

# LES DONNÉES INDIVIDUELLES DISPONIBLES AU NIVEAU EUROPÉEN

*Roxane Silberman*

CNRS & CASD

CNIS - Commission *Système financier et financement de l'économie* – Novembre 2020

Secure Data HuB



CASD C

# LE RAPPROCHEMENT DES DONNÉES FINANCIÈRES GRANULAIRES EN EUROPE

## UNE INTRODUCTION

Présentations de deux projets dans cette session:

- Inexda (International Network for Exchanging Expérience on Statistical Handling of Granular Data)
- Anacredit Banque Centrale Européenne

3 points en introduction de cette séance sur le rapprochement des données financières granulaires en Europe

- Une très brève vue d'ensemble sur l'évolution générale en ce qui concerne la disponibilité pour la recherche des données confidentielles, parmi lesquelles les données financières granulaires
- Les problèmes posés pour l'utilisation des données de plusieurs pays
- Un exemple d'initiative sur ce point - dans le domaine des données de la statistique publique et des données administratives

IDAN - International Data Access Network

## LE CONTEXTE GÉNÉRAL : UNE OUVERTURE CROISSANTE DE L'ACCÈS AUX DONNÉES CONFIDENTIELLES

- Un mouvement d'ouverture croissante des données couvertes par des secrets, partout et notamment en Europe
- Une expérience de près de 20 ans sur ces questions d'accès sécurisé avec une ouverture large des données confidentielles des instituts nationaux de statistiques et des données administratives pour les pays dit à registres
- Des accès au départ sur site puis une évolution massive vers l'accès à distance
- De fortes similitudes sur les principes d'accès
  - Bien résumées par les *5 Safes* (cadre proposé par l'ONS (Office National de Statistique du Royaume-Uni) : chercheur, projet de recherche, données, sécurité de l'accès sécurisé, contrôle de l'anonymisation des résultats
  - Modulation des contraintes de sécurité en fonction du niveau de risque
- Des différences dans les modalités et les procédures :
  - Accréditation: institution + chercheur vs chercheur
  - *Remote execution ou job submission* vs *remote access* avec accès direct aux données, préféré par les chercheurs
  - Pour le *remote access*: accès depuis des *safe rooms* vs bureau des chercheurs dans les universités vs domicile
  - Intégralité des fichiers vs extractions (notamment pour les données administratives)
  - Contrôle systématique des résultats vs contrôle aléatoire

## UNE OUVERTURE QUI CONCERNE AUSSI LES DONNÉES FINANCIÈRES GRANULAIRES

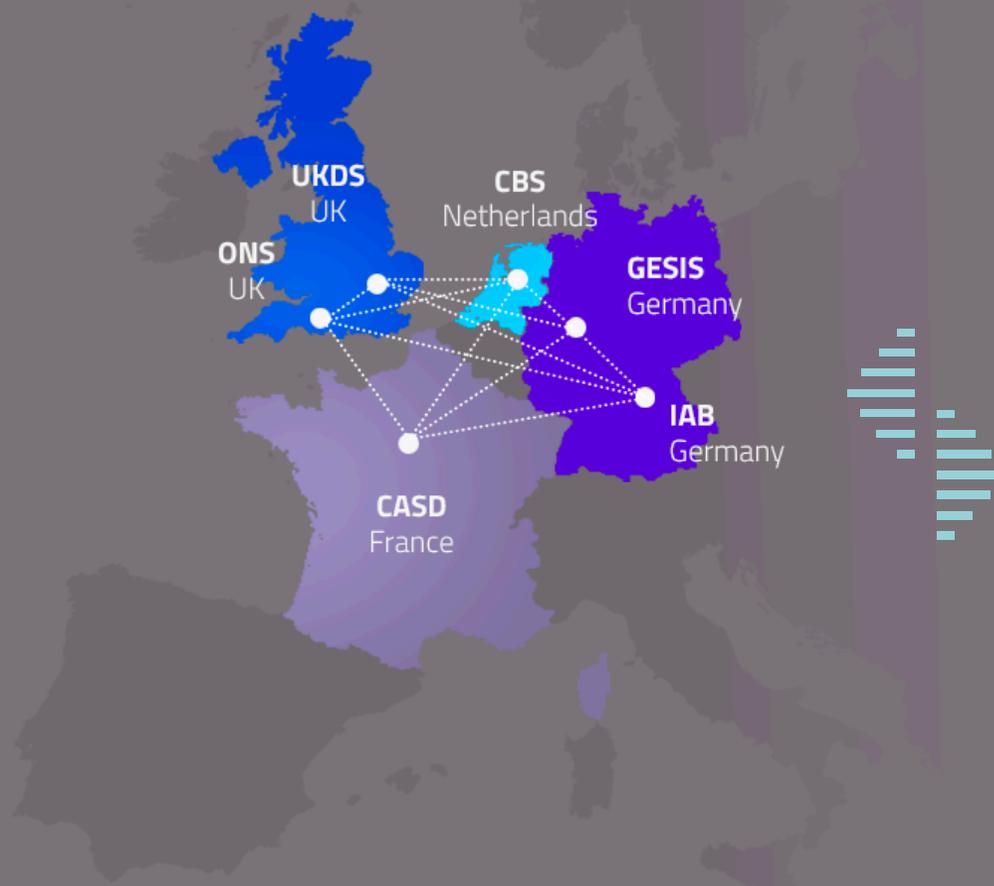
- Une ouverture plus récente
  - Le rapport Capelle-Blancard du groupe de travail du CNIS en 2015
- Un mouvement qui se poursuit depuis
  - Principalement, une formalisation des *Data Room* sur site et de l'ouverture aux chercheurs extérieurs
  - Une progression de l'accès à distance en remote execution ou job submission
    - ex: Bundesbank, Banca d'Italia, Banco di Espana (en projet)
  - Quelques accès en *remote access*
    - Ex: DND, la Banque centrale des Pays Bas, le projet Banque de France d'Open Data Room virtuelle
  - Des coopérations pour combiner données de la statistique publique et données des banques centrales
    - Ex: le centre sécurisé de CBS, l'Institut national de statistiques des Pays-Bas avec la Banque centrale DND (accès via CBS), le projet Banque de France et CASD (Centre d'accès sécurisé aux données)

# LA QUESTION DE L'ACCÈS TRANSNATIONAL ET DE L'UTILISATION CONJOINTE DE DONNÉES DE DIFFÉRENTS PAYS

- Une question cruciale
  - Accès à des données originales
  - Approches comparatives avec données de plusieurs pays
  - Mobilité croissante des chercheurs et travail en collaboration entre équipes de différents pays
- Les aspects juridiques de l'accès transnational aux microdonnées confidentielles pour la recherche
  - Les discussions OCDE pour l'accès transnationale aux données nationales, Eurostat pour l'accès des chercheurs aux microdonnées européennes intégrées
    - La loi en général « silencieuse »
    - Ce qui est en jeu: la sécurité, les équivalences (sanctions en cas de rupture de confidentialité) et les clauses contractuelles
    - L'idée centrale : un « Circle of trust » (OCDE)
  - Le RGDP européen conforte le transfert de données personnelles UE (et hors UE)
- Une croissance de l'accès transnational sous différentes formes
  - Avec déplacement du chercheur : au Royaume-Uni sur les centres sécurisés de ONS-Office National de statistique au Royaume-Uni, à Luxembourg pour les microdonnées Eurostat actuellement
  - Accès distant depuis le pays du demandeur pour Statistics Denmark, CBS-Institut de statistique des Pays-Bas, CASD en France
  - Encore limité pour les banques centrales
- Des difficultés qui demeurent importantes pour l'utilisation de données confidentielles de plusieurs pays
  - Intégration et harmonisation en amont (ex Eurostat) + accord nécessaire des pays pour l'accès
  - Données nationales (ex données administratives): déplacements multiples, équipements d'accès différents, procédures multiples nécessaires + ils ne permettant pas le traitement conjoint de données dans un même environnement de recherche

## UN EXEMPLE D'INITIATIVE DU CÔTÉ DE LA STATISTIQUE PUBLIQUE

- Un background:
  - FP7 DwB-Data without Boundaries 2011-2015 avec 14 pays, 29 partenaires dont plusieurs instituts nationaux de statistique, un pilote pour un réseau intégré de centres sécurisés
  - NORDMAN: 2015-2018, un pilote de 6 pays nordiques permettant le transfert de données confidentielles pour un finalité de recherche entre instituts nationaux de statistiques de données confidentielles
- IDAN- International Data Access Network créé en 2018
  - Objectif: faciliter l'accès transnational aux données confidentielles et permettre aux chercheurs de travailler ensemble avec les sources des différents partenaires/pays via des accès réciproques entre partenaires
  - Un *Coordination Understanding* entre 6 centres sécurisés de 4 pays
    - CASD-Centre d'accès sécurisé aux données, France (coordinateur)
    - CBS-Institut national de statistique, Pays-Bas
    - GESIS- Data Archive, Université de Mannheim & Cologne, Allemagne
    - IAB-FDZ (Institute for Employment Research/Federal Employment Agency, Allemagne
    - ONS-SRS Office National de statistique-Secure Research Service, Royaume-Uni
    - UK Data Service/ Secure Lab, Royaume-Uni



# IDAN

INTERNATIONAL DATA ACCESS NETWORK



Une situation initiale compliquée pour les chercheurs impliquant de multiples déplacements et de grandes difficultés pour travailler sur les données de ces différents pays ....

# INITIAL SITUATION

## CASD

Accreditation (CSS)



Training



Access



## IAB

Accreditation



Training



Access



## CBS

Accreditation



Training



Access



## ONS

Accreditation



Training



Access



## GESIS

Accreditation



Training



Access



## UKDS

Accreditation



Training



Access

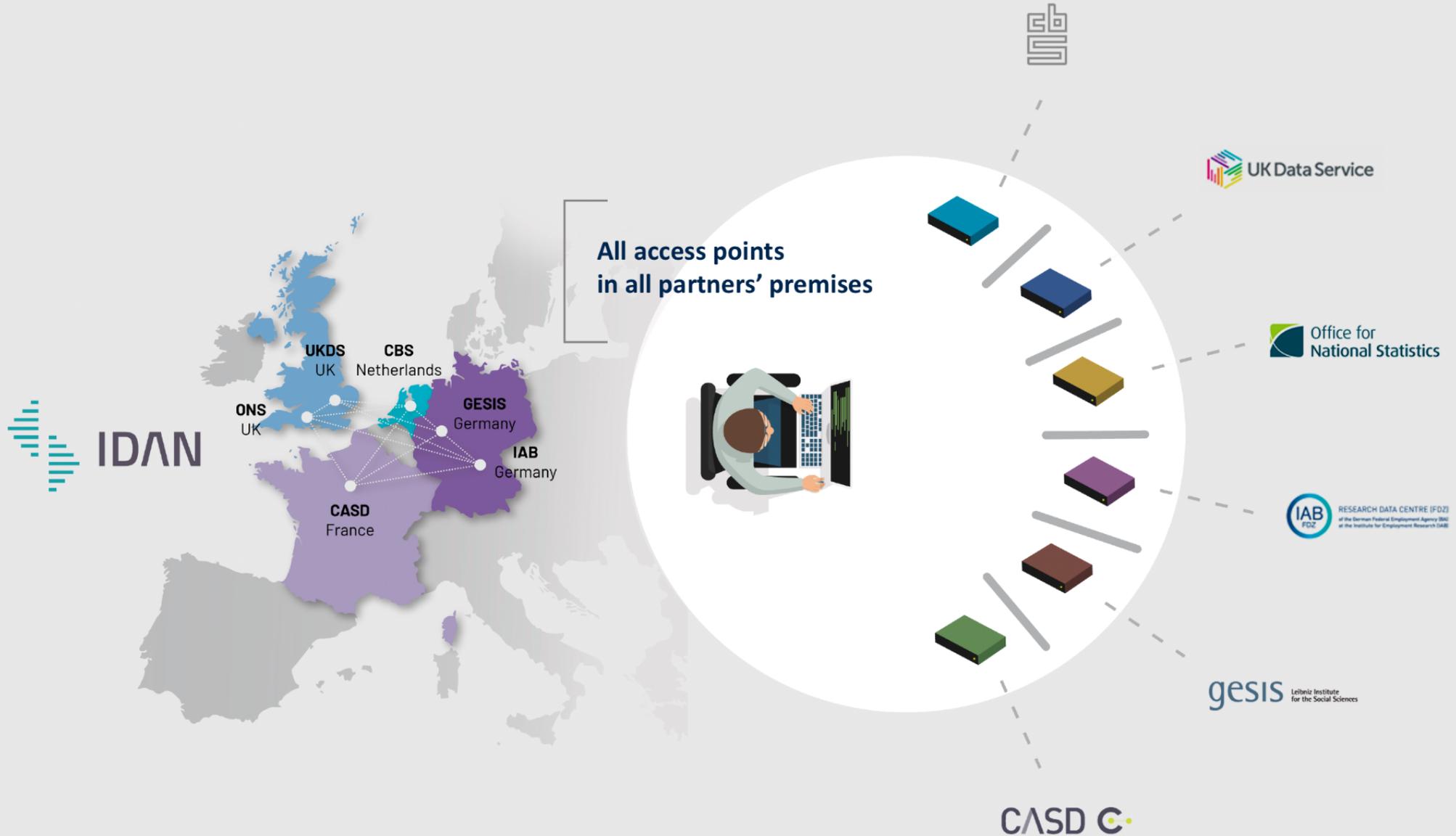


## UNE STRATÉGIE PRAGMATIQUE POUR AVANCER RAPIDEMENT

---

- Centrée sur l'accès : accès réciproques avec 6 points d'accès chez chacun des 6 partenaires
- Pas le " monde idéal " : pas le même équipement, pas le même environnement de recherche, mais tous les accès au même endroit (*Safe Room*) chez chaque partenaire
- L'accréditation, la formation/enrôlement, le contrôle des sorties restent dans les mains des partenaires

# International Data Access network (IDAN)



## Des situations de départ différentes

- *Remote access* transnational à partir des universités: CASD, de n'importe où: CBS
- Accès sur site ou à partir d'une Safe Room mais pas au même niveau de détail que sur site: IAB
- *Remote access* seulement à l'intérieur du pays - UK Data Service
- Accès uniquement sur site - GESIS, ONS-SRS

## Des changements requis incluant

- Passage à un *Remote access* au même niveau de détail que sur site (IAB)
- Passage à un *Remote access* transnational (ONS-SRS, UK Data Service prévu 2021)
- Création d'une Safe Room (CASD, CBS)
- Harmonisation et adaptation des requis de sécurité pour les *Safe Rooms* pour pouvoir implanter les points d'accès des partenaires dans un même lieu

# UNE IMPLANTATION PROGRESSIVE DES POINTS D'ACCÈS ...



FRANCE  
CASD



NETHERLAND  
CBS



GERMANY  
GESIS



GERMANY  
IAB



UK  
ONS



UK  
UKDS



## ET AUSSI

---

- Site web IDAN (<https://idan.network/>)
  - Tableaux synthétiques avec liens sur les procédures d'accréditation et modes d'accès des partenaires
  - Information sur les procédures d'accès aux *safe rooms* des partenaires
  - Liens vers les catalogues de données et *Top tens*
- En cours
- Traduction des metadonnées en anglais sur les sites des partenaires
- Travail sur la comparabilité des données:
  - Un premier travail avec des [tables de comparaisons](#) (CASD-IAB) pour des données françaises et allemandes

## PERSPECTIVES

---

- Des opportunités immédiates: des projets de recherche qui ont démarrés (CASD-IAB-GESIS)
  - Un cadre pour des développements futurs
    - Un contrat multilatéral incluant notamment les réquisits de sécurité et les accords en cas de rupture de confidentialité
    - Harmoniser et partager les formations/enrôlements
    - Un cadre pour des transferts de données pour les projets requérant une utilisation conjointe
    - Une extension progressive à d'autres partenaires
-

## LES INITIATIVES DU CÔTÉ DES DONNÉES FINANCIÈRES GRANULAIRES

---

- INEXDA
  - ANACREDIT
- 
-

**Merci !**  
**Des questions ?**

<https://www.casd.eu/>

<https://idan.network/>

[roxane.silberman@casd.eu](mailto:roxane.silberman@casd.eu)

