Preview

# Preview

edit

## GENERAL INFORMATION

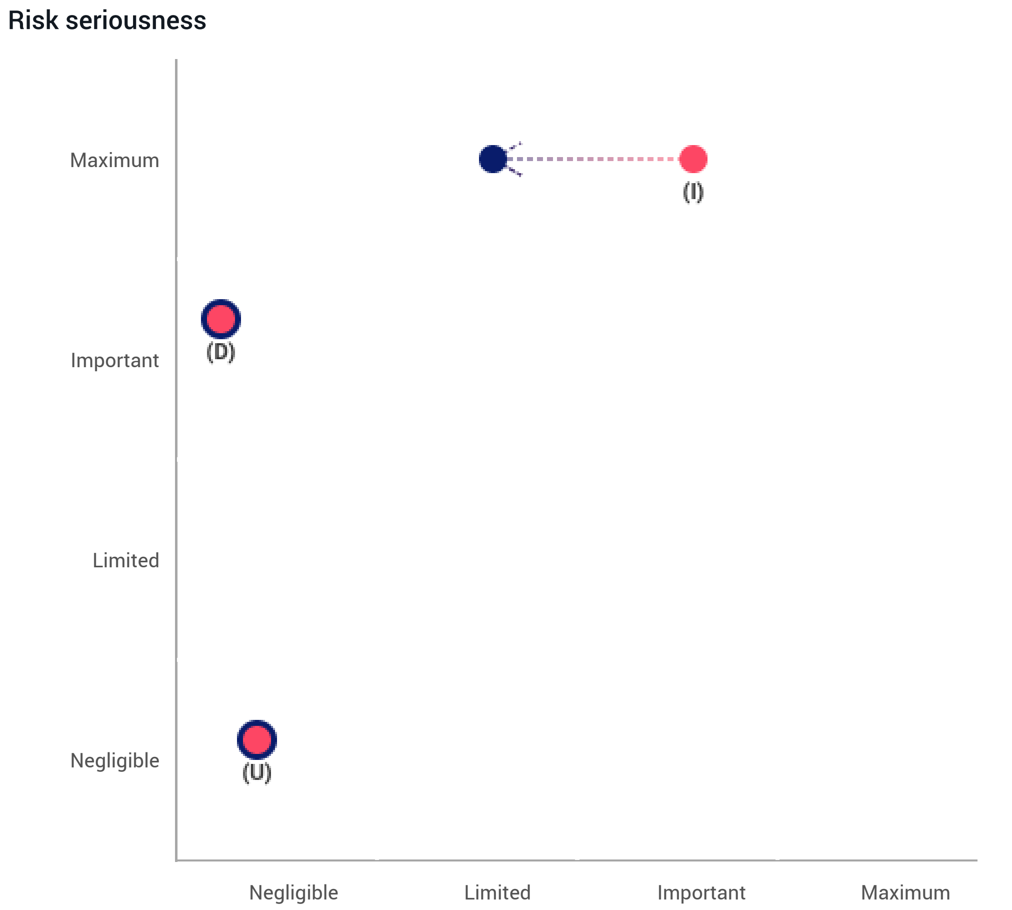
100%

**Editing :** João, Freitas
**Evaluation :** Rui, Gonçalves
**Validation :** João, Vilela

**Status :** Simple validation

# Validation

## Risk mapping



Risk likelihood

- Planned or existing measures
- With the corrective measures implemented
- (I)llegitimate access to data
- (U)nwanted modification of data
- (D)ata disappearance

12/31/21

# Validation

## Action plan

### Overview

Fundamental principles | Planned or existing measures

Fundamental principles

- Purposes
- Legal basis
- Adequate data
- Data accuracy
- Storage duration
- Information for the data subjects
- Obtaining consent
- Right of access and to data portability
- Right to rectification and erasure
- Right to restriction and to object
- Subcontracting
- Transfers

Planned or existing measures

- Encryption
- Anonymisation
- Logical access control
- Paper document security
- Network security

Risks

- Illegitimate access to data
- Unwanted modification of data
- Data disappearance

Improvable Measures
Acceptable Measures

## Fundamental principles

No action plan recorded.

## Existing or planned measures

### Anonymisation

**Action plan / corrective actions :**
Re-anonymize the dataset to grant that the privacy levels are bigger than the utility levels, but at the same time the analytic tools can provide good results.
**Evaluation comment :**
Although anonymized, there is still a risk of information disclosure if it revealed to an attacker, as the anonymization had to take into account the needed utility levels for the prediction tool to yield good results.
**Expected date of implementation :** 1/31/22
**Responsible for implementation :** Data Scientists

### Logical access control

**Action plan / corrective actions :**
Deploy GPS Security Mechanism to track USB Yubi keys of data scientists
**Evaluation comment :**
Deploy a GPS security mechanism that allows the track of the USB Yubi keys. This will reduce the likelihood of the lose of proprietary devices.
**Expected date of implementation :** 1/31/22
**Responsible for implementation :** CISO, CEO

### Paper document security

**Action plan / corrective actions :**
Deploy a IDS rule to block unrecognized e-mail domains.
**Evaluation comment :**
Deploy a IDS rule to block unrecognized e-mail domains. This will prevent social-engineering attacks by outsiders.
**Expected date of implementation :** 1/31/22
**Responsible for implementation :** CISO

**Responsible for implementation :** CISO

### Network security

**Action plan / corrective actions :**
Setup anti install mechanism on employees computers, to prevent install of malware.
**Evaluation comment :**
Setup anti install mechanism on employees computers, to prevent install of malware. Employees that need to install any software will have to check with the CISO before, so he can validate if it is malware or not.
**Expected date of implementation :** 1/31/22
**Responsible for implementation :** CISO

## Risks - Illegitimate access to data

**Action plan / corrective actions :**
- Deploy GPS security mechanism to reduce likelihood of Proprietary Device Steal;
- Deploy IDS rule to prevent likelihood of Social Engineering via e-mail;
- Block installation of software that is not present on the supported software list;
- Re-anonymize dataset to increase privacy but reduce utility levels, as a measure to reduce the likelihood of correlating the anonymized dataset

**Evaluation comment :**
Reverted likelihood risk to Limited
Added action plan
**Expected date of implementation :** 2/28/22
**Responsible for implementation :** CISO, Data Scientists
Taking into account the action plan, how do you re-evaluate the seriousness of this risk (Illegitimate access to data)? **Maximum**
Taking into account the action plan, how do you re-evaluate the likelihood of this risk (Illegitimate access to data)? **Limited**

## Risks - Data disappearance

**Action plan / corrective actions :**
- Improve hardware security mechanisms to protect access to external hard drive in the company facilities.

**Evaluation comment :**
Added action plan
**Expected date of implementation :** 2/28/22
**Responsible for implementation :** CISO
Taking into account the action plan, how do you re-evaluate the seriousness of this risk (Data disappearance)? **Important**
Taking into account the action plan, how do you re-evaluate the likelihood of this risk (Data disappearance)? **Negligible**

# Validation

## TO TRANSLATE - DPO and data subjects opinion

### DPO's name

João Freitas

### DPO's status

The treatment could be implemented.

### DPO's opinion

PIA is valid and actions plans were defined.

### Search of concerned people opinion

Concerned people opinion was requested.

### Concerned people opinions

US Census Bureau

### Concerned people statuses

The treatment could be implemented.

### Concerned people opinions

"Dataset security will increase with the proposed action plan, which means that the privacy of the dataset will too"

## Context

Overview

### What is the processing under consideration?

The 1994 US Data Census dataset is part of one of the machine learning projects of Têrepê, a Portuguese software house company. The purpose of the project is to create a tool that predicts whether or not an American citizen has an income bigger than $50,000 per year. For that, the US Census Bureau provided the dataset for study.

The principal stateholders for the project is the US Census Bureau and Têrepê CISO and DPO.

### What are the responsibilities linked to the processing?

The team allocated for the project is constituted by 5 software engineers, 3 data scientists, 1 product owner and 1 UI designer. They are responsible for developing and mantaining the project for a span of 3 years.

The CISO is responsible for the company overall information security. Together with DPO, they are finding every privacy concern a flaw which might be considered a threat for the user privacy.

### Are there standards applicable to the processing?

Europe:
GDPR - General Data Protection Regulation

America:
California Privacy Protection Act (CalOPPA)
California Consumer Privacy Act (CCPA)

**Evaluation : Acceptable**

## Context

Data, processes and supporting assets

### What are the data processed?

The data processed characterize an individual biological and identity status (age, sex, race, native country, relationship, marital status), social status and education (work class, occupation, weekly work hours, education), net worth (capital gain, capital loss).

The original data was given to the company in an hard-drive, flown by the US Census Bureau. The company keeps the hard-drive in a secure storage inside the company and a copy of the dataset inside the company on-premise servers. Developers can access data via VPN in any location external to the company.

The data is to be stored for a span of 2 years (December 2021 - December 2023), which corresponds to the duration of the project.

### How does the life cycle of data and processes work?

An overview of the project lifecycle can be found in the attachements in the form of a Swimlane diagram (trp-data-census-pii-primary-flows.png).

### What are the data supporting assets?

The company only allows the use of the following technologies and platforms to work with:

- Windows 10, version 1909
- Ubuntu x64 20.04
- Visual Studio Code, version 1.63
- Github Enterprise 2021
- Microsoft Office Tools (Outlook, Word, Powerpoint, Excel, Teams) 2021
- Redis Database, version 6.0.9
- PostgreSQL, version 13.4
- OpenVPN, version 2.5.5

The company does not enforce the use of any programming languages as these are to be decided by the team working on the projects. Any software that team members use that is not present in the supported list is not covered by the company if any issue occurs and thus should be the worker responsibility to pay for damages caused.

**Evaluation : Acceptable**

# Fundamental principles

## Proportionality and necessity

### Are the processing purposes specified, explicit and legitimate?

Processing of the dataset is required to create machine learning tools. For this, analytic tools ingest the dataset in order to classify it and create learning models for the prediction in context. Datascientists are the entities responsible to access the data, and use the analytic tools.

**Evaluation : Acceptable**

### What are the legal basis making the processing lawful?

No direct identifiers are present in the dataset, so datascientists cannot indicate which US citizen an entry in the dataset corresponds to. The dataset does have indirect and sensitive identifiers that can be used for data correlation and thus predict which US Citizen dataset entry corresponds to. To grant that only datascients are allowed to access the data, a dedicated server to access the data was created, which only the data scientists have the keys to access it.

**Evaluation : Acceptable**

### Are the data collected adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')?

The majority dataset attributes are required for building the prediction tool as these relate to the individual capital power. Attributes fnlwgt, education-num can be removed from the dataset as these are metadata attributes that are not benefitial for the prediction.

**Evaluation : Acceptable**

### Are the data accurate and kept up to date?

Data is considered accurate and up to date as it was provided by the US Census Bureau. Beside outliers and unecessary data removal, the dataset is not modified.

**Evaluation : Acceptable**

### What are the storage duration of the data?

Data is stored for the period of 2 years.

**Evaluation : Acceptable**

## Fundamental principles

Controls to protect the personal rights of data subjects

### How are the data subjects informed on the processing?

The US Census Bureau has made a public announcement on their website that the census data will be used for learning about the quality of life of citizens.

**Evaluation : Acceptable**

### If applicable, how is the consent of data subjects obtained?

When filling up the census, citizens were informed that their data will be kept privacy and used with care for learning purposes.

**Evaluation : Acceptable**

### How can data subjects exercise their rights of access and to data portability?

US Citizens can contact for the Chief FOIA Officer at 1-800-432-1494, or by email at pco.policy.office@census.gov. data requests.

**Evaluation : Acceptable**

### How can data subjects exercise their rights to rectification and erasure?

US Citizens can not request data removal from the US Data Census.

**Evaluation : Acceptable**

### How can data subjects exercise their rights to restriction and to object?

Filling the US Data Census is mandatory for US Citizens.

**Evaluation : Acceptable**

**Evaluation : Acceptable**

### Are the obligations of the processors clearly identified and governed by a contract?

US Census Bureau signed a NDA contract with all Têrepê employees, stating that for a period of 5 years, these could not reveal that they are working with the US Data Census or with the US Census Bureau, neither disclose information about the Census.

**Evaluation : Acceptable**

### In the case of data transfer outside the European Union, are the data adequately protected?

The only processing and storage operations performed by Têrepê are conducted in Portugal. Operations conducted in the US or outside the European Union by the US Census Bureau is not a concern of Têrepê.

**Evaluation : Acceptable**

# Risks

## Planned or existing measures

### Encryption

At Têrepê offices, Internet access is barried by a secure firewall that makes sure to encrypt any unencrypted connections with TLS 1.1 and 1.2. Outside Têrepê offices, access to the company infrastructure is done by using a VPN using the OpenVPN software. VPN tunnel is protected using IPSec IKE v2.

There is still a risk of disclosure of information by employees when accessing outside the company in unprotected Wi-Fi networks.

**Evaluation : Acceptable**

### Anonymisation

The US Data Census dataset has been anonymized before processing by analytic tools using the ARX software tool.

**Evaluation : Improvable**
**Action plan / corrective actions :**
Re-anonymize the dataset to grant that the privacy levels are bigger than the utility levels, but at the same time the analytic tools can provide good results.
**Evaluation comment :**
Although anonymized, there is still a risk of information disclosure if it revealed to an attacker, as the anonymization had to take into account the needed utility levels for the prediction tool to yield good results.

### Logical access control

Data scientists are the only employees who have access to the dataset. A dedicated server to access the dataset was created which can only be accessed using keys distributed to these employees in the form of USB Yubi Keys.

There is still a risk of gaining unauthorized access by internal employees or external attackers, if someone manages to grab/steal the USB Yubi Key.

**Evaluation : Improvable**

**Evaluation : Improvable**
**Action plan / corrective actions :**
Deploy GPS Security Mechanism to track USB Yubi keys of data scientists
**Evaluation comment :**
Deploy a GPS security mechanism that allows the track of the USB Yubi keys. This will reduce the likelihood of the lose of proprietary devices.

### Paper document security

Product Owners and UI Designers do weekly reports on the tool being created in order to refine its features. For this they use Microsoft Word to create a report, which is then translated in the .PDF format and distributed to board members for presentations and validations. Every document is stored in a secure storage and destroyed after two years.

There is still a risk of disclosure of information by employees if they share these report documents with those that are not on the meetings, either via Social Engineering attacks in E-mails or by talking about it.

**Evaluation : Improvable**
**Action plan / corrective actions :**
Deploy a IDS rule to block unrecognized e-mail domains.
**Evaluation comment :**
Deploy a IDS rule to block unrecognized e-mail domains. This will prevent social-engineering attacks by outsiders.

### Network security

CISO implented and deployed a network topology that involves an IDS, a DMZ and firewalls to secure and detect unauthorized accesses.

There is still a risk of disclosure of information if employee install malware or a compromised software.

**Evaluation : Improvable**
**Action plan / corrective actions :**
Setup anti install mechanism on employees computers, to prevent install of malware.
**Evaluation comment :**
Setup anti install mechanism on employees computers, to prevent install of malware. Employees that need to install any software will have to check with the CISO before, so he can validate if it is malware or not.

## Risks

### Illegitimate access to data

#### What could be the main impacts on the data subjects if the risk were to occur?

Breach of US Citizens net worth, Public damange, Breach of US Citizens Identity Information, Target of Marketing Companies, Target of Robbers, Political and Social-Economical Power to Russia and China

#### What are the main threats that could lead to the risk?

Social Engineering, Packet Sniffing, Supply-Chain Attacks, Malware, Proprietary Device Thief, Human Interaction

#### What are the risk sources?

Employees, Software

#### Which of the identified planned controls contribute to addressing the risk?

Encryption, Anonymisation, Logical access control, Paper document security, Network security

### How do you estimate the risk severity, especially according to potential impacts and planned controls?

Maximum, Breaching data of the citizens of the worlds largest country would be catastrophic, not only because it would put the country citizens identity and status at risk, but also because it would cause enormous fines to the company, Portugal, United States and it would given political advantage to rival countries such as China and Russia.

### How do you estimate the likelihood of the risk, especially in respect of threats, sources of risk and planned controls?

Important, It is unlikely that the risks may actually happen for two reasons: the company data processing is well secured and the dataset does not direct identifiers, meaning that it would require secret datasets from the US to correlate the data entries.

**Evaluation : Improvable**

**Action plan / corrective actions :**

- Deploy GPS security mechanism to reduce likelihood of Proprietary Device Steal;
- Deploy IDS rule to prevent likelihood of Social Engineering via e-mail;
- Block installation of software that is not present on the supported software list;
- Re-anonymize dataset to increase privacy but reduce utility levels, as a measure to reduce the likelihood of correlating the anonymized dataset

**Evaluation comment :**

Reverted likelihood risk to Limited

Added action plan

Taking into account the action plan, how do you re-evaluate the seriousness of this risk (Illegitimate access to data)? **Maximum**

Taking into account the action plan, how do you re-evaluate the likelihood of this risk (Illegitimate access to data)? **Limited**

## Risks

## Unwanted modification of data

### What could be the main impacts on the data subjects if the risk were to occur?

Public damange

### What are the main threats that could lead to the risk?

Human Interaction, Proprietary Device Thief, Malware

### What are the risk sources?

Employees

### Which of the identified controls contribute to addressing the risk?

Logical access control

### How do you estimate the risk severity, especially according to potential impacts and planned controls?

Negligible, It will not matter if the employees at Têrepê modify the data, as they are creating a predictive tool that is agnostic of the content of the dataset. The US Census Bureau will use whatever dataset they desire from the census

### How do you estimate the likelihood of the risk, especially in respect of threats, sources of risk and planned controls?

Negligible, N/A

**Evaluation : Acceptable**

**Evaluation comment :**

Revert impact and likelihood risk to Negligible

Taking into account the action plan, how do you re-evaluate the seriousness of this risk (Unwanted

Taking into account the action plan, how do you re-evaluate the seriousness of this risk (Unwanted modification of data)? **Negligible**

Taking into account the action plan, how do you re-evaluate the likelihood of this risk (Unwanted modification of data)? **Negligible**

# Risks

## Data disappearance

### What could be the main impacts on the data subjects if the risk were to occur?

Public damange

### What are the main threats that could lead to the risk?

Human Interaction, Malware, Packet Sniffing, Proprietary Device Thief, Social Engineering, Supply-Chain Attacks

### What are the risk sources?

Employees, Software

### Which of the identified controls contribute to addressing the risk?

Encryption, Logical access control, Network security

### How do you estimate the risk severity, especially according to potential impacts and planned controls?

Important, There is no impact if data is lost for the citizens as the US Census Bureau detains the primary source of it. However it would be really bad for the company to lose the dataset, as it would leave a really bad impression for the US Government. It would have a serious impact if that data would be lost in terms of a robbery, but that falls under the *Illegitimate access to data* risk category.

### How do you estimate the likelihood of the risk, especially in respect of threats, sources of risk and planned controls?

Negligible, N/A

**Evaluation : Improvable**

**Action plan / corrective actions :**

- Improve hardware security mechanisms to protect access to external hard drive in the company facilities.

**Evaluation comment :**

Added action plan

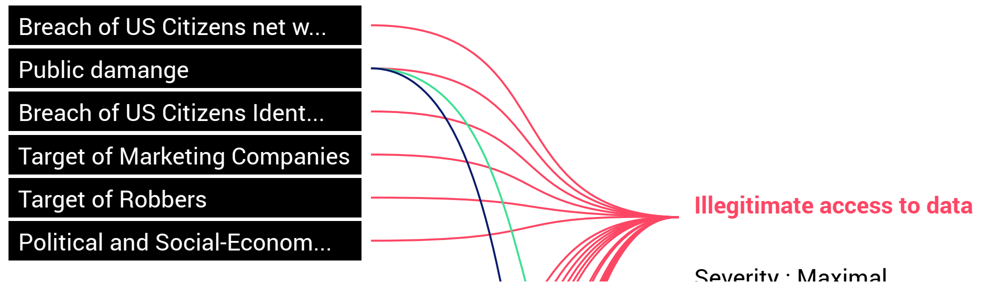
Taking into account the action plan, how do you re-evaluate the seriousness of this risk (Data disappearance)? **Important**

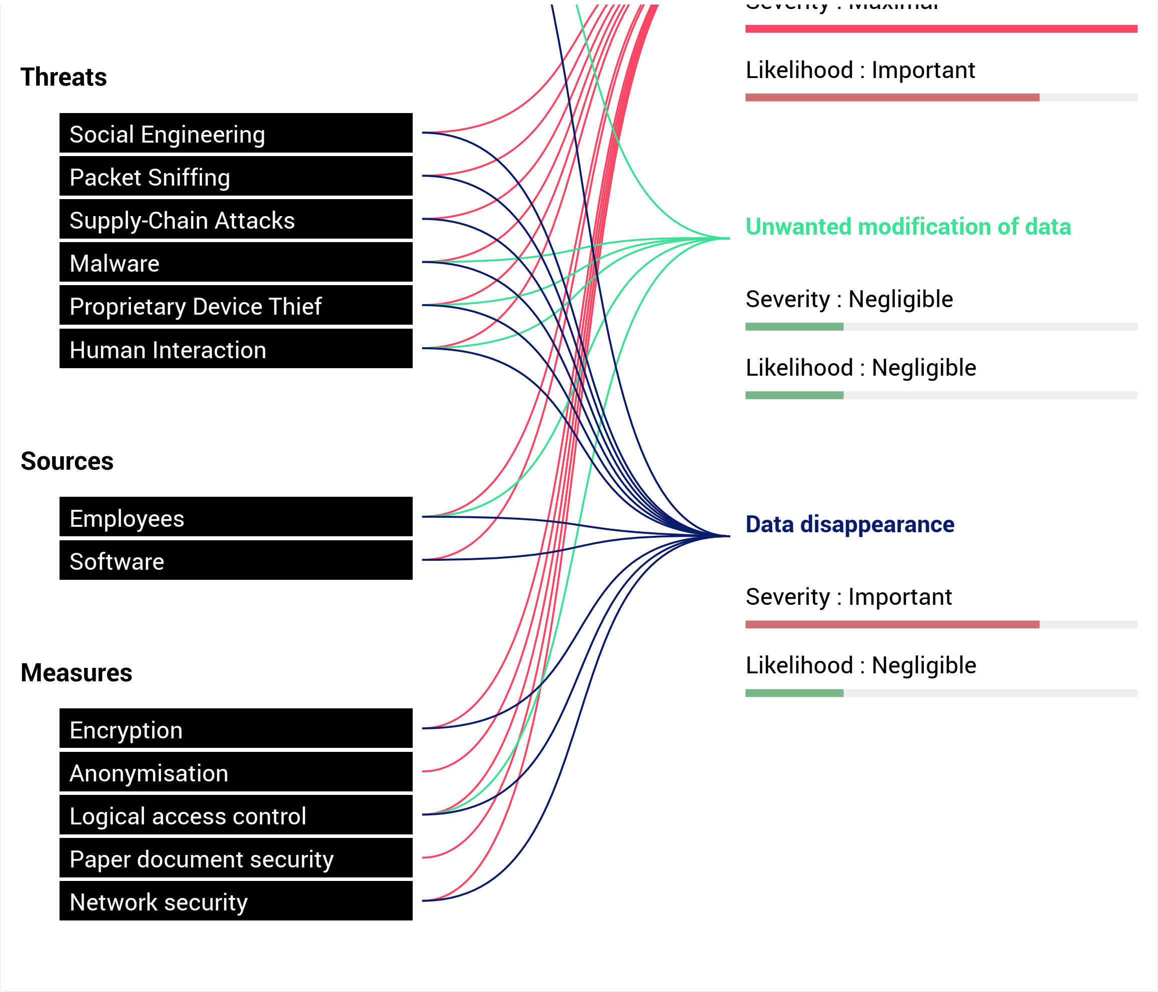
Taking into account the action plan, how do you re-evaluate the likelihood of this risk (Data disappearance)? **Negligible**

# Risks

## Risks overview

**Potential impacts**


Breach of US Citizens net w...
Public damange
Breach of US Citizens Ident...
Target of Marketing Companies
Target of Robbers
Political and Social-Econom...
Illegitimate access to data
Severity : Maximal



Threats

- Social Engineering
- Packet Sniffing
- Supply-Chain Attacks
- Malware
- Proprietary Device Thief
- Human Interaction

Sources

- Employees
- Software

Measures

- Encryption
- Anonymisation
- Logical access control
- Paper document security
- Network security

Severity : Maximal

Likelihood : Important

Unwanted modification of data

Severity : Negligible

Likelihood : Negligible

Data disappearance

Severity : Important

Likelihood : Negligible