

# Enrolment Session

---


*Centre d'Accès Sécurisé aux Données*



# Table of contents

---

- Presentation of the CASD
- Legal framework of data access
- IT
  - Infrastructure
  - Workflows
  - Demonstration
  - Terms and Conditions of Use
  - Information about access card issuing
  - Q&A
- Data
  - Anonymization techniques
  - Confidentiality rules
  - Input/Output
  - Data citation and return of publications
  - Q&A
- Support contacts
- Quiz



# The CASD

# The CASD

---

- **A non profit public interest grouping** composed of six members: GENES, INSEE, CNRS, HEC Paris, École Polytechnique and Banque de France
- **A secure infrastructure to access confidential data** which benefited from “Equipement d'excellence” (EQUIPEX) funding of PIA (Programme d'investissements d'avenir) 1<sup>st</sup> edition
  - **Our main mission:** organize and set up services of secure access to confidential data for users pursuing non-profitable research, study, evaluation or innovation purposes
    - Secondary mission: valorization of the technology in the private sector
  - **Our goal :** provide a highly secure access for accredited data users in the best possible work conditions while minimizing access costs

# The challenges of the secure data access

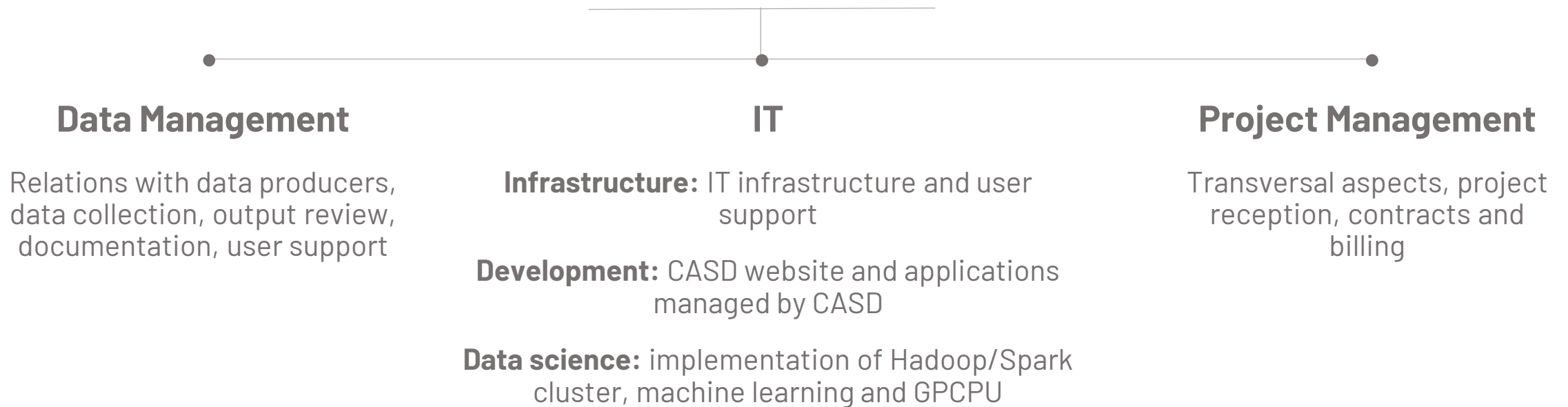
---

- **Security, a real challenge to allow data access:**
    - Ensuring a high level of security **in order for data producers to make more and more data available in confidence**
    - Strong authentication using biometrics
    - The enrolment session is mandatory and must be renewed every 6 years
    - A formal accreditation process before accessing the data
  - **Uses:**
    - To meet the needs of users in terms of work environment (software, configuration...)
    - A shared work environment between project members
  - **Fair treatment of all users**
- ➔ Thanks to this device, for example, tax data access was made possible in 2013
- ➔ The secure system put in place will allow remote access to other confidential data sources as well as the cross-processing of different data sources

# The CASD in figures

---

30 staff members organized in **3 poles**



**Key figures:** <https://www.casd.eu/en/le-centre-daccs-securise-aux-donnees-casd/le-casd/>

# Certifications and CASD commitments

GDPR compliance and CNIL authorization (n°2014-369)



ISO 27001 and ISO 27701 (GDPR) certifications  
« Health data hosting » certification and « RSDS » approval

Regular security audits



ISO 27001

Sécurité de l'information  
FR055849



ISO 27701

Protection des données personnelles  
RGPD / FR060159



HDS

Hébergeur de données de santé  
FR055852



SNDS

Homologation au référentiel de  
sécurité des données de santé

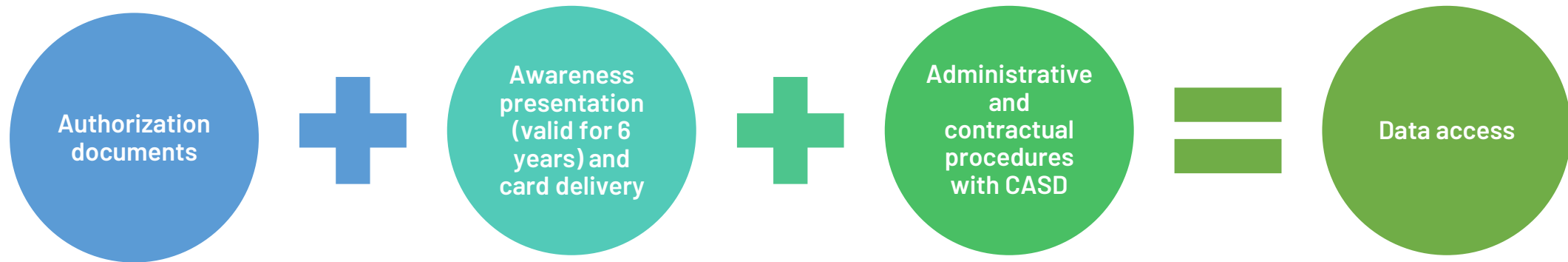


Autorisation de traitement  
2014-369



# Steps towards Data access

---



## Authorization documents:

- Statistical Secrecy Committee (CSS):
  - Documents signed by data producers and national archives
  - DGFIP authorization for tax data
- Data not covered by CSS: direct authorization by the relevant authority





# Legal framework of data access

# Statistical secrecy

---

- Access to data covered by the statistical secrecy is allowed under the conditions of **articles 6 and 7 bis of the 1951 law** (after consulting the Statistical Secrecy committee). These conditions were extended to tax data by the LPF (Livre des procédures fiscales – Tax procedures handbook) and to all administrative database by the CRPA (Code des relations entre le public et l’administration – Code of relations between the public and the administration)
- Data access is granted under cover of the statistical secrecy, the secrecy is shared
- ➔ **No dissemination, neither of individual data nor of results indirectly identifying persons or firms**
- Consequences in case of a secrecy breach (recalled by your commitment):
  - **You are personally liable (data access is strictly personal)**
  - **Criminal sanctions:**
    - articles 226-13 and 226-14 of the penal code (breach of professional secrecy):  
“the disclosure of secret information by a person who is in possession of it either because of his profession or his status, or because of his function or a temporary mission, is punishable by **one year imprisonment and a fine of 15 000 euros**”;
    - articles 226-16 to 226-24 of the Penal code (violations of personal rights resulting from computer files or processing) in case of information related to individual firms
  - **Compensation in civil liability** for caused damages
  - **Not to mention damage to reputation...**

# Processing of personal data

---

➤ **Data processing is subject to the obligations of the French Data Protection Act and the GDPR**

(specific provisions for organizations located on French territory).

➤ **Treatments prohibited by the GDPR:**

- A treatment with the final or intermediate purpose **of re-identifying one or more natural persons**
- A treatment with the final or intermediate purpose of taking **a decision against an identified natural person.**

➤ **Consequences in case of breach:**

- According to the Data Protection Act: up to 5 years of imprisonment and a fine of 300 000 euros (section 5 of Chapter VI of Title II of Book II of the Penal Code)
- According to the GDPR: administrative fines up to 20 000 000 euros or, in the case of companies, 4% of the total annual worldwide turnover of the previous fiscal year.

**Reminder: personal data**

Data referring to individuals, in other words all household sources, and individual firms in firm data sources

# Processing of health data

---

- The processing of health data is subject to the obligations of the Public Health Code (Article L4113-7).
- Treatments prohibited by the Public Health Code :
  - A treatment that would aim at **promoting health products** to health professionals or health institutions
  - A treatment that would result in **the exclusion of benefits or the modification of insurance contributions or insurance premiums** for an individual or group of individuals
- Any treatment that does not comply with the purposes declared to the CNIL is a prohibited treatment.

# Steps to take when processing personal data

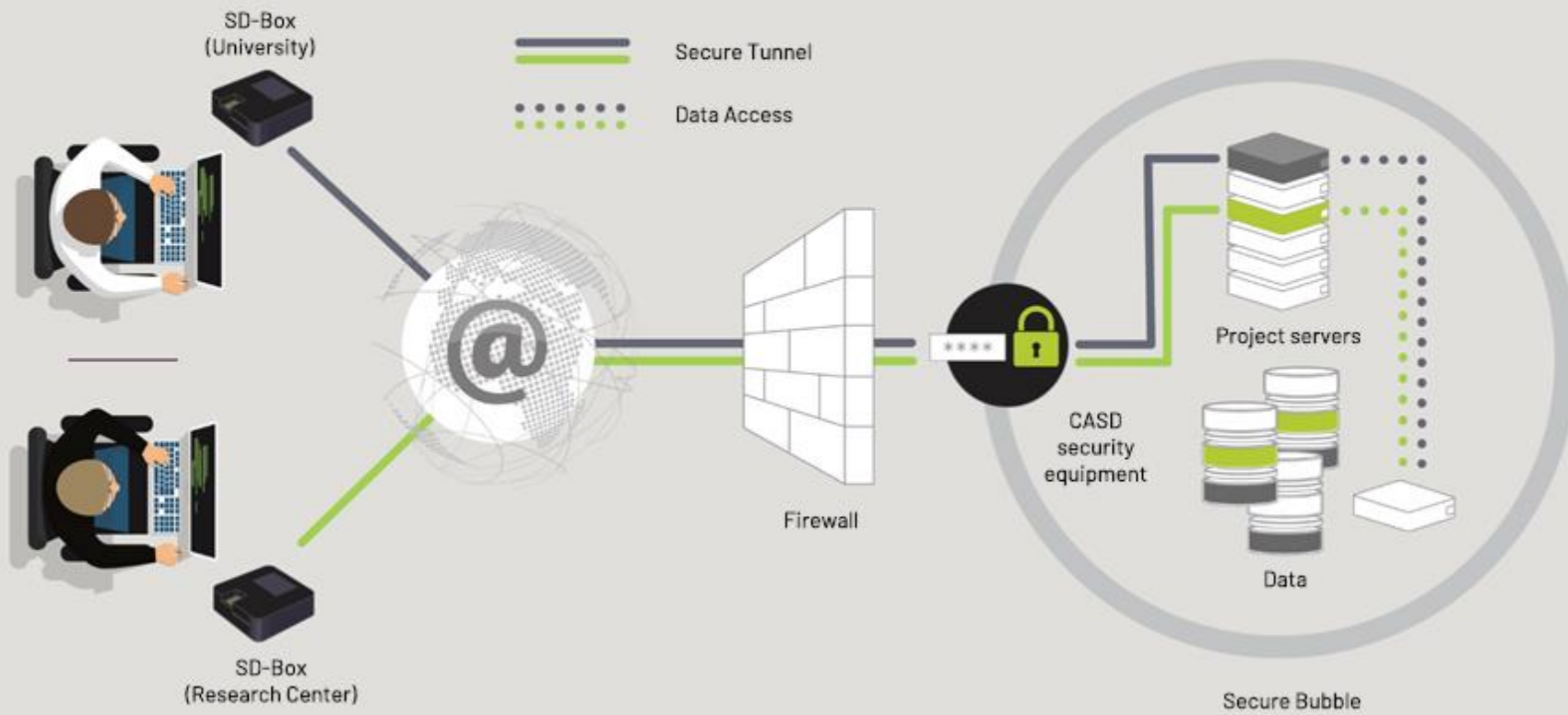
---

- For the processing of personal data, here are the obligations to be complied with depending on the case:
    - registration of the project in the register of the processing operation
    - carrying out a privacy impact assessment (PIA) in the case of the use of so-called sensitive data (Art. 9 of the GDPR)
    - request of a CNIL processing authorization (in particular in the case of health data)
- ➔ *Contact your legal correspondent or your data protection officer when appropriate. More information on the website: **CNIL.fr***





# IT Architecture



# The SD-BOX

Un boîtier autonome et dédié à l'accès

Ecran de paramétrage

Connecteur RJ45  
Connecteur HDMI  
Connecteurs USB

Lecteur biométrique

Lecteur de carte à puce





# Workflows

---

## ➤ Remote access

- With the SD-Box, you can work remotely on confidential data, while guaranteeing to the producer that no file can be retrieved by the user or inserted in the work environment (no copy/paste, no printing, no USB key, no internet connection...)

## ➤ Inputs and outputs

- There are specific workflows for importing or exporting files outside the secure environment, with verification by the Data Management and the IT departments
- Inter-project outputs ➔ transferring files from a project to another

## ➤ Adapted and customizable work environment

- Updates – system or software – on the project environment are done at the request of the project leader

The background of the slide is a dark blue field. Overlaid on this are white, thin, wavy contour lines that resemble topographical map lines. These lines are more densely packed on the left side and become more sparse towards the right. Scattered throughout the image, particularly along the contour lines, are numerous small, light blue dots. The word "Demonstration" is centered horizontally and vertically in a white, sans-serif font.

# Demonstration

# Terms and Conditions of Use (1)

---

- The SD-Box must be in a closed room and only the accredited user may see the screen
- The screen, the keyboard and the mouse must be provided and installed by the local IT manager. No other peripherals should be connected to the SD-Box
- Remove your smartcard from the SD-Box when you are away for even a short time (your calculations will not be interrupted)
- The smartcard is strictly personal
- You must not “lend” your session
- **No photo or video recording!**

# Terms and Conditions of Use (2)

---

- Concerning exports, the user commits to export only non confidential data, it is the user that must do the checking if confidentiality rules are respected
- Only the user is responsible in case of problem (output, disclosure...)
- The user contract specifies how CASD service can be interrupted (updates, maintenance...)
- Computer monitoring mechanisms are implemented in order to ensure compliance with security rules

# Smartcard issuance

---

## ➤ Biometric authentication system :

- Type 2 authentication
- The biometric data stored are encrypted twice with a key stored in the smart card and a key only known by the CASD
- New projects are automatically added to your smart card
- Remotely enroll fingerprints (chargeable option)

➤ In case of suspected loss/theft, notify the CASD as soon as possible. We will take reversible measures to reduce the risk of fraudulent use. If the loss of the smartcard is confirmed, the card will be permanently cut off and you will have to return to the CASD to get a new one.

➤ Reminder : the card delivery is done by appointment, the identification documents accepted are ID card, passport, driving license, residence permit (no photo)

# In essence

---

Security is an important challenge for confidential data access:

- This induces a certain number of constraints for the user that we have tried to make the least “unconformable” possible (dedicated environment with many software, dedicated equipment almost plug and play...) compared to other data access solutions in other countries.
- A large part of our system relies on the **trust placed in users** who can see the data (which is not the case in some other countries...)





# Anonymization techniques

# Data anonymization

---

In order to respect the statistical secrecy of outputs, data must be **anonymized**.

« ***The anonymization, according to the CNIL, is a process using a set of techniques rendering impossible, in practice, any individual identification by any mean and in an irreversible way*** »

→ Method recommended by the CNIL: **generalization**, which means transforming data to make them refer to a set of units instead of a single unit.



# Three requirements

---

In order to be anonymized, data must respect three requirements:

1. **Non-individualization:** it must not be possible to isolate an individual from the dataset
2. **Non-correlation:** it must not be possible to link multiple datasets together concerning the same individual
3. **Non-inference:** it must not be possible to deduce near-certainly new information about an individual

# Anonymization issues: thresholds

Two techniques:

- The aggregation: aggregate sufficiently the data in order not to have only X units per group

Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises	7	300
SME	2	800
Big enterprises	10	5200



Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises & SME	9	1100
Big enterprises	10	5200

- Delete the information when it concerns less than X units

Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises	7	300
SME	2	800
Big enterprises	10	5200



Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises	7	300
SME	S	S
Big enterprises	10	5200

# Anonymization issues: thresholds

Two techniques:

- The aggregation: aggregate sufficiently the data in order not to have only X units per group

Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises	7	300
SME	2	800
Big enterprises	10	5200



Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises & SME	9	1100
Big enterprises	10	5200

- Delete the information when it concerns less than X units

Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises	7	300
SME	2	800
Big enterprises	10	5200
Total	19	6300



Enterprise category	Number of enterprises	Amount R&D (K€)
Micro-enterprises	S	S
SME	S	S
Big enterprises	10	5200
Total	19	6300

- Pay attention to the secondary secrecy !

This technique presents a risk of re-identification because you have to pay attention to the fact that other available data may allow to recalculate the masked value. This is why, what we call the secondary secrecy must be applied: more than one value must be deleted. In the example that we have, if we find the total number of enterprises, we can easily recalculate the deleted number of SME.

# Anonymization issues: the diversification

Achieve a distribution of group characteristics that is sufficiently diverse to reduce the risk of certain or near-certain deductions

Diagnoses taken care of during hospital stays for a given month and a given department						
Age group	Number of patients	Hypertension	Diabetes	Asthma	Cancer	Respiratory deficiency
20-29	13	0	4	13	0	0
30-39	36	6	10	9	5	7
40-49	52	15	9	11	16	8
50-59	49	14	11	6	10	8
60-69	53	12	9	8	11	23
70-79	58	8	31	12	6	56

# Anonymization issues: high contributions

Avoid high contributions for amount variables

Business sector	Number of enterprises	Turnovers
<b>Building construction</b>	467	860 745
<b>Civil engineering</b>	389	1 696 872
<b>Special construction work</b>	804	973 610

Business sector	Maximum turnover	Maximum's percentage
<b>Building construction</b>	256 804	29,83%
<b>Civil engineering</b>	1 531 794	90,27%
<b>Special construction work</b>	41 947	4,30%



# Confidentiality rules

# Confidentiality rules: general rules

---

## ➤ Household data

- **The knowledge of an individual characteristic** cannot lead to **the knowledge of another one on the same individual**
- The rules apply to natural persons (individual and individual enterprises)

## ➤ Firm data

- No less than **3 units** per cell
- A firm cannot account for more **than 85% of the total of an amount**
- The rules apply to SIREN and not to SIRET

## ➤ Agriculture data:

- No less than **3 units** per cell
- A farm cannot account for more than **85% of the total of an amount**
- The rules apply to farms and not to plots

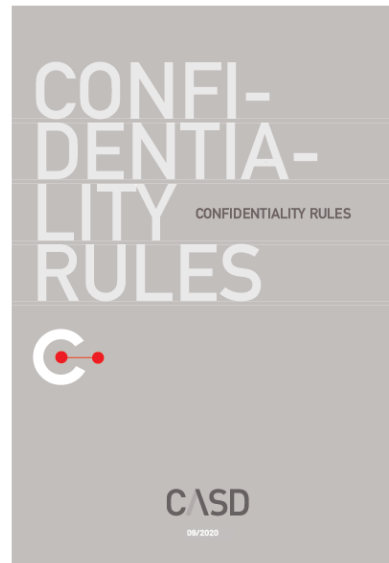
## ➤ Tax data

- Same rules apply for firms
- For household data: no less than **11 individuals**
- For tax-related data: a threshold of **11 units**, a unit cannot account for more than **85% of the total of a cell**

# Confidentiality rules: specific rules

---

- In addition, specific rules per data sources exist, they are defined by the data providers
- Detailed in a document about confidentiality rules (sent by email after the enrolment session)







# Workflows: Inputs and Outputs

# Export: definition

---

- Definition: export of non-confidential data outside of the secure environment
- Two types of exports exist:
  - Manual export : needs CASD review of **every file**.
    - The majority of projects with access to public statistical data.
  - Automatic export : files are directly sent with no prior review by CASD.
    - Mainly projects with access to health data, among others.
- **You are solely responsible for the respect of confidentiality rules in export files**

# Manual export: two processing

---



If the export respects confidentiality rules, and depending on its level of complexity, you should receive it within 48 hours

# Manual export: available on CDAP

---

- You will have an account on CDAP : <https://cdap.casd.eu/>
- On your CASD interface, you can see all your projects
- You will find information on the end of your authorization/subscription, the list of active members etc.
- You will see all exports made in your project and be able to download yours during 15 days
- You will have the information on the number of export credits you have left

# Manual export: counting

---

- At the beginning of your project, you have a pack of **20 export credits** (for all project members, not per one member)
- **One export credit** equals to **30 minutes** of processing time by the Data Management department to check one or multiple exports
- It means you initially have 600 minutes of checking time

Example : 1<sup>st</sup> export: 5 min of checking time (do files without data)

2<sup>nd</sup> export: 35 min of checking time (results tables)

➔ Counting : 40 min of checking time, being 1 export and 10 min use. You have 19 export credits left

- If you reach 20 export credits, you can order an additional pack (10 export credits)

# Manual export: how to proceed?

---

- To request an export, put your files in a .zip archive, right click and choose “Sortie CASD”
- Insert a **file describing** the export. It must indicate:
  - **the data used** so we can directly know which rule must be applied and the **signification of every used variable**
  - for regression, econometric models: the number of observations used
  - for maps and graphs: the population and the definition of the used variables
  - for aggregated results tables: variables description, each cell's number of observations and information on the highest contribution in each cell for results regarding **amount variables** (in a non-confidential control file)

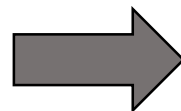
# Control file example

For results regarding amount variables, add columns indicating the **maximum** and **maximum's percentage** of amount variables.

→ Create an additional file for the review, it will be deleted from your output before we send it to you

**Control file**

Business sector	Number	Total turnover	Maximum turnover	Maximum's percentage
<b>Buildings construction</b>	43 467	860 745	256 804	29,83%
<b>Civil engineering</b>	22 389	1 696 872	1 531 794	90,27%
<b>Specialized construction work</b>	61 804	973 610	41 947	4,30%



**Output file**

Business sector	Number	Total turnover
<b>Buildings construction</b>	43 467	860 745
<b>Civil engineering</b>	<b>S</b>	<b>S</b>
<b>Specialized construction work</b>	61 804	973 610

# Inter-projects export

---

- Transfer of an export from one project to another (whether you are a member or not)
  - ➔ Do an export request as indicated previously
  - ➔ In your email requesting the export, specify:
    - ➔ That the export is a transfer between two projects
    - ➔ The **name of the two projects concerned**
  - ➔ After checking, the export will be directly transferred from one project environment to the other

## ➤ **Warning!**

If the project of origin is accredited for different data sources than the project receiving the export, you will need to provide a detailed description of the files to be transferred in order for CASD to verify that the export does not contain data for which the receiving project is not accredited.



# Automatic exports

---

- An automatic export is not reviewed by CASD before being sent to the user. Nevertheless, it will be saved and can be the object of an a posteriori review.
- You will directly receive the files in your mailbox
- A form must be filled engaging you to some obligations (they differ depending on the data accessed) and you have to add a description file to your export.
- Thresholds of size and frequencies are implemented (they differ depending on the used data)
  - Example : 10 exports per day per user with a maximum size of 10 Mo per export

# Import: definition and process

---

- Definition : insertion by CASD into your work environment of files that you send us in a readable format



In the email of import request, specify your **project name** and add a **file description** (file type and its content)

# Import : rules

---

- Import of **individual data** destined to be matched with other CASD data is only possible if it has been explicitly declared in your research project submitted for your authorization
  - Data producer agreement is required to import any data unless it is public data
  - Can only be imported « inactive » files (non executable)
  - It is only possible to import files in formats of the available software in the CASD environment (txt, csv, SAS, R, STATA...)
- **Pay attention** not to forget to encrypt confidential data

# Open access data

---

Some open access data can be retrieved from the folder “Libre Acces” in the folder “Raccourcis”:

➤ Classifications

- GEOFLA, ADMIN-EXPRESS and Contour IRIS of the IGN database
- Sirene: register of enterprises and establishments – stocks on the 1<sup>st</sup> of January of each year
- Geographical Classifications and NAF, NES, PCS, COICOP, CPF, CJ
- Legal population
- ESANE documentation
- INSEE methodology sheets and SAS macro (CALMAR, CUBE, data analysis)
- National Register of Health and Social Establishments (FINESS)
- Drug and Tariff Information Base

➤ CASD documents:

- CASD user guide
  - Confidentiality rules
  - Good practices for export requests
  - Good practices for using the softwares in the secure environment
- Sharing folder



# Data citation and publications sharing

# Data citation

- Cite the used data in your publication!
- A template for citing data and their DOI (Digital Object Identifier) is available on each source webpage in the section “Persistent identifiers”

Example for FARE 2017: Insee & Ministère des Finances (DGFIP)[Producer], Fichier approché des résultats d'Esane - 2017 [Data file], Centre d'Accès Sécurisé aux Données (CASD)[Diffusor], <http://doi.org/10.34724/CASD.42.3127.V1>



**FARE : Fichier approché des résultats d'Esane - 2017**

---

**Producteur :** [Insee & Ministère des Finances \(DGFIP\)](#)

---

**Description :** Le fichier approché des résultats d'Esane contient les informations comptables issues des liasses fiscales mises en cohérence avec les informations provenant de l'enquête Sectorielle Annuelle.

---

**Thème :** Caractéristiques des entreprises

---

**Type de ressource :** Fichiers de données

---

**Habilitation :** [Comité du Secret Statistique](#)

---

**Version :** 1

---

**DOI :** [10.34724/CASD.42.3127.V1](http://doi.org/10.34724/CASD.42.3127.V1)

---

**Citation :** Insee & Ministère des Finances (DGFIP)[Producteur], Fichier approché des résultats d'Esane - 2017 [Fichiers de données], Centre d'Accès Sécurisé aux Données (CASD) [Diffuseur], <http://doi.org/10.34724/CASD.42.3127.V1>

# Publications sharing

---

➤ If your project benefits from a subsidized rate, we ask you to mention the CASD in your publication, in the following terms:

« Access to some confidential data, on which is based this work, has been made possible within a secure environment offered by CASD – Centre d'accès sécurisé aux données (Ref. ANR-10-EQPX-17) » [English Version]

« L'accès à certaines données utilisées dans le cadre de ce travail a été réalisé au sein d'environnements sécurisés du Centre d'accès sécurisé aux données – CASD (Réf. ANR-10-EQPX-17) » [French Version]

➤ Do not forget to inform us about your publication by filling our online form:  
<https://www.casd.eu/en/share-a-new-article/>



The background is a dark blue gradient. On the right side, there are white, thin, wavy contour lines that resemble a topographic map. Scattered across the image, particularly along these lines and in the open spaces, are numerous small, light blue dots. The word "Support" is centered in the middle of the image in a white, clean, sans-serif font.

Support

# CASD Support

## The Data

- Available data
- Data documentation
- Access procedure
- Enrolment
- Opening the right to access the data
- Importing files in your work space
- Results exports

### Data Management Service

01 84 19 69 24

## IT

- Your access tools: biometric card, SD-Box...
- Connection issues
- Technical issues on your project server
- Server configuration and modification: hardware, software

### IT Service

01 84 19 11 37

## Contracts and billing

- Your contracts
- Fees estimate
- Your billing
- Payments means

### Project Management Service

01 70 26 69 32

Email addresses: [service@casd.eu](mailto:service@casd.eu) / [accés.pmsi@casd.eu](mailto:accés.pmsi@casd.eu) / [imports-exports@casd.eu](mailto:imports-exports@casd.eu)

CASD website: <https://www.casd.eu>

# Quiz

---

- You just received an email with an access link
- 10 questions with only **one possible answer**
- The quiz is **mandatory**
- Goal: estimate your understanding and explain the parts which may have caused some understanding difficulties



We thank you for your attention

CASD C