



# LIVRET DU CITOYEN

## LA CHARTE ETHIQUE DES USAGES DES DONNÉES D'EKITIA

Version mise à jour en 03/22



Data et IA en pleine confiance

# LA CHARTE

---

## **PRINCIPE 1** *SCIENCE DES DONNÉES ET SOCIÉTÉ*

### **1.1. Bienfaisance**

Les données doivent être utilisées en vue de contribuer aux 17 Objectifs de Développement Durable définis par les Nations Unies à l'horizon 2030. Leur mise à disposition pour la réalisation de missions d'intérêt public devrait également être favorisée.

### **1.2. Innovation soutenable**

Les données doivent, dans la mesure du possible être collectées, stockées et analysées dans des conditions respectueuses de l'environnement.

### **1.3. Solidarité, diversité et non-discrimination**

L'usage des données ne doit en aucun cas aboutir à créer ou à renforcer une discrimination ou une stigmatisation à l'encontre de certains individus.

### **1.4. Facteur humain**

L'usage des données s'inscrit en toutes circonstances dans une logique de préservation de l'autonomie décisionnelle humaine et de responsabilité humaine.

## **PRINCIPE 2** *SCIENCE DES DONNÉES ET INDIVIDU*

### **2.1. Respect et renforcement de l'autonomie individuelle**

Les personnes dont les données sont utilisées sont toujours dûment informées de cette utilisation, de leurs droits sur ces données et de la finalité de cette utilisation. Sont par ailleurs encouragées les innovations facilitant l'exercice de ces droits par les personnes concernées.



## 2.2. Respect de la vie privée

Les usages des données à caractère personnel sont conformes au cadre juridique français applicable en la matière. Cependant, la possibilité d'anonymiser de telles données en amont de leur utilisation devrait toujours être envisagée.

## PRINCIPE **3** QUALITÉ DES DONNÉES ET SÉCURITÉ DES SYSTÈMES D'INFORMATION

### 3.1. Qualité des données

Les données devraient être exactes, pertinentes et représentatives par rapport à la finalité de leur utilisation ainsi que réutilisables

### 3.2. Sécurité des données et des centres de données

Les données devraient être hébergées dans des centres de données situés sur le territoire de l'Union Européenne et respectueux de normes de sécurité adaptées pour l'usage envisagé.

### 3.3. Robustesse des algorithmes

Les résultats issus d'une analyse de données réalisée à l'aide d'un algorithme devraient toujours être fiables à un haut degré de certitude.

## PRINCIPE **4** TRANSPARENCE

### 4.1. Information claire et accessible

Les porteurs de projets basés sur l'usage de données devraient fournir une information claire et accessible aux utilisateurs finaux sur les éléments suivants : la finalité du projet, les bénéficiaires et les risques du projet, les données utilisées, leur lieu de stockage et la durée de leur conservation.



## 4.2. Explicabilité des algorithmes

Dans le même esprit, la logique des méthodes utilisées pour analyser des données devrait être documentée en vue de pouvoir expliquer de manière intelligible aux utilisateurs finaux le résultat obtenu. La marge d'erreur affectant ce résultat devrait toujours être communiquée.

## 4.3. Auditabilité

Toutes les étapes d'un projet basé sur l'usage de données (ex. sélection des données, sélection des modèles algorithmiques utilisés pour traiter ces données, résultats des évaluations des risques effectuées ...) sont documentées en vue de pouvoir soumettre le projet à un audit.

# PRINCIPE **5** GOUVERNANCE DES DONNÉES DANS UN CADRE DE CONFIANCE

## 5.1. Déclinaison de la Charte en règles

La Charte est détaillée en des critères plus précis permettant sa mise en œuvre concrète dans le cadre de projets numériques. De plus, ces critères constituent le référentiel du Label EKITIA visant à valoriser les projets numériques en conformité avec les présents principes.

## 5.2. Apprentissage collectif

Lorsqu'un projet basé sur l'usage de données implique plusieurs partenaires, ceux-ci devraient partager leurs bonnes pratiques techniques, juridiques et économiques en matière d'usage des données.

## 5.3. Evaluation des bénéfices et des risques

Il est recommandé d'évaluer les bénéfices et les risques en amont de tout projet basé sur un usage de données, en priorisant les

risques sur l'environnement, sur l'organisation de la société et sur les droits individuels. Seuls les projets présentant plus de bénéfices que de risques devraient être mis en œuvre.

#### 5.4. Inclusion des citoyens et des utilisateurs finaux

Les utilisateurs finaux d'outils développés grâce à l'usage de données devraient être impliqués dans la réalisation du projet, par exemple à travers des consultations ou des ateliers de co-construction ... Cela est particulièrement pertinent pour les projets contribuant à la réalisation d'une mission d'intérêt public.

#### 5.5. Intégrité

Toute personne intervenant dans un projet basé sur l'usage de données respecte les règles déontologiques auxquelles elle est soumise, et agit dans un esprit d'intégrité intellectuelle et de coopération.

## PRINCIPE **6** RECIPROCITE

### 6.1. Reconnaissance

Lorsqu'un projet basé sur l'usage de données implique plusieurs partenaires, le chef de file reconnaît explicitement, voire publiquement (dans la limite de la confidentialité et des accords entre les collaborateurs) la contribution de chacun à la réalisation du projet.

### 6.2. Répartition équitable de la création de valeur

La valeur générée grâce à un projet basé sur l'usage de données devrait être équitablement répartie entre tous les acteurs ayant contribué à sa réalisation.

## PRINCIPE **7** ETHIQUE ET URGENCE

En temps de crise sanitaire, environnementale ou sécuritaire, les présents principes constituent un cadre de référence. Toutefois, si les principes liés au respect de la vie privée et à la transparence sont indérogeables, les autres principes peuvent être aménagés afin de répondre efficacement à des problématiques urgentes.

## PRINCIPE **8** REEXAMEN REGULIER DE LA CHARTE

La Charte est un instrument dynamique : elle est régulièrement réexaminée en vue d'être en cohérence avec les évolutions technologiques, légales et éthiques, et avec les attentes des parties prenantes.

# LEXIQUE TECHNOLOGIQUE

## Donnée (numérique)

Toute information pouvant être enregistrée en format numérique.

## Données à caractère personnel

Il s'agit de données dont le traitement est encadré par la loi pour garantir aux personnes une protection adéquate de leur vie privée (que ces données reflètent). Précisément, il s'agit de toute information permettant d'identifier une personne physique directement (nom, prénom, photo, empreinte digitale ...) ou indirectement (n° de sécurité sociale, adresse IP ...)

## Données sensibles

Il s'agit d'une catégorie de données personnelles faisant l'objet d'une protection spécifique : leur traitement est interdit par principe. Précisément, il s'agit de toute information relative à l'origine raciale ou ethnique, aux opinions politiques, aux convictions philosophiques ou religieuses, à l'adhésion syndicale, à la génétique, à la biométrie, à la santé, à l'orientation ou à la vie sexuelle d'une personne.

## Données non personnelles

On en distingue deux sortes, d'une part les données qui sont par nature non rattachables à une personne physique et, d'autre part, les données personnelles qui ont fait l'objet d'une anonymisation. Ces données peuvent librement circuler en Union Européenne (sous réserve que cette circulation ne porte pas atteinte à l'intérêt public d'un État membre)<sup>2</sup>.

<sup>2</sup> Règlement (UE) 2018/1807 du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne.

## Lots de données

Il s'agit d'un ensemble de données formant un « tout » cohérent. Les lots de données sont souvent présentés sous une forme de tableau (les colonnes contiennent les variables (ex. âge) et les lignes les valeurs de ces variables (ex. 18 ans)).

## Cycle de vie d'une donnée

- **Collecte de données** : Tout processus dont l'objectif est de rassembler des données.
- **Stockage de données** : Les données sont stockées dans des centres de données permettant de sécuriser leur accès. La plupart de ces centres sont gérés par des prestataires externes.
- **Traitement de données** : Toute opération permettant d'extraire de l'information à partir des données collectées, quel que soit le procédé utilisé. Il peut notamment s'agir d'analyser des données.

On distingue généralement trois types d'analyses :

**Analyse descriptive** : Elle consiste simplement à décrire le contenu d'un ensemble de données.

**Analyse prédictive** : Elle consiste à prévoir, à l'aide d'une mesure de probabilité, ce qui pourrait survenir dans le futur. Dans le domaine du Big Data, les systèmes d'apprentissage automatique (cf. ci-dessous) sont très utilisés pour faire des prédictions de ce type.

**Analyse prescriptive** : Elle reprend le principe de l'analyse prédictive en y ajoutant la notion d'action. C'est-à-dire que l'analyse va permettre d'identifier les cibles sur lesquelles agir.

- **Archivage de données** : L'archivage consiste à exporter les données inutilisées du lieu où elles sont stockées vers un autre support sécurisé afin de libérer de l'espace de stockage.
- **Partage de données** : Le partage de données consiste à fournir à des tiers intéressés un accès à ses propres données.





## Big Data

Littéralement ces termes renvoient à l'expression de « données massives ». En pratique, le Big Data correspond à la production, par différentes sources, de grands volumes de données. Si ces grands volumes de données permettent de faire des analyses à une échelle nettement supérieure à celle que permettent des lots de données plus petits, la contrepartie est que leur traitement nécessite d'importants moyens techniques (ce qui provoque le retard, voire l'exclusion, des petites structures sur le marché du Big Data).



## Algorithme

Un algorithme est une suite d'instructions en nombre fini permettant d'obtenir un résultat souhaité (l'exemple le plus basique étant celui de la recette de cuisine). Afin de pouvoir être exécutée par un logiciel, cette suite d'instructions est traduite en langage informatique (le langage de programmation). Lorsque les instructions ainsi programmées visent à reproduire une fonction cognitive humaine, on dit du logiciel qu'il est un « système d'intelligence artificielle ». On en distingue principalement deux catégories :

### Systèmes experts (déterministes)

Ils résolvent des problèmes en s'appuyant sur une base de connaissances relatives à un domaine spécifique. Ces connaissances correspondent en fait à toutes les informations relatives à un sujet donné, et sur lesquelles, traditionnellement, un expert humain se fonde pour établir son expertise. Cet ensemble de règles forme le modèle sur la base duquel, à la manière d'un expert du domaine traité, le système peut raisonner.

### Systèmes d'apprentissage automatique (probabilistes)

Ils résolvent des problèmes de manière plus ou moins autonome, en s'appuyant sur un modèle qu'ils auto-génèrent en analysant des données existantes. En d'autres termes, il s'agit de systèmes qui « apprennent » un modèle de raisonnement à partir de données existantes : une phase d'apprentissage est nécessaire, afin que le système affine et optimise son modèle de raisonnement, au fur et à mesure qu'il est confronté à de nouveaux cas.



## Intelligence Artificielle (IA)

Elle sera ici appréhendée en tant que composante des systèmes algorithmiques permettant d'analyser les données.

## Science des données

Il s'agit d'une évolution de la démarche statistique classique, qui consistait à se poser une question, à formuler une hypothèse, à planifier une expérience, à recueillir des données et à estimer les paramètres d'un modèle pour in fine prendre une décision. Avec la science des données les hypothèses sont directement formulées à partir des données, et non plus théoriquement, ce qui constitue un changement de paradigme majeur.

## Économie des données

Les données numériques que nous émettons tous lors de nos activités quotidiennes créent de nouvelles opportunités économiques liées à leur exploitation. Toutefois, dans le cadre d'Ekitia les usages ne se limitent pas à l'économie au sens strict : ils peuvent aussi concerner des innovations sociales, par exemple.



**Ekitia** (ex Occitanie Data) est une association de préfiguration d'un pôle d'économie de la donnée, en cours de transformation en un GIP national. Structurée en Occitanie mais rassemblant des acteurs au-delà, ses membres sont des entreprises privées et publiques, des collectivités, des acteurs académiques, des clusters et pôles de compétitivité. Ekitia propose un **cadre de confiance, éthique et souverain**, destiné à permettre aux acteurs de partager et de croiser leurs données tout en respectant les intérêts des individus et des propriétaires des données.

Pour construire ce cadre, une première version de la **Charte éthique des usages des données** a été publiée en avril 2020. Il ne s'agit toutefois pas d'une version définitive : elle sera enrichie par le fruit d'ateliers de co-construction que nous avons initié, en collaboration avec la **Chaire UNESCO «Ethique, Sciences et Société»**, au cours de l'été 2020.

En effet, après avoir réalisé 6 ateliers avec des experts métier sectoriels (mobilité, santé, énergie, environnement, agriculture, emploi et formation) et pris en considération leurs retours, nous ambitionnons d'organiser des ateliers de co-construction citoyenne, afin de confronter les principes de la Charte à des regards neutres. La prise en compte de ces précieux retours nous a amené à réaliser une deuxième version de notre Charte éthique publiée en mars 2022.



Pour toute information complémentaire ou pour nous  
contacter, retrouvez-nous sur notre site internet :



[www.ekitia.fr](http://www.ekitia.fr)



Ekitia



@Ekitia