

[EN] 12. The Mechanisms of Consistency Control of the dLibra System

The dLibra system periodically carries out certain activities which make it easier to control the consistency of data stored in the system. If the system encounters an inconsistency and cannot solve that problem on its own, it communicates the fact to the system administrator. The following activities are carried out in the dLibra system to control the consistency of the system:

- Publication file consistency check – the dLibra system checks periodically if publication files have been modified without the system knowing about it.

When files are entered in the digital library, the dLibra system calculates so-called checksums for them. Later, the system periodically (by default, every night) verifies if the files it manages still have the same [checksums](#) as when they were sent to the system. If there are few files in the digital library, then the system can verify the content of all of them every night. If, on any night, there are too many to verify their content in the time allotted per night, the system checks those files which have been verified the longest time ago. Each verification ends with a summarizing entry in the server logs, for example:

```
INFO: Consistency check finished. Took 24423717 ms.
Not all versions (only 780) were checked.
24 of total 780 compared versions and stored checksums is probably damaged.
```

In this case, we can see that the consistency verification took exactly 24423717 ms, that is, about 6 hours and 47 minutes. During that time, the system was not able to verify all files – only 780 were checked. From those 780 files, the checksums of 24 were different than the model checksums recorded in the database. During a consistency verification, for every file the checksum of which differs from its model checksum, an additional message is recorded in the logs of the server. The system also records situations in which it is not possible to access a file, for example:

```
ERROR: Problem while calculating digests
java.io.FileNotFoundException:
F:\dlibra\content\files\1.dir\5.dir\22.dir\696.grp\664.pub\76075.fil\76283.ver
(The system cannot find the file specified)
```

The path in the message indicates which publication and file are meant. In this example, the file is .fil, with ID 76075, or – more precisely – its version (.ver) with ID 76283. That file is a part of publication .pub with ID 664, which, in turn, is a part of group publication .grp with ID 696. That publication is in the path of three catalogs (.dir) with IDs: 1, 5, and 22. The provided information makes it possible to find the file from the level of the Editor Application. The identifier of the publication can also be used to access it through the website. For that purpose, the main library address should be complemented with /publication/<publication id>.

- Search index consistency check – the dLibra system checks if search indexes are consistent with indexed object metadata. If they are not, the object in question is re-indexed (to remove the inconsistency).