



DIGIT
Unit B1

FOSSA WP3 - Deliverable n. 3
**Selection of tools to perform periodic inter-
institutional inventories of software
assets and standards**

Date: 25/04/2016
Doc. Version: Final

TABLE OF CONTENTS

TABLE OF FIGURES.....	3
TABLE OF TABLES.....	3
1. DELIVERABLE OVERVIEW	4
1.1. Suggested Readers	Error! Bookmark not defined.
2. APPROACH TO INVENTORY AND SELECTED TOOLS.....	5
2.1. Selection of ETL tools	6
2.2. Selection of relational databases	6
2.3. Selection of Business Intelligence tools	7
2.4. The selected tools in the inventory architecture	7
3. MAIN FEATURES OF THE SELECTED TOOLS.....	9
4. APPENDIX – ABBREVIATIONS AND ACRONYMS	10

TABLE OF FIGURES

Figure 1 - Families of tools for the execution of software components inventory in the various approaches.....	5
Figure 2 - Selected tools for each inventory architecture layer	8

TABLE OF TABLES

Table 1 - Summary of tool ranking by layer	6
Table 2 - Features of the selected inventory tools	9

1. DELIVERABLE OVERVIEW

Taking as a basis the Deliverable 2 of WP3, “*Proposal of tools to perform periodic inter-institutional inventories of software assets and standards*” of the FOSSA Pilot Project, the present deliverable finalizes the choice of the tools to be used in the architecture described in DLV2.

In particular, this deliverable describes the logical process for the selection and the reasons underlying the choice of each tool.

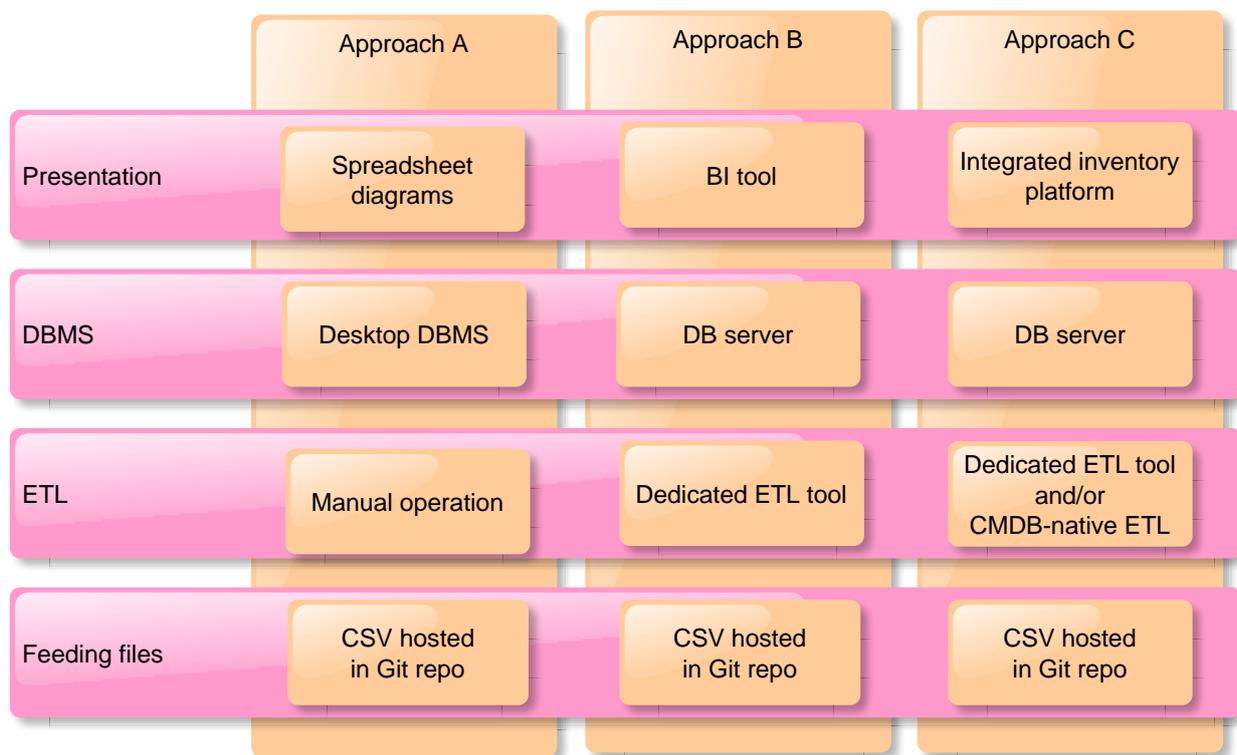
Finally, for each of the selected tools further information is provided, particularly on the respective applicable versions and technical requirements.

2. APPROACH TO INVENTORY AND SELECTED TOOLS

In Deliverable 2, three possible approaches to the execution of Step 1 (Software component Inventory) of the OSS Inventory approach have been identified and scrutinized against the nature of the data sources provided by the European Institutions and the constraints of the Pilot Scenario.

For each of the three approaches, families of tools for each layer of the Software Component Inventory architecture have been identified as per the figure below.

Figure 1 - Families of tools for the execution of software components inventory in the various approaches



The possibility to automatize the approach, while coping with the requirement to maintain the implementation effort in line with the constraints of the Pilot Scenario, led to recommend and choose Approach “B”. However, the implementation of a local database has been preferred to a server database in consideration, again, of its easier implementation combined with adequate coverage of the requirements.

Subsequently, for each family of tools recommendations have been formulated. This starts from a long list of potential candidate tools, filtered through appropriate selection criteria in order to obtain a shortlist that is submitted to a detailed scoring and ranking based on further specific criteria.

In line with the nature of the FOSSA project, only Open Source tools have been shortlisted and submitted to a detailed scoring and ranking.

Exception has been made for Relational Database Management Systems (RDBMS), as they can now rather be considered as a commodity having lost differentiation among the various marketed solutions. Therefore, only a longlist has been defined, within which DIGIT has selected the preferred option, among the open source tools presented in the longlist.

The summary of shortlists with the pertinent ranking are reported in the table below. For the detailed scoring and evaluation of each tool please refer to DLV2, “Proposition of tools to perform periodic inter-institutional inventories of software assets and standards”.

Table 1 - Summary of tool ranking by layer

Layer	Recommended tools
ETL	1°: Talend Open Studio / 2°: Pentaho Data Integration / 3°: Jasper Soft ETL / 4°: KNIME / 5: Apatar
RDBMS	n.a.
Business Intelligence	1°: Pentaho BI Suite / 2°: Eclipse BIRT / 3°: Jasper Report / 4°: Rapid Miner

The proposed shortlists and the pertinent ranking have been discussed with DIGIT B1, for the final selection of the tools.

The selection process has been managed entirely through discussions between DIGIT B1 and the Consortium and finalized in a dedicated meeting.

2.1. Selection of ETL (Extract-Transform-Load) tools

The evaluation of the shortlisted ETL tools has resulted into a quite close scoring of Talend Open Studio and Pentaho Data Integration.

As it has been pointed out in a further detailed analysis¹, although such tools are in principle substantially interchangeable, the point of advantage for Talend compared to Pentaho is however essentially due to the fact that Talend’s API set looks more complete and powerful than Pentaho.

While Pentaho is an absolutely “Java-oriented” platform, Talend appears to be more flexible regarding integration with web services, especially by the compliance with JSR-224/ JAX –WS standard, enabling the integration with XML Web services via Java API.

In addition, although the acquisition of Pentaho by Hitachi Data Group did not change the “open” nature of the product, its future road-map is not yet available.

Based on this further assessment, the choice of Talend Open Studio has been validated.

2.2. Selection of relational database management system (RDBMS)

As mentioned above, the selection of the relational database for the inventory architecture has been made directly from the longlist below:

¹ See DLV2, page 30.

- IBM-DB2, Sybase-ADS, Apple-FileMaker, MariaDB_Community-MariaDB, Microsoft-SQL_Server, ORACLE-MySQL, ORACLE-RDB, SAP_HANA-SAP_AG, PGDG-PostgreSQL, Sybase-SQL_Anyware, Teradata-Teradata.

Out of this longlist, only open source tools, once again, have been taken into consideration. Based on this, DIGIT B1 has directly selected MariaDB, which is actually used by some important open source projects such as Red Hat OpenStack, Fedora, openSUSE and Arch Linux². Furthermore, MariaDB is a fork based on the source code of MySQL, which is actually the most popular open source RDBMS, but whose open source nature has been questioned following its acquisition by Oracle.

2.3. Selection of Business Intelligence (BI) tools

The Pentaho BI suite has been recommended over Eclipse BIRT, another leader in this segment, in consideration of the higher scalability of the former compared to the latter³. DIGIT B1 has validated such recommendation.

An alternative to such choice would have been the visualisation and production of reports directly in the CMDB tool. However, such possibility has been excluded due to the preference for a local relational database.

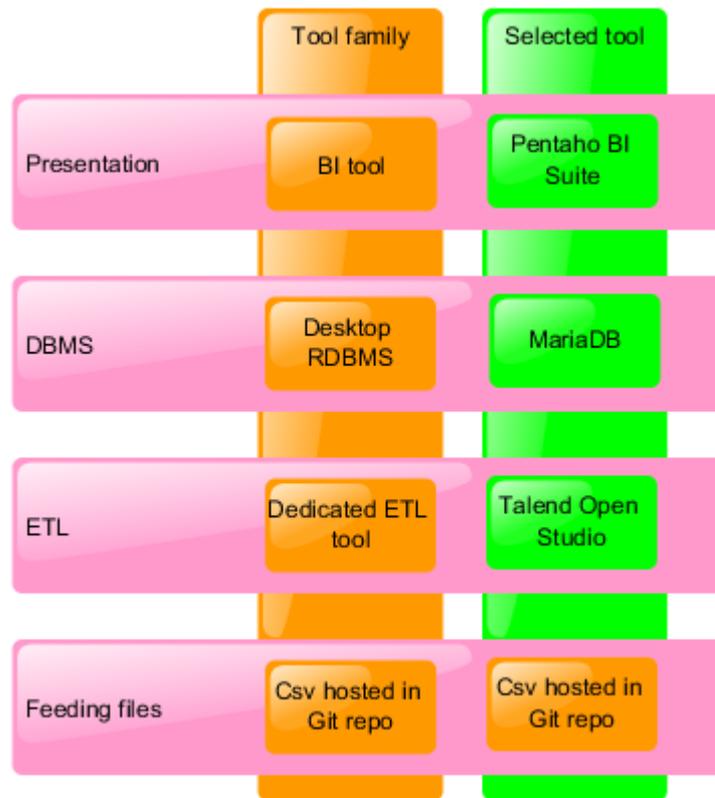
2.4. The selected tools in the inventory architecture

The result of such analysis is summarized in the picture below, showing the selected tool for each layer of the architecture.

² <http://linuxaria.com/article/mariadb-is-conquering-the-desktop-distributions>; <https://mariadb.com/blog/webinar-mariadb-red-hat-how-persistence-pays-openstack-and-you>;

³ DLV2, page 34.

Figure 2 - Selected tools for each inventory architecture layer



3. MAIN FEATURES OF THE SELECTED TOOLS

The table below indicates the main features of the selected tools. For each of them, the most recent stable release has been identified for the installation and set-up.

Table 2 - Features of the selected inventory tools

Function	Software name	Release	Size	Software origin
ETL	Talend Open Studio for Data Intergration	6.1.1	607MB zip file	Zip file from sourceforge
DBMS	MariaDB	10.1	387MB docker image	Docker hub
BI	Pentaho Business Intelligence Server	6.1.0.1-196	1796MB docker image	Docker hub + zip file from sourceforge
	Java (Pentaho dependency)	openJDK 1.7.0		
Runtime	Docker engine	1.11.0	14MB	Dockerproject.org
Orchestration	Docker compose	1.7.0	7,6MB	github.com/docker/compose

4. APPENDIX – ABBREVIATIONS AND ACRONYMS

FOSSA	Free and Open Source Software Auditing
WP	Work Package
DLV	Deliverable
csv	Comma-Separated Values (file format)
DBMS	Data Base Management System
ETL	Extract, Transform, Load
RDBMS	Relational Data Base Management System
BI	Business Intelligence