

Building an Inventory for Open Source Software used by European Public Services

Context and Guidelines

February 2021

EUROPEAN COMMISSION Directorate - General Informatics Directorate B – Digital Business Solutions DIGIT B3 – Reusable Solutions

Help protect the open source software Europe runs on

ALL MODERN DIGITAL INFRASTRUCTURE A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003

All European Public Services use open source software to some extent. But do we know our most critical components?

Sadly, the answer is a clear NO... \rightarrow we don't even have an inventory!

This <u>ISA² Sharing and Reuse Action (2016.31)</u> initiative aims to fix this, and build an initial inventory.

That will help to establish which software is **most critical** to European Public Services \rightarrow so that we can collectively secure it, sustain it and protect it.

We can't do this without you, and we encourage you to participate by sending us *anonymised data*. These slides will explain how you can do that.

You may contact the following for additional information.

- 1. Trasys International: Ksenia BOCHAROVA (<u>eu-oss@trasysinternational.com</u>) (Note: Trasys International is our subcontractor on this project)
- 2. DG for Informatics Unit B3: Miguel DIEZ-BLANCO, Saranjit ARORA and Evangelos TSAVALOPOULOS (<u>DIGIT-OSPO@ec.europa.eu</u>)









Collect Open source software usage data from 30 European Public Services (National, Regional, Towns)

For an initial 30 European Public Services





Control Control Soft Exercises Control Contrel Control Control Control Control Control Control Control Contr

Where do we start?

The European Commission and Council have created three inventories so far, their experience may help you get started.

Step 1: Identify your data sources	We identified workstations, servers, virtual machines, cloud servers
Step 2: Prepare for final output	We identified the fields which were needed, which were optional, and discussed what was needed, paying special regard to what was sensitive data.
Step 3: Data extraction scripts	We wrote multiple scripts for extracting data from systems such as SCCM and Satellite. For some systems it was quite a challenge, so extraction was done in multiple steps.
Step 4: Extract the data	Often we could not distinguish between open source and proprietary software, so we exported both. This was cleansed later. There were also multiple records, e.g. one for each instance, this helped to later obtain a count of the number of instances.
Step 5 : Anonymise data	We got rid of user names, computer names, building locations etc See guidelines

Previous experience

Step 6: Package and send \rightarrow



What if?

Q&As - Answering some questions you may have.

What if we can't extract all our data?	No worries, send us what you can. Our European picture will develop slowly.
We don't understand the needed output?	Call us, we can help you.
We can't separate open source and proprietary?	Don't worry, we will separate it using our knowledge base from earlier projects.
I am worried about security, sensitivity.	We are too. Please see our guidelines on the next page.
What will we get out of this?	You will get (i) your own inventory, (ii) summary European level inventory (iii) an identification of which software is the most critical for running European Public Services.



Guidelines - Anonymisation of data

- User names and file paths: Please avoid showing user names and file paths where the software is
 installed e.g. in users' private folders, like C:\Users\jonesd\...\KeePass.exe. At least the part containing
 user id needs to be stripped out or replaced with a generic string. If the data provider would like to still
 be able to differentiate software from different users, without having clues of who they are, they can
 apply a hashing function (with salt) to the path string or just the user id;
- **Software versions:** This information can be considered sensitive, because they can be exploitable, but only if linked to an organisation;
- Other points e.g. managing organisation, installation path and years of use might be considered sensitive. However, do note that the installation path sometimes provides information about the name of the open source software, because the install file could just be called install.exe, and path could provide a clue e.g. /h/pathname/software_name/install.exe
- A general warning: Please do not include sensitive information about users, machines, locations, IP and MAC addresses or anything else that you may consider to be confidential.



Thank you

Contacts

- 1. Trasys International: Ksenia BOCHAROVA (<u>eu-oss@trasysinternational.com</u>) (Note: Trasys International is our subcontractor on this project)
- 2. DG for Informatics Unit B3: Miguel DIEZ-BLANCO, Saranjit ARORA and Evangelos TSAVALOPOULOS (DIGIT-OSPO@ec.europa.eu)



© European Union, 2021

Unless otherwise noted the reuse of this presentation is authorised under the <u>CC BY 4.0</u> license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Credits: the image used in slide 2 is a courtesy of XKCD.com

