



Comment rendre l'utilisation des données compatible avec le respect de l'éthique de nos organisations ?

Bertrand Monthubert – Professeur, Université de Toulouse
Président du Conseil National de l'Information géolocalisée, d'Ekitia, d'OPenIG
Co-président du Groupe de travail « Gouvernance de la donnée » du Partenariat Mondial de l'IA

- Problématiques d'accès aux données sensibles
- La gouvernance des données : le cadre éthique
- Les technologies d'accroissement de la confidentialité
- Vers des espaces communs de données multi-sectoriels

Problématiques d'accès aux données sensibles

Problème fondamental : opérer un traitement de données qui proviennent d'organisations différentes qui ne peuvent pas/ne veulent pas les transmettre

- ❑ Croisement de données médicales, socio-économiques, et de localisation pour déployer des politiques de prévention et de santé publique
- ❑ Croisement de données relatives au handicap avec des données liées à l'emploi

ekitia | D'où viennent les données ?

DONNEES PUBLIQUES



**COLLECTIVITES,
ADMINISTRATIONS**

Ex : données relatives aux équipements socio-culturels, base d'adresses, subventions accordées...



**ETABLISSEMENTS
PUBLICS OU
CHARGES DE
MISSION PUBLIQUE**

Ex : données de recherche, données spatiales, statistiques publiques...

DONNEES PRIVEES



ENTREPRISES

Ex : base client, données liées aux produits et services mis sur le marché...



INDIVIDUS

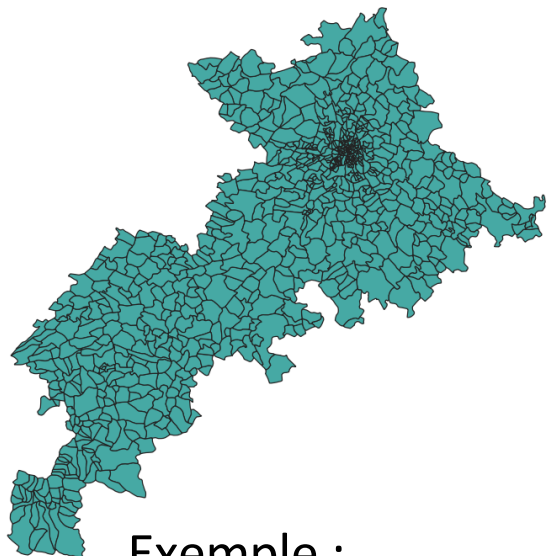
Ex : nom, prénom, adresse email, adresse IP, géolocalisation...

ekitia | Un cas concret : catégorie sociale et cancer

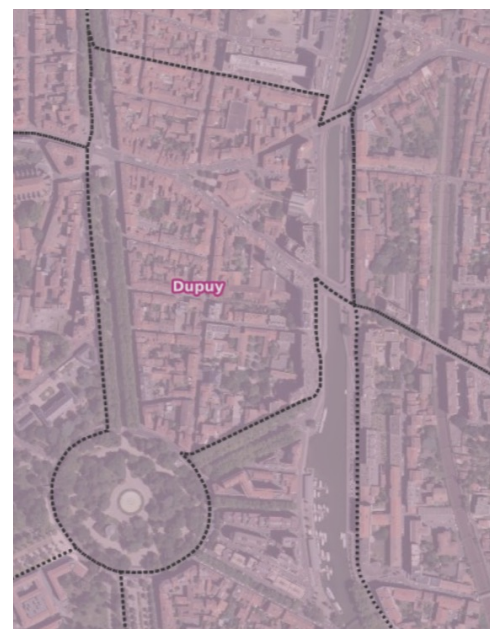
- Etude sur l'impact de la condition sociale sur le degré d'avancement d'un cancer du sein lors d'une première consultation chez un oncologue
- Lamy, S., Molinié, F., Daubisse-Marliac, L. *et al.* Using ecological socioeconomic position (SEP) measures to deal with sample bias introduced by incomplete individual-level measures: inequalities in breast cancer stage at diagnosis as an example. *BMC Public Health* **19**, 857 (2019). <https://doi.org/10.1186/s12889-019-7220-4>



Besoin de croiser des données sociales
et des données de santé

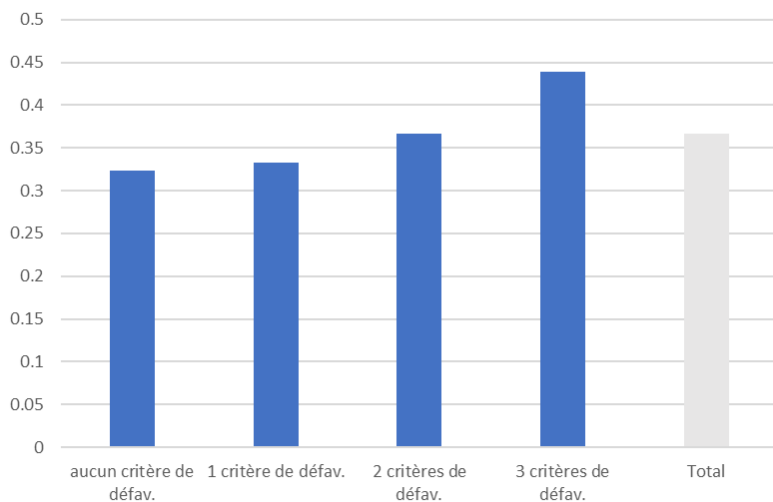


Exemple :
IRIS 1101 « Dupuy »
De la commune de
Toulouse, en Haute
Garonne 31555

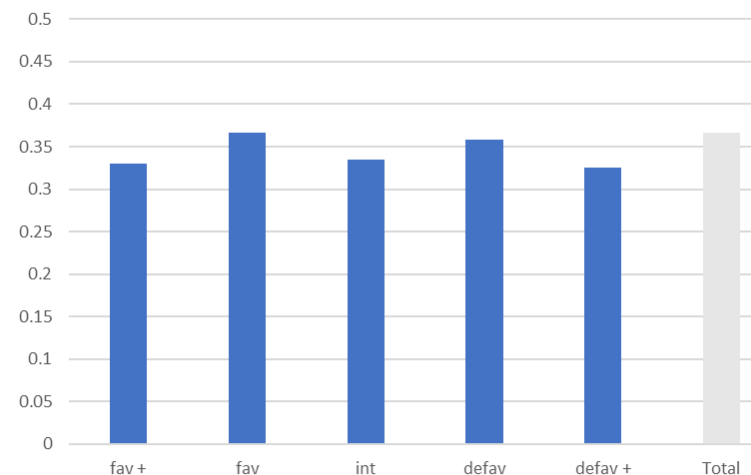


Probabilité de présenter un stade avancé au diagnostic en fonction de la défavorisation sociale chez les patientes atteintes de cancer du sein

Probabilité médiane de présenter un stade avancé au diagnostic parmi les patientes atteintes d'un cancer du sein (cas incidents)



mesure individuelle de la défavorisation sociale



mesure de la défavorisation sociale à l'IRIS

Source : Sébastien Lamy, chercheur INSERM, CERPOP



Ekitia : créer un cadre de confiance pour le
traitement collaboratif des données



EkitIA : un écosystème unique

70
MEMBRES

Entreprises

Logos of DAWEX (data exchange technology), sopra steria, CIVITEO, cleyrop, and orange.

Collectivités et administrations

Logos of La Région Occitanie (Pyrénées - Méditerranée), SiCoval (Communauté d'agglomération du Sud-Est Toulousain), RENNES MÉTROPOLE, and Montpellier Méditerranée métropole.

Pôles, clusters et associations

Logos of LA Mée, aerospace valley, AGROPOLIS INTERNATIONAL, PÔLE MER MÉDITERRANÉE, and Digital113.

Établissements publics ou chargés de mission publique

Logos of SNCF, ENEDIS (L'ELECTRICITE EN RESEAU), cnes, and IGN (INSTITUT NATIONAL DE L'INFORMATION GÉOGRAPHIQUE ET FORESTIÈRE).

Établissements d'enseignement et de recherche

Logos of Inserm (La science pour la santé - From science to health), UNIVERSITÉ DE MONTPELLIER, IRT SAINT EXUPERY, Université de Toulouse, UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3, and cnrs.

DIVERSITE DE SOURCES DE DONNEES (données météorologiques, données géographiques, données de mobilité, données énergétiques, données de santé ...)

DIVERSITE D'INFRASTRUCTURES (de stockage, de calcul ...)

Ekitia | Les piliers de la gouvernance des données



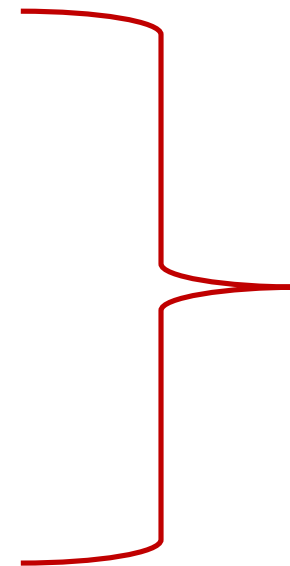
ETHIQUE ET JURIDIQUE



MODELES ECONOMIQUES



SOLUTIONS TECHNIQUES



Faciliter le
partage des
données pour
en permettre
des usages
innovants dans
un cadre de
confiance

Rendre l'éthique actionnable

ekitia | Pourquoi un Label ?

Raison d'être d'Ekitia → développer un cadre de confiance concernant l'usage des données



Tous les outils et services conçus par Ekitia ont pour objectif de promouvoir et de renforcer l'effectivité des principes consacrés au sein de la **Charte éthique des usages des données**



Charte éthique des usages des données



Des principes organisés en 6 thématiques...

- Science des données et société
- Science des données et individu
- Qualité des données et sécurité du système d'information
- Transparence
- Gouvernance dans un cadre de confiance
- Réciprocité

... co-construits et réexaminés régulièrement avec diverses parties prenantes (experts et citoyens)

Document accessible en intégralité via ce lien : <https://www.ekitia.fr/la-charte-ethique-ekitia/>

ekitia | Périmètre du Label Ekitia



- Le Label peut être demandé par **toute structure publique ou privée**, quel que soit son secteur d'activité
- Pour une **solution numérique (site web, plateforme, application, objet connecté, ...)** issue d'un projet porté par cette structure
- Une fois octroyé, le Label est valable pendant **3 ans**

ekitia | Ils nous ont fait confiance



Les technologies d'accroissement de la
confidentialité



Des solutions techniques : les Technologies d'Accroissement de la Confidentialité

1 - Datasets disclosed but not in original form

Anonymisation/ Hashing	Obfuscate personal data into data that cannot identify any particular individual
Differential Privacy	Add "noise" to dataset and still allow statistical analysis to run on it
Synthetic Data Generation	Create artificial data with the same statistical characteristics as the original dataset
Homomorphic Encryption	Enable simple computations to be run on encrypted data, only outputs can be decrypted

2 – No disclosure of data at all, code comes to data

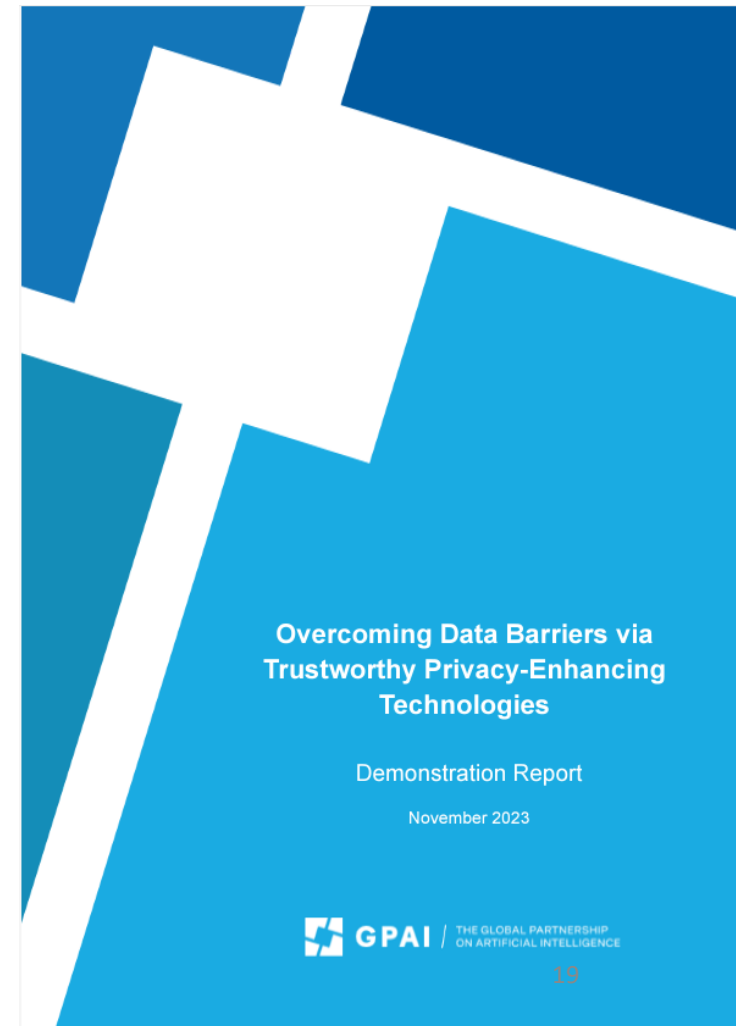
Federated Learning	Train ML models at multiple decentralized devices/servers by exchanging model-weights instead of data
Multiparty Computing	Cryptographic technique that enables computation on distributed nodes
Trusted Execution Environment	Isolated physical or virtual environment that guarantees code and data loaded inside remains protected

Source : PET-enabled Pandemic Model, Final Report, IMDA | MOH | DTC, Singapour, octobre 2023

ekitia | Avantages et inconvénients de l'utilisation des TAC

- **Avantages** : les Technologies d'Accroissement de la Confidentialité permettent d'obtenir des informations sans divulguer les données sensibles d'origine.
- **Limites** :
 - Certaines TAC ont une précision insuffisante
 - Certaines TAC ont un coût de calcul important (chiffrement homomorphe)

➔ il est nécessaire de bien choisir la technologie adaptée à la situation.



Vers des espaces communs
de données multi-sectoriels



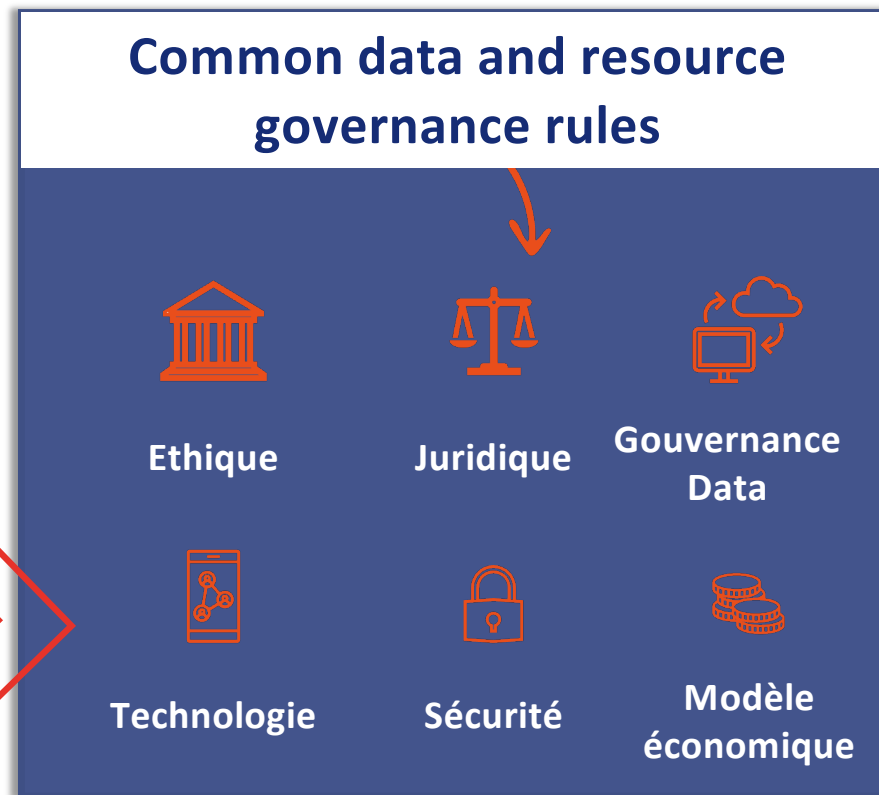
Assembler les pièces du puzzle : les espaces communs de données

Espace commun de données

Une organisation autour de règles de gouvernance, de pratiques et de normes qui permet **d'orchestrer** des ressources numériques entre plusieurs acteurs.

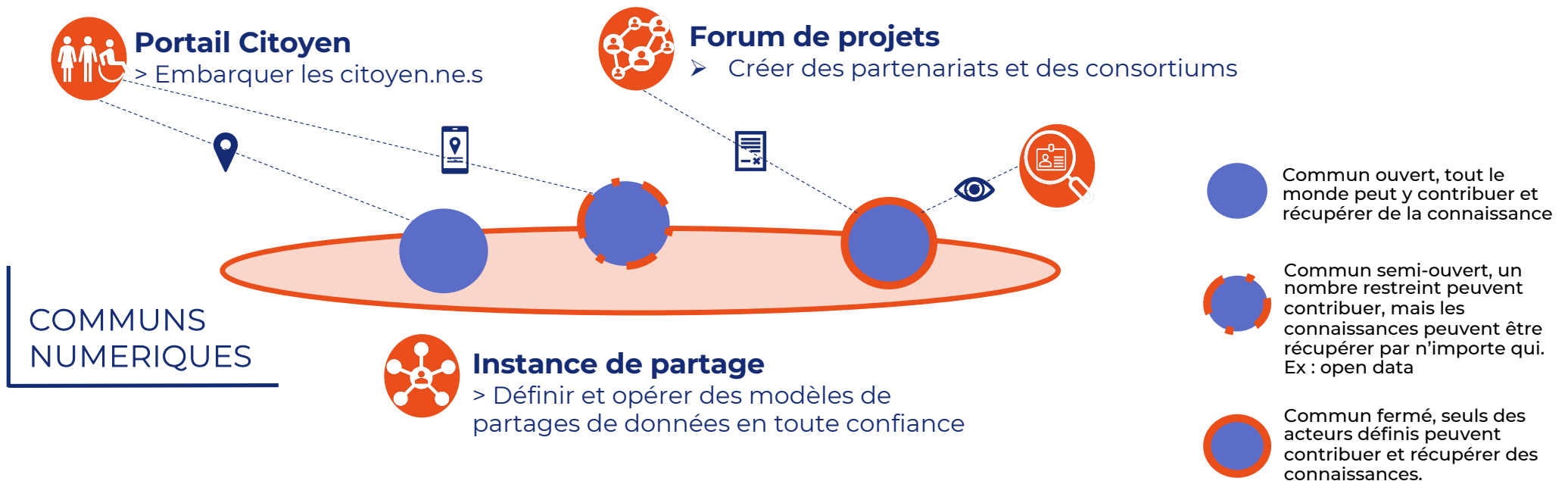
Méthode du livre de règles :

Méthode de co-construction des règles générales de fonctionnement et d'utilisation de l'espace de données par les futurs utilisateurs



Ekitia | Vers un Data Space Territorial Multi-Sectoriel

Projet associant plus de 30 organisations membres d'Ekitia.
Mise en œuvre : 2025



Conclusion



De nouvelles opportunités pour le traitement collaboratif des données

- Les problématiques de respect de la confidentialité ne peuvent se résoudre par des techniques traditionnelles comme l'anonymisation
- Le besoin d'opérer des traitements de données multi-organisations est constant
- De nouvelles technologies permettent d'adopter de nouvelles approches de la problématique du traitement collaboratif de données
- La technologie seule ne résout rien : c'est une approche intégrative qui est nécessaire, associant les aspects éthiques, juridiques, économiques, technologiques
- Les projets d'espaces communs de données offrent de nouvelles opportunités pour le traitement collaboratif des données