



Australian Government
National Archives of Australia

Checksum Checker 3.0.0

User Manual

Version 1.5

RKS: 2013/1309

the archives

naa.gov.au

Document Change Record

| Version | Changed By | Description of Changes | Change Date |
|---------|-------------------|--|-------------|
| 0.1 | Allan Cunliffe | Created (based on <i>Rolling Checksum Checker</i> document, created by Justin Waddell). | 14/01/2010 |
| 0.2 | Allan Cunliffe | Changes with v2 Checksum Checker Service. | 11/02/2010 |
| 1.0 | Allan Cunliffe | Internal review | 25/02/2010 |
| 1.1 | Allan Cunliffe | Removed reference to <i>Rolling Checksum Checker</i> . Replaced with <i>Checksum Checker</i> . | 16/04/2010 |
| 1.2 | Allan Cunliffe | Updated location of files for download. | 18/05/2010 |
| 1.3 | Allan Cunliffe | Changed Data Object to Digital Object. Added information on configuration and the Checker Properties file. | 19/08/2010 |
| 1.4 | Allan Cunliffe | Update to starting the Checksum Checker service. | 09/05/2011 |
| 1.5 | Kirti Chennareddy | Full review and update | 30/06/2013 |

Related Documentation

| Title | Author | Date | RKS |
|---------------------------------------|-------------------|----------|-----|
| Digital Preservation User Manual v1.5 | Kirti Chennareddy | 01/06/13 | |
| Rolling Checksum Checker.odt | Justin Waddell | 01/02/07 | |

Glossary

| Term | Definition |
|-------------------------------------|---|
| AIP | Archival Information Package: A digital record (file, document) with preservation metadata, in a preservation format. |
| Checksum | A numerical value calculated from the contents of a digital object. By comparing a recently determined checksum with an older one, you can tell if the digital object has changed. |
| Checksum Checker | <p>Standalone software that runs regular checksum checks of the AIPs stored in the Digital Archive.</p> <p>Required for the DR Facility to compile reports from the Digital Archive.</p> <p>Formerly called Rolling Checksum Checker.</p> |
| Digital Archive | Archival storage for digital material. |
| Digital Object | <p>An object composed of a set of bit sequences. Once normalised, it becomes an AIP.</p> <p>A Submission Information Package (SIP) in OAIS terms.</p> |
| Digital Preservation Recorder (DPR) | <p>Software that manages the workflow for the NAA digital archiving process. Consists of three distinct stages:</p> <ul style="list-style-type: none"> • Quarantine • Preservation • storage in the Digital Archive. |
| GUI | Graphical User Interface |
| Rolling Checksum Checker | See Checksum Checker. |
| Service | A program designed to perform specific functions without interaction from a user. |
| SMTP | Simple Mail Transfer Protocol. An Internet standard for email transfer across Internet Protocol (IP) networks. |
| Telnet | Teletype Network. An network protocol used on the Internet or a local area network to enable one computer to communicate with another. |

Table of Contents

| | | |
|-----------|--|-----------|
| 1. | INTRODUCTION..... | 5 |
| 2. | FUNCTIONAL OVERVIEW..... | 5 |
| | 2.1 Automated Emails..... | 5 |
| 3. | INSTALLATION..... | 5 |
| | 3.1 Windows..... | 6 |
| | 3.2 Linux..... | 6 |
| 4. | CHECKSUM CHECKER SERVICE..... | 7 |
| | 4.1 Starting the Checksum Checker Service..... | 7 |
| | 4.2 Telnet..... | 7 |
| 5. | APPENDIX A – CHECKSUM CHECKER SEQUENCE FLOW..... | 8 |
| 6. | APPENDIX B – COMMAND LINE CONFIGURATION SETTINGS..... | 9 |
| 7. | APPENDIX C – CHECKER PROPERTIES FILE..... | 11 |
| | 7.1 Email settings..... | 11 |
| | 7.1.1 smtp.server..... | 11 |
| | 7.1.2 smtp.port..... | 11 |
| | 7.1.3 admin.address..... | 11 |
| | 7.2 DPR settings..... | 11 |
| | 7.2.1 rollingchecker.dpr.repository..... | 11 |
| | 7.3 DPR Database Settings..... | 11 |
| | 7.3.1 rollingchecker.db.username..... | 12 |
| | 7.3.2 rollingchecker.db.password..... | 12 |
| | 7.3.3 rollingchecker.db.url..... | 12 |
| | 7.3.4 rollingchecker.db.driverclass..... | 12 |
| | 7.4 General settings..... | 12 |
| | 7.4.1 rollingchecker.period..... | 12 |
| | 7.4.2 rollingchecker.period.unit..... | 12 |
| | 7.4.3 rollingchecker.logfile..... | 12 |
| | 7.5 Thread pool settings..... | 13 |
| | 7.5.1 rollingchecker.thread_pool_size | 13 |
| | 7.5.2 rollingchecker.sleep_time..... | 13 |
| | 7.5.3 rollingchecker.maximum_queued_jobs..... | 13 |
| | 7.6 Sample Checker Properties File..... | 14 |
| 8. | APPENDIX D – TELNET COMMANDS..... | 15 |

1 Introduction

Checksum Checker is a piece of software that is used to monitor the contents of the digital archive for data loss or corruption.

As part of the Digital Preservation Recorder (DPR) workflow, checksums are generated for each Archival Information Package (AIP). Checksum Checker generates a new checksum for each AIP and compares it against the stored checksum. If the checksums do not match, then the AIP is flagged as being corrupt.

Reports about the content of the digital archive can be generated from the DPR. The content of many of these reports is dependent on Checksum Checker being run.

2 Functional Overview

Checksum Checker functions as a service. The Checksum Checker service can be controlled either via a telnet session or the graphical user interface (GUI).

A sequence flow diagram describing how Checksum Checker works is shown in **Appendix A – Checksum Checker sequence flow**.

2.1 Automated Emails

Checksum Checker sends automated emails to a nominated administrator email address, coinciding with certain events (the administrator email address is specified in the Checker Properties configuration file).

The following table describes the automated emails sent by Checksum Checker.

| Event | Email contents |
|-----------------------------|--|
| Checksum Checker is started | <ul style="list-style-type: none">• start time• database |
| An error is detected | <ul style="list-style-type: none">• AIP path and filename• nature of error (for example: <i>file not found or could not read</i>) |
| A run is completed | <ul style="list-style-type: none">• start time• total run time (hours and minutes)• database• where applicable:<ul style="list-style-type: none">• AIPs flagged as having an error• nature of any errors found |

3 Installation

Prior to installing Checksum Checker make sure you have the following:

- a running SMTP server

- administrator email account
- Java version 1.7 or later installed.

3.1 Windows

1. Download the Windows executable from:

<https://sourceforge.net/projects/checksumchecker/files/>

2. Double-click the *Checksum Checker <version> Setup.exe*

Result: The installation wizard is displayed.

3. To start the installation, click **Next**.
4. To complete the installation, click **Install**.

Result: Checksum Checker is installed under *Program Files\National Archives of Australia\Checksum Checker*.

5. To exit the installation wizard, click **Finish**.
6. To start Checksum Checker, double-click the Checksum Checker executable.

3.2 Linux

1. Download the Checksum Checker zip file from:

<https://sourceforge.net/projects/checksumchecker/files/>

2. Extract the *checksum-checker.jar* file to a directory on the machine containing *dprclient.jar*.

4 Checksum Checker Service

The Checksum Checker service runs in the background, checking the digital repository for any errors. The output of these checks can be viewed from the command line or the Checksum Checker GUI. The Checksum Checker service runs on port 9362.

4.1 Starting the Checksum Checker Service

The following procedure describes how to start the Checksum Checker service:

1. From the command line, go to the location of *checksum-checker.jar*
2. Enter the following:

```
java -cp checksum-checker.jar:dprclient.jar
au.gov.naa.digipres.rollingchecker.CheckerServer -c <
path_to_checker.properties>
```

Result: The Checksum Checker service starts.

Note: The above procedure assumes that the:

- *checksum-checker.jar* file is in the same location as the *dprclient.jar* file
- contents of the Checker Properties configuration file are correct (refer to **Appendix C – Checker Properties File**). You can override some or all of the settings in the Checker Properties configuration file by passing the additional arguments in step 1 above. Refer to **Appendix B – Command Line Configuration Settings** for more information.
- Delete the checker.lock file if re-running the service.

4.2 Telnet

Telnet can be used to control the Checksum Checker service. For example, via a telnet session, you can pause or start the Checksum Checker service.

The telnet commands available are:

- pause a run
- resume a run
- get the status of a run
- see run history.

To connect to the Checksum Checker service via telnet:

1. From the command line, enter:

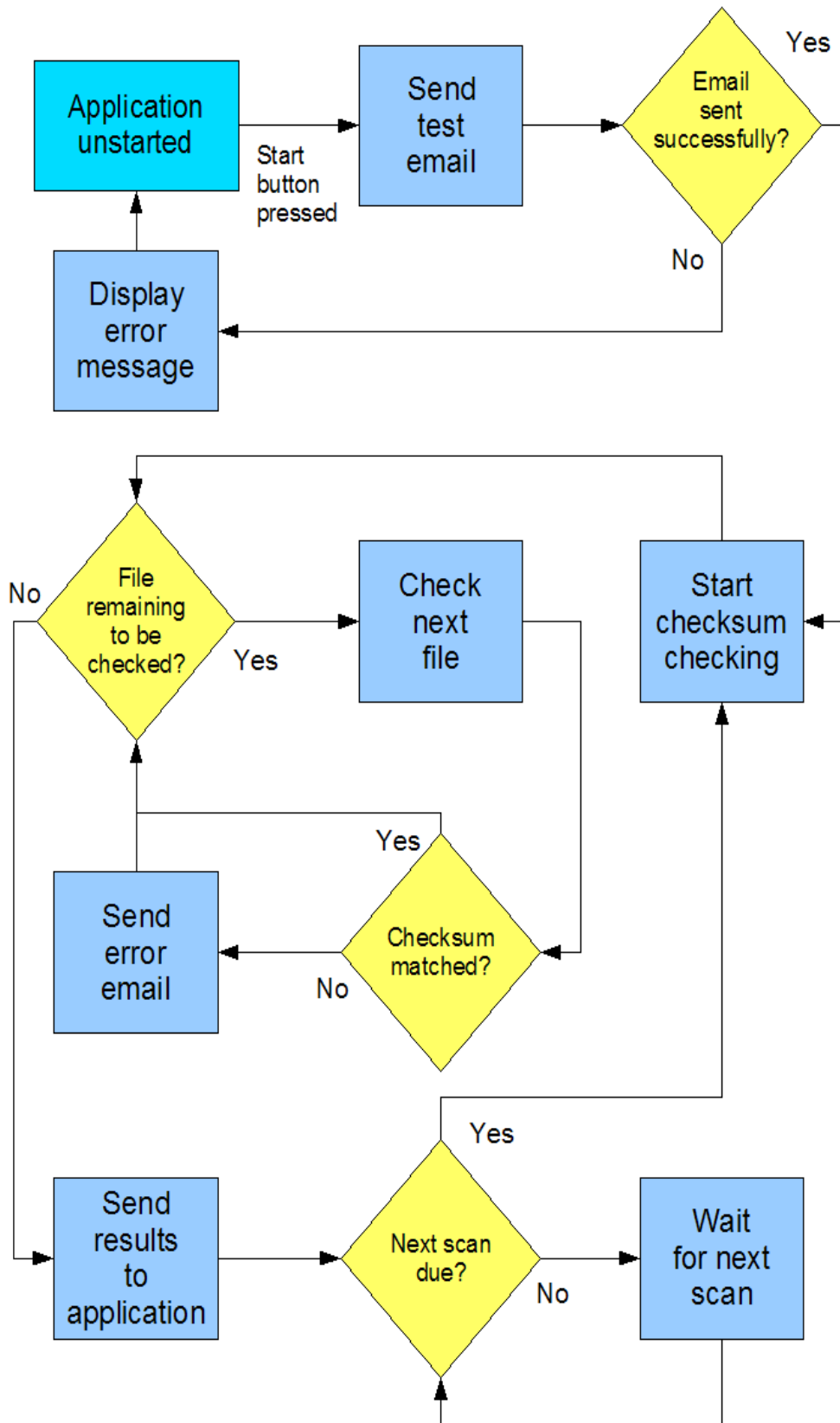
```
telnet <machine_running_Checksum_Checker_service> 9362
```

2. For help, enter:

```
help
```

Result: The telnet help is displayed. Refer to **Appendix D – Telnet commands** for more information.

5 Appendix A – Checksum Checker sequence flow



6 Appendix B – Command Line Configuration Settings

When running the Checksum Checker Service from the command line you must include a reference to the Checker Properties configuration file (called *checker.properties*). You can also specify a number of settings which will override the settings in the Checker Properties configuration file.

To access a list of the configuration settings:

1. From the command line, go to the location of the *checksum-checker.jar* file
2. Enter:

```
java -jar checksum-checker.jar -help
```

Result: A list of arguments is displayed:

Usage:

```
CheckerServer [--config-file=<path to configuration file>]
[--smtp-server=<smtp server>]
[--smtp-port=<smtp port>]
[--smtp-admin-address=<admin email address>]
[--username=<db username>] [--password=<db password>] [--url=<db
URL/Connection string>]
[--driver-class=<db driver class>]
[--repository=<path to repository>] [--period=<period>]
[--period-units=<period units>] [--log=<path to log file>]
```

Arguments:

-c <PATH> | --config-file=<PATH> Config file to use.

-s <ADDRESS> | --smtp-admin-address=<ADDRESS> SMTP server for sending emails.

-p <PORT> | --smtp-port=<PORT> SMTP server port.

-e <EMAIL ADDRESS> | --smtp-admin-address=<EMAIL ADDRESS> Email address to send email notifications to.

-U <USERNAME> | --username=<USERNAME> Database username to connect to the dpr database as.

-P <PASSWORD> | --password=<PASSWORD> Database password.

-C <URL/CONNECTION STRING> | --url=<URL/CONNECTION STRING> The URL/Connection string used to connect to the database.

-d <DRIVER CLASS> | --driver-class=<DRIVER CLASS> The class path of the Driver.

-D <INTEGER> | --period=<INTEGER> Set the period.

-u <PERIOD UNIT> | --period-units=<PERIOD UNIT> Set the period units {second, hour, day}.

-l <PATH> | --log=<PATH> Set the log file.

-h | --help Show this information.

-v | --version Show the version.

7 Appendix C – Checker Properties File

The Checker Properties file is used to configure Checksum Checker. You can not start Checksum Checker without reference to a valid Checker Properties file.

If you have installed Checksum Checker with the Windows installer or the DPSP installer, the Checker Properties file (called *checker.properties*) is located in the Checksum Checker directory. Most of the details in the file will be correct for your local environment. However, you will still need to enter an administrator email address and the location of the DPR repository. Depending on your setup, you may also need to modify the SMTP server and port settings.

The following is a description of the settings in the Checker Properties file:

7.1 Email settings

These settings tell Checksum Checker where to send email notifications.

7.1.1 smtp.server

The name, address or IP address of your SMTP server.

7.1.2 smtp.port

Set the SMTP port for sending email. This is usually port 25.

7.1.3 admin.address

Set the email address of the person responsible for monitoring the output of Checksum Checker. This will take the form of:

`<user.name>@<domain>`

7.2 DPR settings

7.2.1 rollingchecker.dpr.repository

Path to the repository is no longer needed in the checker.properties file when using Checksum Checker version 3.0.0. This version is designed to work with DPR 6.0.0 where DPR Server is used for the transfer of files to the repository. Repository path is set in the repository_locations table in the DR database as part of the DPR server configuration. For more details, please refer to DPR User Manual v 1.5.

If using Read\Write mounts for transfer of files with DPR 6.0.0 or if using any earlier versions of DPR, use checksum checker v 2.0.5.

7.3 DPR Database Settings

These settings tell Checksum Checker where your DPR repository (database) is located.

7.3.1 rollingchecker.db.username

Enter the user name you use to log in to the QF, PF or DR facilities in DPR.

7.3.2 rollingchecker.db.password

Enter the password you use to log in to the QF, PF or DR facilities in DPR.

7.3.3 rollingchecker.db.url

This is the database URL (or JDBC URL) of your DPR database. It should have the form of:

```
jdbc:postgresql:[servername]/[databaseName]
```

Note: All standard database URLs should commence with the string **jdbc**.

7.3.4 rollingchecker.db.driverclass

Enter:

```
org.postgresql.Driver
```

7.4 General settings

These settings govern how often Checksum Checker performs its checks and where the log file is created.

7.4.1 rollingchecker.period

This is how often Checksum Checker will check the repository for errors.

Note: The actual time interval is determined by *period* **and** *unit* settings. For example, if you enter **30** in the *period* setting and then **hour** in the *unit* setting, Checksum Checker will run every 30 hours.

7.4.2 rollingchecker.period.unit

This is the unit of time which determines how often Checksum Checker will check the repository for errors. The available units are:

- second
- hour
- day.

7.4.3 rollingchecker.logfile

Enter the location and name to be used for the log file.

Note: The path to the log file must be valid. The log file does not need to exist as it will be created when the Checksum Checker Service is started.

7.5 Thread pool settings

7.5.1 rollingchecker.thread_pool_size

This is the maximum number of active threads in the pool. Default value is 1000.

7.5.2 rollingchecker.sleep_time

Number of milliseconds that an idle thread is suspended. Default value is 100.

7.5.3 rollingchecker.maximum_queued_jobs

Maximum number of runnable tasks that can queue up awaiting execution. Default value is 100.

7.6 Sample Checker Properties File

The following is an example of a *checker.properties* file that would be used in a Windows environment. **The settings are examples only - you need to adjust them to suit your own environment.**

Note: In Windows, it is important to escape all backslashes (“\”) with another backslash.

```
# Email settings
smtp.server=localhost
smtp.port=25
admin.address=admin@localhost

# DPR settings
#rollingchecker.dpr.repository=//dpr_data/dr

# DPR Database settings
rollingchecker.db.username=dpr
rollingchecker.db.password=dpr
rollingchecker.db.url=jdbc:postgresql://localhost:5432/dr
rollingchecker.db.driverclass=org.postgresql.Driver

# General settings
rollingchecker.period=30
rollingchecker.period.unit=second
rollingchecker.logfile=//tmp/checker.log

#Thread pool settings
rollingchecker.thread_pool_size=1000
rollingchecker.sleep_time=100
rollingchecker.maximum_queued_jobs=100
```

8 Appendix D – Telnet commands

The available Checksum Checker telnet commands are:

`status, s`

Print information about the current status of the checksum checker

`history, results`

Print a summary of previous checker run results

`pause, p`

Pause the checksum checker

`resume, start, r`

Resume the checksum checker

`help, h, ?`

Print this message

`quit, exit, q`

Close the connection to the server (the server will continue to run)