

Collaboration Centre d'Accès Sécurisé aux Données (CASD) & Banque de France

CNIS – COMMISSION SYSTÈME
FINANCIER ET FINANCEMENT
DEL'ÉCONOMIE

NOV. 2020

Kamel GADOUCHE (CASD)
Vincent DUPRAT (Banque de France)

Sommaire

- 1. Brève présentation du CASD**
- 2. Problématiques techniques et fonctionnelles**
- 3. Collaborations en cours**
- 4. Problématiques juridiques**
- 5. Perspectives de collaboration**

1. Brève présentation du CASD

□ Le CASD est un groupement d'intérêt public à but non lucratif

- **Regroupant** l'Etat représenté par l'Insee, le Genes, le CNRS, l'Ecole polytechnique et HEC Paris
- **25 personnes** : Services IT et Datascience, Data Management, Project Management
- **Mission principale** : mise à disposition sécurisée de données de recherche à des fins non lucratives de recherche, d'étude, d'évaluation et d'innovation.
- **Tiers de confiance pour la mise à disposition sécurisée** : le CASD n'a aucune propriété sur les données qui lui sont confiées contractuellement.
- **Niveau de sécurité certifié** : ISO 27001 (sécurité de l'information), ISO 27701 (protection des données à caractère personnel – RGPD), Hébergeur de données de santé.
- **Audit de sécurité techniques réguliers** : réalisés par des « ethical hackers » de sociétés certifiées par l'Agence nationale de la sécurité des systèmes d'informations
- **Gouvernance** : un conseil scientifique, un comité des producteurs et un comité de suivi de la sécurité de l'information
- **Usages** : 300 sources de données mises à disposition, 500 institutions utilisatrices et plus de 1500 chercheurs utilisant le CASD

2. Problématiques techniques et fonctionnelles

❑ Des différences structurelles

	CASD	ODR
Mode d'accès aux données	Distant	Sur site
Tarifcation	Payant	Gratuit
Hachage de l'identifiant	Non	Oui

❑ Un besoin de données CASD et Banque de France pour les chercheurs

- Qui peut parfois être servi via deux accès séparés/ parallèles aux données
- Posant problème pour les recherches impliquant des appariements / croisements entre les données CASD et les données Banque de France
- Des chercheurs qui parfois s'autocensurent sur la source « minoritaire » de données

3. Collaborations en cours

- ❑ **Fourniture des données pour l'Inspection générale des finances**
 - Signature d'une convention entre le CASD et la Banque de France, spécifique à un usage (étude en lien avec la crise sanitaire) et un usager (l'IGF)
 - Contexte juridique favorable, l'IGF étant assimilée au Ministère des Finances
 - Données transmises gratuitement par la Banque de France et mises à disposition non anonymisées pour l'IGF dans l'environnement CASD

- ❑ **Pilote sur base d'échanges croisés de donnée**
 - Pour une exploitation des données dans l'un ou l'autre des environnements
 - Impliquant une anonymisation des données Banque de France pour un chercheur « lambda » et par extension une anonymisation des données CASD devant être croisée avec les données Banque de France et ce quelque que soit l'environnement d'exploitation
 - Ce modus operandi doit s'appuyer sur une convention d'échange de données actuellement à l'étude par la Direction Juridique de la Banque de France (sur le périmètre des données EFI, ECO et ECEIS dans un premier temps)

4. Problématiques juridiques

- ❑ **La Banque de France et/ou l'EuroSystème doit tenir compte des spécificités juridiques :**
 - Chaque jeu de données a potentiellement un corpus de texte propre imposant des contraintes
 - Hachage (pseudonymisation)

- ❑ **Il y a des dispositions particulières pour les données bancaires :**
 - Pour les données disposant d'un cadre européen, les dispositions doivent être respectées à un niveau national (soit pseudonymisation ou agrégation)
 - Une revue des textes nationaux est nécessaire pour étudier les possibilités pour les autres catégories de données

- ❑ **Il n'y a pas à l'heure actuelle de projet de plateforme commune ou d'interconnexion structurelle des deux plateformes**

- ❑ **L'orientation est d'explorer une collaboration d'échange de données :**
 - Étude projet par projet : en fonction des caractéristiques du projet notamment les contraintes liées aux sources mobilisées
 - Régime juridique des données financières concernées
 - Accord du producteur pour la transmission à l'ODR
 - Données fiscales exclusivement accessibles par le CASD (loi ESR2013)

5. Perspectives de collaboration

□ L'ODR 2.0 va amoindrir les différences structurelles

	CASD	ODR ₂	ODR (pour rappel)
Mode d'accès aux données	Distant	Distant	Sur site
Tarifcation	Payant	Payant	Gratuit
Hachage de l'identifiant	Non	Partiel	Oui

- Accès distant dans les deux cas (mais exigences de sécurité non alignées à date) : mise en œuvre de la recommandation n° 12 du GT CNIS accès aux données bancaires et financières
- Service payant dans les deux cas (mais tarification différente)
- Réflexion en cours sur l'assouplissement de l'anonymisation (pas d'application systématique sur l'ensemble des données de l'ODR)

□ Les projets d'échanges croisés de données seront facilités avec l'ODR 2.0