
Le CASD : un dispositif d'accès sécurisé à des données confidentielles très détaillées

Gérer ses données sur le plateau de Saclay

11 février 2025

Rémy Marquier



Le CASD

- **Un Groupement d'Intérêt Public (GIP) à but non lucratif** composé de six membres : GENES, INSEE, CNRS, HEC Paris, Ecole Polytechnique et Banque de France
- **Un dispositif sécurisé d'accès à des données confidentielles** ayant bénéficié du financement « Equipement d'Excellence » (EQUIPEX) de la 1^{ère} vague du Programme d'investissements d'avenir (PIA)
 - **Notre mission principale** : organiser et mettre en œuvre des services d'accès sécurisé aux données confidentielles pour les utilisateurs poursuivant des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation
 - Mission secondaire : valorisation de la technologie dans le secteur privé
 - **Notre objectif** : fournir un accès hautement sécurisé aux utilisateurs de données dans les meilleures conditions de travail possibles tout en minimisant les coûts facturés

Fichier brut

nom	prénom	rue	ville	département	région	Tel	salaire	poste	Siret employeur
Dupont	Paul	Albert Camus	Orsay	91	IDF	6025212	45072	Gardien	15624857954
Durant	Julie	des sentiers	Palaiseau	91	IDF	6959849	58233	Dirigeant	59864785668
Morel	Pierre	Marie Curie	Serre	78	IDF	6446827	23392	Ouvrier	95684759666

Fichier « Secure Use File »

Pseudonymiser avec risque élevé

CASD

Ilôts (Iris)	ville	département	région	salaire	poste	Siret employeur
0103	Orsay	91	IDF	45072	Gardien	15624857954
0141	Palaiseau	91	IDF	58233	Dirigeant	59864785668
0102	Serre	78	IDF	23392	Ouvrier	95684759666

Fichier FPR « Scientific use file »

Anonymiser avec faible risque

Diffusion très restreinte

département	région	salaire	poste	Siret employeur
91	IDF	45072	Gardien	15624857954
91	IDF	58233	Dirigeant	59864785668
78	IDF	23392	Ouvrier	95684759666

Fichier open data « Public use file »

Anonymiser avec risque nul

Diffusion publique

région	Salaire (K€)	CSP
IDF	40-50	Employés
IDF	50-60	Chefs entreprise
IDF	20-30	Ouvrier

Les enjeux de l'accès sécurisé aux données

- **La sécurité, un réel enjeu pour permettre l'accès aux données :**
 - Garantir un niveau de sécurité élevé **pour que les producteurs de données mettent à disposition davantage de données en confiance**
 - Authentification forte par biométrie pour un accès distant
 - L'atelier de sensibilisation (enrôlement) est obligatoire et à renouveler tous les 6 ans
 - Processus formel d'habilitation avant d'accéder aux données
- **Usages :**
 - Répondre aux besoins des utilisateurs en termes d'environnement de travail (logiciels, configuration...)
 - Partage au sein d'un projet
- **Égalité de traitement entre tous les utilisateurs**

→ Grâce à ce dispositif, par exemple, l'accès aux données fiscales a été rendu possible

→ **La sécurité mise en place va permettre l'accès à distance à d'autres sources de données confidentielles ainsi que le traitement croisé de différentes sources**

Conformité **RGPD** et **autorisation CNIL** (n°2014-369)



Certifications **ISO 27001** et **ISO 27701** (RGPD)
Certification « **hébergeur de données de santé** » et homologation « **RSDS** »

Audits techniques réguliers de sécurité réalisés



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

CNIL.

Autorisation de traitement
2014-369

Fonctionnement standard



Les démarches préalables de demande

- Le cas général : autorisation du producteur après avis du CSS
- Pour les données de santé : démarches spécifiques auprès de la CNIL
- Pour les appariements sur données personnelles : nécessité de réaliser un PIA
- Le CASD vérifie (et accompagne) la conformité avant toute mise en œuvre

Concrètement

Portail d'accès aux données

https://cdap.casd.eu

Connectez-vous

C\SD

FR EN

CDAP

Confidential Data Access Portal

Le portail CDAP (portail d'accès à des données confidentielles) regroupe 3 modules permettant de faire des **demandes d'accès à des données sécurisées** ainsi que de pouvoir faire des recherches sur les données déjà utilisées sur le CASD et d'autres types de données sécurisées.

Comité Du Secret Statistique (CSS)

Bulles sécurisées pour les données des séjours hospitaliers (ATIH)

Bulles sécurisées pour les données du Système National de Données de Santé (SNDS)

Quelles données recherchez-vous?

Thèmes

- Économie - Conjoncture - Comptes nationaux
- Démographie
- Revenus - Pouvoir d'achat - Consommation
- Santé - Conditions de vie - Société
- Marché du travail - Salaires
- Banque - Finance

EEC : Enquête Emploi en Continu CSS →

EXTRAPRO : Base de données Extrapro, base de sondage de l'enquête auprès des sortants de contrats de professionnalisation CSS →

I-MILO : Système d'information des missions locales CSS →

OEDIPE MILO : Appariement ad-hoc Œdipe-I-MILO-DADS CSS →

Quelques producteurs de données

Le système statistique public

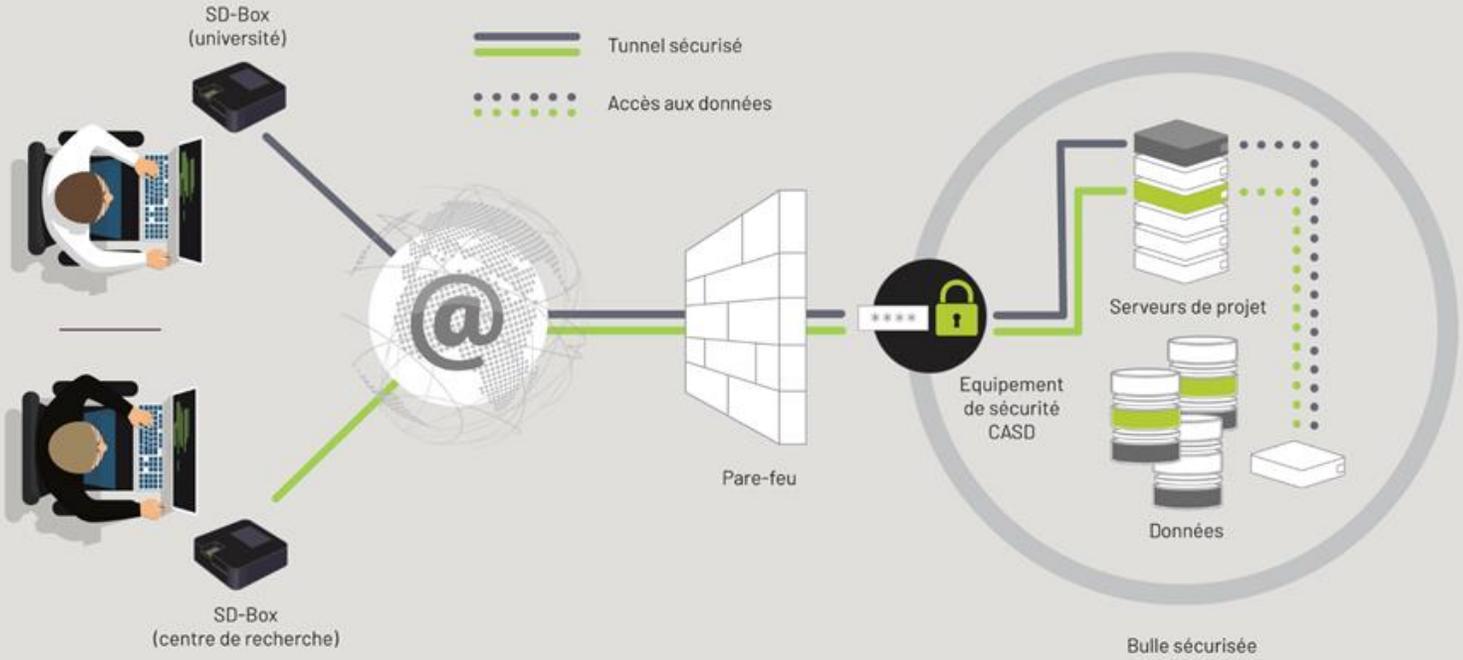
- INSEE
- Ministère des finances (DGFIP)
- Ministère du travail (DARES)
- Ministère de la santé (DREES)
- Ministère de l'agriculture (SSP)
- Ministère de la justice (SSER)
- Ministère de l'intérieur (SSMSI)
- Ministère de l'enseignement supérieur (SIES)
- Ministère de l'environnement (SDES)
- Ministère de la Fonction publique (DGAFP)
- Douanes (DGDDI)
- Institut national de la jeunesse et de l'éducation populaire (INJEP-MEDES)

D'autres administrations

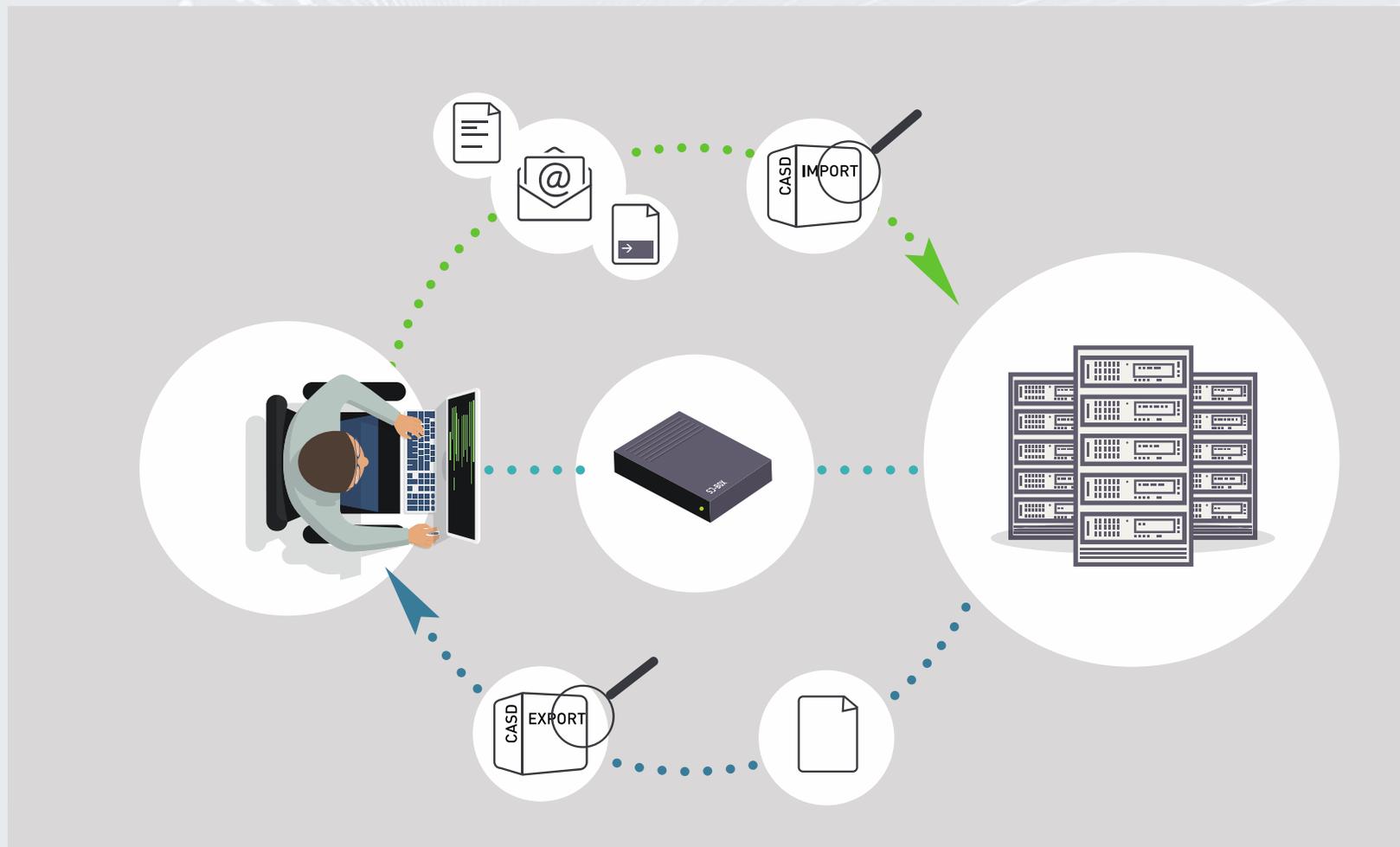
- Banque de France
- Caisse Nationales d'Allocations Familiales (CNAF, Données DSER)
- Mutualité sociale agricole (MSA)
- Caisse Nationale d'Assurance Maladie (données SNDS)
- Agence Centrale des Organismes de Sécurité Sociale (ACOSS)
- Agence Nationale pour l'Information sur le Logement (ANIL)
- Observatoire du Développement Rural (ODR)
- Banque Publique d'Investissement (BPI)
- Agence Technique de l'Information sur l'Hospitalisation (ATIH, données PMSI)
- Centre d'études et de recherches sur les qualifications (Céreq)
- Institut de recherche en économie de la santé (IRDES)
- Institut national de la santé et de la recherche médicale (INSERM – Cohorte Constances)
- Santé Publique France (SPF)

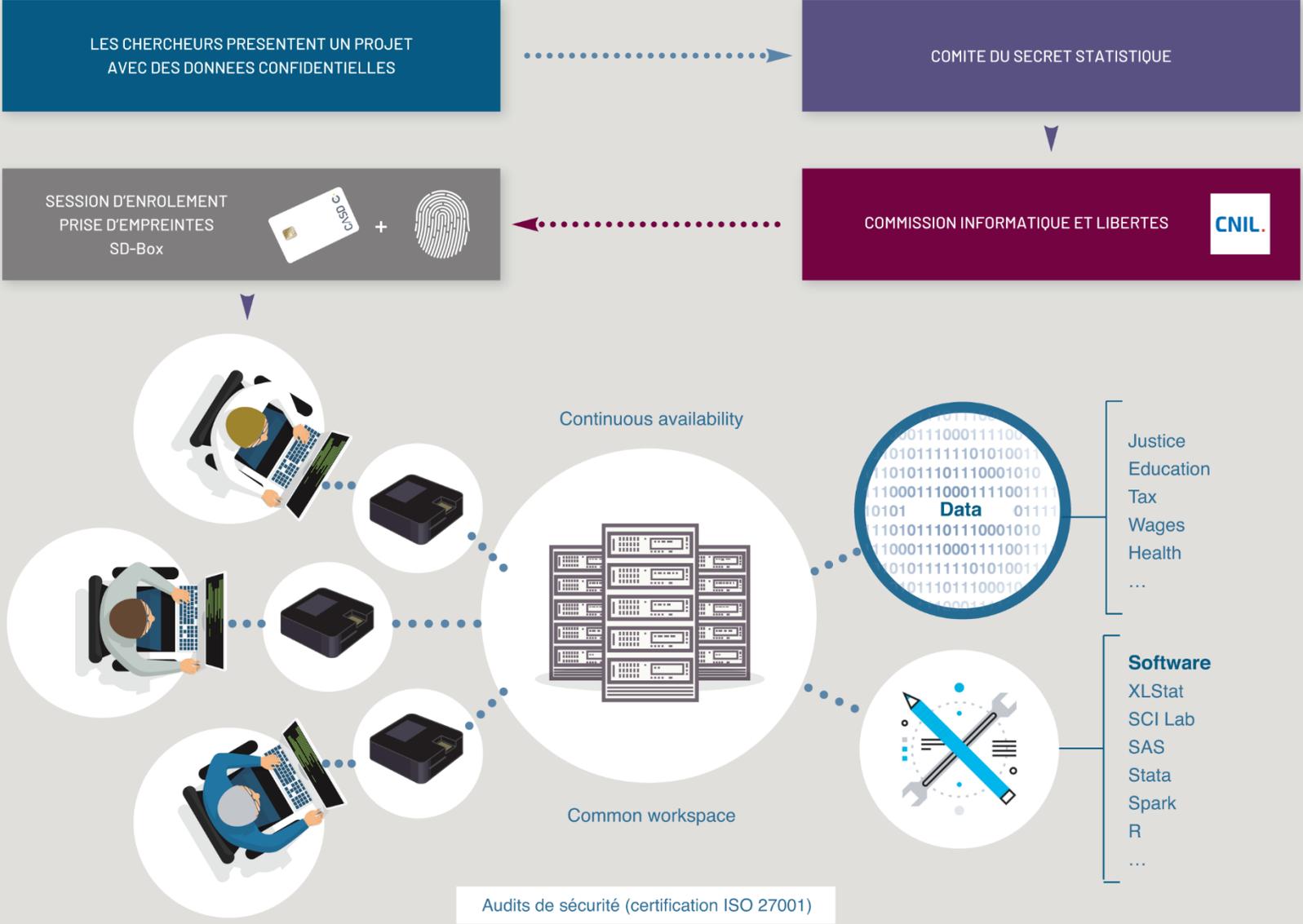


Dispositif technique



Workflows

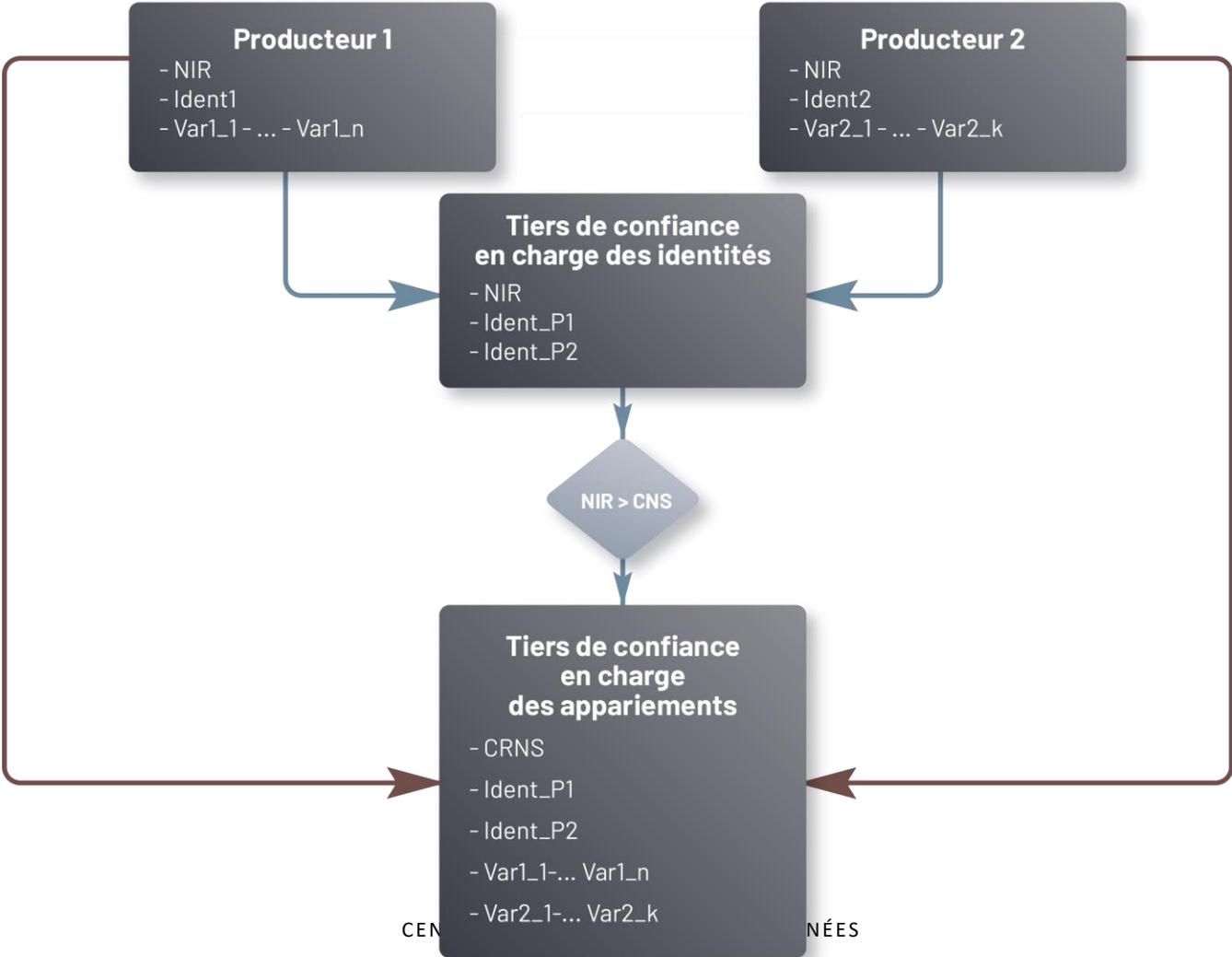


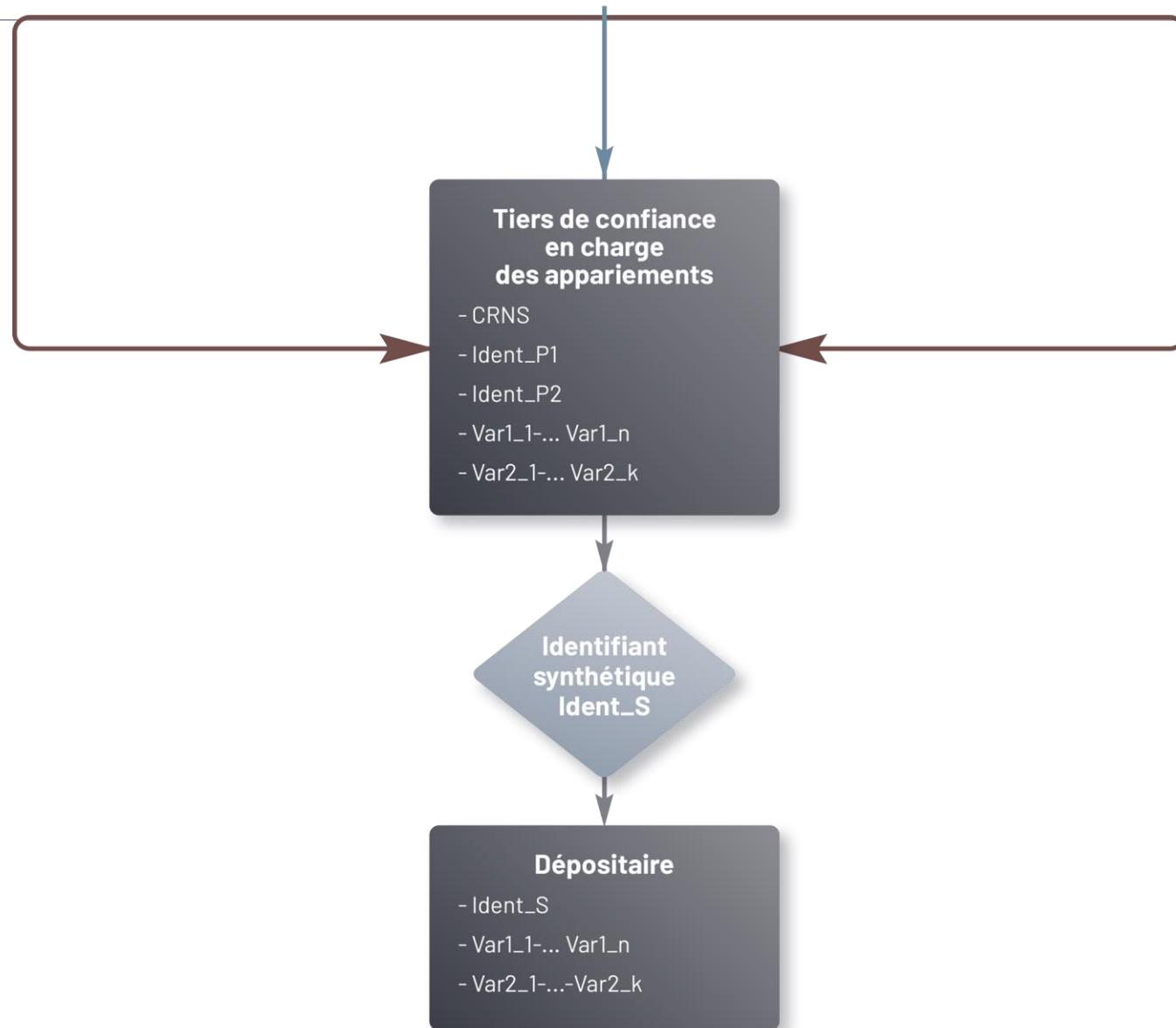


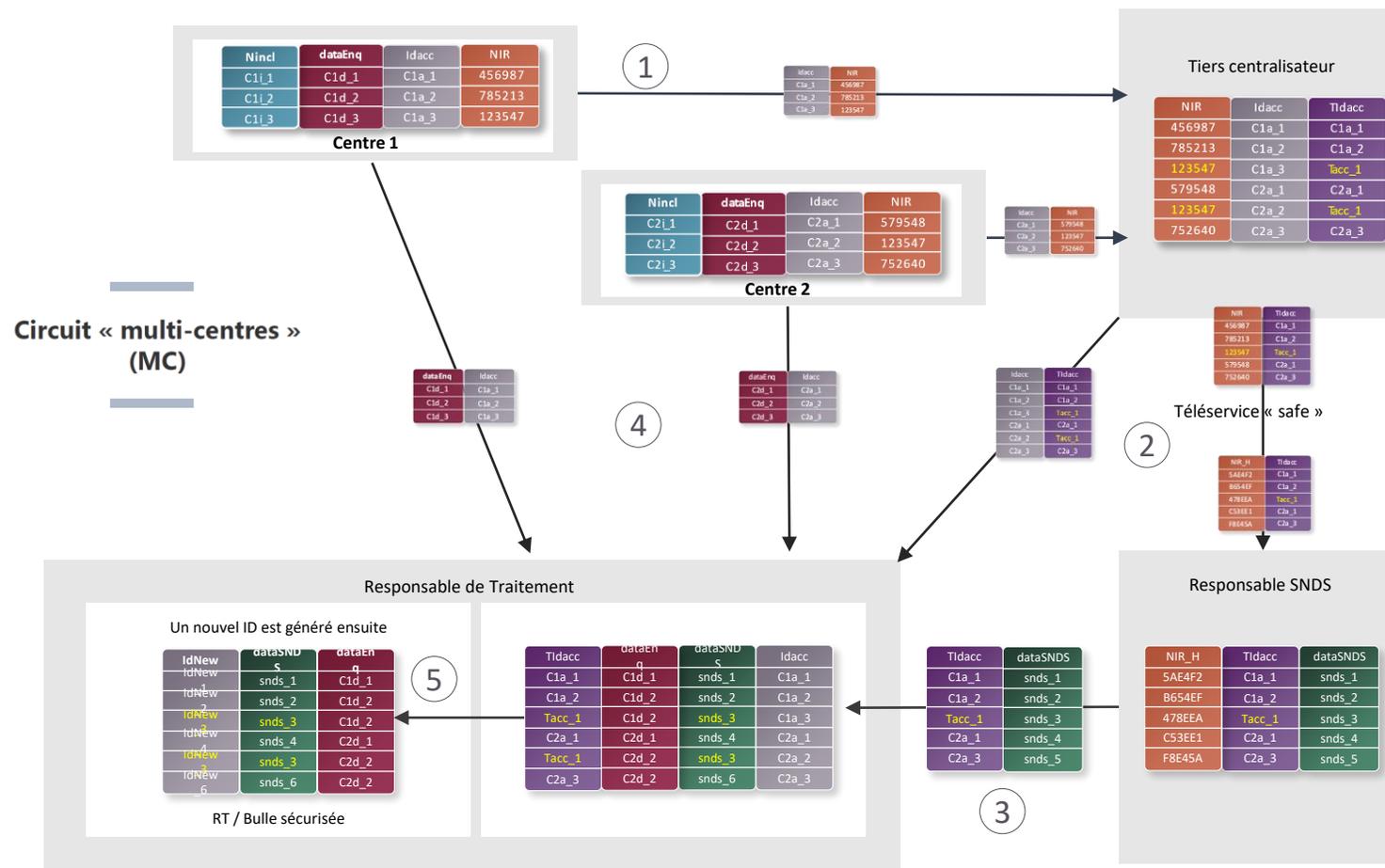
Quelques chiffres

- 700 projets en 2024
 - Les 2/3 des projets travaillent sur des sources de producteurs différents
→ décloisonnement des données
- 1800 utilisateurs actifs
- Plus de 600 sources de données
- Près de 6000 sorties de résultats vérifiées

Des opérations d'appariement complexes







Quelques projets

Déterminants des choix d'orientation dans un système centralisé : analyse du système APB et évaluations contrefactuelles



Ce projet analyse le système APB (Admission Post-Bac) utilisé pour allouer les étudiants au sein de l'enseignement supérieur. Nous étudions en particulier le rôle joué par l'information fournie aux étudiants concernant leur probabilité d'admission au sein des différentes options, ainsi que par les préférences des étudiants pour les différentes universités et types de formation. Nous étudions dans un deuxième temps la trajectoire des étudiants, ainsi que l'impact de l'information fournie aux étudiants au moment de l'admission sur leur réussite universitaire et orientation future. Nous évaluons enfin dans quelle mesure le nouveau système Parcoursup résulte en une allocation plus efficace entre les différentes formations.

Toulouse School of Economics

Membres : DE GROOTE Olivier, FABRE Anais, LUFLADE Margaux, MAUREL Arnaud

Performances des firmes multinationales et effets agrégés



Au cours de la décennie passée, la Commission Européenne a régulièrement pointé du doigt l'existence d'excédents des comptes courants de certains pays de l'Union, notamment de l'Allemagne (mais également des Pays-Bas, du Danemark et de la Suède). Cette critique revêt l'idée qu'un excédent du compte courant ou de certains de ses éléments pourrait avoir un impact négatif sur le développement économique du pays concerné. Ces excédents sont principalement dus à un déséquilibre de la balance commerciale et aux revenus nets des investissements directs étrangers (IDE). Alors que les conséquences micro- et macro-économiques d'un déséquilibre de la balance commerciale ont déjà été étudiées, les conséquences des IDE sur le développement

économique des pays émetteurs et bénéficiaires fait encore débat.

Kiel Institute for the World Economy

Membres : GORG Holger, HORNOK Cécilia, MARCHAL Léa

Management de l'innovation et innovations environnementales



Dans le cadre de nos recherches sur le management de l'innovation, nous souhaitons connaître les déterminants au niveau micro-économique des innovations environnementales. Pour cela, nous allons nous servir des données CIS (enquêtes communautaires) mais il nous faut également savoir quelle est la part de recherche et développement dans ces processus, pour répondre à la question suivante : quelles sont les caractéristiques de firmes qui s'engagent dans de tels projets ?

École Supérieure pour le Développement Économique et Social

Membres : CHASSAGNON Virgile, HANED-

CHASSAGNON Naciba

L'impact de l'emploi et du logement sur les trajectoires de vie des migrants: fécondité et statut marital



Ce projet étudie l'effet de l'emploi et du logement sur le comportement familial. Dans un premier temps, nous analyserons l'évolution des trajectoires de vie en terme de fécondité et de statut marital en comparant différents groupes d'immigrants. Ceci permettra d'avoir plus d'information sur les trajectoires de vie et les schémas familiaux dominants chez les descendants d'immigrants d'après-guerre et les immigrants plus récents (après 1990). Nous examinerons ensuite comment les changements en terme d'emploi et de logement influencent les transitions de statut marital et de fécondité. Enfin, nous simulerons les trajectoires (vie familiale et fertilité) pour les descendants d'immigrants de l'après-guerre et les immigrants récents afin d'examiner comment les trajec-

toires familiales changent en fonction de conditions économiques et sociales.

University of St Andrews

Membres : DELAPORTE Isaure, HARRISON Joseph, KULU Hill, PAILHE Ariane

Epidémiologie de l'hypobétalipoprotéinémie en population générale (HYPOBETA)



L'hypobétalipoprotéinémie (HBL) constitue un groupe de maladies du métabolisme des lipoprotéines, < 5ème centile de la population pour l'âge et le sexe, du taux de cholestérol LDL-cholestérol (LDL-C) ou d'apolipoprotéine B (APOB). L'étude HYPOBETA a pour objectif d'établir le profil phénotypique des sujets ayant une HBL en population générale en comparaison à des sujets ayant des taux normaux de LDL-C. L'étude cherchera notamment à comparer la prévalence : - des altérations de l'homéostasie du glucose (prédiabète, diabète) - des complications hépatiques et notamment la stéatose, la cirrhose et le carcinome hépato-cellulaire - des pathologies cardiovasculaires - des pathologies de la vision dont la cataracte - de troubles neuropsychologiques (dépression, risque suicidaire, troubles cognitifs) - de troubles de la fertilité chez la femme Outre une meilleure description de l'HBL, l'étude HYPOBETA permettra indirectement d'obtenir des informations pertinentes sur les potentielsrisques des traitements hypolipémiants, grâce aux informations sur les conséquences cliniques d'un taux de LDL-C à long-terme.

Centre Hospitalier Universitaire de Nantes

Membres : CARIUO Bertrand, GORONFLOT Thomas, WARGNY Matthieu

Merci

CASD

