

Reptor Manual

Thomas Zastrow, 2017



Table of Content

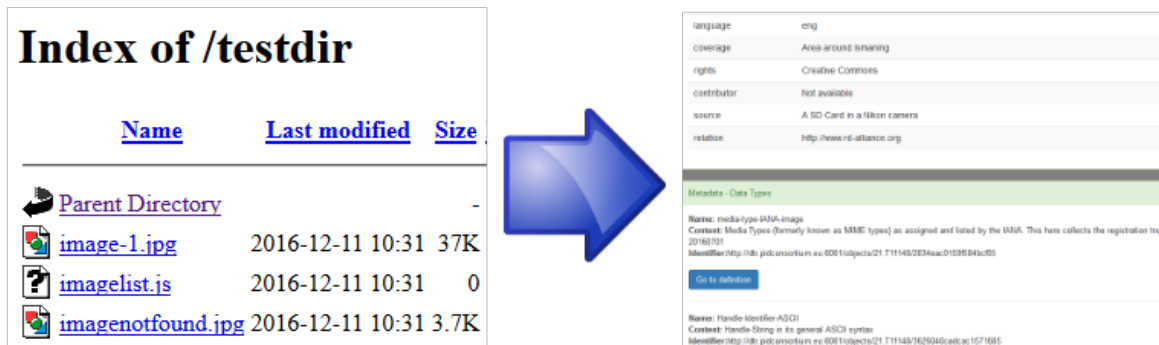
1. What it is	3
2. How it works	3
3. Installation.....	4
4. Configuration	4
5. The Admin Interface	5
6. Functionality	5
6.1. Persistent Identifiers (PIDs)	5
6.1. Metadata	5
6.1.1. Plain Text.....	5
6.1.2. Dublin Core	5
6.1.3. Types.....	6
6.1.4. Taco Properties Files	6
6.2. RDA Collections API.....	6
6.2.1. Creating a new collection	6
6.2.2. Showing all collections.....	6
6.2.3. Adding a new item to a collection	7
6.2.4. Show all items in a collection.....	7
6.2.5. Delete a member of a collection	7
6.2.6. Delete a whole collection	7
6.3. OAI-PMH.....	7
6.4. ResourceSync	8
7. Permissions	8
8. Using Docker	8
9. Cloud Storage Support	8

by Thomas Zastrow, 2017

1. What it is

Reptor is a PHP application which turns a webserver into a data repository. It demonstrates the functionality of a modern *data repository* along the recommendations of the **Research Data Alliance** (RDA)¹. By adding *metadata* and *Persistent Identifiers* to a folder with *Object Data*, Reptor turns simple folders on a webserver into a *Digital Object*.

After being installed, Reptor offers:



- PID handling, independent of the used PID system (DOI, Handle, URN etc.). The *shortref.org* service is preconfigured and can be used to assign Handle based PIDs to Reptor's Digital Objects
- Untyped metadata handling by supporting several types of serialised metadata (plain text, key-value pairs [Dublin Core etc.], XML formats, etc.)
- Integration of Taco Property files
- Integration of Data Type Registries (DTR)
- Support for virtual collections (experimental)
- Access to the object data (files and folders) stored on the filesystem and taking care of permission management
- OAI-PMH
- ResourceSync (partially)
- Social media integration (Facebook, Twitter)
- Admin interface

All data is stored in one or more filesystems the webserver has access to. No database is necessary. Replication and backup is easy doable by just copying over the root folder of the Reptor installation.

2. How it works

Reptor consists of a bunch of PHP, JavaScript and CSS files. The main work for the users view is done by the file **index.php** which takes some parameters and then displays recursively the content of the folder called **data** as *Digital Objects*. Also external filesystems which are mounted into **data** are exposed. Metadata, PIDs etc. will be rendered automatically as well as dissemination technologies like OAI-PMH and social media buttons will be shown to the user. The *admin*

¹<https://www.rd-alliance.org>

interface can be used to create metadata and PIDs as well as uploading new data to the webserver.

3. Installation

Clone the GitHub repository into an empty folder on a PHP enabled webserver:

```
git clone https://github.com/TomZastrow/reptor.git
```

Beside the basic Reptor files, this will also create some basic example entries in the “data” folder: this folder is holding the whole data, any other (external) filesystem can be mounted on any mountpoint below “data”.

Please note: Reptor does not contain a user management currently. But there is an Admin panel.

Make sure that the “admin” folder is secured by the webserver! If you are using Apache Webserver, you will find here² a tutorial about securing a folder. If you don't need the Admin Panel for manual adding and editing data, you can delete the admin folder.

4. Configuration

Configuration is done in the file “config.php”. Most of these settings are coming along with a sensible default setting but can be adjusted accordingly to individual needs:

```
$config['dataTypeRegistry'] = "http://dtr.pidconsortium.eu:8081/objects/";
```

The URL to a data type registry: per default, GWDGs DTR is used

```
$config['namePIDFile'] = "pid.txt";
```

The name of the file which contains the PID of the current directory

```
$config['nameMetadataText'] = "metadata.txt";
```

The name of the file which contents is interpreted and displayed as metadata text

```
$config['namesDataTypes'] = "metadata.types";
```

The name of the file which contains the data types, related to the data type registry mentioned above, to describe the content of the current directory

```
$config['nameDCFile'] = "metadata.dc";
```

This file contains Dublin Core metadata in Key-Value format

```
$config['nameTacoProperties'] = "metadata.properties";
```

```
$config['nameCollectionItems'] = "collection.txt";
```

```
$config['template'] = "default";
```

The CSS- and HTML template which is used to render the pages. Can be found in subfolder “templates”.

```
$config['showEvaluation']=true;
```

Defines if an evaluation against the RDA specifications is shown (true) or not (false) at the bottom of the page

```
$config['showAdminLogin']=true;
```

```
$config['shortrefUrl'] = 'https://lindat.mff.cuni.cz/services/shortener/api/v1/handles';
```

```
$config['showCollectionItemsAsLinks'] = true;
```

² <https://www.digitalocean.com/community/tutorials/how-to-use-the-htaccess-file>

```
$config['showMissingParts']=false;  
$config['repositoryTitle'] = "RDA Data Repository";  
$config['repositoryEmail'] = "thomas.zastrow@mpcdf.mpg.de";  
$config['repositoryIdentifier'] = "shortref.org";  
$config['showFacebookLikeButton'] = true;  
$config['showTwitterButton'] = true;
```

5. The Admin Interface

If you have not deleted the „admin“ folder, the admin interface is accessible via the subfolder „/admin“ of the Reptor installation. In the configuration file, it is also possible to activate a link to the admin interface in the header line.

Here, all three parts of a Digital Object can be created and edited:

- PID Management
- Metadata
- Object Data

6. Functionality

6.1.Persistent Identifiers (PIDs)

Reptor is currently not able to manage PIDs on its own. You need an external PID provider, which can be a Handle based service (for example DOI) or any other PID system like ARK³. The PID should point to:

```
http://yourserver.com/path/to/reptor/subfolder/subfolder
```

where „subfolder/subfolder“ is the relative path to the folder you want to assign a PID to. Please note that the folder „data“ as the starting folder of the data repository is not part of the path. Also „index.php“ is not part because we assume that your server is configured in such a way that „index.php“ is delivered by default if no other filename is mentioned.

6.1.Metadata

Metadata is stored in files in any (sub-) directory of the „data“ dir. Depending on the filename, the metadata will be handled differently.

6.1.1.Plain Text

Its content will be displayed as it is in the „Metadata“ box of that directory. Any HTML, XML or whatever content is possible.

6.1.2.Dublin Core

Key-value pairs as described in Dublin Core. Will be included in OAI-PMH output!

³A good starting point is the free-of-cost service Shortref (www.shortref.org)

6.1.3.Types

semantic explanations of the content of the current directory. Makes use of the „Data Type Registry“ at GWDG

6.1.4.Taco Properties Files

The Reptor Repository supports the rendering of Property Files created by the Taco System⁴.

6.2. RDA Collections API

Reptor supports partially the RDA Collection API. A collection is represented by any kind of items (links, strings, numbers, PIDs, ...) in a file with a defined name in a folder (configuration variable “nameCollectionItems”, per default set to “collections.txt”). The path to the folder represents the name of the collection.

The admin interface allows manual editing a collection. In addition, the RDA Collections API is implemented in the file “collections/api.php” and can be used to manipulate collections via RESTStyle web services.

The starting point for any web service call is the full path to “collections/api.php” plus the stop word “collections”. Any other paths behind it will be interpreted as path to a collection (path to a subfolder in “data”) and / or a call to its members.

The RDA collection API implements write operations. As the admin interface, these should be secured. The Reptor software doesn’t implement any user management at its own, so its a good idea to secure the write operations of the collections API via .htaccess directives. The following code in a .htaccess file limits the write access to a valid user, while keeping the GET operations free:

```
AuthType Basic
AuthName "Please log in"
AuthUserFile .htpasswd

<Limit POST PUT DELETE>
Require valid-user
</Limit>
```

6.2.1.Creating a new collection

The path to the new to created collection has to be specified, HTTP method is POST. The following call will create a new collection if it not already exists: if both folder and collection file does not exists, Reptor will create both. If the folder but not the file exists, only the latter will be created. If both are already in place, an error is thrown.

The following curl command creates a new collection in the folder “demoCollection”:

```
curl -X POST http://localhost:8000/collections/api.php/collections/demoCollection
```

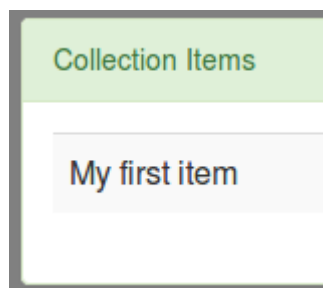
After executing this command, a new folder should appear on Reptors frontpage, containing

⁴ <http://www.mpcdf.mpg.de/services/data/annotate>

nothing so far.

6.2.2. Showing all collections

The following GET command shows all collections which are available in the Reptor instance:



```
curl -X GET http://localhost:8000/collections/api.php/collections
```

6.2.3. Adding a new item to a collection

Members can be added by executing a a POST call to the “members” section of a collection, specifying the new member (here “My first item”) in the body of the call as Json construct:

```
curl -X POST -d '{"id" : "My first item"}'  
http://localhost:8000/collections/api.php/collections/demoCollection/members
```

6.2.4. Show all items in a collection

```
curl -X GET http://localhost:8000/collections/api.php/collections/demo/members
```

6.2.5. Delete a member of a collection

```
curl -u reptor:rda -X DELETE  
http://localhost:8000/collections/api.php/collections/demo/members/One%20more%20item
```

6.2.6. Delete a whole collection

```
curl -u reptor:rda -X DELETE http://localhost:8000/collections/api.php/collections/demo
```

6.3. OAI-PMH

The “Open Archive Initiative - Protocol for Metadata Harvesting” (OAI-PMH) is implemented in Reptor and can be used to offer the metadata for other repositories, catalogue systems etc. for harvesting. The Reptor implementation supports the following OAI-PMH verbs:

- Identify
- ListMetadataFormats

- ListSets
- ListIdentifiers
- ListRecords

The OAI-PMH implementation is complete, it validates against the OAI-PMH validator at oaipmh.com⁵. Currently, it only offers “Dublin Core” metadata for harvesting.

6.4.ResourceSync

ResourceSync⁶ is a specification for synchronization of resources between servers. Its implementation in Reptor is rudimentary.

7. Permissions

Permissions can be handled via the webserver. In the case of Apache, .htaccess can be used: just put a correct configured file in a subdirectory of “data” and the Apache webserver will take care of the correct handling of its content.

Following RDAs recommendation: the metadata and the PID in this directory will be displayed always! Only accessing the content itself follows under the rules defined in .htaccess.

8. Using Docker

The Reptor software needs nothing else than a PHP enabled webserver to run. On DockerHub, you will find a bunch of images which could be used. Here is one way:

- Clone the Reptor software as mentioned in the chapter before from GitHub to your local machine
- Download the “eboraas/apache-php” Docker image:

```
docker pull eboraas/apache-php
```
- Run the image and mount your local clone of the GitHub repository:

```
docker run -p 9999:80 -v /path/to/git-clone/:/var/www/html -d eboraas/apache-php
```

Here, “9999” is the port on your host machine where the Reptor software is being made available. and “/path/to/git-clone” points to the directory where you made a clone of the Reptor GitHub repository. Now point your browser to:

<http://localhost:9999>

9. Cloud Storage Support

Reptor's data just has to be stored somewhere below its „data“ directory. Here, any other (remote) filesystem can be locally mounted. Reptor will then handle these filesystems just as local (sub-) directories. Make sure that the webserver has the file permissions to access the mounted filesystems.

⁵ <http://validator.oaipmh.com/>

⁶ <http://www.openarchives.org/rs/1.0.9/resourcesync>

Reptor Manual

Via this mechanism, a user can add / edit data easily on a remote filesystem while any change is instantaneously visible in Reptor.