



# GUIDE POUR L'OUVERTURE DES DONNÉES PUBLIQUES EN TUNISIE



---

**Direction de la publication :** Khaled Sellami, DG de l'Unité de l'administration Electronique, Présidence du Gouvernement

**Coordination :** Emilie Becle et Asma Sabri, Expertise France

**Rédacteurs :** Kamel Malleh et Charfeddine Yaakoubi

**Vérification :** Rim Garnaoui et Sonia Gharbi, Unité de l'administration Electronique, Présidence du Gouvernement

**Design :** CreaWorld.tn

**Remerciements :** Stéphane Trainel et Sasun Saugy, Ministère de l'économie, des finances et de la relance, Secrétariat General, France

Ce guide a été réalisé dans le cadre du projet PAGOF d'expertise France et financé par l'Agence Française de Développement (AFD)

**Janvier 2022 ©**

<b>1. CADRE CONCEPTUEL DE L'OPEN DATA</b> .....	<b>4</b>
1.1. QU'EST-CE QU'UNE DONNÉE ? .....	5
1.2. QU'EST-CE QU'UNE INFORMATION ? .....	5
1.3. QU'EST-CE QUE LA CONNAISSANCE ? .....	6
1.4. QU'EST-CE QUE L'OPEN DATA ? .....	6
1.5. CARACTÉRISTIQUES DE L'OPEN DATA .....	7
1.6. QUELS SONT LES PRINCIPES DE L'OUVERTURE DES DONNÉES ? .....	8
1.7. QUELS SONT LES NIVEAUX D'OUVERTURE DES DONNÉES ? .....	9
1.8. QUELS SONT LES AVANTAGES DE L'OPEN DATA ? .....	10
<b>2. APERÇU SUR L'OPEN DATA EN TUNISIE</b> .....	<b>13</b>
2.1. LE CADRE STRATÉGIQUE .....	14
2.2. LE CADRE LÉGAL .....	16
2.3. LE CADRE ORGANISATIONNEL .....	18
2.4. LE CADRE TECHNOLOGIQUE .....	22
2.5. LES CAS D'USAGE DES DONNÉES OUVERTES .....	24
<b>3. PROCESSUS D'OUVERTURE DES DONNÉES</b> .....	<b>27</b>
3.1. LA PRÉPARATION DES JEUX DE DONNÉES À OUVRIR .....	28
3.2. LES PLATEFORMES OPEN DATA .....	45
3.3. LE PLAN D'ACTION ANNUEL DES DONNÉES OUVERTES .....	49
3.4. LE RAPPORT D'ACTIVITÉ ANNUEL DES DONNÉES OUVERTES .....	53
3.5. L'INVENTAIRE DES DONNÉES PUBLIQUES .....	54
3.6. INSTAURER UNE DÉMARCHE BASÉE SUR LA RÉUTILISATION .....	56

# INTRODUCTION

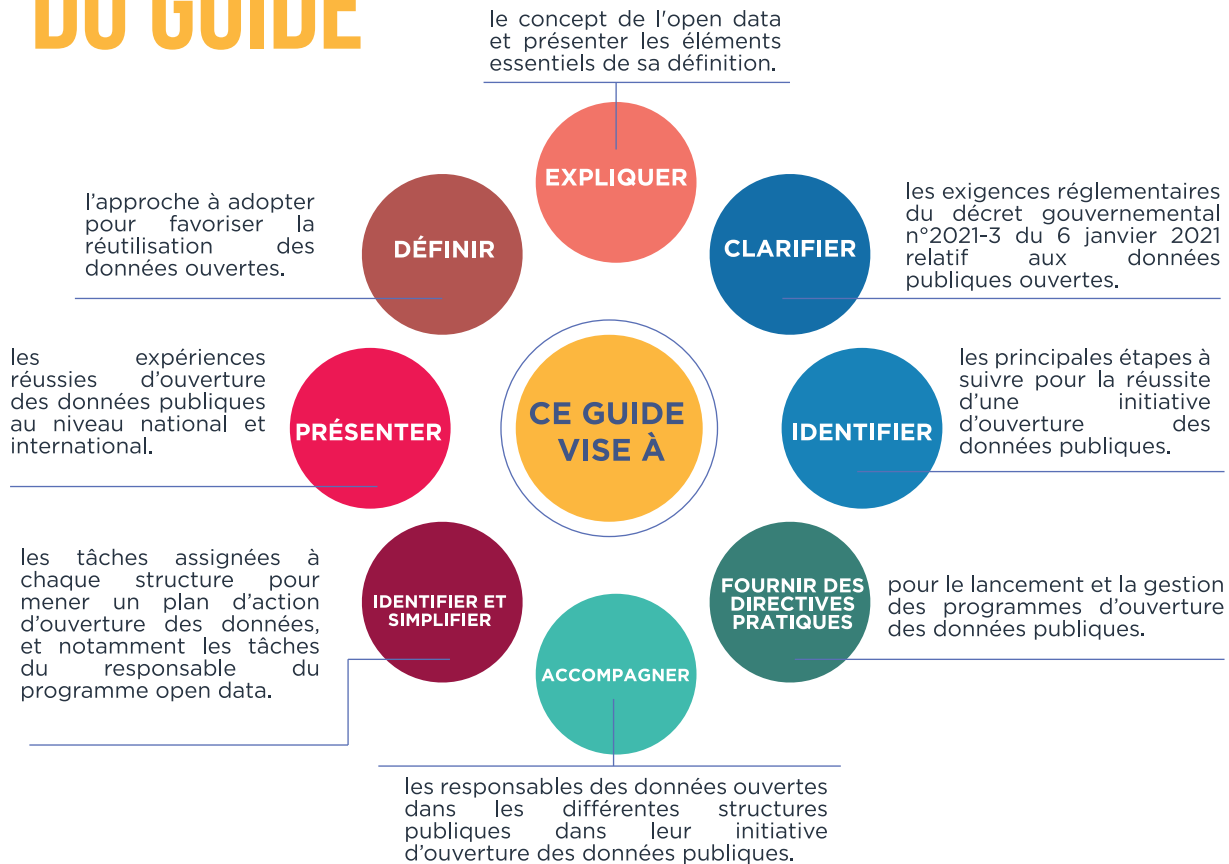
Le concept d'open data, né en 1995 dans le milieu de la recherche scientifique, s'est rapidement répandu pour intégrer la sphère de la gestion publique. À partir de 2009, les initiatives d'ouverture de données ont commencé à s'étendre dans divers pays tels que les États-Unis, le Royaume-Uni, le Canada qui ont annoncé de nouvelles initiatives visant à ouvrir leurs informations et données publiques.

La Tunisie a rejoint ce mouvement depuis 2012 avec le lancement du portail national des données ouvertes qui s'inscrit dans le cadre de sa nouvelle politique axée sur la transparence de l'action publique et l'OpenGov. L'adhésion au Partenariat pour un Gouvernement ouvert (PGO) en 2014 et l'adoption de la stratégie Smart Gov 2020 ont permis de propulser l'ouverture des données au cœur des engagements de la modernisation de l'administration tunisienne.

Avec la promulgation du décret gouvernemental n°2021-3 du 6 janvier 2021 relatif aux données publiques ouvertes, la Tunisie a franchi une nouvelle étape lui permettant de concrétiser sa vision de l'open data et sa stratégie OpenGov. Cette stratégie représente également un enjeu réel pour pouvoir passer de simples initiatives sectorielles à une politique open data plus affirmée instaurant ainsi le principe d'ouverture des données par défaut.

Cette nouvelle approche implique des changements culturels et institutionnels puisque ce ne sont pas seulement les données qui s'ouvrent avec l'open data, mais aussi les entités publiques et leurs pratiques.

# LES OBJECTIFS DU GUIDE



# PORTÉE DU GUIDE

Ce guide comporte 3 chapitres

1

Le Cadre  
conceptuel de  
l'Open data



2

Aperçu sur  
l'Open data en  
Tunisie



3

Processus  
d'ouverture  
des données



# 1. CADRE CONCEPTUEL DE L'OPEN DATA

# 1. CADRE CONCEPTUEL DE L'OPEN DATA

## 1.1. QU'EST-CE QU'UNE DONNÉE ?

Une donnée est une information dans sa forme la plus élémentaire, c'est « une description élémentaire d'une réalité »<sup>1</sup>.

Les données sont le résultat de mesure ou d'observation. Elles ne sont ni structurées, ni contextualisées, ni commentées. Elles représentent ainsi un fait brut, non traité et non interprété.

## 1.2. QU'EST-CE QU'UNE INFORMATION ?

C'est un niveau plus élevé qui repose sur le traitement des données dans le but de créer une information intelligible et utile.

Ainsi, l'information correspond à un ensemble de données contextualisées et structurées suite à un processus de traitement de données (quantitatives ou qualitatives)

Quelle est la différence entre une donnée et une information ?

L'information se base sur une interprétation de la donnée. C'est le produit de la source « donnée ».

<sup>1</sup> <https://bit.ly/3t2MP4W>

### EXEMPLES



Les données météo



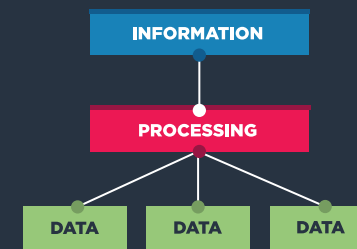
Les horaires des transports



Les données des élections



Les cartes routières



### EXEMPLES



Température moyenne

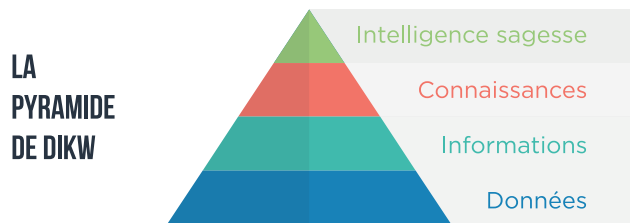


Taux de remplissage des barrages



## 1.3. QU'EST-CE QUE LA CONNAISSANCE ?

La connaissance est le résultat du traitement de l'information permettant de l'organiser, la synthétiser et/ou la résumer pour la rendre plus facilement compréhensible et l'adopter en tant que base dans la prise de décision. Quand on ajoute une signification aux informations cela devient des connaissances.



## 1.4. QU'EST-CE QUE L'OPEN DATA ?

L'open data consiste à rendre accessible au grand public des données produites par les institutions publiques ou le secteur privé, en permettant leur réutilisation libre et gratuite.

Selon cette définition, une donnée ouverte est une donnée :



**Accessible**



**Exploitable**



**Réutilisable  
par tous**



**Libre de droits**

EXEMPLE

**Fumer  
tue**

### LA DÉFINITION LÉGALE DES DONNÉES OUVERTES

Les données ouvertes doivent être libres de droits. Par conséquent, il ne doit y avoir aucune restriction quant à leur utilisation et redistribution.

N'importe quel utilisateur devrait être libre d'utiliser, de modifier, de combiner et de partager les données, même à des fins commerciales.

## 1.5. CARACTÉRISTIQUES DE L'OPEN DATA



### LICENCE OUVERTE

Les données doivent être publiées sous une licence légale qui permet à quiconque d'y accéder, de les utiliser et de les partager.



### ACCÈS OUVERT

Les données doivent être accessibles à tous sans aucune barrière technique ou financière.

Exp: la publication de données en format (PDF) rend le traitement difficile pour les utilisateurs.



### FORMAT OUVERT

La publication de données dans un format électronique exploitable par un ordinateur (directement ou via une API).

Ces caractéristiques sont confirmées dans l'article 3 du décret gouvernemental n°2021-3 du 6 janvier 2021 relatif aux données publiques ouvertes, qui définit les données publiques ouvertes comme étant

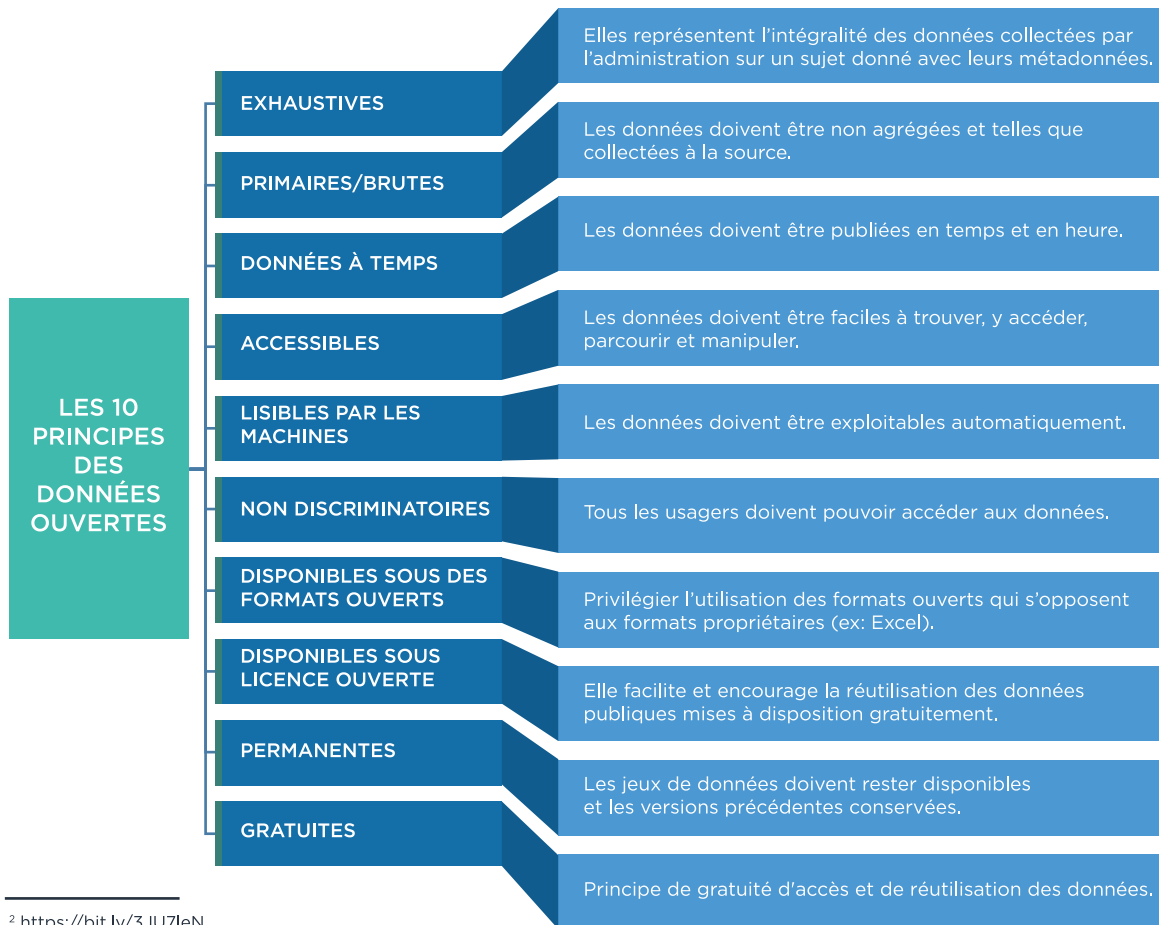


### ARTICLE 3

**« données publiques librement accessibles et réutilisables par le public, disponibles sous un format ouvert, lisibles par machines et non soumises à des restrictions liées aux droits d'autrui dont la protection de sa vie privée, ses données personnelles, sa propriété intellectuelle et d'autres droits couverts par une protection légale. »**

## 1.6. QUELS SONT LES PRINCIPES DE L'OUVERTURE DES DONNÉES ?

Les 10 grands principes de l'ouverture des données selon la Sunlight foundation<sup>2</sup> se résument comme suit:

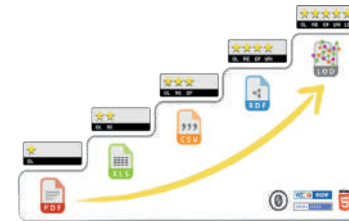


<sup>2</sup> <https://bit.ly/3JU7leN>

## 1.7. QUELS SONT LES NIVEAUX D'OUVERTURE DES DONNÉES ?

Pour mesurer les niveaux d'ouverture des données de point de vue technique et légal, on utilise généralement l'échelle « 5 étoiles » de Tim Berners-Lee, principal inventeur du Web et instigateur des LinkedData.

Cette échelle se base sur une évaluation de 5 niveaux afin de mesurer le potentiel d'exploitation des données et leur interconnexion.



**La première étoile :** elle est octroyée à toutes données pouvant être qualifiées comme ouvertes, c'est-à-dire celles dont le contenu est disponible sous licence ouverte et ce sous n'importe quelle forme comme un fichier PDF, JPEG. Toutefois, la réutilisation des données peut s'avérer difficile.



**La 2<sup>ème</sup> étoile :** elle est octroyée aux données réutilisables facilement, toutefois le format de publication est un format propriétaire ou fermé, comme Excel.



**La 3<sup>ème</sup> étoile :** les données doivent être disponibles sous un format structuré et lisible par une machine qui n'est pas limitée à un pack logiciel en particulier. Exemple le format CSV, JSON, XML...

*On privilégiera l'utilisation du format CSV pour le partage de données ouvertes puisque c'est un format facile à comprendre, hautement réutilisable et lisible par des machines.*



**La 4<sup>ème</sup> étoile :** elle requiert la création et l'utilisation d'identifiants URI sur le Web pour toutes les données.

*Le format privilégié pour ce niveau d'ouverture est le RDF.*



**La 5<sup>ème</sup> étoile** introduit le concept d'interconnexion des données entre elles pour y ajouter du contexte, c'est le concept du Web de données.

## 1.8. QUELS SONT LES AVANTAGES DE L'OPEN DATA ?

L'open data représente une opportunité à plusieurs niveaux : économique, politique, scientifique, social...

### 1.8.1. AU NIVEAU POLITIQUE

L'open data:

- Favorise la transparence et la redevabilité.
- Vise à revitaliser la démocratie en encourageant la participation citoyenne et à mieux impliquer les citoyens dans les débats publics.
- Permet de connecter les citoyens et de créer un dialogue plus éclairé sur les politiques publiques et les réformes à entreprendre.

### 1.8.2. AU NIVEAU ÉCONOMIQUE

L'open data:

- Contribue à la croissance économique en offrant des opportunités aux entreprises quelles que soient leurs tailles.
- Crée des emplois.
- Favorise l'innovation.
- Crée de nouveaux services à haute valeur ajoutée.
- Améliore l'efficacité des entreprises et identifie des sources d'économies.



## EXEMPLES D'IMPACT ÉCONOMIQUE

- ✓ L'ouverture des données gouvernementales peut aider à débloquer entre 3 000 à 5 000 milliards de dollars américains par an pour l'économie mondiale dans sept secteurs (énergie, transport, santé, éducation...)<sup>3</sup>
- ✓ Entre 2016 et 2020, la taille du marché des données ouvertes devrait augmenter de 37% pour atteindre une valeur de 75,7 milliards d'euros en 2020.<sup>4</sup>
- ✓ Entre 2016 et 2020, 25 000 emplois directs dans le domaine de l'Open Data seront créés.<sup>5</sup>
- ✓ La réduction des coûts de l'administration de €1.7 milliards dans les 28 pays de l'UE d'ici 2020.<sup>6</sup>
- ✓ Une étude publiée en 2015 a révélé que les entreprises britanniques utilisant des données ouvertes ont enregistré un chiffre d'affaires annuel réuni de plus de £92 milliards<sup>7</sup> et emploient plus de 500 000 personnes<sup>8</sup>.
- ✓ En Finlande, les PME qui utilisent l'open data ont grandi à un taux de 15% plus vite que les autres PME<sup>9</sup>.

<sup>3</sup> Manyika, James and others (2013). Open data: Unlocking innovation and performance with liquid information. McKinsey Global Institute.

<sup>4</sup> Commission Européenne « Creating Value through Open Data » Nov 2015

<sup>5</sup> Commission Européenne « Creating Value through Open Data » Nov 2015

<sup>6</sup> Commission Européenne « Creating Value through Open Data » Nov 2015

<sup>7</sup> <https://theodi.org/article/open-data-means-business/>

<sup>8</sup> Commission Européenne « Creating Value through Open Data » Nov 2015

<sup>9</sup> <https://bit.ly/3sp7LnU>

### 1.8.3. AU NIVEAU ADMINISTRATIF

#### L'open data:

- Constitue un catalyseur pour la transformation numérique de l'administration publique.
- Représente une source clé pour améliorer le processus d'élaboration des politiques publiques.
- Aide à l'amélioration de la qualité des services publics et contribue à l'efficacité de l'administration.
- Permet de valoriser le rendement de l'administration en montrant au public la qualité des services rendus.
- Promeut l'instauration de nouvelles possibilités de croisement et d'exploitation des données internes à l'administration.



### 1.8.4. AU NIVEAU DE L'INNOVATION ET DE LA RECHERCHE SCIENTIFIQUE

#### L'open data:

- Stimule l'innovation et la recherche scientifique et ce en franchissant les obstacles qui entravent l'accès, l'usage et le partage des données.
- Aide à sauver des vies grâce au développement de la recherche scientifique et à mieux agir en temps de crises, de catastrophes et d'épidémies (Exemple de l'épidémie du Covid 19)



## **2. APERÇU SUR L'OPEN DATA EN TUNISIE**



## 2. APERÇU SUR L'OPEN DATA EN TUNISIE

### 2.1. LE CADRE STRATÉGIQUE

#### 2.1.1. AU NIVEAU POLITIQUE

Dans le cadre du projet d'Appui à la mise en œuvre du e-Government et de l'Open Government, la Tunisie a adopté une stratégie et une feuille de route du programme «Smart Gov 2020».

L'une des huit (8) orientations stratégiques de la mission du programme consiste à ***la mise en place d'un cadre d'ouverture des données favorisant la transparence et la réutilisation des données publiques.***

Cette orientation s'articule autour des initiatives suivantes :

**INTRODUIRE LE PRINCIPE DE DONNÉES PUBLIQUES OUVERTES PAR DÉFAUT.**

**METTRE EN PLACE DES PROCÉDURES POUR L'INVENTAIRE, LA COLLECTE, LA PUBLICATION ET LA QUALITÉ DES DONNÉES.**

**ADOPTER UNE APPROCHE SECTORIELLE POUR PROMOUVOIR L'UTILISATION ET LE DÉVELOPPEMENT DES DONNÉES OUVERTES EN TUNISIE.**

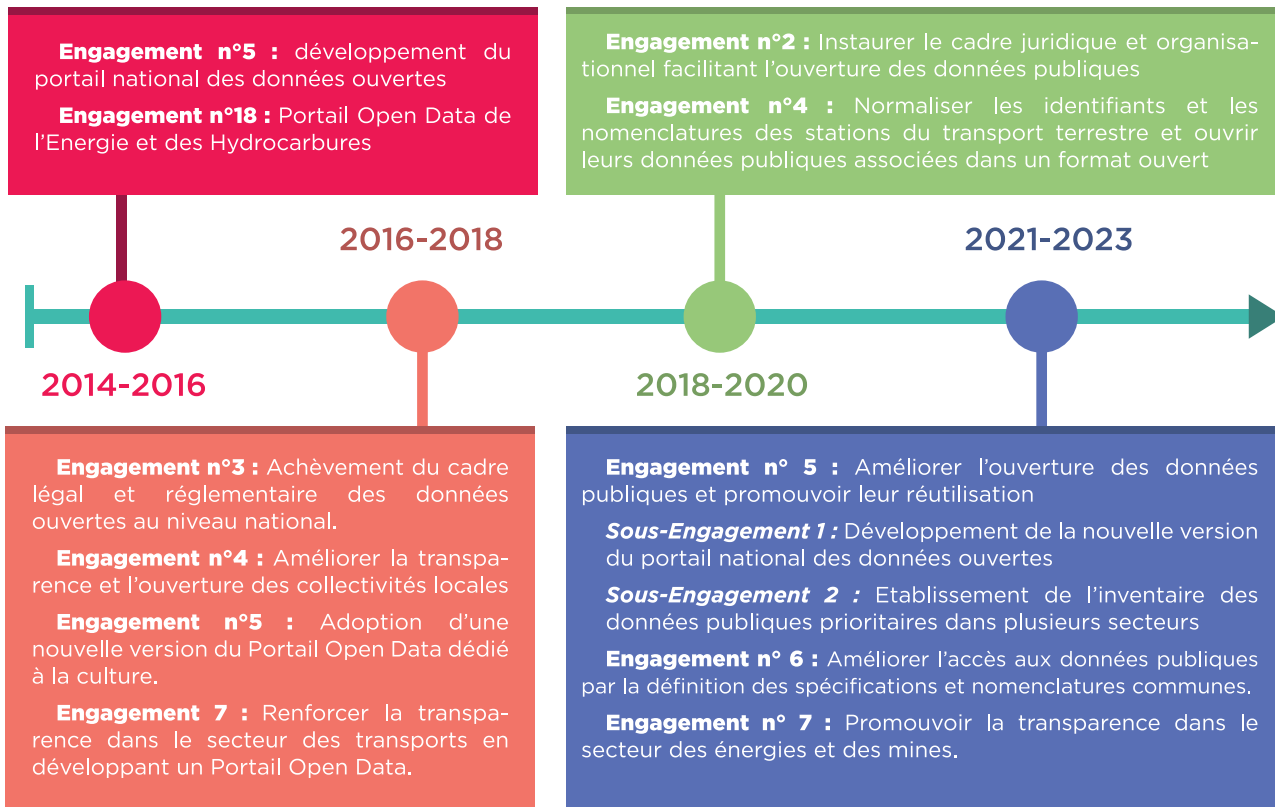
Suite à cette orientation générale, 7 projets structurants ont été définis :

1. Mise en place d'un cadre institutionnel pour les données ouvertes.
2. Élaboration des procédures et standards pour les données ouvertes.
3. Inventaire et industrialisation de la publication des données ouvertes.
4. Programme de communication et formation portant sur les données ouvertes.
5. Groupes de travail thématiques sur les données ouvertes.
6. Appui à la mise en œuvre de six projets pilotes thématiques.
7. Mise en place d'un registre national unique et ouvert des adresses.

L'avancement est notable quant à la réalisation de ces projets.

## 2.1.2. LE PARTENARIAT POUR UN GOUVERNEMENT OUVERT

La Tunisie a rejoint le partenariat pour un gouvernement ouvert en 2014. Depuis, elle a adopté 4 plans d'action qui ont tous comporté des engagements relatifs aux données publiques ouvertes.



## 2.1.3. LE RAPPORT D'ÉVALUATION DE L'ÉTAT DE PRÉPARATION DE LA RÉPUBLIQUE TUNISIENNE EN VUE D'UNE INITIATIVE D'OUVERTURE DES DONNÉES PUBLIQUES (2017)<sup>10</sup>

Ce rapport a permis de fournir un état des lieux en matière d'ouverture de données publiques en Tunisie et de proposer un plan d'action séquencées sur le court, le moyen et le long terme pour la mise en œuvre efficace d'un programme national de données ouvertes.

## 2.2. LE CADRE LÉGAL

### 2.2.1. LA CONSTITUTION DE 2014

La constitution tunisienne de 2014 a tenu à intégrer plusieurs dispositions relatives à la transparence dont l'open data est l'un des principaux leviers.

**L'article 15** « L'Administration publique est au service du citoyen et de l'intérêt général. Elle est organisée et agit conformément aux principes de neutralité, d'égalité et de continuité du service public, et conformément aux règles de transparence, d'intégrité, d'efficacité et de redevabilité. »

**L'article 32** « L'Etat garantit le droit à l'information et le droit d'accès à l'information. »

### 2.2.2. LA LOI RELATIVE AU DROIT D'ACCÈS À L'INFORMATION

Le droit à l'information s'est consolidé avec la loi organique n°2016-22 du 24 mars 2016 relative au droit d'accès à l'information qui a contribué à instaurer un socle juridique considéré comme un facteur déterminant et inhérent à la réussite des initiatives d'ouverture de données, en particulier à travers la diffusion proactive des données.

Cette loi dispose dans son article 6 que «Les organismes assujettis aux dispositions de la présente loi, sont tenus de publier, d'actualiser, de mettre périodiquement à la disposition du public, **dans une forme utilisable**, les informations suivantes... ».

Ainsi, toutes les structures soumises à cette la loi peuvent être considérées comme productrices des données publiques ouvertes (**environ 10 000 structures en Tunisie**).

<sup>10</sup> <http://www.ogptunisie.gov.tn/fr/wp-content/uploads/2021/08/Etude-Initiative-Open-Data-Tunisie.pdf>

## 2.2.3. LE DÉCRET GOUVERNEMENTAL N°2021-3 DU 6 JANVIER 2021 RELATIF AUX DONNÉES PUBLIQUES OUVERTES

C'est le texte réglementaire qui organise l'ouverture des données publiques (open data) en Tunisie et qui explicite les caractéristiques techniques et réglementaires à adopter.

Ce décret a pour but de :

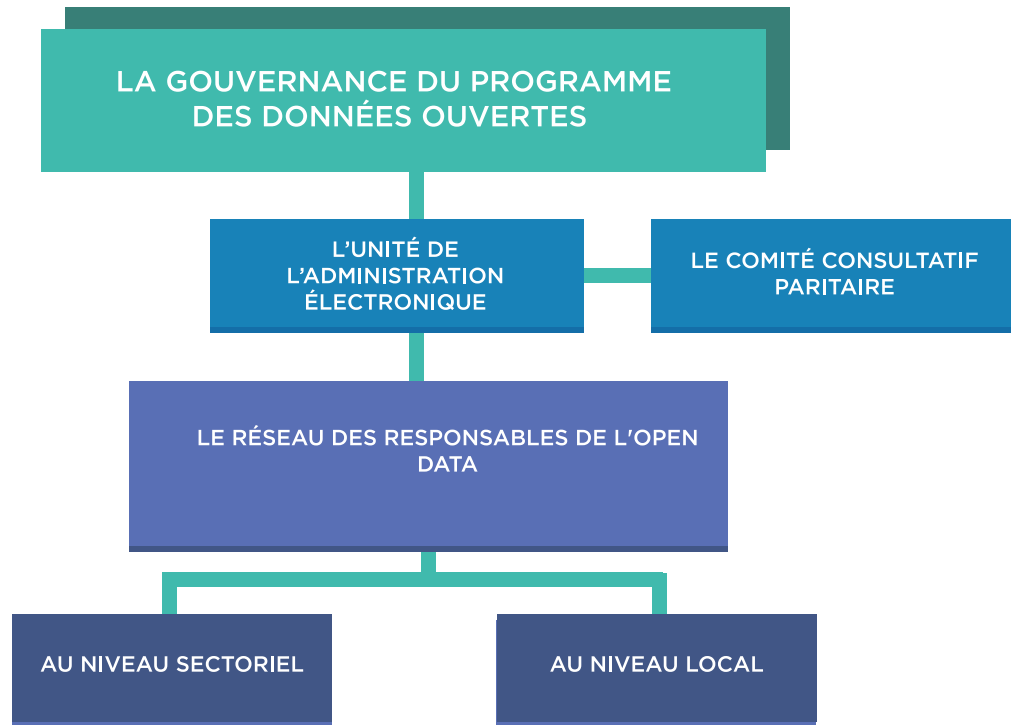
Promouvoir	Soutenir	Moderniser
les principes de transparence et de reddition des comptes.	la participation du public à l'élaboration des politiques publiques ainsi qu'au suivi de leur mise en œuvre et de leur évaluation.	l'administration et améliorer la qualité et l'efficacité des services publics.
<b>Contribuer</b>	à la mise en place d'un cadre approprié pour favoriser le développement économique et créer des opportunités d'emploi supplémentaires, notamment en stimulant la création des start-ups qui œuvrent pour le développement de nouveaux cas d'usages innovants en exploitant des données publiques.	

Ce décret comporte 5 chapitres :

- 1<sup>er</sup> chapitre** : relatif aux dispositions générales qui exposent les objectifs de ce décret et les définitions des concepts en relation avec l'open data.
- 2<sup>ème</sup> chapitre** : relatif à la gouvernance du programme des données publiques ouvertes.
- 3<sup>ème</sup> chapitre** : relatif à la publication des données.
- 4<sup>ème</sup> chapitre** : relatif aux métadonnées.
- 5<sup>ème</sup> chapitre** : relatif à la qualité des jeux de données.

## 2.3. LE CADRE ORGANISATIONNEL :

Selon le décret gouvernemental n° 2021-3, la gouvernance du programme des données publiques ouvertes s'articule autour de 3 intervenants :



## 2.3.1. L'UNITÉ DE L'ADMINISTRATION ÉLECTRONIQUE :

Créée à la Présidence du Gouvernement par le décret n°2005-1894 du 2005, l'unité de l'administration électronique est chargée notamment de :

### PRÉPARER

Le **plan d'action annuel**, publié sous un format ouvert, qui comprend les travaux et projets qui doivent être mis en œuvre au niveau national selon un calendrier précis

La **stratégie pour stimuler le développement de nouveaux cas d'usages** de données ouvertes et assurer sa pérennisation.

Le **rapport d'activité annuel** publié avant la fin du mois de mars de chaque année qui comprend un état sur les activités et les résultats obtenus au niveau national concernant l'ouverture des données publiques au cours de l'année écoulée.

### ORGANISER ET COORDONNER

Les **travaux du réseau** des responsables de l'open data.

Le **support technique** pour la mise en œuvre du programme open data avec les différents organismes publics.

### ASSISTER ET ADMINISTRER

Assister les **structures publiques à développer et mettre en œuvre leur propre plan d'action annuel** sur les données ouvertes. Ce plan comprend une liste des activités relatives aux données ouvertes que chaque organisme public prévoit de mettre en œuvre l'année suivante.

**Administrer le portail national** des données ouvertes.

### VALORISER

**Evaluer** les résultats obtenus au niveau national concernant l'ouverture des données.

**Promouvoir** la politique nationale d'ouverture des données

**Stimuler le développement de nouveaux cas d'usages.**

**Communiquer** sur les résultats réalisés au niveau de l'ouverture des données.



## 2.3.2. LE COMITÉ CONSULTATIF PARITAIRE :

Il est créé auprès du ministre chargé de la modernisation administrative, chargé du suivi de la mise en œuvre du programme de données ouvertes.

La composition et le mode de travail du comité sont déterminés par arrêté du ministre chargé de la modernisation administrative.

La nomination de ses membres et sa composition seront renouvelées périodiquement tous les trois ans.

### LES TÂCHES DU COMITÉ CONSULTATIF PARITAIRE

Donner un avis consultatif sur les travaux réalisés par l'unité de l'administration électronique.

La promotion de l'Open Data au sein de l'administration et de la société civile dans son ensemble.



## 2.3.3. LE RÉSEAU DES RESPONSABLES DU PROGRAMME DES DONNÉES OUVERTES

Composé des responsables de l'open data dans les différents organismes publics. L'Unité de l'administration électronique est chargée de la coordination des travaux de ce réseau.

La structure de l'open data au niveau sectoriel ou local est basée essentiellement sur le responsable du programme des données ouvertes.

Cependant il est possible de créer des groupes consultatifs au niveau de chaque secteur pour coordonner les efforts.

Chaque organisme public est tenu de désigner un responsable du programme des données ouvertes.

## La Désignation :

Il est désigné par le premier responsable de chaque organisme public. Le chargé d'accès à l'information peut être désigné en cette fonction à condition qu'il dispose des compétences requises à exercer les tâches assignées.

## Les Tâches :

**Participer** aux travaux du réseau des responsables de l'open data.

**Suivre** la mise en œuvre des résultats inclus dans les procès-verbaux des réunions du réseau des responsables de l'open data et présenter tous les documents y afférents à l'approbation du responsable de l'organisme.

**Préparer** un plan d'action annuel portant sur les activités des données ouvertes de l'organisme.

**Coordonner** la publication des données sous format ouvert et sur le portail adopté par l'organisme public.

**Coordonner** la préparation et la mise à jour de l'inventaire de données de l'organisme et le publier sur le portail des données publiques ouvertes qu'il utilise, tout en donnant la possibilité à l'utilisateur d'inclure ses commentaires.

**Elaborer** et mettre à jour le plan de suivi et d'évaluation du programme de données publiques ouvertes au niveau de l'organisme.

**Préparer** un programme de formation spécifique aux données publiques ouvertes au profit des cadres et des agents publics de l'organisme concerné.

**Promouvoir** le programme des données publiques ouvertes au sein de l'organisme et l'activer au niveau du secteur concerné.

**Préparer** le rapport annuel sur les activités des données publiques ouvertes, le publier sur le site internet de l'organisme concerné et le transmettre à l'Unité.



## 2.4. LE CADRE TECHNOLOGIQUE :

### 2.4.1. UN PORTAIL NATIONAL INTERCONNECTÉ :

Depuis le lancement des premières initiatives d'ouverture des données, la Tunisie a adopté un modèle qui se base sur **un portail fédéré des données ouvertes au niveau national interconnecté à plusieurs portails sectoriels** des ministères précurseurs en open data.

Cette approche stipule que les fichiers de données et/ou les services d'API sont gérés par les entités productrices des données, tandis que les métadonnées restent fournies au catalogue central/national pour offrir des capacités d'accès et de recherche au niveau national et pour devenir le point d'accès unique à toutes les données publiques ouvertes.

La diversité technologique des plateformes existantes, couplée à un degré d'interopérabilité assez élevé, représente un atout incontournable de l'initiative d'ouverture des données en Tunisie. Toutefois, cette multitude technologique exige une bonne maîtrise de solutions technologiques afin d'assurer une interconnexion continue avec la plateforme nationale.

### 2.4.2. DES INITIATIVES CITOYENNES D'OUVERTURE DES DONNÉES PUBLIQUES

L'initiative tunisienne se caractérise par le fait que **l'ouverture des données ne se limite plus aux seuls acteurs publics mais elle inclue plusieurs initiatives citoyennes lancées par des représentants de la société civile** à l'instar de : openBALADIATI, data4Tunisia et openGEOdata.

L'interconnexion entre ces initiatives est rendue possible par la standardisation des processus de production et de publication des données et ce en se basant sur les standards DCAT et schema.org qui visent à normaliser la description des catalogues ainsi que les jeux de données ouvertes.

### 2.4.3. UNE OFFRE VARIÉE DES DONNÉES OUVERTES

En Tunisie, l'offre des données ouvertes est très variée:

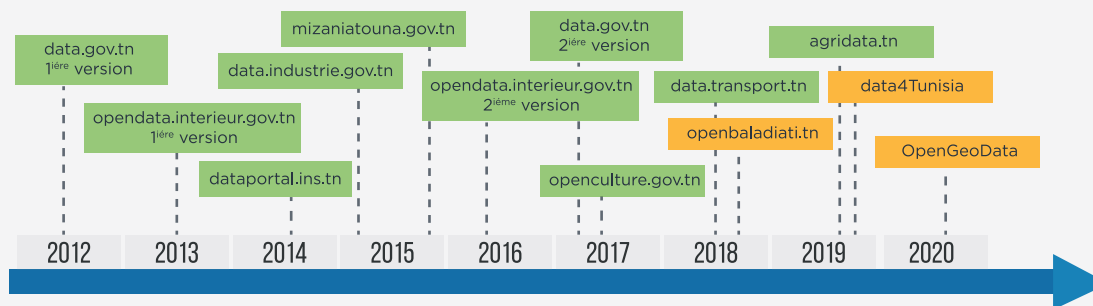
- On y trouve des données géographiques, des données de transport en utilisant le GTFS statique et le GTFS real-time,
- On y trouve des données accessibles en mode API et celles disponibles en téléchargement des fichiers CSV, Excel etc.....

Toutefois, l'ouverture n'a pas dépassé le 3<sup>ème</sup> niveau sur l'échelle de 5 étoiles de « Tim berners Lee ».

## LISTE DES PORTAUX DE DONNÉES PUBLIQUES OUVERTES EN TUNISIE

	STRUCTURE	URL	DATE DE LANCEMENT
1	Portail national Open Data	<a href="http://www.data.gov.tn">http://www.data.gov.tn</a>	2012
2	Ministère de l'intérieur	<a href="http://opendata.interieur.gov.tn">http://opendata.interieur.gov.tn</a>	2013
3	Ministère des finances	<a href="http://www.mizaniatouna.gov.tn">http://www.mizaniatouna.gov.tn</a>	2015
4	INS (Institut National des Statistiques)	<a href="http://dataportal.ins.tn">http://dataportal.ins.tn</a>	2014
5	Ministère de l'industrie, des mines et de l'énergie	<a href="http://data.industrie.gov.tn">http://data.industrie.gov.tn</a>	2015
6	Ministère des affaires culturelles	<a href="http://www.openculture.gov.tn">http://www.openculture.gov.tn</a>	2017
7	Ministère des transports	<a href="http://data.transport.tn">http://data.transport.tn</a>	2018
8	Ministère de l'Agriculture, des ressources hydrauliques et de la pêche maritime	<a href="http://agridata.tn">http://agridata.tn</a>	2019
9	Municipalités	<a href="http://openbaladiati.tn">http://openbaladiati.tn</a>	2018

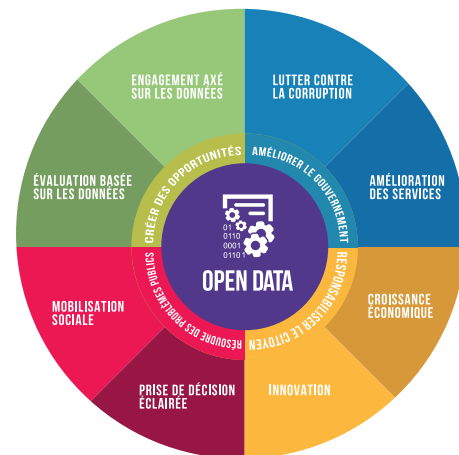
### LES PORTAUX OPEN DATA EN TUNISIE



## 2.5. LES CAS D'USAGE DES DONNÉES OUVERTES

Les cas d'usage consistent en la réutilisation de données publiques ouvertes par des tiers à d'autres fins que celles de la mission de service public pour laquelle les données ont été produites ou reçues.

Le succès d'un programme d'ouverture des données dépend fortement du niveau d'interaction avec toutes les parties prenantes, en particulier la société civile et le secteur privé, soit à travers la consultation des données, soit à travers le développement d'applications ou l'organisation des hackathons et autres événements open data.



### 2.5.1. LES CAS D'USAGE AU NIVEAU NATIONAL

En Tunisie, plusieurs initiatives de réutilisation des données ouvertes ont pu être exécutées à savoir :

- **L'étude sur le secteur des hydrocarbures en Tunisie**<sup>11</sup> a pu être réalisée suite à la publication des contrats pétroliers sur la plateforme open data du ministère de l'industrie et grâce aux données ouvertes sur la production des hydrocarbures actualisées quotidiennement.
- **La création de Dataviz** : Plusieurs dataviz ont pu être créés grâce à la publication des données ouvertes comme celle de la situation journalière des barrages publiée sur le portail agridata<sup>12</sup>, ou celle du budget municipal dans 31 municipalités publié sur la plateforme openbaladiat<sup>13</sup>.
- **L'organisation de Hackathons** : Les hackathons sont des événements durant lesquels des développeurs volontaires se réunissent pendant une période de temps donnée afin de travailler sur des projets de programmation informatique de manière collaborative en utilisant des données ouvertes.

On cite à cet égard l'« OpenGovDataHack2020 » autour de la réutilisation des données publiques ouvertes organisé par l'Unité de l'Administration Electronique, ou Onshorhack en 2019 organisé par l'association Onshor ou l'open geodata hackathon organisé par AGEOS et The Tunisian E-Governance Society.

<sup>11</sup> <https://bit.ly/3s5tjWy>

<sup>12</sup> <https://bit.ly/3H1FGYO>

<sup>13</sup> <https://bit.ly/3dWQaeO>

- **Des recettes financières supplémentaires** : la commune de Bouarada a pu recouvrer un montant de 85 000 dinars de dettes suite à la publication des données ouvertes sur les montants dus envers les concessionnaires du marché hebdomadaire de la municipalité , ce qui lui a permis de programmer de nouveaux projets.<sup>14</sup>

## 2.5.2. LES CAS D'USAGE AU NIVEAU INTERNATIONAL

### Dans le secteur du transport :



- En 2007, Transport for London TFL a décidé de publier un nombre important de jeux de données dans un format ouvert, gratuitement et en temps réel sur les horaires des transports publics.<sup>15</sup>

Une étude réalisée par Deloitte en 2017 a démontré qu'il existe désormais plus de 600 applications alimentées par ces données et utilisées par 42% des Londoniens. Ces applications ont permis de réaliser des économies de 70 à 95 millions de livres sterling par an en termes de gain de temps.

### Dans le secteur de la santé :



- L'épidémie d'Ebola a touché l'Afrique de l'Ouest en 2014. Au début, les informations sur les cas d'Ebola et les efforts de riposte étaient dispersés entre divers intervenants, ce qui a retardé et limité l'efficacité des interventions.

Trois intervenants - le Centre national de réponse à Ebola de la Sierra Leone (NERC), l'Échange de données humanitaires des Nations Unies (HDX) et Ebola GeoNode - ont considérablement amélioré la qualité et l'accessibilité des informations utilisées par les organisations humanitaires et les décideurs politiques afin d'améliorer les actions et de faire face à la crise<sup>16</sup>.

### Dans le secteur de l'éducation :



- <http://www.mejoratuescuela.org> : Il s'agit d'une plateforme citoyenne mexicaine qui fournit des informations sur les performances scolaires des différentes écoles au Mexique.

Elle aide les parents à choisir la meilleure option éducative pour leurs enfants, leur permet d'exiger une éducation de meilleure qualité et leur fournit des outils pour s'impliquer dans la scolarisation de leurs enfants. Elle met également à la disposition des citoyens des données pour identifier les foyers de corruption dans le domaine scolaire. Les données disponibles sur le site ont été utilisées dans un rapport qui a révélé une corruption généralisée dans le système éducatif mexicain.

<sup>14</sup> <https://bit.ly/3sc4ZSS>

<sup>15</sup> <https://bit.ly/3DZtWDj>

<sup>16</sup> <https://bit.ly/3IR1ruY>

### **3. PROCESSUS D'OUVERTURE DES DONNÉES**

# 3. PROCESSUS D'OUVERTURE DES DONNÉES

## 3.1. PRÉPARATION DES JEUX DE DONNÉES À OUVRIR

### 3.1.1. C'EST QUOI UN JEU DE DONNÉES ?

En open data, un jeu de données (dataset) est une collection cohérente de données. Il est composé de ressources qui représentent les entités porteuses de données et de métadonnées et qui décrivent et caractérisent ce jeux de données ainsi que ses ressources.

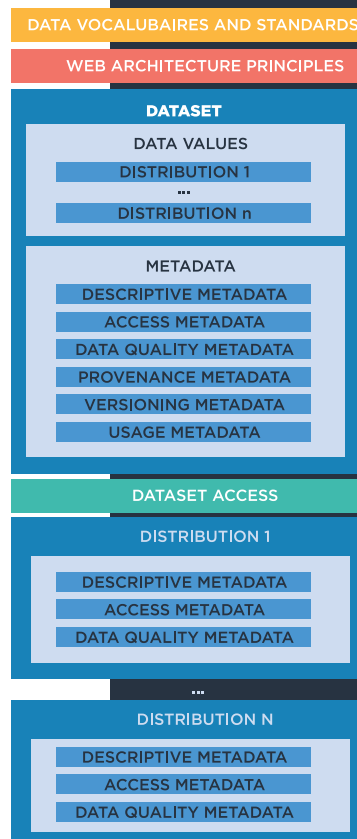
**Les ressources peuvent prendre l'une des formes suivantes : des fichiers de données, des liens sur le web, des points de terminaison d'APIs, etc...**

Les métadonnées permettent de fournir le contexte dont on a besoin pour exploiter les données ouvertes et ce en répondant aux questions suivantes : quoi ? qui ? où ? quand ? et pourquoi ?

**Généralement, le cycle de vie de métadonnées est plus grand que le cycle de vie des jeux de données. Les métadonnées peuvent être créées avant que les jeux de données ne soient créés.**

Un jeu de données peut avoir une ou plusieurs ressources et une ressource ne peut appartenir qu'à un seul jeu de données.

Pour chaque ressource, on trouve aussi des métadonnées pour décrire les données.



<sup>17</sup> <https://bit.ly/3vJQNIX>

Figure 1 Composition d'un jeu de données (dataValue et metaData) <sup>(17)</sup>

### 3.1.2. QUELLES MÉTADONNÉES POUR UN JEU DE DONNÉES ?

Il est recommandé d'utiliser des modèles standardisés de métadonnées favorisant l'interopérabilité, la découverte et la réutilisation des données à ouvrir à l'instar de :

- **DCAT** qui est un vocabulaire RDF visant à faciliter l'interopérabilité entre les catalogues de données .
- **Schema.org** par lequel on désigne un vocabulaire permettant d'ajouter des informations selon un format de micro-données et favorisant le référencement sur le web.



**Il est recommandé que les plateformes open data proposent l'intégration du protocole Opendgraph pour une meilleure intégration sur les réseaux sociaux.**

Généralement, les métadonnées doivent comprendre des éléments clefs qui permettent de définir le contexte et de fournir tous les renseignements de base concernant un jeu de données spécifique. Ces éléments clefs constituent le vocabulaire commun de métadonnées élaboré par l'UAE (article N° 16 du décret open data).

Les gestionnaires des plateformes open data (sectorielles, citoyennes, etc...) doivent utiliser ce vocabulaire commun pour référencer leurs jeux de données auprès du portail national et être en conformité avec le décret des données ouvertes.

Les acteurs publics disposant de modèles de métadonnées spécifiques (les données géographiques, les données de statistiques, ou autres), peuvent proposer leurs modèles sous forme d'un champ à renseigner dans le schéma commun. Ainsi, une fois ce champ est renseigné, il sera visible et accessible aux utilisateurs finaux comme étant une ressource spécifique au jeu de données.

#### LES CARACTÉRISTIQUES DES MÉTADONNÉES

##### OBLIGATOIRE

métadonnées obligatoires

##### FACULTATIF

métadonnées optionnelles

##### IMPLICITE

métadonnées générées automatiquement

##### BILINGUE

métadonnées en plusieurs langues

## Les métadonnées nécessaires à la description d'un jeu de données :

Un titre	obligatoire / bilingue	<p>Le titre du jeu de données doit être précis et spécifique.</p> <p>Il est recommandé d'utiliser un vocabulaire compréhensible par les utilisateurs finaux.</p> <p><b>Mauvais titre :</b> Budget annuel</p> <p><b>Bon titre :</b> Budget annuel de la commune de Bouarada pour la période 2011 - 2021</p>
Une description	obligatoire / bilingue	<p>La description permet d'obtenir des informations détaillées sur le contenu et la structure des ressources publiées ainsi que sur le contexte de production et/ou de collecte des données, des hypothèses de travail, etc.</p> <p>Les descriptions répondent généralement aux questions suivantes :</p> <ul style="list-style-type: none"><li>● Que contient le jeu de données ? Combien il y a de ressources ?</li><li>● À quoi sert le jeu de données ? Quelle est sa raison d'être ?</li><li>● Comment les données sont structurées ? À quoi correspondent les colonnes des ressources ? Existe-t-il un dictionnaire des données ?</li><li>● Comment lire le jeu de données ? Comment ouvrir les fichiers ?</li></ul> <p>Répondre à ces questions dans votre description permet aux utilisateurs de récupérer et de manipuler vos données facilement.</p>
Une licence	obligatoire	<p>Les licences définissent les règles de réutilisation des jeux de données publiés.</p> <p>En choisissant une licence de réutilisation, vous vous assurez que le jeu de données publié sera réutilisé selon les conditions d'usage que vous avez définies.</p>
Une fréquence de mise à jour	obligatoire	<p>La fréquence de mise à jour correspond à la fréquence à laquelle vous prévoyez mettre à jour les données publiées,</p>
Des mots-clefs	obligatoire / bilingue	<p>Les mots-clés caractérisent un jeu de données. Généralement, on ne dépasse pas 5 mots-clefs par jeu de données avec possibilité d'introduire des mots composés.</p> <p>Les mots-clefs apportent un meilleur référencement et une meilleure classification du jeu de données. À partir de chaque mot-clé, on peut obtenir la liste des jeux de données pour lesquels le mot clé a été également assigné.</p>



Une couverture temporelle	facultatif	La couverture temporelle indique la portée dans le temps des données publiées.
Une couverture spatiale	facultatif	La granularité spatiale indique le niveau de détail géographique le plus fin que peuvent couvrir vos données.
Une date de création (et/ou une date de première publication)	implicite	Permet de savoir à quelle date le jeu de données a été créé sur la plateforme. On peut même distinguer entre la date de création et celle de la première publication sur la plateforme.
Une date de dernière mise à jour	implicite	Permet de savoir à quelle date, le jeu de données a été actualisé sur la plateforme. C'est une information nécessaire mais pas suffisante. Ce qui nécessite de prévoir une somme de contrôle (checksum).
Un producteur	implicite	Permet d'indiquer le producteur du jeu de données (structure publique/organisme)
Un point de contact	facultatif	Permet d'indiquer la personne à contacter pour en savoir plus sur le jeu de données ou bien pour proposer des améliorations, etc...
Un lien SLUG	implicite	C'est la version « machine » du titre du jeu de données. Le slug change à chaque fois que le titre du jeu de données change.
Un lien Permanent	implicite	Lien avec un identifiant permanent immuable pendant toute la durée de vie d'un même jeu de données. C'est la manière la plus fiable et la plus pérenne pour partager <b>le lien d'un jeu de données</b> .

## Les métadonnées nécessaires à la description d'une ressource :

Un titre	obligatoire / bilingue	Le titre du jeu de données doit être précis et spécifique. Il est recommandé d'utiliser un vocabulaire compréhensible par les utilisateurs finaux.
Une description	obligatoire / bilingue	La description permet d'obtenir des informations détaillées sur le contenu et la structure de la ressource, le contexte de production et/ou de collecte des données, les hypothèses de travail, les unités utilisées... etc.
Une URL permanente	implicite	Un lien qui ne change pas pendant tout le cycle de vie d'une ressource.
Une URL vers le conteneur des données de la ressource	obligatoire	un lien direct vers le conteneur des données de la ressource. Ce lien peut changer à l'actualisation de la ressource (fichier, lien, API...).
Une taille en bytes	Implicite / facultative	Permet d'indiquer la taille de la ressource une fois uploadée sur le serveur.
Un format	Implicite	JSON, XML, CSV, API, WEB, etc...
Un type mime	Implicite / facultative	Permet d'indiquer le type de fichier associé à la ressource.
Une somme de contrôle	Implicite	Une somme de contrôle pour les métadonnées de la ressource et une autre facultative pour les données permettant d'identifier les changements.

### 3.1.3. QUELLES LICENCES CHOISIR ?

Sans licence, une fois publiées, les données ouvertes sont certes accessibles mais il se peut que les utilisateurs n'aient ni l'autorisation de les utiliser ni de les partager. Ainsi, les utilisateurs se trouvent dans une zone de flou juridique et par conséquent, les données ne sont plus considérées comme véritablement ouvertes.

#### Ce que dit le décret open data :

Le décret open data dans son article n°11 a instauré une limitation des licences utilisables par les acteurs publics afin de publier des jeux de données en open data. Ainsi, pour chaque jeu de données, le producteur de données doit utiliser une licence qui doit prendre en compte les caractéristiques suivantes :

- ✓ L'obligation de référencer le producteur de données ainsi que la date des données utilisées par les utilisateurs des données.
- ✓ La non-révocabilité du droit de réutilisation des données.
- ✓ La possibilité d'utiliser les données à des fins commerciales.

#### Les types de licences

Face aux multiples licences existantes, les acteurs publics peuvent se sentir un peu désorientés par les obligations et les droits accordés aux réutilisateurs, qui ne sont certes pas toujours les mêmes.

Pour la majorité des cas, il est recommandé d'utiliser en premier lieu la licence qui sera adoptée par l'U.A.E, une fois publiée. Pour les cas spécifiques, vous pouvez utiliser l'une des licences suivantes.



## Les licences Creative Commons « CC »



Les licences Creative Commons constituent un ensemble de licences régissant les conditions de réutilisation et de distribution d'œuvres. On peut utiliser les 2 versions de licences ouvertes dans le domaine des données ouvertes.

Ces dernières peuvent s'appliquer à différents types d'œuvres.



### Paternité « Attribution » (CC-BY v4.0) :

L'œuvre peut être librement utilisée, à condition de l'attribuer à l'auteur. Cela ne signifie pas que l'auteur est en accord avec l'utilisation faite de ses œuvres.



### Paternité et Partage à l'identique ('Attribution and share-alike'): (CC-BY-SA v4.0)

Elle est applicable dans le cas où vous souhaiteriez partager à l'identique non seulement les données dérivées, mais aussi les œuvres produites à partir de ces données.

## L'Open Data base Licence « ODbL »

### ODbL

C'est une licence de style copyleft qui permet de copier, de modifier, de faire un usage commercial, sous trois conditions : citer la source ; redistribuer sous des conditions de partage identique les modifications ; maintenir ouverte techniquement la base de données que vous redistribuez, qu'elle soit modifiée ou non.

L'ODbL ne couvre que les bases de données. Son partage à l'identique se propage uniquement à des bases de données, mais jamais aux œuvres produites à partir de ces bases de données.

### QUE CHOISIR ALORS ?



Il est important de rester sur un principe simple et conforme au décret open data et à son esprit :

- ✓ Par défaut, associer à votre jeu de données la licence adoptée par l'UAE, sinon la licence CC-BY.
- ✓ ODbL pour les bases de données lorsqu'il y a un motif d'intérêt général démontré.
- ✓ CC-BY-SA là où un partage à l'identique serait aussi d'intérêt général démontré.

### 3.1.4. COMMENT IDENTIFIER UN JEU DE DONNÉES À OUVRIR ?

Théoriquement, plus il y a de données ouvertes, mieux c'est. Mais étant donné les ressources limitées des acteurs publics, il est important de prioriser l'ouverture des jeux de données. Ainsi, la différence entre le succès et l'échec d'une initiative d'ouverture de données réside dans l'ouverture des données à fort impact parmi celles produites ou collectées par une organisation.

Par ailleurs, on doit garder à l'esprit que la valeur des données réside dans leurs utilisations. En effet, personne n'a intérêt à ouvrir en priorité des données de moindre valeur, qui ne seront pas utilisées par la suite.

Afin de caractériser les données prioritaires, il est important de raisonner en fonction du point de vue des propriétaires de données (interne à l'organisation voire même inter départemental), mais aussi et surtout en fonction du point de vue des ré-utilisateurs.

#### Point de vue des producteurs des données

En se référant au point de vue du producteur des données, différents facteurs peuvent justifier l'ouverture d'une donnée. Plus précisément, un jeu de données peut être considéré de grande valeur lorsqu'il réunit un ou plusieurs critères mentionnés ci-dessous :

- Il contribue à plus de transparence.
- La publication (voire l'ouverture) est soumise à une obligation légale.
- L'existence d'une liaison directe avec la mission principale (objectifs stratégiques et/ou opérationnels) de l'acteur public pour plus de redevabilité.
- La capacité de l'optimisation des services et de la rationalisation des dépenses publiques.

Les éléments de réponse à ces questions peuvent se trouver dans les documents stratégiques et budgétaires de l'acteur public. Cependant, trouver des réponses à ces questions, ne doit pas être vu comme étant l'unique responsabilité de la personne en charge de l'ouverture des données au sein d'une organisation. Cette dernière doit mener un dialogue avec les fonctionnaires non seulement au sein de la même organisation mais aussi d'une manière interdépartementale au sein du gouvernement lui-même et ce en faveur de l'identification, du développement des cas d'usage internes et de la proposition des solutions aux problèmes identifiés.

De même, certaines priorités peuvent s'imposer brusquement, comme le démontre la crise du COVID-19 ou certaines catastrophes naturelles.

### Point de vue des réutilisateurs de données

La valeur d'un jeu de données dépend principalement de son utilisation et de son potentiel de réutilisation, ce qui peut effectivement conduire à la génération d'une haute valeur économique et/ou sociale directe ou indirecte. Ce potentiel dépend en premier lieu de la taille de la communauté des utilisateurs finaux ; et en second lieu de la nature des systèmes et/ou services qui vont consommer ce jeu de

données une fois publié générant ainsi la création de nouveaux services.

Toutefois, il est difficile de prévoir les cas de réutilisations possibles de données avant leur ouverture. Ainsi, les responsables des données ouvertes dans chaque organisation publique sont invités à identifier, dès le départ, les utilisateurs finaux (si ce n'est pas encore fait) et de mener un dialogue direct avec eux durant toute la démarche d'ouverture des données.

Pour conclure, l'identification des données prioritaires passe obligatoirement par l'instauration d'un dialogue qui doit engager toutes les parties prenantes. Ce dialogue doit se focaliser, entre autres, sur les activités liées au développement des cas de réutilisation internes et externes. i

Le cas d'usage du jeu de données de la situation journalière des barrages publié par le ministère de l'agriculture et plus précisément l'Observatoire Nationale de l'agriculture (ONAGRI), constitue un bon exemple d'un cas de réutilisation des données ouvertes. La visualisation proposée dans le cadre du programme ONSHOR a été adoptée par les services du ministère pour devenir un outil incontournable de suivi et de monitoring de la situation journalière des barrages en Tunisie.

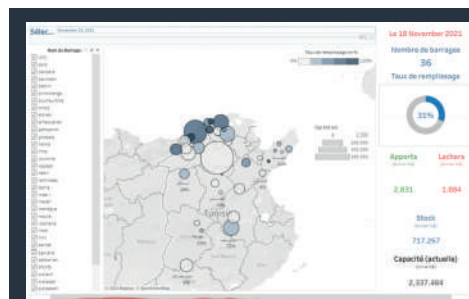


Figure 2 : Visualisation du jeu de données de la situation journalière des barrages

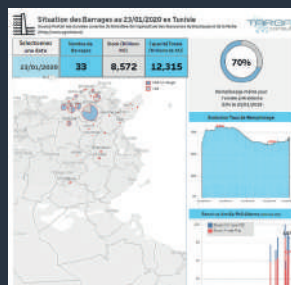


Figure 3: Visualisation développée par Traga Consult

« Targa Consult » était le premier cabinet de conseil spécialisé en intégration de systèmes d'information qui a proposé une première réutilisation de ce jeu de données sous forme d'une « data visualisation ».

Ce jeu de données reste de loin le premier jeu de données le plus visité sur la plateforme agridata.tn<sup>[18]</sup>.

<sup>18</sup> <https://bit.ly/3CsJEwr>

### 3.1.5. COMMENT PRÉPARER UN JEU DE DONNÉES POUR L'OUVERTURE ?

Le choix de la modalité de publication appropriée dépend de l'étendue des données ainsi que de leur fréquence de mise à jour. Les responsables open data dans chaque organisme public sont libres de choisir l'option qui leur convient. Cependant, certaines méthodes sont plus recommandées que d'autres et ce en fonction des caractéristiques de leurs données.

Vous êtes invités à contacter l'UAE qui vous aidera volontairement à définir le mode de publication le plus approprié.





- 1. approche** Généralement, si vous disposez d'un faible nombre de jeux de données avec une faible fréquence de mise à jour, la saisie manuelle directe sur le formulaire web de la plateforme nationale open data vous sera recommandée. Ce scénario sera également recommandé si le producteur des données ne dispose pas de systèmes d'information qui ne favorisent pas l'extraction et la mise à jour automatique.
- 2. approche** Si vous disposez de plus de 10 jeux de données avec une fréquence de mise à jour régulière, l'approche automatisée vous sera conseillée. Cette approche se base sur l'exploitation des APIs (d'écriture) offerts par la plateforme nationale ou bien par une autre plateforme sectorielle (à condition que cette dernière soit référencée sur la plateforme nationale).
- 3. approche** Dans le cas où le nombre de jeux de données serait important avec une actualisation fréquente, il serait recommandé d'étudier la mise en place d'une plateforme spécifique d'open data avec la mise en place de la technique de moissonnage des données vers la plateforme nationale.

i

Il est à noter que la première approche est celle qui sera toujours recommandée par l'UAE en premier lieu. En fonction de l'évolution du nombre de jeux de données, de leurs fréquences de mise à jour et de l'existence de systèmes d'information, l'organisme public peut donc à tout moment modifier le mode de publication.

### 3.1.6. ÉTAPES DE PRÉPARATION DES DONNÉES POUR L'OUVERTURE

Quel que soit le mode de publication choisi, les étapes de préparation ne changent pas. Elles consistent généralement à :

	<b>l'identification</b> des données à ouvrir (voir section suivante),
	<b>la préparation</b> des données à ouvrir (préparation manuelle ou extraction automatique sous forme d'un fichier CSV, JSON, XML, RDF, un lien sur la toile, un point de terminaison API) sous forme d'une ou de plusieurs ressources,
	<b>la création</b> du jeu de données sur la plateforme en remplissant le formulaire métadonnées associées sur la plateforme open data,
	<b>l'ajout</b> des ressources (des données) et leurs métadonnées associées.

À la base, un jeu de données est simplement une présentation structurée de données, telle qu'une feuille de calcul. Cependant, les données ne sont pas initialement structurées sous la forme recommandée. Cela implique parfois de transformer les données du format d'origine en un format différent.

Lorsque la source initiale du jeu de données émane d'un système d'information, il est nécessaire d'extraire et de convertir les données vers des formats facilitant leur utilisation. L'extraction et la transformation peuvent se faire aussi à partir des fichiers Excel ou autres sources numériques.

**Il est recommandé d'automatiser l'opération réalisée afin de faciliter la mise à jour des données publiées. Cette automatisation instaure un processus pérenne baptisé « processus d'extraction, transformation et publication : E.T.P. ».**

#### **Un jeu de données doit être enregistré dans un format ouvert !**

L'avantage de publier des données dans un format ouvert réside dans le fait qu'il sera plus facile pour quelqu'un d'autre d'analyser et de remixer ces données avec d'autres jeux de données.

*Etant donné que les données ne peuvent pas toujours être exprimées sous forme de feuille de calcul, il faut alors envisager d'autres formats. Il y a trois structures clés à connaître : tabulaire, hiérarchique et en réseau.*



### | Les données tabulaires :

C'est la structure la plus répandue. Les données sont organisées dans des colonnes et des lignes qui listent des valeurs (dans des cellules). Pour cette structure de données, il est recommandé d'opter pour le format CSV.



Les plateformes proposant une solution d'API performante garantissent la génération des fichiers CSV à partir de fichier Excel (à l'instar de la solution CKAN).

Pour cette structure, il est important de distinguer entre les deux formes « larges et longues ». Les deux tableaux ci-dessous représentent le même contenu et les mêmes valeurs selon deux formes différentes.

Il s'agit s'une forme large facilement compréhensible par l'être humain.

**Production de l'huile d'olive**

Region	2019	2020	2021
Nord	100 000	105 000	95 000
Centre	120 000	102 000	98 000
Sud	80 000	90 000	96 000



**Production de l'huile d'olive**

Region	Année	Production
Nord	2019	100 000
Centre	2019	120 000
Sud	2019	80 000
Nord	2020	105 000
Centre	2020	102 000
Sud	2020	90 000
Nord	2021	95 000
Centre	2021	98 000
Sud	2021	96 000

Ce tableau renvoie à la forme longue, préférée par les spécialistes de données (logiciels d'analyse et de visualisation de données).



Il est recommandé d'offrir ces deux formes pour les jeux de données de type tabulaire, celles-ci peuvent constituer deux ressources pour un même jeu de données.

Les langages comme python, R ... permettent facilement de basculer entre ces deux formes.

### | Les données hiérarchiques :

Si le jeu de données suit une structure où les points sont liés sous forme d'arborescence, alors dans ce cas, le format hiérarchique sera recommandé. Les données hiérarchiques exigent l'utilisation d'un format tel que JSON, XML et GeoJSON, KML, etc.... Il s'agit généralement des données géographiques, organigramme, etc...



### | Les données en réseau :

Les données structurées en réseau permettent d'établir des rapports entre n'importe quelle combinaison d'éléments dans n'importe quelle direction. Le format le plus utilisé dans ce cas est le RDF (Ressource Description Framework).



## Les données d'un jeu de données doivent être nettoyées !

Afin de favoriser le traitement automatisé des jeux de données publiés en open data, les données doivent être à la fois :

- **Primaires (raw)** : présentées dans le format le plus simple et élémentaire possibles,
- **Propres (clean)** : en utilisant un formatage de données uniforme et standardisé, sans entrées manquantes, avec des données dans chaque champ et en évitant au maximum les erreurs.

L'obtention des données brutes et propres peut être un défi suite à une conversion d'un fichier existant en un fichier à télécharger dans le cadre d'un jeu de données open data. Il est particulièrement judicieux de rechercher des éléments tels que des cellules fusionnées et des formules qui peuvent empêcher la lecture des données.

Date
01-déc-21
12 Novembre 2020
14/09/2019
15/01/2021



Date
2021-12-01
2020-11-12
2019-09-14
2021-01-15

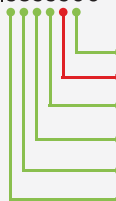
Le standard **ISO8601** pour les dates

« 2018-01-01T13:00:00Z »

« 2018-01-01T15:00:00+02:00 »

« 2018-01-01T15:00:00.15+02:00 »

**Latitude** : 46.5833300° **Longitude**: 0.3333300°



Précision à 11 cm

Précision à 1 m

Précision à 11 m

Précision à 110 m

Précision à 1 km

Précision à 10 km

Le système **WGS84** pour les coordonnées géographiques

« +37.5665, +126.9780 »

## L'encodage du fichier

Il est conseillé d'utiliser l'encodage UTF-8 qui permet d'encoder l'ensemble des caractères du répertoire universel des caractères codés (notamment les caractères contenant des accents ou les caractères spéciaux).

### L'en-tête des colonnes (pour le format tabulaire)

Dans un fichier tabulaire, la première ligne du fichier peut être utilisée pour nommer chaque colonne et donner des informations sur les données associées. Au lieu d'indiquer "Colonne n°2", il serait mieux de donner un nom de colonne explicite. Le nom des colonnes doit figurer sans majuscule, ni abréviations, ni accents, ni espaces (qui doivent être remplacés par « \_ ») afin de faciliter la manipulation des fichiers d'une manière automatisée.

### Gestion des champs non attribués

Il est possible qu'un champ ne soit pas attribué. Il convient alors de le laisser vide plutôt que d'y attribuer la valeur 0. Car le zéro correspond à une valeur qui peut changer l'interprétation des données.

## 3.1.7. QUELLES TECHNIQUES D'ANONYMISATION POUR UN JEU DE DONNÉES ?

### C'est quoi une donnée personnelle ?

Loi organique n° 2004-63 du 27 juillet 2004 portant sur la protection des données à caractère personnel

**Article 4** «on entend par données à caractère personnel toutes les informations quelle que soit leur origine ou leur forme et qui permettent directement ou indirectement à identifier une personne physique ou la rendent identifiable, à l'exception des informations liées à la vie publique ou considérées comme telles par la loi. »

**Article 5** « Est réputée identifiable, la personne physique susceptible d'être identifiée, directement ou indirectement, à travers plusieurs données ou symboles qui concernent notamment son identité, ses caractéristiques physiques, physiologiques, génétiques, psychologiques, sociales, économiques ou culturelles ».

### En quoi consiste l'anonymisation des données ?

L'anonymisation est une solution pour rendre exploitable les données personnelles dans le respect des droits et libertés des personnes. C'est un traitement qui consiste à utiliser un ou plusieurs techniques de manière à rendre impossible toute identification de la personne par n'importe quel moyen de manière irréversible<sup>19</sup>.

<sup>19</sup> <https://bit.ly/3sNNDMf>

## Que dit le décret Open Data :

L'anonymisation des données est la modification du contenu ou de la structure de ces données afin de rendre très difficile ou impossible l'identification des personnes ou des entités concernées.

Lorsque les données à ouvrir contiennent des données personnelles, elles doivent subir un traitement d'anonymisation avant leur publication voire leur ouverture.

Cependant, il existe trois hypothèses où la publication des jeux de données contenant des données personnelles peut avoir lieu sans anonymisation :

- dans le cas où « une disposition législative contraire » l'autoriserait ;
- lorsque les personnes intéressées sont consentantes ;
- si les données ont trait à la vie publique et sont considérées comme nécessaires à l'information du public.

## Des techniques d'anonymisation utiles mais insuffisantes:

<b>Suppression d'attributs</b>	La suppression d'attributs fait référence à la suppression des colonnes (attributs) dans les ressources d'un jeu de données (une « colonne » contenant une donnée personnelle).
<b>Masquage des caractères</b>	Le masquage des caractères est le changement de certains caractères en utilisant un symbole constant (par exemple « * » ou « x »). Le masquage est généralement partiel, c'est-à-dire appliqué uniquement à certains caractères de l'attribut (Ex : 7598654 devient XXXX654). Le changement peut se faire par un caractère vide (Ex : 7598654 devient 654), dans ce cas il est recommandé de préciser qu'on est en train d'afficher uniquement un certain nombre de caractères de l'attribut (qui correspond dans ce même cas à trois caractères).
<b>Généralisation</b>	À titre d'exemple : convertir l'âge d'une personne en tranche d'âge, ou convertir un emplacement précis en un emplacement moins précis. Cette technique engendre une réduction délibérée de la précision des données.
<b>Agrégation de données</b>	Il s'agit de convertir un ensemble de données à partir d'une liste d'enregistrements en valeurs résumées.
<b>Permutation des valeurs</b>	Le but de permutation est de réorganiser les données dans le jeu de données de telle sorte que les valeurs d'attribut individuelles restent toujours représentées dans le jeu de données sans pour autant correspondre aux enregistrements d'origine. Cette technique peut être partielle ou complète pour cibler tous les attributs.
<b>Perturbation des données</b>	Les valeurs d'une ressource de jeu de données d'origine sont modifiées pour être légèrement différentes (effet de bruit sur les données).

➡ En se limitant à ces techniques, il est souvent facile d'identifier les individus en liant les données à d'autres sources.

L'exemple le plus répandu est celui de la ré-identification en 1997 des données médicales du gouverneur du Massachusetts « William Weld ». En effet, LATANYA SWEENEY <sup>[20]</sup> a démontré qu'il est facile d'identifier des individus en combinant les données spécifiques aux patients **considérées comme étant anonymes** et en faisant une simple jointure avec les données des électeurs. Ainsi, ils sont parvenus à identifier que 6 personnes avaient la même date de naissance que le gouverneur, que seulement 3 d'entre eux étaient des hommes et que cette personne était la seule ayant un code postal à 5 chiffres.

### Comment concevoir un processus d'anonymisation ?

Une solution d'anonymisation doit être construite cas par cas et adaptée aux usages prévus.

Le groupe de l'Article 29 « G29 », qui regroupe les autorités de protection des données européennes, a publié en 2014<sup>[21]</sup> un avis sur les principales techniques d'anonymisation et ce afin d'expliquer comment les mettre en œuvre.

Il a également proposé la réunion des trois critères suivants :

- ✓ **l'individualisation** : est-il toujours possible d'isoler un individu ?
- ✓ **la corrélation** : est-il possible de relier des jeux de données distincts concernant un même individu ?
- ✓ **l'inférence** : peut-on déduire de l'information sur un individu ?

➡ Dans le cas contraire, une analyse détaillée des risques de ré-identification est à envisager

### Anonymisation par application des techniques de floutage : K-anonymity, L-diversity

L'application des techniques de floutage permet de garantir à la fois la non sensibilité et la persévérance de l'utilité des données à ouvrir.

Pour appliquer les techniques de K-anonymity , L-diversity, on doit commencer par identifier la classification des attributs.

<sup>20</sup> <https://bit.ly/35CeEcG>

<sup>21</sup> <https://bit.ly/3IT3Nt6>

## Caractéristiques des colonnes

Généralement, on peut trouver quatre groupes distincts de classification :

**1. Identifiant** : C'est un attribut qui permet d'identifier un enregistrement de façon unique. Il s'agit souvent de la clé primaire, qui peut être simple ou composée.

➔ **Ces attributs doivent être pseudonymisés, masqués ou bien supprimés.**

**2. Quasi-identifiant** : Un quasi-identifiant ne peut pas permettre l'identification d'un enregistrement unique s'il est exploité d'une manière isolée. Cependant, lorsque ces attributs sont combinés, ils peuvent permettre l'identification d'enregistrement unique.

➔ **Exemple : un ensemble composé d'attributs [Sexe, Age, Adresse]**

**3. Sensible** : Il s'agit des données que les individus ne veulent généralement pas publier, comme les informations médicales ou les salaires. Les données que l'on veut conserver « secret ».

**4. Autre** : Il s'agit d'autres données figurant dans le fichier de données et qui ne sont pas sensibles.

### Choix des valeurs de K et L :

On dit qu'un jeu de données est K-anaonymisé s'il y a au moins « K » enregistrements pour n'importe quel groupe d'enregistrement partageant le même ensemble d'attributs quasi-identifiants.

On dit qu'un jeu de données est L-diversifié s'il y a au moins « L » valeurs sensibles pour chaque classe / groupe de quasi-identifieur. Ainsi, on dit que les données sensibles sont L diversifiées.

## Exemple

Les données ci-dessous correspondent à 3-anonymity & 1-diversity

Identifiant	Quasi-identifiant			Sensible
	Pseudonyme	Année de naissance	Sexe	
1000001	1971	Femme	Préfecture d'Oujda-Angad	1751
1000002	1971	Femme	Préfecture d'Oujda-Angad	1809
1000003	1971	Femme	Préfecture d'Oujda-Angad	2450
1000004	1971	Femme	Préfecture d'Oujda-Angad	2015
1000005	1974	Femme	Province de Berkane	1089
1000006	1974	Femme	Province de Berkane	2400
1000007	1974	Femme	Province de Berkane	2010
1000008	1977	Homme	Préfecture de Fès	1876
1000009	1977	Homme	Préfecture de Fès	1645
1000010	1977	Homme	Préfecture de Fès	1956
1000011	1978	Homme	Province d'El Hajeb	1655
1000012	1978	Homme	Province d'El Hajeb	1655
1000013	1978	Homme	Province d'El Hajeb	1655

Pour chaque triplet de quasi-identifiant [année de naissance, Sexe, Gouvernorat], on trouve au moins 3 individus.

=>  $K = 3$

Cependant, pour les trois derniers enregistrements (appartenant à la même classe de quasi-identifiant), la donnée sensible est la même « 1655 », il s'agit d'une (1) seule valeur sensible.

=>  $L = 1$

On applique la technique d'agrégation sur l'attribut « gouvernorat » pour passer au niveau « région » :

➡ Région = Nord [Tunis + Ariana]

➡ Région = Centre [Sousse + Monastir]

Et on applique la technique de généralisation sur l'attribut « année de naissance » pour obtenir les deux classes ci-dessous : [1970-1974] et [1975-1979]

On obtient le résultat ci-dessous :

Identifiant	Quasi-identifiant			Sensible
	Pseudonyme	Année de naissance	Sexe	
1000001	[1971-1974]	Femme	Oriental	1751
1000002	[1971-1974]	Femme	Oriental	1809
1000003	[1971-1974]	Femme	Oriental	2450
1000004	[1971-1974]	Femme	Oriental	2015
1000005	[1971-1974]	Femme	Oriental	1089
1000006	[1971-1974]	Femme	Oriental	2400
1000007	[1971-1974]	Femme	Oriental	2010
1000008	[1975-1979]	Homme	Fès-Meknès	1959
1000009	[1975-1979]	Homme	Fès-Meknès	1617
1000010	[1975-1979]	Homme	Fès-Meknès	2011
1000011	[1975-1979]	Homme	Fès-Meknès	1655
1000012	[1975-1979]	Homme	Fès-Meknès	1655
1000013	[1975-1979]	Homme	Fès-Meknès	1655

Les données correspondent maintenant à 6-anonymity & 4-diversity (Après transformation et génération de la nouvelle version à ouvrir).

Il est par conséquent impossible d'identifier un individu et son salaire avec certitude, même en croisant ces informations avec une autre source de données (jeu de données).

## 3.2. LES PLATEFORMES OPEN DATA

Pour se conformer aux exigences du décret gouvernemental n°2021-3 ayant trait aux données ouvertes, les organismes publics en Tunisie sont tenus de publier leurs jeux de données directement sur le portail national des données ouvertes.

### 3.2.1. QUE VEUT-ON DIRE PAR PUBLICATION DIRECTE SUR LE PORTAIL NATIONAL ?

Il s'agit de l'insertion manuelle des jeux de données (métadonnées et données) en utilisant l'espace réservé à chaque organisme public.

Cette approche exige l'enregistrement des métadonnées dans un formulaire dédié et le téléchargement des fichiers porteurs de données ou bien l'insertion des liens vers les données.

Par ailleurs, la publication directe peut se faire également en adoptant une approche automatisée qui exploite les APIs offerts par le portail national.

À cet effet, l'organisme public doit solliciter l'équipe de l'UAE pour la création d'un compte sur la plateforme nationale et/ou la génération d'une clé d'API pour l'alimentation automatisée.

**Comment peut-on choisir la meilleure approche ?**

#### ACTUALISATION AUTOMATISÉE

Si la fréquence d'actualisation des jeux de données est très élevée (journalière, temps réel, etc...).

Si je dispose des bases de données dont je peux extraire les données à publier.

#### INSERTION ET ACTUALISATION MANUELLE

Si je ne dispose pas de base de données,

Si le nombre de jeux de données n'est pas élevé.

Si l'actualisation n'est pas fréquente.

Un même producteur de données publiques ouvertes peut combiner entre les deux approches.



## Interopérabilité entre les plateformes spécifiques et la plateforme nationale

Les organismes publics qui utilisent une plateforme open data spécifique doivent assurer l'interopérabilité de cette dernière avec la plateforme nationale.

Le moissonnage permet de référencer sur la plateforme nationale des jeux de données déjà publiés sur d'autres catalogues. Ainsi, un organisme public n'a pas besoin d'importer à la main ses jeux de données un par un.

Le moissonnage sur la plateforme nationale ne fonctionne qu'avec certains formats de métadonnées. Pour vérifier si votre catalogue spécifique peut être référencé sur le portail national, vous serez invités à coordonner avec l'équipe de l'UAE.

Si le moissonnage n'est pas assuré, le producteur de données ouvertes n'est pas en conformité par rapport aux exigences du décret n°2021-3 et plus précisément à celles de l'article 14.

### 3.2.2. QUEL PORTAIL CHOISIR POUR LES DONNÉES OUVERTES ?

#### Ce que dit le décret open data :

L'article 13 stipule que « en cas d'utilisation d'un portail spécifique de données ouvertes, les organismes doivent fournir les ressources humaines et financières nécessaires à la mise en place et à la gestion de ce portail.»

- ➔ Il doit disposer des ressources humaines et financières nécessaires à la mise en place et à la gestion de la nouvelle plateforme.
- ➔ Il est appelé à utiliser une solution open source.
- ➔ Il doit assurer l'interopérabilité avec la plateforme nationale.
- ➔ La plateforme doit proposer des mécanismes de participation et d'interaction avec les utilisateurs.

#### Les plateformes open data devraient inclure des mécanismes de participation tels que :

- une rubrique pour les demandes des nouveaux jeux de données ; les demandes soumises ainsi que leurs réponses doivent être publiées sur le portail,
- un espace de commentaires pour chaque jeu de données,
- une rubrique pour les propositions d'amélioration de la plateforme open data,
- un forum de discussion.

### Autres exigences fonctionnelles :

En plus des exigences réglementaires, ce guide propose une liste de fonctionnalités obligatoires et recommandées pour choisir un portail open data. Cette liste peut être enrichie par chaque organisme public au regard de ses objectifs et priorités.

Un portail open data est donc un espace de présentation des données publiques ouvertes avec les fonctions suivantes :

- le lieu de stockage des données (un entrepôt de données)
- l'indexation et la recherche des jeux de données (métadonnées et données),
- la présentation détaillée des jeux de données (jeu de données et ressources, métadonnées et données) sous différents formats, y compris API,
- la visualisation et la présentation graphique des données quand cela est possible.

### Solutions de Données Ouvertes couramment utilisées :

**CKAN** <sup>[22]</sup> est le catalogue de données open source le plus utilisé par les gouvernements. Cette solution peut être installée sur un serveur Linux.

Le catalogue CKAN est conçu pour publier et gérer des données au moyen d'une interface utilisateur ou d'une API. D'un point de vue technique, il s'agit de code en Python pour le back-end et de JavaScript pour la partie interface.

La mise en place de catalogue de données ouvertes à base de CKAN peut être rapide mais la personnalisation et l'optimisation demandent plus d'investissement qu'une solution clé en main. Une communauté de développeurs conséquente assure une évolution constante de la technologie et des outils.

En dehors des outils de recherche, de visualisation, de cartographie classique et d'interrogation API, CKAN propose aussi des extensions externes accessibles sur ce site : <https://github.com/ckan/ckan/wiki/List-of-extensions>

Exemples : <https://catalog.data.gov/>; <https://data.gov.uk/>;



<sup>22</sup> <https://ckan.org>

En Tunisie, à la date d'élaboration de ce guide, il existe 5 plateformes sectorielles à savoir <http://agridata.tn/>; <http://datatransport.tn/>; <http://openbaladiati.tn/>; <http://catalog.industrie.gov.tn/>; <http://www.openculture.gov.tn/>;

**DKAN** <sup>[23]</sup> est une variante de CKAN qui repose sur DRUPAL <sup>[24]</sup>, un système de gestion de contenu très populaire qui est écrit en PHP au lieu de Python. DRUPAL possède sa propre architecture modulaire avec des milliers de modules disponibles en téléchargement. Ce logiciel permet également de personnaliser les modules et possède une grande communauté de développeurs.

The logo for dkan, consisting of the lowercase letters 'dkan' in a bold, blue, sans-serif font.

**uDATA** <sup>[25]</sup> c'est la plateforme développée par EtatLab, l'agence de service numérique de l'Etat Français. Elle est développée en Python et facilement adaptable au besoin.

Pour faire tourner une plateforme uData en mode standalone, vous avez besoin d'un simple serveur linux ou MacOS tout en suivant la documentation en ligne. Par la suite, il faut juste travailler sur la personnalisation. Il est aussi possible de déployer la plateforme avec Docker.

The logo for uDATA, featuring a stylized blue fingerprint icon to the left of the word 'UDATA' in a blue, sans-serif font.

Le code source est accessible sur : <https://github.com/opendatateam/udata>

Exemples: <https://www.data.gouv.fr> ; <https://data.public.lu> ; <https://dados.gov.pt>

**GeoNode** <sup>[26]</sup> est une plateforme open source qui vise à développer des systèmes d'informations géospatiales (GIS) et à déployer des infrastructures de données géo-spatiales. GeoNode a été conçue pour être facilement étendue et modifiable, elle peut également être intégrée dans une autre plateforme existante.

The logo for GeoNode, featuring a stylized blue asterisk icon to the left of the word 'GeoNode' in a blue, sans-serif font.

Exemples : <https://geonode.wfp.org>

<sup>23</sup> <https://getdkan.org>

<sup>24</sup> <https://www.drupal.org>

<sup>25</sup> <https://udata.readthedocs.io/en/latest>

<sup>26</sup> <https://geonode.org>

## 3.3. LE PLAN D'ACTION ANNUEL DES DONNÉES OUVERTES

### 3.3.1. QU'EST-CE QU'UN PLAN D'ACTION NATIONAL :

Un plan d'action national d'ouverture des données publiques est un document qui regroupe l'ensemble des projets et des activités relatifs à l'ouverture des données à mettre en œuvre l'année suivante au niveau national selon un calendrier précis.

#### RÉFÉRENCES LÉGALES :

Articles 4 et 6 du Décret gouvernemental n° 2021-3 du 6 janvier 2021, relatif aux données publiques ouvertes.

### 3.3.2. QUE DOIT CONTENIR UN PLAN D'ACTION ANNUEL :

Chaque acteur public doit inclure dans son plan d'action annuel deux parties à savoir les activités qui seront mises en œuvre l'année suivante et une liste des jeux de données les plus importants à ouvrir l'année prochaine.

#### 3.3.2.1. LES ACTIVITÉS À INCLURE :

Cette partie doit inclure l'ensemble des activités qui seront mises en œuvre l'année suivante, dans l'optique de renforcer et d'améliorer le programme national d'ouverture de données. Les activités peuvent être regroupées selon les trois (03) dimensions ci-dessous :

##### Accroître l'accès aux données ouvertes :

- ▶ Développer des activités pour promouvoir les initiatives d'ouverture des données existantes.
- ▶ Développer le réseau des responsables open data et data champions.
- ▶ Développer des activités pour améliorer la qualité des données.
- ▶ Améliorer l'interopérabilité et la conformité des données ouvertes aux standards internationaux.
- ▶ Développer des activités pour la conduite et la publication de l'inventaire des données détenues par les acteurs publics.

## Instaurer une culture « Ouvert par défaut et protégé si nécessaire »

- ▶ Doter les fonctionnaires du gouvernement d'un état d'esprit axé sur les données.
- ▶ Développer les compétences en matière de gouvernance des données auprès des fonctionnaires et décideurs.
- ▶ Intégrer les données ouvertes dans la conception de projets, programmes, services, systèmes et politiques
- ▶ Publier plus de jeux de données sous format ouvert sans indiquer leur utilisation potentielle par des tiers.
- ▶ Fournir des solutions techniques pour l'extraction, le traitement et la publication automatisée des données ouvertes.
- ▶ Mettre en place des mécanismes de protection et de contrôle pour gérer la publication massive des jeux de données (données protégées, etc...)

## Accroître le potentiel de réutilisation :

- ▶ Identifier et prioriser les données à fort impact de réutilisation.
- ▶ Démontrer la valeur des données ouvertes (économiques, sociales et politiques).
- ▶ Permettre des connexions bidirectionnelles entre les utilisateurs finaux et les acteurs publics.
- ▶ S'engager en permanence avec la communauté des données pour identifier les priorités et fluidifier la circulation des idées de réutilisation provenant des utilisateurs finaux (les besoins, les demandes).
- ▶ Développer des partenariats et organiser des concours, des hackathons, des datajams, etc....,
- ▶ Permettre aux utilisateurs de données de partager leurs expériences, outils et codes sources.

### LE PLAN D'ACTION ANNUEL COMPREND-T-IL LES ACTIONS RÉCURRENTES ET HABITUELLES ?

Le plan d'action peut inclure des actions permanentes à mener chaque année. Mais ce n'est obligatoire que si l'on souhaite mettre en valeur cette activité répétitive.



Exemple : l'actualisation des jeux de données ouverts est une action récurrente qui ne peut pas apparaître dans un plan d'action annuel d'ouverture des données même si l'acteur public souhaite mettre en valeur cette activité.



Exemple : l'amélioration de la plateforme nationale ou bien sectorielle est une action récurrente qui peut apparaître dans un plan d'action annuel d'ouverture des données.

### 3.3.2.2. LA LISTE DES JEUX DE DONNÉES LES PLUS IMPORTANTS À OUVRIR :

Dans cette section, l'acteur public est invité à lister les jeux de données les plus importants à ouvrir l'année suivante. Cette liste sera sous forme tabulaire et elle peut contenir les éléments ci-dessous:

■	Le nom du service interne responsable de l'ouverture de la donnée.
■	L'intitulé du jeu de données (en deux langues si possible).
■	Une description détaillée (en deux langues si possible).
■	Le ou les thèmes au niveau de la plateforme nationale qui seront associés à ce jeu de données.
■	L'échéance d'ouverture en précisant le mois pour l'année prochaine.
■	La fréquence d'ouverture qui sera proposée.
■	L'estimation budgétaire (s'il y a lieu).
■	La référence du jeu de données sur l'inventaire des données publiques (chaque jeu de données mentionné dans un plan d'action doit figurer dans l'inventaire des données publiques de l'acteur public).
■	L'énumération des cas de réutilisations potentielles (internes et/ou externes).

#### Quels jeux de données à insérer dans un plan d'action annuel ?

L'identification des jeux de données à ouvrir l'année prochaine nécessite l'implication de toutes les parties prenantes de l'écosystème de l'open data telles que la société civile, le secteur académique, le secteur privé, la communauté des développeurs, etc.... Tout en prenant en considération les jeux de données identifiés dans l'inventaire des données publiques non encore ouvertes et dont le niveau de priorisation est élevé.

**Dans un plan d'action des activités d'ouverture des données, on peut inclure aussi les jeux de données à ouvrir à moyen terme (dans les deux ou les trois prochaines années). Cela permettrait de partager la vision stratégique du programme open data d'une organisation.**

### 3.3.3. QUELLE MÉTHODOLOGIE POUR UN PLAN D'ACTION RÉUSSI ?

Le responsable de l'open data au sein de chaque organisation doit créer une équipe de travail pour l'élaboration du plan d'action de l'année suivante. Il est recommandé d'impliquer les personnes ayant participé à la mission de l'inventaire des données publiques.

Il peut prévoir des rencontres/interviews entre la communauté des utilisateurs des données et les utilisateurs finaux identifiés.

Il peut envoyer des questionnaires en ligne aux acteurs de l'écosystème open data.

Il peut adopter une démarche hybride regroupant les deux approches susmentionnées.

La mission d'inventaire des données publiques ainsi que son enrichissement doit figurer dans tous les plans d'action comme une activité nécessaire mais insuffisante.

#### LES CLÉS DE SUCCÈS D'UN PLAN D'ACTION



Fixer des objectifs SMART à atteindre & prévoir des délais réalistes,



Faire participer toutes les parties prenantes impliquées,



Elaborer un document simple, clair, opérationnel.

## 3.4. LE RAPPORT D'ACTIVITÉ ANNUEL DES DONNÉES OUVERTES

### 3.4.1. QU'EST-CE QU'UN RAPPORT D'ACTIVITÉ ANNUEL ?

Un rapport d'activité est un instrument de suivi ancré dans le présent mais il constitue aussi une projection dans le futur. Il représente en effet une synthèse de l'activité et des résultats obtenus concernant l'ouverture des données publiques au cours de l'année écoulée et en même temps il nous aide à prendre des décisions pour le futur. Selon le décret open data, qui est tenu de préparer un Rapport d'Activité ?

L'UAE doit préparer un rapport d'activité annuel de l'année écoulée. Ce rapport doit être publié avant la fin du mois de mars sur le site web du ministère en charge de la modernisation administrative et du portail national des données publiques ouvertes.

Chaque Responsable Open Data est tenu de préparer un rapport d'activité au niveau de son organisation publique et de le communiquer à l'UAE dans les meilleurs délais ().

### 3.4.2. QUEL CONTENU À METTRE DANS UN RAPPORT D'ACTIVITÉ ?

Évaluation générale de la performance de l'année précédente en termes d'ouverture des données et des activités qui y sont liées en se référant au plan d'action annuel de l'année écoulée.

Forces et faiblesses identifiées lors du processus d'ouverture des données.

Coût total de mise en œuvre et implémentation du programme d'ouverture des données.

Perspectives et recommandations : il est recommandé d'ajouter au niveau national les recommandations du comité consultatif paritaire.

### 3.4.3. COMMENT RÉUSSIR UN RAPPORT D'ACTIVITÉ ?

Pour réussir un rapport d'activité, il serait judicieux de mettre en place en début de chaque année des mécanismes pour assurer le suivi et l'implémentation des activités fixées dans le plan d'action de l'année A-1.

Ces mécanismes de suivi peuvent prendre la forme d' :

organisation des réunions périodiques de coordination (par exemple chaque mois) avec les parties impliquées afin d'évaluer l'état d'avancement, d'identifier les difficultés à surmonter, etc...

élaboration des rapports trimestriels ou semestriels de mise en œuvre du programme open data.



## 3.5. L'INVENTAIRE DES DONNÉES PUBLIQUES

### 3.5.1. QU'EST-CE QU'UN INVENTAIRE DES DONNÉES PUBLIQUES ?

Un inventaire des données publiques est un document qui contient une liste des jeux de données produits et/ou collectés par chaque organisme dans le cadre de l'exercice de ses activités. Chaque jeu de données de l'inventaire est identifié par un ensemble de métadonnées à l'instar de l'identification de la donnée (son intitulé, sa description, et quelques mots clés, etc...), sa classification, sa périodicité de collecte/production, son propriétaire et diverses autres informations.

**En bref :** C'est une liste contenant uniquement des métadonnées sur les données produites et/ou collectées et non des données réelles.

### 3.5.2. QUELLE DÉMARCHÉ POUR LA CONDUITE D'UN INVENTAIRE DES DONNÉES PUBLIQUES ?

L'UAE a mis en place une plateforme « PDIMS » pour faciliter la conduite de l'inventaire des données publiques détenues par les organismes.

**Les Composantes du système d'inventaire des données:**

#### COMPOSANTE POUR L'ADMINISTRATION « PDIMS »

Consiste à digitaliser le processus d'inventaire des données à travers l'informatisation des différentes étapes à savoir :  
L'identification des données, l'enregistrement des métadonnées de chaque jeu de données identifié, la gestion des organismes publics et des utilisateurs, le tableau de bord pour le suivi des activités d'inventaire de la structure.

#### COMPOSANTE POUR L'UTILISATEUR FINAL « REGISTRE DES DONNÉES »

Consiste à la publication du registre des données afin de donner aux utilisateurs finaux une vue d'ensemble des données publiques inventoriées.

Il incombe aux structures publiques d'identifier les jeux de données à inclure dans le registre d'inventaire.

<p><b>DÉFINITION DU PÉRIMÈTRE ET FIXATION DES DÉLAIS</b></p>	<p>Le responsable Open Data « ROD » décidera des directions internes qui seront invitées à participer à la mission de l'inventaire. Au niveau de la plateforme, il doit ajouter ces services internes à savoir les directions générales, directions centrales, directions, services, etc.... Il doit également ajouter les organismes publics sous-tutelle si le ROD représente un ministère.</p>	<p>»</p>
<p><b>CONSTITUTION DES ÉQUIPES</b></p>	<p>Le responsable Open Data prendra en charge la création des comptes de collaborateurs au sein de son organisme ainsi que les comptes des responsables « Open Data » des organismes sous-tutelle.</p> <p>Pour le département identifié à l'issue de la première étape, on doit créer une équipe pour la conduite de la mission de l'inventaire.</p>	<p>»</p>
<p><b>IDENTIFICATION DES DONNÉES « BRAINSTORMING »</b></p>	<p>Chaque département interne sera invité à organiser des réunions de brainstorming dont l'objectif est d'identifier des idées sur les données détenues par les organismes publics. La plateforme PDIMS offre un outil de carte mentale pour l'enregistrement des outputs des réunions de brainstorming. Elle permet d'enregistrer les idées des données ainsi que le nom de la personne qui sera en charge pour renseigner les métadonnées.</p>	<p>»</p>
<p><b>ENREGISTREMENT DES MÉTADONNÉES</b></p>	<p>Cette étape consiste à enregistrer les métadonnées de chaque jeu de données identifié dans la phase précédente. Trois catégories de métadonnées ont été définies (clés, de priorisation, avancées)</p>	<p>»</p>
<p><b>PUBLICATION DU REGISTRE ET COLLECTE DES FEEDBACKS</b></p>	<p>Cette étape consiste à la validation du travail réalisé et à la publication du registre de l'inventaire. Le système permettra aussi la publication de l'inventaire juste après la clôture de la quatrième étape c.-à-d. une fois que les métadonnées clés ont été renseignées. Cette décision incombe au responsable open data dans chaque organisation.</p>	<p>»</p>

## 3.6. INSTAURER UNE DÉMARCHE BASÉE SUR LA RÉUTILISATION

### 3.6.1. QU'EST-CE QUE LA RÉUTILISATION DANS LE DOMAINE DE L'OPEN DATA ?

La réutilisation désigne l'utilisation des données ouvertes par des tiers généralement pour d'autres fins que celles de la mission principale de l'acteur public qui a rendu ces dernières accessibles gratuitement sous format ouvert et libre de droit.

Elle peut prendre la forme d'une visualisation des données, d'une cartographie, d'un article à publier et/ou d'une recherche, d'une application web et/ou mobile, d'un nouveau et/ou d'une amélioration d'un service, d'une économie et/ou d'une optimisation de l'action publique.

### 3.6.2. QUELLES SONT LES ÉTAPES POUR UNE DÉMARCHE RÉUSSIE EN FAVEUR DE LA RÉUTILISATION DES DONNÉES OUVERTES ?

Quelle que soit la nature de l'acteur public, on peut identifier des étapes en commun favorisant la création d'impact à travers la réutilisation des données ouvertes à savoir :

*l'identification  
des réutilisateurs  
externes/internes*

*la priorisation des  
enjeux et de la  
demande  
d'ouverture*

*l'ouverture  
progressive et  
continue des jeux  
de données*

*l'accompagnement  
et la valorisation*

Cette démarche peut varier selon l'activité des organisations publiques et en fonction des ressources humaines et matérielles mises à leurs dispositions.

#### 3.6.2.1. IDENTIFIER LES RÉUTILISATEURS ?

Le faible taux de **réutilisation** peut être justifié par le fait que les acteurs publics abordent la réutilisation à partir de leur propre héritage institutionnel et se préoccupent beaucoup plus de la conformité législative que des besoins et des attentes des **réutilisateurs**.

## Réutilisateurs

Producteurs de données	Promoteurs de données ouvertes	Data-scientists	Data-journalistes
Développeurs	Chercheurs	Citoyens	Associations

Le responsable Open Data doit d'emblée essayer d'identifier les réutilisateurs potentiels. Cette activité nécessite au préalable la compréhension de la chaîne de réutilisation des données ouvertes qui repose en général sur 4 étapes : la recherche, l'interprétation, la transformation et l'intégration.

Ainsi, en tant qu'acteur public, je dois commencer par le repérage des acteurs clés et, en particulier, identifier les premiers réutilisateurs potentiels en essayant d'apporter des réponses aux éléments ci-dessous :

- Qui est à la recherche des données que je produis et ou je collecte ?
- Qui demandent l'accès aux données produites ou collectées par mon organisation ?

### 3.6.2.2. PRIORISER LES ENJEUX ET LES DEMANDES D'OUVERTURE ;

Une fois que les réutilisateurs sont identifiés et repérés, le meilleur moyen pour démarrer le dialogue avec eux consiste à effectuer le recensement des données ayant un intérêt important en commun entre le(s) producteur(s) et ces réutilisateurs. Ce recensement concernera en particulier les données qui déclenchent des idées créatives et qui proposent des solutions aux problèmes et enjeux identifiés dans le cadre de ce dialogue ou même antérieurement.

**Cette activité doit prendre en considération les questionnements ci-dessous :**

Les contraintes de l'acteur public : d'ordre technique, financier, juridique, etc... en relation avec les données identifiées.

Les motivations et les contraintes des réutilisateurs identifiés.

Et surtout, l'accès aux autres sources complémentaires de jeux de données ouverts et nécessaires pour le développement des réutilisations en question (**difficile à garantir si l'offre reste assez limitée**).

Dans cette étape, il est important de s'interroger aussi sur les partenariats utiles à développer. Il s'agirait de partenariats avec des organisations internationales, avec des opérateurs privés, avec des associations actives et même de partenariats croisés avec d'autres acteurs publics, etc...

### APPORTS DU PARTENARIAT

#### APPORT FINANCIER

Les partenariats peuvent apporter le financement nécessaire à toute les étapes d'ouverture des données.

#### APPORT TECHNIQUE

Cet apport pourrait aller jusqu'à l'outsourcing de certaines activités.

#### 3.6.2.3. OUVRIR LES PREMIERS JEUX DE DONNÉES ;

La troisième étape consiste à ouvrir les données **progressivement** et **régulièrement** couvrant contenu et conteneur (les données, les métadonnées et la documentation).

En général, il est plus utile de s'intéresser à un petit nombre de données en profondeur plutôt qu'à un nombre très large de manière superficielle. Des livraisons progressives et régulières, même si elles ne sont pas parfaites, assurent un feedback régulier des réutilisateurs. Cela permet aux acteurs publics de corriger rapidement les problèmes identifiés et d'adapter l'offre à la demande réelle.

Pour assurer la viabilité et la pérennité des usages, les réutilisateurs exigent, à moyen et à long terme, un accès permanent aux jeux de données en favorisant l'automatisation, une mise à jour régulière, un niveau de qualité acceptable et tout cela doit s'appliquer à l'échelle des données et des métadonnées.

#### 3.6.2.4. ACCOMPAGNER ET VALORISER ;

Ouvrir des jeux de données ne suffit pas. L'acteur public doit également savoir orienter, stimuler et valoriser les usages de ses données. Les partenariats pourraient être très efficaces à ce stade.

Les moyens sont nombreux, citons à titre d'exemples : les appels à projets, l'animation de réseaux d'innovateurs, les hackathons, les datajams, les ateliers créatifs sous toutes leurs formes.

