



SD-Box : boîtier sécurisé d'accès

► **EQUIPEX CASD** Centre d'accès sécurisé aux données

Valoriser les données de tous horizons

En proposant un accès sécurisé à une multitude de données sensibles sur les entreprises et particuliers français, l'Equipex CASD intéresse un nombre croissant de chercheurs et de chargés d'études du secteur privé et public.

Permettre l'accès à un nombre considérable de données très détaillées sur les individus, les ménages et les entreprises tout en garantissant des conditions de sécurité et de confidentialité très élevées. Telle est la mission du CASD (Centre d'Accès Sécurisé aux Données). A l'heure où les projets Big Data se multiplient, le CASD met à disposition des chercheurs de tous horizons et de la data science une technologie unique au service de la valorisation des données sensibles.

Le CASD est une des directions du GENES (Groupe des Ecoles Nationales d'Economie et de Statistique), un établissement public dont dépendent notamment l'ENSAE (Ecole nationale de la statistique et de l'administration économique) ou l'ENSAI (Ecole nationale de la statistique et de l'analyse de l'information) à Rennes. La structure a vu le jour en 2008 et a ensuite bénéficié de l'important effet de levier que représente le PIA (Programme d'investissements d'avenir). « Nous avons 30 utilisateurs en 2010. On en totalise aujourd'hui environ 1 500 », se réjouit Kamel Gadouche, directeur du CASD.

Avec ses innombrables informations recueillies au fil des enquêtes socio-économiques, l'INSEE représente le premier fournisseur de données pour le centre, et une véritable mine d'or pour de nombreuses organisations désireuses de mener des recherches en exploitant cette masse d'éléments collectés. « Or, ces informations sont par nature indirectement identifiantes, pouvant permettre par divers croisements de remonter aux données individuelles, ce que la technologie du CASD interdit. L'Insee a donc bien sûr une obligation de secret par rapport à celles-ci et doit veiller à conserver l'anonymat des individus, ménages ou entreprises concernés par ces données. Nous avons alors imaginé une offre, un service, s'appuyant sur un dispositif



**OUVRIRE L'ACCÈS
AUX DONNÉES
DE MANIÈRE
SECURISÉE EST
UN VÉRITABLE
CATALYSEUR
POUR FAIRE PROGRESSER
LA CONNAISSANCE**

**Kamel Gadouche,
directeur du CASD.**

securisé qui permette la mise à disposition de ces données très convoitées, sans risque de divulgation », explique le directeur.

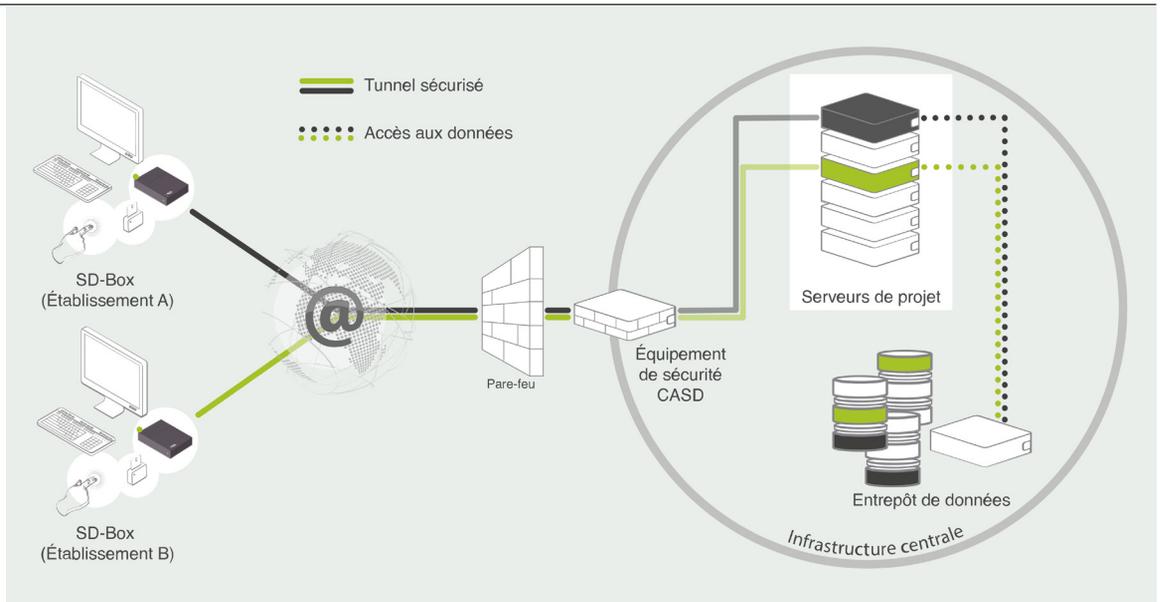
Grâce à une authentification forte par certificat et biométrie, le chercheur se connecte aux serveurs du CASD organisés en « bulle » hautement sécurisée, par le biais d'un boîtier numérique dans laquelle il insère une carte à puce. Il peut ensuite travailler à distance. En aucun cas, l'utilisateur ne peut exporter les données et ne peut sortir des résultats sans passer par une phase de contrôle manuelle ou

automatique garantissant l'anonymisation.

POUR DES ÉTUDES MULTI-DOMAINES

En plus des données issues de l'Insee, les informations hébergées proviennent de l'administration fiscale, du secteur de l'agriculture, du ministère de la Justice, de l'Éducation, de l'Environnement, du Travail ou encore de la Banque Publique d'Investissement. A cela s'ajoute un nouveau domaine, vaste et prometteur : les données de santé. Elles sont aujourd'hui très convoitées, le passage par une autorité habilitante, en l'occurrence la CNIL, est une obligation pour pouvoir avoir accès aux informations. Les conditions de sécurité d'accès à ces données sont aussi très drastiques. C'est pourquoi le CASD peut apporter une réponse aux organismes qui n'ont pas les moyens d'investir dans des dispositifs et des développements coûteux pour l'accès aux données de santé. Jusque-là par exemple, les laboratoires pharmaceutiques n'avaient pas accès aux données mais l'entrée en vigueur d'une nouvelle loi votée en 2016 leur permet d'obtenir un accès indirect, par l'intermédiaire d'un bureau d'études extérieur, pour la réalisation d'études à leur demande. La CNIL est également un interlocuteur obligatoire pour d'autres familles de données impliquant un caractère personnel, mais dans un cadre moins contraignant. Il s'agit alors d'une déclaration indiquant le recours aux données en passant par le CASD. Les données d'entreprise ne sont pas concernées par ce type de procédure.

« Nous mettons notre technologie brevetée au service d'entreprises qui souhaitent mettre à disposition leurs données. C'est le cas par exemple de RTE (Réseau de Transport d'Électricité) qui détient de grands volumes d'informations stratégiques concernant les consommations électriques des sociétés. RTE a souhaité disposer d'un



datalab dans lequel ils peuvent collaborer avec des consultants extérieurs ou des chercheurs sans que ces derniers aient à venir dans leurs locaux », illustre Kamel Gadouche. D'autres groupes comme BNP Paribas ou Michelin ont manifesté des intérêts analogues.

Les producteurs issus de la sphère publique mettent leurs données à disposition gratuitement pour contribuer à l'amélioration de la connaissance grâce à la recherche scientifique et permettre leur exploitation par des chargés d'études extérieurs à leur administration. Ces ressources sont ainsi utilisées par des chercheurs pour leurs sujets d'études. « Dans ce cas, les producteurs de données n'ont pas de droit de regard sur les réalisations des chercheurs, à la différence des situations où les producteurs sont des acteurs privés, car ces derniers sont alors commanditaires des études menées », indique Kamel Gadouche.

La demande auprès du CASD connaît une forte croissance depuis quelques années. Au-delà des clients et utilisateurs français, les entreprises, chercheurs ou organisations européennes occupent une place toujours plus importante.

DES PROJETS D'AVENIR

« Nous avons investi depuis 2012 dans un projet relatif au Big Data et à la Data Science qui s'intitule TeraLab, mené en partenariat avec l'Institut Mines-Telecom. Il consiste en une plateforme dans laquelle on trouve un compartiment hyper-sécurisé

**BEAUCOUP
D'ENTREPRISES
SOUS-EXPLOITENT
LE POTENTIEL
DE LEURS DONNÉES
POUR DES PROBLÈMES
DE SÉCURITÉ, ALORS
QUE DES SOLUTIONS
ÉPROUVÉES EXISTENT**

administré par le GENES et doté de la technologie du CASD (CASD - TeraLab) dont la vocation est de mettre à disposition des entreprises intéressées un environnement enrichi de technologies big data qu'ils peuvent expérimenter. Pour les premiers pas d'une structure vers les technologies Big Data, nous proposons aussi un accompagnement pour mettre en place, élaborer l'environnement de travail selon les besoins exprimés et sécuriser l'accès. Nous travaillons également avec la filiale Datastorm du Genes qui propose du conseil en datascience sur les aspects métiers », souligne Kamel Gadouche.

De nombreux projets d'appariements entre des sources de données différentes sont également opérés. Il s'agit par exemple de procéder à des croisements de données socio-économiques et de données de santé.

« Dans ces cas de figure où on apparie les données au niveau de l'individu, la CNIL se montre bien sûr très vigilante. Mais cela permet par exemple de mesurer l'impact de la santé sur les trajectoires professionnelles, et obtenir des éléments

sur lesquels il n'existe pas de chiffres aujourd'hui. Il s'agit là d'un fabuleux potentiel, tant pour la recherche que pour le citoyen ou le décideur politique », poursuit le directeur.

La France s'est lancée plus tardivement que d'autres pays dans l'exploitation massive de données, ce qui représente paradoxalement un avantage à l'heure actuelle. Le CASD est un centre d'accès sécurisé donnant accès de façon uniforme à un grand nombre de sources de données. Mener des études croisées à distance impliquant des données du ministère de la Justice et de l'Insee devient ainsi possible de façon plus aisée et dans les meilleures conditions de travail possibles pour l'utilisateur. A l'étranger, différents centres physiques ont vu le jour par le passé pour traiter, de manière isolée, les informations de la Justice, de l'Education, des Statistiques, du Travail... Ce développement en silos peut actuellement être un vrai frein pour des études innovantes multi-domaines. Ce contexte explique l'importance du nombre de chercheurs étrangers qui se tournent désormais vers les services du CASD. ■

CASD 
CENTRE D'ACCÈS SÉCURISÉ AUX DONNÉES

CONTACT CASD – GENES

5 avenue Henry le Châtelier
91170 Palaiseau
<https://www.casd.eu>