



Institut national de la statistique
et des études économiques



Le centre d'accès sécurisé aux données (CASD)

ORSAY– 30 janvier 2013



GROUPE DES ÉCOLES NATIONALES D'ÉCONOMIE ET DE STATISTIQUE

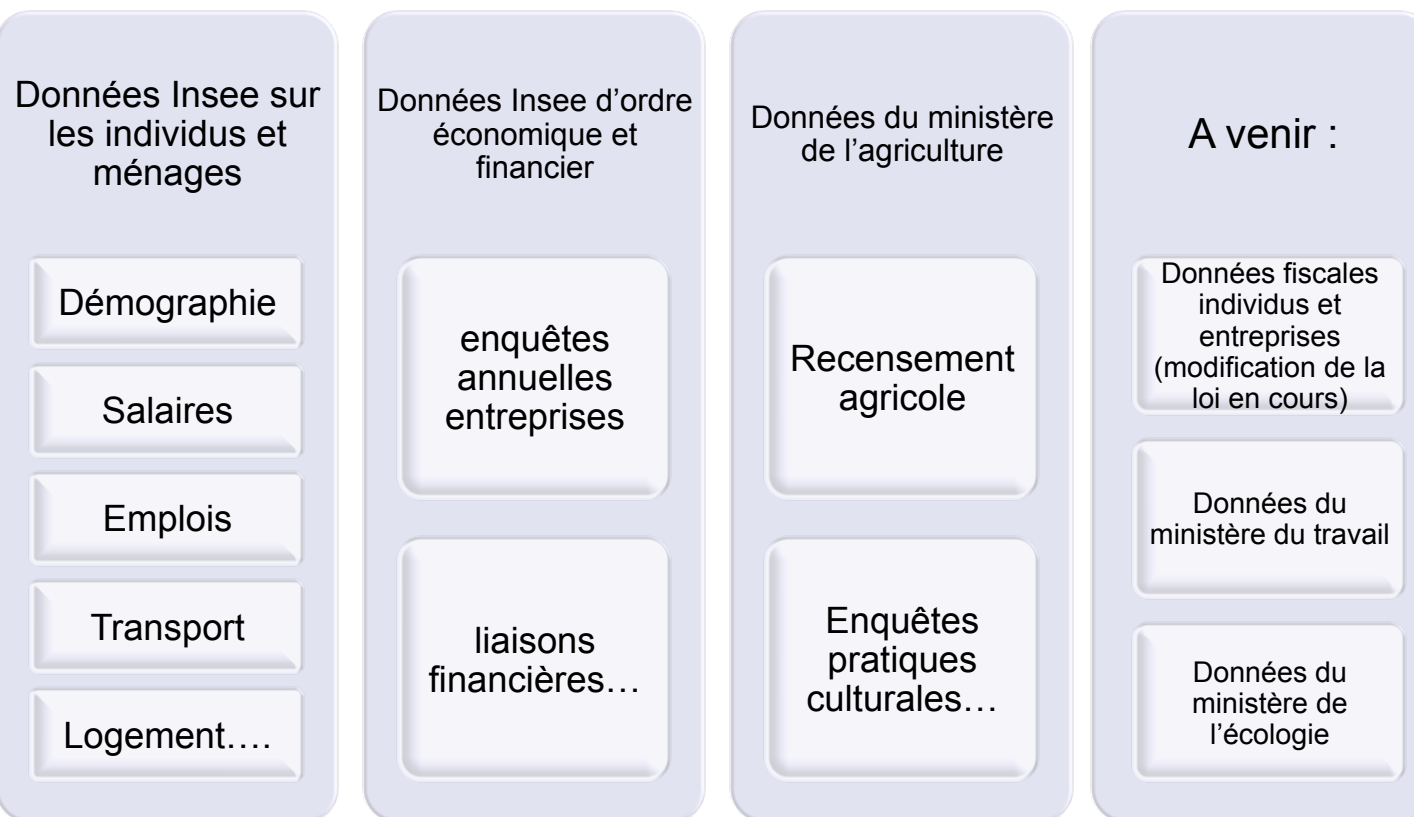


Pourquoi le CASD ?

- **Un nombre important de demandes d'accès** aux données individuelles très détaillées faites par les chercheurs depuis de nombreuses années.
- Ce type d'accès **existe depuis longtemps** dans la plupart des autres pays.
- Mais surtout, le CASD doit sa création à la volonté de l'Insee d'ouvrir le plus largement possible l'accès à ses données aux chercheurs pour **mieux les valoriser**.
- Verrou juridique : La modification en 2008 de la loi sur le secret statistique (loi n°51-711 du 7 juin 1951) a **rendu possible juridiquement** l'accès des chercheurs à ces données.
- Verrou technique : Des contraintes de sécurité importantes nécessaires pour ouvrir cet accès : le projet CASD

Quelles données sur le CASD

- Ce sont des données individuelles très détaillées indirectement nominatives : qui concernent de nombreux domaines



Un patrimoine très riche de données de qualité, structurées et documentées

Evolution des sources disponibles au sein du CASD

THEME	INTITULE
Consommation et modes de vie	Enquête Budget de Famille
	Enquête Emploi du Temps
	Enquête Logement
	Enquête Transport
Population et démographie	Recensement de la population
	Mortalité des adultes
	Enquête Etat Civil (naissance)
	Échantillon démographique permanent
Emploi et salaires	Enquête emploi (annuelle et en continu)
	Déclarations annuelles des données sociales
	Enquête sur le coût de la main et la structure des salaires
	Enquête structure des emplois
	Panel européen des ménages (enquête longitude)
Indicateurs sociaux	Statistiques sur les ressources et les conditions de vie
Autres	Changement organisationnel et informatisation



GENES

Evolution des sources disponibles au sein du CASD

THEME	INTITULE
Environnement et énergie	Enquête sur les études et les investissements pour protéger l'environnement
	Enquête sur les consommations d'énergie dans l'industrie
Infrastructure	Changements organisationnels et les technologies de l'information et de la communication
Salaires, activités des entreprises, liaisons financières	Déclarations de données sociales au niveau établissements et déclarations de données au niveau entreprises
	Enquêtes annuelles entreprises
	Enquête annuelle de production
	Système unifié de statistiques d'entreprises
	Enquête sur les liaisons financières les entreprises françaises du secteur privé

Evolution des sources disponibles au sein du CASD

➤ **Données acquises et en cours d'acquisition depuis le dernier comité de pilotage (30/05/2012)**

Enquêtes/Sources	Descriptifs	Années	Acquisition en cours
Budget famille DADS	Fichier détail complet	1979 1985	1968
	panel	2009	
Enquête emploi	Enquête emploi en continu concernant les individus répondants aux questionnaires logement et individuel et les libellés des professions et activités principales déclarées dans l'enquête .	2009 2010	
Patrimoine			2010
RP	Recensement de la population des exploitations principales pour toutes les années sauf en 2008 avec complémentaire ainsi que de la documentation.	1975 1982 1990 1999 2008	
Emploi du temps	Fichier de diffusion	2006 2010	1986 1999

Avancement du projet

Evolution du nombre de projet au sein du CASD

➤ Données acquises et en cours d'acquisition depuis le 25 juin 2012

Enquêtes/Sources	Spécificité	Années	Acquisition en cours
Budget Famille		1973-1974	
DADS	Postes	1993/2010	
Enquête Emploi	Annuelle	1982-1990; 1991-1998 (en cours de conversion)	
	Mod. conditions de travail	1995/2001	
Condition de vie et sécurité		2007 à 2011	
Patrimoine			2010
Recensement		1968 (principal)	
Enquête Statistique sur les Ressources de vie et Condition de Vie		2009	
Emploi du temps			1986
Histoire Familiale		1990	

Avancement du projet

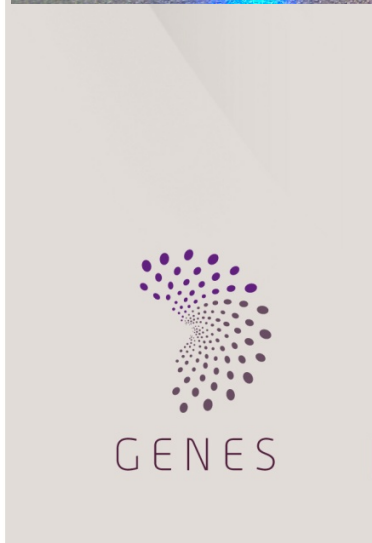
Données SSP (agriculture) acquises au sein du CASD

THEME	INTITULE SOURCE
Exploitation des cultures et du sol	Enquête sur les pratiques culturelles Grandes cultures
	Enquête sur les pratiques culturelles Bassin Versant
	Enquête sur les pratiques culturelles prairies
	Enquête sur les pratiques culturelles viticulture
	Coordonnées des points référencés dans les enquêtes pratiques culturelles
	Enquête sur l'occupation des sols
Production animal	Enquêtes bâtiments d'élevage: Bovins, Porcins, Volailles
Recensement	Enquête de recensement agricole
Energie	Enquête sur la consommation et la production d'énergie des entrepreneurs des territoires (EDT) et des coopératives d'utilisation de matériel agricole (Cuma)
	Enquête annuelle sur les consommations d'énergie
Structure des exploitations	Enquête sur les structures de la production légumière
	Enquête sur la structure des vergers
Industries agro-alimentaires	Enquêtes annuelles d'entreprises industrie Agro-alimentaires
	Enquête annuelle entreprise Entreprises et établissements
	Enquête sectorielle annuelle Industries agro-alimentaires

Et dans quel contexte technique ?

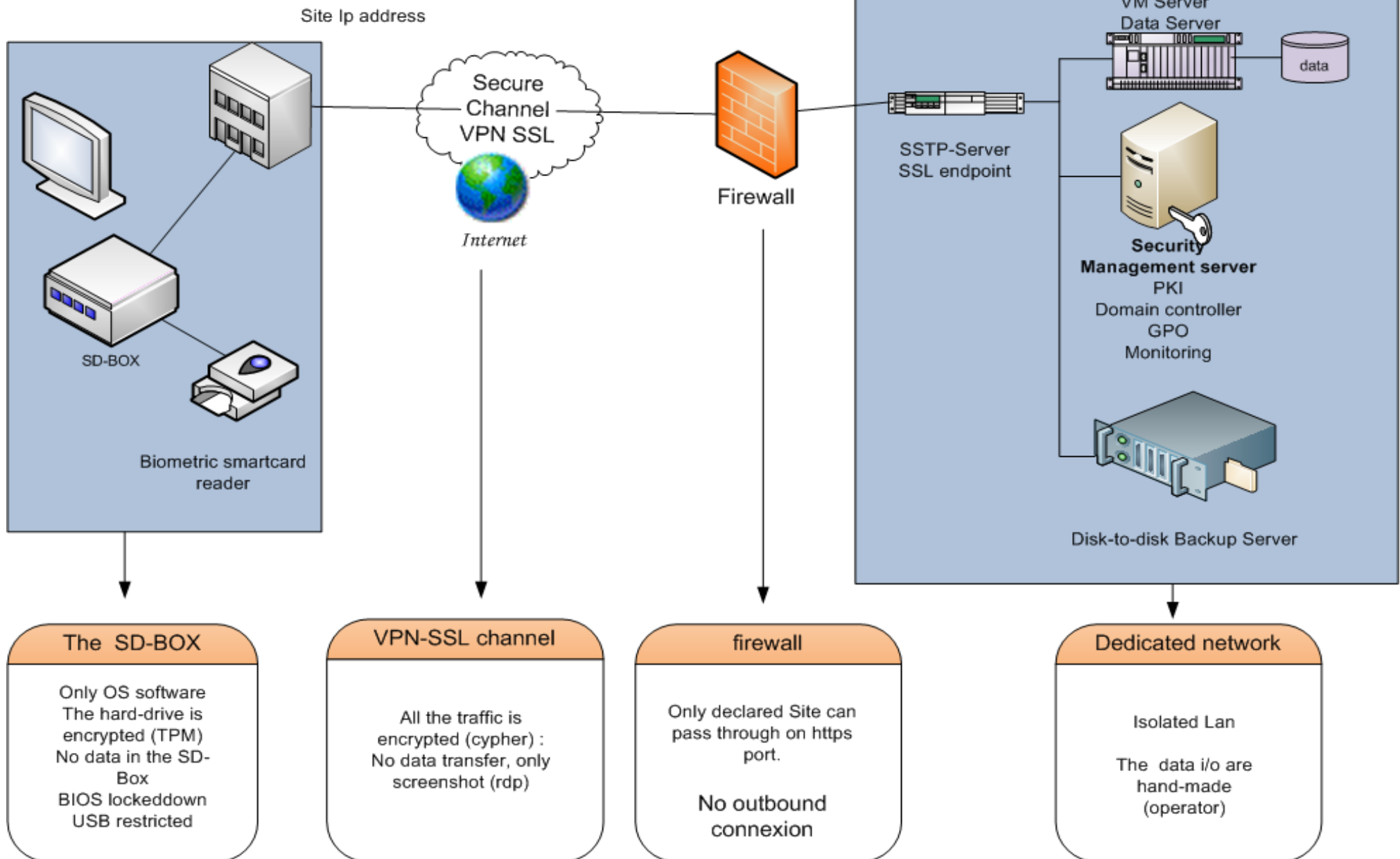
- Les premiers centres d'accès sécurisé physiques sont apparus il y a plus de 20 ans en Amérique du Nord (Cornell, StatCanada..)
- Depuis les années 2000, certains pays ont mis en place des solutions d'accès sécurisé à distance pour les chercheurs (Danemark, suède, Pays-Bas...)
- Ces solutions reposent sur l'installation sur l'ordinateur du chercheur de logiciels particuliers, mais finalement :
 - elles ne sont pas assez sécurisées (le poste socle est maître)
 - elles sont coûteuses et complexes à mettre en œuvre
 - en installation, compatibilité, évolution, assistance...
- De ce constat, il est apparu qu'il fallait raisonner différemment : au lieu de raisonner en termes de logiciel, nous avons raisonné en termes de matériel : 9

la SD-Box



CASD C SD BOX

CASD Architecture



La SD-Box™



- C'est un boîtier informatique entièrement conçu par le Genes pour :
 - Ne servir que pour permettre l'accès distant des chercheurs
 - être très sécurisé (tout est crypté, l'authentification est biométrique)
 - être verrouillé aussi bien matériellement que logiciellement
 - empêcher toute sortie de données (usb, imprimante...)
 - être très simple à installer (le chercheur vient la chercher)
 - être très simple à remplacer en cas de problème
 - pouvoir être surveillée et mise à jour à distance
- Nous sommes le premier pays à avoir réalisé cette approche :
 - Cette solution spécifique a été brevetée
 - Un audit de sécurité poussé a été récemment réalisé et a montré son excellent niveau de sécurité.



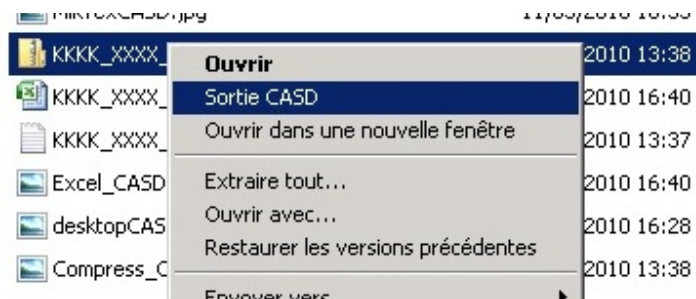
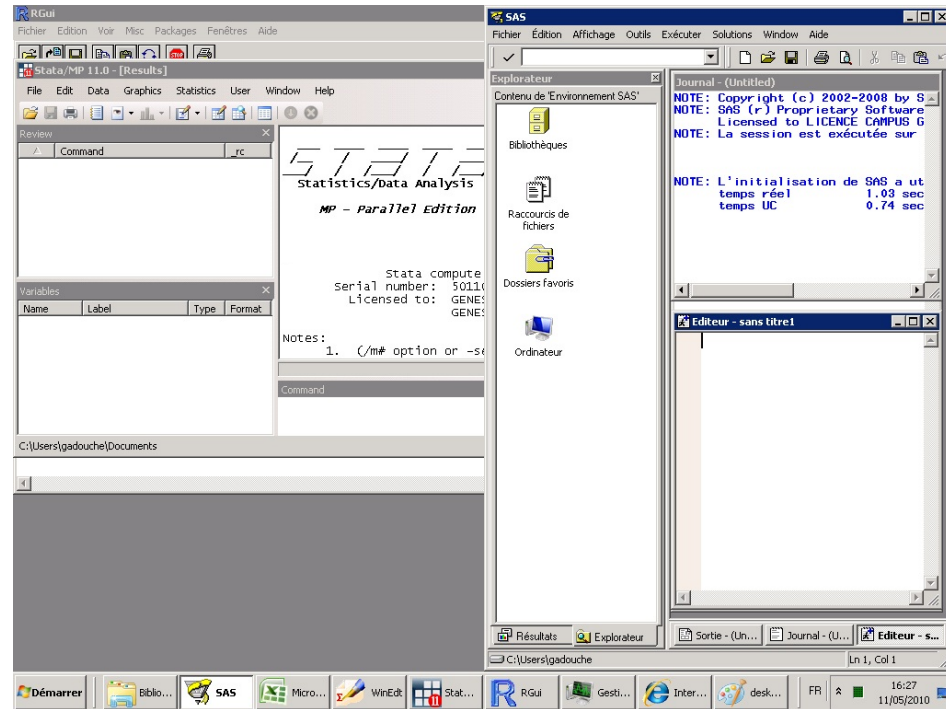
Comment utiliser le CASD ?

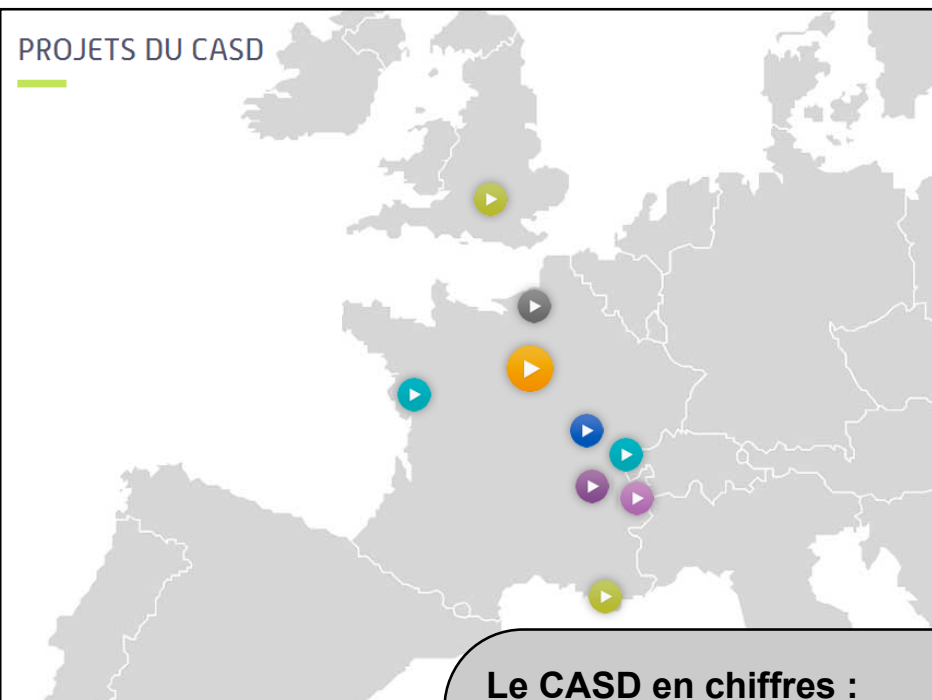
- **Pour travailler sur des données confidentielles, un chercheur doit :**
 - Faire un dossier au comité du secret statistique qui émet un avis qui est ensuite suivi d'une décision de la direction des archives.
 - Suivre une séance d'une demi journée au GENES de sensibilisation aux aspects juridiques, informatiques et concernant le secret statistique
- Le GENES fournit alors une SD-Box : appareil spécifique pour accéder à distance aux données et aux logiciels scientifiques pour travailler : notion de bulle de travail.
- Le chercheur peut travailler sur les données dans sa bulle mais ne peut pas :
 - Récupérer des données sans contrôle par un opérateur
 - Entrer des données sans contrôle par un opérateur
 - Accéder à des données auxquelles il n'est pas autorisé
- Une offre large de logiciels scientifiques est offerte au chercheur dans son environnement de travail



Utilisateurs, contrats et charte

- Séances d'enrôlement obligatoire
 - Sensibilisation juridique
 - Sensibilisation informatique
- Charte d'utilisation signée par le chercheur :
 - Accord de confidentialité
 - Conditions d'utilisation : ne pas céder son accès (carte), etc...
- Contrat signé avec l'établissement hébergeant le boîtier
 - Contrainte d'emplacement du boîtier (pas sur un lieu de passage, écran non lisible par d'autres personnes, local fermant à clé...)
 - Identification de l'établissement par son adresse postale et son adresse réseau
- Extension des accès aux plateformes universitaires de données :
 - Des points d'accès CASD seront disponibles dans les PUD
 - Cela permettra un meilleur accompagnement
 - Des formations sur les données seront organisées
 - Portail en cours de développement





Le CASD en chiffres :

- 322 chercheurs en France et à l'étranger (UK, Italie),
- Répartis sur 143 projets de recherche.
- Une trentaine de sources de données disponibles sur le CASD

A titre de comparaison :

- Au Danemark : environ 1 000 chercheurs
- Aux Pays-Bas : environ 400 chercheurs
- En Suède : Environ 420 chercheurs

QUELQUES INSTITUTS ACCEDANT AU CASD

Instituts	Ville	Nombre de Projets
Centre de données-socio politiques de Sciences Po	Paris	5
Maison des Sciences Economiques- Centre d'Economie de la Sorbonne (Sorbonne)	Paris	3
Paris School of Economics	Paris	8
Groupe de Recherche en Economie Quantitative d'Aix-Marseille	Marseille	4
Centre de Recherche en Economie et Statistique	Paris	18
Institut National d'Etudes Démographique	Paris	5
Royal College University de Londres	Londres	2
Groupe d'Analyse et de Théorie Economique - Lyon - Saint-Etienne	Lyon	4
Centre d'analyse stratégique	Paris	2
Géographie-Cités	Paris	3



GENES

Mise en place de la facturation en 2012

- Une mise en place de la facturation à partir d'octobre 2012

Les principales modalités de facturation et de paiement seront les suivantes :

- La durée minimale d'un projet sera d'un an sauf cas particulier
- La facturation et le paiement associés peuvent être annuels ou uniques
- Il sera possible de mettre l'abonnement en veille dans des conditions précises.
- une institution pourra régler les factures de plusieurs chercheurs
- Un contrat qui stipulera les conditions d'utilisation, les modalités et les obligations du CASD et du chercheur ;



Recherche de subventions par le CASD

- Le CASD et ses partenaires financeront l'ensemble des coûts fixes (c'est-à-dire l'ensemble des coûts indépendants du nombre de projets et de chercheurs) par le biais de subventions (Equipex, ANR...)
- Seuls sont partiellement facturés les coûts variables dépendant du projet et du nombre de chercheurs
- les doctorants et post-doctorants bénéficieront de réductions financées par le CASD et ses partenaires (après avis du conseil scientifique du CASD)
- Un schéma de facturation appliqué par palier :
 - Permet une meilleure anticipation pour le chercheur
 - Si des subventions/participations sont **prix en prendra compte** (en diminuant l
 - Les premières années, le différentiel se charge par les partenaires et equipex



		Tarif sur la période 2012 - 2014	Tarif sur la période 2015 - 2016	Tarif sur la période 2017 - 2018	Tarif à partir de 2019
Forfait Standard	L'ensemble des services "Standard" du CASD Configuration 1 (<i>cf. ci-dessous</i>) 20 Entrées/Sorties 1 SD-Box 1 Lecteur biométrique Accès pour 1 utilisateur/chercheur	83,00 € HT / mois	125,00 € HT / mois	167,00 € HT / mois	238,00 € HT / mois
Suppléments	Utilisateur 2	+ 37,00 € HT / mois			
	Utilisateur 3	+ 35,00 € HT / mois			
	Utilisateur 4	+ 30,00 € HT / mois			
	Utilisateur 5	+ 25,00 € HT / mois			
	Utilisateur 6 et au-delà	+ 20,00 € HT / mois			
	Point d'accès (SD-Box + Lecteur biométrique) supplémentaire	+ 35,00 € HT / mois			
	Pas de point d'accès (si déjà existant)	- 35,00 € HT / mois			
	Configuration matérielle 2 (<i>cf. ci-dessous</i>)	+ 69,00 € HT / mois			
	Configuration matérielle 3 (<i>cf. ci-dessous</i>)	+ 138,00 € HT / mois			
	Service de conservation (Espace de 20Go)	5,00 € HT / mois			
	Pack de 10 Imports/exports supplémentaires	300,00 € HT			
	Import ou export supplémentaire si total supérieur à 30	100,00€ HT par import ou export			
	Installation d'un logiciel spécifique	sur devis			
	Carte d'accès supplémentaire	25,00€ HT			
	Réparation / Non restitution du matériel	780,00€ HT			

Comparaison avec les autres pays

- le tarif (projet à 238€/mois) du CASD est entre 5 et 6 fois moins élevé que celui de CBS (Pays-Bas) dont les tarifs sont en ligne et qui dispose d'un système assez proche.
- En Angleterre, L'ONS (office statistique britannique) applique une tarification de 60€ par jour d'accès, ce qui représente en moyenne annuelle observée par utilisateur environ 2300€ (38 jours).
- Aux USA, au NORC (National Opinion Research Center, University of Chicago), le coût annuel d'un point d'accès aux données, partagé en moyenne par 30 chercheurs, est de 210K€ par an, soit environ 7000€ par chercheur et par an.
- En Suède, le coût annuel par chercheur est de 2600€.



Une visibilité nationale internationale

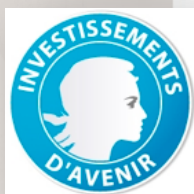


Institut national de la statistique
et des études économiques



Réseau Quetelet

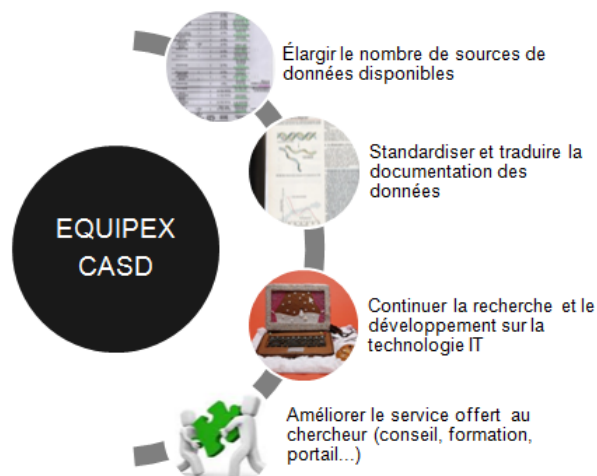
Réseau français des centres de données pour les sciences sociales
French Data Archives for social sciences



GENES

Le CASD, lauréat EQUIPEX

- Le projet CASD est l'un des lauréats de la première vague de l'appel à projet « équipement d'excellence » **equipex** :
 - 4 M€ des investissements d'avenir ont ainsi été attribués au projet
 - Consortium : Genes, Insee, GIS Quetelet, ENS Cachan, École polytechnique et HEC
- Une opportunité pour réussir un développement ambitieux du CASD, selon plusieurs axes de travail :



GRUPE DES ÉCOLES NATIONALES D'ÉCONOMIE ET DE STATISTIQUE

et

Deux projets européens



20 pays ayant pour objectif de favoriser l'accès aux micro-données par les chercheurs et de réaliser un réseau de centres d'accès sécurisés.



Projet EssNet DARA regroupant plusieurs instituts de statistique européens (Allemagne, Portugal, Angleterre..) pour la mise en place d'un pilote de centre d'accès sécurisé européen à Eurostat.

Au moins 3 rencontres internationales sur ce sujet ont lieu chaque année

Perspectives

- **Big Data** : explorer et analyser les données du CASD en les enrichissant des données du Big Data :
 - Acquisition par le CASD de données libres du Big Data (réseaux sociaux...)
 - Travailler sur les aspects 'sécurité' et juridiques sur des plateformes Big Data en construction ou sur des projets R&D Big Data
- **Data Matching** : Créer un environnement sécurisé pour pouvoir utiliser différentes sources de données de manière conjointe dans le respect des libertés individuelles et des lois :
 - une méthodologie complexe à mettre en œuvre pour garantir l'anonymat et la sécurité
 - Un potentiel infini d'enrichissement de la donnée
- **Concevoir un réseau de CASD fédéré**: Etendre le CASD à un réseau de centres d'accès sécurisé accessibles uniquement par des SD-Box sécurisées
 - un niveau élevé et homogène de sécurité
 - Une industrialisation de fonction de production et d'exploitation
 - Une allocation dynamique des ressources virtualisées