a workflow to be initiated prior that will dictate whether the content is eligible for deletion (in this example).

Centera integration details

Oracle Records DB 10g (version 10.1.2.2 – released in March 2006) uses the 3.1 Patch 1 Centera SDK version. Concurrency is achieved using a multi-threaded model – one thread per read request. As mentioned previously, Content DB uses the Oracle database as its central repository for metadata and content, so all read and write operations will involve a database operation. The average file size is on the order of 300 KB, although OCS will support BLOB content of up to 4 GB.

The concept of a “folder” in Content DB is a purely logical one, so the content must exist in this logical structure before it is operated on by the Oracle Records DB application. Therefore, it is important to remember that the content has to first be ingested into Content DB (that is, in the Oracle database as a BLOB) before being moved to Centera.

Conclusion

The EMC / Oracle partnership offers the first solution for online availability of records and content lifecycle while remaining compliant with the most stringent regulated environments. The Oracle Records Database platform manages the creation, policy definition, management, and disposition of records. The tight integration with EMC Centera provides users a comprehensive networked storage system for content that, once committed, cannot be change and yet once available for disposition, can be deleted. This joint solution offered by Oracle and EMC provides an easy way for users to leverage Oracle Records DB using a robust records management interface that takes advantage of Centera’s retention capabilities. The solution includes EMC Centera content-addressable storage and Oracle Records DB and provides the foundation for managing, securing, accessing, archiving, and integrating critical business information into your business processes.

References

For more information on Centera and Records DB:

- <http://www.EMC.com/products/systems/centera.jsp>
- <http://www.oracle.com/technology/products/recordsdb/index.html>
- <http://www.oracle.com/technology/products/contentdb/pdf/emc-centera-solutionsheet.pdf>