



ENVIRONNEMENTS DE CALCUL SÉCURISÉS

9e Journée scientifique des cohortes Constances et Gazel

Kamel GADOUCHE
Directeur du CASD

26 mai 2026

CASD

CASD

CASD

CASD

SECURITE



EXPERIENCE
UTILISATEUR



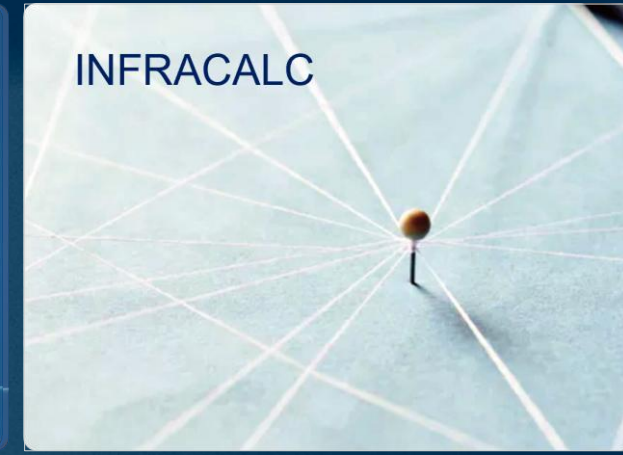
SANTE



IA & LLM



INFRACALC



CONSORTIUM



GROUPE
ENSAE
ENSAI

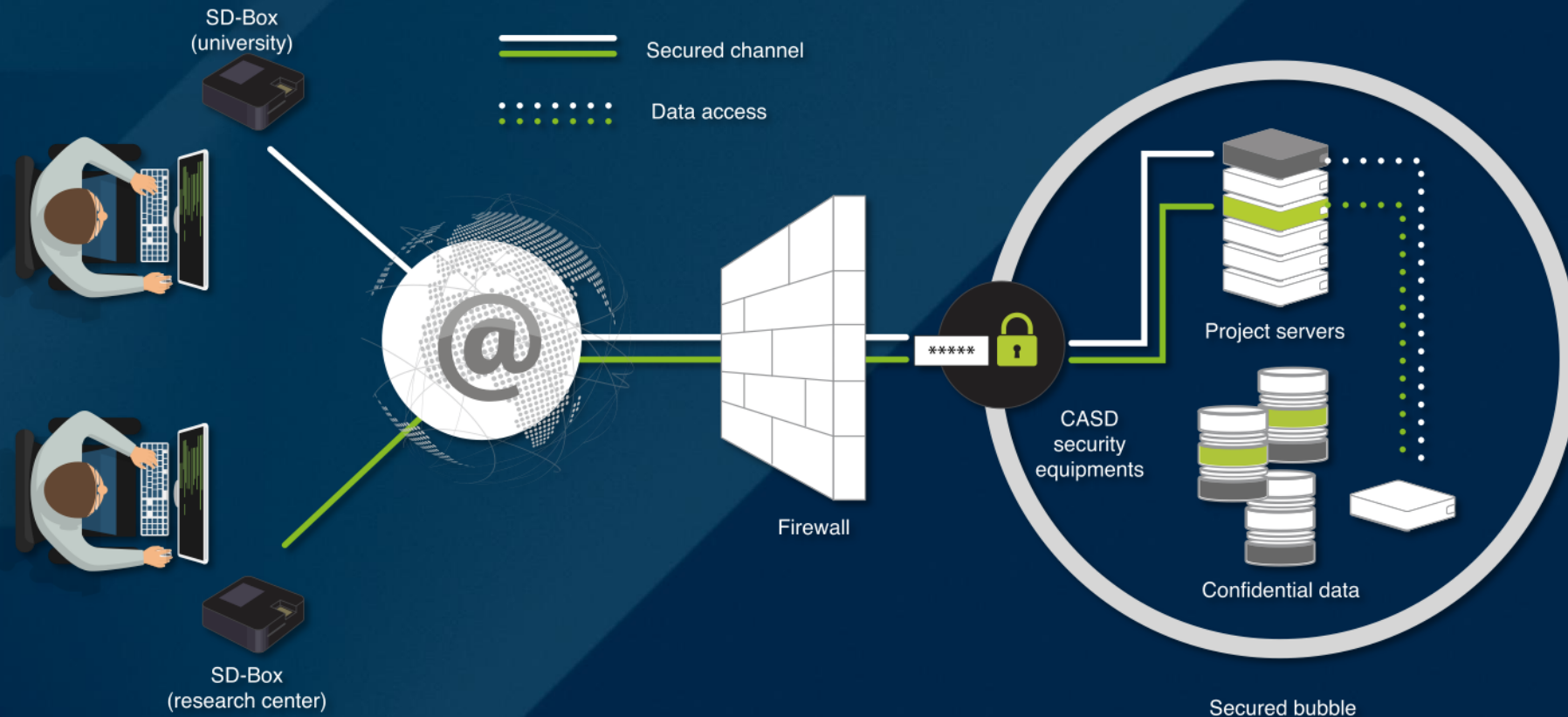


Groupement d'intérêt public
Personne publique à but non lucratif

CASD

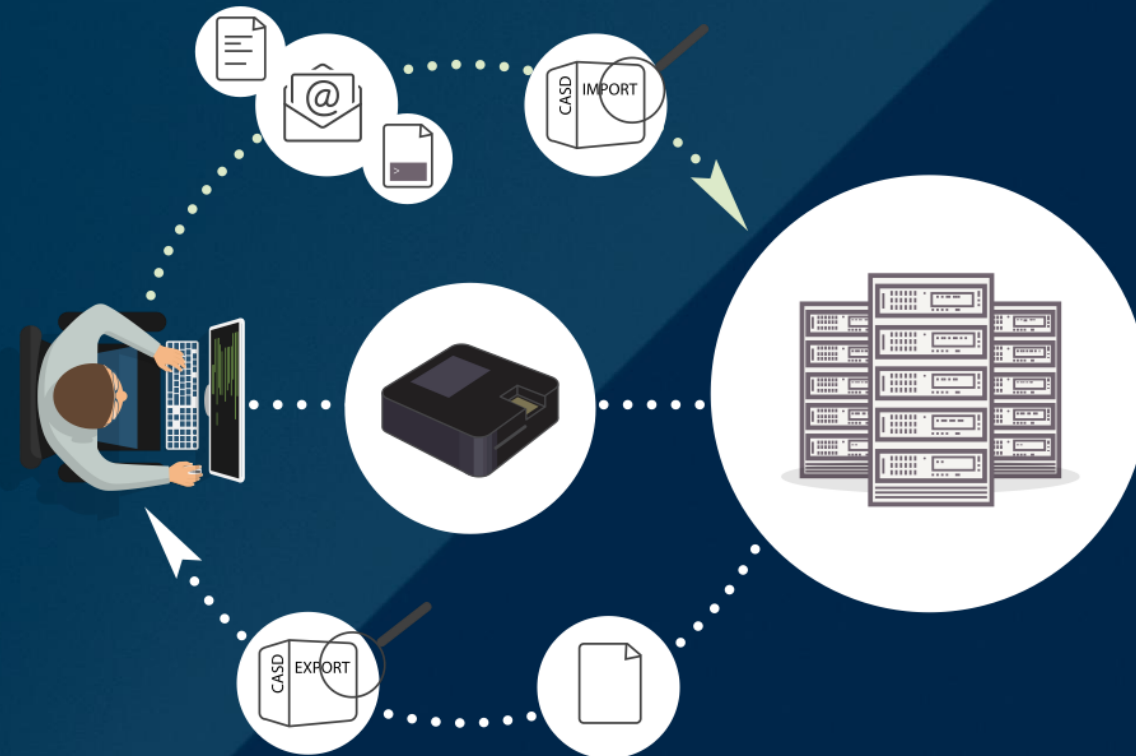
MISSIONS

1. Opérer des environnements de calcul sécurisés (bulles sécurisées)



MISSIONS

2. Valoriser la technologie et l'expertise dans le secteur privé



Insee
Mesurer pour comprendre


DIRECTION GÉNÉRALE DES
FINANCES PUBLIQUES


MINISTÈRE
DE LA JUSTICE
*Liberté
Égalité
Fraternité*


MINISTÈRE
DU TRAVAIL
ET DE L'EMPLOI
*Liberté
Égalité
Fraternité*


MINISTÈRE
DE LA SANTÉ
ET DE L'ACCÈS
AUX SOINS
*Liberté
Égalité
Fraternité*


MINISTÈRE
DE L'ÉDUCATION
NATIONALE
*Liberté
Égalité
Fraternité*


MINISTÈRE
DE LA TRANSITION
ÉCOLOGIQUE
*Liberté
Égalité
Fraternité*


SÉCURITÉ SOCIALE
l'Assurance
Maladie

bpifrance

 Meta

 ATIH | Nos données
au service
de la Santé


BANQUE DE FRANCE
EUROSYSTEME



Nos données
au service
de la Santé



« 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... »

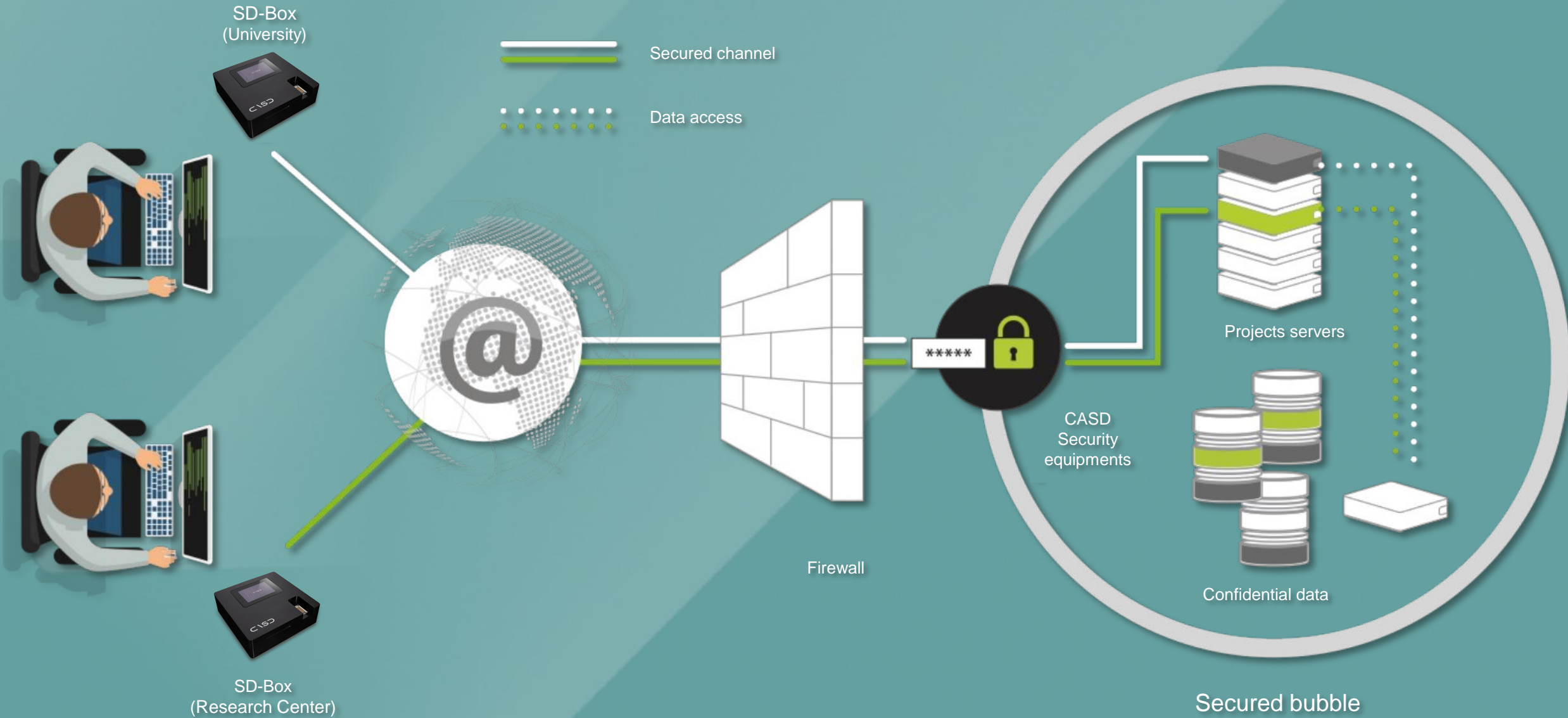
Science

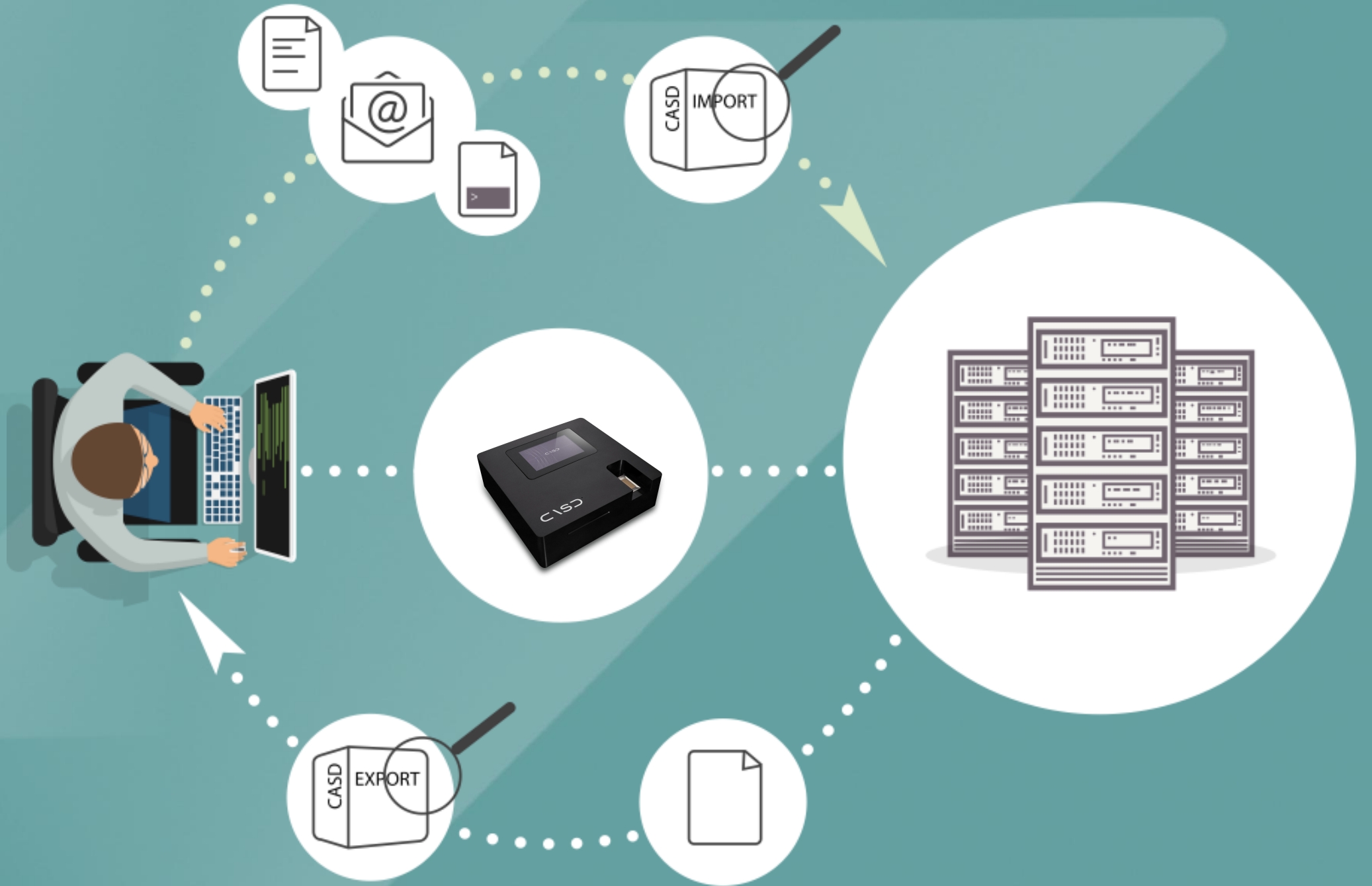
« CASD :
une bulle pour protéger les données »

Le Monde



Accès à distance à des bulles de calcul sécurisées





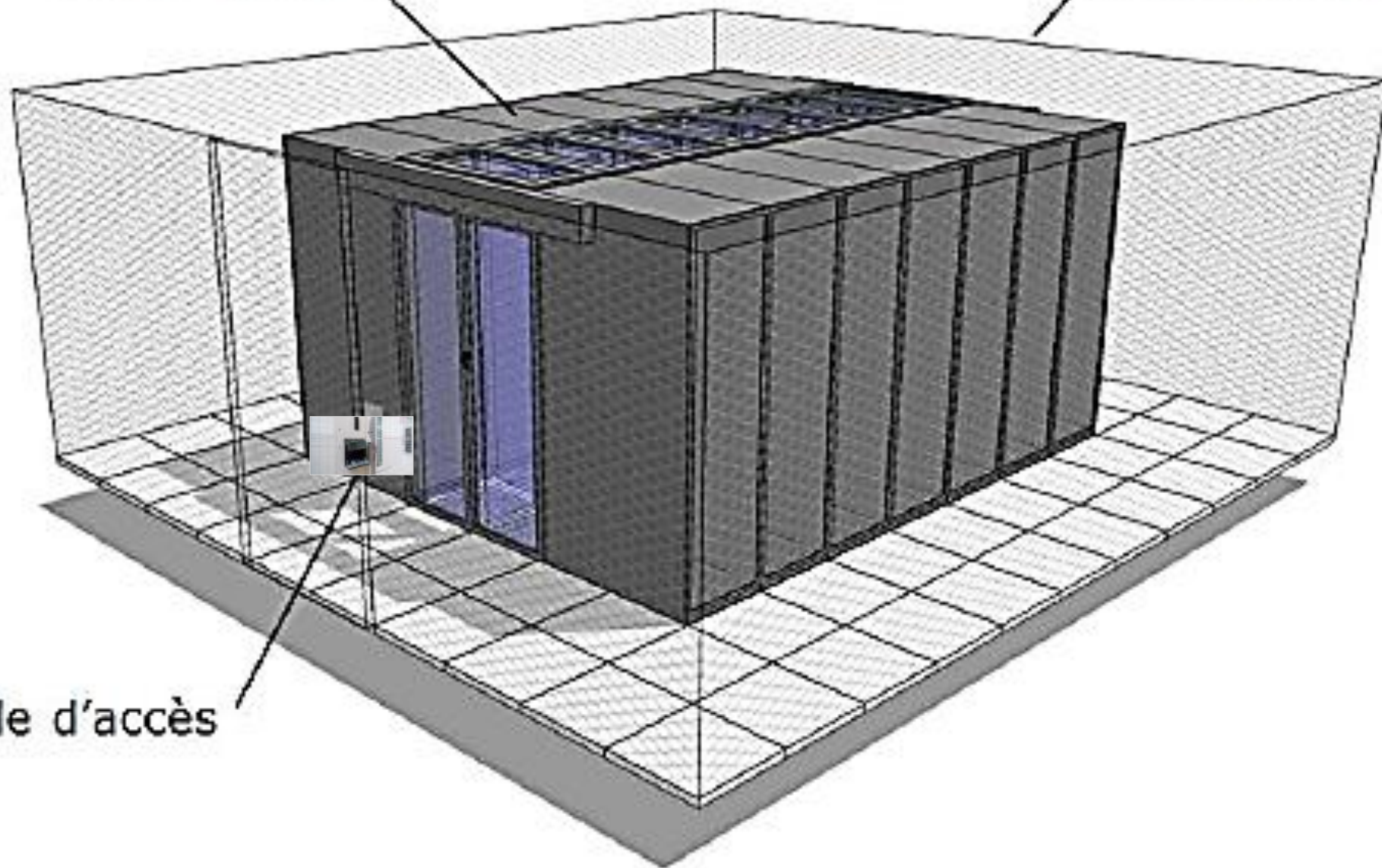


Datacenters



Couloir Froid

Cage de sécurité



Contrôle d'accès



HandKey II
HandKey II
HandKey II

CASD

SECURITE

CASD : Premier hébergeur de données en Europe certifié officiellement RGPD (Art. 42)



RGPD
(Art. 42) selon le référentiel Europrivacy



Europrivacy
Applicant

CNIL.
Autorisation de traitement
2014-369



ISO
27001
Sécurité de l'information
FR078964

ISO
27701
Protection des données personnelles
Privacy / FR060212



HDS
Hébergeur de données de santé
FR060212

SNDS
Homologation au référentiel
de sécurité des données de santé

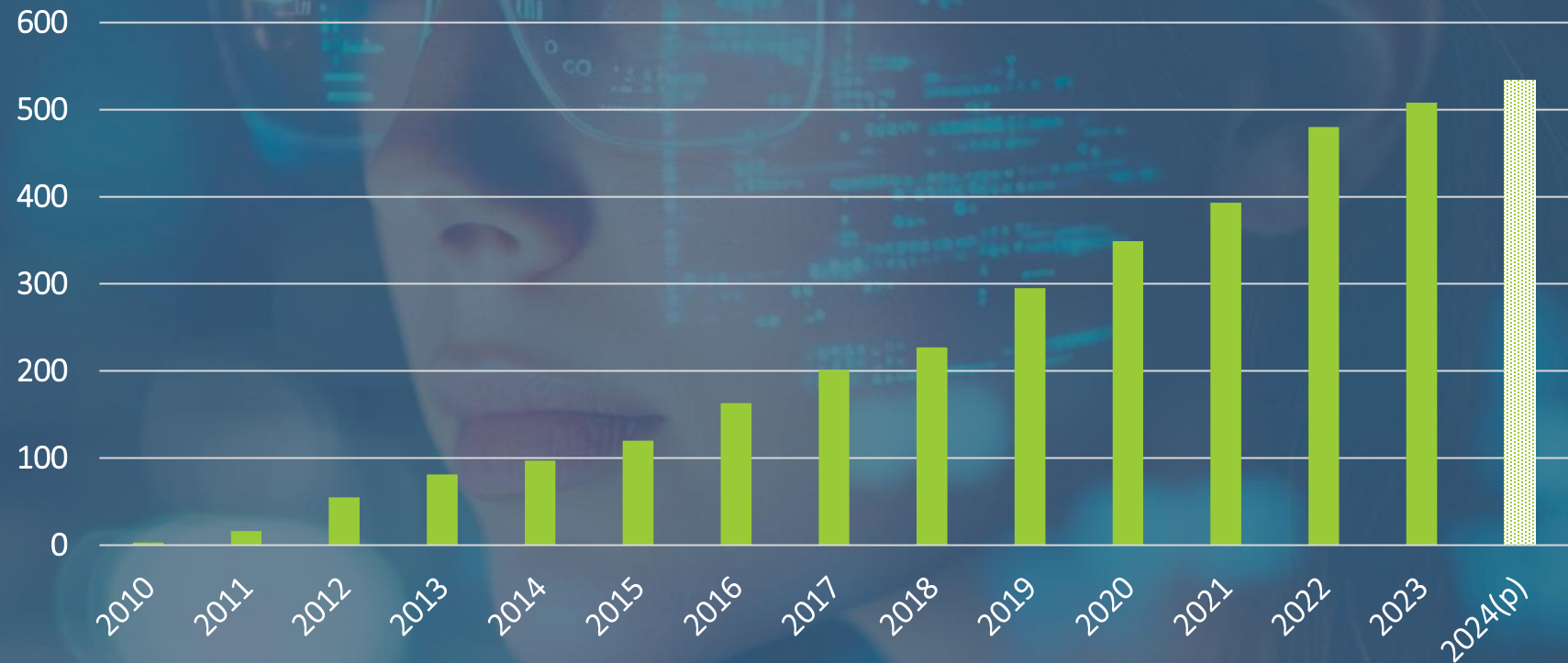


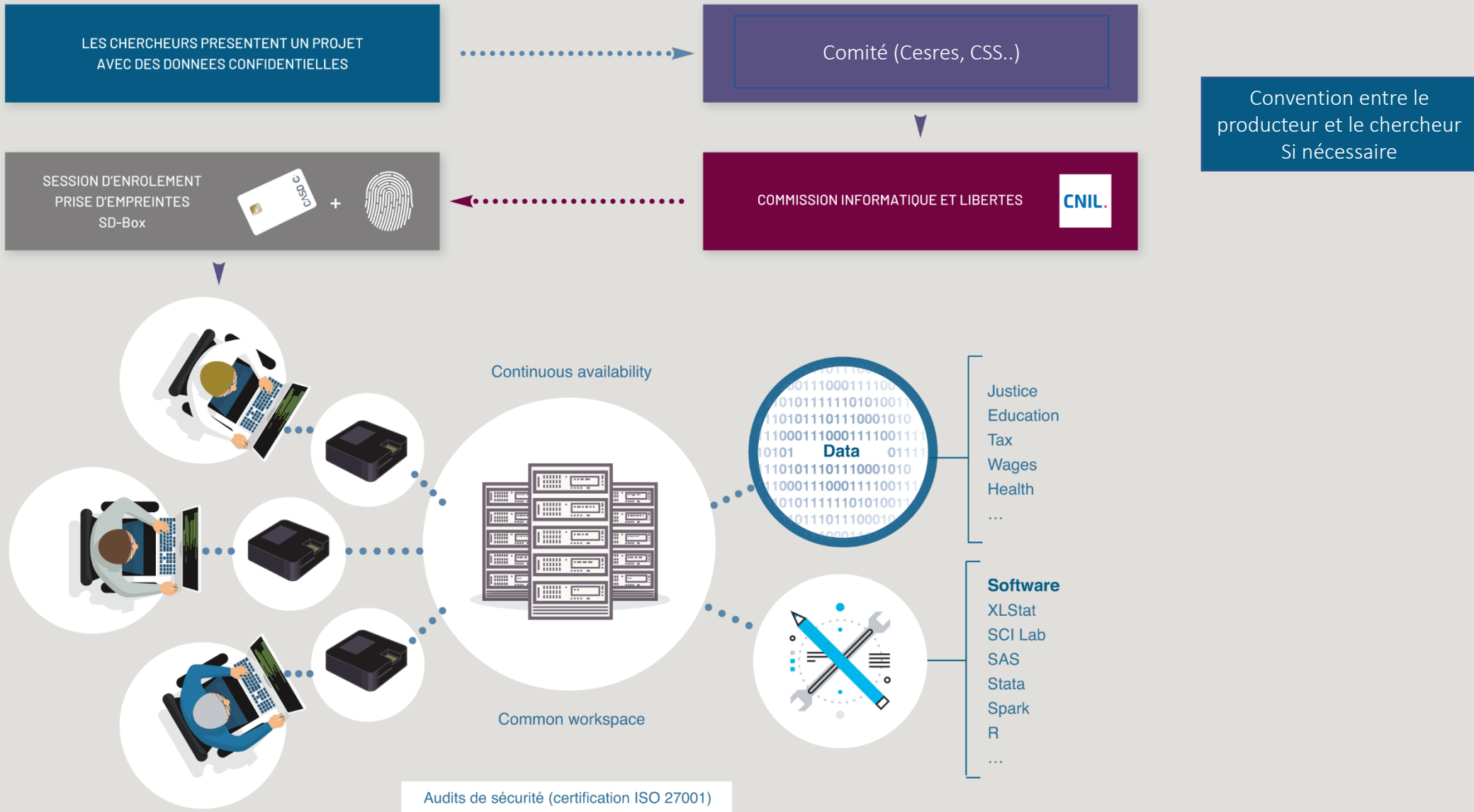
MORE SECURITY, MORE DATA

655

DATASOURCES EN 2025
+9% par rapport à 2024

Nombre de sources de données mises à disposition au CASD





Plus de 80
logiciels
et packages



Onyxia



TensorFlow



DuckDB

VTL

TREVAS^{JS}
VTL ENGINE



TS-DEMO002

04.Demo_geo.ipynb A

```

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)

```

```

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()

```

main+ 0 0 5 0 CRLF Cell 1 of 12 6:20 PM 6/12/2024

Home

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 (pykernel)

```

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)

[ ]: 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()

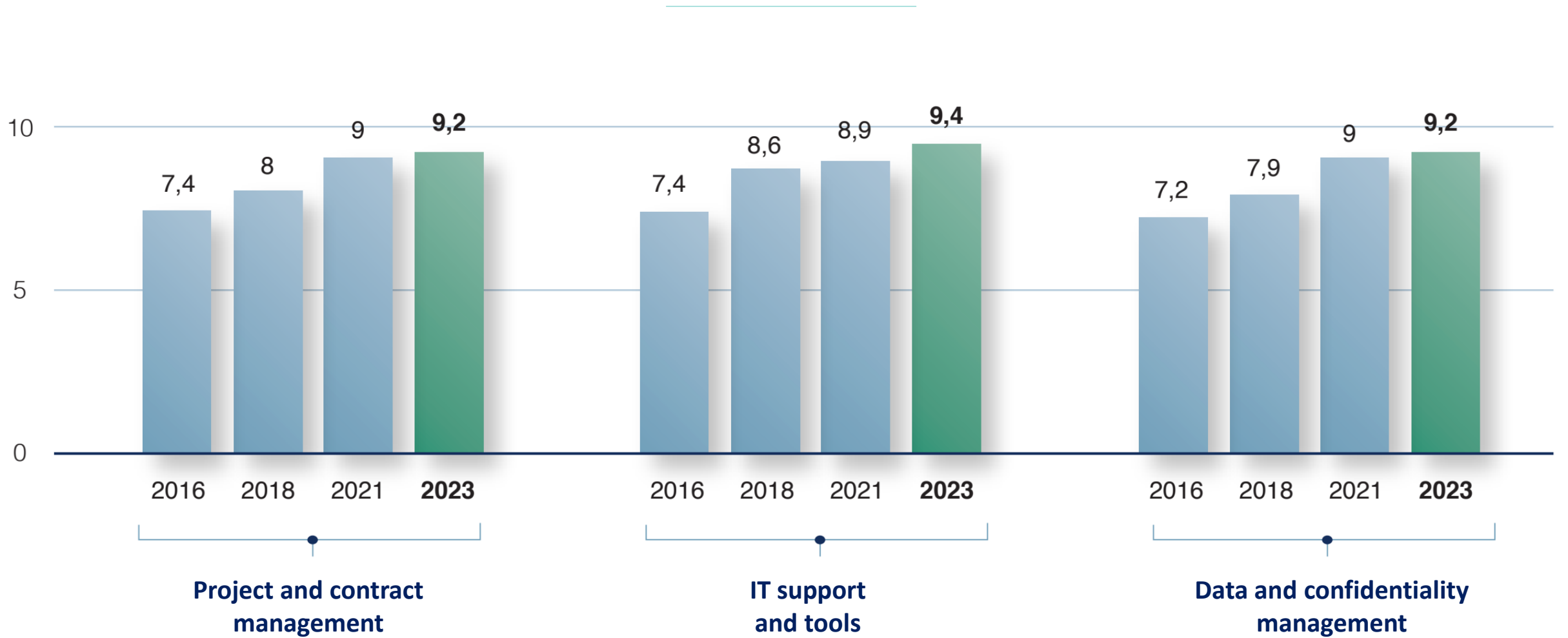
```

2 Read shape file and build geo pandas dataframe

Percentage of the polpulation who has Cancers

6:21 PM 6/12/2024

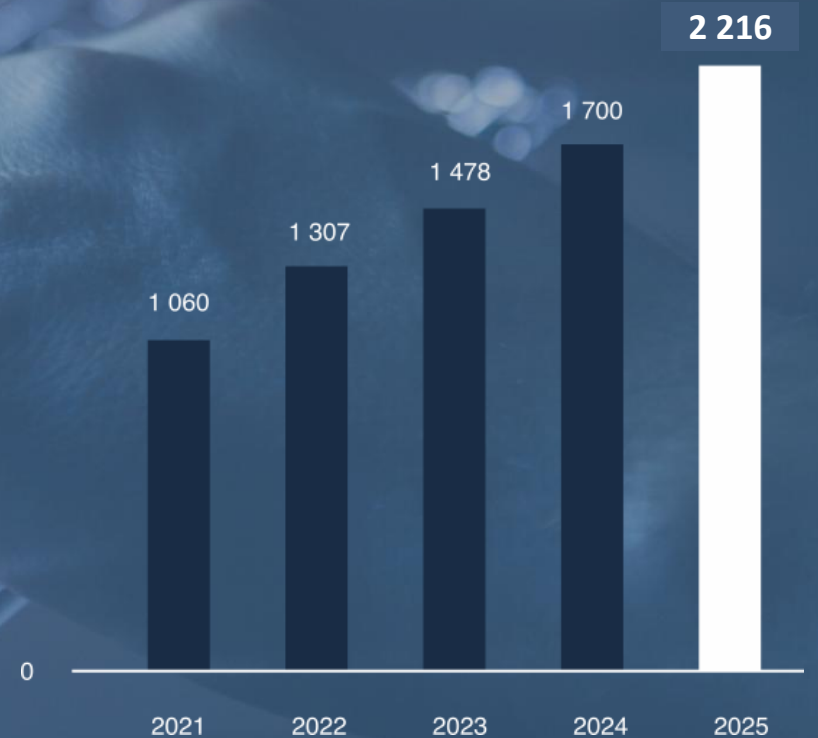
ENQUÊTE DE SATISFACTION CONCERNANT LES SERVICES DU CASD
2016, 2018, 2021 AND 2023



UTILISATEURS

2 216

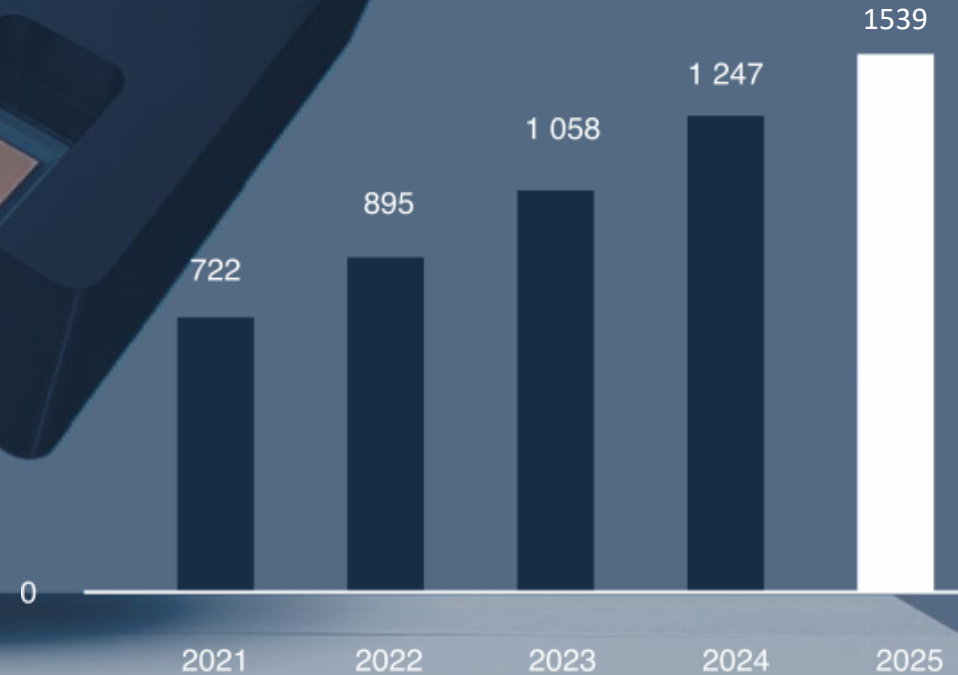
COMPTES UTILISATEURS ACTIFS EN 2025
+16% PAR RAPPORT À 2024



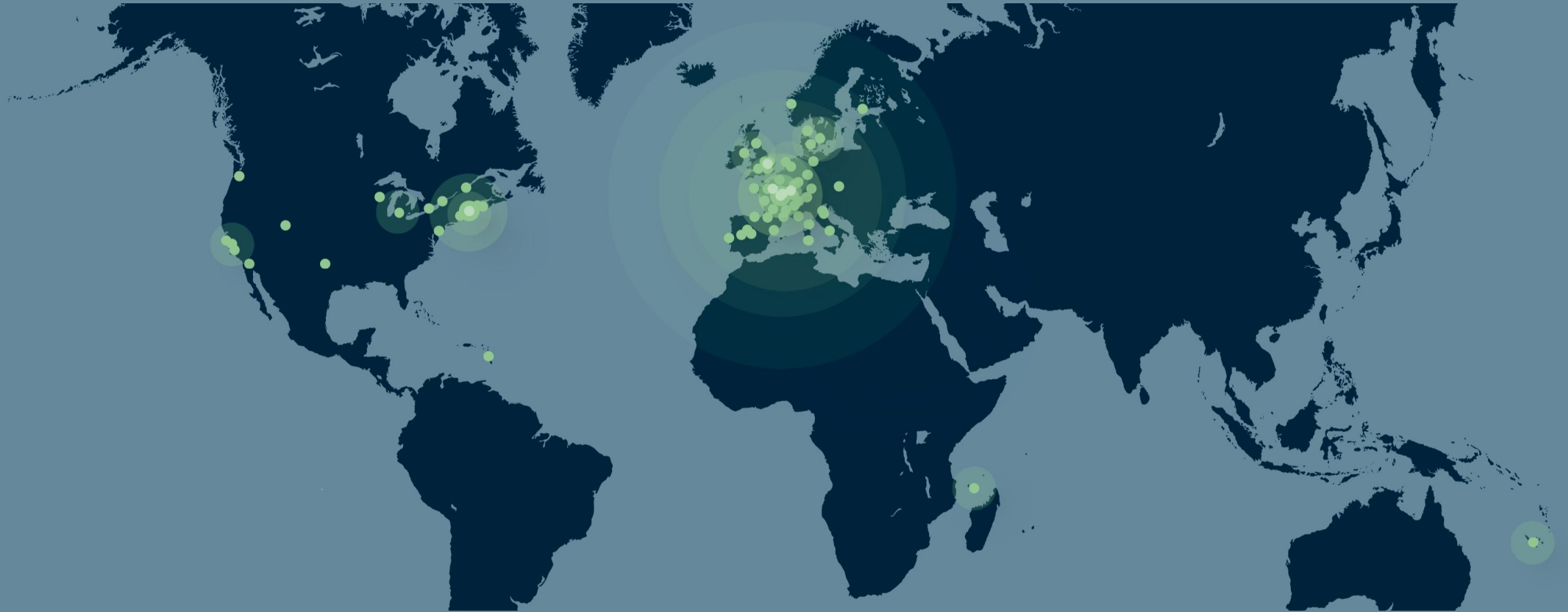
SD-BOX

1 539

SD-BOX DÉPLOYÉES EN 2025
+14% PAR RAPPORT À 2024



SD-BOX



Documentation DDI

Filtrer les sources ↻

x Santé – Handicap – Dépendance
x

Source

Sélectionnez des sources ▼

Millésimes

1954
●
●
 2025

Producteur

Sélectionnez des producteurs ▼

NOMBRE DE RÉSULTATS

34

Q

SOURCE PMSI - Médecine, chirurgie, obstétrique et odontologie (MCO)

La médecine, chirurgie, obstétrique et odontologie fait l'objet d'un recueil d'informations administratives, médicales et économiques réalisé à chaque séjour hospitalier ou clinique d'un patient.

Thème : Santé – Handicap – Dépendance
Millésime(s) : 2025, 2024, 2023, 2022, 2021, 2020, 2019, 2018, 2017, 2016, 2014

Projets : 49
Publications : 7

SOURCE Cohorte Constances et Gazel (Cohorte Constances et Gazel)

Les cohortes CONSTANCES et GAZEL collectent des données qui couvrent diverses dimensions et proviennent de différentes source du domaines de la santé.

Thème : Santé – Handicap – Dépendance
Millésime(s) :

Projets : 42
Publications : 1

SOURCE Données du Système national des données de santé (SNDS)

Le SNDS regroupe les principales bases de données de santé publiques existantes. Chaque projet bénéficie d'extractions suivant sa thématique de recherche et les autorisations obtenues.

Thème : Santé – Handicap – Dépendance
Millésime(s) :

Projets : 25
Publications : 0

Exemples en santé

- Entrepôts de données de santé (EDS)
 - DOREMY (APHP) : leucémies aiguës Myéloïdes pédiatriques
 - I-CML-Ped (CHU Poitiers) : leucémies périodiques chroniques pédiatriques
 - Cancers en Alsace (Groupe Hospitalier de la Région de Mulhouse et Sud-Alsace)
- Registres et cohortes
 - FRENCHIE (APHP) : infarctus du myocarde
 - Constances et Gazel
- Essai clinique apparié SNDS
 - DIABSURGE (CHU Lille) : patients obèses diabétiques de type 2
- Ministère de la santé : Direction de la recherche, des études, de l'évaluation et des statistiques

GAZEL Cohorte



20 000

Volontaires dans la cohorte

intégration de pipelines de données

37 ans de suivi longitudinal

EDS

Entrepôt de données de santé

CONSTANCES Cohorte

655 000

Volontaires dans la cohorte
Une des plus grande cohorte d'Europe



cohorte
CONSTANCES
Améliorer la santé de demain



plusieurs dizaines de projets de recherche

Appariements au niveau individuel

SNDS

Carrière professionnel

Données de formulaires et d'examens de santé

Environnement

Etc.

Cutting edge tech

expertise en appariement

Metadata (DDI)

Pipelines

Traitement de données de
gros volumes



Comparaison de l'état de santé de sujets par le VIH à celui de la population générale, en prenant en compte les conditions socio-économiques et le mode de vie de sujets (JASMIN)

Institut National de la Santé et de la Recherche Médicale - PARIS



Etude de facteurs de risque cardiométabolique non traditionnels dans la cohorte Constances

UMS 011 Inserm-UVSQ "Cohortes épidémiologiques en population" - VILLEJUIF



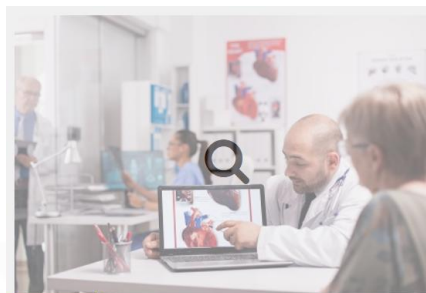
Conséquences de la sclérose en plaques sur les parcours professionnels en France

Ecole des Hautes Etudes en Santé Publique - RENNES



Description et coûts de la prise en charge de la stéatose hépatique non alcoolique (NASH) en France (chez les patients diabétiques et non diabétiques) (quid nash)

Assistance publique - Hôpitaux de Paris - PARIS



Analyse du dépistage systématique de la cardiopathie ischémique chez les patients diabétiques de type 2 à très haut risque cardiovasculaire

Centre Hospitalier Universitaire de Bordeaux - TALENCE



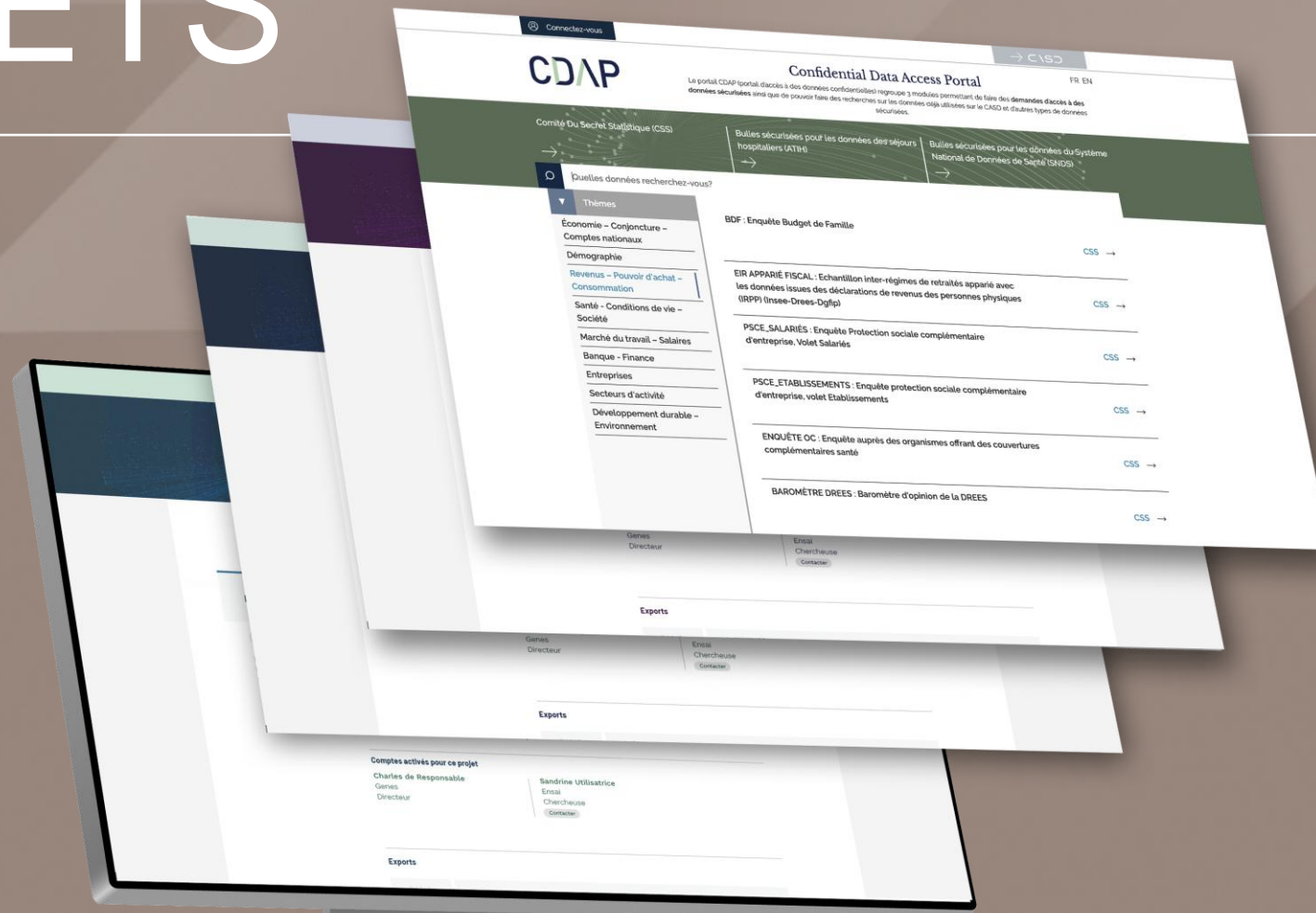
Utilisation de la mammographie, du test de recherche de sang occulte dans les selles et de la coloscopie au-delà de l'âge limite pour le dépistage des cancers en Europe

Luxembourg Institute of Health - STRASSEN LUXEMBOURG - LUXEMBOURG



188 Projets
en santé

PROJETS



Développements R&D / IA

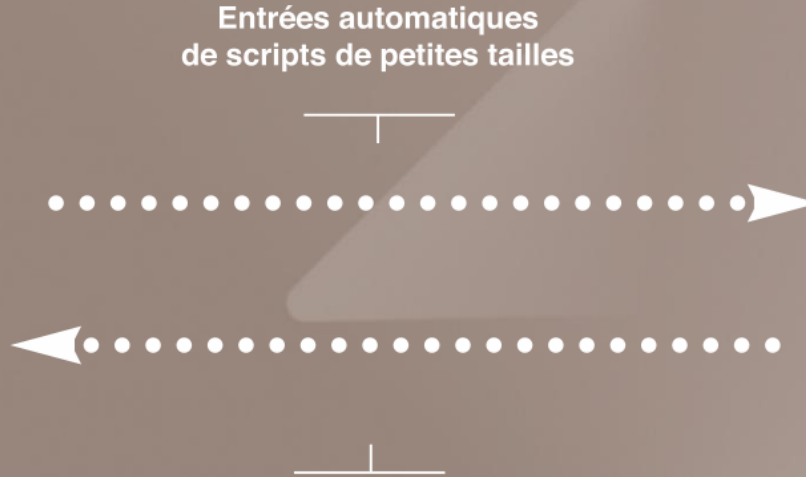
FONCTIONNALITÉS DÉVELOPPÉES

Enrôlement sécurisé à distance



IP domicile/ IP fixe

Entrées automatiques
de scripts de petites tailles



Bulle sécurisée

Session recording

Scellement par blockchain des logs
Algorithme d'analyse des sessions
(indexation)

Outils datascience

Spark, Onyxia...

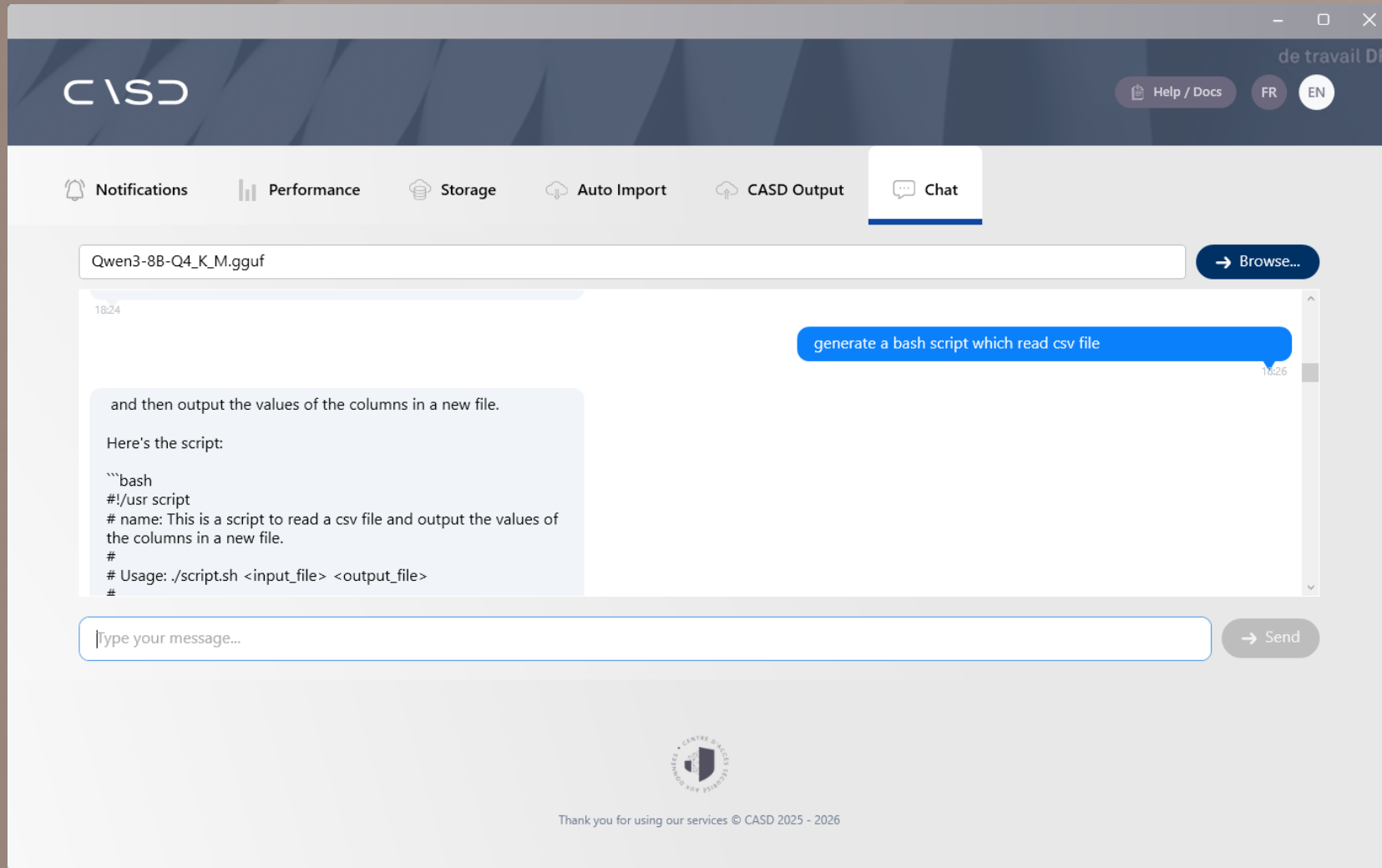
Algorithme de machine learning sur les sorties de fichiers : Colysée

Notation des sorties sur un « risque » de non respect du secret

Vérification manuelle des sorties les plus à risque et d'un échantillon des autres sorties

Alimentation du modèle de ML pour l'entraînement

LLM – assistant virtuel



The screenshot displays the CASD web interface with a chat window open. The chat window shows a user input field containing the filename "Qwen3-8B-Q4_K_M.gguf" and a "Browse..." button. The chat history shows a user message at 18:24: "generate a bash script which read csv file". The assistant's response at 18:26 includes the instruction "and then output the values of the columns in a new file." and provides a bash script. The script is as follows:

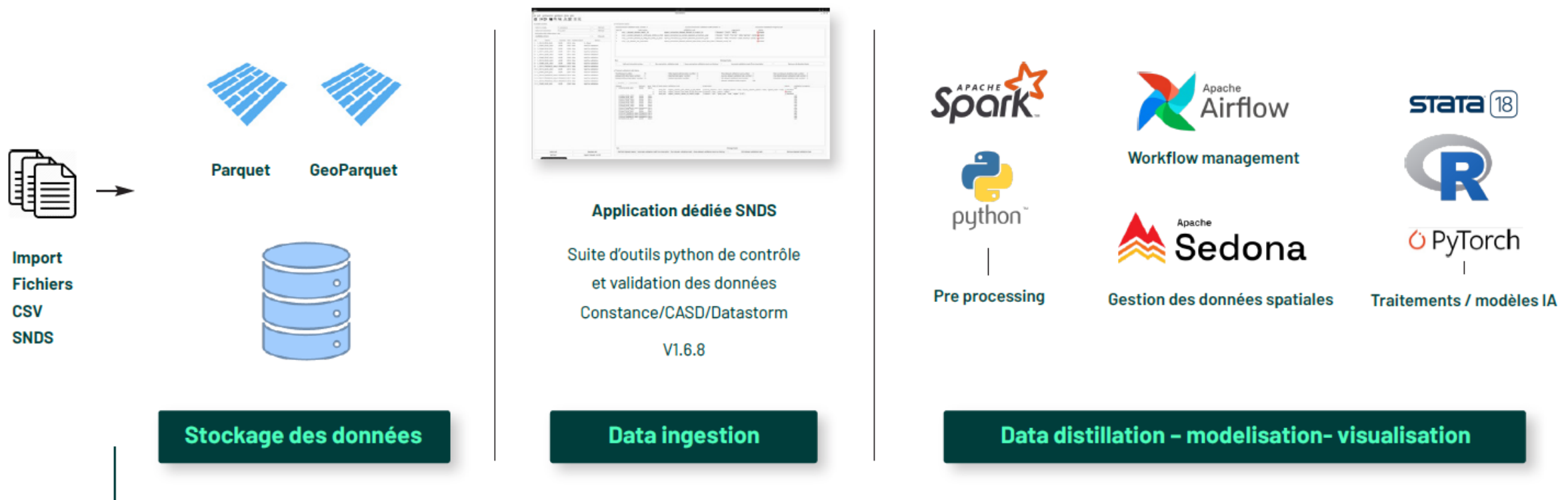
```
Here's the script:

```bash
#!/usr script
name: This is a script to read a csv file and output the values of
the columns in a new file.
#
Usage: ./script.sh <input_file> <output_file>
#
```

At the bottom of the interface, there is a text input field with the placeholder "Type your message..." and a "Send" button. The footer of the page features the CASD logo and the text "Thank you for using our services © CASD 2025 - 2026".

# PLATEFORME DU PROJET GRAPH4HEALTH

Le projet GRAPH4HEALTH vise à traiter 10 années quasi exhaustives de données du SNDS (DCIR et PMSI en population entière entre 2008 et 2018) appariées avec les données des praticiens, qui vise à étudier les mécanismes d'allocation des patients aux acteurs de l'offre de soins, via des méthodes d'analyse de graphes (réseaux au sens mathématique) et d'IA. De nombreux premiers résultats sont très prometteurs en matière de publication.



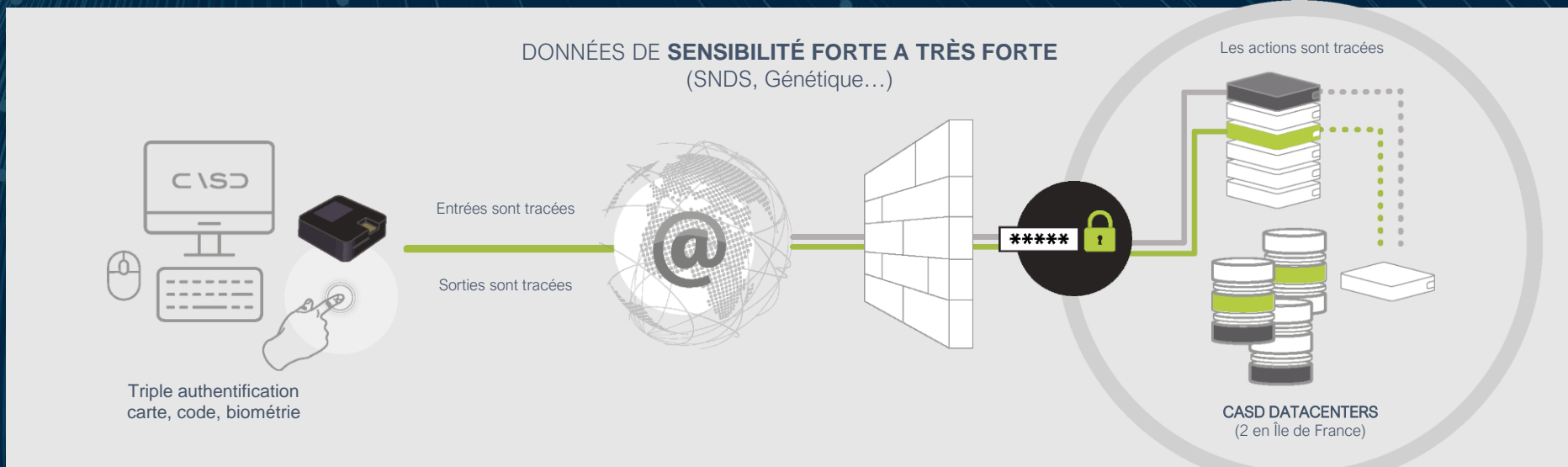
 **OpenMetadata**

**Data governance : Metadata management et data lineage**

INFRA $\Delta$ CALC 

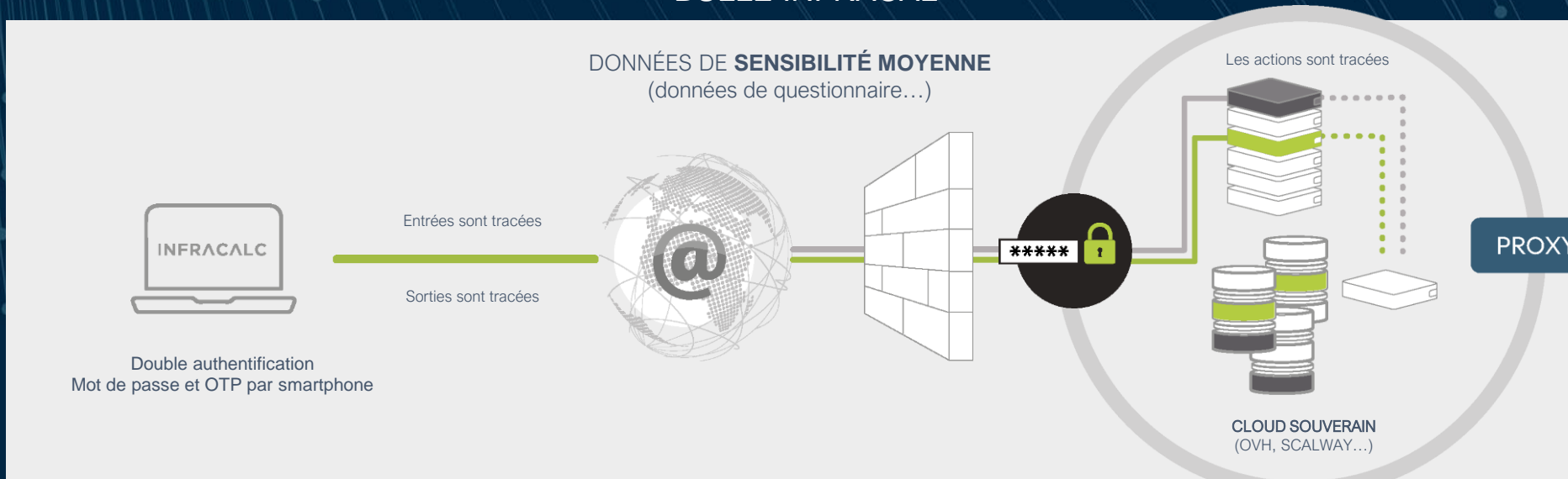
Des bulles sécurisées  
pour des données  
moins sensibles

## BULLE CASD



PAS  
D'ACCÈS  
EXTERNE

## BULLE INFRACAL



API  
EXTERNE  
CONSULTABLE

INFRACALC



pour des données moins sensibles (pilote)

---

Accès depuis le poste de travail (sans boîtier)

Double authentification

Bulles sécurisées avec accès potentiel API

Réutilisation des développements : import, exports, sessions

Hébergement dans un cloud souverain

Gouvernance des données (CASD ou INFRACALC)