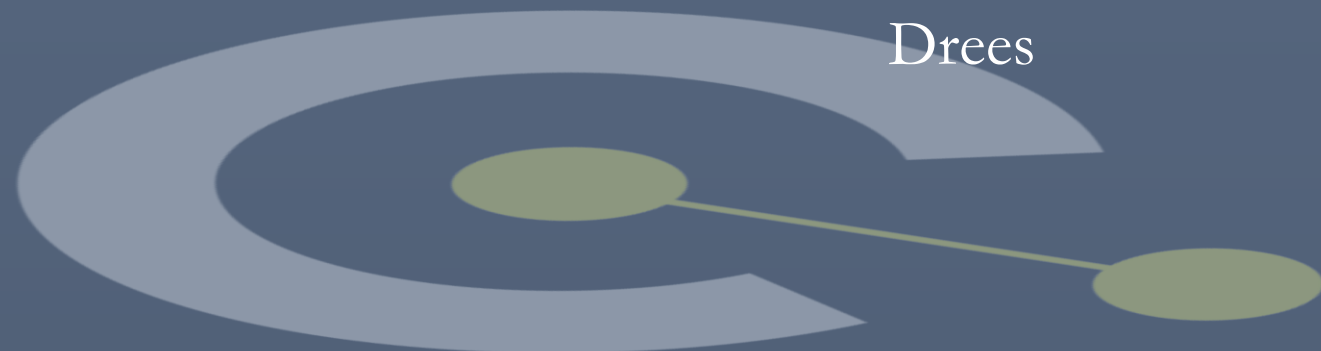


Présentation du centre d'accès sécurisé aux données CASD

Séminaire ouverture des données de santé
le 10 décembre 2014

Drees



présenter un équipement qui existe aujourd'hui qui permet de donner un accès sécurisé :

Aux données de l'Insee, données fiscales, données du ministère de la justice, de l'agriculture, de l'éducation nationale, de la BPI...
soit plus de **100 sources de données dé-identifiées**

À plus de **800 chercheurs** répartis dans plus de 300 projets de recherche

Depuis plus de **200 point d'accès** en France et en Europe (UK, GER, NL, SP, DNM)



- Un besoin d'accès aux données les plus détaillées possible exprimé par la communauté de la recherche.
- Des producteurs de données montrant leur volonté d'ouvrir l'accès à ces données (dé-identifiées).
- En général, cette confrontation de la demande et de l'offre se heurte à deux verrous :
 - Verrous réglementaires et/ou législatifs (autorisation, accréditation, résultats...)
 - Verrous techniques liés à la sécurité des données avec les risques juridiques et d'image associés en cas de fuite
- Le CASD cherche à lever les verrous techniques.

Quels risques liés à la diffusion de fichiers ?

Les données peuvent être **disséminées**:
volées, copiées, publiées sur internet



Les **infrastructures d'accès** à ces données peuvent être **vulnérables** à diverses menaces (virus, vers...)

La diffusion de ces données est soumise à diverses **contraintes juridiques encadrées** (Cnil...)



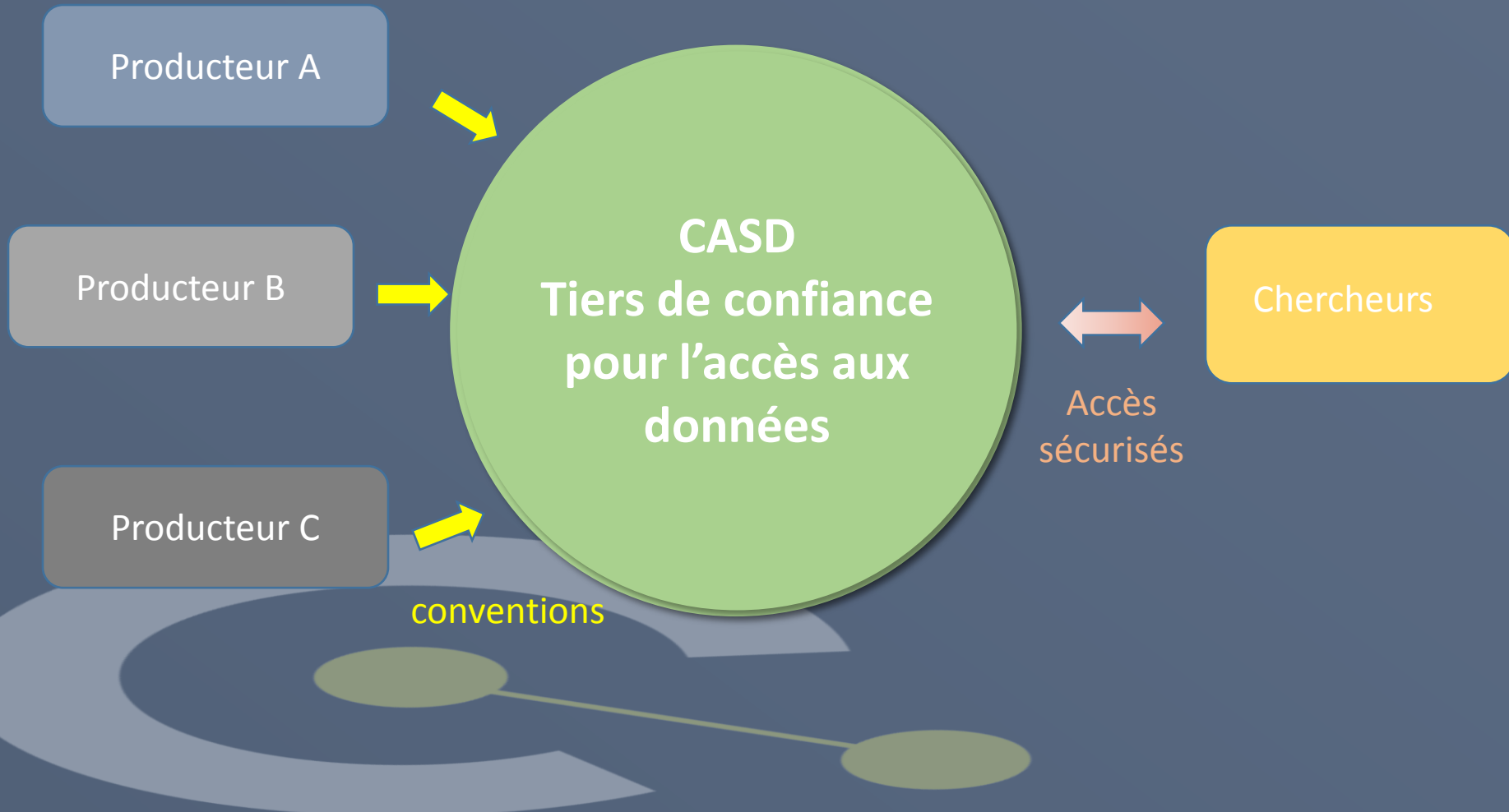
L'identité des utilisateurs accédant à ces données **n'est pas garantie**.

Les différentes actions affectant ces données ne sont **pas suffisamment tracées**.



Données à disposition, autorisations, accreditations, modalités de vérification des sorties

Utilisation des données





Un dispositif **hermétique** pour limiter la **dissémination** du matériau sensible

Permettant de **visualiser** le matériau sensible

Permettant de **manipuler confortablement** le matériau sensible

Permettant de **disposer de tous les outils nécessaires** aux différentes manipulations à mener sur le matériau

Une **procédure contrôlée** des insertions/extractions de matériau sensible dans/depuis ce dispositif

Un boîtier sécurisé d'accès



La SD-Box.

Hautement sécurisée et dédiée à l'accès sécurisé aux données sensibles

Authentification forte et incontournable

Verrouillée physiquement et logiquement, pour limiter la récupération de fichiers

Pas de données sensibles stockées sur la SD-Box

Des boîtiers standardisés

- Fiabilité (configuration unique, stable et validée)
- Déploiement simple et économique
- Besoin très limité d'assistance

Facile à déployer

- Nécessite un écran, un clavier & une souris
- Nécessite une connexion à internet
- Simple à configurer
- Pas d'incidence sur le reste du SI



La SD-Box

La Bulle : notre infrastructure étanche

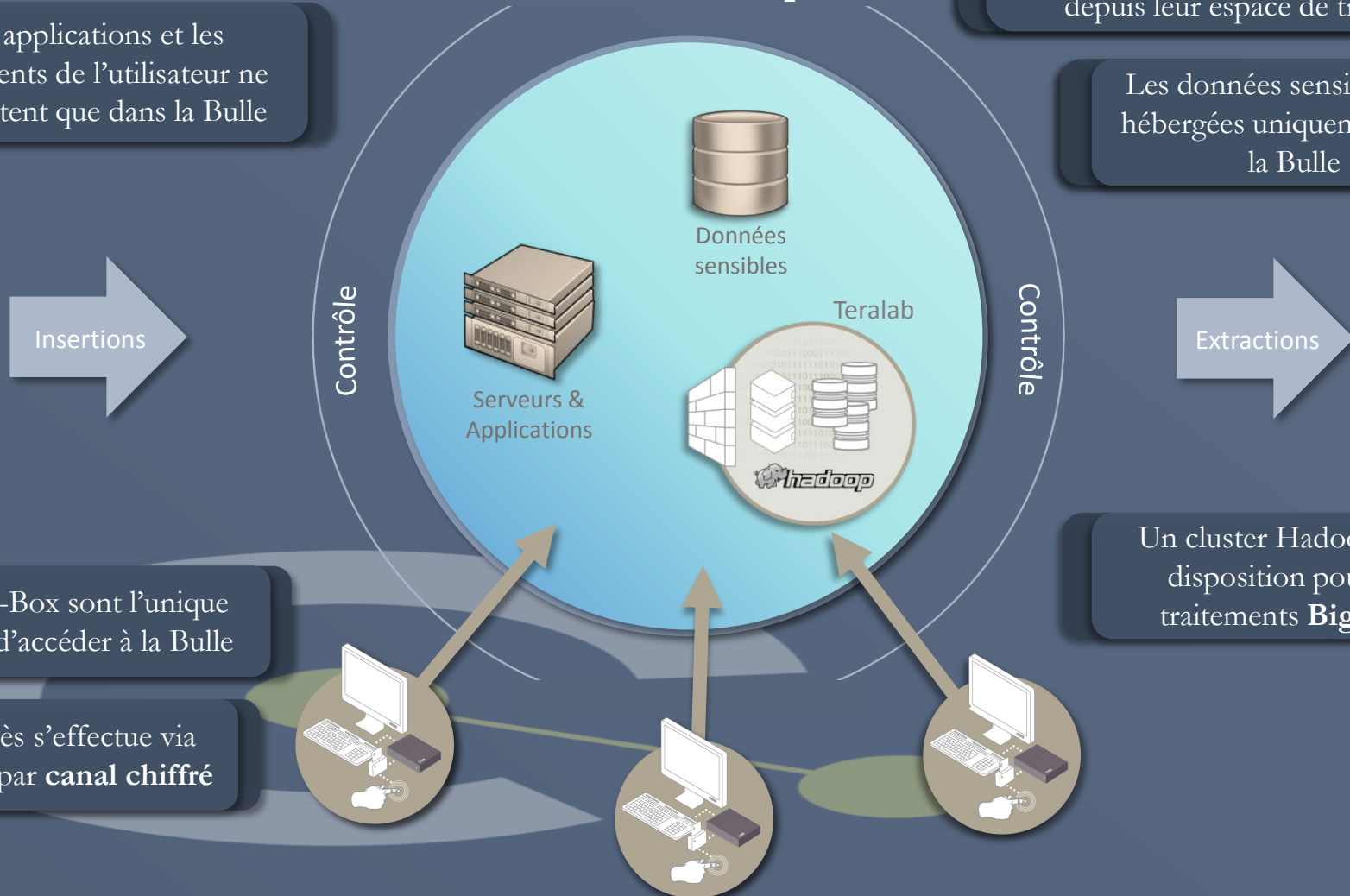
La Bulle est **un ensemble étanche de serveurs sécurisés**

Les applications et les traitements de l'utilisateur ne s'exécutent que dans la Bulle

Les insertions / extractions de données sont **contrôlées**. Les utilisateurs n'ont pas accès à internet depuis leur espace de travail.

Les données sensibles sont hébergées uniquement dans la Bulle

Bulle hermétique



Insertions

Contrôle

Serveurs & Applications

Données sensibles

Teralab

hadoop

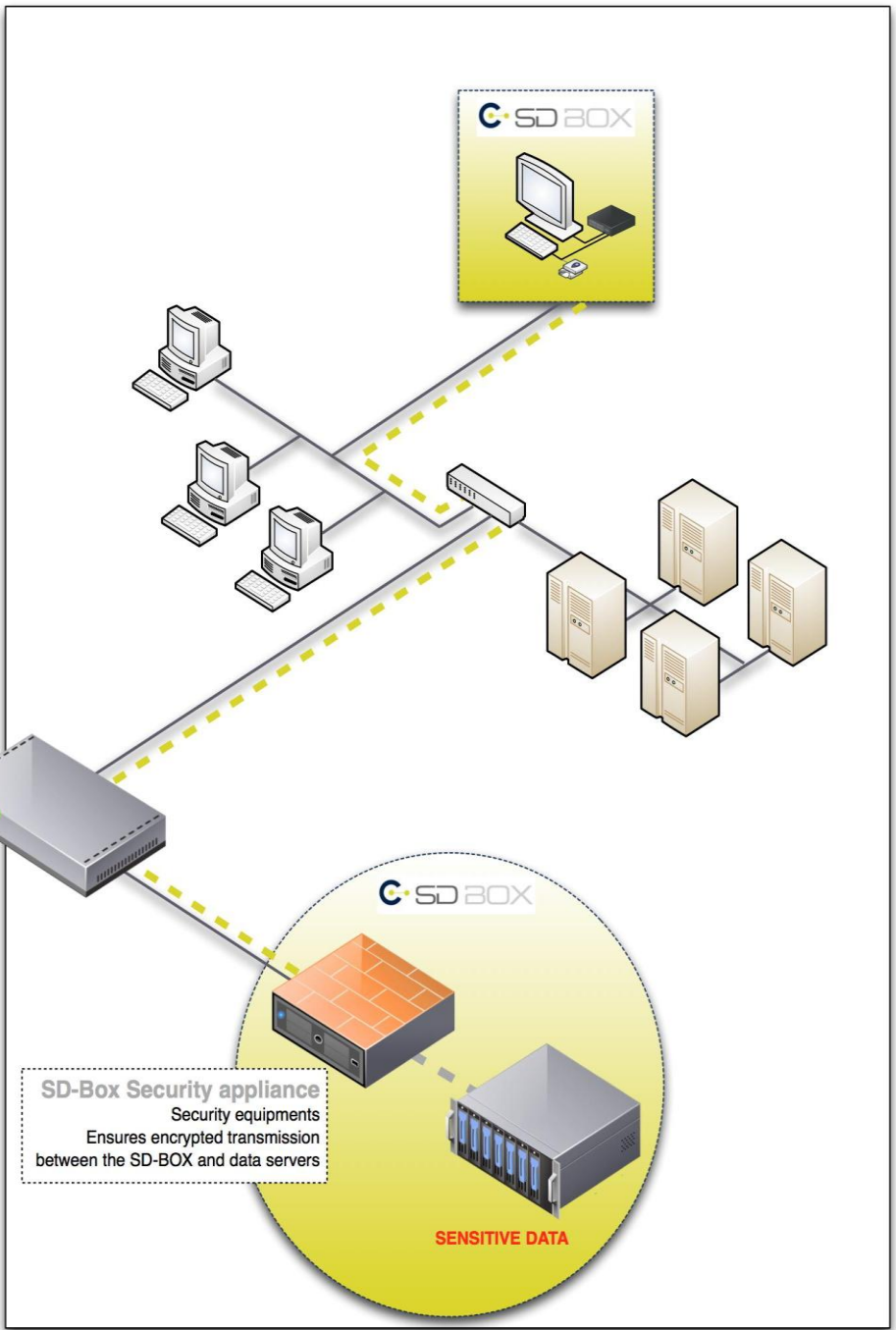
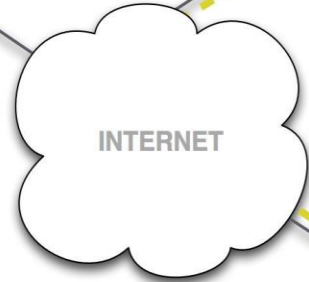
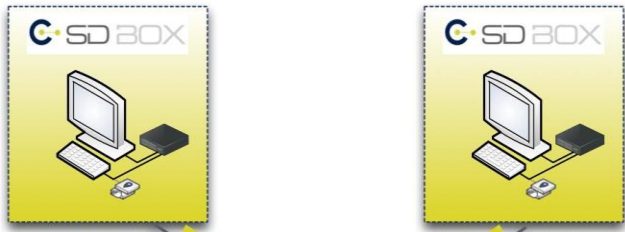
Contrôle

Extractions

Les SD-Box sont l'unique moyen d'accéder à la Bulle

Cet accès s'effectue via internet par canal chiffré

Un cluster Hadoop est à disposition pour les traitements **BigData**



— INTERNET LINK

- - - VPN-SSL Channel
All the traffic is encrypted (cypher)
No data transfer, screen display only

C-SD BOX

The SD-BOX System
 Biometric smartcard reader
 The hard drive is encrypted (TPM)
 No data in the SD-BOX
 Bios Lockdown
 USB Restricted

SD-Box Security appliance
 Security equipments
 Ensures encrypted transmission
 between the SD-BOX and data servers

SENSITIVE DATA

The screenshot displays a Windows desktop environment with two primary software applications running:

- Stata/MP 11.0 - [Results]:** Shows the Stata logo and 'MP - Parallel Edition'. The command window contains the text:


```
Stata compute
Serial number: 50110
Licensed to: GENES
GENES
```
- SAS:** Shows a journal window with the following text:


```
NOTE: Copyright (c) 2002-2008 by SAS
NOTE: SAS (r) Proprietary Software
Licensed to LICENCE CAMPUS G
NOTE: La session est exécutée sur

NOTE: L'initialisation de SAS a ut
temps réel      1.03 sec
temps UC        0.74 sec
```

The taskbar at the bottom includes icons for Démarrer, Biblio..., SAS, Micro..., WinEdt, Stat..., RGui, Gest..., Inter..., and desk... The system tray shows the date 11/05/2010 and time 16:27.

File Name	Action	Timestamp
KKKK_XXXX_	Ouvrir	2010 13:38
KKKK_XXXX_	Sortie CASD	2010 16:40
KKKK_XXXX_	Ouvrir dans une nouvelle fenêtre	2010 13:37
Excel_CASD	Extraire tout...	2010 16:40
desktopCAS	Ouvrir avec...	2010 16:28
Compress_C	Restaurer les versions précédentes	2010 13:38
	Envoyer vers...	

Le CASD est un équipement qui comprend :

- Une infrastructure informatique centrale « étanche » (IICE) / bulle
- Des boîtiers spécifiques d'accès, les SD-Box™, garantissant cette étanchéité en étant l'unique moyen d'accès à l'IICE.

Avec la SD-Box, l'utilisateur peut travailler à distance sur les données confidentielles tout en garantissant au producteur :

- qu'aucun fichier ne puisse être récupéré par l'utilisateur (pas de copier/coller, d'impressions, d'insertion de clé usb...)
- qu'il s'agit bien de la personne habilitée qui se connecte sur la SD-Box (authentification biométrique)...

Avec la SD-Box, l'utilisateur peut :

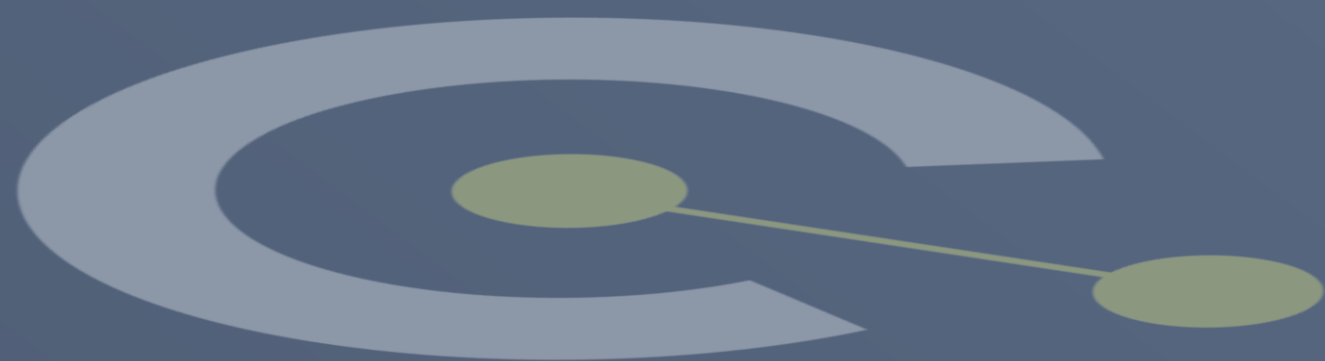
- réaliser des traitements interactifs à distance avec des logiciels statistiques 64 bits : SAS, Stata, R, Gauss, Matlab, latex, excel...
- demander des sorties de résultat finaux ou intermédiaires

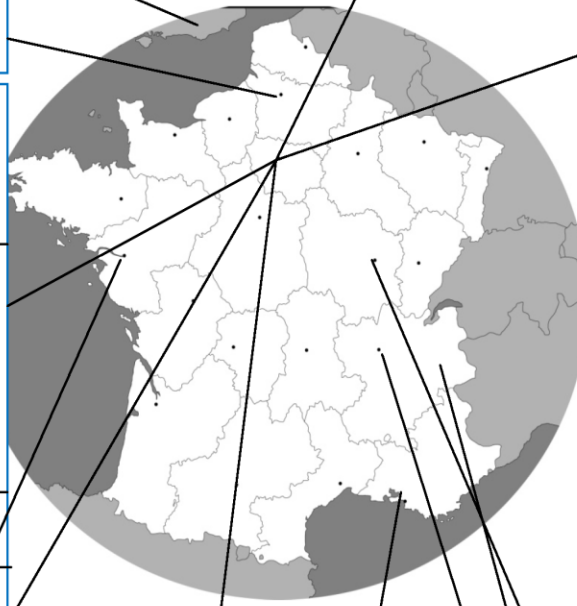
Un dispositif biométrique autorisé par la **CNIL** (délibération n°2014-369)



Une certification **ISO 27001** (sécurité des systèmes d'information) en cours d'implémentation

Des audits de sécurité indépendants et réguliers





Londres LSE-DISTRIB
Comparaison de la distribution des hauts revenus entre le Royaume-Uni et la France
Brian BELL (15 déc 2009)

Londres LSE-RENDNAL
Les rendements salariaux de l'obligation scolaire en France et au Royaume-Uni
Julien GRENET (18 mars 2010)

Amiens CRIISEA-SUCRE
Analyse des impacts d'une mutation institutionnelle majeure pour l'industrie du sucre
Jean-Louis GIRARD (15 déc 2009)

Paris 15 CEPII-KIMMAT
Impact du capital immatériel sur la croissance, la productivité et la compétitivité
Laurence NAYMANN (6 oct 2009)

Paris 15 CEPII-COMINTS
Lien entre globalisation, « skill upgrading » dans l'entreprise et inégalités des salaires
Maria BAS, M-C MITARITONNA, S. JEAN, A. TRITAH et M. CARRE (15 déc 2009)

Paris 2 ERMES-EPARSAL
Objectifs de l'épargne salariale
Fathi FAKHFAKH (15 décembre. 2010)

Paris 6 GEOCITE-RURAMAP
Gentrification et vieillissement dans les campagnes françaises
Pierre PISTRE (7 oct. 2010) [et Catherine Rhein]

Paris 7 SCIENCES-PO-EFFDISP
Effets des dispositifs ZUS, ZRU, ZFU (avec trois axes)
Th. MAYER, C. MALGOUYRES, F. MAYNERIS, L. PY et E. WASMER (18 mars 2010)

Paris 13 CREDOC-EMPSERP
L'emploi dans les services aux personnes âgées : mythes et réalités
Anne LOONES et Isa ALDEGHI (17 juin 2010)

Paris 16 DAUPHINE-KMULTI
Performances des multinationales françaises à l'international
Frédéric PELTRAULT, Alexandre GAZIANOL et Estelle DHONT (15 décembre. 2010)

Nantes MSH-EMPESCO
L'emploi dans l'économie sociale
Lionel PROUTEAU, Franck BAILLY, C. LE RENDU-LIZEE (17 juin 2010)

Malakoff CREST-SALSA
Dimensions subjectives des rémunérations
Olivier GODECHOT, Michel GOLLAC (15 déc 2009)
+ Croissance des inégalités dans le secteur de l'industrie financière
O. GODECHOT (15 déc 2009)

Malakoff CREST-BASEL
Absence des enseignants de français et maths sur la réussite au brevet
Magali BEFFY et Arnaud MAUREL (15 déc 2009)

Malakoff CREST-DYNAM
Dynamique des territoires, structure des emplois et commerce international
Francis KRAMARZ et Ferdinando MONTE (18 mars 2010)

Malakoff CREST-DYCARPUB
Attractivité fonction publique et caractéristiques de changement de secteur
Thierry KAMIONKA et Olivier FILTRIAU (17 juin 2010)

Paris 1 / Banque de France SAMIC-SURENDE et SAMIC-CYCLISAL
Surendettement (I) et réponses des salaires face au cycle éco. (II)
Henri FRAISSE (18 mars 2010) avec 5 chercheurs

Paris SDPE-POLCOMP
Politique de compétitivité entreprises
J-C LOMONACO et D. BASTIDE (15 déc 2009 / oct. 2010)

Marseille GREQAM-RESTOS
Impact de l'aide et du recours à l'aide offerte aux restaurateurs
Stephen BAZEN, L. BOUVARD, X. JOUTARD, P-H BONO et P. GRANIER (15 déc 2009)

Marseille GREQAM-FEMPLVI
Rôle des villes / discrimination des femmes sur le marché du travail
Pierre-Philippe COMBES, Cecilia GARCIA-PENALOSA et Shamalia NAWAZ (18 mars 2010)

Marseille GREQAM-VOISIN
Stratégies de discriminations par le voisinage
Benoît SCHMUTZ (7 oct. 2010)

USA Université du Michigan-ENTSECU
Entreprises et Etat – Providence : préférences des employeurs
Raj ARUNACHALAM et Sara WATSON (15 déc 2009)

Paris 20 INED-MOBRESI
Mobilité résidentielle au niveau individuel
Laurent GOBILLON et Matthieu SOLIGNAC (15 déc 2009)

Paris 20 INED-ECOAGGL
Impact des économies d'agglomération sur la productivité
-Id-, Sébastien ROUX et P-Ph COMBES (17 juin 2010)

Paris 20 INED-TIES
Caractéristiques de la population immigrée d'origine turque
Ana R. M. MATIAS, Patrick SIMON, Ch. HAMEL (17 juin 2010)

Paris 13 MSE-PSE-CONCSAL
Concurrence marché des biens & discrimination marché du travail
Clémence BERSON (6 oct 2009)

Paris 13 LAVUE-RVIMPAC
Impact rénovation urbaine Lille sur mobilités et diversifications
Jean-Pierre LEVY et Julien SALINGUE (7 oct 2010)

Paris 14 Jourdan-PSE-ENTFAM
Gestion de la main d'œuvre dans les entreprises familiales
Eve CAROLI, T. BREDA, A. BASSANINI, A. REBEYRIOUX (15 déc 2009)

Paris 14 Jourdan-PSE-RECPAR
Comportements de recherche d'emploi / salaire, horaire, contrat
Muriel ROGER et Sébastien ROUX (18 mars 2010)

Paris 14 Jourdan-PSE-ENCLAVE
L'hétérogénéité de l'impact de la politique de Zones Franches Urbaines : le rôle de l'enclavement géographique
Miren LAFOURCADE, Anthony BRIANT et Benoît SCHMUTZ (18 mars 2010)

+ Jourdan PSE-ETALURB
Étalement urbain et consommations énergétiques
-Id-, Thierry MAYER et C. BLAUDIN DE THE (17 juin 2010)

Nogent sur Marne CIREG-GES
Politiques de performance énergétique secteurs transport et bât.
Laure LAMPIN (6 oct 2009)

Dijon CESAER-CLIMAT
Sensibilité au climat des différents groupes sociaux et de la vulnérabilité des ménages et des territoires
Jean CAVAILHES, P. WAVRESKY, M. HILAL (15 déc 2009) + Cyrus GROUT (déc. 2010)

Grenoble LEPII-PARCOURS
Sécurisation des parcours et dialogue social/précarité dans emploi
Jon B. ZUBIRI-REY (15 déc 2009)

Lyon 2 Ecully GATE-SEGURB
Analyse des choix résidentiels des ménages
Florence GOFETTE-NAGOT et Modibo SIDIBE (15 déc 2009)

Lyon 2 Ecully GATE-EMPPPOP
Accès à l'emploi dans les territoires de la politique de la ville
-Id- -N. M. S. et Nathalie HAVET (18 mars 2010)

International Projects

2 London School of Economics projects:

- Comparing High Income Distribution between the United Kingdom and France
- Wage-based Education Obligation Yields in France and the United Kingdom

1 Royal College of London Project

- Working Hours as an Adjustment Mechanism: The Cases of Flexible and Rigid Labor Markets

London-Paris-Milan Project

Designation: Why are industrial groups different? Domestic labor and capital markets in French industrial groups

1 Erasmus University Rotterdam project

- Work, Family, and Happiness

1 Florence Project – European University Institute

Designation: Local Multiplier Effect and the Impact of Local Shocks on Labor Demand in France

London-Paris-Milan Project

Designation: Why are industrial groups different? Domestic labor and capital markets in French industrial groups

1 Malakoff Project (American researchers in collaboration with the CREST [Economics and Statistics Research Center])

Designation: Businesses and the Welfare State



Number of International Researchers : 11
+ 3 on the DwB Project

London-Paris-Milan Project

Designation: Why are industrial groups different? Domestic labor and capital markets in French industrial groups

Titre de la publication	Source de données utilisées	Comité de lecture ?	Parue / En cours de soumission	Année principale de la réalisation	CASD cité
Public Housing and Residential Segregation of Immigrants in France, 1968-1999	RP 1999	Oui	Population 66 (1), 2011, 169-194	2008 - 2009	Oui
Renouveaux des campagnes françaises : évolutions démographiques, dynamiques spatiales et recompositions sociales	RP 2007, ENL 1992 à 2006	Oui	Disponible sur : http://halshs.archives-ouvertes.fr/tel-00764869/	2008 - 2012	Oui
Modibo Sidibé Structural Empirical Models of Spatial Inequalities : Housing Choices, Policies and Generational Consequences	Enquêtes Logement 1984 à 2006 / Enquêtes Emploi en continu 2003 à 2008		Thèse (dec 2011) / Lyon 2	2010 - 2011	Oui
Évaluation des effets des politiques ciblées sur les territoires : le cas des Zones Franches Urbaines, C. Malgouyres, T. Mayer, F. Mayneris et L. Py	DADS-FICUS	Oui	Parue, rapport pour la DARES	2010 - 2012	Non
The impact of urban enterprise zones on establishment location decisions: Evidence from French ZFUs, CEPR DP 9074 et IRES DP 2012-19, T. Mayer, F. Mayneris et L. Py	FICUS	Non	Parue	2010 - 2012	Oui
Les effets de relocalisation des Zones Franches, T. Mayer, F. Mayneris et L. Py	FICUS	Non	LIEPP Policy Brief n°4	2010 - 2012	Non
The relocation effects of enterprise zones, T. Mayer, F. Mayneris et L. Py	FICUS	Non	Vox column	2010 - 2012	Non
Sex and the City : Gender Gaps in Labor Markets and Economic Geography	Panel DADS	Oui		2010 - 2012	
Andrea Bassanini, Thomas Breda, Eve Caroli, and Antoine Rebérioux, Working in Family Firms: Paid Less But More Secure? Evidence from French Matched Employer-Employee Data	DADS, DMMO, REponse, DIANE	Oui	A paraître en avril 2013 Industrial and Labor Relations Review, Vol 66 No 2	2011	Oui
Florence Goffette-Nagot, Modibo Sidibé, Logement social et accession à la propriété, Economie et Prévision	Enquêtes Logement 1984 à 2006	Oui	En cours de publication	2011	Oui
Modibo Sidibé, The Contribution of Housing to the Dynamics of Inequalities	Enquêtes Logement 1992, 1996 et 2002	Non	Paru, Working paper GATE, 2012-15	2011	Non
La consommation d'énergie des ménages en France	Enquêtes Logement	Non	Paru, rapport final Convention MEEDDM	2011	Oui
Les émissions directes de CO ₂ des ménages selon leur	Enquêtes	Oui	Paru	2011	Oui



Titre de la publication	Source de données utilisées	Comité de lecture ?	Parue / En cours de soumission	Année	CASD cité ?
Les émissions directes de CO ₂ des ménages selon leur localisation, Le Point Sur n° 137	Enquêtes Logement	Oui	Parue	2011	Oui
Goffette-Nagot Florence, Charlot Sylvie, Dujardin Claire, Havet Nathalie, Sidibé Modibo, 2012, Accès à l'emploi dans les territoires de la politique de la ville : un appariement entre emplois et populations Rapport pour la DARES "évaluation des effets des politiques ciblées sur des territoires"	Enquêtes Emploi en Continu 2003 à 2008, Panel DADS	Oui	Rapport final rendu à la Dares en juin 2012	2011 - 2012	Oui
The Great Compression of the French Wage Structure	DADS-EDP	Oui	En cours de soumission	2011 - 2012	Oui
Sorting and local wage and skill distributions in France	Panel DADS	Oui	Parue en 2012, vol 42 p. 913-930, Regional Science and Urban Economics	2011 - 2012	Oui
The neighbor is king: customer discrimination in the housing market, avec Bruno Decreuse, Benoît Schmutz et Alain Trannoy	Enquête Logement, recensement	Oui	En cours de soumission	2011 - 2012	Oui
Quarante ans de ségrégation et d'incorporation des immigrés, 1968-2007	Série des 6 RP de 1968 à 2007	Non	Paru le 7 janvier 2013	2012	Oui
Time is on my side: 40 years of immigrant segregation in France, 1968-2007	Série des 6 RP de 1968 à 2007	Oui	En révision après 1e soumission	2012	Oui
État de la ségrégation et discriminations au logement en France	Série des 6 RP de 1968 à 2007	Non	A paraître	2012	Non
Can tax breaks beat geography? Lessons from the French enterprise zone experience	Enquête Nationale Transports et Déplacements	Oui	En cours de soumission	2012	Oui
The Neighbor is King. Customer discrimination in the housing market	Enquête Nationale Logement	Oui	En cours de soumission	2012	Oui
Is sprawling residential behavior influenced by climate?	Enquêtes Logement	Oui	En cours de soumission au J. of Urban Economics	2012	Oui
le goût des ménages français pour un climat chaud et l'étalement urbain	Panel DADS	Oui	En cours de soumission à Environment and Resource Economics	2012	Oui
La qualité de l'emploi dans l'ESS. Etude exploratoire sur la région des Pays de la Loire	DADS	Oui	Parue	2012	Oui

Composition du Conseil Scientifique (CS)

Présidente Roxane SILBERMAN, Secrétaire générale du Comité de concertation pour les données en sciences humaines et sociales (CCDSHS), Directrice du GIS Réseau Quetelet, Directrice de recherche au CNRS

Alain Trognon, Conseiller scientifique du CASD, Directeur du GRECSTA (groupe de recherche en économie et statistique), Membre du CREST-ENSAE (centre de recherche en économie et statistique), Professeur associé, Université Paris 1

Membres

▪ **Cunéo Philippe**, Directeur de la méthodologie et de la coordination statistique et internationale, Insee

▪ **Benoit Dostie**, Professeur agrégé, directeur de l'Institut d'économie appliquée, HEC Montréal

▪ **Brigitte Dormont**, Professeur à l'université Paris Dauphine, Directrice du laboratoire d'économie et de gestion des organisations de santé, LEDa-LEGOS, Université Paris-Dauphine

▪ **Olivier Godechot** Sociologue, Chargé de recherche CNRS au Centre Maurice Halbwachs et au laboratoire de sociologie quantitative

▪ **Fabien Jobard** Directeur du Centre de recherche sociologiques sur le droit et les institutions pénales (CESDIP), chargé de recherche CNRS

▪ **Thierry Magnac** Chercheur Toulouse School of Economics, Membre de l'Institut d'Economie Industrielle (IDEI), Professeur d'économie, Université Toulouse 1 Sciences sociales

▪ **Catherine Quantin** Professeur Université de Bourgogne, Chef de service, PU-PH dans le service biostatistique et information médicale, Université de Bourgogne

▪ **Muriel Roger**, chargée de recherche, INRA

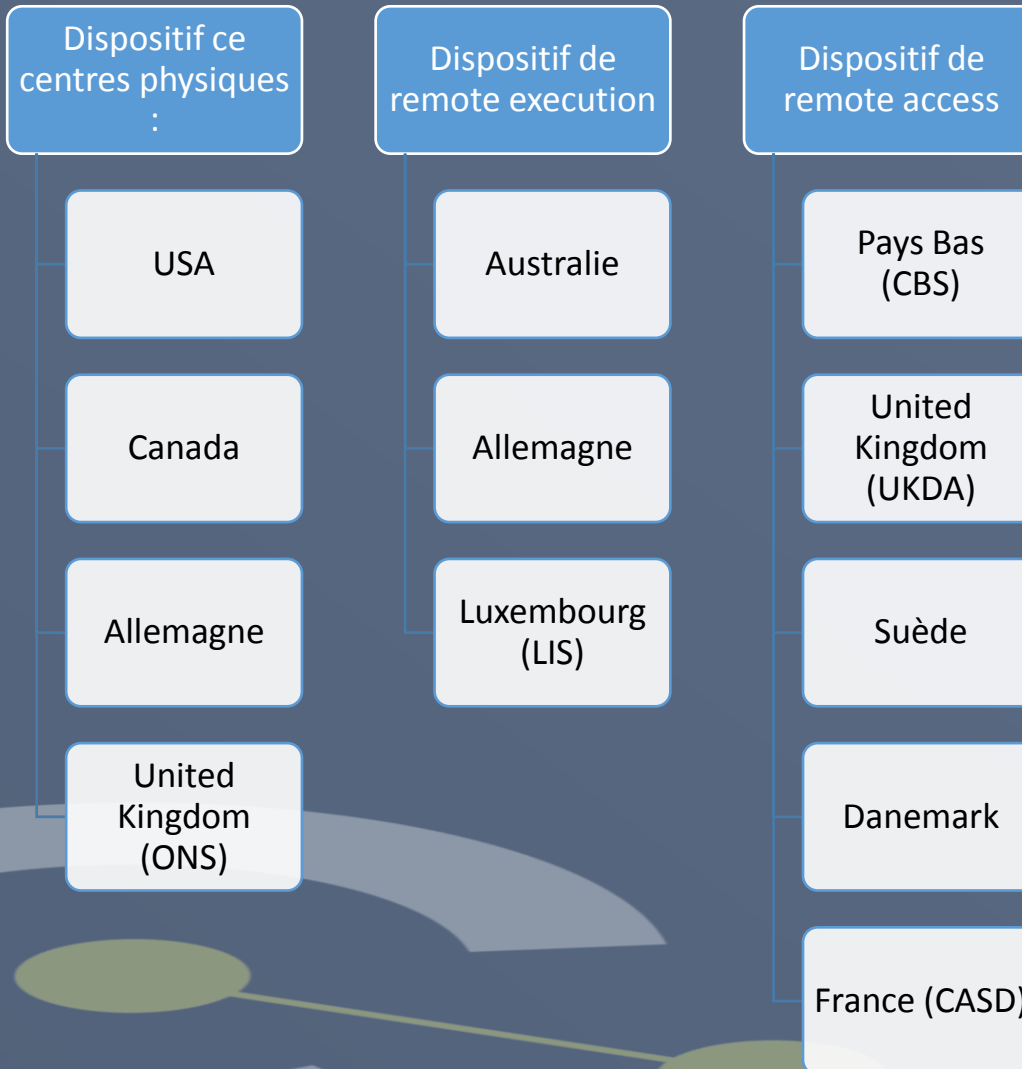
▪ **Sébastien Roux**, membre du CREST-ENSAE

▪ **Georges Alter, IPCSR, Director**

▪ **Dominique Joye, université de Lausanne**

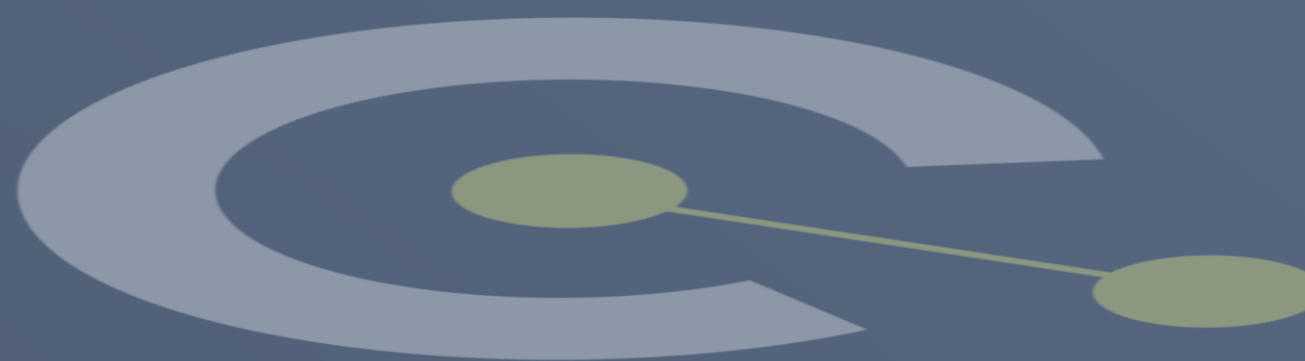
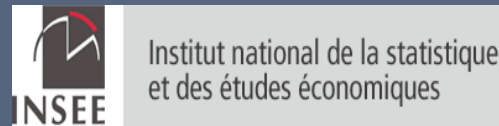
▪ **Lars Vilhuber, Senior Research Associate, Cornell University**

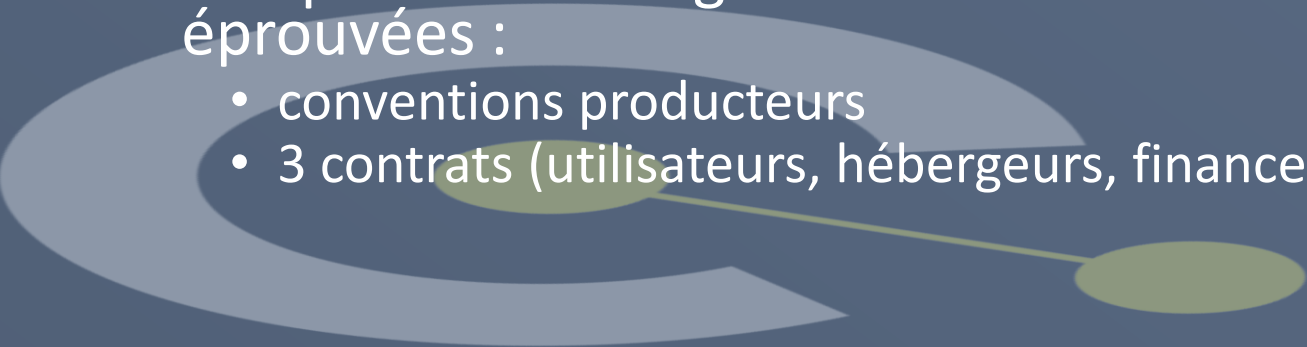
▪ **Tanvi Desai, Data Manager, London School of Economics**



Une entité du **Genes** (Groupement des Ecoles Nationales d'Economie et de Statistiques) :

- Des ingénieurs et datascientists pour l'infrastructure
- Des ingénieurs R&D
- Des statisticiens avec le savoir faire de l'Insee



- Un équipement aux spécifications à la croisée entre les besoins des chercheurs et les besoins de sécurité des producteurs de données
 - Un investissement important en R&D (4 Millions d'euro) – brevet
 - Un acteur du secteur public (Genes)
 - Une expérience de 6 ans de fonctionnement (et d'améliorations associées)
 - Une croissance forte et une conception dès le départ adaptée à la montée en charge.
 - Des procédures organisationnelles et contractuelles éprouvées :
 - conventions producteurs
 - 3 contrats (utilisateurs, hébergeurs, financeurs)
- 

Merci de votre attention

