Replication Manager Installation Manual

Installation and configuration instructions

Version 1.0.31
This document describes the installation and configuration of the Replication Manager software, part of the CDI/RSM ordering system of SeaDataNet.
The current document can be found at:

References documents


Table of contents

1. Requirements ........................................................................................................................................... 5
   1.1. Operating system .......................................................................................................................... 5
   1.2. Java version ....................................................................................................................................... 5
   1.3. Tomcat ................................................................................................................................................ 6
   1.4. Network access .............................................................................................................................. 6
   1.5. Internet browser ............................................................................................................................. 6
   1.6. Supported databases ....................................................................................................................... 6

2. Installation .................................................................................................................................................. 8
   2.1. Install Sheet ....................................................................................................................................... 9
   2.2. Backup current version .................................................................................................................. 10
   2.3. Installation ....................................................................................................................................... 10
   2.4. Configuration ..................................................................................................................................... 11
       2.4.1. Security configuration (web.xml file) ..................................................................................... 11
       2.4.2. RM Configuration file (RMConfiguration.properties) .............................................................. 13
       2.4.3. Log configuration (log4j2.xml file) ......................................................................................... 16
       2.4.4. Configuration validation ......................................................................................................... 17

3. Maintenance .............................................................................................................................................. 18
   3.1. Log access ......................................................................................................................................... 18
   3.2. External resources ........................................................................................................................... 20
   3.3. Synchronization and checks .......................................................................................................... 20
   3.4. Database ............................................................................................................................................ 21

4. Usage ....................................................................................................................................................... 22
   4.1. Population ......................................................................................................................................... 22
   4.2. New metadata and data submission ............................................................................................... 22
   4.3. Restricted data requests process ..................................................................................................... 22

Illustrations

Figure 1 Default page (Summary) with configuration not set or invalid ....................................................... 11
Figure 2 Organization example .................................................................................................................. 14
Figure 3 Summary Page, Maintenance section .......................................................................................... 18
Figure 4 Logs access ................................................................................................................................... 18
Figure 5 external resources maintenance ................................................................................................ 20
Figure 6 Batches ready to send .................................................................................................................. 23
Introduction

The Replication Manager replaces the Download Manager in the new SeaDataCloud ordering system.

What does not change:
The coupling table is still the way to organize data declaration, using moduses 1 to 3.

What’s new:
- the Replication Manager is a unique Tomcat web application (no more batches)
- it communicates with two other components: Maris (IM/RSM) and EUDAT (cloud)
- it manages CDIs and unrestricted data ingestion: unrestricted data are stored in a cloud
- it manages restricted data orders (no more orders on unrestricted data, as they are available on the cloud)


1. Requirements

1.1. Operating system

The Replication Manager can be installed on both the Windows and Linux platform.

1.2. Java version

The Replication Manager requires at least Java Oracle JRE version 1.8u151, but lower than 1.9.

One of the recommended Java SE JRE can be downloaded from http://www.oracle.com/technetwork/java/javase/downloads/index.html.

Check if Java is available by executing command ‘java –version’ in the command line of windows or terminal of *nix distribution.
1.3. Tomcat

The Replication Manager requires at least Tomcat 8.5.31.

Download the Tomcat web server at https://tomcat.apache.org/download-80.cgi

The tomcat installation is not in the scope of this document. Please follow the Tomcat installation instructions from the official documentation:
https://tomcat.apache.org/tomcat-8.5-doc/index.html

1.4. Network access

The server shall be able to connect to:

- IM
  importmanager.seadatanet.org (IP 77.87.163.212), port 443
- BODC
  vocab.nerc.ac.uk, port 80
- CSR list
  seadata.bsh.de/isoCodelists/sdnCodelists/csrCodeList.xml

The server shall allow connections from:

- IM
  77.87.163.212
- RSM
  83.163.127.252
- EUDAT
  sdctest.csc.fi (195.148.30.56), port 80

1.5. Internet browser

The Replication Manager Dashboard can be accessed via a web browser.

Supported browsers:

- Google Chrome >= 59.0.3071
- Mozilla firefox >= 52.2.0

1.6. Supported databases

The Replication Manager can handle data and/or coupling table from a database, see [2].

The drivers loaded by the Replication Manager are displayed on the log file at each start.
Available drivers are:

**Oracle:**      ojdbc14 v12.2.0.3.0
**MySQL:**      mysql-connector-java v5.1.26
**PostgreSQL:** jdbc3 v9.1-901
**MS SQLServer:** sqljdbc4 v4.0
**Jtds:**        jtds v1.2.6
2. Installation

This section describes how to install the Replication Manager.

If data preparation is not already done, follow the Replication Manager User Manual [2] before perform installation.
2.1. Install Sheet

Replication Manager 1.0.31

This sheet aim is to gather all important information about your RM installation. Please fill it during the installation, this will be useful for assistance.

**ADMINISTRATIVE INFORMATION**

Partner name *Name of the Data Centre*

<table>
<thead>
<tr>
<th>edm code EDMO code of the Data Centre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Technical contact *The person who will be notified by the NAGIOS monitoring system that the Download Manager is no longer replying to the monitoring requests thus it is considered offline*

Name:  
email:  

**Approval contact** *(for data requests approbation)*. E-mail address of the contact at the Data Centre that will deal with data requests. *Login is the SeaDataNet ID obtained from the AAA authority server.*

Name:  
Login:  
email:  

**TECHNICAL INFORMATION about the server on which the RM is installed**

*Operating system*

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Java version*

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Tomcat version*

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Tomcat port*

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Server IP address and host name Public IP address and host name if any of the server on which the RM is installed*

IP address:  
Host name:  

*Proxy Is your network behind a proxy?*

YES / NO
2.2. Backup current version

To be ready to move backwards if needed, be sure that you save the previous Replication Manager application.

- backup previous Replication Manager webapp directory
  
  <tomcat_webapps>/ReplicationManager in a safe place

- backup database file in a safe place:
  
  o click on the BACKUP button in the Summary page, Database section
  
  o backup the file created in the backup directory (the database file path is in the RMConfiguration.properties file, given with the parameter “embeddedDatabase_backupDirectory_path”)

2.3. Installation

Download the lastest release of the Replication Manager at

https://www.seadatanet.org/Software/Replication-Manager

Put the Replication Manager war file (ReplicationManager.war) in your Tomcat webapps directory (override the previous one if exists). The application will be automatically deployed by Tomcat.

By default, the access is allowed only for localhost and Maris. You will see in the §2.4.1 how to modify security filters.

Open a web browser in the machine hosting the RM and check that the application is deployed successfully using the Replication Manager About page:

http://localhost:<port>/ReplicationManager

If a problem occurs, check the Tomcat catalina.out log file (default path is <tomcat_directory>/logs/catalina.out)

Check the configuration in the Replication Manager Summary page:

http://localhost:<port>/ReplicationManager/Summary

The result should be the page below:
2.4. Configuration

The configuration is set in the application RMConfiguration.properties, web.xml and eventually log4j2.xml files.

The modifications written in the RMConfiguration.properties, log4j2.xml and web.xml file are automatically reloaded by the server. The reload takes a few seconds and can be checked in the log:

```
******************** RMListener contextDestroyed ********************
[...]
******************** RMListener INITIALIZATION ********************
```

2.4.1. Security configuration (web.xml file)

The web.xml configuration file is located in the RM WEB-INF directory:

```
<tomcat_webapps>/ReplicationManager/WEB-INF/web.xml
```

The RM uses IP filters to allow or refuse to process requests from clients.

See also https://tomcat.apache.org/tomcat-8.0-doc/config/filter.html#Remote_Address_Filter

The web.xml file contains

- filters definitions (<filter> tags)
- filters mappings(<filter-mapping> tags)
A filter is applied on each RM access:

- Dashboard pages ("dashboard" section)
- Logs (RM LOG and RM CHECKER LOG sections)
- Files harvesting (HARVESTING section)
- RM API calls (API section)

Three filters are defined and should be sufficient:

- SDNSystem (Maris, EUDAT, localhost)
- Maris (also contains localhost)
- Eudat (also contains localhost)
- Local

Local contains only the localhost IP (ie allow only the machine on which the RM is installed).

You can add local IPs, or a mask representing your local network, separated by a pipe (|).

Each filter is identified by its name (<filter-name> tag).

```xml
<filter>
  <filter-name>local</filter-name> <!-- local only; add you local IP or network mask here -->
  <filter-class>org.apache.catalina.filters.RemoteAddrFilter</filter-class>
  <init-param>
    <param-name>allow</param-name>
    <param-value>127\.d+\.d+\.d+|::1|0:0:0:0:0:0:0:1|192.168.1.2</param-value>
  </init-param>
</filter>
```

Add you local IP address (or network mask) to all filters. Example: if your IP is 192.168.1.2, add it in local only filter as below:

```xml
<filter>
  <filter-name>local</filter-name> <!-- local only; add you local IP or network mask here -->
  <filter-class>org.apache.catalina.filters.RemoteAddrFilter</filter-class>
  <init-param>
    <param-name>allow</param-name>
    <param-value>127\.d+\.d+\.d+|::1|0:0:0:0:0:0:0:1|192.168.1.2</param-value>
  </init-param>
</filter>
```

Then add also this IP in the Maris, EUDAT ans SDNSystem filters.
2.4.2. RM Configuration file (RMConfiguration.properties)

This file is located in the <tomcatWebApps>/ReplicationManager/WEB-INF/classes directory. The RMConfiguration.properties file contains the RM configuration that shall be customized by each Data Centre.

If you need to change a parameter value in the RMConfiguration.properties file, just refresh the Dashboard Summary page after you have modified the file.

If you need to change some configuration elements outside the configuration file (create a directory, for example), you will need to launch the configuration checker manually, using the “RELOAD” button on the Dashboard Summary page.

Important elements of this configuration are the directories paths. The figure below shows what are these path and how they can be organized.
Requested paths in configuration

All paths are independent, you can choose your own organization

Example of organization

Figure 2 Organization example
All paths shall be absolute paths.
The resulting RMConfiguration.properties file is shown below.

In **red**: fields that you need to modify
In **brown**: fields that you may want to modify (optional)

```
# REPLICATION MANAGER CONFIGURATION
#--------------------------------------------------------------------------------
#TEST MODE : 0 is production, 1 is test
test_mode=1

# EDMO CODE
edmo_code=xxxxx

# WORKFLOW ORGANIZATION
#--------------------------------------------------------------------------------
readyToSendCDIs_path=/home/ReplicationManager/workspace/readyToSendCDIs
tmpDirectory_path=/home/ReplicationManager/workspace/tmp
queueDirectory_path=/home/ReplicationManager/workspace/queue

# ARCHIVES
archive_path=/home/ReplicationManager/ARCHIVES

# PRODUCTION DATA FILES
production_path=/home/ReplicationManager/PRODUCTION

# RM DATABASE
embeddedDatabase_path=/home/ReplicationManager/RMDatabase/rmDatabase.odb
embeddedDatabase_backupDirectory_path=/home/ReplicationManager/RMDatabase/backups

# DATA ORGANIZATION
data_path=/home/ReplicationManager/RMData/data
mapping_files_path=/home/ReplicationManager/RMData/mapping

# COUPLING TABLE
# Coupling table type : 0 if coupling table is in file or 1 if coupling table is in database
# for coupling in file only : coupling table file path
coupling_table_type=0
coupling_table_file_path=/home/ReplicationManager/RMData/coupling.txt

# for coupling in database only : database connexion
#coupling_table_connection=jdbc:oracle:thin:@195.178.224.89:1312:database_name
#coupling_table_user=user
```
2.4.3. Log configuration (log4j2.xml file)

The log4j2 configuration file is located in the RM WEB-INF/classes directory:
<tomcat_webapps>/ReplicationManager/WEB-INF/classes/log4j2.xml

The default RM log files path is the log directory in the default tomcat path:
<tomcat_directory>/logs/
You can customize using absolute paths, as shown below.
Eg: fileName="/home/RM/logs/replicationManager.log"

The default configuration creates a new log file as soon as the current one exceeds 250Mo.
The paths that can be customized are in red.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Configuration status="WARN">
    <Appenders>
        <Console name="Console" target="SYSTEM_OUT" follow="true">
            <PatternLayout>
                <pattern>%d{ISO8601}[GMST][\[-]RM\]]-%p-%m (%c{3}%l) %n</pattern>
            </PatternLayout>
        </Console>
        <RollingFile name="RollingFile" fileName="${sys:catalina.home}/logs/replicationManager.log" filePattern="${sys:catalina.home}/logs/replicationManager.log.%d{yyyy-MM-dd}.%i">
            <PatternLayout>
            </PatternLayout>
        </RollingFile>
    </Appenders>
</Configuration>
```
2.4.4. Configuration validation

Once the configuration is done, go to the RM dashboard Summary page and click on the “RELOAD” button:

http://<host>:<port>/ReplicationManager/

The Configuration in the Summary page shall be valid:

**Configuration**

![Configuration is valid](last reload: 20190607T12:31:48Z) **RELOAD**

RM configuration is valid

If the configuration is not valid

- check the error messages in the “Custom Parameters” table (status column)
- fix the parameter(s) value(s) in the RMConfiguration.properties file if needed
- check the files and directories
- click on RELOAD button

3. Maintenance

Information about the system are available on the Summary page, “System” section:

- Operating system
- Tomcat version
- Java version
- RM version

Several maintenance tools are available from the Summary page, “Maintenance” section:

- logs access
- BODC and other external resources versions and update launcher
- synchronization tools
- database backup launcher

![Figure 3 Summary Page, Maintenance section](image)

3.1. Log access

In the dashboard Summary page, the “Logs” section gives URLs to access the current RM log and RM Checker log.

![Figure 4 Logs access](image)
The log paths are given in the log4j2 xml configuration file (§2.4.3).

Direct URLs are:
http://<host>:/<port>/ReplicationManager/RMLog
http://<host>:/<port>/ReplicationManager/RMLogChecker

Older logs can be accessed using URL parameters date and index:
- Date = the log file date
- Index: the log file number in this day

The logs pages are not auto-refresh. Click on F5 to refresh.

The log section RM logs list opens a page listing all the existing log files.
3.2. External resources

This section displays the BODC vocabularies version and the CSR list file last update date.

To update all, click on the “UPDATE” button.

Figure 5 external resources maintenance
No message is displayed on the page, but the updates can take a while.

Wait 10mn and refresh the page to display the new versions.

3.3. Synchronization and checks

The RM dashboard Synchronization section offers the following functionalities:

- Launch a local check (check coupling table)
- Synchronize (check consistency between local coupling table, local database and IM catalogue)
- Populate the system

WARNING: The population should be launched only on first RM installation!
3.4. Database

The objectDB database used by the RM is stored in a file. This file can be backed up using the “BACKUP” button in this section.

The database file path is defined in the RMConfiguration file with parameter `embeddedDatabase_path`. The backup directory is defined by the parameter `embeddedDatabase_backupDirectory_path`. 
4. Usage

4.1. Population

Once the RM is correctly configured, it has to be populated using metadata and data already existing before RM installation.

Population process will:

- Download the catalogue from the IM
- Save metadata information in the RM embedded database
- Generate and store in ARCHIVE the restricted data

The metadata CDI files and unrestricted data are stored on the cloud by Maris.

4.2. New metadata and data submission

The “Ready To Send” directory is the directory defined by the “ReadyToSendCDIs_path” parameter in the RMConfiguration.properties file. This path can be read from the Summary page, in the “Custom Parameters table”.

Zips shall be placed in the “Ready To Send” directory, so they become visible in the Batches In Progress page.

From this page, Data Centre Manager can select and submit batches. He can follow the workflow using “Batches cancelled”, “Batches in production” and “Local CDI Ids” pages.

4.3. Restricted data requests process

Incoming restricted data requests will automatically be processed by the RM.

Logs shall be accessed in log file (no information in the dashboard).
Choose the CDIs batches that you want to submit (by clicking on the rows), then click on the “Submit” button.

The batch will start the automatic workflow:

- Local preparation -> in queue
- Ingestion -> Current batch (one at a time)
- Workflow ended -> cancelled or in production

At each time, click on the batch name to display a detailed page.