File Repository Documentation

Release 2

Wolnosciowiec Team

Contents:

1	First steps	3				
2	Manual installation	5				
3	Installation with docker					
4	Development environment setup	11				
5	Post-installation	13				
6	Configuration reference6.1Application configuration6.2Permissions list6.3Docker container extra parameters6.4PostgreSQL support	17 17 21 26 26				
7	Docker, releases and versioning	27				
8	Using postman to manage the application	29				
9	Authorization 9.1 Creating a token 9.2 Searching tokens 9.3 Looking up a token 9.4 Revoking a token 9.5 Managing authentication using console commands 9.6 Token load hierarchy	33 33 35 37 38 39 41				
10	Files storage 10.1 Security	43 44 45 46 47 49				
11	Backup 11.1 Getting started	51 51				

17	From	a authors	89
16	Gene	ral guide for Administrators, DevOps and Developers	87
	15.2	Docker container concept	83
	15.1	Introduction	
15		access	83
	14.7	Without docker	81
	14.6	Using docker container	80
	14.5	Setup	80
	14.4	Slack/Mattermost notifications	80
	14.3		79
	14.2	· ·	77
- •	14.1		71
14	Bahu	h API client	71
	13.1	SecureCopy API endpoints	70
13		reCopy	69
	12.2	Endpoints	01
		Quick start in steps	61 61
12		mumUI	61
	11.0	Managing collections from shell	59
	11.5	Data replication	58
	11.4		55
	11.3		
		Managing collections	
	44.0		

File Repository is a modern API application dedicated for storing files. It is able to use various storage backends including AWS S3, Dropbox, Google Drive and just filesystem. Lightweight, requires just PHP7 and at least SQLite3 or MySQL (other databases can be also supported in future due to using ORM).

Main functionality:

- Strict access control, you can generate a token that will have access to specific actions on specific items
- Store files where you need; on AWS S3, Minio.io, FTP, local storage and others...
- Deduplication for non-grouped files. There will be no duplicated files stored on your disk
- Backups management, you can define a collection of file versions that can rotate on adding a new version
- API + lightweight frontend
- Ready to integrate upload forms for your applications. Only generate token and redirect a user to an url

Contents: 1

2 Contents:

First steps

To start using the application you need to install PHP 7.4 with extensions listed below. The dependencies are managed by Composer - it will also validate your environment for required extensions and PHP version.

You can also use a ready-to-use docker container instead of using host installation of PHP, if you have a possibility always use a docker container.

Application requirements:

- PHP 7.4 or newer (with extensions: ctype, fileinfo, curl, json, openssl, pdo, pdo_mysql, pdo_pgsql, iconv)
- NodeJS 12.x + NPM (for building simple frontend at installation time)
- SQLite3, MySQL 5.7+ or PostgreSQL 10+
- Composer (PHP package manager, see packagist.org)
- make (GNU Make, we use it for build scripts)

 $Notice: For\ PostgreSQL\ configuration\ please\ check\ the\ configuration\ reference\ at\ :ref: `postgresql_support'\ page$

Manual installation

At first you need to create your own customized .env file with application configuration. You can create it from a template .env.dist.

Make sure the **APP_ENV** is set to **prod**, and that the database connection settings are valid. On default settings the application should be connecting to a SQLite3 database placed in local file, but please keep in mind, that this is not optimal for production usage.

```
cd server
cp .env.dist .env
edit .env
```

To install the application - download dependencies, install database schema use the make task install.

```
make install_frontend
```

All right! The application should be ready to go.

Now set up an NGINX + PHP-FPM or Apache to redirect all traffic to point at /public/index.php

For more help please visit: https://symfony.com/doc/current/setup/web_server_configuration.html

When you have the web server up and running, you can check the health check endpoint.

```
# "test" is defined in "HEALTH_CHECK_CODE" environment variable
curl http://localhost/health?code=test
```

Installation with docker

There are at least four choices:

- Use quay.io/riotkit/file-repository image by your own and follow the configuration reference
- Generate a docker-compose.yaml using *make print VARIANT="gateway s3 postgres postgres-persistent"* in environment basing on it
- Copy the env environment from this repository and adjust to your needs
- Take a look at our compose in env directory and at configuration reference, then create a Kubernetes or other type deployment

Proposed way to choose is the prepared docker-compose environment that is placed in env directory.

Preparing configuration of the environment:

Before you will run the environment we suggest to take a look at few configuration variables possibly most important for you at the beginning.

- SECURITY_ADMIN_TOKEN: Will create an admin token for you automatically during container startup
- BASE URL: Application URL (in web browser)

```
edit ./env/.env
```

```
# For production usage please pick a fixed version
# List of available tags: https://quay.io/repository/riotkit/file-repository?tab=tags
FILE_REPOSITORY_VERSION=latest-build
BASE_URL=http://localhost:8000

# # Docker Environment settings
# COMPOSE_ENV_NAME=fr_tests

# This token will be added during File Repository container startup
```

```
# please CHANGE IT BEFORE STARTING an application on production
# See: https://www.uuidgenerator.net/
# See: Linux command "uuidgen" from e2fsprogs package
SECURITY_ADMIN_TOKEN=ca6a2635-d2cb-4682-ba81-3879dd0e8a77
# Code required to access health check endpoint
# http://localhost/health?code=test
HEALTH_CHECK_CODE=test
# PostgreSQL settings (when using PostgreSQL)
POSTGRES_USER=riotkit
POSTGRES_DB=rojava
POSTGRES_PASSWORD=rojava
# MySQL/MariaDB settings (when using MySQL/MariaDB)
MYSQL_DB=rojava
MYSQL_PASSWORD=rojava
MYSQL_USER=riotkit
MYSQL_ROOT_PASSWORD=solidarity-forever
# Minio Settings
MINIO_DOMAIN=s3-minio-test
MINIO_ACCESS_KEY=international_workers_association
MINIO_SECRET_KEY=six_hour_workingday_for_everybody
# Bahub client settings
BAHUB_VERSION=dev
# Backup collections
# To generate a collection type:
   ./bin/console backup:create-collection -b 5 -o 10mib -c 2gib -s delete_oldest_
→when_adding_new -f test.gz
COLLECTIONS_POSTGRES_ID=946348f2-8f3c-4cf0-8827-650fb044ed39
COLLECTIONS_POSTGRES_SINGLE_DB_ID=74803fcf-6661-41b4-b063-58e8f614661a
COLLECTIONS_POSTGRES_BASE_ID=38358f25-b63e-4d70-bce2-71cda2258f70
COLLECTIONS_WWW_FILES_ID=7682aba1-38f5-4f20-a7c1-d939b3a5b928
COLLECTIONS_MYSQL_ID=43472cc8-a44b-489e-bb43-a867669a5a2c
```

Starting the example environment:

```
cd ./env
make up VARIANT="gateway s3 postgres persistent"
```

Generating a docker-compose example file:

```
cd ./env
make print VARIANT="gateway s3 postgres persistent"
```

Production tips:

- Use external non-containerized database, do backups. If you want to use containers then use replication
- Do not use SQLite3 for production. Use PostgreSQL or MySQL instead
- Mount data as volumes. Use bind-mounts to have files placed on host filesystem (volumes can be deleted, bind-mounted files stays anyway)
- Use SECURITY_ADMIN_TOKEN environment variable to setup an administrative token to be able to log-in into the application
- For automation, use *POST_INSTALL_CMD* to execute console commands to create collections and tokens with ids your applications expects

Development environment setup

For development purposes use the "dev" configuration, which mounts the application into the docker container, in effect all changes are present in the application immediately without a rebuild.

You can also run the application with PostgreSQL and/or with S3 as a storage.

```
cd env
make up VARIANT="test"
# with PostgreSQL as a database
make up VARIANT="dev test postgres"
# bind application on port 80
make up VARIANT="dev test postgres gateway"
# keep all of the changes between environment restarts
make up VARIANT="dev test postgres postgres-persistent gateway"
# to have a good, production type configuration
make up VARIANT="s3 postgres postgres-persistent gateway"
# to have a production type configuration, that can be behind reverse proxy (do not_
→expose ports itself to host)
make up VARIANT="s3 postgres postgres-persistent"
# to have server + Bahub client container and it's test containers
make up VARIANT="dev test postgres bahub-test"
make sh@bahub # here you can perform test backups upload/restore
```

Please check out the detailed instruction in the README file.

Post-installation

At this point you have the application, but you don't have access to it - except if you use docker container and specify the *SECURITY_ADMIN_TOKEN*, then docker container would create an admin token for you. **You will need to generate an administrative access token (if you dont have one already)**, to be able to create new tokens, manage backups, upload files to storage. To achieve this goal you need to execute a simple command.

You need to execute ./bin/console auth:generate-admin-token in the project directory.

So, when you have an administrative token, then you need a second token to upload backups. It's not recommended to use administrative token on your servers. **Recommended way is to generate a separate token, that is allowed to upload a backup to specified collection**

To do so, check all available roles in the application:

```
GET /auth/roles?_token=YOUR-ADMIN-TOKEN-HERE
```

Note: If you DO NOT KNOW HOW to perform a request, then please check the postman section

You should see something like this:

```
"upload.images": "Allows to upload images",
       "upload.videos": "Allows to upload video files",
       "upload.documents": "Allows to upload documents",
       "upload.backup": "Allows to submit backups",
       "upload.all": "Allows to upload ALL types of files regardless of mime type",
       "upload.enforce_no_password": "Enforce no password for all uploads for this_
⇔token",
        "upload.enforce_tags_selected_in_token": "Enforce token tags. In result every...
→uploaded file will have tags specified in token regardless if they were sent in_
⇒request",
       "upload.only_once_successful": "",
       "security.authentication_lookup": "User can check information about ANY token
       "security.search_for_tokens": "User can browse\/search for tokens",
       "security.overwrite": "User can overwrite files",
       "security.generate_tokens": "User can generate tokens with ANY roles",
       "security.use_technical_endpoints": "User can use technical endpoints to_
→manage the application",
       "security.revoke_tokens": "User can expire other token, so it will be not,
→valid anymore",
       "security.administrator": "Special: Marking - tokens with this marking will...
→not be able to be revoked by non-administrators",
       "security.create_predictable_token_ids": "Allow to specify token id when_
→creating a token",
       "deletion.all_files_including_protected_and_unprotected": "Delete files that...
→do not have a password, and password protected without a password",
       "view.any_file": "Allows to download ANY file, even if a file is password,
→protected",
       "view.files_from_all_tags": "List files from ANY tag that was requested, else...
→the user can list only files by tags allowed in token",
       "view.can_use_listing_endpoint_at_all": "Define that the user can use the,
→listing endpoint (basic usage)",
       "securecopy.stream": "Can use SecureCopy at all?",
       "securecopy.all_secrets_read": "Read SecureCopy secrets: Encryption method,_
→password, initialization vector. With following role can read secrets of any token.
⇒in the system.",
       "collections.create_new": "Allow person creating a new backup collection",
       "collections.create_new.with_custom_id": "Allow to assign a specific id, when,
⇔creating a collection",
       "collections.allow_infinite_limits": "Allow creating backup collections that...
→have no limits on size and length",
       "collections.modify_details_of_allowed_collections": "Edit collections where,
→token is added as allowed",
       "collections.modify_any_collection_regardless_if_token_was_allowed_by_
→collection": "Allow to modify ALL collections. Collection don't have to allow such,
→token which has this role",
       "collections.view_all_collections": "Allow to browse any collection.
→regardless of if the user token was allowed by it or not",
       "collections.can_use_listing_endpoint": "Can use an endpoint that will allow_
→to browse and search collections?",
       "collections.manage_tokens_in_allowed_collections": "Manage tokens in the...
→collections where our current token is already added as allowed",
       "collections.delete_allowed_collections": "Delete collections where token is,
→added as allowed",
       "collections.upload to_allowed_collections": "Upload to allowed collections",
       "collections.list_versions_for_allowed_collections": "List versions for...
⇒collections where the token was added as allowed",
```

```
"collections.delete_versions_for_allowed_collections": "Delete versions only_ 

from collections where the token was added as allowed"
}
}
```

To allow only uploading and browsing versions for assigned collections you may choose:

```
POST /auth/token/generate?_token=YOUR-ADMIN-TOKEN-THERE
{
    "roles": ["upload.backup", "collections.upload_to_allowed_collections",
    "collections.list_versions_for_allowed_collections"],
    "data": {
        "tags": [],
        "allowedMimeTypes": [],
        "maxAllowedFileSize": 0
    }
}
```

As the response you should get the token id that you need.

Remember the tokenId, now you can create collections and grant access for this token to your collections. Generated token will be able to upload to collections you allow it to.

Check next steps:

- 1. Collection creation
- 2. Assigning a token to the collection

That's all.

Configuration reference

6.1 Application configuration

When setting up application without a docker a .env file needs to be created in the root directory of the application. The .env.dist is a template with example, reference values. If you use a docker image, then you may use those variables as environment variables for the container.

```
# This file is a "template" of which env vars need to be defined for your application
# Copy this file to .env file for development, create environment variables when,
→deploying to production
# https://symfony.com/doc/current/best_practices/configuration.html#infrastructure-
\rightarrowrelated-configuration
# -----
# Symfony framework specific configuration
# -----
# possible values: prod, test, dev. Use "prod" for public instances. Use "test"...
→installing the application and changing its configuration
#APP_ENV=prod
#APP_SECRET=faec5e5fcf0ff499f53ecc30bdffc653
#TRUSTED_PROXIES=127.0.0.1,127.0.0.2
#TRUSTED_HOSTS=localhost, example.com
# -----
# Technical
# application domain (if empty, then it will be taken from Host header. If Host,
→header is invalid, then fill this variable)
#APP_DOMAIN=
# how long can take request if it is expected, that it will take long (eg. file_
<del>→upload)</del>
                                                                      (continues on next page)
```

```
# please consider there if you have request buffering enabled in webserver, nginx has.
→this by default
#LONG_EXECUTION_TIME=300
#TEMP_DIRECTORY=/tmp
#HTTP_TIMEOUT=30
#READ_ONLY=false
# Health check, will expose an endpoint http://your-app/health?code=some-code-there
# use it to connect your application into the monitoring. It will report storage, _
\rightarrow database health.
#HEALTH_CHECK_CODE=
# -----
# Token settings
# -----
# default expiration time for generated tokens
#TOKEN_EXPIRATION_TIME="+30 minutes"
# used for tokens encryption (tokens can be generated by external applications)
#ENC_TOKEN_PHRASE=test123
# Checksum generation salt (important for the security, needs to be not guessable)
#ENC SALT=some-salt
# -----
# Application connection details
# -----
# SQLite3 example
#DATABASE_PATH=./var/data.db
#DATABASE_DRIVER=pdo_sqlite
# MySQL example
#DATABASE_HOST=db_mysql
#DATABASE_USER=riotkit
#DATABASE_PASSWORD=riotkit
#DATABASE_NAME=file_repository
#DATABASE_VERSION=5.7
#DATABASE DRIVER=pdo mysql
#DATABASE PORT=3306
# PostgreSQL example
#DATABASE_USER=riotkit
#DATABASE PASSWORD=riotkit
#DATABASE_NAME=file_repository
#DATABASE_DRIVER=pdo_pgsql
#DATABASE_HOST=/var/run/postgresql
#DATABASE_HOST=192.168.1.161
#DATABASE_PORT=5432
# -----
# Backups feature
# -----
#BACKUP ONE VERSION MAX SIZE=4GB
```

```
#BACKUP_COLLECTION_MAX_SIZE=15GB
#BACKUP_MAX_VERSIONS=5
# Anti-Hotlink protection
# (NOTICE: Requires application to be restarted or ./bin/console cache:clear to be,
⇔executed)
#ANTI_HOTLINK_PROTECTION_ENABLED=true
#ANTI_HOTLINK_RESTRICT_REGULAR_URLS=false
#ANTI_HOTLINK_URL=/stream/{accessToken}/{expirationTime}/{fileId}
#ANTI HOTLINK CRYPTO=md5
#ANTI_HOTLINK_SECRET_METHOD="\$http_x_expiration_time\$filename\$http_remote_addr MY-
→ AWESOME-SUFFIX"
# -----
# Storage
# Storage type
#
   local: local filesystem
#
  awss3: AWS S3, Minio.io or other compatible with AWS S3 interface
   ftp: FTP/FTPS
#
#FS_ADAPTER=local
# === Local filesystem ===
# path, where for local filesystem the files should be stored. %kernel.root_dir% is_

→ the application main directory

#FS_LOCAL_DIRECTORY="%kernel.root_dir%/../var/uploads"
# (?) see: http://flysystem.thephpleague.com/docs/adapter/local/
#FS LOCAL LAZY=0
# file locking; 1 - LOCK_SH, 2 - LOCK_EX, 3 - LOCK_UN, 4 - LOCK_NB
#FS_LOCAL_WRITEFLAGS=
# how to handle symbolic links: 1 - skip links, 2 - disallow links
#FS LOCAL LINKHANDLING=
# permissions chmod eg. 0755
#FS_LOCAL_PERMISSIONS=
# === S3 ===
# set to enable S3 interface as adapter (supports Minio, and possibly Amazon AWS S3)
```

```
#FS_ADAPTER=awss3v3
# (advanced) Symfony service name of an S3 Client
#FS_AWSS3V3_CLIENT=s3_client
#FS_AWSS3V3_BUCKET=misc
#FS_AWSS3V3_PREFIX=
#FS_AWSS3V3_OPTIONS_ENDPOINT=http://localhost:9000
#FS_AWSS3V3_VERSION=latest
#FS_AWSS3V3_REGION=eu-central-1
#FS_AWSS3V3_KEY=some-key
#FS_AWSS3V3_SECRET=some-secret
# === FTP ===
# there you can enable FTP adapter
#FS_ADAPTER=ftp
# FTP host where to connect to, IP address or hostname
#FS_FTP_HOST=localhost
# FTP port to connect to
#FS_FTP_PORT=21
# FTP authorization: username
#FS_FTP_USERNAME=user
# FTP authorization: user's password
#FS_FTP_PASSWORD=password
# FTP directory where files will be stored in
#FS_FTP_ROOT=/
# Use SSL? (FTPS instead of FTP)
#FS_FTP_SSL=1
# Timeout
#FS_FTP_TIMEOUT=120
# Permissions for owner eg. 0755
#FS FTP PERMPRIVATE=
# Permissions for public (others) eg. 0400
\#FS\_FTP\_PERMPUBLIC=
# Passive mode?
#FS_FTP_PASSIVE=0
# -----
# SecureCopy
# -----
#REPLICATION_MODE=primary
```

6.2 Permissions list

You can get a permissions list by accessing an endpoint in your application:

```
GET /auth/roles?_token=test-token-full-permissions
```

There is also an always up-to-date permissions list, taken directly from the current version of the application for which the documentation you are browsing.

How to read the list by example:

```
/** Allows to upload images */
public const ROLE_UPLOAD_IMAGES = 'upload.images';
```

Legend:

- Between /** and */ is the description
- upload.images is the role name that you need to know

```
<?php declare(strict_types=1);</pre>
namespace App\Domain;
* List of roles which could be required for a temporary token
    There are 2 types of roles:
       - GRANT
        - RESTRICTION
   They are distinguished because, we want to GRANT all rights to administrator for
→example, excluding RESTRICTION
   roles.
* @codeCoverageIgnore
*/
final class Roles
   public const TEST_TOKEN
                                      = 'test-token-full-permissions';
   public const INTERNAL_CONSOLE_TOKEN = 'internal-console-token';
   // upload/creation
   11
   11
   /** Allows to upload images */
   public const ROLE_UPLOAD_IMAGES
                                             = 'upload.images';
   /** Allows to upload video files */
   public const ROLE_UPLOAD_VIDEOS
                                               = 'upload.videos';
   /** Allows to upload documents */
   public const ROLE_UPLOAD_DOCS
                                              = 'upload.documents';
    /** Allows to submit backups */
```

(continues on next page)

6.2. Permissions list 21

```
public const ROLE_UPLOAD_BACKUP
                                             = 'upload.backup';
   /** Allows to upload ALL types of files regardless of mime type */
   public const ROLE_UPLOAD
                                            = 'upload.all';
   /** Enforce no password for all uploads for this token */
   public const ROLE_UPLOAD_ENFORCE_NO_PASSWORD = 'upload.enforce_no_password';
   /** Enforce token tags. In result every uploaded file will have tags specified in.
→token regardless if they were sent in request */
   public const ROLE_UPLOAD_ENFORCE_TOKEN_TAGS = 'upload.enforce_tags_selected_in_
→token';
   public const ROLE_UPLOAD_ONLY_ONCE_SUCCESSFUL = 'upload.only_once_successful';
   //
   //
   // authentication and tokens
   //
   /** User can check information about ANY token */
   public const ROLE_LOOKUP_TOKENS = 'security.authentication_lookup';
   /** User can browse/search for tokens */
   public const ROLE_SEARCH_FOR_TOKENS
                                               = 'security.search_for_tokens';
   /** User can overwrite files */
   public const ROLE ALLOW OVERWRITE FILES
                                               = 'security.overwrite';
   /** User can generate tokens with ANY roles */
   public const ROLE GENERATE TOKENS
                                                = 'security.generate_tokens';
   /** User can use technical endpoints to manage the application */
   public const ROLE_USE_TECHNICAL_ENDPOINTS = 'security.use_technical_endpoints
   /** User can expire other token, so it will be not valid anymore */
   public const ROLE REVOKE TOKENS = 'security.revoke tokens';
  /** Special: Marking - tokens with this marking will not be able to be revoked by...
→non-administrators */
   public const ROLE_ADMINISTRATOR
                                               = 'security.administrator';
   /** Allow to specify token id when creating a token */
   public const ROLE_CREATE_PREDICTABLE_TOKEN_IDS = 'security.create_predictable_
→token_ids';
   //
   // deletion
   //
   //
  /** Delete files that do not have a password, and password protected without a...
→password */
  public const ROLE_DELETE_ALL_FILES = 'deletion.all_files_including_
                                                                      (continues on next page)
→protected_and_unprotected';
```

```
11
   11
   // browsing
   //
   /** Allows to download ANY file, even if a file is password protected*/
   public const ROLE_VIEW_ALL_PROTECTED_FILES = 'view.anv file';
   /** List files from ANY tag that was requested, else the user can list only files.
→by tags allowed in token */
   public const ROLE BROWSE LIST OF FILES BY ANY TAG = 'view.files from all tags';
   /** Define that the user can use the listing endpoint (basic usage) */
   public const ROLE ACCESS LISTING ENDPOINT = 'view.can use listing endpoint_at_all
';
   11
   // SecureCopy domain
   //
   11
   /** Can use SecureCopy at all? */
   public const ROLE_SECURE_COPY_READ_DATA_STREAM = 'securecopy.stream';
   /** Read SecureCopy secrets: Encryption method, password, initialization vector...
→With following role can read secrets of any token in the system. */
   public const ROLE_READ_SECURE_COPY_SECRETS = 'securecopy.all_secrets_read';
   // collections
   //
   11
   /** Allow person creating a new backup collection */
   public const ROLE_COLLECTION_ADD = 'collections.create_new';
   /** Allow to assign a specific id, when creating a collection */
   public const ROLE_COLLECTION_CUSTOM_ID = 'collections.create_new.with_custom_id';
   /** Allow creating backup collections that have no limits on size and length */
   public const ROLE COLLECTION ADD WITH INFINITE LIMITS = 'collections.allow_
→infinite_limits';
   /** Edit collections where token is added as allowed */
   public const ROLE_MODIFY_ALLOWED_COLLECTIONS = 'collections.modify_details_of_
→allowed_collections';
   /** Allow to modify ALL collections. Collection don't have to allow such token.
→which has this role */
   public const ROLE_COLLECTION_MODIFY_ANY_COLLECTION = 'collections.modify_any_
→collection_regardless_if_token_was_allowed_by_collection';
   /** Allow to browse any collection regardless of if the user token was allowed by
                                                                        (continues on next page)
→it or not */
```

6.2. Permissions list 23

```
public const ROLE COLLECTION VIEW ANY COLLECTION = 'collections.view all
⇔collections';
   /** Can use an endpoint that will allow to browse and search collections? */
   public const ROLE_CAN_USE_LISTING_COLLECTION_ENDPOINT = 'collections.can_use_
→listing_endpoint';
   /** Manage tokens in the collections where our current token is already added as.
⇒allowed */
   public const ROLE_CAN_MANAGE_TOKENS_IN_ALLOWED_COLLECTIONS = 'collections.manage_
→tokens_in_allowed_collections';
   /** Delete collections where token is added as allowed */
   public const ROLE CAN DELETE ALLOWED_COLLECTIONS = 'collections.delete_allowed_
⇔collections';
   /** Upload to allowed collections */
   public const ROLE_CAN_UPLOAD_TO_ALLOWED_COLLECTIONS = 'collections.upload_to_
→allowed_collections';
   /** List versions for collections where the token was added as allowed */
   public const ROLE_LIST_VERSIONS_FOR_ALLOWED_COLLECTIONS = 'collections.list_
→versions_for_allowed_collections';
   /** Delete versions only from collections where the token was added as allowed */
   public const ROLE_DELETE_VERSIONS_IN_ALLOWED_COLLECTIONS = 'collections.delete_
→versions_for_allowed_collections';
   /** Collection manager: Create, edit, delete collections */
   public const GROUP_COLLECTION_MANAGER = [
       self::ROLE_COLLECTION_ADD,
       self::ROLE_COLLECTION_ADD_WITH_INFINITE_LIMITS,
       self::ROLE_COLLECTION_MODIFY_ANY_COLLECTION,
       self::ROLE_COLLECTION_VIEW_ANY_COLLECTION,
       self::ROLE_CAN_USE_LISTING_COLLECTION_ENDPOINT,
       self::ROLE_CAN_MANAGE_TOKENS_IN_ALLOWED_COLLECTIONS,
       self::ROLE_MODIFY_ALLOWED_COLLECTIONS,
       self::ROLE_CAN_UPLOAD_TO_ALLOWED_COLLECTIONS,
       self::ROLE_LIST_VERSIONS_FOR_ALLOWED_COLLECTIONS,
       self::ROLE DELETE VERSIONS IN ALLOWED COLLECTIONS
   ];
   public const GRANTS_LIST = [
       self::ROLE_UPLOAD_IMAGES,
       self::ROLE_UPLOAD_DOCS,
       self::ROLE_UPLOAD_BACKUP,
       self::ROLE_UPLOAD,
       self::ROLE_LOOKUP_TOKENS,
       self::ROLE_SEARCH_FOR_TOKENS,
       self::ROLE_ALLOW_OVERWRITE_FILES,
       self::ROLE_GENERATE_TOKENS,
       self::ROLE USE TECHNICAL ENDPOINTS,
       self::ROLE_DELETE_ALL_FILES,
       self::ROLE VIEW ALL PROTECTED FILES,
       self::ROLE BROWSE LIST OF FILES BY ANY TAG,
       self::ROLE_ACCESS_LISTING_ENDPOINT,
       self::ROLE_REVOKE_TOKENS,
```

(continues on next page)

24

```
self::ROLE_ADMINISTRATOR,
    // collections
    self::ROLE_COLLECTION_ADD,
    self::ROLE_COLLECTION_CUSTOM_ID,
    self::ROLE_COLLECTION_ADD_WITH_INFINITE_LIMITS,
    self::ROLE_CAN_DELETE_ALLOWED_COLLECTIONS,
    self::ROLE_COLLECTION_MODIFY_ANY_COLLECTION,
    self::ROLE_MODIFY_ALLOWED_COLLECTIONS,
    self::ROLE_COLLECTION_VIEW_ANY_COLLECTION,
    self::ROLE_CAN_USE_LISTING_COLLECTION_ENDPOINT,
    self::ROLE_CAN_MANAGE_TOKENS_IN_ALLOWED_COLLECTIONS,
    self::ROLE_CAN_UPLOAD_TO_ALLOWED_COLLECTIONS,
    self::ROLE LIST VERSIONS FOR ALLOWED COLLECTIONS,
    self::ROLE_DELETE_VERSIONS_IN_ALLOWED_COLLECTIONS,
    // securecopy
    self::ROLE_SECURE_COPY_READ_DATA_STREAM,
    self::ROLE_READ_SECURE_COPY_SECRETS
];
public const RESTRICTIONS_LIST = [
    self::ROLE_UPLOAD_ENFORCE_NO_PASSWORD,
    self::ROLE_UPLOAD_ENFORCE_TOKEN_TAGS,
    self::ROLE_UPLOAD_ONLY_ONCE_SUCCESSFUL
];
public static function getRolesList(): array
    return \array_merge(self::GRANTS_LIST, self::RESTRICTIONS_LIST);
 * The test token is available only in APP_ENV=test
 * @param string|null $tokenId
 * @return bool
public static function isTestToken(?string $tokenId): bool
    return $tokenId === static::TEST_TOKEN;
}
 * Internal token is used only in CLI commands
 * Cannot be used within any remote access (eq. via HTTP)
 * @param string|null $tokenId
 * @return bool
*/
public static function isInternalApplicationToken(?string $tokenId): bool
    return $tokenId === static::INTERNAL_CONSOLE_TOKEN;
}
```

6.2. Permissions list 25

6.3 Docker container extra parameters

Parameters passed to docker container are mostly application configuration parameters, but not only. There exists extra parameters that are implemented by the docker container itself, they are listed below:

Name and example	Description
SENTRY_DSN=url-here	(optional) Enables integration with sentry.io, so all failures will be logged
	there
SECU-	(optional) Create admin auth token of given UUIDv4 on container startup
RITY_ADMIN_TOKEN=	

6.4 PostgreSQL support

- 1. Required extensions to enable in PostgreSQL:
- uuid-ossp (CREATE EXTENSION "uuid-ossp";)
- 2. Configuration example:

UNIX Socket example:

```
DATABASE_URL: ""

DATABASE_HOST: "/var/run/postgresql"

DATABASE_NAME: "rojava"

DATABASE_PASSWORD: "rojava"

DATABASE_USER: "riotkit"

DATABASE_DRIVER=pdo_pgsql

DATABASE_CHARSET=UTF8

DATABASE_COLLATE=pl_PL.UTF8

DATABASE_VERSION=10.10
```

IPv4 example:

```
DATABASE_URL: ""
DATABASE_HOST: "192.168.2.161"
DATABASE_NAME: "rojava"
DATABASE_PASSWORD: "rojava"
DATABASE_USER: "riotkit"
DATABASE_USER: "riotkit"
DATABASE_DRIVER=pdo_pgsql

DATABASE_CHARSET=UTF8
DATABASE_COLLATE=pl_PL.UTF8
DATABASE_VERSION=10.10
```

3. "SQLSTATE[21000]: Cardinality violation: 7 ERROR: more than one row returned by a subquery used as an expression"

This is an unresolved issue in the Doctrine DBAL driver that we use. To work around it, please create a separate database, user and use default schema "public" for the application.

Docker, releases and versioning

Images are hosted quay.io

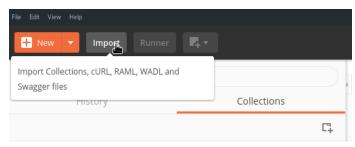
The versions are created from tags, when a code is considered stable, then it is tagged and released as stable. There are also ALPHA and RC (release candidate) versions tagged for testing purposes, often close to the stable release date.

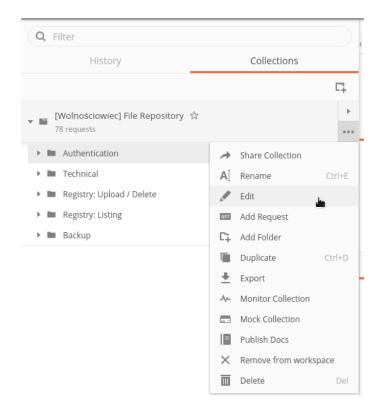
Please see https://semver.org/ for how we version the application.

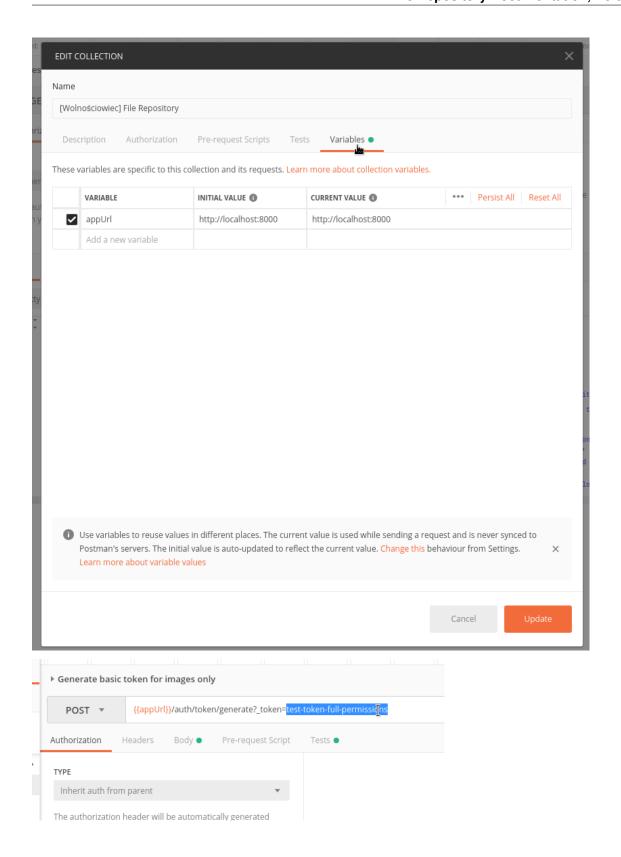
quay.io/riotkit/file-repository:version
quay.io/riotkit/bahub:version
quay.io/riotkit/kropot-cli

Using postman to manage the application

Postman is an API client that allowing to send HTTP requests. You can use it, when you do not have any other graphical application, that could be acting as a client of the File Repository.







ile Repository Documentation, Release 2	
ine ricpository bocumentation, ricicase 2	

CHAPTER 9

Authorization

File Repository is an API application, so there is no user account identified by login and password, there are **ACCESS TOKENS**.

An access token is identified by long UUIDv4, and has assigned information about the access, such as:

- List of actions that are allowed (eg. file uploads could be allowed, but browsing the list of files not)
- Allowed tags that could be used when uploading (optional)
- Allowed file types (mime types) when uploading (optional)
- List of allowed IP addresses that could use this token (optional)
- List of allowed User-Agent strings (optional)
- Maximum allowed file size (optional)
- · Token expiration date

To authorize in the API you need to provide the token in one of those methods: - Using a query parameter "_token" eg. /some/url?_token=123 - Using a HTTP header "X-Auth-Token" - Using an environment variable "FILE REPOSITORY TOKEN"

9.1 Creating a token

Check out the *Permissions list* for a complete list of permissions.

Parameters	
name	description
roles	A list of roles allowed for user. See permissions/configuration reference page
data.tags	List of allowed tags to use in upload endpoints (OPTIONAL)
data.allowedMimeType	esList of allowed mime types (OPTIONAL)
data.maxAllowedFileS	izNumber of bytes of maximum file size (OPTIONAL)
data.allowedUserAgen	tsList of allowed User-Agent header values (ex. to restrict token to single browser) (OP-
	TIONAL)
data.allowedIpAddress	esList of allowed IP addresses (ex. to restrict one-time-token to single person/session) (OP-
	TIONAL)
data.secureCopyEncry	ptEmMention method in SecureCopy mechanism (if using) (OPTIONAL)
data.secureCopyEncry	ot Emockeption key in SecureCopy component. If active, then client using this token will be
	downloading encrypted files (zero-knowledge) (OPTIONAL)
expires	Expiration date, or "auto", "automatic", "never". Empty value means same as "auto"
id	Custom UUIDv4 (requires: security.create_predictable_token_ids role or to be an admin)

```
POST /auth/token/generate?_token=your-admin-token-there

{
    "roles": ["collections.create_new", "collections.manage_tokens_in_allowed_
    -collections"],
    "data": {
        "tags": [],
        "allowedMimeTypes": ["image/jpeg", "image/png", "image/gif"],
        "maxAllowedFileSize": 14579,
        "allowedUserAgents": ["Mozilla/5.0 (X11; Linux x86_64; rv:57.0) Gecko/
    -20100101 Firefox/57.0"],
        "allowedIpAddresses": ["192.168.1.10"]
    },
        "expires": "2020-05-05 08:00:00"
}
```

Example response:

```
"status": true,
   "error_code": null,
   "http_code": 200,
   "errors": [],
   "context": [],
   "message": "Token created",
   "token": {
       "id": "ca6a2635-d2cb-4682-ba81-3879dd0e8a77",
       "active": true,
       "expires": "2020-05-05 08:00:00",
       "expired": false,
        "data": {
            "tags": [],
            "allowedMimeTypes": ["image/jpeg", "image/png", "image/gif"],
            "maxAllowedFileSize": 14579,
            "allowedIpAddresses": ["192.168.1.10"],
            "allowedUserAgents": ["Mozilla/5.0 (X11; Linux x86_64; rv:57.0) Gecko/
\rightarrow20100101 Firefox/57.0"],
            "secureCopyEncryptionKey": "",
            "secureCopyEncryptionMethod": ""
```

```
},
    "roles": [
        "collections.create_new",
        "collections.add_tokens_to_allowed_collections"
]
}
```

Required roles:

• security.generate_tokens

9.2 Searching tokens

Finds tokens matching search criteria.

```
GET /auth/search?_token=your-token-there&q=&limit=50&page=1
```

Example response:

```
"status": true,
"error_code": null,
"http_code": 200,
"errors": [],
"context": {
    "pagination": {
        "page": 1,
        "perPageLimit": 5,
        "maxPages": 7
},
"message": "Matches found",
"data": [
    {
        "id": "1c2c84f2-d488-4ea0-9c88-d25aab139ac4",
        "active": true,
        "data": {
            "tags": [],
            "allowedMimeTypes": [],
            "maxAllowedFileSize": null,
            "allowedIpAddresses": [],
            "allowedUserAgents": [],
            "secureCopyEncryptionKey": "",
            "secureCopyEncryptionMethod": ""
        "roles": [
            "upload.images"
        ]
    },
        "id": "669d4918-b156-412d-9c89-ba01d6eef9d4",
        "active": true,
        "data": {
            "tags": [],
```

```
"allowedMimeTypes": [],
        "maxAllowedFileSize": null,
        "allowedIpAddresses": [],
        "allowedUserAgents": [],
        "secureCopyEncryptionKey": "",
        "secureCopyEncryptionMethod": ""
    },
    "roles": [
        "security.generate_tokens"
    ]
},
    "id": "fad05629-51f6-4ddf-b21a-315a1451670d",
    "active": true,
    "data": {
        "tags": [],
        "allowedMimeTypes": [],
        "maxAllowedFileSize": null,
        "allowedIpAddresses": [],
        "allowedUserAgents": [],
        "secureCopyEncryptionKey": "",
        "secureCopyEncryptionMethod": ""
    },
    "roles": [
        "upload.images"
    1
},
    "id": "3235ad82-666f-4963-a751-b4dff3168c4c",
    "active": true,
    "expires": "2020-05-05 08:00:00",
    "expired": false,
    "data": {
        "tags": [
            "user_uploads.u123",
            "user_uploads"
        ],
        "allowedMimeTypes": [
            "image\/jpeg",
            "image\/png",
            "image\/gif"
        "maxAllowedFileSize": 100,
        "allowedIpAddresses": [],
        "allowedUserAgents": [],
        "secureCopyEncryptionKey": "",
        "secureCopyEncryptionMethod": ""
    },
    "roles": [
        "upload.images"
    ]
},
    "id": "dafe83fa-7813-4d84-a625-16c6657fec9f",
    "active": true,
    "data": {
        "tags": [],
```

```
"allowedMimeTypes": [],
    "maxAllowedFileSize": null,
    "allowedIpAddresses": [],
    "allowedUserAgents": [],
    "secureCopyEncryptionKey": "",
    "secureCopyEncryptionMethod": ""
    },
    "roles": [
        "collections.create_new",
        "collections.manage_tokens_in_allowed_collections"
    ]
}
```

Required roles:

- security.search_for_tokens
- security.authentication_lookup

9.3 Looking up a token

GET /auth/token/D0D12FFF-DD04-4514-8E5D-D51542DEBCFA?_token=your-admin-token-there

Example response:

```
"status": true,
"error_code": null,
"http_code": 200,
"errors": [],
"context": [],
"message": "Token found",
"token": {
    "id": "ca6a2635-d2cb-4682-ba81-3879dd0e8a77",
    "active": true,
    "data": {
        "tags": [],
        "allowedMimeTypes": [],
        "maxAllowedFileSize": 0,
        "allowedIpAddresses": [],
        "allowedUserAgents": [],
        "secureCopyEncryptionKey": "",
        "secureCopyEncryptionMethod": ""
    "roles": [
        "security.administrator",
        "upload.images",
        "upload.documents",
        "upload.backup",
        "upload.all",
        "security.authentication_lookup",
        "security.search_for_tokens",
        "security.overwrite",
```

```
"security.generate_tokens",
            "security.use_technical_endpoints",
            "deletion.all_files_including_protected_and_unprotected",
            "view.any_file",
            "view.files_from_all_tags",
            "view.can_use_listing_endpoint_at_all",
            "security.revoke_tokens",
            "collections.create_new",
            "collections.create_new.with_custom_id",
           "collections.allow_infinite_limits",
           "collections.delete_allowed_collections",
           "collections.modify_any_collection_regardless_if_token_was_allowed_by_
→collection",
           "collections.modify_details_of_allowed_collections",
           "collections.view_all_collections",
            "collections.can_use_listing_endpoint",
            "collections.manage_tokens_in_allowed_collections",
            "collections.upload_to_allowed_collections",
            "collections.list_versions_for_allowed_collections",
            "collections.delete_versions_for_allowed_collections",
            "securecopy.stream",
           "securecopy.all_secrets_read"
       1
   }
```

Required roles:

• security.authentication_lookup

9.4 Revoking a token

DELETE /auth/token/D0D12FFF-DD04-4514-8E5D-D51542DEBCFA?_token=your-admin-token-there

Example response:

```
"status": true,
"error_code": null,
"http_code": 201,
"errors": [],
"context": [],
"message": "Token was deleted",
"token": {
    "id": null,
    "active": true,
    "expires": "2020-05-05 08:00:00",
    "expired": false,
    "data": {
        "tags": [],
        "allowedMimeTypes": [],
        "maxAllowedFileSize": 0,
        "allowedIpAddresses": [],
        "allowedUserAgents": [],
```

```
"secureCopyEncryptionKey": "",
            "secureCopyEncryptionMethod": ""
        "roles": [
            "security.administrator",
            "upload.images",
            "upload.documents",
            "upload.backup",
            "upload.all",
            "security.authentication_lookup",
            "security.search_for_tokens",
            "security.overwrite",
            "security.generate_tokens",
            "security.use_technical_endpoints",
            "deletion.all_files_including_protected_and_unprotected",
            "view.any_file",
            "view.files_from_all_tags",
            "view.can_use_listing_endpoint_at_all",
            "security.revoke_tokens",
            "collections.create_new",
            "collections.create_new.with_custom_id",
            "collections.allow_infinite_limits",
            "collections.delete_allowed_collections",
            "collections.modify_any_collection_regardless_if_token_was_allowed_by_
⇒collection",
            "collections.modify_details_of_allowed_collections",
           "collections.view_all_collections",
           "collections.can use listing endpoint",
            "collections.manage_tokens_in_allowed_collections",
            "collections.upload_to_allowed_collections",
            "collections.list_versions_for_allowed_collections",
            "collections.delete_versions_for_allowed_collections",
            "securecopy.stream",
            "securecopy.all_secrets_read"
       1
   }
```

Required roles:

· security.revoke_tokens

9.5 Managing authentication using console commands

Tokens can be easily generated without touching the cURL or browser or any API client. Just use the console.

9.5.1 Generating an unlimited administrative token

Probably first time when you set up the **File Repository** you may want to create a token, that will allow you to fully manage everything. We already knew about such case and we're prepared for it! ;-)

```
./bin/console auth:generate-admin-token
Generating admin token...

(continues on next page)
```

In this case "1B3B15EC-18E9-45DD-846B-42C5006E872A" is your administrative token, pssst... keep it safe!

Note: You can also generate a token with custom tokenId using the -id switch

Note: Use –ignore-error-if-token-exists in scripts

9.5.2 Generating a normal token

It is considered a very good practice to minimize access to the resources. For example the server which will be storing backups on the **File Repository** should only be allowed to send backups, not deleting for example.

For such cases you can generate a token that will allow access to specified collections and limit actions on them.

```
./bin/console auth:create-token --help
Description:
 Creates an authentication token
  auth:create-token [options]
Options:
      --roles=ROLES
      --tags=TAGS
      --mimes=MIMES
      --max-file-size=MAX-FILE-SIZE
  -i, --id[=ID]
      --expires=EXPIRES
                                      Example: 2020-05-01 or +10 years
      --ignore-error-if-token-exists Exit with success if token already exists. Does_
→not check strictly permissions and other attributes, just the id.
  -h, --help
                                      Display this help message
  -q, --quiet
                                      Do not output any message
  -V, --version
                                      Display this application version
      --ansi
                                      Force ANSI output
     --no-ansi
                                      Disable ANSI output
  -n, --no-interaction
                                     Do not ask any interactive question
  -e, --env=ENV
                                      The Environment name. [default: "test"]
      --no-debug
                                      Switches off debug mode.
                                      Increase the verbosity of messages: 1 for_
  -v|vv|vvv, --verbose
→normal output, 2 for more verbose output and 3 for debug
Help:
 Allows to generate a token you can use later to authenticate in application for a_{\scriptscriptstyle \perp}
→specific thing
```

Example of generating a token with specified roles:

```
./bin/console auth:create-token --roles upload.images,upload.enforce_no_password --
→expires="+30 minutes" --id="A757A8CB-964F-4F7B-BB70-9DB2CF524BBA"
Form:
[Role] -> upload.images
[Role] -> upload.enforce_no_password
"status": true,
    "error_code": null,
    "http_code": 201,
    "errors": [],
    "context": [],
    "message": "Token created",
    "token": {
        "id": "A757A8CB-964F-4F7B-BB70-9DB2CF524BBA",
       "active": true,
       "expired": false,
        "expires": {
            "date": "2020-02-22 11:19:57.604976",
            "timezone_type": 3,
            "timezone": "UTC"
        "data": {
            "tags": [],
            "allowedMimeTypes": [],
            "maxAllowedFileSize": 0,
            "allowedIpAddresses": [],
            "allowedUserAgents": [],
            "secureCopyEncryptionKey": "",
            "secureCopyEncryptionMethod": ""
        "roles": [
            "upload.images",
            "upload.enforce_no_password"
       ]
    }
```

Note: When you not specify the –id, then the id will be generated automatically

Note: Use -ignore-error-if-token-exists in scripts

9.5.3 Deleting expired tokens

Delete expired tokens to clean up the database out of bloat. This should be a scheduled periodic job in a cronjob.

```
./bin/console auth:clear-expired-tokens
[2019-02-05 08:07:01] Removing token 276CCE10-00C5-4CB6-9F9A-87934101BACE
```

9.6 Token load hierarchy

Authorization can be provided in multiple ways, including query string, headers and environment variables. To build a perfect setup it is necessary to know, how the File Repository is fetching the token value, which source is in priority.

9.6.1 Loading priority

It's a top list, first match wins.

- 1. _token in query string eg. ?_token=xyz is used
- 2. token header
- 3. **x-auth-token** header
- 4. FILE_REPOSITORY_TOKEN environment variable

9.6.2 Use cases: Static assets serving

Best practice is to have each file, each collection secured with a token. You can generate a **viewer token**, and set it as an environment variable on given endpoints, or on whole application.

Using NGINX, Apache 2 or other webserver you can deny access to some routes, on other routes set a default access token - by enforcing a header or environment variable. The webserver proxies also gives a possibility to strip out request data, for example the headers and query string parts.

CHAPTER 10

Files storage

The file storage is like a bag of files, there are no directories, it's more like an object storage. When you put some file it is written down on the disk, and it's metadata is stored in the database.

Files could be tagged with some names, it's useful if the repository is shared between multiple usage types. The listing endpoint can search by tag, phrase, mime type - the external application could use listing endpoint to show a gallery of pictures for example, uploaded documents, attachments lists.

In short words the **File Storage** is a specialized group of functionality that allows to manage files, group them, upload new, delete and list them.

10.1 Security

10.1.1 Access

File can be PUBLIC or PRIVATE, the **public** attribute of input data that is sent together with file means the file will not be listed by listing endpoint (unless the token is not an administrative token).

Password protection could be used to protect from downloading the file content by not authorized person, and also it will anonymize the file in public listing if the person who lists the files will not know the password.

10.1.2 Uploading restrictions

When you give user a temporary token to allow to upload eg. avatar, then you may require that the file will not have a **password**, and possibly enforce to select some tags as mandatory.

Extra roles, that can restrict the token		
name description		
upload.enforce_no_password	Enforce files uploaded with this token to not have a password	
up-	Regardless of tags that user could choose, the tags from token will be copied	
load.enforce_tags_selected_in_tokento each uploaded file		

10.2 Uploading

Files could be uploaded in three ways - as RAW BODY, as POST form field and as URL from existing resource in the internet.

Common parameters for all endpoints	
name	description
tags	List of tags where the file will be listed
public	Should be listed/searched? (true/false)
password	Optionally allows to protect access to the file and it's metadata
encoding	Allows to upload encoded file, example values: base64, '' (helpful for frontend implemen-
	tation)

10.2.1 From external resource by URL

Endpoint specific parameters	
name	description
fileUrl	URL address to the file from the internet

```
POST /repository/image/add-by-url?_token=some-token-there
{
    "fileUrl": "http://zsp.net.pl/files/barroness_logo.png",
    "tags": [],
    "public": true
}
```

10.2.2 In RAW BODY

Endpoint specific parameters	
name	description
filename	Filename that will be used to access the file later

```
POST /repository/file/upload?_token=some-token-here&fileName=heart.png
< some file content there instead of this text >
```

Notes:

- Filename will have added automatically the content hash code to make the record associated with file content (eg. heart.png -> 5Dgds3dqheart.png)
- Filename is unique, same as file
- If file already exists under other name, then it's name will be returned (deduplication mechanism)

10.2.3 In a POST form field

Endpoint specific parameters	
name	description
filename	Filename that will be used to access the file later

```
POST /repository/file/upload?_token=some-token-here&fileName=heart.png

Content-Type: multipart/form-data; boundary=----WebKitFormBoundary7MA4YWxkTrZu0gW
-----WebKitFormBoundary7MA4YWxkTrZu0gW
Content-Disposition: form-data; name="file"; filename=""
Content-Type: image/png
------WebKitFormBoundary7MA4YWxkTrZu0gW--
... file content some where ...
```

10.3 Downloading

When you upload your file you will always get an URL address in the JSON response, but the download endpoints has more to offer than it looks on first view. Let's explain additional things you can do with the download endpoint.

Features:

- Bytes range support, files could be downloaded partially, videos can be rewinded while streamed
- · Big files support
- Content type is sent, so the browser knows the file size and can show the progress bar
- · Optional password protection

Common parameters for all endpoints	
name	description
password	Password to access the file, optionally if the file is password protected

10.3.1 Regular downloading

It's very simple.

```
\label{lem:general} $$\operatorname{GET /repository/file/d3beb8a9f0some-file-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optional-password-optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt?password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.password=optie-name-there.txt.pa
```

10.3.2 Downloading using alias defined in ids_mapping.yaml

Aliases are allowing to access files by other names, they can be defined in ./config/ids_mapping.yaml file. It's very helpful feature when you migrate from other storage application to File Repository.

Example ids_mapping.yaml file:

10.3. Downloading 45

```
"oh-my-alias-there": "d3beb8a9f0some-file-name-there.txt"
```

Example request:

```
GET /repository/file/oh-my-alias-there
```

10.3.3 Downloading using hotlink protection

Hotlink protection is allowing to generate personalized download urls by combining eg. user's IP address, some salt, file name and timestamp. Such link cannot be shared with other users.

Note: Hotlink protection endpoint also supports aliasing.

Example request:

```
\label{eq:GET/stream/531ce1f1d5d242cd5005b3758d3b5435/2219788800/d3beb8a9f0some-file-name-sthere.txt
```

The format of the URL is defined in the environment variables:

```
ANTI_HOTLINK_PROTECTION_ENABLED=true
ANTI_HOTLINK_RESTRICT_REGULAR_URLS=false
ANTI_HOTLINK_URL=/stream/{accessToken}/{expirationTime}/{fileId}
ANTI_HOTLINK_CRYPTO=md5
ANTI_HOTLINK_SECRET_METHOD="\$http_x_expiration_time\$http_test_header MY-AWESOME-
$\times\SUFFIX"
```

It means you can change it, so the URL will be different. {expirationTime} is optional, but very helpful.

Short explanation:

The {accessToken} is generated by hashing with eg. md5 the filled-up ANTI_HOTLINK_SECRET_METHOD.

Example: Given ANTI_HOTLINK_SECRET_METHOD is "\$http_x_expiration_time\$http_test_header MY-AWESOME-SUFFIX" We send a request with {expirationTime} = 123 and a header Test-Header = HELLO

So, the secret would be "123HELLO MY-AWESOME-SUFFIX", now we have to hash using selected crypto - md5. md5(23HELLO MY-AWESOME-SUFFIX) = 531ce1f1d5d242cd5005b3758d3b5435

It means that we have URL: /stream/531ce1f1d5d242cd5005b3758d3b5435/123/d3beb8a9f0some-file-name-there.txt

10.4 Aliasing filenames (migrating existing files to File Repository)

Filename in **File Repository** is created based on file contents hash + name submitted by user. To allow **easier migration of your existing files**, the **File Repository** allows to create *aliases* to files you upload.

10.4.1 Scenario

Let's assume that you have a file named "Accidential-Anarchist.mp4", and your website shows a player that points to https://static.iwa-ait.org/Accidential-Anarchist.mp4 Now you want to migrate your storage to use **File Repository**, so the **File Repository** will store and serve the files with help of your webserver.

To keep old links still working you need to:

- Set up a URL rewrite in your webserver (eg. NGINX or Apache 2) to rewrite the FORMAT OF THE URL, example: /education/movies/watch?v=... to /repository/file/...
- You have a file "Accidential-Anarchist.mp4", after uploading to File Repository it will have different name eg. "59dce00bcAccidential-Anarchist.mp4", you can create an alias that will point from "Accidential-Anarchist.mp4" to "59dce00bcAccidential-Anarchist.mp4"

10.4.2 Practice, defining aliases

To start you need to create a file *config/ids_mapping.yaml*, where you will list all of the aliases in YAML syntax.

Example:

Notice: You need to restart the application (or execute ./bin/console cache:clear -env=prod) after applying changes to this file

10.5 Hotlink protection - personalizing URLs for your visitors

If for any reason you need to secure your content from being distributed outside of your website, then you need a hotlink protection. Hotlink protection gives your website a **control over who can see the video, image or any other resource that is kept on File Repository**.

10.5.1 Preparing your website and File Repository configuration

A website that is displaying eg. a video player that would play a video from **File Repository** need to point to a personalized URL address especially generated for your page visitor.

At first let's look at the URL format, you need to define a URL format that will point to protected files. Below there are multiple examples, you can configure the URL however you want, **this you need to adjust in your .env file or in environment variables in Docker container**.

```
# example 1
ANTI_HOTLINK_URL=/stream/{accessToken}/{expirationTime}/{fileId}

# example 2
ANTI_HOTLINK_URL=/video/{accessToken}/{expirationTime}/{fileId}

# example 3
ANTI_HOTLINK_URL=/watch/{fileId}, {accessToken}, {expirationTime}

# example 4
ANTI_HOTLINK_URL=/watch/{accessToken}/{fileId}
```

So, let's take a look at the most interesting part - the access token generation.

Each visitor on your page needs to get a unique access token that will allow to see the file content only for him/her. To generate such access token we need to DEFINE A COMMON FORMAT that your application will use and File Repository will understand.

```
ANTI_HOTLINK_SECRET_METHOD="\$http_x_expiration_time\$http_x_real_uri\$http_x_remote_

addr MY-AWESOME-SUFFIX"
```

Following example is combining most important variables, why?

- \$http_x_real_uri to restrict this token only to single file (this header may be required to be set on NG-INX/Apache level)
- \$http_x_remote_addr to restrict access to single IP address
- MY-AWESOME-SUFFIX this one definitely you should change to a SECRET you only know. It will prevent anybody from generating a token
- \$http_x_expiration_time optionally validate the passed input data in the url

Generally the rule with the variables is simple as in NGINX, but a little bit more extended to give better possibilities.

Variable tem	Variable templates	
name description		
\$http_xxx	In place of xxx put your normalized header name eg. Content-Type would be content_type	
\$server_xxx	Everything what is in PHP's \$_SERVER, including environment variables	
\$query_xxx	Everything what is in query string (query string in URL is everything after question mark)	

10.5.2 Practical example of generating access token on your website

Assuming that you have following configuration:

```
ANTI_HOTLINK_PROTECTION_ENABLED=true

ANTI_HOTLINK_RESTRICT_REGULAR_URLS=false

ANTI_HOTLINK_CRYPTO=md5

ANTI_HOTLINK_SECRET_METHOD="\$http_x_expiration_time\$filename\$http_remote_addr MY-

AWESOME-SUFFIX"

ANTI_HOTLINK_URL=/stream/{accessToken}/{expirationTime}/{fileId}
```

That would be an example code that could generate URL addresses in your application:

```
<?php
$fileId = 'Accidential-Anarchist.mp4';

(continues on next page)</pre>
```

Chapter 10. Files storage

```
$expirationTime = time() + (3600 * 4); // +4 hours
$rawToken = $expirationTime . $fileId . ($_SERVER['REMOTE_ADDR'] ?? '') . ' MY-

AWESOME-SUFFIX';

$hash = hash('md5', $rawToken);
echo 'URL: /stream/' . $hash . '/' . $expirationTime . '/' . $fileId;
```

10.6 Listing and searching

Each file can be found by using a search endpoint. Password protected files are censored, if the correct password was not entered in the search field.

Note: Files can be named and tagged, marked as public/private, password protected.

Parameters	
name	description
page	Page number
limit	Limit results on single page
password	Password for password-protected files
searchQuery	Search phrase, a word, multiple words to be searched for in the file name
tags	List of tags to filter by (array)
mimes	List of mimes to filter by (array)

Example request:

```
GET /repository?_token=your-auth-token&page=1&limit=20
```

CHAPTER 11

Backup

Backup *collections* allows to store multiple *versions* of the same file.

Each submitted version has automatically incremented version number by one.

Example scenario with strategy "delete_oldest_when_adding_new":

```
Given we have DATABASE dumps of iwa-ait.org website
And our backup collection can contain only 3 versions (maximum)

When we upload a sql dump file THEN IT'S a v1 version
When we upload a next sql dump file THEN IT'S a v2 version
When we upload a next sql dump file THEN IT'S a v3 version

Then we have v1, v2, v3

When we upload a sql dump file THEN IT'S a v4 version
But v1 gets deleted because collection is full

Then we have v2, v3, v4
```

From security point of view there is a possibility to attach multiple tokens with different access rights to view and/or manage the collection.

11.1 Getting started

The workflow is following:

- 1. You need to have an access token that allows you to create collections
- 2. Create a collection, remember it's ID (we will call it collection_id later)
- 3. (Optional) Allow some other token or tokens to access the collection (all actions or only some selected actions on the collection)
- 4. Store backups under a collection of given collection_id

5. List and download stored backups when you need

11.1.1 Versioning

Each uploaded version is added as last and have a version number incremented by one, and a ID string generated.

For example: There is a v1 version, we upload a new version and a new version is getting a number v2

Later any version could be accessed by generated **ID** string or **version number** (in combination with the **collection ID**)

11.1.2 Collection limits

Each collection could either be a infinite collection or a finite collection.

Below are listed limits for finite collections:

Limits	
limit	description
maxBackupsCount	Maximum count of versions that could be stored
maxOneVersionSize	Maximum disk space that could be allocated for single version
maxCollectionSize	Maximum disk space for whole collection (summary of all files)

11.1.3 Permissions

There could be multiple tokens with different permissions assigned to the collection.

Example use case: Generated "Guest token" with download-only permissions could be safe to share between administrators. The "Upload token" could be used by the server to automatically upload new versions without permissions to delete other versions and without need to modify collections limits. "Management token" with all of the permissions for managing a collection.

11.2 Managing collections

To start creating backups you need a collection that will handle ONE FILE. The file may be a zipped directory, a text file, SQL dump or anything you need.

11.2.1 Collection creation

To add any backup you need a **collection** at first. Collection is a container that keeps multiple versions of same file (for example your database dump from each day). Collection additionally can define limits on length, size, type of uploaded file, and tokens which have access to it at all.

Example request:

```
POST {{appUrl}}/repository/collection?_token=test-token-full-permissions
{
    "maxBackupsCount": 5,
    "maxOneVersionSize": 0,
```

```
"maxCollectionSize": "250MB",
   "strategy": "delete_oldest_when_adding_new",
   "description": "iwa-ait.org database backup",
   "filename": "iwa-ait-org.sql.gz"
}
```

In the response you will receive a collection **ID** that will be required for editing collection information, assigning tokens and uploading files.

There are two strategies. **delete_oldest_when_adding_new** is automatically deleting older backup versions when a maxBackupsCount is reached and a new backup is submitted. **alert_when_backup_limit_reached** will raise an error when submitting a new version to already full backup collection.

Notes:

- Put zero values to disable the limit
- Supports "simulate=true" parameter that allows to send a request that will not create any data, but only validate submitted data
- You'r token will be automatically added as token allowed to access and modify the collection

Required permissions:

collections.create_new

Optional permissions:

• collections.allow_infinite_limits (allows to create an infinite collection, it means that you can eg. upload as much files as you like to, and/or the disk space is unlimited)

11.2.2 Collection editing

```
PUT {{appUrl}}/repository/collection?_token=test-token-full-permissions

{
    "collection": "SOME-COLLECTION-ID-YOU-RECEIVED-WHEN-CREATING-THE-COLLECTION",
    "maxBackupsCount": 5,
    "maxOneVersionSize": 0,
    "maxCollectionSize": "250MB",
    "strategy": "delete_oldest_when_adding_new",
    "description": "iwa-ait.org database backup (modified)",
    "filename": "iwa-ait-org.sql.gz"
}
```

Notes:

- The collection size cannot be lower than it is actual in the storage (sum of existing files in the collection)
- You need to have global permissions for managing any collection or to have token listed as allowed in collection you want to edit

Required permissions:

· collections.modify_details_of_allowed_collections

Optional permissions:

• collections.allow_infinite_limits (allows to edit an infinite collection, it means that you can eg. upload as much files as you like to, and/or the disk space is unlimited)

• collections.modify_any_collection_regardless_if_token_was_allowed_by_collection (gives a possibility to edit a collection even if token is not attached to it)

11.2.3 Deleting

To delete a collection you need to at first make sure, that there are no backup versions attached to it. Before deleting a collection you need to manually delete all backups. It's for safety reasons.

```
DELETE {{appUrl}}/repository/collection/SOME-COLLECTION-ID?_token=test-token-full-
--permissions
```

Required permissions:

· collections.delete allowed collections

Optional permissions:

• collections.modify_any_collection_regardless_if_token_was_allowed_by_collection (gives a possibility to edit a collection even if token is not attached to it)

11.2.4 Fetching collection information

You can fetch information about collection limits, strategy, description and more to be able to edit it using other endpoints.

```
GET {{appUrl}}/repository/collection/SOME-COLLECTION-ID?_token=test-token-full-

→permissions
```

Notes:

• You need to have global permissions for managing any collection or to have token listed as allowed in collection you want to fetch

Required permissions:

• (just the token added as allowed for given collection)

Optional permissions:

• collections.modify_any_collection_regardless_if_token_was_allowed_by_collection (gives a possibility to edit a collection even if token is not attached to it)

11.3 Authorization

Multiple tokens with different permissions could be assigned to the single collection. You may create a token for uploading backups, deleting backups and for managing collection limits separately.

11.3.1 Assigning a token to the collection

```
POST /repository/collection/{{collection_id}}/token?_token={{collection_management_

→token}}
{
```

```
"token": "SO-ME-TO-KEN-TO-ADD"
}
```

Legend:

- {{collection management token}} is your token that has access rights to fully manage collection
- {{collection_id}} is an identifier that you will receive on collection creation (see collection creation endpoint)

Required permissions:

• collections.manage tokens in allowed collections

11.3.2 Revoking access to the collection for given token

Legend:

- {{token_id}} identifier of a token that we want to disallow access to the collection
- {{collection_management_token}} is your token that has access rights to fully manage collection
- {{collection_id}} is an identifier that you will receive on collection creation (see collection creation endpoint)

Required permissions:

· collections.manage_tokens_in_allowed_collections

11.4 Backups: Upload, deletion and versioning

Assuming that you have already a collection and an access token, then we can start uploading files that will be versioned and stored under selected collection.

11.4.1 Uploading a new version to the collection

You need to submit **file content in the HTTP request body**. The rest of the parameters such as token you need to pass as GET parameters.

```
POST /repository/collection/{{collection_id}}/backup?_token={{token_that_allows_to_

→upload_to_allowed_collections}}
.... FILE CONTENT THERE ....
```

Pretty simple, huh? As the result you will get the version number and the filename, something like this:

```
"status": "OK",
   "error_code": null,
   "exit_code": 200,
   "field": null,
   "errors": null,
   "version": {
```

```
"id": "69283AC3-559C-43FE-BFCC-ECB932BD57ED",
       "version": 1,
        "creation_date": {
           "date": "2019-01-03 11:40:14.669724",
           "timezone_type": 3,
           "timezone": "UTC"
       },
        "file": {
           "id": 175,
           "filename": "ef61338f0dsolidarity-with-postal-workers-article-v1"
   },
   "collection": {
       "id": "430F66C3-E4D9-46AA-9E58-D97B2788BEF7",
       "max_backups_count": 2,
       "max_one_backup_version_size": 1000000,
       "max_collection_size": 5000000,
        "created_at": {
           "date": "2019-01-03 11:40:11.000000",
           "timezone_type": 3,
           "timezone": "UTC"
       },
       "strategy": "delete_oldest_when_adding_new",
       "description": "Title: Solidarity with Postal Workers, Against State,
→Repression!",
       "filename": "solidarity-with-postal-workers-article"
   }
```

Required permissions:

• collections.upload to allowed collections

11.4.2 Deleting a version

A simple DELETE type request will delete a version from collection and from storage.

```
DELETE /repository/collection/{{collection_id}}/backup/BACKUP-ID?_token={{token}}
```

Example response:

```
"status": "OK, object deleted",
   "error_code": 200,
   "exit_code": 200
}
```

Parameters		
type	name	description
bool	simulate	Simulate the request, do not delete in real. Could be used as pre-validation
string	_token	Standard access token parameter (optional, header can be used instead)

Required permissions:

• collections.delete_versions_for_allowed_collections

11.4.3 Getting the list of uploaded versions

To list all existing backups under a collection you need just a collection id, and the permissions.

```
GET /repository/collection/{{collection_id}}/backup?_token={{token}}
```

Example response:

```
"status": "OK",
   "error_code": null,
   "exit_code": 200,
   "versions": {
       "3": {
            "details": {
                "id": "A9DAB651-3A6F-440D-8C6D-477F1F796F13",
                "version": 3,
                "creation_date": {
                    "date": "2019-01-03 11:40:24.000000",
                    "timezone_type": 3,
                    "timezone": "UTC"
                },
                "file": {
                    "id": 178,
                    "filename": "343b39f56csolidarity-with-postal-workers-article-v3"
            },
            "url": "https://my-anarchist-initiative/public/download/
→343b39f56csolidarity-with-postal-workers-article-v3"
       "4": {
            "details": {
                "id": "95F12DAD-3F03-49B0-BAEA-C5AC3E8E2A30",
                "version": 4,
                "creation_date": {
                    "date": "2019-01-03 11:47:34.000000",
                    "timezone_type": 3,
                    "timezone": "UTC"
                },
                "file": {
                    "id": 179,
                    "filename": "41ea3dcca9solidarity-with-postal-workers-article-v4"
           },
            "url": "https://my-anarchist-initiative/public/download/
→41ea3dcca9solidarity-with-postal-workers-article-v4"
   }
```

Required permissions:

• collections.list_versions_for_allowed_collections

11.4.4 Downloading uploaded versions

Given we upload eg. 53 versions of a SQL dump, one each month and we want to download latest version, then we need to call the fetch endpoint with the "latest" keyword as the identifier.

If there is a need to download an older version of the file, a version number should be used, eg. v49

There is also a possibility to download a last copy from the bottom, the oldest version available using keyword first.

In case we have an ID of the version, then it could be inserted directly replacing the alias keyword.

```
GET /repository/collection/{{collection_id}}/backup/69283AC3-559C-43FE-BFCC-

→ECB932BD57ED?password=thats-a-secret&_token={{token}}
```

Parame	eters	
type	name	description
bool	redirect	Allows to disable HTTP redirection and return JSON with the url address instead
string	password	Password required for requested FILE (please read about passwords in notes section)
string	_token	Standard access token parameter (optional, header can be used instead)

Required permissions:

- · collections.list_versions_for_allowed_collections
- (knowing the password for the collection file)

Notes:

• The password for the file is inherited from collection, but it may be different in case when the collection would have changed the password, old files would not be updated!

11.5 Data replication

File Repository does not replicate the data on application level as it does not make sense to do so. There are specialized filesystems such as S3, Glusterfs or DRBD, specialized database servers that handles primary-replica and primary-primary replication.

When setting up the replication you need to remember, that the application itself cannot keep any state locally, that's the rule of the replication. Often people forget about the application cache, which is not so obvious.

There are 3 areas that needs to be replicated, so the application could be scaled:

- · Storage backend
- Database
- Cache

11.5.1 Choosing scalable storage backends for File Repository

Min.io (using S3 adapter)

- Amazon S3 (using S3 adapter)
- GlusterFS (local filesystem)
- DRBD (local filesystem)
- Ceph (local filesystem)

11.5.2 Selecting a scalable database

Any modern database server supports the replication, it's up to you to pick the best. At RiotKit we are preferring PostgreSQL. Please note that SQLite3 is a tiny scale in-file database that does not scale.

- PostgreSQL
- MySQL
- Oracle
- Microsoft SQL Server

11.6 Managing collections from shell

To allow automating things there are shell commands, those do not require authorization and have the same parameters as API endpoints.

11.6.1 Creating collections

The command will return just a collection id on success. On failure a json is returned.

Example success output:

```
./bin/console backup:create-collection -d "Some test collection" -f "backup.tar.gz" -

→b 4 -o 3GB -c 15GB

48449389-E267-497E-A6F4-EAC91C063708
```

Example failure output:

```
"error_code": 4003,
    "http_code": 400,
    "errors": {
         "maxCollectionSize": "max_collection_size_is_lower_than_single_element_size"
      },
      "collection": null,
      "context": []
}
```

 $Note:\ Use-id\ parameter\ to\ assign\ custom\ id\ for\ the\ collection.\ Helpful\ in\ deployment\ automation.$

Note: Use –ignore-error-if-token-exists in scripts

CHAPTER 12

MinimumUI

Although that **File Repository** is an API project, it has a few HTML endpoints which are allowing to upload files. MinimumUI idea is to allow to use *File Repository* as a fully standalone microservice, with easy to use embeddable upload forms on any website.

12.1 Quick start in steps

- 1. Your application needs to have a possibility to *create tokens* in **File Repository** on backend side (no one should see your administrative token).
- 2. For each user you need to generate a temporary token with minimal permissions (eg. upload only, with restrictions for password, mime types, tags etc.)
- 3. On your website you need to redirect user to the file repository upload form (MinimumUI endpoint) with specifying the "back" parameter in query string, so the user will go back on your website again and pass the uploaded file URL
- 4. You need to validate the URL from your user, if it comes eg. from proper domain where File Repository runs

12.2 Endpoints

Following endpoints are just displaying a static HTML page, that acts as a client to the API. No any endpoint is implementing any additional access rights, if the user does not have access to perform some action, then the page would display, but the backend will respond with an error.

If you need to restrict the file size, mime type, allowed tags or others, then you need to specify it in the access token that will be used in the UI.

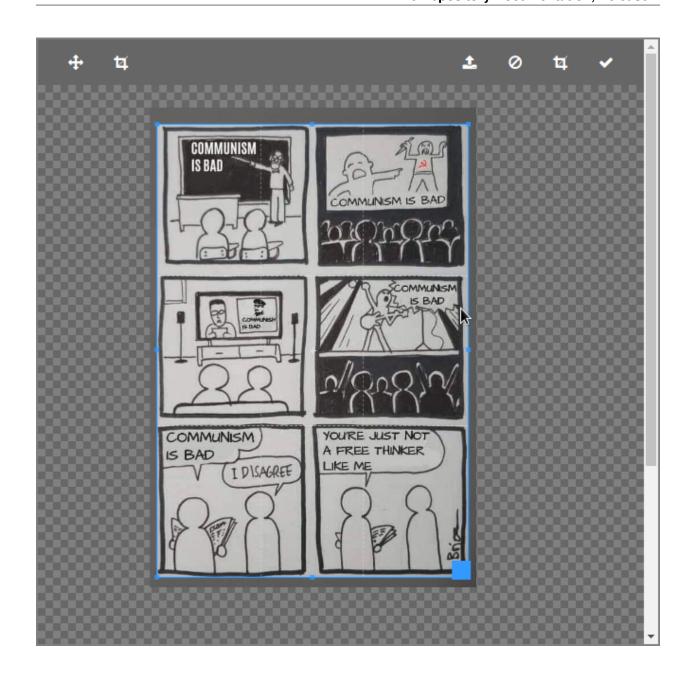
Roles used by the endpoints	
name	description
upload.enforce_no_password	Enforce the file to be uploaded without a password
upload.enforce_tags_selected_in_token	Tag uploaded file with tags specified in the token, regardless of user
	choice
upload.images	Upload images

12.2.1 Image Upload

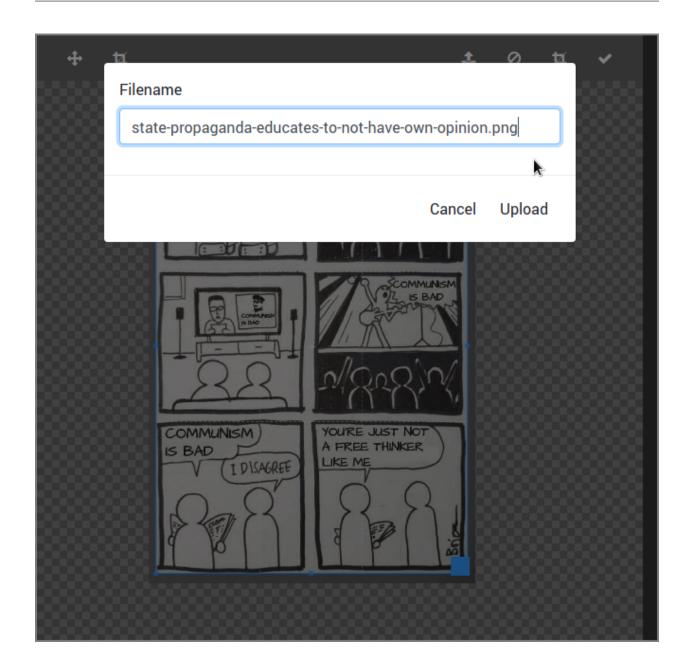
The image upload endpoint allows to upload whole file as is, or with cropping it. Cropper supports an aspect ratio, that could be specified in the query string.

Extra parameters in query string	
name	description
ratio	Aspect ratio for the images eg. 16/9 is 1.77, so it would be ?ratio=1.77
back	URL address to redirect the user on success. FILE_REPOSITORY_URL phrase will be replaced
	with the uploaded file URL
_token	Access token

In the browser access URL: /minimum.ui/upload/image?_token=TOKEN-THERE



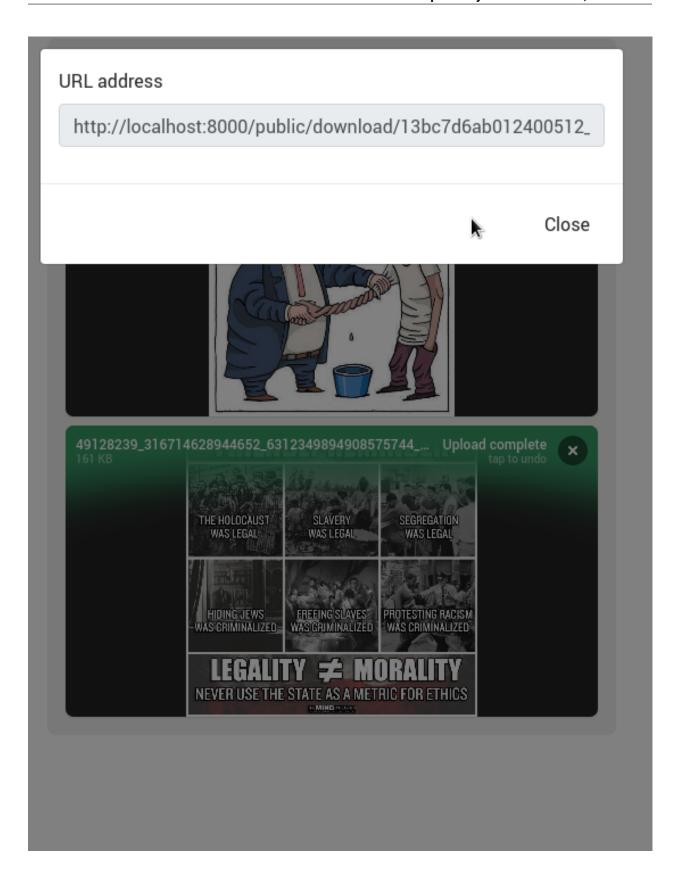
12.2. Endpoints 63



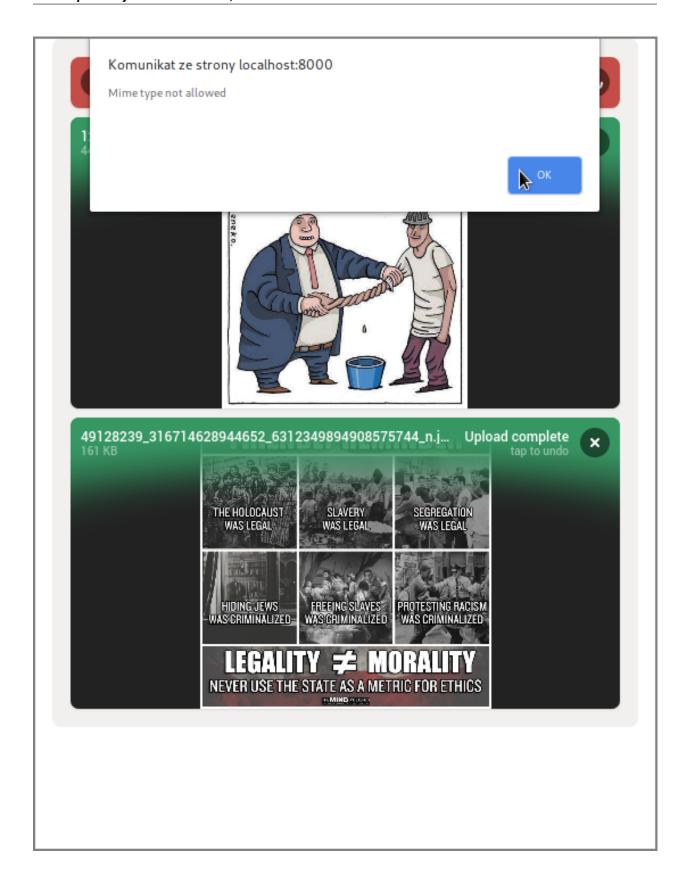
12.2.2 File upload

File upload offers a multiple file upload, with drag & drop and fancy animations.

In the browser access URL: /minimum.ui/upload/file?_token=TOKEN-THERE



12.2. Endpoints 65



66

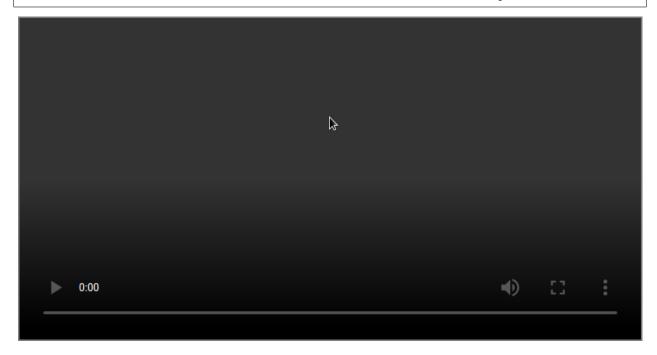


12.2. Endpoints 67

12.2.3 Video watching

File Repository is able to serve video files with possibility to rewind them, that's the responsibility of the download endpoint. MinimumUI exposes additional endpoint with a HTML5 <video> tag, so the video could be embedded easily on other website.

In the browser access URL: /minimum.ui/watch/video/some-file-name.mp4



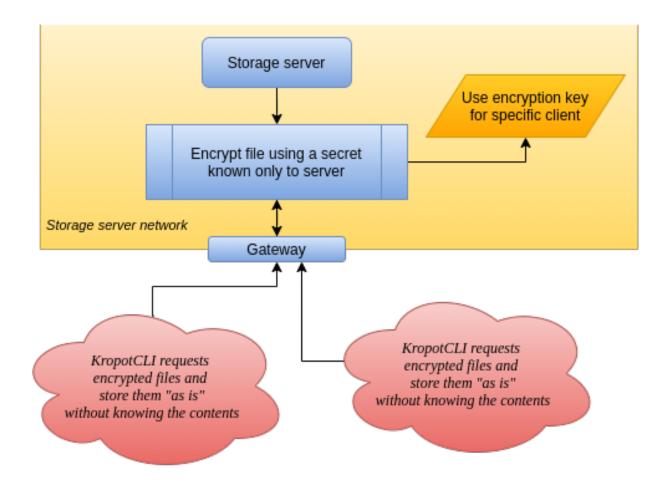
SecureCopy

Storage mirroring with additional layer of security - encryption on the server side.

Features:

- Each client have it's own permissions.
- Encryption credentials are per-client (in token generated in File Repository)
- Black-box/Zero-knowledge encryption, the client only retrieve data
- External, ready to use client application "kropot-cli" that will help you to better redistribute your bread ;-)

The difference between "rsync" type tools and File Repository's Secure Copy	
File Repository	Other tools such as rsync
Token based authorization	Requires SSH access
Independent of storage filesystem (s3, local, other net-	Requires local disk access
worked)	
Encryption on server side, without sharing the key to	Impossible to perform a on-fly encryption, when client
client	requests the files
Requires additional setup time, requires database,	Less maintenance, no database required, less frequent up-
maintenance time	dates
Brand new tool, nobody recognizes it	Everybody know the basic UNIX tools, less entry thresh-
	old



13.1 SecureCopy API endpoints

Copying mechanism exposes a list of files with a pagination/buffer, and a download endpoint.

The data in download endpoint can be encrypted or not, it depends on who is actually requesting the data. If a person with token, that has encryption enabled is requesting a file, then it will be encrypted. The person does not need to know the token, it will only receive a black-box data without having a knowledge what's inside.

Roles used by the endpoints	
name	description
securecopy.stream	Can stream list of all files in storage, it's metadata and file content (in encrypted or not encrypted form)
secure-	(Administrative) Can read all tokens encryption secrets
copy.all_secrets_read	

Bahub API client

Bahub is an automation tool for uploading and restoring backups. Works in shell, can work as a docker container in the same network with scheduled automatic backups of other containers, or can work as an UNIX daemon on the server without containerization.

```
>> You are in a docker container shell, where bahub is installed and configured

Examples of usage:
    bahub --debug backup some_redis_storage
    bahub list some_redis_storage
    bahub --debug restore some_redis_storage v1
    bahub restore some_redis_storage latest

    Check out configuration: cat /bahub.conf.yaml
    The variables between ${ } are fetched from environment, check it: type 'env'

bash-5.0# bahub backup some_redis_storage
{'version': 3, 'file_id': 'D422ED5E-EE0F-4D44-BD3D-20CAB0374D36', 'file_name': '5ad8eaa8laredis-data-files.tar-v3.gz'}
bash-5.0# bahub restore some_redis_storage latest
{"status": "OK"}
bash-5.0#
```

14.1 Configuration reference

There are 3 sections:

- Access: Describes authorization details, name it eg. server1 and put url and token
- Encryption: Encryption type and password (if any) to encrypt your files stored on File Repository
- Backups: Describes where is your data, how to access it and under which COLLECTION to send it to **File Repository**
- Recoveries: Recovery plans. A policy + list of "backups" to restore within a single command

Example scenario:

- 1. You have a server under https://backups.iwa-ait.org and token "XXX-YYY-ZZZ-123", you name it "ait_backups" under **access** section
- 2. You want to have encrypted backups using AES 256 CBC, then you add "ait_secret" under **encryption** with passphrase "something-secret" and type "aes-256-cbc"
- 3. Next you want to define where is the data, in our example it's in a docker container under /var/lib/mysql and we want to send this data to collection "123-456-789-000". You should reference "ait_backups" access and "ait_secret" as the encryption method for your backup there.

14.1.1 Environment variables

If you want to use environment variables, use bash-like syntax \${SOME_ENV_NAME}.

NOTE: In case you will not set a variable in the shell, then application will not start, it will throw a configuration error.

14.1.2 Application configuration

Notice: Below example uses environment variables eg. \${DB_HOST}, you may want to replace them with values like localhost or others

```
# Access information about the storage server (File Repository), the URL, token
# In "backups" section you should use name of an access as a reference in field
→ "access"
accesses:
   backup_one:
       url: http://localhost:8000
        token: test-token-full-permissions
# Allows to define an encryption and identify it with name
# In "backups" section you should use just the name as reference in "encryption" key...
⇔of a backup
encryption:
    # Example of AES-256-CBC encryption (uses OpenSSL)
      Decryption command example:
        cat 0ad166b4cfbackup.tar-v5.gz| openss1 enc -aes-256-cbc -pbkdf2 -d -pass_
→pass:some-encryption-key-here > ./backup.tar.gz
    #
    enc1:
       passphrase: some-encryption-key-here
       method: aes-256-cbc
                                 # possible values: aes-256-cbc, aes-256-ecb, aes-128-
→cbc, aes-128-ecb, des-ecb
        #encrypt_cmd: openssl enc -%method% -pass pass:%pass%
        #decrypt_cmd: openssl enc -d -%method% -pass pass:%pass%
   none:
       passphrase: ""
       method: ""
```

```
# Backups - list of backups, each backup = single archive = single collection.
           To backup invoke the shell command with a backup name eq. bahub backup,
→www_uploads
           To restore invoke eq. bahub restore www_uploads latest
backups:
    # Online / Hot-backup of docker container example
      Copies data from running container.
    # NOTICE: May be dangerous to your data, please make sure you know what you do.
               In most cases please choose "docker_volumes" method for safe, offline.
⇒backup.
   docker_hot_volumes_example:
       type: docker_hot_volumes
        container: "test_www"
       access: backup_one
       encryption: none
        collection_id: "${COLLECTION_ID}"
       paths:
            - /var/www
            - /var/log
        # optional
        #tar_pack_cmd: "tar -czf %stdin% %paths%"
        #tar_unpack_cmd: "tar -xzf %stdin% %target%"
        #docker_bin: "sudo docker"
    # Offline backup of docker container
      Stops the container, copies the data, then starts it again
      Fully safe method of backup for all types of services. Makes a downtime, but
→guarantees a non-corrupted
      backup data.
   www_docker_offline:
       type: docker_volumes
        container: "test_www"
        access: backup_one
        encryption: enc1
        collection_id: "${COLLECTION_ID}"
       paths:
           - /etc
           - /var/lib/mysql
           - /var/log
        # optional
        docker bin: "sudo docker"
        tar_pack_cmd: "tar -czf %stdin% %paths%"
        tar_unpack_cmd: "tar -xzf %stdin% %target%"
        temp_image_name: "alpine:3.9"
```

```
temp_image_cmd: "apk add --update xz bzip2 && sleep 3600"
   # MySQL online backup using mysqldump
      Can backup a single database or all databases from any MySQL server, local,
⇒in-docker or remote.
      Notice: Restoring backups is only possible for SINGLE DATABASES
   mysql_native_single_database:
       type: mysql
       host: "${DB_HOST}"
       port: 3306
       user: root
       password: root
       database: "${DB_DATABASE}"
       access: backup_one
       encryption: none
       collection_id: "${COLLECTION_ID}"
   # MySQL online backup using mysqldump (DOCKER)
      Can backup a single database or all databases from any MySQL server, local,
⇒in-docker or remote.
   # Notice: Restoring backups is only possible for SINGLE DATABASES
   mysql_docker_single_database:
       type: mysql
       host: localhost
       port: 3306
       user: root
       password: root
       database: "${DB_DATABASE}"
       container: "test_mysql" # this one is required to use a docker container
       access: backup_one
       encryption: none
       collection_id: "${COLLECTION_ID}"
       # optional:
       #tar_pack_cmd: "tar -czf %stdin% %paths%"
       #tar_unpack_cmd: "tar -xzf %stdin% %target%"
   # Docker command output online backup
       Executes command in a Docker container and captures output
   docker_command_output:
       type: docker_output
       container: "test_www"
       command: "cat /bin/sh"
       restore_command: "cat - > /tmp/sh.restored"
       access: backup one
       encryption: enc1
       collection_id: "${COLLECTION_ID}"
```

```
# optional
       #docker_bin: "sudo docker"
   # PostgreSQL backup using pg_dumpall. For restore psql is used.
   postgres:
       type: postgres
       access: backup_one
       encryption: enc1
       collection_id: "${COLLECTION_ID}"
       host: "localhost"
       port: 5432
       user: "${POSTGRES_USER}"
       password: "${POSTGRES_PASSWORD}"
       database: ""
   # PostgreSQL backup using pg_dump to dump single database. For restore psql is.
\hookrightarrowused.
   postgres_single_db:
       type: postgres
       access: backup_one
       encryption: enc1
       collection_id: "${COLLECTION_ID}"
       host: "localhost"
       port: 5432
       user: "${POSTGRES_USER}"
       password: "${POSTGRES_PASSWORD}"
       database: "some_db_name"
   # PostgreSQL backup using pg_basebackup command (replication mode backup)
   postgres_base:
       type: postgres_base
       access: backup_one
       encryption: enc1
       collection_id: "${COLLECTION_ID}"
       host: "localhost"
       port: 5432
       user: "${POSTGRES_USER}"
       password: "${POSTGRES_PASSWORD}"
       database: "${POSTGRES_DB}"
       server shutdown cmd: "sudo docker stop fr_tests_postgresql_1" # can there be.
→pg_ctl used, or systemctl also
       server_start_cmd: "sudo docker start fr_tests_postgresql_1"
       restore_dir: "${CONFIG_DIR}/.data/postgresql"
       old_dir: "${CONFIG_DIR}/.data/postgresql-old"
       ownership: "70:70" # postgres user id and group, can be also names. For
→docker containers it is recommended to use ids
   # Execute a local command
```

```
Executes command on this machine/environment on local shell and captures,
→output
    #
   local_command_output:
        type: command_output
        command: "cat /bin/bash"
        restore_command: "cat - > /tmp/bash.restored"
        access: backup_one
        encryption: enc1
        collection_id: "${COLLECTION_ID}"
    # Local file/directory backup
      Packs a local directory or file
    some_local_dir:
        type: directory
        paths:
            - /tmp/test
        access: backup_one
        encryption: enc1
        collection_id: "${COLLECTION_ID}"
        # optional
        #tar_pack_cmd: "tar -czf %stdin% %paths%"
        #tar_unpack_cmd: "tar -xzf %stdin% %target%"
# Recovery plans
  Restores multiple backups in order, using single command
  Possible values:
     policy:
        - restore-whats-possible: Ignore things that cannot be restored, restore what...
\hookrightarrow is posible
       - stop-on-first-error: Restore until first error, then stay as it is
recoveries:
        policy: restore-whats-possible
        definitions: all
    some_selected_only:
        policy: stop-on-first-error
        definitions:
            - some_local_dir
            - local_command_output
# Optional: Error handlers, very helpful to track backup errors
           Be careful! Your shell commands for backup/restore may fail some time, it
\hookrightarrow's worth to have a notification.
           You can set up a free account on sentry.io, install a Sentry instance on,
→your server, or use eg. Mattermost/Slack notifications
#error_handlers:
```

```
# remote_sentry:
# type: sentry
# url: "${SENTRY_IO}"

#
# Optional: Notifications. Can notify when backup starts, ends, there is an error etc.
#
#notifiers:
# mattermost:
# type: slack
# url: "https://xxxxx"
```

14.2 Basic usage

Bahub is offering basic operations required to automate backup sending and receiving, not managing the server.

14.2.1 Sending a backup

```
$ bahub --config ~/.bahub.yaml backup some_local_dir
{'version': 72, 'file_id': 'E9D7103D-1789-475E-A3EE-9CF18F51ACA4', 'file_name':
    '2b2e269541backup.tar-v72.gz'}
```

14.2.2 Listing stored backups

```
$ bahub --config ~/.bahub.yaml list some_local_dir
{
    "v71": {
        "created": "2019-02-10 14:27:52.000000",
        "id": "1684C60D-28B0-4818-A3EC-1F0C47981592"
    },
    "v72": {
        "created": "2019-02-11 07:54:52.000000",
        "id": "E9D7103D-1789-475E-A3EE-9CF18F51ACA4"
    }
}
```

14.2.3 Restoring a backup

Restoring latest version:

```
$ bahub --config ~/.bahub.yaml restore some_local_dir latest
{"status": "OK"}
```

Restoring version by number:

```
$ bahub --config ~/.bahub.yaml restore some_local_dir v71
{"status": "OK"}
```

14.2. Basic usage 77

Restoring version by id:

```
\ bahub --config ~/.bahub.yaml restore some_local_dir 1684C60D-28B0-4818-A3EC- \ 1F0C47981592 {"status": "OK"}
```

14.2.4 Recovery from disaster

In case you need to quickly recover whole server/environment from backup - there is a **RECOVERY PLAN**. A recovery plan is:

- List of backups to restore (names from section "backups")
- Policy of recovery (eg. recover everything, or stop on failure)

```
# Recovery plans
   Restores multiple backups in order, using single command
#
#
   Possible values:
#
      policy:
#
        - restore-whats-possible: Ignore things that cannot be restored, restore what_
\hookrightarrow is posible
#
        - stop-on-first-error: Restore until first error, then stay as it is
#
recoveries:
    default:
        policy: restore-whats-possible
        definitions: all
    plan_2:
        policy: stop-on-first-error
        definitions:
            - local_command_output
```

```
$ bahub --config ~/.bahub.yaml recover default
```

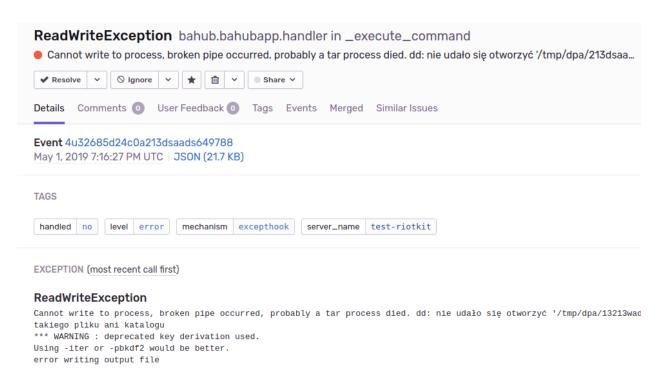
14.2.5 Making a snapshot of multiple services at once

Snapshot works exactly in the same way as **recovery from diaster**, but it's inverted. Instead of downloading a copy, it is actually uploading.

NOTICE: Be very careful, as this is a single command to backup everything, remember about the backups rotation

```
[2019-04-01 07:17:43,052][bahub][DEBUG]: shell(set -o pipefail; sudo docker exec _
→test_1 /bin/sh -c "tar -czf - \"/etc\" "| openssl enc -aes-128-cbc -pass_
[2019-04-01 07:17:45,672][bahub][DEBUG]: Request: https://api.backups.riotkit.org/
[2019-04-01 07:17:45,672][bahub][DEBUG]: response({"status":"OK", "error_code":null,
→"version":1, "creation_date":{"date":"2019-04-01 05:17:45.492490", "timezone_type":3,
→ "timezone": "UTC"}, "file":{"id":110, "filename": "cd06f449fdtest-v2"}}, "collection":{
→backup_version_size":2000000000, "max_collection_size":8000000000, "created_at":{ "date
→":"2019-03-24 21:29:14.000000","timezone_type":3,"timezone":"UTC"},"strategy":
→ "delete_oldest_when_adding_new", "description": "TEST", "filename": "test"}})
[2019-04-01 07:17:45,673] [bahub] [INFO]: Finishing the process
  "failure": [],
  "success": [
     "test"
```

14.3 Monitoring errors with Sentry



Bahub uses shell commands to take some data, pack it and encrypt. What if any of those commands will fail? What if there are no enough permissions? What if the directory does not exist? All of those are good reasons to have set up a monitoring.

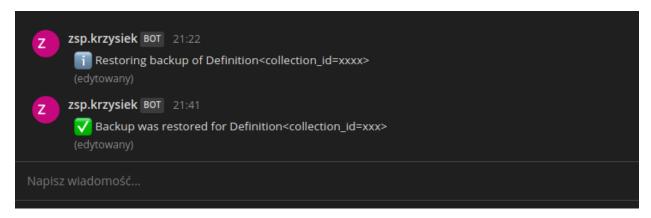
Almost each application failure can be catched and sent to analysis. Don't worry about the privacy, you can use your

own Sentry instance.

To enable the monitoring you need to have a ready-to-use Sentry instance/account and a error_handler configured in Bahub.

```
error_handlers:
    remote_sentry: # name it as you want
        type: sentry
        url: "https://some-url"
```

14.4 Slack/Mattermost notifications



Each event such as upload success, restore success, or a failure can emit a notification.

```
notifiers:
    mattermost: # name it as you want
        type: slack # compatible with Slack and Mattermost
        url: "https://xxxxx"
```

14.5 Setup

Bahub can be running as a separate container attached to docker containers network or manually as a regular process. The recommended way is to use a docker container, which provides a working job scheduling, installed dependencies and a warranty that it will work.

14.6 Using docker container

There exists a repository **quay.io/riotkit/bahub**, for testing purposes you can pick "latest-build" tag, for production it is recommended to pick a fixed version. Example installation on the simplest construction - docker-compose is placed in "examples/client" directory in the repository.

Running a demo

We prepared a demo environment that has a Bahub client, a Redis container and the File Repository server in one network. It's recommended to have client and server separated on different servers.

Take a look around, and create your own docker-compose, kubernetes or plain docker setup basing on our demo configuration.

```
git clone https://github.com/riotkit-org/file-repository
cd file-repository/examples/client
make start
make sh
```

14.7 Without docker

Use Python's PIP to install the package, and run it.

```
pip install bahub
bahub --help
```

14.7. Without docker 81

Shell access

File Repository usage can be automated using shell commands. There are not so many commands, but basic usage could be automated using scripts.

It is very helpful to create initially all required tokens and collections, so you can connect applications quickly and start doing backups or just file storing.

15.1 Introduction

Symfony Console is accessible in the main directory under ./bin/console When using docker you need to get into the container shell to execute it, in our example server environment you need to execute make sh to get into the server's shell.

```
./bin/console backup:create-collection -d "Some test collection" -f "backup.tar.gz" - \Rightarrowb 4 -o 3GB -c 15GB
```

If something is not working as expected, there is an error and you would like to inspect it, then please add a "-vvv" switch to increase verbosity.

15.2 Docker container concept

Our container allows to execute commands during startup to help you with the application setup. With this feature you can create expected tokens, collections on application startup without need to send any HTTP requests or even log in to the shell. It's an automation you will love.

Example in docker-compose syntax

```
version: '2.3'
services:
    filerepository:
        image: quay.io/riotkit/file-repository:${FILE_REPOSITORY_VERSION}
```

```
environment:
           # With this token you can do everything
           SECURITY_ADMIN_TOKEN: "4253f6e5-5c0b-4888-8027-d36bf02eee04"
           # Create a two backup collections, so right after startup you can run a
→backup, WHY NOT? :-)
           # Please notice, that you can easily use there environment variables
           POST_INSTALL_CMD:
               ./bin/console backup:create-collection --ignore-error-if-exists --max-
\rightarrowbackups-count=5
                    --max-one-version-size=10mib --max-collection-size=2gib --
→strategy=delete_oldest_when_adding_new
                    --filename=postgres.sql.gz --id=3dfa4ea9-1cec-4e24-b773-
→1cefb9c112c2;
                ./bin/console backup:create-collection --ignore-error-if-exists --max-
⇒backups-count=5
                    --max-one-version-size=5kib --max-collection-size=50kib --
→strategy=delete_oldest_when_adding_new
                   --filename=postgres-single-db.sql.gz --id=${COLLECTIONS_POSTGRES_

SINGLE_DB_ID;
```

15.2.1 Technical shell commands

For better experience and less frustration we implemented a few commands that improves deployment and overall application maintenance. Also the Symfony Framework, that we use implements tons of useful shell utilities worth looking at.

Wait for database

```
# wait for application's database to be ready
./bin/console health:wait-for:database --timeout=300
```

Wait for other File Repository instance to be up

Health check

Does same thing as HTTP health check endpoint, but does not require authorization.

```
./bin/console health:check
```

Dump database connection information

Very useful in debugging configuration of the database. The command dumps already parsed parameters of the Doctrine DBAL driver.

```
$ ./bin/console doctrine:connection:info
==> Parameters:
^ array:12 [
 "driver" => "pdo_pgsql"
  "charset" => "UTF8"
  "default_dbname" => "rojava"
  "dbname" => "rojava"
  "host" => "/var/run/postgresql"
  "password" => "rojava"
  "user" => "riotkit"
  "port" => "5432"
  "path" => "./var/data.db"
  "driverOptions" => []
  "serverVersion" => "11"
  "defaultTableOptions" => array:2 [
   "charset" => "UTF8"
   "collate" => "pl_PL.UTF8"
 ]
]
==> Database:
^ "rojava"
==> Platform:
^ "postgresql"
```

General guide for Administrators, DevOps and Developers

There is a general guide on how to maintain a backup server, what is the common approach to setup a server from Riotkit template and more.

The guide is on a separate repository: https://github.com/riotkit-org-education/guide

 $Check\ also\ our\ RiotKit\ Education\ organization\ at\ https://github.com/riotkit-org-education\ ,\ where\ we\ teach\ basic\ and\ mid-advanced\ things.$



From authors

Project was started as a part of RiotKit initiative, for the needs of grassroot organizations such as:

- Fighting for better working conditions syndicalist (International Workers Association for example)
- Tenants rights organizations
- Various grassroot organizations that are helping people to organize themselves without authority

Technical description:

Project was created in *Domain Driven* like design in PHP 7, with Symfony 4 framework. There are API tests written in *Postman* and unit tests written in *PhpUnit*. Feel free to submit pull requests, report issues, or join our team. The project is licensed with a MIT license.

RiotKit Collective