

7. Integrity and Replication

Data integrity checking is performed primarily by the Drupal framework (through its Schema API) in conjunction with its database back-end, MariaDB: CRUD operations are handled by the Schema API, offering an abstraction layer for database operations on PECE/Drupal data structures, and the database server guarantees integrity through ACID (atomicity, consistency, isolation and durability) conditions for all data transactions. For automatic checking the integrity of database tables, we use the extension module “dba” which allows for checking, reporting, and repairing data corruption on a regular basis.

Data replication can be handled in many ways on PECE: 1) automated replication between production, testing, and backup instances for redundancy and/or performance (for advanced PECE administrators using our VM distribution: we discuss this configuration in the “PECE Technical Specification” document); 2) scheduled, automated server “snapshot” generation performed by the hosting service company to save the state of a particular instance; and last but not least 3) using PECE Open API to replicate the data of a particular instance. This last option allows for easy integration with large-scale data repositories as described in the section on “Metadata extraction” of this document. For administrators with *nix expertise, replication is also conveniently done with Drush (and batch operations using shell scripting).

```
# Replicating all the artifacts of a particular type, i.e. "images"
$ drush ne-export -t images -f images_output.txt
```

```
# Replicating all the artifacts of all types
$ for i in {images, text, audio, video, etc.}; \
    do drush ne-export --type $i >> "$i".output.txt; \
    done
```

```
# Importing all the artifacts of a particular type
$ drush node-export-import --file=$filename.output.txt
```

```
# Export and import the entire instance for replication/redundancy
$ drush archive-dump default --destination=PECE.tar.gz
$ drush archive-restore PECE.tar.gz
```

```
# Export and import the database only
$ drush sql-dump > PECE_db.sql
$ drush sql-cli < PECE_db.sql
```

This command returns all the images with their respective metadata for replication purposes. In order to replicate binary files, it is necessary to also execute `wget` if replication of the PECE Image artifacts is successful. Please observe that checksum verification for binary files is currently not supported, it is a planned feature for PECE version 2.0:

```
# Replicating all the artifacts of a particular type, i.e. "images"
# Copying all the respective public binary image files as well
$ drush ne-export --type image >> images_output.txt && \
    wget --no-certificate -r -ll -A "gif, jpg, png, svg" \
    https://your-domain.org/sites/default/files/
```

Instruction for developers: replication depends on the contributed module “node-export” (7.x-3.0). Database check, report, and repair functionalities rely on the module “dba” (7.x-1.0). The script for automatic checking of database table integrity can be installed via admin interface at: “administer >> database >> run script”. The script “PECE_dba_table_repair.txt” can be found in the appendix.