

Ouverture des données publiques au Maroc

**Guide méthodologique pour la mise en
œuvre d'un inventaire de données au sein
d'une structure publique de
l'Administration marocaine**

Version 1.3 - février 2023

Ce guide a été élaboré avec l'appui de la
Banque Mondiale

Table des Matières

Version	3
Introduction	3
Rôle de l'inventaire de données dans l'ouverture des données publiques	3
Acteurs d'un inventaire de données.....	6
Enrôlement des structures publiques.....	7
Mise en œuvre de l'inventaire.....	8
Structure organisationnelle & approche stratégique	8
Sensibilisation et mobilisation	10
Processus de cartographie.....	10
Classification des données.....	11
Priorisation des données	12
Publication de l'inventaire	12
Interaction avec les réutilisateurs.....	13
Les métadonnées à répertorier	13
Calendrier d'exécution d'un inventaire de données	19

Version

La présente version de ce document est disponible à l'adresse : www.data.gov.ma.

Introduction

Ce document est un manuel à destination des Responsables Open Data (ROD) au sein des structures publiques. L'objectif de ce manuel est d'apporter à ces ROD un support opérationnel dans la mise en place et la gestion d'un inventaire de données ouvertes au sein de leurs structures respectives.

Ce manuel est divisé en quatre parties. La première partie présente l'importance de l'inventaire de données dans la mise en œuvre d'une initiative Open Data. La deuxième partie présente les acteurs de l'inventaire de données. La troisième partie propose une méthodologie de mise en œuvre de l'inventaire. Enfin, la dernière partie liste l'ensemble des métadonnées à collecter pour chaque entrée de l'inventaire.


Ce manuel est un document vivant qui devrait s'enrichir au fur et à mesure des retours d'expérience des ROD, des défis qu'ils rencontrent et des stratégies qu'ils mettent en place pour les résoudre. Les commentaires et les propositions d'évolutions sont à envoyer à l'adresse suivante : Opendata@add.gov.ma.

Rôle de l'inventaire de données dans l'ouverture des données publiques

Les données ouvertes en tant que concept trouvent leur origine dans la conviction que les données publiques devraient être librement disponibles pour être publiées, utilisées et réutilisées sans limitations ni contrôle. L'idée principale de cette approche est que la disponibilité de ces données dans un format approprié permettrait leur valorisation à la fois au sein de l'Administration mais également par l'ensemble des acteurs de la société. L'exploitation et la valorisation de ces données amènent à plusieurs impacts mesurables que ce soit en termes d'efficacité de l'action publique, de création de richesse, d'emplois, de services pour les citoyens et de connaissances.

L'émergence de ces impacts est directement liée à la fois au volume de données publiées, à leur pertinence pour les réutilisateurs (internes ou externes à l'Administration), au format de la publication et à la qualité des données.

Le processus de publication de données ouvertes est schématiquement représenté dans le diagramme suivant :



Etapes	Acteurs	Tâches
Identifier	Resp. Open Data (ROD) - Gestionnaires & Producteurs de données	- Rencontrer les agents - Cartographier les données - Documenter les données & métadonnées
Valider la publication	ROD - Responsables	- Evaluer les jeux de données - Prioriser l'ouverture - Valider l'ouverture
Extraire	ROD - DSI - Gestionnaires de données - Prestataires SI	- Négocier avec les prestataires - Reconstituer schémas - Développer outils d'extraction
Editer	ROD - Gestionnaires & Producteurs de données	- Convertir en format lisible par une machine - Vérifier et Améliorer la qualité - Anonymiser
Standardiser	ROD - Gestionnaires & Producteurs de données	- Convertir dans un format ouvert - Transformer les données selon les standards techniques & référentiels adoptés
Publier	ROD - Service Informatique - Gestionnaires & Producteurs de données	- Publier les données sur le portail
Maintenir	ROD - Gestionnaires & Producteurs de données	- Corriger les erreurs - Mettre à jour - Répondre aux questions et commentaires

Pour conduire ce processus et atteindre l'objectif d'un grand volume de données de qualité publiées, les structures publiques et leurs agents doivent répondre à plusieurs questions essentielles :

- Quelles sont les données disponibles au sein de la structure et quel est leur emplacement ?
- Quelle est la qualité de ces données ?
- Parmi ces données, quelles sont celles qui peuvent être publiées et partagées (avant ou après traitement) ?
- Quelles sont celles qui sont les plus pertinentes et les plus utiles ?
- Quelles sont les étapes à suivre pour procéder à la publication des données sous un format optimum qui favorisera la réutilisation de ces données ?
- Sous quelles licences de réutilisation peuvent être publiées ces données ?

Le rôle d'un inventaire de données est exactement de répondre à ces questions épineuses et de résoudre des différentes problématiques administratives rencontrées dans la publication de données afin d'en faciliter la réalisation technique. L'objectif premier est de transformer le processus de publication de données en un problème uniquement technique réalisable par n'importe quel agent sans nécessité de validation hiérarchique. En détail, le rôle de l'inventaire est de :

1. Référencer l'ensemble du patrimoine de données de la structure publique¹.
2. Répertorier les résultats de l'analyse de chaque source de données. Cette analyse couvre les sujets suivants :
 - a. La collecte des informations relatives aux données (voir la section 6 de ce document « Métadonnées ») ;

¹ Cette étape permettra, le cas échéant, de détecter des duplications d'efforts potentiels.

- b. La classification des données ;
 - c. La qualité des données ;
 - d. Les traitements à effectuer sur les données avant publication ;
 - e. L'intérêt de ces données pour les réutilisateurs potentiels.
3. Associer une licence de réutilisation de ces données en fonction du cadre légal des licences de réutilisation et de l'origine des données.
 4. Mettre à disposition des réutilisateurs (internes et externes à l'Administration) les informations de l'inventaire afin de faciliter les demandes de publication.

Concernant le point 2.c, l'évaluation de la qualité d'un jeu de données s'intéresse à la qualité intrinsèque des données ainsi que leur capacité à être exploitées sans difficulté et à être croisés avec d'autres jeux de données. Les critères de qualité impliquent en particulier les éléments suivants :

- Les données sont publiées avec une licence de réutilisation ouverte ;
- Les données utilisent des formats ouverts ;
- Les données sont structurées et exploitables informatiquement (machine-readable)
 - Les données numériques tabulaires sont sous format long.
- Les données sont des données brutes désagrégées ;
- Les données sont à jour au regard du cycle de collecte ;
- Les données utilisent les standards de données adoptés par le comité de pilotage au niveau national ;
- Les données utilisent les référentiels communs de l'administration et les référentiels sectoriels ;
- Les données sont documentées ;
- Les données sont formellement valides au regard du dictionnaire de données.

Une des priorités de l'inventaire est, d'une part, d'évaluer dans quelle mesure les données répondent à ces critères, et d'autre part, d'identifier les transformations (transcodage, restructuration, anonymisation, etc.) à appliquer pour maximiser la facilité de réutilisation des données avant leur publication.

Définition : Inventaire de données vs. Annuaire de données

Le terme « inventaire de données » nécessite d'être défini précisément. Le terme « inventaire » est issu du domaine de la gestion des stocks et représente généralement l'exécution d'une tâche de référencement d'un ensemble de produits (ici des sources de données). Ce référencement est généralement une tâche administrative exécutée de façon ponctuelle sur une courte période, parfois par des personnels extérieurs à la structure et qui sert principalement aux gestionnaires de stock. Cette définition est peu appropriée pour le cas des données. En effet, d'une part, il est quasi impossible dans les grosses structures de conduire un processus exhaustif de référencement des données réparties sur des dizaines d'ordinateurs (voir même dans des documents papiers) dans des dizaines de bureaux. D'autre part, l'objectif n'est pas de conduire régulièrement une évaluation du patrimoine de données, mais de mettre à jour, au fur et à mesure, l'outil en fonction de l'évolution du cycle de vie des

données (arrêt de certains jeux de données, modification, mise en place de nouvelles collectes, etc.). Enfin, l'inventaire de données est un outil utile à l'ensemble de la structure et à tout agent qui recherche des données. De ce fait, un terme adapté pourrait être « annuaire de données » qui décrit de façon plus appropriée, l'utilisation de cet outil et les principes de mise à jour au cours du temps.

Portée des données

Une des questions principales à résoudre avant de s'engager dans la mise en place de l'inventaire est de savoir quelles données doivent être incluses dans l'inventaire. A partir des objectifs définis dans la première partie de cette section, il est important que l'inventaire :

- Soit un outil utile au sein de la structure et donc :
 - Couvre les données collectées et les données produites par la structure ;
 - Couvre toutes les catégories de données y compris celles qui ne peuvent pas être rendues publiques (mais qui sont utilisées/utilisables en interne).
- Permette de détecter les duplications d'efforts de collecte/production. Pour ce faire, l'inventaire se focalisera uniquement sur les données primaires collectées ou produites par la structure elle-même (et non les données en provenance d'autres structures).

Acteurs d'un inventaire de données

La réalisation d'un inventaire de données implique plusieurs acteurs internes et externes à la structure :

- **COPIL Open Data et ADD:** Le COPIL Open Data et l'ADD définissent et adoptent les éléments structurants de l'inventaire de données qui :
 - Assurent l'homogénéité des données d'inventaires entre les structures publiques, notamment au niveau de la liste des métadonnées à répertorier ;
 - Mettent en place, en coordination avec les autorités compétentes, un cadre légal commun en particulier sur les licences de données ;
 - Mettent en place les outils nécessaires à la réalisation de l'inventaire. Ces outils incluent notamment les modèles (« template ») de collecte de données et la plateforme d'inventaire ;
 - Développent les capacités des ROD pour exécuter les différentes tâches de l'inventaire.
- **Premier Responsable de la structure :** Le rôle du premier responsable de la structure est double :
 - Mobiliser l'ensemble des collaborateurs concernés et positionner la mise en œuvre de l'inventaire comme une action stratégique de la structure ;
 - Valider formellement l'inventaire afin que les informations contenues, notamment la classification des données fassent référence au sein de la structure.

- **ROD** : Le ROD est la cheville ouvrière de l'inventaire. Son rôle est de coordonner l'ensemble des activités liées à la réalisation de l'inventaire et d'assurer le suivi et l'évaluation. Dans les petites structures, le ROD pourrait exécuter lui-même les tâches.
- **Chargé d'accès à l'information** : Le chargé d'accès à l'information est un acteur important de l'inventaire. Par son rôle, il a une visibilité globale sur les requêtes soumises à sa structure par l'ensemble des acteurs de la société marocaine. D'une part, il est en partie informé des sources de données existantes, et d'autre part, il pourrait être à même capable d'identifier la pertinence des données et donc contribuer à la priorisation de leur publication. Enfin, il contribue à la classification des données en apportant son expertise en matière de la mise en œuvre de la loi n°31-13 relative au droit d'accès à l'information.
- **Responsable des données personnelles** : Le responsable des données personnelles contribue à la classification des données en apportant son expertise sur la mise en œuvre de la loi n°09-08 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel notamment dans l'identification et la gestion des données personnelles dans les données de l'inventaire.
- **Responsable de la cybersécurité** : Le responsable de la cybersécurité contribue à la classification des données en apportant son expertise en matière de la loi n°05-20 relative à la cybersécurité et l'identification des ressources de données classifiées.
- **L'entité en charge des archives** : L'entité en charge des archives contribue à l'identification des documents, informations et données qui peuvent être publiées en relation avec la loi n°69.99 relative aux archives.
- **Service Informatique, Direction des statistiques & Gestionnaires de données, ou équivalent** : Ces acteurs sont des contributeurs potentiels à l'identification des sources de données. Ils seront donc les acteurs prioritairement impliqués lors de la phase de cartographie des sources de données.
- **Analystes/statisticiens/Réutilisateurs externes/autres** : Ces acteurs sont les réutilisateurs potentiels des données, et donc les utilisateurs naturels de l'inventaire. Ils contribuent à la priorisation des publications.

Enrôlement des structures publiques

Le principe élémentaire de l'inventaire est de construire une cartographie globale et complète de l'ensemble du patrimoine de données de l'Administration marocaine. Cette cartographie globale est constituée de l'ensemble des inventaires de chaque structure publique. Il est évident que plus l'inventaire est riche plus il est utile, c'est-à-dire que les données référencées généreront plus de valeur ajoutée grâce à leurs croisements et analyses. Il est donc essentiel que toutes les structures publiques s'engagent dans la constitution de leur propre inventaire.

Cependant, il est également important de considérer deux points :

1. Comme toute innovation ou toute révolution notamment celle relative au secteur numérique, les structures publiques ne réagiraient pas toutes de la même manière à la demande de mise en place d'un inventaire. Certaines structures, notamment les pionnières en la matière, ont probablement déjà commencé à répertorier leur

patrimoine de données et donc s'engageraient naturellement dans ce processus et présenteraient des résultats rapides. Ensuite, d'autres structures, voyant les premiers résultats des pionniers seront motivées à rejoindre le mouvement et ainsi de suite jusqu'à une éventuelle adhésion des autres structures audit mouvement. Il est essentiel de prendre compte cette dimension de motivation dans la sélection des premières structures à implémenter l'inventaire. Les structures qui n'ont pas de champions internes, ou ayant des contraintes, pourraient produire des résultats de faible qualité, même si elles ont des données importantes. Il est donc essentiel de s'appuyer en premier lieu sur les structures pionnières et de mettre en place une approche en cycles qui intègre des plaidoyers et la présentation des succès des structures pionnières auprès des structures qui ne sont pas encore engagées.

2. L'inventaire de données au sein d'une structure publique n'est qu'un outil et élément d'une stratégie globale de mise en œuvre d'une initiative Open Data au sein de cette structure. Il n'est utile et indispensable que si les agents de cette structure l'utilisent pour analyser et exploiter des données pour la prise de décision, la définition des politiques publiques, etc. Il est donc important de mettre en place au sein des structures pionnières, en même temps que l'inventaire, les autres éléments (développement de capacités techniques, conduite du changement, développement de cas d'usage) de l'initiative Open data. En d'autres termes, il conviendrait de solliciter les structures publiques pionnières pour qu'elles définissent une stratégie Open data qui intègre l'inventaire et les autres éléments. En l'absence des autres éléments, l'intérêt d'une structure publique pour mobiliser ses responsables et agents afin de mettre en place un inventaire serait faible.

Mise en œuvre de l'inventaire

Cette section présente l'approche proposée pour la mise en place de l'inventaire qui s'organise autour de trois axes :

- La définition des priorités stratégiques et la mise en place d'une structure organisationnelle ;
- La mobilisation et la sensibilisation des personnels ;
- La cartographie des données ;
- La classification des données ;
- La priorisation des données ;
- La publication de l'inventaire.

Ces étapes sont détaillées ci-dessous.

Structure organisationnelle & approche stratégique

La mise en place d'un inventaire au sein d'une structure publique nécessite de conduire des investigations au sein de toutes ses entités. Pour certaines structures de taille importante, cette approche devrait se faire étape par étape. Il conviendrait donc au premier responsable de la structure de définir, en collaboration avec le ROD, les priorités pour le déploiement de l'inventaire.

Quatre options principales peuvent être considérées :

- **Approche stratégique** : Cette approche consiste à mener une réflexion sur les priorités de la structure, ou ses défis, et identifier les sujets sur lesquels une approche « données » pourrait apporter des solutions. A partir de l'identification du sujet, il serait possible d'identifier les données nécessaires pour conduire la tâche, et orienter le développement de l'inventaire sur la recherche de ces données.
- **Approche organisationnelle** : Cette approche consiste à sélectionner un ou plusieurs directions/départements et conduire l'inventaire en leur sein. Cette démarche serait ensuite étendue aux autres sous-structures au fur et à mesure.
- **Approche pragmatique** : Cette approche consiste à focaliser les efforts d'abord sur les sources de données facilement identifiables au sein de la structure. Il s'agit par exemple de travailler avec le service informatique, la direction des statistiques, le service communication, etc. pour identifier facilement les données.
- **Approche basée sur la demande** : Cette approche consiste à mener une enquête en amont du démarrage de la tâche inventaire, et ce, auprès des réutilisateurs potentiels internes à la structure, au sein du gouvernement et auprès des acteurs non-gouvernementaux. Cette enquête permettrait de mettre en lumière les thèmes et les données pour lesquels une forte demande existe et à laquelle il convient de répondre en priorité.

Il est important de noter que ces approches ne sont pas mutuellement exclusives. Au contraire, il est recommandé de les associer pour obtenir des meilleurs résultats et structurer l'activité.

L'ordre de présentation ci-dessous représente l'articulation optimum de la structuration de la tâche :

1. Dans un premier temps, et dans le cadre du plan d'action annuel Open Data de la structure, le premier responsable de la structure définit, en collaboration avec le ROD, les priorités stratégiques.
2. A partir de là, le ROD identifie les entités (directions/départements/divisions, etc.) à impliquer.
3. Ensuite, au sein de chaque entité, l'approche pragmatique se concentrerait d'abord sur les sources de données facilement identifiables avant d'étendre l'activité à toute la structure.
4. Enfin, et en parallèle, des consultations avec les réutilisateurs permettraient de structurer la recherche sur des données spécifiques.

Conformément à ses missions, le ROD est le responsable de la coordination de la tâche inventaire. Cependant, il est fortement recommandé de mettre en place un groupe de travail multi-expertise (légal, technique, etc.) pour organiser et implémenter la tâche. Chaque membre du groupe sera en charge d'une partie de la cartographie, et l'ensemble du groupe sera en charge des tâches de classification et du catalogage des données. En fonction de l'ampleur de la tâche, et de la volonté du premier responsable de la structure de couvrir rapidement un nombre important d'entités au sein de sa structure, il serait nécessaire de

mettre en place plusieurs sous-groupes de travail ; un sous-groupe au sein de chaque direction par exemple ; avec un représentant de chaque sous-groupe au sein du groupe de travail global de la structure, coordonné par le ROD.

Sensibilisation et mobilisation

Une fois que l'approche est finalisée et que le ou les groupes de travail est/ sont identifié(s), la première étape consisterait à mobiliser et sensibiliser l'ensemble des responsables et agents de la structure qui vont être impliqués.

Cette sensibilisation prendra plusieurs formes :

- **Une approche formelle** : Il est recommandé au premier responsable de la structure publique de sensibiliser les collaborateurs concernés à l'importance de la tâche pour la structure (réunion formelle, note de service...) et de nommer officiellement (note de service...) les membres du ou des groupes de travail ainsi que leurs missions et prérogatives.
- **Une approche didactique** : En dehors de l'aspect formel, il est également recommandé de sensibiliser les responsables et agents impliqués sur l'intérêt de l'inventaire, les opportunités et bénéfices, et l'intérêt des approches Open data. La compréhension de l'exercice et son apport potentiel sont des éléments motivants pour les responsables et agents et favorisent leur implication et contribution.

Processus de cartographie

Une fois la sensibilisation des personnels effectuée, la cartographie des sources de données² pourra débuter.

Dans chaque entité impliquée, il est recommandé d'adopter une approche pragmatique :

1. Cartographier les sources de données facilement identifiables :
 - a. Les systèmes d'information et e-services ;
 - b. Les données déjà publiées (statistiques...) et leurs sources ;
 - c. Les annuaires (services publics, habilitations...) ;
 - d. Les données interministérielles (budget, ODD, marchés publics...) ;
 - e. Les données utilisées pour les rapports, les prises de décision, et le suivi & évaluation ;
 - f. Les données mentionnées dans des accords de partage ;
 - g. Les demandes d'accès à l'information.
2. Cartographier ensuite le reste des données digitales actives qui existent (démarche bureau-par-bureau).
3. Cartographier enfin les données historiques et les données sous format papier.

La cartographie des données consiste à relever un ensemble d'informations (métadonnées) sur les données. La liste des métadonnées prioritaires et des métadonnées recommandées est décrite dans la section « Les métadonnées à répertorier ».

²Pour rappel, les données à cartographier sont toutes les données primaires de la structure, publiques ou non, qui sont collectées par la structure.

Il serait probablement difficile de renseigner certains champs ; ou d'autres champs nécessiteraient des activités complémentaires (par exemple le dictionnaire de données ou le schéma de validation). Les informations non disponibles ou qui nécessitent ces actions particulières devraient être notées. Dans un premier temps, l'attention devrait être portée sur les métadonnées obligatoires et ensuite sur les métadonnées recommandées.

Concernant les sections « Classification des données » et « Ouverture du jeu de données », ces éléments pourraient être discutés avec le gestionnaire de la source de données et renseignés si l'information existe. Dans le cas contraire, ces points sont traités dans les étapes suivantes.

Classification des données

La classification des données est le processus qui permet de définir qui peut accéder à quelles données.

L'objectif de ce processus est double :

1. Identifier les données :
 - a. Qui peuvent être rendues publiques directement.
 - b. Qui peuvent être rendues publiques après expurgation ou anonymisation.
 - c. Qui nécessitent des habilitations spécifiques pour pouvoir être consultées (données à accès restreint).
2. Identifier la liste des traitements à appliquer pour que les données de catégorie (b) puissent être rendues publiques.

NB : Les données des catégories (a) et (b) (après traitement) entrent dans le cadre de l'Open Data. Les données des catégories (b) (sans traitement) et (c) sont répertoriées dans l'inventaire parce qu'elles peuvent être exploitées par les responsables et agents habilités en utilisant les mêmes techniques.

Au Maroc, la classification des données est définie principalement par 3 législations :

- **La Loi n°05-20 relative à la cybersécurité**³ (et son décret d'application n°2-21-406 du 15 juillet 2021) qui définit la classification des actifs informationnels en fonction de l'impact potentiel des incidents de cybersécurité.
- **La Loi n°31-13 relative au droit d'accès à l'information**⁴ qui définit, de façon précise, les informations publiques qui peuvent être publiées sur demande ou de façon proactive.
- **La Loi n°09-08 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel**⁵ qui fournit un cadre réglementaire pour la protection de la vie privée et identifie en particulier les données qui rentrent

³ <https://www.dgssi.gov.ma/fr/content/loi-ndeg-0520-relative-la-cybersecurite.html#:~:text=La%20loi%2005%2D20%20relative,des%20collectivit%C3%A9s%20territoriales%2C%20des%20%C3%A9tablissements>

⁴ <https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/107094/131706/F-985611108/MAR-107094.pdf>

⁵ <http://adala.justice.gov.ma/production/html/Fr/156780.htm>

dans le cadre de cette loi et les traitements qui doivent être appliqués avant publication des données.

Le processus de classification des données devrait être effectué en tenant compte des dispositions y afférentes de chaque loi, et le cas échéant, de ses textes d'application.

Si à l'issue de ce processus, il ressort que les données doivent être anonymisées, la sélection et l'application de traitements d'anonymisation spécifiques devra être discuté avec des responsables/ agents ayant les capacités appropriées sur ces sujets, et si nécessaire, avec les autorités compétentes chargées de veiller à la mise en œuvre de ces lois.

Priorisation des données

Le principe de la priorisation des données est d'identifier parmi toutes les données référencées dans l'inventaire, les plus pertinentes, c'est-à-dire celles qui ont le plus gros potentiel de valorisation. L'objectif de cette priorisation est de guider la structure dans la focalisation de son effort de publication de données. Elle permet à la structure de savoir sur quels jeux de données elle doit porter son effort de publication pour maximiser l'impact de la mise en place de l'inventaire. Ce critère de potentiel de valorisation ou potentiel d'impact est subjectif et il est défini en fonction du nombre de réutilisateurs potentiels au sein de la structure, du gouvernement ou parmi les acteurs non-gouvernementaux. Dans la section « Métadonnées », paragraphe « Ouverture du jeu de données », une méthode est proposée pour calculer un score de priorité qui prend en compte à la fois la demande pour le jeu de données, la qualité des données et l'effort nécessaire à fournir pour publier ce jeu sous format ouvert. Le calcul de ce score nécessiterait une analyse approfondie des différentes métadonnées collectées sur le jeu de données, et des échanges avec les réutilisateurs potentiels.

Publication de l'inventaire

Comme expliqué dans l'introduction de ce document, l'objectif de l'inventaire est de servir d'annuaire pour l'ensemble des responsables/ agents concernés de la structure et pour tous les réutilisateurs potentiels de données internes ou externes à l'Administration. Il est donc essentiel que les données de l'inventaire soient mises à disposition des utilisateurs au travers d'un outil qui favorisera la recherche et l'identification de jeux de données spécifiques. Cette mise à disposition nécessite deux étapes :

1. **Validation de l'inventaire** : Dans un premier temps et pour que l'inventaire puisse servir de référence administrative, il est primordial que les données de l'inventaire et en particulier les informations de classification soient validées par le premier responsable de la structure qui en assure la responsabilité. De cette manière, tout agent qui reçoit une requête de données ou a l'intention de publier des données peut se référer à l'inventaire pour connaître cette classification.
2. **Mise à disposition au travers d'une plateforme appropriée** : Une fois l'inventaire validé (et à chaque modification qui nécessite une nouvelle validation, c'est-à-dire à chaque fois qu'une information de classification change⁶), il peut être publié et mis à

⁶ Insertion d'une nouvelle entrée dans l'inventaire ou changement de classification d'une entrée

disposition des réutilisateurs potentiels de données. L'objectif de l'inventaire étant de servir d'annuaire, la plateforme hébergeant ces données d'inventaire doit fournir tous les outils adéquats pour faciliter la recherche et l'identification des différents enregistrements qui composent l'inventaire. En l'absence d'une plateforme dédiée, il est recommandé de rassembler toutes les données d'inventaires validées dans un jeu de données et le publier.

Interaction avec les réutilisateurs

L'inventaire de données est un outil important d'interaction avec les réutilisateurs notamment en ce qui concerne la priorisation des publications, ou des demandes d'éclaircissement sur la documentation ou le niveau de classification. Les contributions des réutilisateurs sur les données qu'ils recherchent sont également des informations précieuses qui facilitent la recherche et la cartographie des sources de données. De ce fait, il serait essentiel de s'assurer que la plateforme retenue pour la publication de l'inventaire supporte ces interactions et offre a-minima un service de requête de données et un service de commentaire sur l'inventaire.

Les métadonnées à répertorier

Les métadonnées marquées avec un astérisque simple (*) sont les métadonnées proposées comme devant être obligatoirement renseignées.

Les métadonnées marquées avec un astérisque double (**) sont les métadonnées qu'il est fortement recommandé de fournir à la fois en français et en arabe. Ces métadonnées pourraient également être fournies en amazigh et en anglais pour faciliter la recherche des jeux de données aux réutilisateurs utilisant ces langues comme langues primaires de recherche.

Les métadonnées marquées avec un astérisque triple (***) concernent les métadonnées qui devront être fournies par sélection d'une valeur dans un ensemble de valeurs et la liste des valeurs prédéfinies sera traduite dans toutes les langues.

A notre qu'il est essentiel dans un premier temps de renseigner les métadonnées obligatoires. Dans un deuxième temps, les autres métadonnées pourraient être renseignées au fur et à mesure. Par exemple, le champ sur les traitements à appliquer aux données avant publication pourrait être renseigné lors de la première publication lorsque ces traitements sont développés et mis en œuvre.

Informations sur le jeu de données

- **Importance** : essentielle pour le référencement.
- **Fréquence** : dans tous les inventaires.
- **Contenu**
 - *Nom/Titre* (*) (**): Nom du jeu de données ;
 - *Description* (*) (**): Brève description du jeu de données ;
 - *Secteur* : Secteur concerné par les données (exemple : élevage, primaire, secondaire...);

- *Mots-clés/Thème (*) (***)* : Mots-clés ou thèmes illustrant le jeu de données ;
- *Date de création (*)* : Date de la première collecte ou production du jeu de données ;
- *Date de fin de support/Clôture* : Date à partir de laquelle les données ne sont plus collectées ou produites ;
- *Date de dernière mise à jour des données (à la date de la dernière mise à jour de l'inventaire)* : Date de la dernière collecte ou production du jeu de données ;
- *Date de publication* : Date à laquelle le jeu de données a été ou sera publié pour la première fois sur le portail national de données ouvertes ou sur le portail de la structure concernée ;
- *URL* (si le jeu de données est publié sur le portail national de données ouvertes ou un portail sectoriel) : URL générique sur le portail en question qui permet d'accéder à la description du jeu de données.

Informations sur le producteur de données

- **Importance** : essentielle pour faciliter les contacts entre utilisateurs et producteurs.
- **Fréquence** : dans tous les inventaires.
- **Contenu** :
 - *Organisation en charge (*) (***)* : Structure publique (Ministère/département...) ayant la responsabilité du jeu de données ;
 - *Nom et coordonnées du responsable métier (*)* : Coordonnées du responsable administratif du jeu de données ;
 - *Nom et coordonnées du responsable technique (*)* : coordonnées du gestionnaire technique des données (la personne qui vérifie la qualité et gère la publication).

Information sur les données

- **Importance** : essentielle pour l'exploitation des données.
- **Fréquence** : beaucoup des attributs ci-dessous apparaissent dans les métadonnées des jeux de données publiés.
- **Contenu** :
 - *Couverture géographique (*)* : portée géographique des données dans le jeu de données.
 - *Niveau de désagrégation (*)*
 - *Géographique* : Niveau administratif minimum (national, régional, provincial/préfectoral, communal...) présent dans le jeu de données ;
 - *Autre* : liste des critères de désagrégation autre que géographique présent dans le jeu de données (genre, tranche d'âge, catégorie sociale professionnelle...);
 - *Couverture temporelle (*)* : portée temporelle des données dans le jeu de données ;

⁷Pour le cas du Maroc, chaque structure publique a probablement déjà son nom traduit dans les langues principales. Il suffit donc de fournir une des traductions, les autres pouvant être automatiquement récupérées et stockées dans le descriptif de chaque structure publique.

- *Couverture historique* : liste des différentes versions historiques du jeu de données. Par exemple, un jeu de données peut contenir des informations sur une année donnée (couverture temporelle : annuelle) et peut être disponible pour une période d'années (couverture historique) ;
- *Langues disponibles (*)* : Langue(s) dans laquelle ou lesquelles le jeu de données est disponible ;
- *Format de stockage* : Format sous lequel est stocké le jeu de données : format papier, format digital (fichier) ou système d'information (inclus les bases de données) (*)
 - Si les données sont stockées sous format digital, indiquer les informations suivantes :
 - *Format du fichier* : Excel, Geojson....
 - Si les données sont stockées dans un système d'information, indiquer les informations suivantes :
 - *Fonctionnalité d'export disponible* : Est-ce que le système d'information dispose de fonctionnalités d'export des données (Oui/Non) ;
 - *Export configurable sur les champs, les colonnes, les formats d'export* : Décrire les caractéristiques d'export des données (s'il est possible par exemple de sélectionner les champs, les colonnes, les formats d'export...).
- *Emplacement de stockage* : emplacement physique des données ;
- *Processus d'extraction automatique en place* : Indiquer, en cas de publication du jeu de données sur un portail de données ouvertes, si un processus automatique de publication est en place lorsque les données sont mises à jour (Oui/Non/Non applicable) ;
- *Dictionnaire de données* : Description et typage des colonnes et des indicateurs/variables ainsi que les nomenclatures et référentiels utilisés ;
- *Schéma de validation* : Description sous le format Table Schema⁸ du dictionnaire de données pour automatiser la validation formelle des données ;
- *Format lisible par une machine/données structurées (*)* : est-ce que les données (données stockées dans un fichier ou données exportées d'un système d'information) correspondent à la définition de données structurées lisibles par une machine ? (Oui/Non/NA) ;
- *Taille des données* : Taille en méga-octet du jeu de données ;
- *Nombre d'enregistrements* : Nombre d'enregistrement dans le fichier (par exemple nombre de lignes dans un fichier tabulaire en dehors de la ligne d'entête de description) ;

Informations sur la collecte de données

- **Importance** : Les informations sur le cycle de collecte sont essentielles pour évaluer l'âge et l'état de mise à jour des données.

⁸ <https://specs.frictionlessdata.io/table-schema>

- **Fréquence** : dans beaucoup d'inventaires.
- **Contenu** :
 - *Périodicité de collecte/production (*)* : Périodicité à laquelle les données sont collectées/produites (quotidiennement, hebdomadairement, annuellement, tous les 5 ans...);
 - *Date de disponibilité des données* : Date à partir de laquelle les données sont prêtes à être utilisées ou publiées après leur collecte/production et le processus de validation. Par exemple si les données sont collectées annuellement, elles peuvent être disponibles à partir du mois de mars de l'année suivante ;
 - *Méthode de collecte / production des données* : Toute information sur la manière de collecte ou production des données : définition d'un échantillon statistique, déclaratif, automatique (capteurs).

Classification des données

- **Importance** : essentielle pour appuyer les responsables/ agents en charge de la publication.
- **Fréquence** : Dépend de la portée de l'inventaire (inutile si l'inventaire ne concerne que les données publiques).
- **Contenu** :
 - *Classification selon la loi n°05-20 (*)* : comment sont classifiées les données au regard de cette loi : Très secret / Secret / Confidentiel Diffusion restreinte / Non classifié.
 - *Niveau d'accès selon la loi n°31-13 (*)* : Restriction de publication des données au regard de cette loi : Public / Public après traitement/ Accès restreint.
 - *Justification du niveau d'accès selon la loi n°31-13 (*)* : Justification de la classification au regard de cette loi.
 - *Statut de validation du niveau d'accès (*)*: est-ce que la classification a été validée par le responsable de la structure ?
 - *Licence de réutilisation (*)* : quelle licence de réutilisation de données est associée aux données (inclut accès payant/gratuit) ?

Ouverture du jeu de données

- **Importance** : Aide à la priorisation des données.
- **Fréquence** : Informations qui sont prises en compte dans les inventaires les plus récents.
- **Contenu** :
 - *Traitements avant publication* : Documentation des traitements à appliquer aux données avant publication (avec références aux scripts et autres codes utilisés):
 - Extraction des données.
 - Anonymisation : est-ce qu'une anonymisation est requise et quelle technique est appropriée ?

- Transformation en données structurées lisibles par une machine.
 - Transcodage du jeu de données en format ouvert.
 - Transcodage des données en utilisant les référentiels, nomenclatures et standards de données sélectionnées.
- *Qualité des données* : ce champ contient un score reflétant la qualité des données. Le score sur 20 est à calculer selon le tableau suivant :

Critère	Score
Données lisibles par une machine	0 ou 3
Données existantes sous format ouvert	0 ou 3
Données mises à jour suivant la périodicité	0 ou 3
Données désagrégées	0 si les données sont agrégées 1 si les données sont désagrégées géographiquement au niveau régional 2 si les données sont désagrégées géographiquement au niveau provincial 3 si les données sont désagrégées géographiquement au niveau communal 4 si les données sont désagrégées sur une autre dimension (genre...) ou si ce critère ne s'applique pas (par exemple des données textuelles comme un texte de loi)
Dictionnaire des données est fourni et est complet	0 ou 3 Le score de 3 est donné si ce critère ne s'applique pas (par exemple des données textuelles comme un texte de loi)
Existence d'un schéma de validation	0 ou 2 Le score de 2 est donné si ce critère ne s'applique pas (par exemple des données textuelles comme un texte de loi)
Données utilisant les standards techniques	0 ou 1 Le score de 1 est donné si ce critère ne s'applique pas (par exemple des données textuelles comme un texte de loi)
Données utilisant des référentiels	0 ou 1 Le score de 1 est donné si ce critère ne s'applique pas (par exemple des données textuelles comme un texte de loi)

- *Accessibilité des données* : ce champ contient un score reflétant la facilité de publication des données. Le calcul d'un score sur 10 se fait suivant le tableau suivant :

Critère	Score
Les données sont disponibles sous format digital	0 si les données sont sous format papier 1 si les données sont sous format digital (Excel...) 2 si les données sont sous format digital et sont mises à jour automatiquement 3 si les données sont dans un système d'information
Le système d'information offre des fonctionnalités d'extraction automatique	0 si l'obtention des données sont difficiles (numérisation, ...) 1 si les données peuvent être récupérées facilement manuellement (fichier Excel...)

	2 si la récupération des données peut être automatisée (Le système d'information offre des fonctionnalités d'extraction, les fichiers Excel sont générés automatiquement...)
Les fonctionnalités d'extraction permettent d'obtenir des données structurées lisibles par une machine	0 ou 1 Le score de 1 est donné si les données sont nativement lisibles par une machine, ou les fonctionnalités d'extraction des données permettent de générer des données lisibles par une machine
Les fonctionnalités d'extraction permettent d'obtenir des données répondant aux standards techniques	0 ou 1 Le score de 1 est donné si les données utilisent nativement les standards techniques de données ou les fonctionnalités d'extraction des données permettent de générer des données qui utilisent les standards ou ce critère n'est pas applicable au jeu de données (cas d'un texte de loi par exemple)
Les fonctionnalités d'extraction permettent de générer les métadonnées importantes (dictionnaire, schéma de données...)	0 ou 1 Le score de 1 est donné si les données sont nativement générées avec un ensemble de métadonnées
Les données ne nécessitent pas de traitement d'anonymisation	0 ou 2 Le score de 2 est donné si aucun traitement d'anonymisation n'est requis

- *Valeur du jeu de données* : ce champ contient un score reflétant le potentiel de valorisation et de réutilisation du jeu de données. Le calcul d'un score sur 20 se fait suivant le tableau suivant :

Critère	Score
Les données sont utilisées au sein de la structure	0 Les données ne sont pas utilisées 1 Les données sont principalement utilisées par le gestionnaire des données 2 Les données sont partagées largement au sein de la structure concernée 4 les données sont utilisées dans des processus de rapportage réguliers importants de la structure
Les données sont utilisées/demandées au sein de l'Administration	0 Les données ne sont pas utilisées/demandées en dehors de la structure concernée 2 Les données sont partagées avec des structures publiques 5 Les données font partie des échanges régulés avec d'autres structures 8 Les données sont fournies à des organisations internationales dans le cadre d'accords nationaux ou internationaux
Les données sont utilisées/demandées par des acteurs non-gouvernementaux	0 Les données ne sont pas utilisées/demandées en dehors de l'Administration 2 points par catégorie de demandeurs <ul style="list-style-type: none"> • Acteurs de l'innovation (startups) et secteur privé • Société civile • Recherche • Média

Calendrier d'exécution d'un inventaire de données

L'inventaire de données d'une structure publique est un outil dynamique et vivant qui évolue en permanence en fonction de nouvelles collectes ou de productions de données, ou leur arrêt, ou la découverte de nouvelles sources de données non identifiées précédemment. Les processus décrits dans ce document ont pour objectif de développer une première version exhaustive de l'inventaire.

Dans un deuxième temps, dès que l'ensemble des personnels sont sensibilisés et formés, et une fois qu'un premier inventaire complet des données est finalisé, la phase de régime de croisière de l'inventaire démarrera où les entrées seront mises à jour en fonction de l'évolution des sources de données.

Pour les petites structures (quelques dizaines d'agents), il serait possible de conduire entièrement la phase de déploiement sur un délais court. Mais pour la majorité des structures publiques, la mise en place d'un premier inventaire exhaustif des données nécessiterait plus d'efforts et de ressources. De ce fait, il est recommandé de conduire cette phase en cycles successifs pour qu'à la fois les responsables/ agents et la structure publique puissent voir l'intérêt et l'impact de cette tâche. Il est donc recommandé de diviser la réalisation de l'inventaire en étapes (voir la sous-section Structure organisationnelle & approche stratégique) et de conduire les étapes successives au cours de plusieurs cycles annuels et en incluant en parallèle la publication de données, le développement de cas d'usage de réutilisation. Un cycle prendrait donc la forme suivante :

1. Définition d'un objectif stratégique global pour un plan d'action annuel ;
2. Réalisation de la partie de l'inventaire relative à l'objectif stratégique ;
3. Publication de données en fonction de l'objectif stratégique et de l'inventaire ;
4. Développement de cas d'usage (internes et externes) qui démontrent l'intérêt et l'impact de l'inventaire et de la publication de données.

Concernant le calendrier d'exécution de l'inventaire dans un cycle, il est proposé de mettre en place un calendrier court :

Activités	S1 ⁹	S2	S3	S4	S5	S6	S7	S8
Constitution d'un groupe de travail								
Identification des cibles								
Sensibilisation et mobilisation des cibles								
Cartographie, classification et documentation des données								
Priorisation des données								
Publication de l'inventaire								

⁹ S=Semaine