

Données sensibles pour la recherche scientifique

CASD

1^{er} décembre 2025
ANDOR

Kamel Gadouche
Directeur du CASD

CASD

SENSITIVE DATA



SECURITY



CASD



USER EXPERIENCE



DATA GOVERNANCE



OUTCOME



1^{ER} décembre 2025

Assises Nationales des Données de la Recherche - ANDOR 2025

SENSITIVE DATA

Health records (SNDS, Constances Cohort)

Educational records (Parcoursup)

Tax Data (IR, FIRM)

Criminal records (Casier)

Professional career and wages

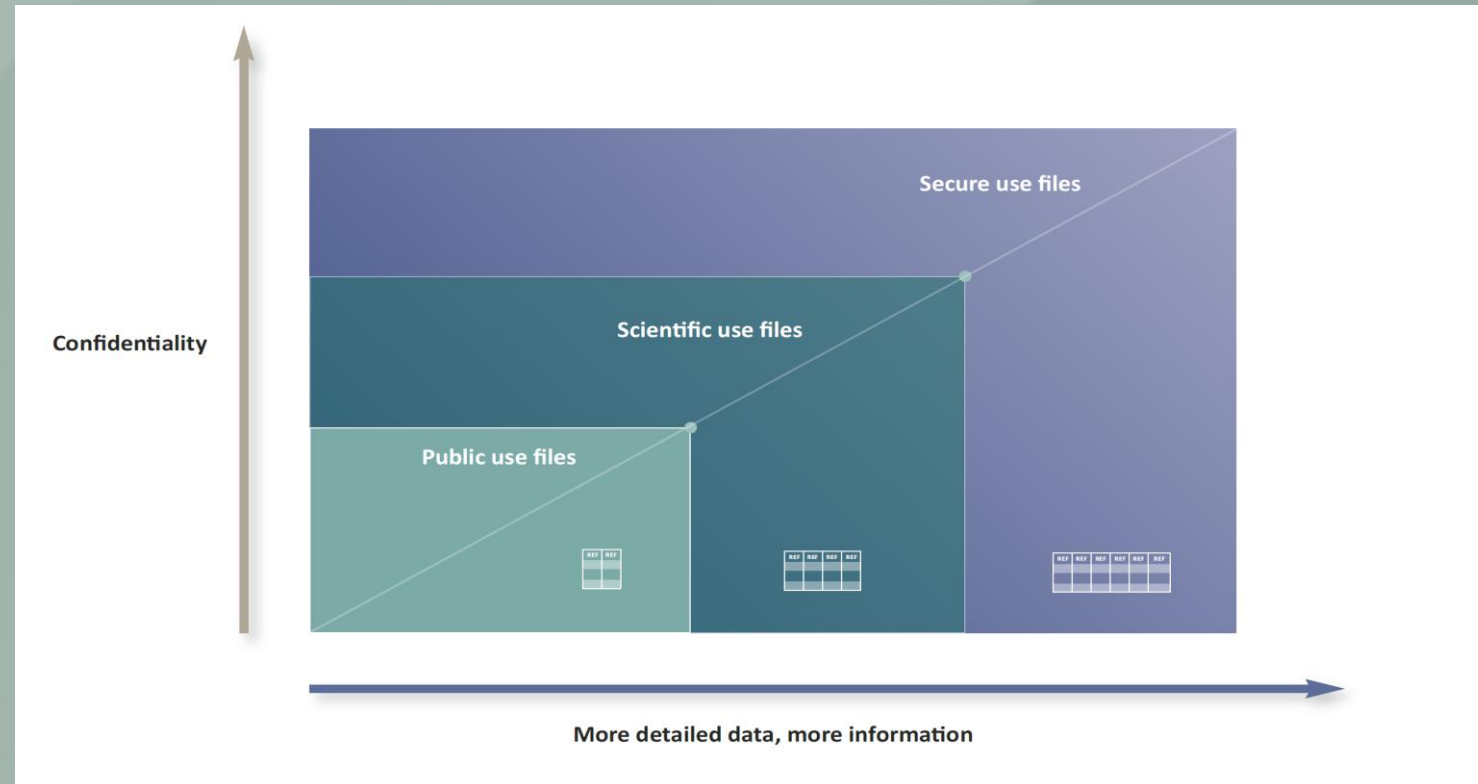
etc.

SENSITIVE DATA

Sensitive data can also be :

- Very detailed data
- Exhaustive

Even pseudonymized (no direct Identifiers) it could be still possible to re-identify data subject !



DID I FORGET
SOMETHING ?

Yesterday (11/30/2025)

Le Monde

Aurélie Ouss Il faut autoriser le rapprochement sécurisé des données judiciaires avec d'autres fichiers publics

Le 13 octobre, Gérard Darmanin a adressé une lettre aux agents du ministère de la justice, présentant son projet de réforme pénale. Le garde des sceaux y défend en particulier un projet de loi destiné à assurer une « sanction utile, rapide et effective ». Une part de cette réforme concerne l'encadrement du sursis et l'exécution plus systématique des peines, dans la logique selon laquelle une meilleure exécution des sanctions contribuerait à une justice plus efficace.

Depuis plusieurs décennies, la réponse française à la délinquance repose quasi exclusivement sur le renforcement des sanctions. Mais le premier effet de telles politiques est l'explosion de la population carcérale. En 2023, 21300 récidivistes ont été condamnés à un emprisonnement avec sursis. Si toutes ces peines étaient exécutées en détention, cela représenterait une hausse d'environ 20 % des entrées en prison, dans un contexte où les maisons d'arrêt comptent aujourd'hui 164 détenus pour 100 places. Une telle politique aggraverait la surpopulation carcérale, alors même que la France a été condamnée en 2020 par la

L'économiste appelle à croiser les informations du casier judiciaire avec celles d'autres administrations, sous contrôle strict et uniquement à des fins de recherche. Selon elle, cela permettra de mieux cerner les causes de la délinquance et de prévenir la récidive, afin de sortir d'une logique répressive

Cour européenne des droits de l'homme pour les conditions indignes imposées aux détenus.

L'histoire récente l'a montré : les peines planchers, instaurées en 2007 pour durcir les sanctions en cas de récidive, n'ont pas produit la baisse de criminalité espérée. Plus de fermeté n'a pas conduit à plus de sécurité, mais seulement à plus de prison. Les recherches empiriques sur les politiques pénales confirment ce constat : la sévérité des peines est bien moins dissuasive qu'on ne l'imagine, quel que soit le contexte étudié.

Autres leviers de prévention

La France se prive encore trop souvent de la possibilité d'évaluer d'autres leviers de prévention de la délinquance, notamment autour de l'emploi, du logement ou de la formation. Aux Etats-

Unis, par exemple, plusieurs études ont montré qu'offrir des jobs d'été aux adolescents réduit la délinquance. En exploitant des données administratives en Virginie, mes travaux ont démontré que le passage par la prison et le casier judiciaire pèsent durablement sur les parcours socio-économiques. Etant donné la richesse des données administratives françaises, des travaux similaires seraient possibles – mais la loi rend de telles études difficiles.

L'article 777-3 du code de procédure pénale interdit de croiser les données du casier judiciaire avec des informations issues d'autres administrations, même à des fins de recherche. Si cette protection répond à des préoccupations légitimes en matière de libertés individuelles, elle a une conséquence paradoxale : elle empêche d'étudier la manière dont les politiques sociales influencent la délinquance, ou dont les sanctions pénales affectent ensuite la vie économique et sociale des personnes condamnées. Depuis 2020, les données du casier sont disponibles au Centre d'accès sécurisé aux données, mais seulement en isolation [sur un fichier isolé, sans qu'elles puissent être comparées à d'autres données].

Nous ne voyons pas les effets sociaux de nos politiques pénales, ni les effets de nos politiques sociales sur la délinquance.

Pourtant, les choses commencent à évoluer. Dans certains contextes très encadrés, la statistique publique permet d'apparier des données issues de différentes administrations. Par exemple, dans le cadre d'un projet mené à l'Institut des politiques publiques sur les trajectoires scolaires et professionnelles des mineurs ayant été détenus, un appariement entre des données des ministères de la justice, de l'éducation et du travail est en train de voir le jour. Ce sont des avancées qui montrent qu'il est possible de concilier protection des libertés et production de connaissances, mais elles restent encore rares. Elles devraient encourager à aller plus loin et à systématiser ces pratiques.

Culture de l'évaluation

La France a déjà expérimenté les peines planchers, puis leur abrogation, sans en tirer de leçons précises. Nous commençons tout juste à explorer les interactions entre politique pénale et politique sociale. Réduire la délinquance ne dépend pas seulement du quantum ou de l'exécu-

tion des peines : cela suppose de comprendre comment l'école, l'emploi ou le logement influencent la récidive, et comment une condamnation pénale freine ensuite l'accès à ces ressources.

Une solution existe : autoriser, sous contrôle strict et uniquement à des fins de recherche, l'appariement sécurisé des données judiciaires avec d'autres fichiers publics. D'autres pays en Europe, en Scandinavie notamment, l'ont fait sans compromettre les libertés individuelles.

Si la France veut une justice à la fois crédible et efficace, elle doit continuer à développer une culture de l'évaluation dans le domaine pénal : non pas pour mesurer la fermeté des peines, mais pour identifier ce qui réduit réellement la délinquance. C'est à cette condition qu'elle pourra passer d'une politique de réaction à une politique de prévention ; d'une logique de punition à une logique de compréhension. ■

Aurélie Ouss est économiste et professeure de criminologie à l'université de Pennsylvanie



D'AUTRES PAYS EN EUROPE, EN SCANDINAVIE NOTAMMENT, L'ONT FAIT SANS COMPROMETTRE LES LIBERTÉS INDIVIDUELLES



Unis, par exemple, plusieurs études ont montré qu'offrir des jobs d'été aux adolescents réduit la délinquance. En exploitant des données administratives en Virginie, mes travaux ont démontré que le passage par la prison et le casier judiciaire pèsent durablement sur les parcours socio-économiques. Etant donné la richesse des données administratives françaises, des travaux similaires seraient possibles – mais la loi rend de telles études difficiles.

L'article 777-3 du code de procédure pénale interdit de croiser les données du casier judiciaire avec des informations issues d'autres administrations, même à des fins de recherche. Si cette protection répond à des préoccupations légitimes en matière de libertés individuelles, elle a une conséquence paradoxale : elle empêche d'étudier la manière dont les politiques sociales influencent la délinquance, ou dont les sanctions pénales affectent ensuite la vie économique et sociale des personnes condamnées. Depuis 2020, les données du casier sont disponibles au Centre d'accès sécurisé aux données, mais seulement en isolation [sur un fichier isolé, sans qu'elles puissent être comparées à d'autres données].

Nous ne voyons pas les effets sociaux de nos politiques pénales, ni les effets de nos politiques sociales sur la délinquance.

Pourtant, les choses commencent à évoluer. Dans certains contextes très encadrés, la statistique publique permet d'apparier des données issues de différentes administrations. Par exemple, dans le cadre d'un projet mené à l'Institut des politiques publiques sur les trajectoires scolaires et professionnelles des mineurs ayant été détenus, un appariement entre des données des ministères de la justice, de l'éducation et du travail est en train de voir le jour. Ce sont des avancées qui montrent qu'il est possible de concilier protection des libertés et production de connaissances, mais elles restent encore rares. Elles devraient encourager à aller plus loin et à systématiser ces pratiques.

Culture de l'évaluation

La France a déjà expérimenté les peines planchers, puis leur abrogation, sans en tirer de leçons précises. Nous commençons tout juste à explorer les interactions entre politique pénale et politique sociale. Réduire la délinquance ne dépend pas seulement du quantum ou de l'exécu

[Actualités](#)[Économie](#)[Vidéos](#)[Débats](#)[Culture](#)[Le Goût du Monde](#)[Services](#)

DÉBATS • POLICE ET JUSTICE

TRIBUNE

Aurélie Ouss
Economiste

Aurélie Ouss, économiste : « Il faut autoriser le rapprochement des données judiciaires avec d'autres fichiers publics »

Dans une tribune au « Monde », la professeure de criminologie appelle à croiser les informations du casier judiciaire avec celles d'autres administrations, sous contrôle strict et uniquement à des fins de recherche. Selon elle, cela permettra de mieux cerner les causes de la délinquance et de prévenir la récidive, afin de sortir d'une logique répressive.

Publié aujourd'hui à 06h00 |  Lecture 3 min.

 Lire plus tard



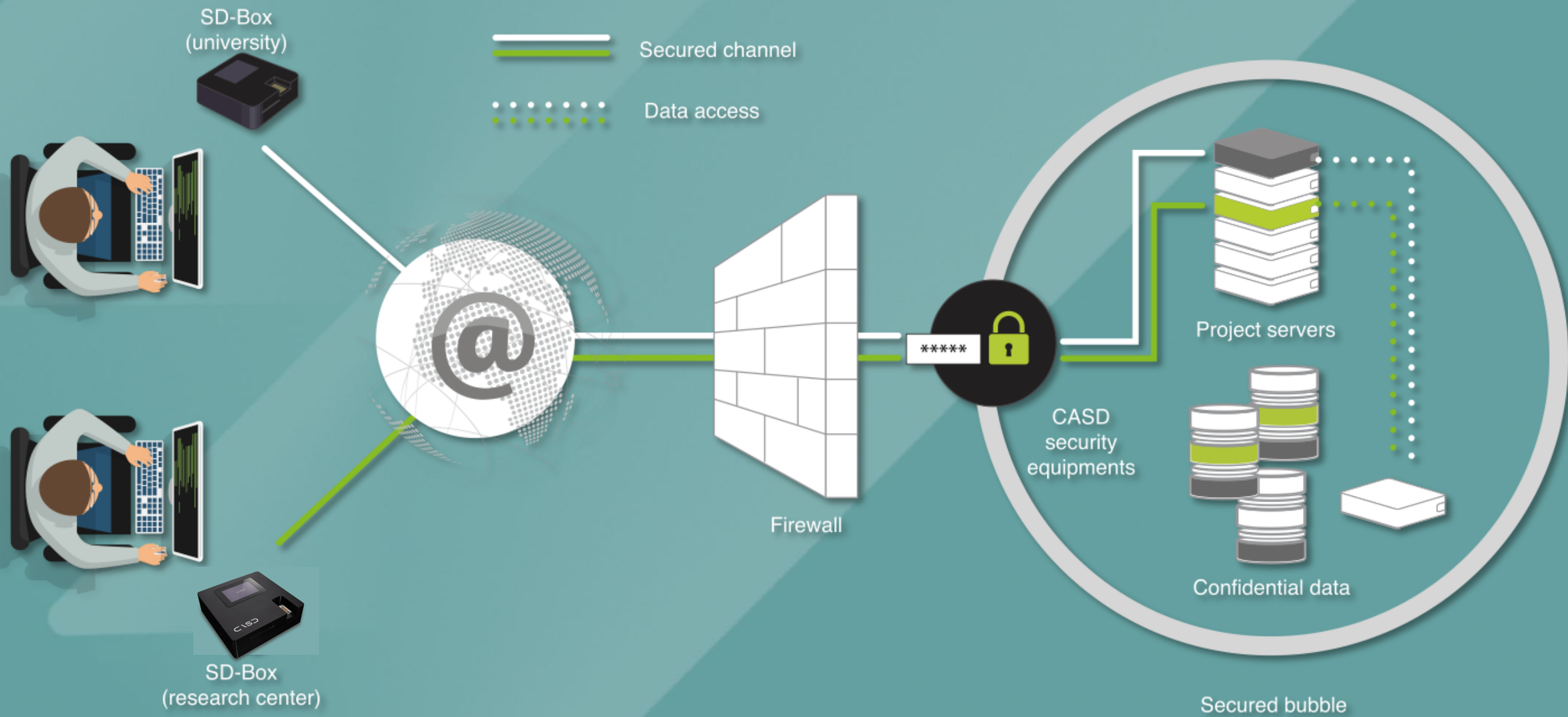
Sensitive and very detailed data can even
be much more valuable for research when
they are linked

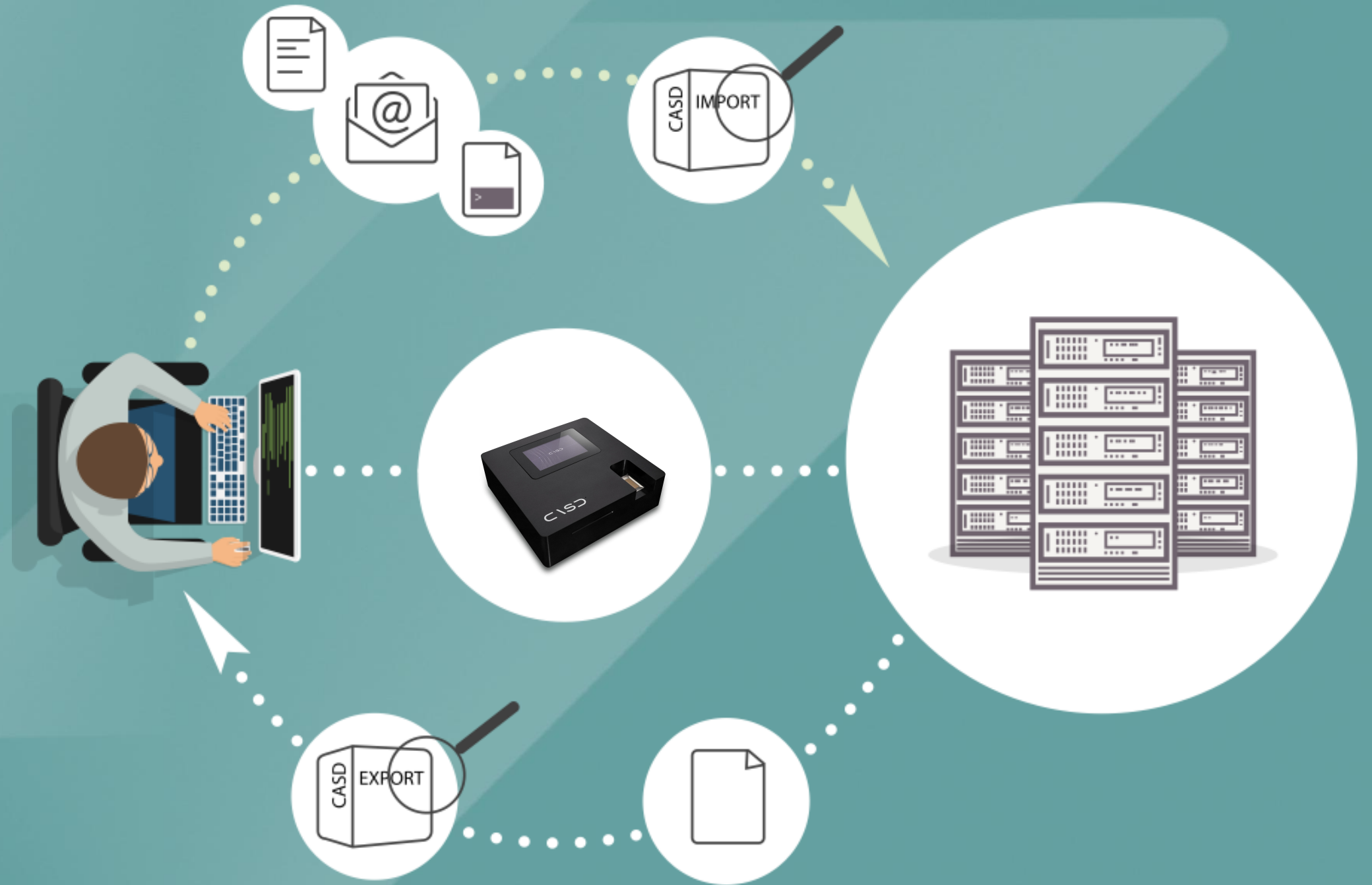
BUT MUCH MORE SENSITIVE

To make everything easier,
the data producers remain responsible,
not researchers, for the data they collect

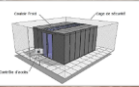
How to convince data
producers and enable
research on sensitive data ?







Where are the servers?



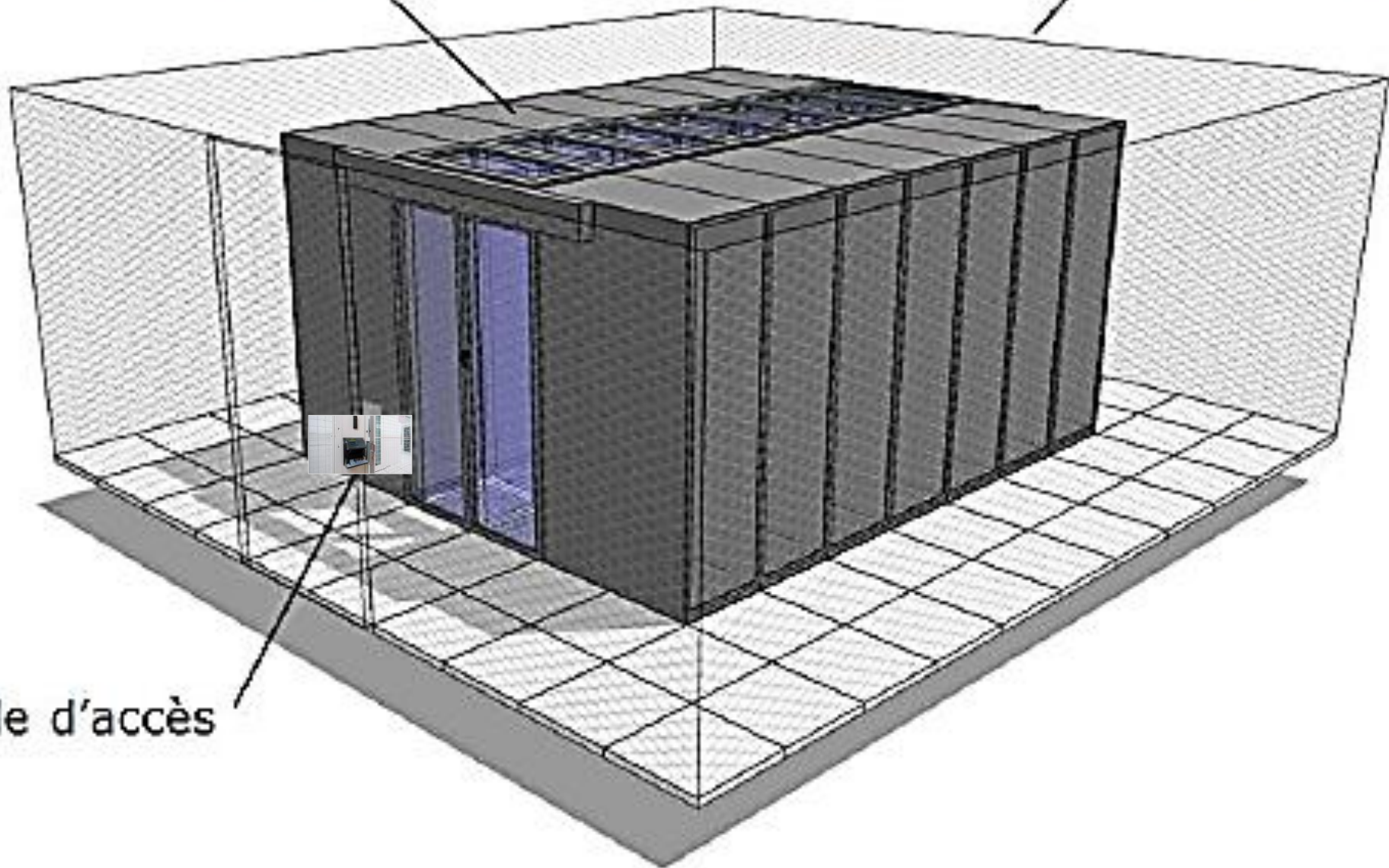
Datacenters



Couloir Froid

Cage de sécurité

Contrôle d'accès





Il presente documento è riservato e non deve essere divulgato a terzi. È vietata espressamente la ristampa o l'uso non autorizzato senza permesso scritto dalla Direzione.

SECURITY

**CASD : First data hosting service
to obtain official
GDPR certification (Art. 42)**



International privacy symposium, Venice, May 2025


Europrivacy
GDPR
Certification

CASD



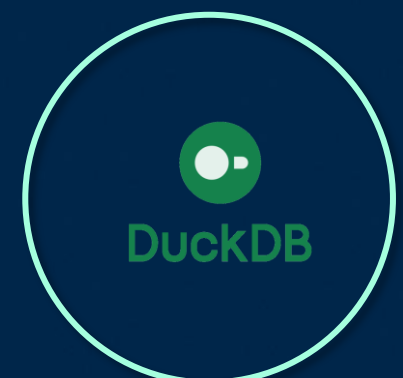
CONSORTIUM



GROUPE
ENSAE
ENSAI



Groupement d'intérêt public (40 employees)



TS-DEMO002

04.Demo_geo.ipynb A

04.Demo_geo.ipynb > ...

+ Code + Markdown | ▶ Run All ⌂ Restart ⌂ Clear All Outputs ⌂ Variables ⌂ Outline ...

demo (Python 3.10.13)

```
disease_to_plot = 'Cancers'
cancer_res_filtered = cancer_res[cancer_res['patho_niv1'] == disease_to_plot]
```

[28] ✓ 0.0s Python

```
merged = gdf.set_index('code').join(cancer_res_filtered.set_index('departement'))
print(merged.columns)
```

[29] ✓ 0.0s Python

```
Index(['GID_2', 'GID_0', 'COUNTRY', 'GID_1', 'NAME_1', 'NL_NAME_1', 'NAME_2',
       'VARNAME_2', 'NL_NAME_2', 'TYPE_2', 'ENGTPE_2', 'CC_2', 'HASC_2',
       'geometry', 'patho_niv1', 'perc', 'malades', 'pop'],
      dtype='object')
```

variable_to_plot = "perc"

```
# Define the plot color
norm = mcolors.Normalize(vmin=merged[variable_to_plot].min(), vmax=merged[variable_to_plot].max())
cmap = plt.colormaps.get_cmap('RdYlGn_r') # Utilisation de 'RdYlGn_r' pour aller du vert au rouge

# Create the map
fig, ax = plt.subplots(1, 1, figsize=(10, 8))
merged.plot(column=variable_to_plot, cmap=cmap, norm=norm, legend=True, ax=ax)

ax.set_axis_off()


# define the axe size
cbar = ax.get_figure().get_axes()[1]
cbar.set_aspect(30)

# Add a title to the plot
ax.set_title(f'Percentage of the polpulation who has {disease_to_plot}', fontdict={'fontsize': '10', 'fontweight': '3'})

# show the plot
plt.show()
```

[30] ✓ 0.3s Python

Percentage of the polpulation who has Cancers



04.Demo_geo

localhost:8888/notebooks/Public/Documents/DemoDuckData/04.Demo_geo.ipynb

jupyter 04.Demo_geo Last Checkpoint: 2 hours ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

```
cancer_res.head()
```

```
[ ]: # step3: calculate the percentage of different type of diseases among all population
cancer_res.insert(1, 'perc', cancer_res['malades']/cancer_res['pop'] * 100)

cancer_res.head(5)
```

2 Read shape file and build geo pandas dataframe

```
[ ]: gadm_path = f"(root_dir)/gadm41_FRA/gadm41_FRA_2.shp"
gdf = gpd.read_file(gadm_path)
gdf['code'] = gdf['CC_2'].astype(str).str.zfill(2)

[ ]: disease_to_plot = 'Cancers'
cancer_res_filtered = cancer_res[cancer_res['patho_niv1'] == disease_to_plot]

[ ]: merged = gdf.set_index('code').join(cancer_res_filtered.set_index('departement'))
print(merged.columns)
```

[12]: variable_to_plot = "perc"

```
# Define the plot color
norm = mcolors.Normalize(vmin=merged[variable_to_plot].min(), vmax=merged[variable_to_plot].max())
cmap = plt.colormaps.get_cmap('RdYlGn_r') # Utilisation de 'RdYlGn_r' pour aller du vert au rouge

# Create the map
fig, ax = plt.subplots(1, 1, figsize=(10, 8))
merged.plot(column=variable_to_plot, cmap=cmap, norm=norm, legend=True, ax=ax)


ax.set_axis_off()

# define the axe size
cbar = ax.get_figure().get_axes()[1]
cbar.set_aspect(30)

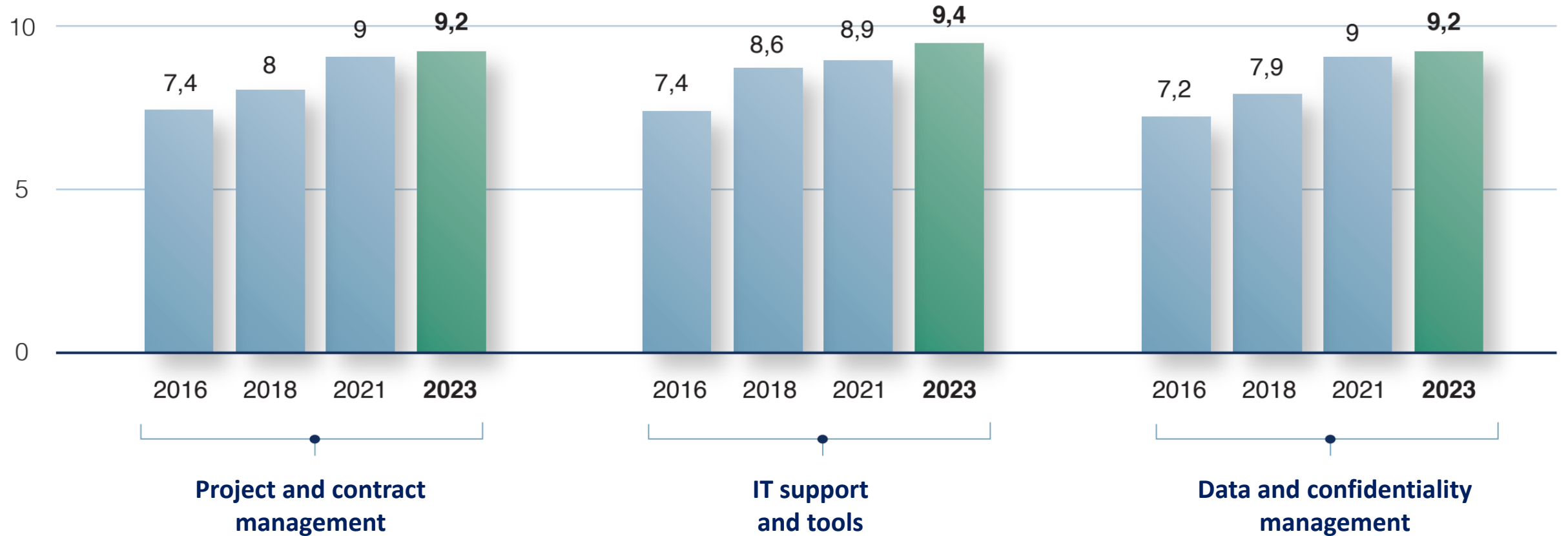
# Add a title to the plot
ax.set_title(f'Percentage of the polpulation who has {disease_to_plot}', fontdict={'fontsize': '10', 'fontweight': '3'})

# show the plot
plt.show()
```

Percentage of the polpulation who has Cancers



USERS' RATINGS BASED ON SURVEY ABOUT CASD SERVICES 2016, 2018, 2021 AND 2023

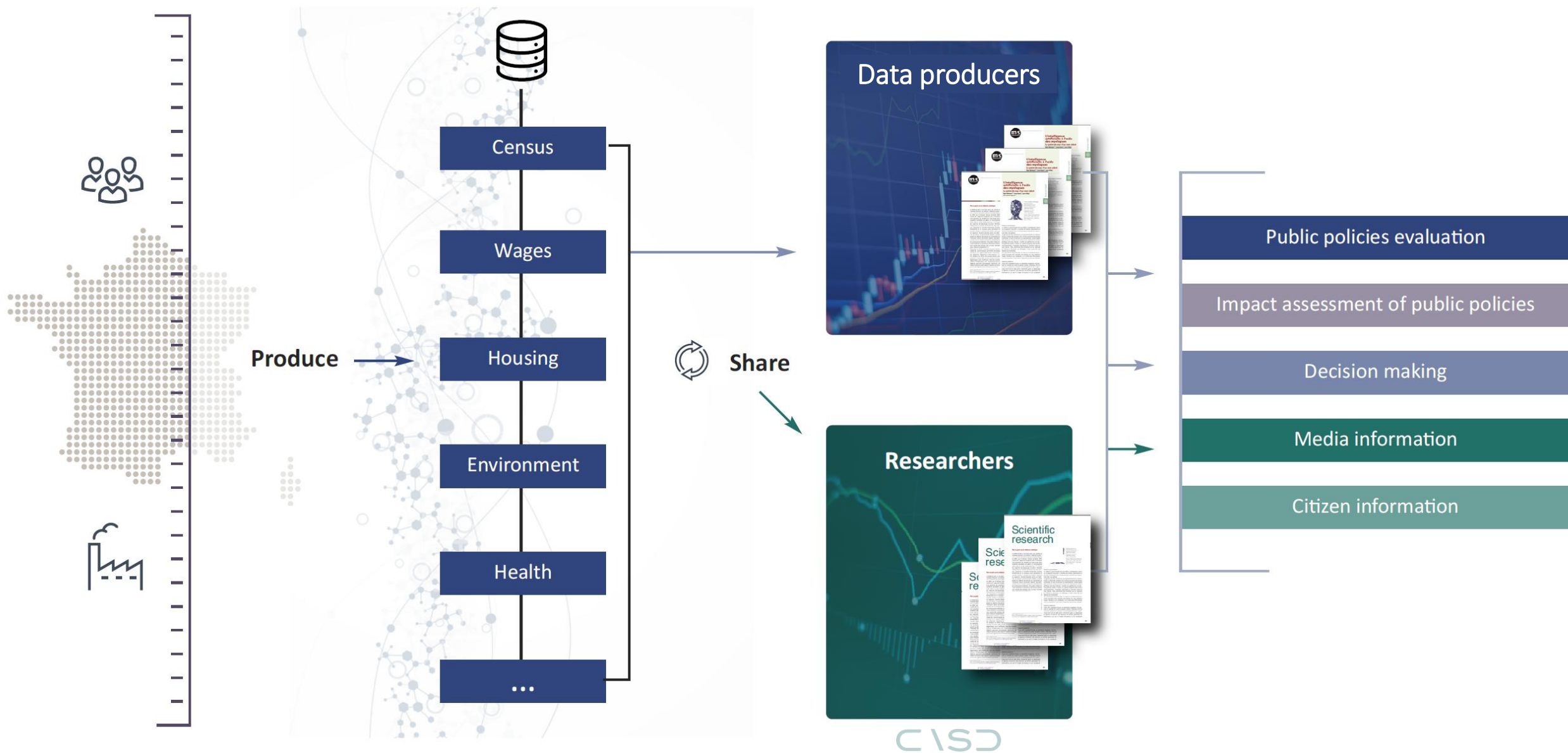


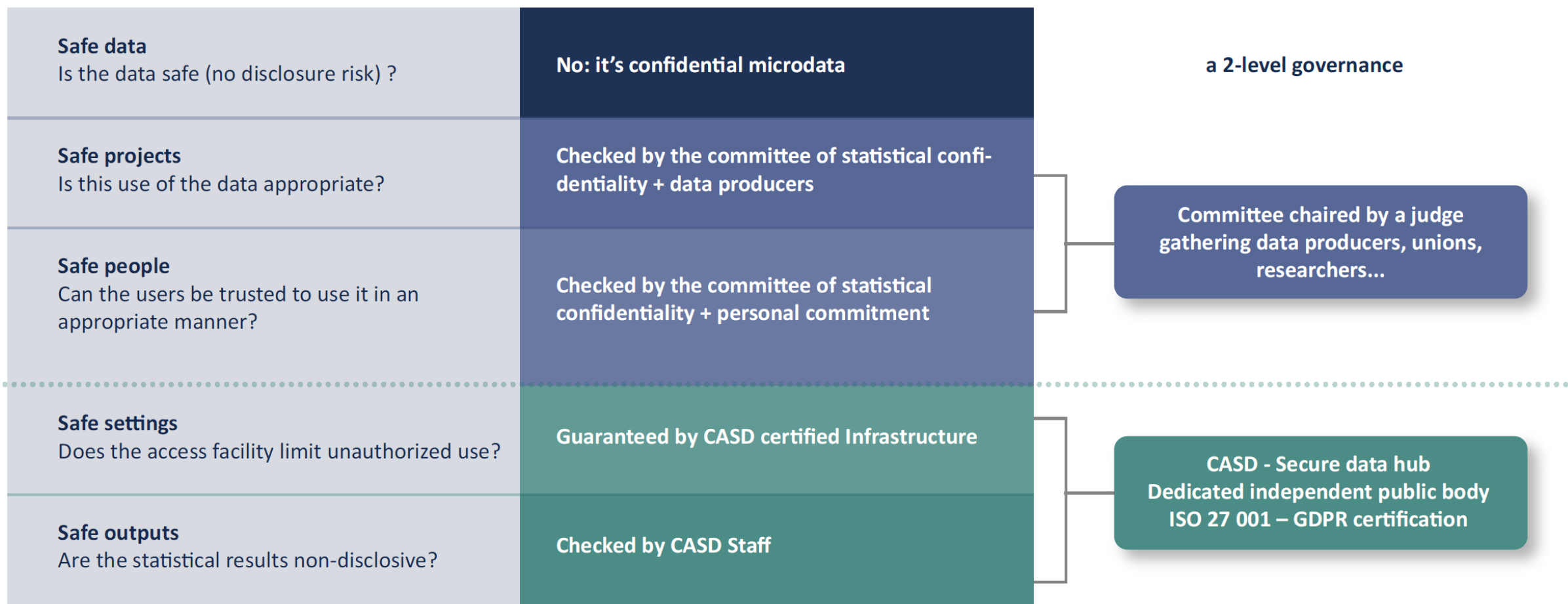
Measuring

Micro data

Studies and research

Better understanding





The data producers remain the owners of the data and have the final say: they (ie. DG of NSOs, DG of Tax Directorate) decide who can get access and for which purpose.

(*) Tanvi Desai, Felix Ritchie and Richard Welpton, Five Safes: designing data access for research, University of the West of England, 2016

National Institute of
Statistics and
Economic Studies
(INSEE)

Ministry of Finance

Ministry of Labour

Ministry of Agriculture

Ministry of Justice

Ministry for Health

Ministry of
Environment

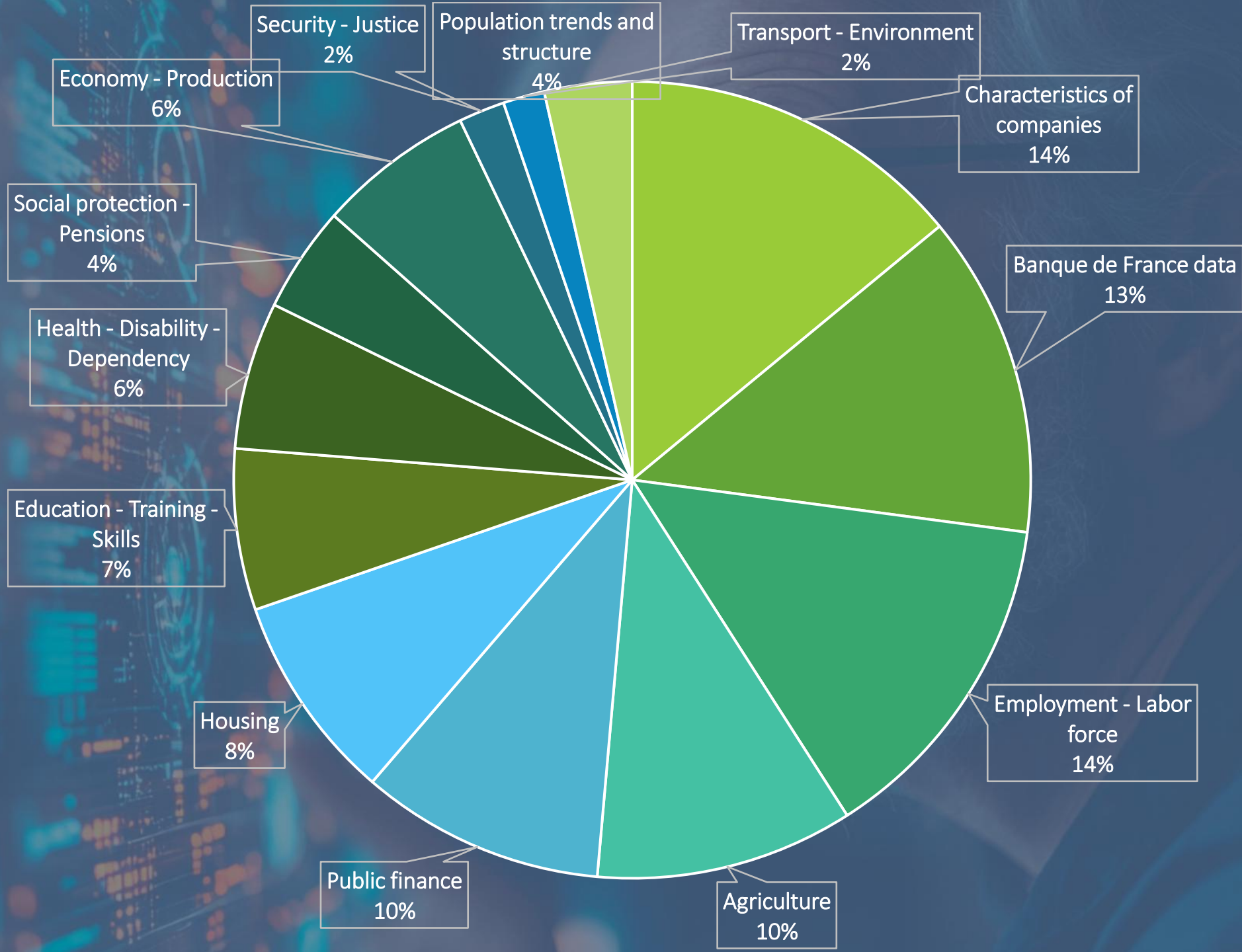
CONSTANCES
COHORT

Public Investment
Bank

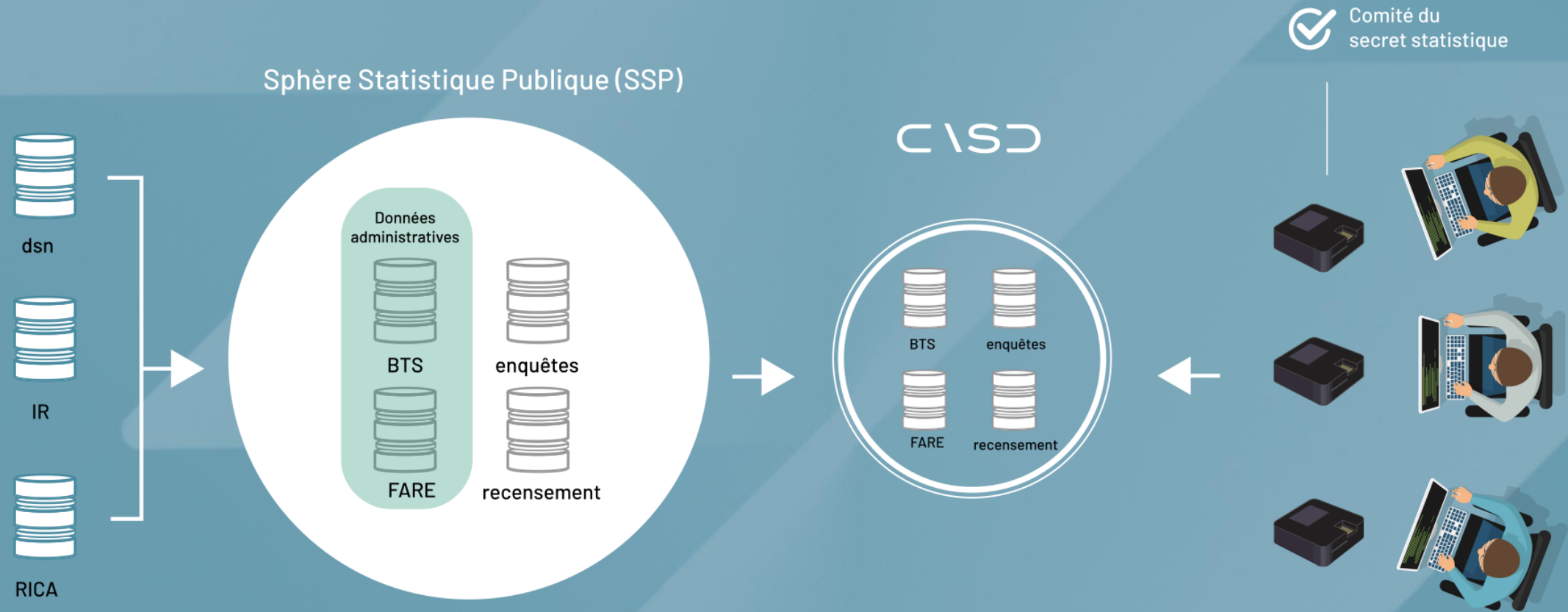
Meta (Facebook and
Instagram) and
Bytedance (TikTok)

Healthcare data

Banque de France

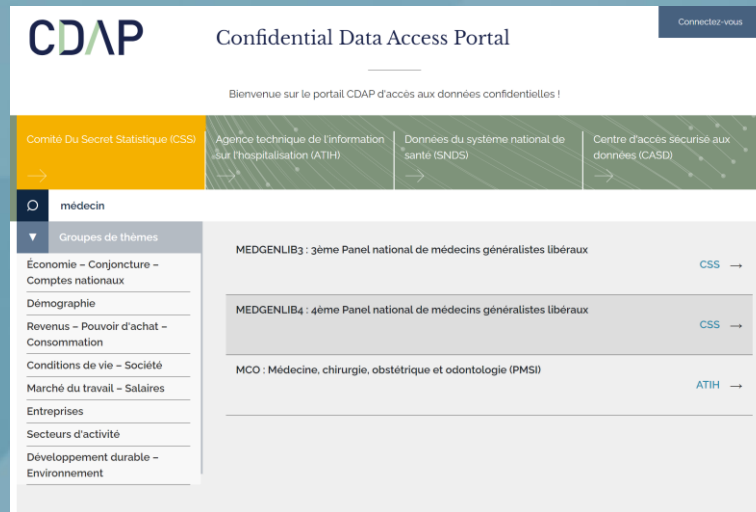


COMITE DU SECRET STATISTIQUE





Accreditation



CDAP

Confidential Data Access Portal

www.cdap.casd.eu

Statistical
Confidentiality
Commitee

Banque
de
France

Data
producer

BPI,
CNAF,
ACOSS...



Cross-sector

Agreements

with 40+ data producers

Unifying
The
accreditation
procedure

Allowing researchers to apply only once
for many datasources from many data
producers

Unifying
the secured
environment

Allowing working on many datasources from
many data producers in the same
environnement

LINKING data from different Administrative sources has proved essential for research

CASD : 72% of projects carried out in the CASD's secure environment uses data from more than one dataproducer

Example : 80% of projects from academic researchers using data from Banque de France is a joint use with other data.



Cross-sector

« CASD provides controlled access to detailed information, including tax and medical data »

The New York Times

« CASD is an example of the type of infrastructure needed to support this kind of analyses. »
« ...to grant third-parties access to sensitive health data... »

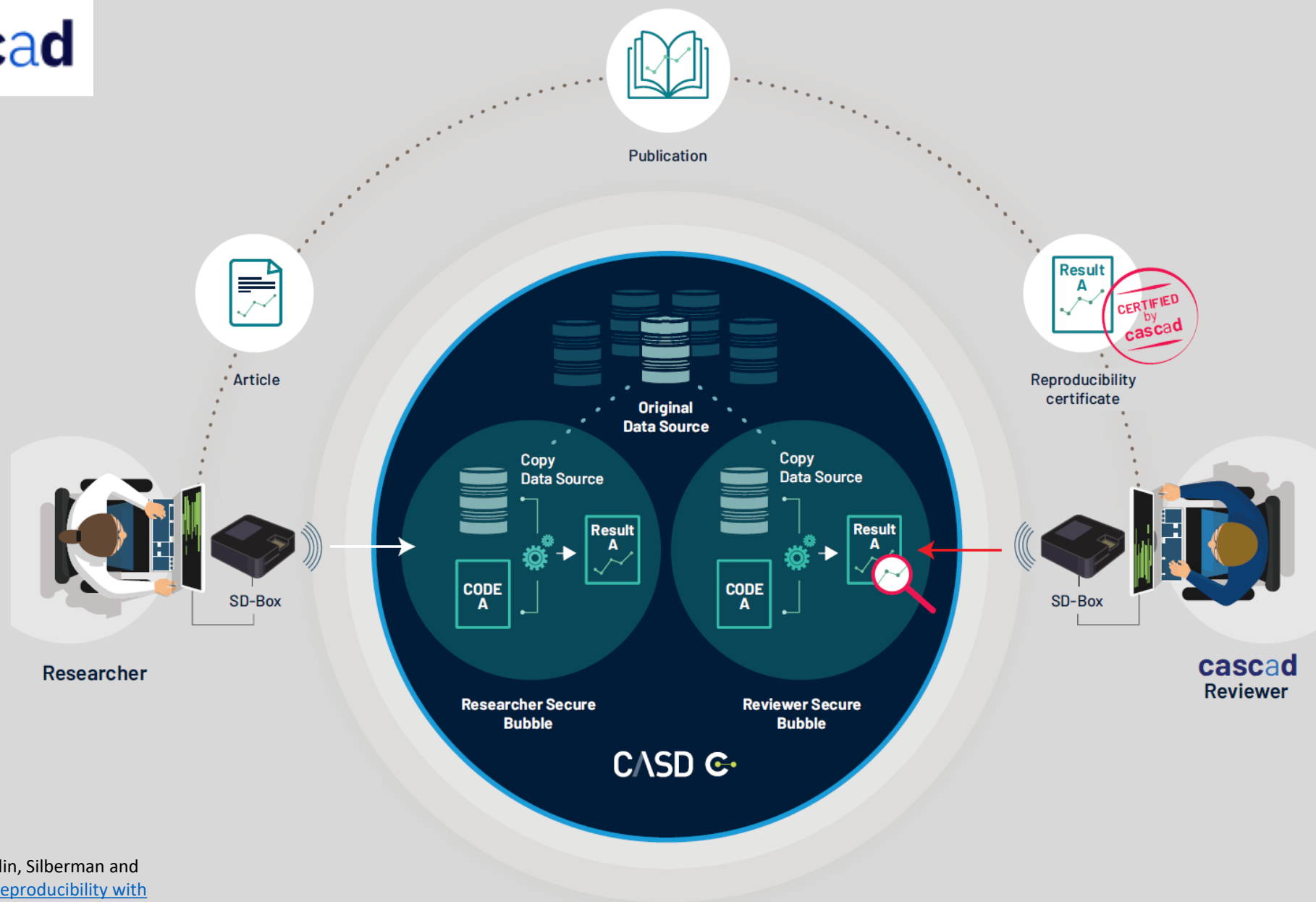
nature

« CASD, a single entry point to a large number of data producers... »

Science

« CASD :
une bulle pour protéger les données »

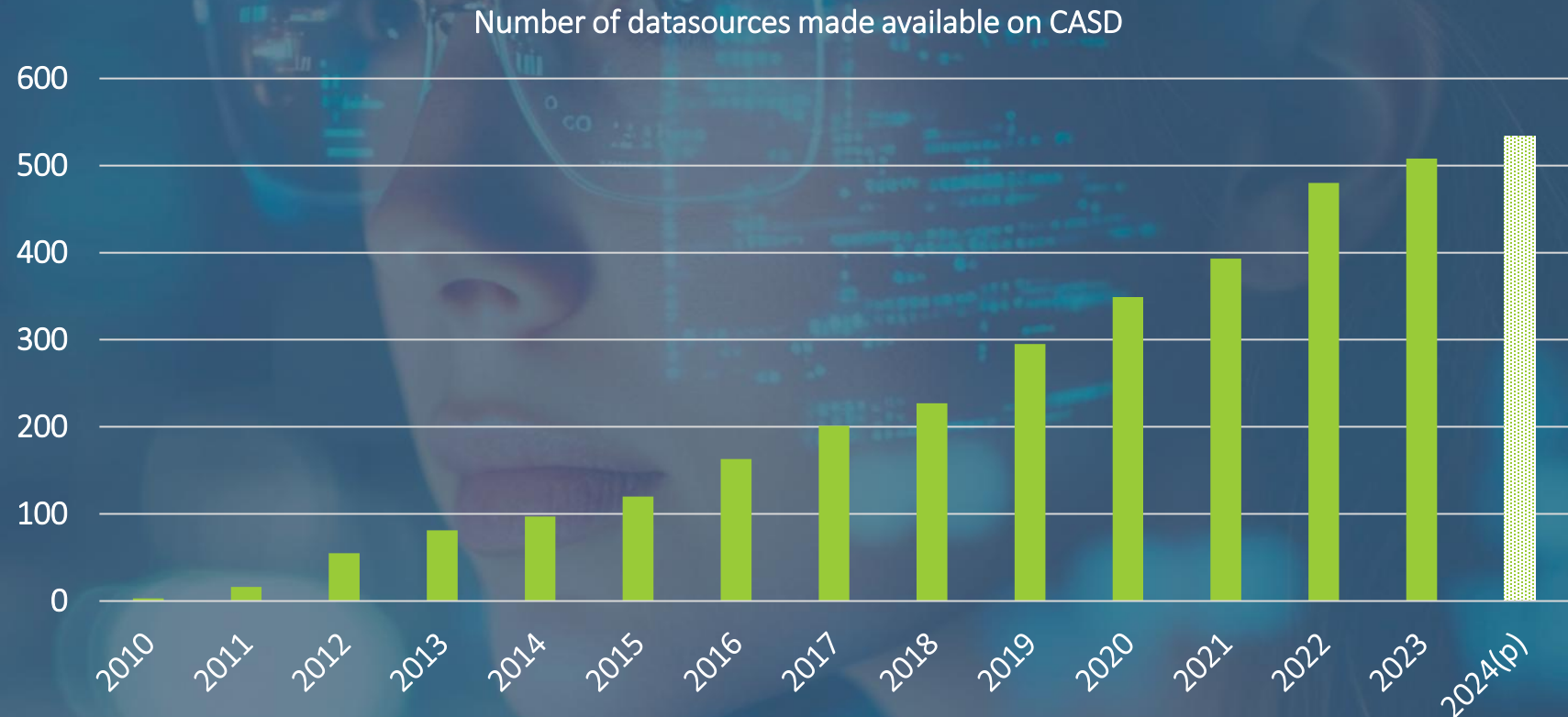
Le Monde



MORE SECURITY, MORE DATA

655

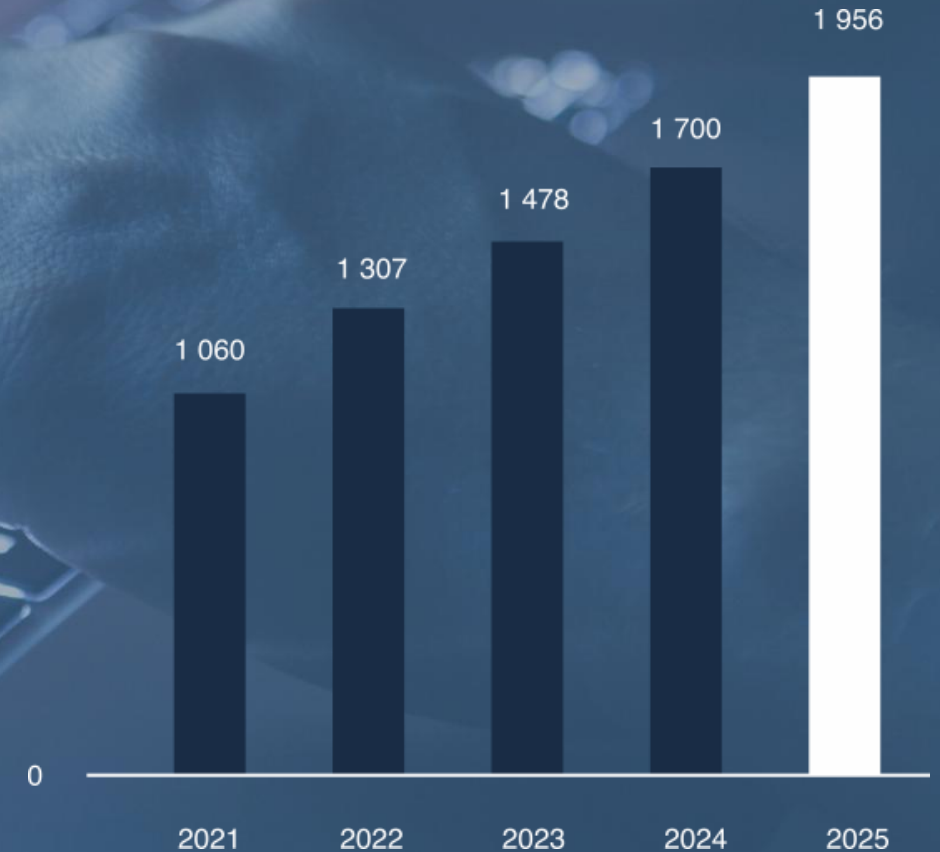
DATASOURCE IN 2025
+9% compared to 2024



ACTIVE USERS

1 956

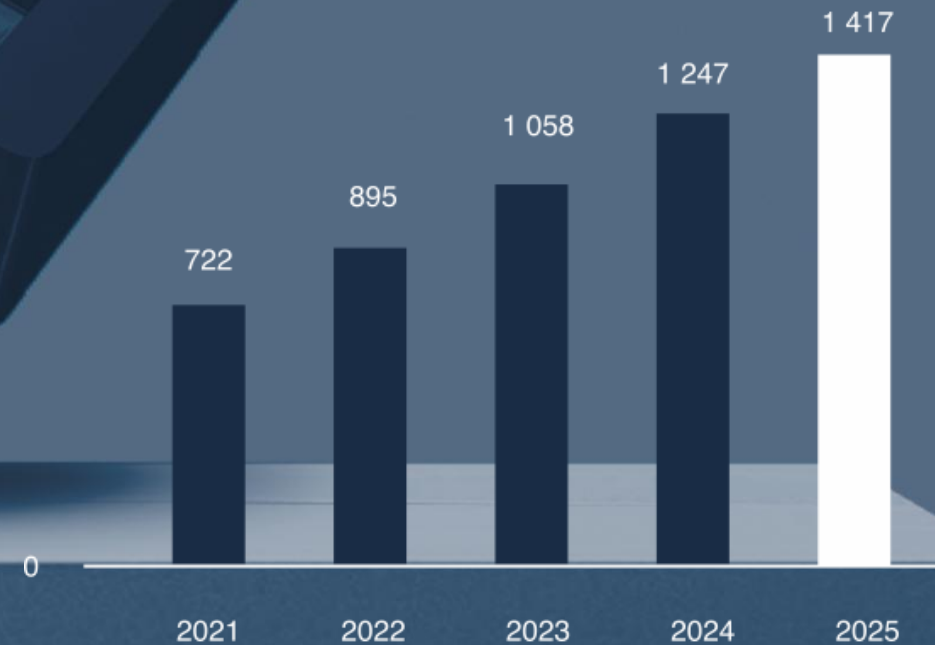
ACTIVE USER ACCOUNTS IN 2025 (mai)
+9% compared to 2024



SD-BOX

1 417

SD-BOX OPERATING IN 2025
+14% compared to 2024

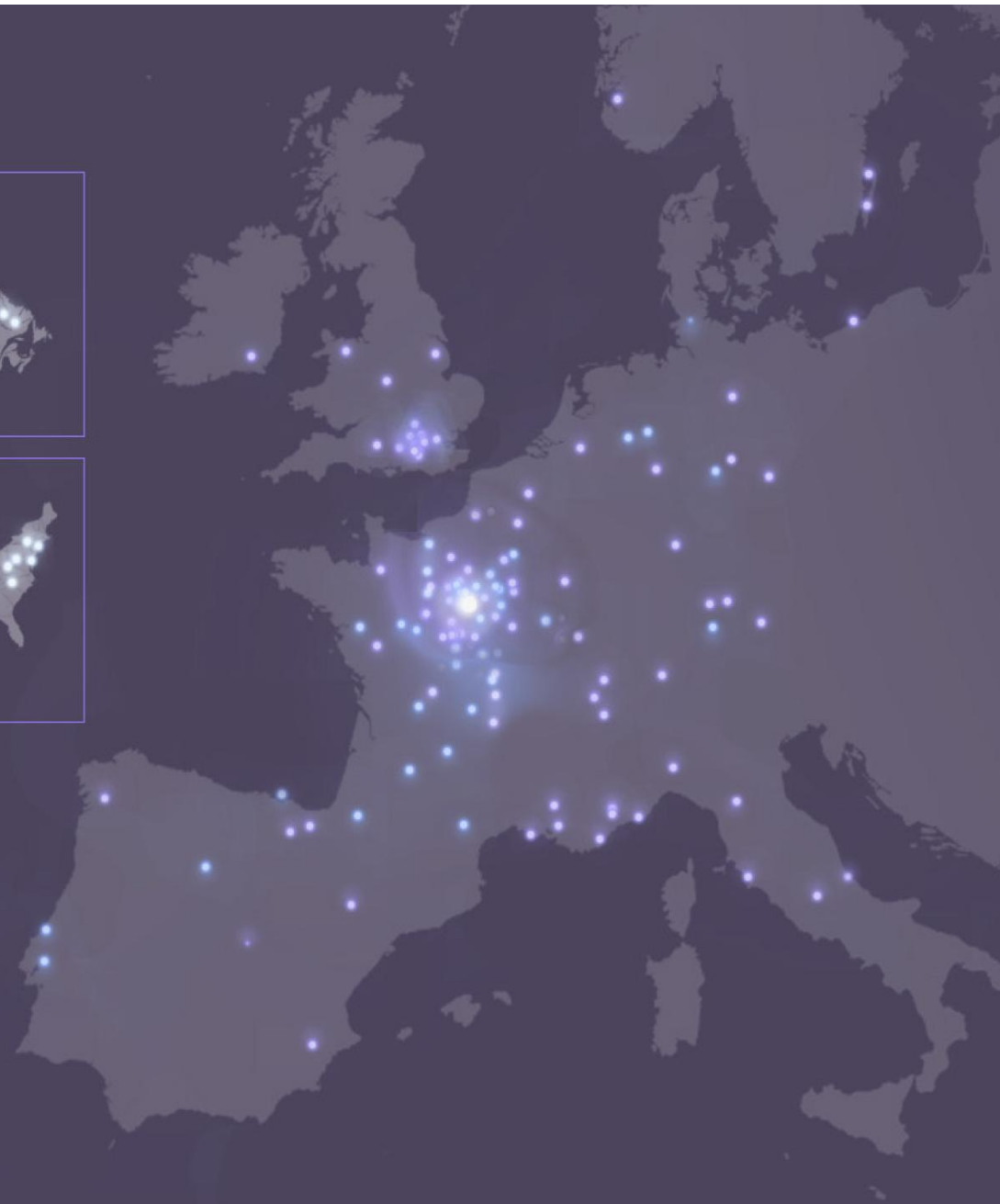




Canada



US



Researchers

From France and all across Europe and North America
Currently hosting about 700 research projects

Some other institutions :

- The Court of Auditors
 - The Treasury
 - European Central Bank
 - OECD secretariat
-

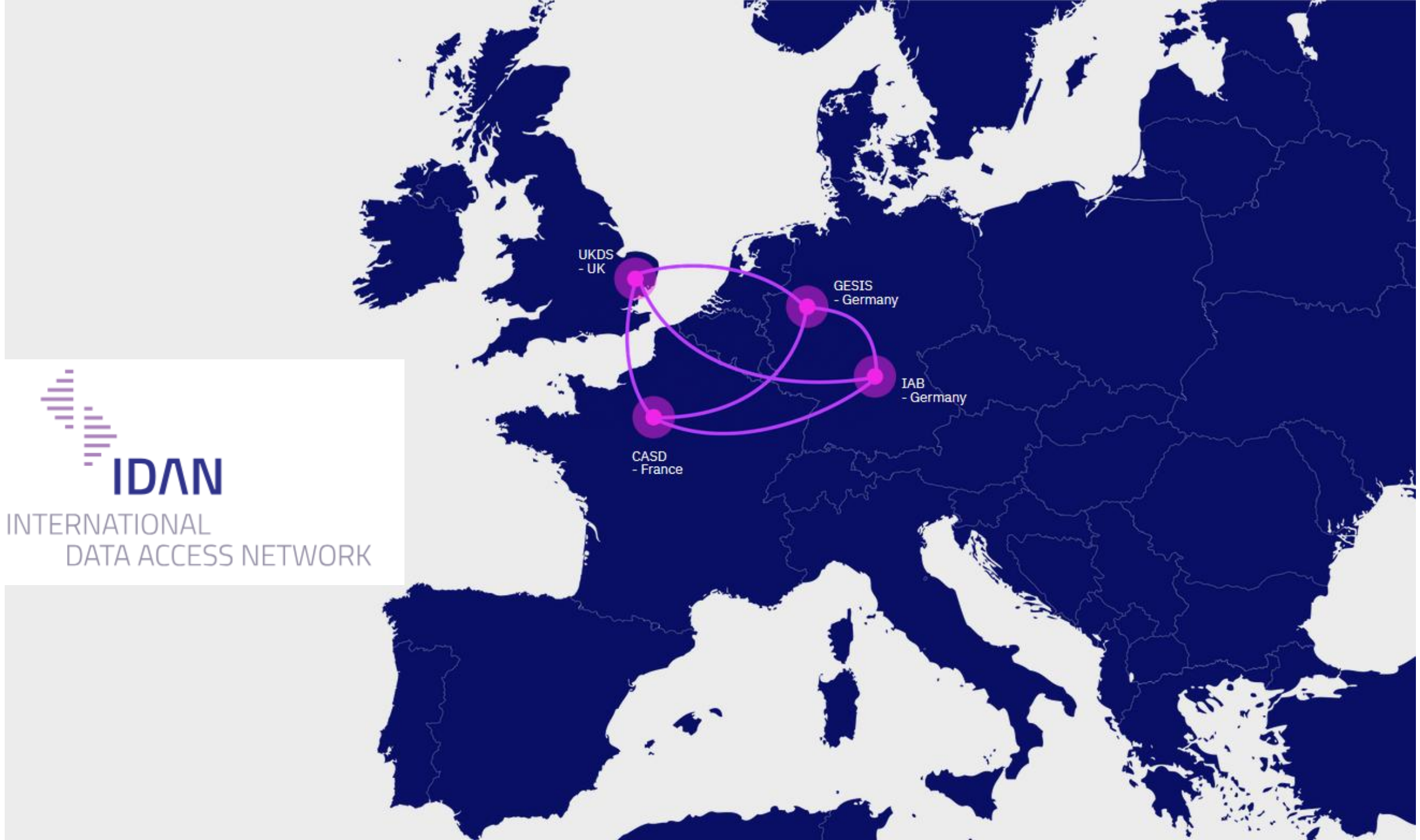
The annual CASD budget is about 5 M€ with some public fundings

The average annual fee for a standard research project is about **4 300€**

The CASD technology is also valued in the private sector



IDAN
INTERNATIONAL
DATA ACCESS NETWORK





Projects



Social status of farmers involved in short food supply circuits' short supply and distribution chains

CESAER - DIJON - 2014



CERGAM - AIX-EN-PROVENCE - 2018



Sustainable towns and the successor project

EIFER - KARLSRUHE - 2018



Knowledge extraction from time and space data through "time and space data mining" methods

IRSTEA - ANTONY - 2017



Explaining the Detailedness of Interorganizational Imitation

Tilburg University - TILBURG - 2017



Studying the erosion of the tax base

DG Trésor - PARIS - 2017



An essay on inequalities and property transfer taxation - France 1980 - 2012

PSE - PARIS - 2017



The costs of research and employment in science: the case for France

IPGP - PARIS - 2017



Projects



Innovation, competition and competitiveness, the role of human asset and energy in the dynamic of French companies

Sciences Po - PARIS - 2014



MEMENTO cohort project carried out by INSERM U988

INSERM - PARIS - 2016



Integration of migrants into the French labour market: the importance of geographical factors

OCDE - PARIS - 2018



An evaluation of hospital activity

Le Point - PARIS - 2017



Finance, organization of labor and corporate social and environmental responsibility

EconomiX - NANTERRE - 2014



Rapid global growth towards the 2020 horizon

BPI France - MAISONS-ALFORT - 2016



Corporate demographics and their establishments in the Paris region

IFSTTAR - MARNE LA VALLEE - 2017



Corporate organisation, wages and production factor dynamics

UCLA - LOS ANGELES - 2018



Relationship between the inappropriateness prescription and functional and cognitive performances, among early preclinical signs of Alzheimer disease and related diseases patients

Hospices Civils de Lyon -
VILLEURBANNE - 2017



Project carried out by the company named "Carte Blanche Partenaires"

Carte Blanche - PARIS - 2016



Project carried out by the company named "INTUITIVE SURGICAL"

Intuitive surgical - VELIZY - 2016



MEMENTO cohort project carried out by INSERM U988

INSERM - PARIS - 2016



Impact of stress on cognitive and brain biomarkers in Memento cohort

INSERM INSERM CIC1401 -
BORDEAUX - 2017



Project carried out by the company named "OpusLine"

OpusLine - PARIS - 2017



Project carried out by the company named "ALTENSE Consulting"

ALTENSE - LANDAS - 2016



Project carried out by the company named "CEMKA EVAL"

CEMKA - BOURG-LA-REINE - 2016

Projects



META PROJECT



7

PILOT PROJECTS
ON INSTAGRAM AND
FACEBOOK DATA

CANADA, AUSTRALIA
AND US (SD-BOX)

RESEARCH ON
WELL-BEEING
IN PARTICULAR FOR TEENAGERS

+ MCL
Meta Content Library
partnership
to facilitate access
to real time data

CONSTANCES PROJECT

770 000

subjects in the cohort (health)
On of the largest cohort in Europe



cohorte
CONSTANCES
Améliorer la santé de demain



Many innovative research projects

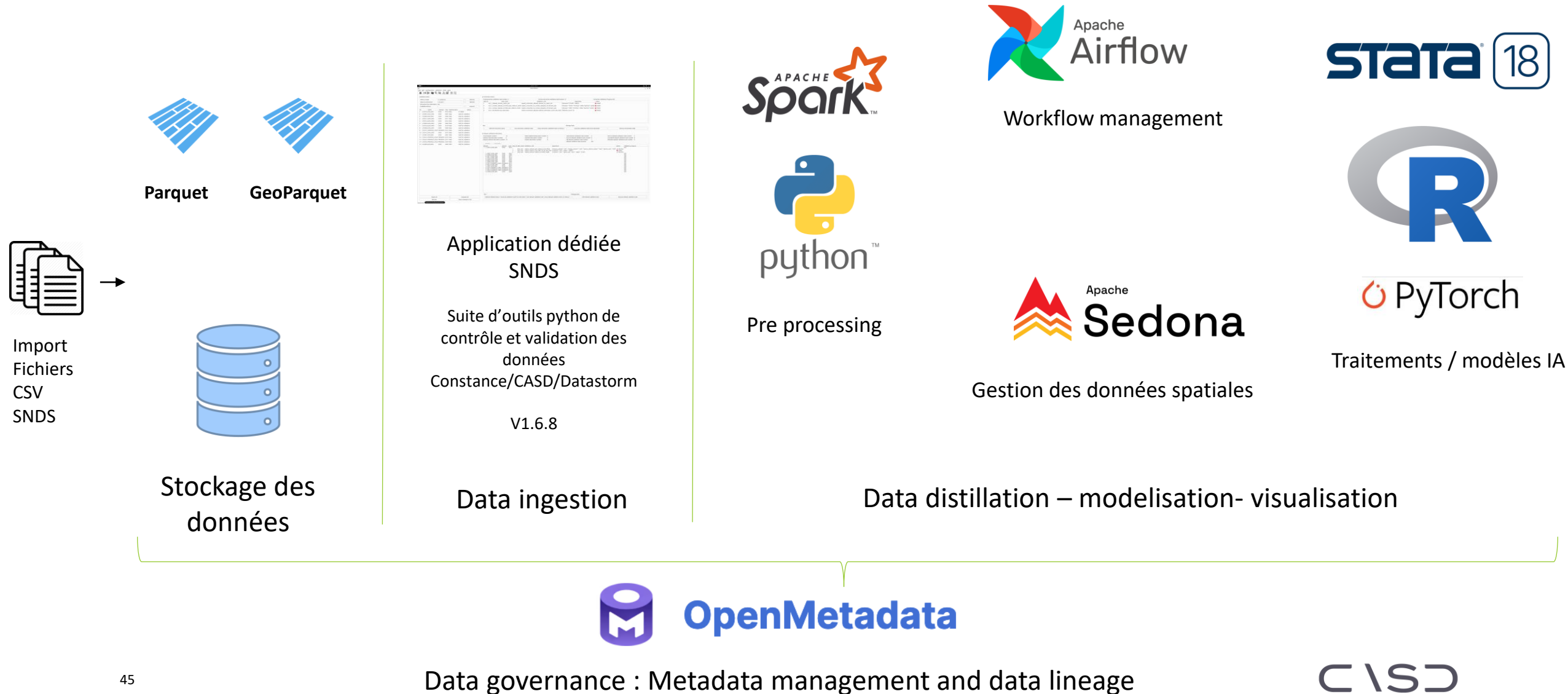
Individual data matching

Professional career
Environmental conditions
DNA
Etc.

Cutting edge tech

expertise on data-
matching
Metadata (DDI)
Pipelines
High volume processing

GRAPH4HEALTH AI PLATFORM (LLM) : CREST, CASD, ESSEC, CONSTANCES



CONCLUSION

- Many things have been done since 2010 thanks to :
 - The data producers involvement for researchers
 - The researchers :
 - help to convince decision makers for opening access to data
 - co-financed the operating costs
 - Produced hundreds of publications
 - The members of the consortium (voluntary funding)
 - The data protection authority (CNIL), the comité du secret statistique

CONCLUSION

- Many things still to do :
 - Get fundings to lower the prices for researchers
 - Simplify the data matching procedure from a legal aspect
 - Foster the secure usage of AI

MERCI

C\SD

1^{er} décembre 2025
ANDOR

Kamel Gadouche, Directeur