

Atelier de formation sur l'Open Data au profit des organismes et institutions publics relevant de la Région Guelmim Oued Noun

Data.gov.ma

Session de formation sur l'Open Data

Ordre du jour:

Séance 1: Concepts et principes de l'Open Data

Séance 2: Plateformes Open Data

Séance 3: Standards techniques de l'Open Data

Séance 4: Inventaire, classification et référentiels

Séance 5: Cas pratique

Séance 6: Défis de mise en oeuvre et solutions

Session de formation sur l'Open Data

Séance d'ouverture

9h30-9h40

Session de formation sur l'Open Data

Séance 1: Concepts et principes de l'Open Data

9h40-10h10

Session 1: Concepts de l'Open Data

Introduction

Une vision Royale

Discours Royal à l'occasion de l'ouverture de la première session de la cinquième année législative de la onzième législature

La grande transformation que nous cherchons à réaliser au niveau du développement territorial exige un changement concret dans les **mentalités et les méthodes de travail**, ainsi qu'une véritable culture des résultats, fondée sur **des données de terrain précises** et l'utilisation des technologies numériques.

الخطاب الملكي السامي بمناسبة افتتاح الدورة الأولى من السنة التشريعية الخامسة من الولاية التشريعية الحادية عشر

"فالتحول الكبير، الذي نسعى إلى تحقيقه على مستوى التنمية الترابية، يتطلب تغييرا ملموسا في العقليات، وفي طريقة العمل، وترسيخا حقيقيا **لثقافة النتائج؛ وذلك بناء على معطيات ميدانية دقيقة**، وباستعمال **التكنولوجيات الرقمية**".



صَلَّيْهِ وَسَلِّمْ وَنَحْمُهُ وَرَحْمَةُ اللَّهِ

Session 1: Concepts de l'Open Data

Introduction

Le paradoxe de Simpson

Ici exemple scolaire, mais applicable à différentes études statistiques (vaccins, traitements médicaux, sport, marketing, sociologie, ...)



Lycée A



350 élèves



273 ont réussi
Taux de réussite:

78%



Lycée B



350 élèves



289 ont réussi
Taux de réussite:

83%

Quel est le meilleur lycée, d'un point de vue de réussite académique?

Session 1: Concepts de l'Open Data

Introduction

Le paradoxe de Simpson

Ici exemple scolaire, mais applicable à différentes études statistiques (vaccins, traitements médicaux, sport, marketing, sociologie, ...)



Lycée A



350 élèves

87 Economie
263 Science



273 ont réussi
Taux de réussite:

78%

Filière Economie: 81 (/87)
Filière Science: 192 (/263)

93%
73%



Lycée B



350 élèves

270 Economie
80 Science



289 ont réussi
Taux de réussite:

83%

Filière Economie: 234 (/270)
Filière Science: 55 (/80)

87%
69%

Session 1: Concepts de l'Open Data

Introduction

Le paradoxe de Simpson

Ici exemple scolaire, mais applicable à différentes études statistiques (vaccins, traitements médicaux, sport, marketing, sociologie, ...)



Analyses scolaire

Taux de réussite
Taux d'admission



Vaccins & épidémiologie

Taux de létalité COVID-19
Efficacité des vaccins



Démographie

Naissances, Mortalités
Analyses démographiques



Sport

Taux de victoires
Analyse de performances



Justice

Inculpation et infractions
Plaintes, affaires



Marketing

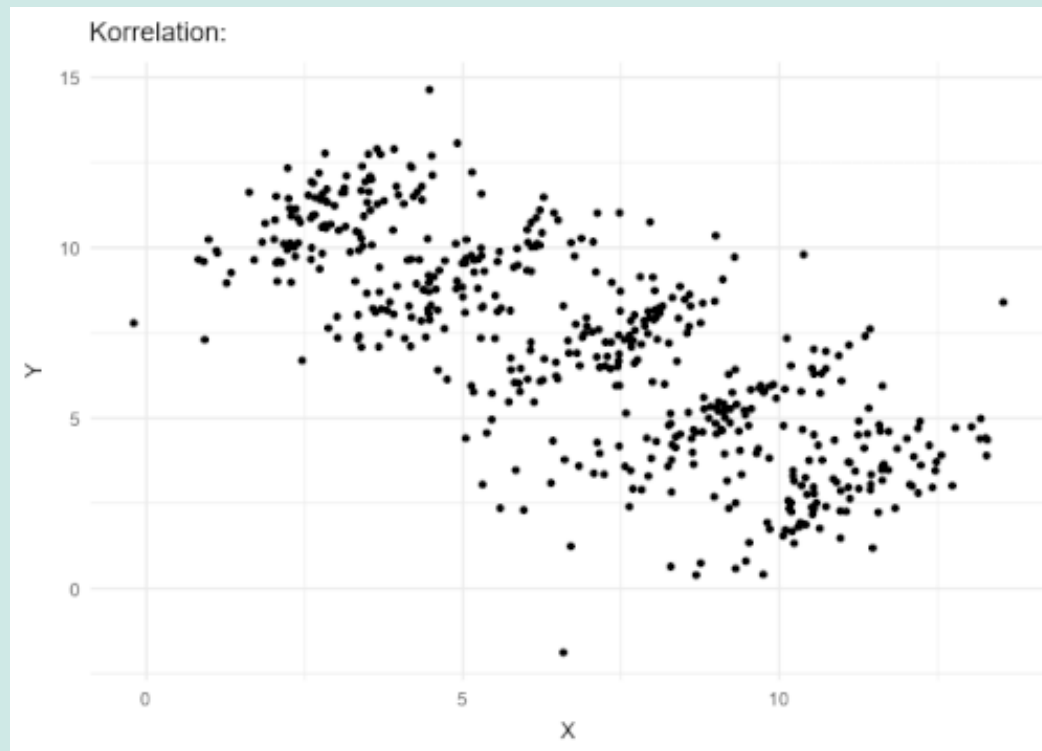
Conversion par source de
trafic
Analyse de revenus

Session 1: Concepts de l'Open Data

Introduction

Le paradoxe de Simpson

Sous forme visuelle



Session 5: Complexité d'un projet Open Data

Moment de reflexion



Data disponible

Est elle dans un système, ou archives, ou non collectée? Quel cout pour la collecter?



Data fiable

Exactitude de la donnée



Data détaillée

Le niveau de détail doit être suffisant pour permettre une analyse fiable

Exemple du paradoxe de Simpson



Data & Contexte

La donnée ouverte doit être complètement expliquée (moyen de mesure, population mesurée, données complémentaires, ..)

Exemple de l'avion



Data légale

Confidentialité, sécurité, sensibilité



Data aux bons formats

La donnée ouverte doit être complètement expliquée (moyen de mesure, population mesurée, données complémentaires, ..)

Exemple du RIB

Session 1: Concepts de l'Open Data

Introduction du formateur

Firas Ezzaki - Expert en Transformation Digitale - 20 ans d'expérience



Banques

Sécurisation du patrimoine vital de la Banque

Crédits, Garanties, Main levées, Contentieux, Moyens de paiement



Banques

KYC, Data Client, Conformité, Lutte Anti Blanchiment

Dossiers client, Justificatifs, Transactions, Courrier



Assurances & Retail

Flux indemnisation des sinistres (Auto, Santé, ..)

Dossiers de remboursement, factures, justificatifs



Retraites

Calcul des cotisations & bénéfices sociaux

Bordereaux de déclaration, états de précompte, listings en masse



Administrations

Digitalisation du patrimoine institutionnel

Fonds d'archive, registres légaux (civil, légal, juridique, foncier, financier)



Administrations

Digitalisation des procédures légales & fiscales

Déclarations légales (entreprises, Tribunaux de commerce, ...), fiscales (impôts & DGI)



Grandes Distribution

Flux Marketing Direct & Opérations de fidélité

Offres de remboursement, lutte anti fraude, programmes de fidélité



Santé

Organisation Mondiale de la Santé (Statistiques)

Certificats de naissance et décès produits par les médecins

Session 1: Concepts de l'Open Data

Maroc Digital 2030



[Source \(MTNRA\)](#)

Session 1: Concepts de l'Open Data

Maroc Digital 2030

Un objectif ambitieux



Digitalisation des Services Publics

Passer de la 113e à la 50e place mondiale sur le classement EGDI d'ici 2030, grâce à des services numériques accessibles et centrés sur l'utilisateur.



Inclusion Numérique

Offrir des services numériques simples d'utilisation pour tous les citoyens, y compris en milieu rural et pour les personnes en situation de handicap.



Dynamisation de l'économie numérique

Créer 240 000 emplois directs, générer 100 milliards de MAD pour le PIB, et positionner le Maroc comme un producteur de numérique.



Gouvernance Participative

Adopter une approche collaborative impliquant le secteur public, le privé et la société civile, avec un soutien méthodologique du Ministère de la Transition Numérique.

[Source \(MTNRA\)](#)

Session 1: Concepts de l'Open Data

Introduction à l'Open Data

“Data is the new Oil” -

(“La donnée est le nouveau pétrole”)

Clive Humby - 2006

Mathématicien britannique



Session 1: Concepts de l'Open Data

Introduction à l'Open Data

“Data is the new Oil” (“La donnée est le nouveau pétrole”) - Clive Humby - 2006



Ce que le mathématicien britannique souligne, c'est que la donnée au format brut n'a de valeur que si elle est transformée pour produire une valeur ajoutée (comme le pétrole).

La comparaison avec le pétrole est doublement pertinente:

[Source \(Data is the new oil\)](#)

Le pétrole a été l'origine de révolutions dans le transport, l'énergie, l'industrie,

La data sera à l'origine d'autres révolutions dans de nombreux domaines (IA par ex.),

Mais surtout ...

.... Sa valeur réside dans son **potentiel**, pas dans le produit brut



75\$



300\$



150 000 \$

brut

carburant

cosmétique

Session 1: Concepts de l'Open Data

Introduction à l'Open Data

Une data déjà produite

La plupart des secteurs (publics ou privés) collectent de la donnée pour offrir leurs services



Naissance



Education



Santé



**Travail &
Fiscalité**



Justice

Données collectées :

nom, date et lieu de naissance, filiation (parents)

Données collectées :

nom de l'élève, âge, adresse, parcours scolaire, résultats académiques

Données collectées :

antécédents médicaux, traitements reçus, couverture sociale

Données collectées :

revenu, profession, contributions fiscales, cotisations sociales.

Données collectées :

casiers judiciaires, infractions, plaintes déposées

Data isolée vs. Information de valeur: Evolution de la capacité de décision

Donnée brute

Data isolée: (brute)

Exemple: Nombre de consultations pour des symptômes grippaux dans un hôpital

Information

Data combinée dans le même secteur:

Exemple: Donnée précédente combinée à

- + **Consultations pour symptômes grippaux** (plusieurs hôpitaux & régions)
- + **Vente de médicaments contre la grippe**

Data combinée dans d'autres secteurs:

Exemple: Donnée précédente combinée à

- + **Education:** Taux d'absentéisme dans les écoles
- + **Economie:** Données sur les perturbations dans les entreprises (arrêts maladie, baisse de productivité).
- + **Météorologie:** Données sur l'évolution climatique (températures basses, humidité élevée)

Session 1: Concepts de l'Open Data

Introduction à l'Open Data

L'Open Data, simplement

La définition de l'Open Data est très simple. Le challenge consiste à le mettre en oeuvre sur la durée, et selon les bonnes pratiques

DATA

Une donnée représente **une information brute, sans traitement.**

Exemple : date, montant, pourcentage, adresse, etc...

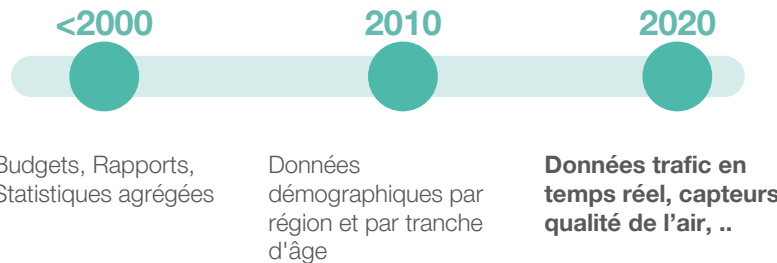
Son exploitation et analyse va permettre de produire de l'information

Exemple : Analyses, statistiques, utilisation ...

OPEN DATA

“Pratique de rendre les données publiques **disponibles** et **accessibles**, de **manière libre**, **gratuite** et **réutilisable**”

Une Open Data qui a évolué dans le temps:



Session 1: Concepts de l'Open Data

Introduction à l'Open Data

Les débuts de l'Open Data (Années 50-60)

Signature de la FOIA

1966 par le président Johnson



Né d'une **pression cumulative** :

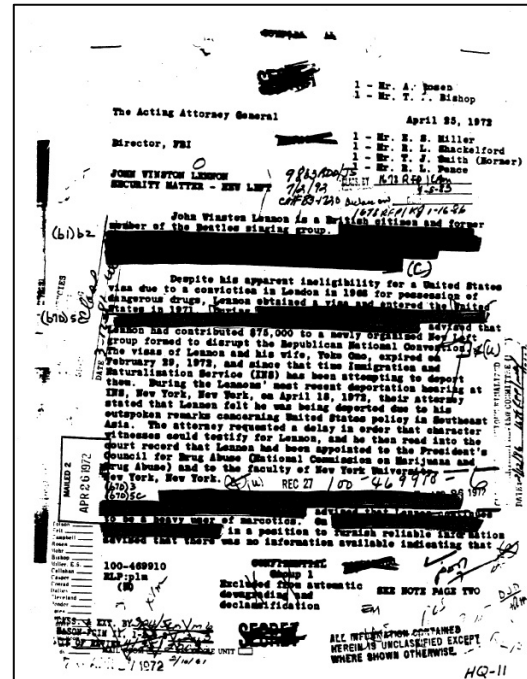
- abus de secret d'État,
- peur d'un gouvernement incontrôlé,
- volonté de la presse et du Congrès d'instaurer un **droit démocratique à l'information**.

✓ **Transparence pour les droits civiques**

✗ **Pas pour l'innovation**

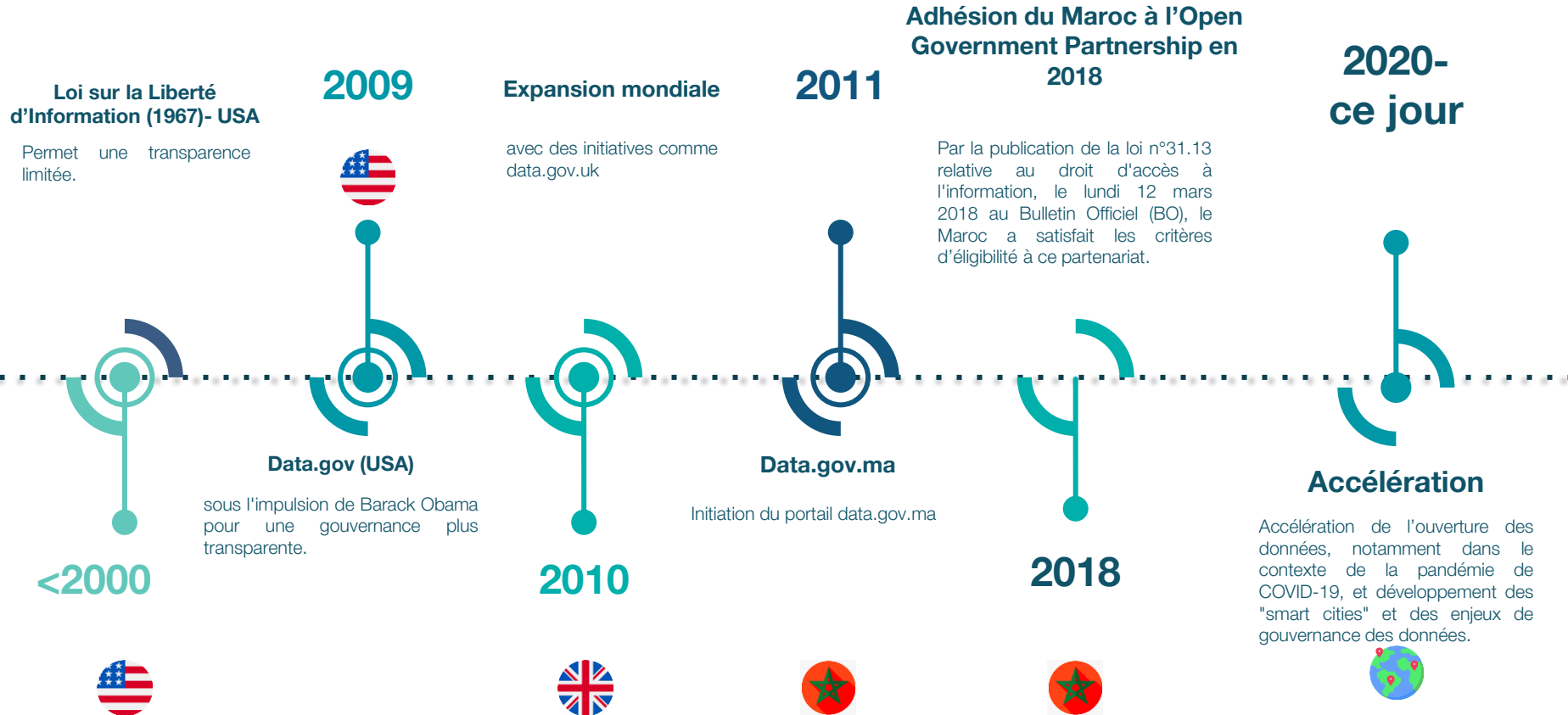
✓ **Destiné aux journalistes**

✗ **Réactive et non proactive**



Session 1: Concepts de l'Open Data

Introduction à l'Open Data



Session 1: Concepts de l'Open Data

Des utilisations dans tous les domaines

Différents secteurs, différentes utilisations

L'Open Data peut être utilisée dans de nombreux domaines. Tant que la Data est disponible et ouverte, les possibilités d'innovation sont illimitées.



Gestion Urbaine

Applications de cartographie, de gestion des voies publiques, ...



Développement durable

Cartographie des stations électriques, tableaux de bord environnementaux, etc.



Mobilité & transport

Applications de mobilité en temps réel, Analyse des flux



Santé

Applications d'assistance de soins, de remboursement, de cartographie, prévisions sanitaires



Economie

Etudes & recherches économiques favorisant les investissements, Prévisions économiques



Tourisme

Applications d'assistance aux visiteurs (ex. Coupe du Monde 2030)

[Source \(Open Data Barometer\)](#)

Session 3: Bénéfices socio économiques de l'Open Data

Importance et bénéfices

Des bénéfices à tous les niveaux

La donnée ouverte confère des bénéfices en cascade, permettant d'être un véritable moteur de croissance et d'amélioration.



Transparence gouvernementale

L'Administration publique gagne en crédibilité et en confiance citoyenne en rendant sa donnée publique



Recherche & Innovation

A partir de la data ouverte, les startups, chercheurs et acteurs de l'innovation peuvent développer de nouvelles solutions et services



Optimisation du service public

Grâce à l'Open Data, l'administration peut optimiser ses ressources et améliorer sa qualité de service, en recourant à l'innovation sur la data

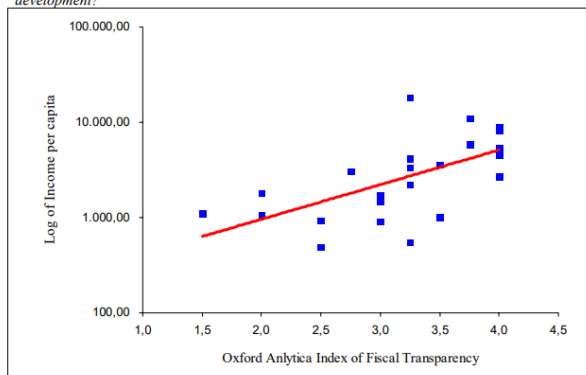
Session 3: Bénéfices socio économiques de l'Open Data

Transparence

La transparence, bien plus qu'une valeur morale

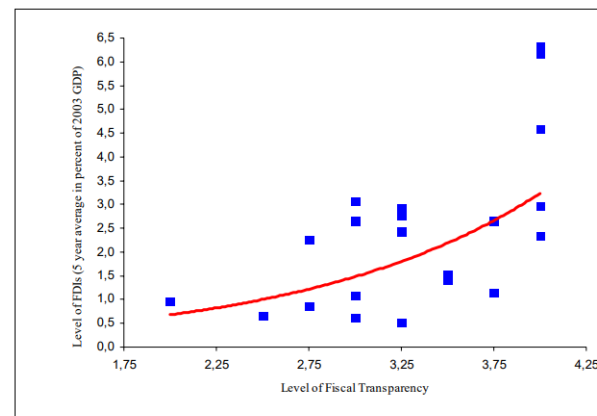
Lien transparence - richesse

Figure 2: Fiscal Transparency and GDP per capita: Is Transparency a pattern of development?



Data: Oxford Analytica

Lien transparence - Investissements



Data: Oxford Analytica and FDI-inflows from UNCTAD

[Lien](#)

Session 3: Bénéfices socio économiques de l'Open Data

Transparence



**Les cuisiniers
s'appliquent**



**Les clients sont en
confiance**

Session 3: Bénéfices socio économiques de l'Open Data

Importance et bénéfices

1 **Transparence** gouvernementale

Transparence



Meilleure compréhension par les citoyens des politiques publiques, des décisions administratives, et l'utilisation des fonds publics.

Bonne gouvernance



Visibilité sur la manière dont les ressources publiques sont utilisées

Engagement civique



Citoyens plus impliqués dans la vie publique



**Transparence
gouvernementale**



Recherche &
Innovation

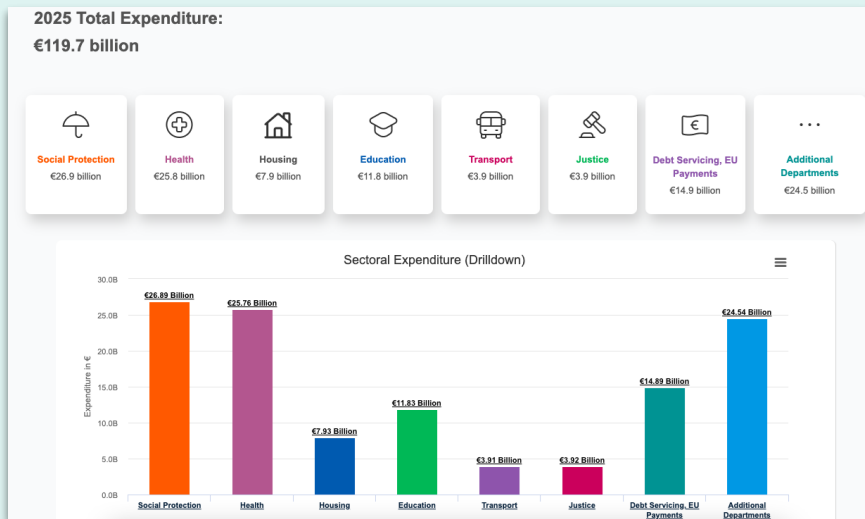


Optimisation du
service public

Session 3: Bénéfices socio économiques de l'Open Data

Exemples

1 Transparence gouvernementale (exemple)



Source ([Where your money goes - Ireland](#))

Where your money goes (Irlande): Accès aux citoyens irlandais, à la répartition des dépenses gouvernementales

- **Répartition par secteur :** Les dépenses sont détaillées par domaines tels que la protection sociale, la santé, l'éducation, le logement, les transports et la justice.
- **Données annuelles :** Les dépenses pour différentes années

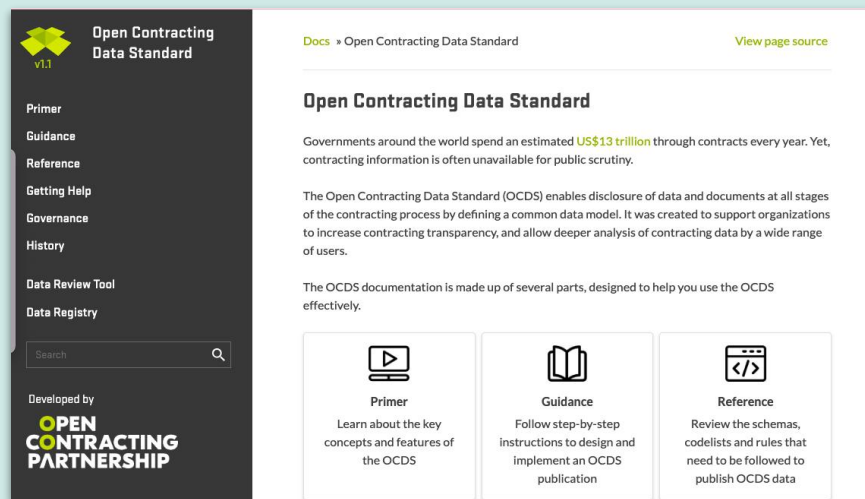
- Visibilité par les citoyens de l'usage de l'argent des impôts
- Analyse des tendance budgétaires sur une décennie

Session 3: Bénéfices socio économiques de l'Open Data

Exemples

Le cas Mexicain (OCDS)

Au Mexique, la transparence sur les marchés publics a fait l'objet d'innovation avec des plateformes offrant des analyses avancées sur les marchés publics



The screenshot shows the Open Contracting Data Standard (OCDS) website. The left sidebar contains a navigation menu with links: Primer, Guidance, Reference, Getting Help, Governance, History, Data Review Tool, and Data Registry. Below the menu is a search bar and the text 'Developed by OPEN CONTRACTING PARTNERSHIP'. The main content area is titled 'Open Contracting Data Standard' and includes a breadcrumb 'Docs » Open Contracting Data Standard' and a 'View page source' link. The text describes the OCDS as a common data model for public contracts, mentioning that governments spend an estimated US\$13 trillion annually. It also states that the OCDS documentation is designed to help users use the standard effectively. At the bottom, there are three icons representing different sections: Primer (a play button icon), Guidance (a book icon), and Reference (a code icon).

[Source \(OCDS Mexico\)](#)

Open Contracting Data Standard (Mexique, Honduras) Mise à disposition aux citoyens des données sur les contrats publics

- **Données sur le processus de passation des marchés** : L'OCDS couvre l'ensemble du cycle de vie des marchés publics, incluant la planification, la passation, l'attribution, la contractualisation et l'exécution des contrats
- **Documents clés associés** : appels d'offres, contrats signés, rapports d'évaluation et les certificats d'achèvement

- Meilleure transparence
- Vue personnalisable sur les transactions publiques

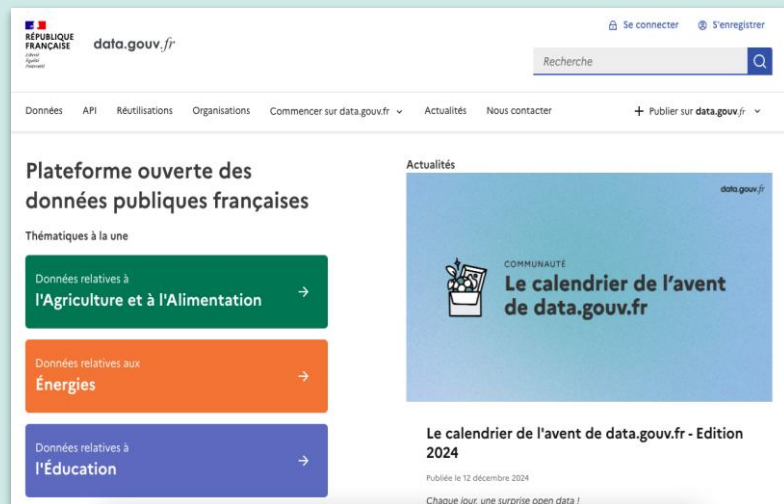
Session 3: Bénéfices socio économiques de l'Open Data

Exemples

Le cas data.gouv.fr



L'expérience en France a permis de publier des données sur un vaste champs de domaines, augmentant la transparence publique à un niveau non atteint précédemment



Source (Data France)

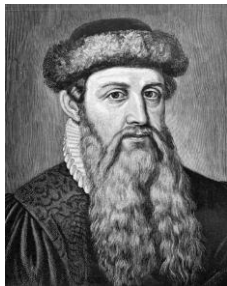
Data.gouv.fr (France) : Accès aux citoyens, entreprises et chercheurs à +30 000 jeux de données publiques.

- **Données administratives et économiques :** Budgets publics, indicateurs macroéconomiques, emploi, commerce.
- **Données géographiques et environnementales :** Cartes, données SIG, qualité de l'air,
- **Données sociales, sanitaires et éducatives :** Population, santé, logement, éducation
- **Données de transport et culturelles :** Réseaux de transport, horaires, infrastructures, patrimoine culturel,, etc.

- Suivi des dépenses publiques
- Gestion des transports en commun
- Etudes de l'empreinte carbone.
- Utilisation des données climatiques, économiques, ou sanitaires pour des études de modélisation ou d'analyse

Session 3: Bénéfices socio économiques de l'Open Data

Recherche & Innovation



Johannes Gutenberg

Inventeur de l'imprimerie
1393-1468



*A inventé bien plus qu'un outil:
A permis la circulation du savoir*



Session 3: Bénéfices socio économiques de l'Open Data

Recherche & Innovation

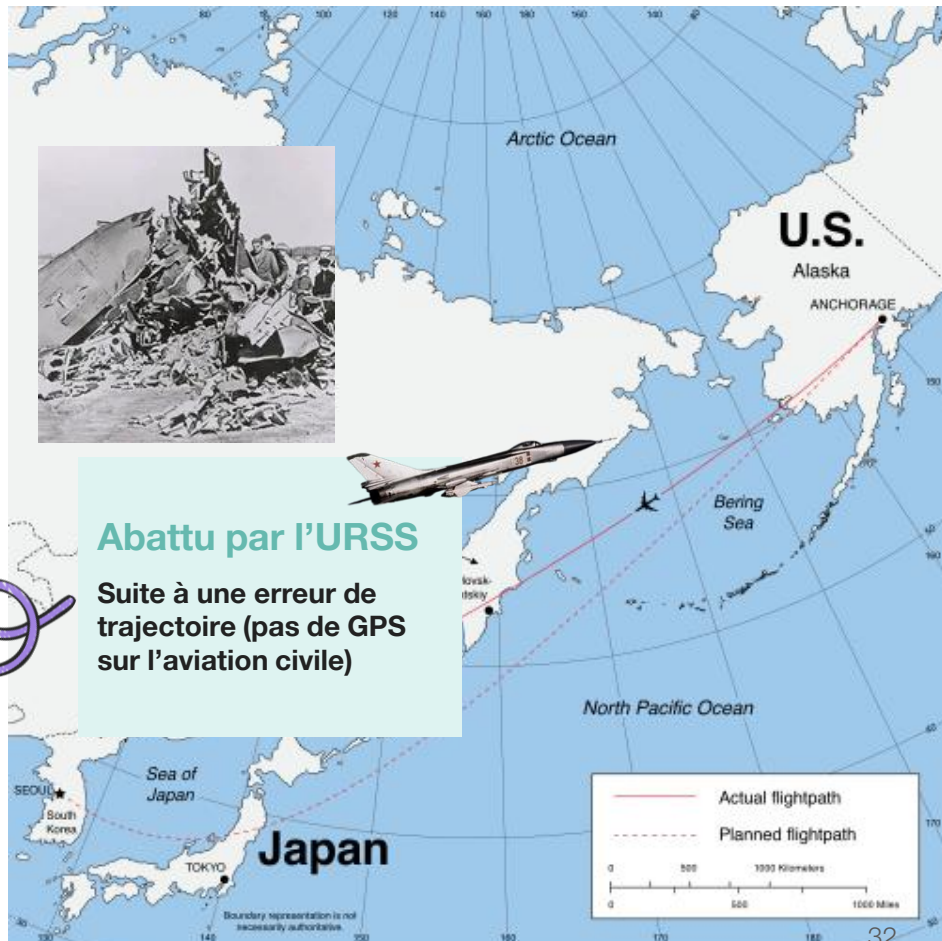


<1983

GPS: usage militaire seul

Vol Korean 007

Anchorage (Alaska) - Seoul
(Corée du Sud)



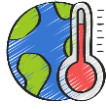
Abattu par l'URSS

Suite à une erreur de
trajectoire (pas de GPS
sur l'aviation civile)

Ouverture du GPS (1983)



Agriculture



Climat



Environnement



Urbain

Session 3: Bénéfices socio économiques de l'Open Data

Importance et bénéfices



Transparence
gouvernementale



Recherche &
Innovation



Optimisation du
service public

2

Favorisation de la recherche et de l'innovation

Accélérer la recherche



Terrain de coopération pour les chercheurs, les institutions publiques, et les entreprises qui souhaitent collaborer à des projets communs, en particulier dans les domaines de la **recherche scientifique**, de **l'environnement**, de **la santé**.

Nouveaux services et applications



Applications de mobilité, des services de santé, des solutions de gestion environnementale, etc.

Création de nouvelles entreprises



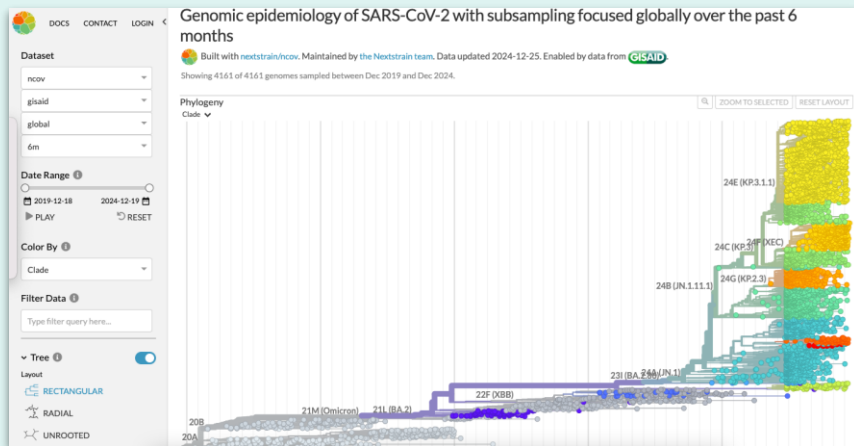
Lancement de nouvelles entreprises sans avoir à investir massivement dans la collecte de données.

Session 3: Bénéfices socio économiques de l'Open Data

Exemples

2

Favorisation de la recherche et de l'innovation (exemple)



[Source \(Next strain\)](#)

Nextstrain.org : collecte des données mondiales de la pandémie de Covid

- **Séquences génomiques** : Analyse les séquences génétiques complètes des agents pathogènes
- **Données épidémiologiques** : Les dates et lieux de collecte des échantillons
- **Métadonnées associées** : La lignée de l'agent pathogène, les mutations spécifiques présentes

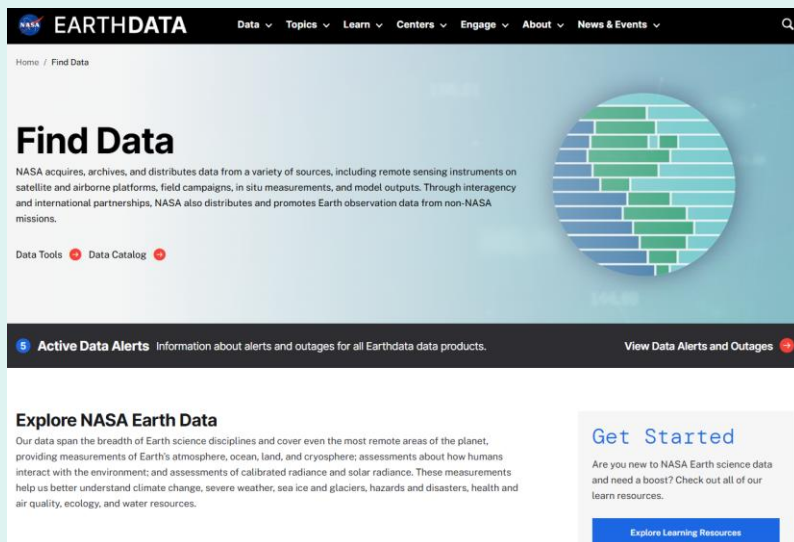
- Meilleure gestion des hôpitaux
- Interventions ciblées dans les lieux les plus contaminés
- Suivi de la diffusion géographique des souches virales
- Projection de la tendance en terme de contagions

Session 3: Bénéfices socio économiques de l'Open Data

Exemples

2

Favorisation de la recherche et de l'innovation (exemple)



[Source \(NASA\)](#)

NASA (Earth Data)

- **Catalogue unifié** : Mesures d'observations de la Terre (satellites, avions, capteurs)
- **Accès libre**

- Permet à chercheurs externes et étudiants de reproduire analyses physiques (ex. découverte du boson de Higgs)
- Alimente la recherche en data science haute performance
- Sert de support pédagogique dans des milliers d'universités

Session 3: Bénéfices socio économiques de l'Open Data

Amélioration du service public



Session 3: Bénéfices socio économiques de l'Open Data

Amélioration du service public

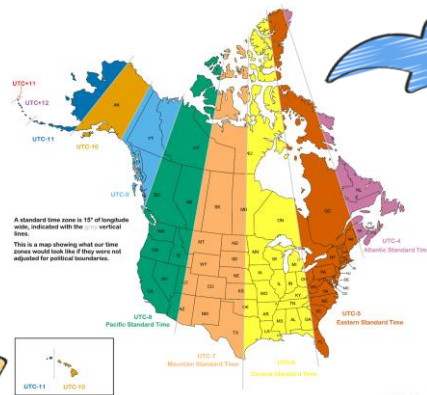
L'invention du train a permis la standardisation du Temps



Avant le train
Chaque ville avait sa propre heure (soleil)



Généralisation du train
Pontualité, flux passagers, risques d'accidents:
⇒ **nécessité de standardiser le temps**



1883
1er fuseau horaire (USA et Canada)
Par les compagnies de chemin de fer



1884
Adoption du GMT
Greenwich time

Session 3: Bénéfices socio économiques de l'Open Data

Importance et bénéfices

3

Amélioration de l'efficacité du Secteur Public

Prévisions et planification



Une meilleure prévision des tendances (démographiques, économiques, environnementales) afin d'adapter les politiques publiques ou les stratégies d'entreprises

Réduction des coûts de collecte et de gestion des données



Éviter l'effort dupliqué pour la collecte et l'analyse

Amélioration des services publics



Développement des services publics qui répondent mieux à leurs attentes



Transparence
gouvernementale



Recherche &
Innovation



Optimisation du
service public

Session 3: Bénéfices socio économiques de l'Open Data

Exemples

Le cas HandiMap



Avec Handimap, l'efficacité dans la mobilité des personnes en situation de handicap a été renforcée grâce à une plateforme proposant des itinéraires et des cartographies adaptés, facilitant des déplacements inclusifs et optimisés.

Handimap
LA VILLE ACCESSIBLE À TOUS
PAR SOMEWARE

Offre Réalisations

Des cartographies et itinéraires adaptés à chacun

Handimap est la seule solution capable d'intégrer l'ensemble de vos données pour proposer des cartographies de l'accessibilité de votre territoire et des itinéraires adaptés à chacun.

Développée par **Someware**, l'offre Handimap s'adresse aux acteurs de la ville, de la mobilité et du tourisme :

- Pour cartographier l'accessibilité de la voirie, des ERP et des transports, avec des outils adaptés à vos services techniques
- Pour communiquer sur l'accessibilité du territoire auprès des citoyens et visiteurs (loisirs / affaires), avec notamment le calculateur d'itinéraires accessibles Handimap, seule solution prenant en compte l'accessibilité des trottoirs, traversées et chemins

[Demander une démo →](#)

Handi Map (France - Région Lorient)

Description du projet :

Handimap est une application qui aide les personnes à mobilité réduite à planifier des itinéraires adaptés en ville. Elle prend en compte les obstacles urbains tels que les escaliers, les trottoirs non accessibles, ou l'absence d'ascenseurs dans les transports publics.

Données utilisées :

- Données géographiques sur les infrastructures urbaines (largeur des trottoirs, pentes, présence de rampes, etc.).
- Localisation des équipements publics accessibles (toilettes, parkings réservés, ascenseurs).
- Données des transports publics (horaires, accessibilité des stations).
- Signalements citoyens d'obstacles ou d'infrastructures inaccessibles.

Impact :

- **Social** : Amélioration de l'inclusion urbaine et de la mobilité des personnes en situation de handicap.
- **Économique** : Meilleure utilisation des infrastructures accessibles, avec une satisfaction accrue des utilisateurs.
- **Opérationnel** : Aide aux municipalités pour identifier et prioriser les améliorations nécessaires dans les infrastructures.

[Source \(HandiMap\)](#)

Session 3: Bénéfices socio économiques de l'Open Data

Exemples

Le cas FixMyStreet



Avec FixMyStreet, l'efficacité opérationnelle dans la gestion des infrastructures locales a été améliorée grâce à une plateforme permettant aux citoyens de signaler directement les problèmes aux autorités, facilitant des interventions plus rapides et ciblées.

[Source \(FixMyStreet\)](#)

Fix My Street (UK)

Description du projet :

"Fix My Street" est une plateforme permettant aux citoyens de signaler facilement les problèmes d'infrastructure publique (nids-de-poule, lampadaires défectueux, déchets, etc.) directement aux autorités locales via une application ou un site web.

Données utilisées :

- Localisation géographique des signalements (coordonnées GPS).
- Données des infrastructures municipales (routes, lampadaires, réseaux d'eau).
- Statuts des interventions (dates de réparation, coût estimé).

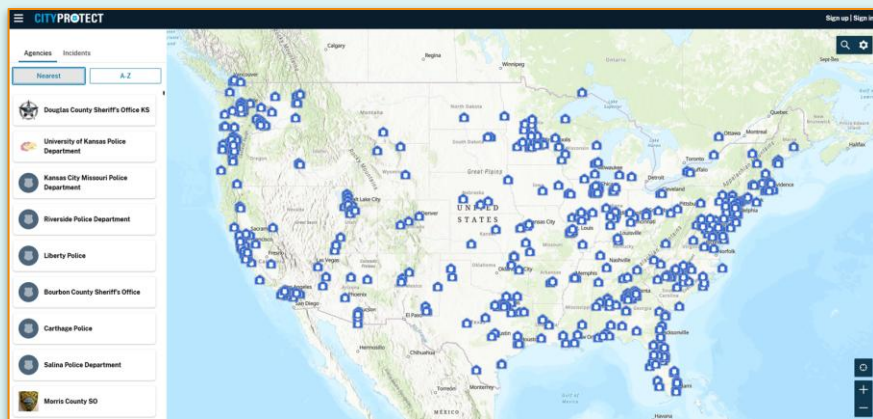
Impact :

- **Social** : Renforcement de l'engagement citoyen et amélioration de la communication avec les autorités locales.
- **Économique** : Optimisation des ressources municipales en priorisant les interventions.
- **Opérationnel** : Réduction des délais de traitement des problèmes grâce à un suivi transparent.

Session 3: Bénéfices socio économiques de l'Open Data

Exemples

3 Amélioration de l'efficacité du Secteur Public (exemple)



[Source \(City Protect\)](#)

City Protect est une application qui permet aux citoyens de signaler des problèmes de sécurité publique (zones mal éclairées, comportements suspects, incidents) directement aux autorités locales.

- **Signalements citoyens** (localisation GPS, descriptions, photos/vidéos).
- **Données des infrastructures urbaines** (éclairage public, caméras de surveillance).
- **Statistiques de criminalité** et incidents signalés.
- **Données en temps réel des autorités** (statuts des interventions, rapports de police).

- **Social** : Amélioration de la sécurité des citoyens grâce à des interventions rapides et ciblées.
- **Opérationnel** : Optimisation des ressources des forces de sécurité et des municipalités.
- **Prévention** : Identification des zones sensibles pour des actions préventives (ex. installation de caméras, renforcement de l'éclairage).

Session 3: Bénéfices socio économiques de l'Open Data

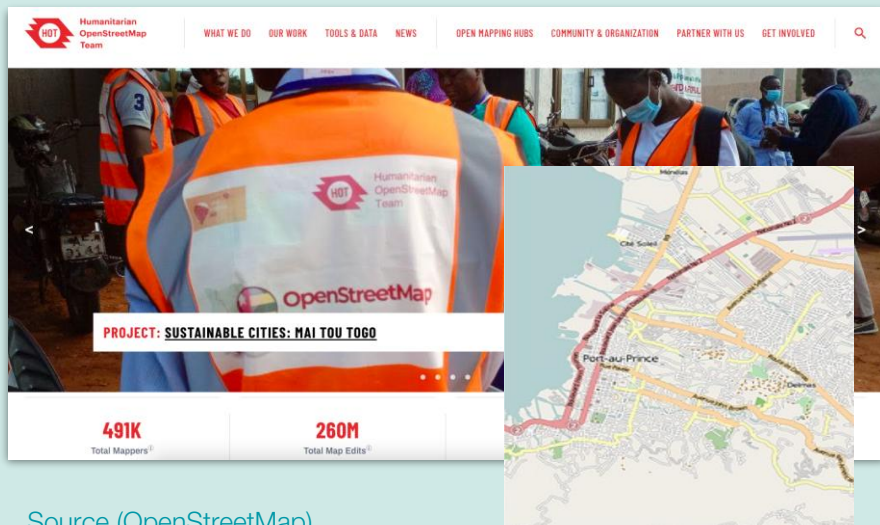
Exemples

Le cas OpenStreetMap



Avec OpenStreetMap, l'efficacité opérationnelle des secours en cas de catastrophe naturelle a été renforcée grâce à une cartographie collaborative et actualisée, permettant de mieux coordonner les interventions et d'optimiser les ressources.

exemples: Tremblement de terre en Haïti (2010) (800 points en 2 jours)



Source ([OpenStreetMap](https://www.hotosm.org/))

Des organisations telles que *Humanitarian OpenStreetMap Team* utilisent des données ouvertes pour aider lors des catastrophes naturelles

- Données géospatiales pour cartographier les zones sinistrées en temps réel.
- exemple:
En deux jours, plus de 800 modifications ont été effectuées, transformant le plan de Port-au-Prince en une carte détaillée.

- Amélioration de la réactivité des équipes de secours
- Réduction du temps de réponse lors des crises humanitaires.

Session de formation sur l'Open Data

Séance 2: Plateformes Open Data

10h10-11h00

Session 2: Plateformes Open Data

Les producteurs

Why Invest in Morocco

A strong and stable economy

A continuous and continuing growth
3.1% 10-year average

A growing domestic demand
Induced by the growth of population and Gross Domestic Income per Capita

A controlled inflation
Morocco is **#1 among 141 countries** in inflation stability (less than 2%)



Economy

Why Invest in Morocco

Qualified & competitive H.R.

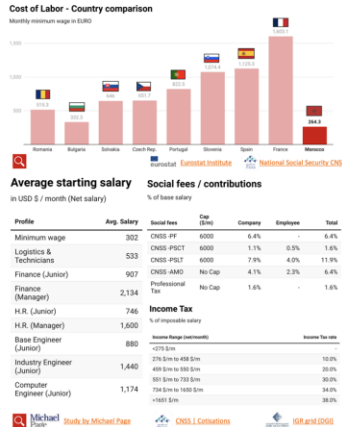
Morocco benefits of a young population opened to the world, educated and trained through the vast network of universities and professional training institutions.

33% Of population is between 20 and 45 years

150 k University graduates per year

11 k Graduating engineers per year

423 Secondary Education or Training Institutions



Why Invest in Morocco

Sustainable

37% of energy is from renewable source in 2020 - 52% target in 2030

Morocco's per capita emission of CO2 is less than half of world average

#1 most attractive country in African, #16 worldwide

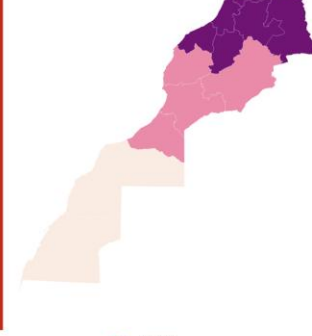
By Renewable Energy Country Attractiveness Index 2021

recai
Renewable Energy Country Attractiveness Index

Energy Cost per region

€/MWh

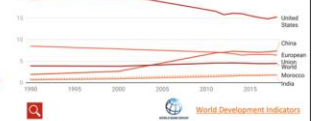
0.02 0.04



MOROCCO NOW Sustainable | Morocco Now

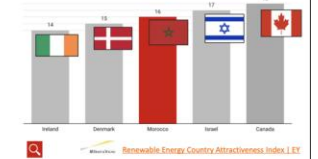
CO2 emission

metric ton per capita



Renewable Energy Attractiveness Index

(BECA) Country ranking

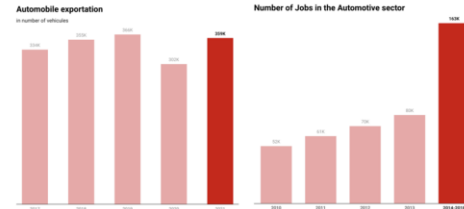


Why Invest in Morocco

Case of the Automotive industry in Morocco

A growing industry

The automotive sector is a growing sector in the Moroccan industry, maintaining its export vocation and creating more and more jobs



Evolution of automotive production from 2012 to 2018

Country comparison - in % growth rate



Evolution of Automobile production capacity

from 1960 to 2023 in number of vehicles



Session 2: Plateformes Open Data

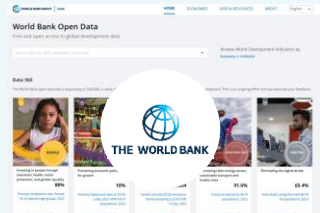
Les producteurs

Qui produit l'Open Data?




Les ONG

Publication de données collectées par des ONG (Banque Mondiale, ONU, FMI, ...)




Banque Mondiale

Indicateurs économiques: PIB, Education, ... [Lien](#)




Nations Unies

Statistiques multi-thèmes (population, commerce, énergie, agriculture)




Fonds Monétaire Int'l

PIB réel, inflation, balance des paiements, réserves, taux de change




Organisation Mondiale de la Santé

Indicateurs économiques: PIB, Education, ...



UNICEF

Données sur survie de l'enfant, nutrition, éducation, protection



UNESCO

Données comparatives sur éducation, science, culture, communication

Session 2: Plateformes Open Data

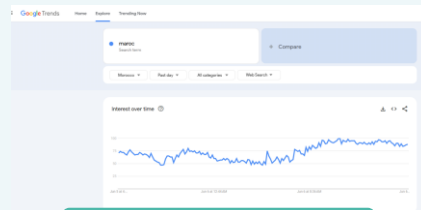
Les producteurs

Qui produit l'Open Data?



Les Entreprises

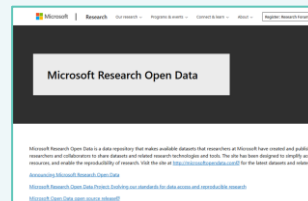
Certes rares, mais peuvent également publier certaines données



Google

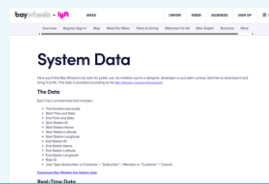
Statistiques de recherche
Données lors du Covid (mobilité)

[Lien](#)



Microsoft

Statistiques multi-thèmes (population, commerce, énergie, agriculture)



Lyft

Temps de trajet et vitesses agrégés
(origine / destination)



Orange

Relevés anonymisés d'appels & SMS
(via accords de recherche)

Session 2: Plateformes Open Data

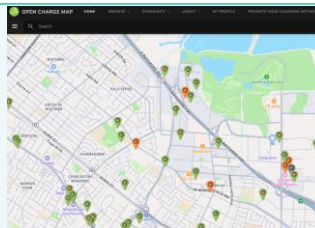
Les producteurs

Qui produit l'Open Data?



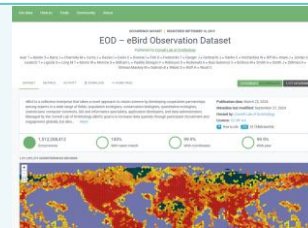
Les particuliers/Société civile

Publication participative de données dans différents domaines (recherche, nature, transport, ...)



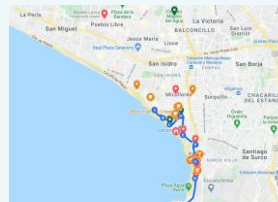
Cartographie

Cartographie de services essentiels, zones sinistrées, ...
Ex. Open streetmap



Recherche scientifique

Statistiques multi-thèmes (population, commerce, énergie, agriculture)



Wayz / Google Maps

Statistiques multi-thèmes (population, commerce, énergie, agriculture)

Session 2: Plateformes Open Data

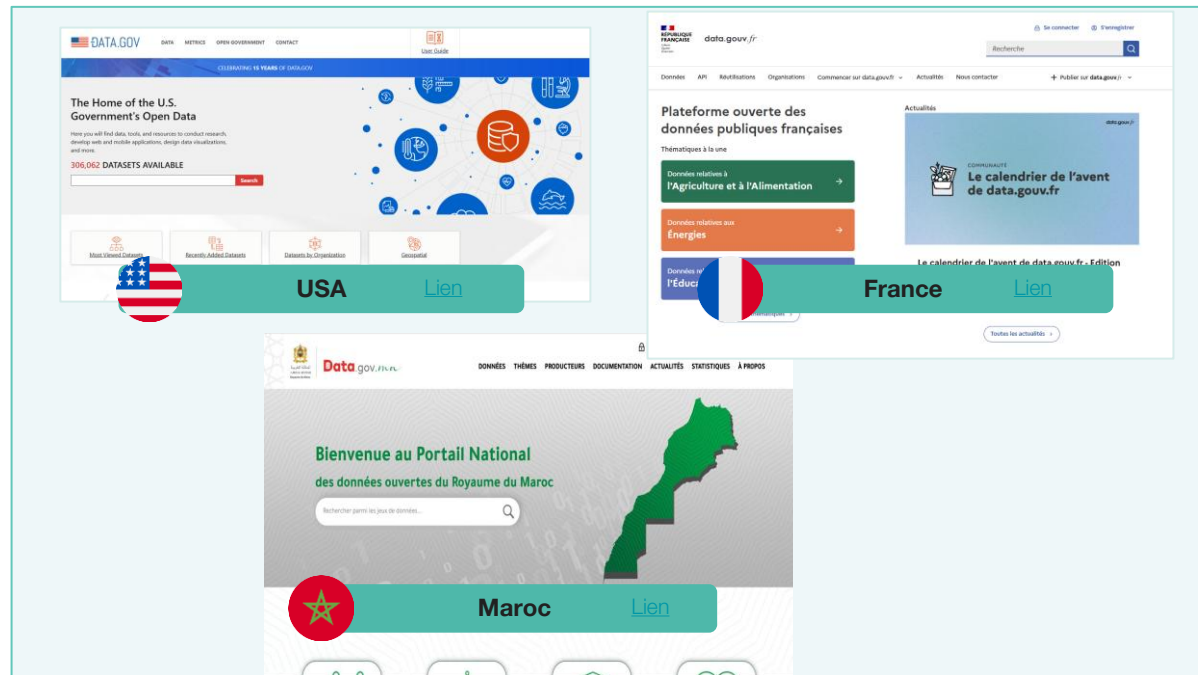
Les producteurs

Qui produit l'Open Data?



Le gouvernement

Publication de données gouvernementales
(Finance, Culture, Education, Santé, ..)

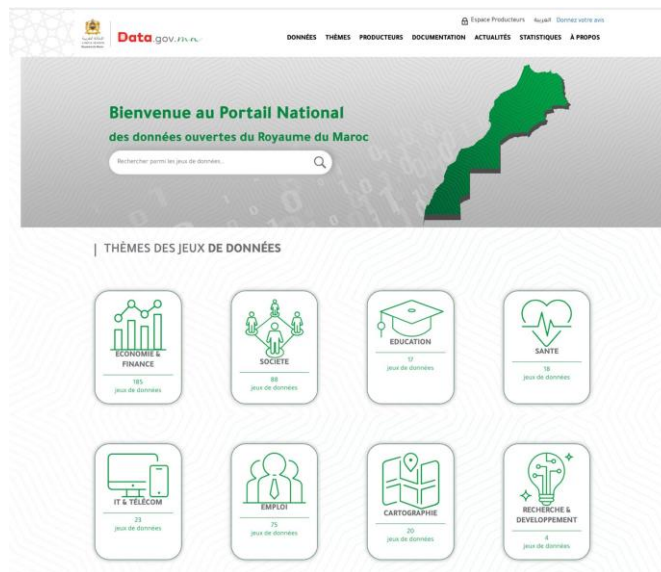


Session 2: Plateformes Open Data

Le portail national

Une démarche amorcée

Le Portail National compte déjà 600+ jeux de données, publiés par 40+ organismes publics



Source (Portail National Open Data)

682

Jeux de données

1 535

Ressources

47

Producteurs de données

166 096

Fichiers téléchargés

2 524 474

Pages vues

Session 2: Plateformes Open Data

Le portail national

Producteurs de données actuels

Le Portail National compte déjà 600+ jeux de données, publiés par 42 organismes publics

ROYAUME DU MAROC



MINISTÈRE DE LA JUSTICE

المملكة المغربية
ROYAUME DU MAROC



وزارة التربية الوطنية
والتعليم الأولي والابتدائي
Ministère de l'Éducation
Nationale, de l'Enseignement
Primaire et de l'Enseignement
Secondaire



وزارة إعداد التراب الوطني والتعمير
وسياطة المدينة
Ministère de l'Aménagement du
Territoire National,
de l'Urbanisme, de l'Habitat et de la
Politique de la Ville

المملكة المغربية
ROYAUME DU MAROC



وزارة التضامن والتنمية
الاجتماعية والتضامن
والتضامن
Ministère de la SOLIDARITÉ
DE L'INSERTION SOCIALE ET DE LA FAMILLE

المملكة المغربية
مجلس النواب
الملك محمد السادس
ROYAUME DU MAROC
Assemblée Nationale



الضمان الاجتماعي
CNS
Le devoir de vous protéger

IMANOR
المعهد المغربي للتقييس
Institut Marocain de Normalisation



NARSA
الوكالة الوطنية للسلامة الطرقية
National Road Safety Agency



وزارة الانتقال الرقمي
وإصلاح الإدارة



وزارة إعداد التراب الوطني والتعمير
وسياطة المدينة
Ministère de l'Aménagement du
Territoire National,
de l'Urbanisme, de l'Habitat et de la
Politique de la Ville



وزارة الفلاحة والصيد البحري
والتشجير
Ministère de l'Agriculture, de la Pêche
Maritime et des Eaux et Forêts



MTAESS
Données ouvertes du Ministère du Tourisme,
de l'Artisanat, et de l'Économie Sociale et
Solidaire.

أرشيف المغرب
Archives du Maroc



Fondation
المؤسسة
Nationale
الوطنية
des Musées
للمتاحف
des Musées



MAECAMRE
Données ouvertes des Affaires Étrangères,
de la Coopération Africaine et des
Marocains Résident à l'Étranger.



المملكة المغربية
ROYAUME DU MAROC
وزارة التعليم العالي والبحث العلمي والتكنولوجيا
Ministère de l'Enseignement Supérieur,
de la Recherche Scientifique et de la Technologie

المملكة المغربية
وزارة الإجماع الاقتصادي
والمقاولات الصغرى، والتشغيل والكفاءات

#ADD
Agence de Développement du Digital

ANM
البنك المغربي لمخزون الرصيد
ANM
ANM

anrt
agence nationale de réglementation
des télécommunications
الوكالة الوطنية لتنظيم المواصلات
التواصلية

المملكة المغربية
وزارة الشباب والثقافة والتواصل
Ministère de la Jeunesse, de la Culture et de la Communication



Royaume du Maroc
Ministère de l'Industrie et du Commerce



المملكة المغربية
وزارة الصناعة والتجارة



الصندوق المغربي للتقاعد
Caisse Marocaine des Retraites

بنك المغرب
BANK AL-MAGHRIB

Inppic
المعهد الوطني للتخطيط
National Institute of Planning



وزارة الصحة والحمية الاجتماعية
Ministère de la Santé et de la Protection Sociale



Royaume du Maroc
Ministère de l'Économie et des Finances

Royaume du Maroc
CNOPS
La solidarité c'est bon pour la santé

المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN



acaps
الجمعية المغربية لعلوم الاقتصاد والتجارة
Association Marocaine de Sciences Économiques et Commerciales

Session 2: Plateformes Open Data

Le portail national

Parmi les producteurs actuels des institutions régionales

Plusieurs régions publient déjà sur le Portail National



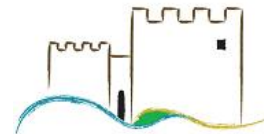
جهة بني ملال خنيفرة
ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ
Région Béni Mellal Khénifra



جهة طنجة - تطوان - الحسيمة
ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ
RÉGION TANGER-TÉTOUAN-AL HOCEIMA



وكالة الحوض المائي لسوس ماسة
ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ
Agence du Bassin Hydraulique de Souss Massa



الوكالة الحضرية لتسفي - اليوسفيّة
ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ
AGENCE URBAINE SAFI - YOUSOUFIA



غرفة التجارة والصناعة والخدمات لجهة مراكش أسفي
ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ
CHAMBRE DE COMMERCE, D'INDUSTRIE ET DE SERVICES DE LA RÉGION MARRAKECH-SAFI

المركز الجهوي للاستثمار سوس ماسة
ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ
Centre Régional d'Investissement Souss Massa



المركز الجهوي للاستثمار • درعة تافيلالت
ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ ⵜⴰⵎⴻⵔⴰ
Centre Régional d'Investissement • Drâa-Tafilalet

Session 2: Plateformes Open Data

Le portail national

Une multitude de catégories & secteurs

Les jeux de données disponibles dans le portail permettent de concevoir des solutions dans tous les domaines



Agriculture

Types de data:
Production végétale
Cheptels bovins



Assurance

Types de data:
Statistiques Retraites,
Mutualité, AMO



Cartographie

Types de data:
Liste et positions
d'infrastructures,
administrations, codes
postaux, ..



Culture

Types de data:
Liste des musées, Centres
sportifs, maisons de
jeunesse, campings, ..



Economie

Types de data:
Statistiques & Indicateurs
(Cout de la vie,
importations, ...)



Education

Types de data:
Liste des établissements
par spécialité, filières



Emploi

Types de data:
Indicateurs Actifs, postes
de responsabilité, ..



Equipement

Types de data:
Liste des ports, barrages,
infrastructures, ..



Habitat

Types de data:
Prix de matière première,
Montant des loyers,
honoraires des agents, ..



IT & Telecom

Types de data:
Parc internet/mobile,
qualité de service, traffic, ..



Jeunesse

Types de data:
Colonies de vacances,
centre sociaux, ..



Justice

Types de data:
Statistiques & indicateurs
sur les activités de
tribunaux

[Source \(Portail National Open Data\)](#)

Session 2: Plateformes Open Data

Le portail national

Comité de Pilotage national de l'Open Data

Un Comité de Pilotage dédié à l'Open Data a été mis en place en décembre 2019 pour superviser et veiller à la mise en œuvre du chantier Open Data au niveau national.

Principales missions du Comité de Pilotage

- Le pilotage et la supervision de la mise en œuvre du chantier Open Data à l'échelle nationale ;
- L'accompagnement de toutes les initiatives des organismes et institutions publics concernés relatives à l'ouverture des données publiques ;
- La mobilisation de l'ensemble des parties prenantes ;
- Le suivi et l'évaluation des performances de la mise en œuvre du plan d'actions Open Data.

Comité actuellement composé de:

- Ministère de l'Economie et des Finances;
- Ministère de la Jeunesse, de la Culture, et de la Communication (Département de la Communication) ;
- Ministère de la Transition Numérique et de la Réforme de l'Administration;
- Haut-Commissariat au Plan;
- Instance Nationale de la Probité, de la Prévention et de la Lutte contre la Corruption ;
- Agence Nationale de Réglementation des Télécommunications ;
- Agence de Développement du Digital.

[Source \(Portail National Open Data\)](#)

Session 2: Plateformes Open Data

Acteurs

Pilotage et Coordination

Le Comité de Pilotage Open Data est coordonné par l'Agence de Développement du Digital, en tant qu'entité chargée de mettre en œuvre la stratégie de l'Etat en matière de développement du digital et de promouvoir des outils numériques et le développement de leur usage auprès des usagers.



Acteur central de l'Open Data

- Mise en œuvre du plan d'action national Open Data
- Gestion du portail national
- Guides, Méthodologies, Formations
- Assistance technique

Acteurs clés qui insufflent la vision, définit la stratégie pour la mise en œuvre d'une démarche Open Data

[Source \(Portail National Open Data\)](#)

Session 2: Plateformes Open Data

Acteurs

Producteurs de données

Les données ouvertes sont principalement produites par les institutions publiques. La Feuille de Route nationale organise les rôles au sein de l'administration en matière de production et publication de données ouvertes:



Structures publiques

- Administrations
- Entreprises publics
- Collectivité territoriales

Avec la désignation de :



Responsable Open Data (ROD)

l'acteur qui transforme la vision du responsable en plan d'actions, coordonne les activités, met en œuvre les différents outils et assure le suivi et évaluation



Gestionnaire de Données (GdD)

acteurs techniques clés de la publication de données. Ils sont en charge de la préparation et de la publication des données qu'ils gèrent

Session 2: Plateformes Open Data

Acteurs

Réutilisateurs de données

Les données ouvertes produites peuvent être exploitées par plusieurs acteurs, pour différents objectifs:



Acteurs de l'innovation

Concevoir des solutions et services innovants



Structures publiques

Améliorer l'efficacité du service public, analyses, prédictions, benchmarks



Secteur privé

Réaliser des études de marché, développer de nouveaux services



Médias

Améliorer la qualité de l'information délivrée



Société civile

Mener des enquêtes et analyses



Secteur académique

Mener des recherches et études universitaires

Session 2: Plateformes Open Data

Objectifs nationaux

Par le biais de ces dispositifs, le Maroc ambitionne de:



Promouvoir la transparence à l'échelle nationale

Par la publication de diverses données relatives au fonctionnement du service public et autres secteurs



Stimuler l'innovation et la création de richesse

Par l'encouragement à l'investissement, le développement de nouvelles solutions, produits & services à valeur ajoutée



Soutenir le développement économique et social

Par des solutions qui favorisent l'inclusion sociale, la réduction des inégalités, le renforcement des opportunités économiques dans des régions défavorisées, ..

Comment atteindre ces objectifs ?

Exposer des données produites par les administrations, les établissements publics et les collectivités territoriales au niveau du **portail national Open data** et/ou les portails des organismes concernés afin qu'elles puissent être **exploitées** par des acteurs de **l'innovation** pour apporter de la valeur

Session 2: Plateformes Open Data

Réalisations

Bref résumé de ce qui a été fait jusqu'à présent



- **Désignation des RODs et GdDs** (plus de 180 ROD/GdD chez plus de 100 organismes publics)
- Mise en place de groupes de travail



- **Upgrade du Portail national** de l'Open Data (www.data.gov.ma)
- **Mise à disposition de + de 600 jeux de données** sur le portail
- **Activation de l'API** de la mise à jour automatique des données du HCP sur le portail Open Data, et le travail est en cours pour les autres producteurs ;



- **Intégration de la catégorie « Open Data »** dans le cadre du prix Arriyada organisé en marge du GITEX Morocco 2024 ;
(5 administrations ont été récompensées suite à leur partage actif sur le portail)
 - Le Ministère de l'Economie et des Finances ;
 - Le Ministère de l'Éducation nationale, du Préscolaire et des Sports ;
 - Le Ministère de la Justice ;
 - Le Ministère de l'Équipement et de l'Eau ;
 - Le Ministère de l'Aménagement du Territoire national, de l'Urbanisme, de l'Habitat et de la Politique de la Ville.

- Ateliers de concertation avec les acteurs privés, le monde académique et de la société civile pour **encourager la réutilisation des données ouvertes**
- **Ateliers de sensibilisation et de formation** au profit des responsables des organismes publics
- Mise à disposition de **manuels de formations** et de procédures sur le portail national
- **Hackathon Open Data** afin d'exploiter la valeur que peut apporter la data



Plus de **1200 bénéficiaires** des ateliers de sensibilisation et de formation au niveau central et régional

Session 2: Plateformes Open Data

Réalisations

Hackathon Open Data du 18-19 février '25

Equipes et thèmes

13 équipes de 6, composées à partir de projets autour de l'OPEN DATA. Elles ont été accompagnées et encadrées pendant 2 jours pour la conception des solutions selon la méthode Design Thinking/Pitch



**Thème
Tourisme**



**Thème
Agriculture**



**Thème
Mobilité**



**Thème
Economie**



Thème Santé



**Thème
Environnement**



**Thème
Education**



**Thème
Artisanat /
Culture**



Session 2: Plateformes Open Data

Réalisations



Session 2: Plateformes Open Data

Réalisations

Pendant le Hackathon

Projets et équipes gagnantes

🏆 **1er Prix – Hospitranfert** : Plateforme d'interconnexion hospitalière pour les médecins.

🏆 **2e Prix – Mizaniyatouna** (Open Budget Lab) : Solution facilitant l'accès aux données sur les dépenses et recettes publiques.

🏆 **3e Prix – Aquacoeur** : Plateforme optimisant la consommation d'eau en agriculture pour soutenir les ressources hydriques.

🏆 **Prix Coup de Cœur du Jury : "Rabeh"**, outil d'aide aux investisseurs et porteurs de projets, offrant un accès à des données stratégiques sur les besoins et opportunités d'investissement avant, pendant et après la Coupe du Monde.

🏆 **Prix spécial Jeune Talent** : Adam Benchlih (15 ans). Le jury a tenu à récompenser M. Adam Benchlih, venu d'Agadir pour participer au hackathon



Pause café

11h00-11h15

Session de formation sur l'Open Data

Séance 3: Standards techniques de l'Open Data

11h15-12h30

Session 3: Standards et Plateformes techniques

Standards techniques: un pilier essentiel

Les standards techniques jouent un rôle clé pour garantir que les données ouvertes soient exploitables, fiables et réutilisables par tous les utilisateurs. Voici les raisons majeures qui justifient leur importance :

Interopérabilité



Permet à différents systèmes, plateformes et applications d'échanger et de traiter des données sans obstacles techniques.

Réutilisabilité



Les données conformes aux standards peuvent être manipulées et analysées à l'aide d'outils standardisés.

Qualité et Transparence



Les standards garantissent une présentation claire et une structuration fiable des données, augmentant leur crédibilité.

Efficacité



Évite de réinventer des processus pour chaque publication, ce qui diminue les coûts et efforts techniques.

Innovation et Développement de Nouveaux Services



Les standards techniques stimulent la créativité en facilitant la conception de nouvelles applications et services innovants.

Prévisibilité et Planification



Les données conformes aux standards permettent une analyse fiable des tendances pour la prise de décision stratégique.

Un guide des standards Open Data élaboré par l'ADD est fourni ici: [Lien vers le guide](#)

Session 3: Standards et Plateformes techniques

Critère 1: Formats de fichiers

Pour permettre au plus grand nombre d'utiliser un fichier de données, il est nécessaire que ce fichier soit dans un format documenté et public, aussi appelé « format ouvert », manipulable par des outils standard et ne nécessitant pas l'acquisition de logiciels spécifiques.

Pour les données tabulaires

CSV ✓

XLS ✓ ⓘ

Pour les images

PNG ✓

PSD ✗

JPEG ✓

TIFF ✗

Pour les images vectorielles

SVG ✓

AI ✗

Pour les API

OAS ✓

REST ✓

Pour les données géospatiales

Geojson ✓

DXF ✗

Geotiff ✓

GML ✓

KML ✓

WFS ✓

Voir également les normes
NM ISO 1913126 et ISO
19119

Pour les données structurées

XML ✓

AZW ✗

JSON ✓

RDF ✓

ODF ✓

EPUB ✓

Session 3: Standards et Plateformes techniques

Critère 2: Données tabulaires

Il est nécessaire de garder en tête que ces données sont destinées à être lues par des machines, et non pas à être ergonomiques pour un humain

Table 4. Gender Development Index										

Session 3: Standards et Plateformes techniques

Critère 3: Encodage des données

Parmi une multitude d'encodages possibles (ASCII, UTF), l'encodage défini au niveau international est UTF-8

Encodage ASCII / Latin-1 (ISO8895)

- 128 caractères dédiés à l'anglais ou spécifique aux langues latines
- Source de problèmes d'incompatibilité des langues

Où sont les caractères accentués ?

Encodage UTF-8

- 2 millions d'entrées

Où sont les caractères accentués ?

Pour éviter ces problèmes, UTF-8 est l'encodage défini au niveau international qui permet de couvrir l'ensemble des caractères de tous les langages.

C'est également l'encodage recommandé dans le CGI.

Session 3: Standards et Plateformes techniques

Critère 4 : Types de données (valeurs numériques)

Certains types de données présentent un défi particulier car peuvent être présentées de façons différentes. Il est donc essentiel de standardiser les structures dont sont présentées certaines informations

Cas des valeurs numériques,

par exemple le nombre 1.5

- Dans les pays anglosaxons: **“1.5”**
- Dans les pays francophones: **“1,5”**

par exemple le nombre 2 500 000

- Dans les pays anglosaxons: **“2,500,00”**
- Dans les pays francophones: **“2 500 000”**

l'organisme de standardisation IEEE a défini une norme internationale (IEEE 754-201938) qui:

- impose le « . » comme séparateur de décimal
- interdit l'utilisation de séparateur de milliers

(par exemple « 1001.5 » respecte cette norme).

Session 3: Standards et Plateformes techniques

Critère 4 : Types de données (Dates)

Certains types de données présentent un défi particulier car peuvent être présentées de façons différentes. Il est donc essentiel de standardiser les structures dont sont présentées certaines informations

Cas des valeurs dates,

par exemple le 12 janvier 2005

- Dans les pays anglosaxons: **“01/12/2005”** ou **“01-12-2005”**
- Dans les pays francophones: **“12/01/2005”** ou **“12-01-2005”**

Plusieurs formats de dates longues existent

- Par exemple 12/05/2022 2:57pm , 2002-05-12 12:27:58 ou 2022-05-12T14:57:00Z
- Nécessité de tenir compte des fuseaux horaires

L'organisme de standardisation ISO a défini une norme (ISO 860139), norme également adoptée par l'IMANOR (NM ISO 860140) qui fixe un format de date universel

AAAA-MM-JJTHH:MM:SS(.sss)Z

exemple : 2022-05-12T14:57:00Z pour le 12 mai 2022 à 14:57:00 UTC).

Ce format est référencé dans le CGI.

Session 3: Standards et Plateformes techniques

Critère 4 : Types de données (n° téléphone)

Certains types de données présentent un défi particulier car peuvent être présentées de façons différentes. Il est donc essentiel de standardiser les structures dont sont présentées certaines informations

Cas des valeurs numéros de téléphone,

Groupages différents selon les pays

- Au Maroc: 05 37 11 11 11
- Aux Etats Unis: 555-333-1234

Indicatifs internationaux

- +212 661 111 123 ou +212 (6) 61 111 123 ou 00212661111123, etc...

L'Union Internationale des Télécommunications (UIT) a adopté un standard (ITU E.16441) qui

impose une représentation unique des numéros de téléphone avec les caractéristiques suivantes

- Le numéro commence par le code international du pays sans le signe +
- Puis le numéro complet sans groupage et sans séparateur directement après le code pays et sans les chiffres optionnels locaux

Les exemples précédents s'écriraient donc :
212661111123

Session 3: Standards et Plateformes techniques

Critère 4 : Types de données (Coordonnées géographiques)

Certains types de données présentent un défi particulier car peuvent être présentées de façons différentes. Il est donc essentiel de standardiser les structures dont sont présentées certaines informations

Cas des valeurs coordonnées géographiques,

Il existe 2 façons de transcrire des coordonnées

- Format décimal : 34.0209° N, -6.8416° W
- Format degrés, minutes, secondes (DMS) : 34° 1' 15" N, 6° 50' 30" W

Il existe plusieurs systèmes géodésiques

- WGS 84, NAD 83, PZ-90, GCI-02, BD-09

Données atomiques: il est recommandé de fournir ce type de données séparées sous plusieurs colonnes

Le plus utilisé est WGS 84:

Standard global : Utilisé pour les systèmes de navigation par satellite, notamment **GPS**.

Universalité : Compatible avec la majorité des cartes et systèmes de localisation modernes.

Précision : Offre une représentation précise de la forme de la Terre (ellipsoïde) et des coordonnées géographiques.

Adoption internationale : Reconnu comme référence par de nombreuses organisations, y compris l'ONU, l'aviation civile (OACI), et les services maritimes.

Session 3: Standards et Plateformes techniques

Récapitulatif des critères techniques

Dimension	Critère	Standard recommandé
Format de fichiers	Données tabulaires	CSV - RFC 4180
	Image bitmap	PNG(*), JPEG (*), TIFF (*)
	Image vectorielle	SVG (*)
	Texte	TXT(*), RTF (*), HTML(*)
	Données géospatiales	Geojson, geotiff ou shapefile
	Données structurées/hierarchiques	XML(*), JSON (*), EPUB
Encodage	Encodage des fichiers	UTF-8 (*)
Type de données	Valeurs numériques	IEEE 754-2019
	Date	NM ISO 8601 (*)
	Numéro de téléphone	ITU E.164
	Coordonnées géographiques	Système géodésiques WGS 84 (GPS)
	Données textuelles	Utilisation de référentiel
	Données atomiques	Les données ayant plusieurs composantes comme les données géographiques sont séparées (une colonne par composante)
	Valeur manquante	Cellule/champs vide
Unités	Unité de mesure	Système international de données - ISO 80000-1

Session 3: Standards et Plateformes techniques

Transformer la donnée pour la rendre homogène

Il est nécessaire de garder en tête que ces données sont destinées à être lues par des machines, et non pas à être ergonomiques pour un humain. Il faut transformer/formater les données selon des standards définis

Normaliser

Les espaces, les majuscules, les accents, la ponctuation

Valeurs numériques

Appliquer la norme IEE 754-2019
Décimales avec le '.', pas de séparateurs de milliers

Dates

Harmoniser les structures de date
Utiliser la norme 8601
Préciser le fuseau horaire
Préférer le fuseau UTC

Numéros de téléphone

Appliquer la norme ITU E.164
Ex. 212661123123

Coordonnées géographiques

Système géodésiques WGS 84 (GPS)
Séparer les valeurs en colonnes

Données textuelles

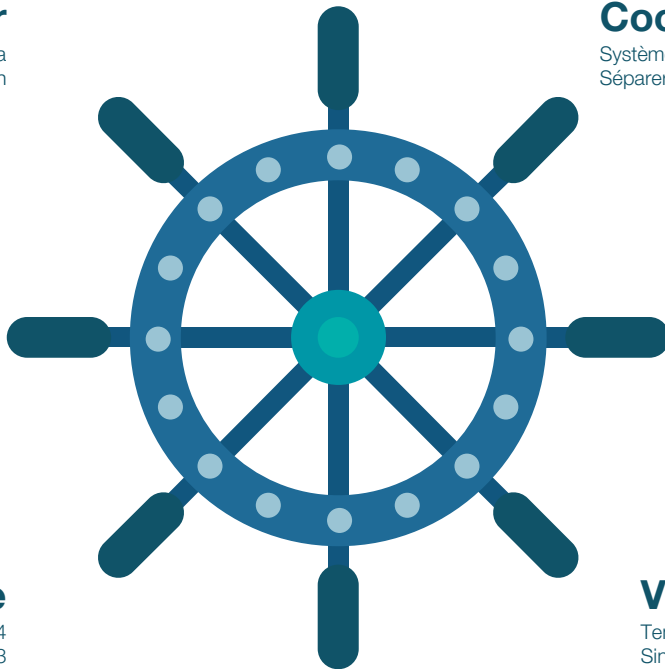
Privilégier l'utilisation de référentiels (ex. registre de commerce, codes postaux, ..)

Données atomiques

Séparer les données composées de sous données
(ex. GPS) en plusieurs colonnes

Valeurs manquantes

Tenter de récupérer la donnée
Sinon laisser vide



Session 3: Standards et Plateformes techniques

Anonymiser les données

Les données permettant d'identifier une personne, directement ou indirectement doivent être anonymisées, en vertu de la loi 09-08

01

Pseudonymisation

Transformation des données

Champs : Noms, identifiants uniques (ID), adresses e-mail

Exemple (avant -> après) :

John Doe
-> User123



02

Masquage

Suppression partielle

Champs : Numéros de téléphone, cartes de crédit

Exemple (avant -> après) :

1234-5678-9012-3456 ->
XXXX-XXXX-XXXX-3456



03

Généralisation

Réduction de précision

Champs : Adresses, dates de naissance

Exemple (avant -> après) :

12 Avenue des FAR ->
Avenue des FAR



04

Perturbation

Ajout de bruit

Champs : Salaires, données financières

Exemple (avant -> après) :

45,000
-> 44,987



04

Suppression

Élimination complète

Champs : Champs sensibles inutiles

Exemple (avant -> après) :

123 Avenue des FAR -
> (vide)



04

K-Anonymisation

Regroupement

Champs : Âge, code postal

Exemple (avant -> après) :

29
-> 20-30



05

Randomisation

Modification aléatoire

Champs : Adresses e-mail, noms

Exemple (avant -> après) :

jane.doe@example.com ->
zane.qwe@example.com



06

Tokenisation

Remplacement par un jeton

Champs : Identifiants uniques, numéros de compte

Exemple (avant -> après) :

AB123456
-> TK982345



Session 3: Standards et Plateformes techniques

Choix de la méthode d'anonymisation

La bonne méthode d'anonymisation selon le contexte

Selon le type de data, le niveau de réidentification, et l'usage cible prévu, il est possible d'appliquer la bonne méthode:



Type de données

Les types de données définissent la sensibilité et le besoin de protection :

Si ce sont par exemple des données permettant directement l'identification (Nom/Prénom, Adresse, Numéro de téléphone, ...)

Techniques associées :

- **Suppression** : Retirer les informations directement identifiables (ex. : noms dans un fichier médical).
- **Pseudonymisation** : Remplacer les identifiants directs par des codes (ex. : un numéro aléatoire au lieu d'un nom).
- **Généralisation** : Réduire la granularité des données (ex. : transformer une date de naissance en année de naissance).



Niveau de risque de réidentification

Le risque de réidentification dépend de la sensibilité des données et des techniques de croisement:

Si par exemple on a un risque de réidentification à partir des éléments croisés (ex. profession + localisation permettent une identification).

Techniques associées :

- **K-anonymité** : Assure qu'au moins K individus partagent les mêmes attributs pour éviter la réidentification (ex. : au moins 10 personnes avec la même combinaison de sexe et âge).
- **Perturbation** : Ajouter du bruit statistique (ex. : altérer légèrement les valeurs numériques pour masquer les données exactes).



Usage prévu des données

L'usage final des données détermine la précision nécessaire et les contraintes éthiques :

Si par exemple l'usage prévu est à des fins statistiques:

Techniques associées :

- **Agrégation** : Transformer des données individuelles en statistiques globales (ex. : taux moyen d'une maladie par région).

Séance 4: Inventaire, classification et référentiels

12h30-13h30

Séance 4:

Inventaire, classification et référentiels

Section 4.1:

Inventaire & classification

Session 4: Inventaire, classification et référentiels

Démarche générale

L'inventaire comme base de travail

Un inventaire de données, dans le cadre de l'Open Data, est un recensement systématique des données disponibles au sein d'une organisation, accompagné d'une description détaillée (source, format, accessibilité, métadonnées, statut légal). Il permet d'identifier les données publiables, d'évaluer leur qualité et de préparer leur ouverture tout en respectant les contraintes juridiques.



Identifier:

- Cartographier les données
- Documenter la data et metadata

Qui?

ROD, Gestionnaires & Producteurs de data

Valider:

- Evaluer la data
- Prioriser & valider l'ouverture

Qui?

ROD

Extraire:

- Développer les outils d'extraction

Qui?

ROD, DSI, GdD, Prestataires SI

Editer:

- Rendre lisible par machine
- Anonymiser

Qui?

ROD, DSI, GdD, Prestataires SI

Standardiser:

- Convertir au format ouvert
- Adopter les standards

Qui?

ROD, Gestionnaires & Producteurs de data, SI

Publier:

- Développer les outils d'extraction

Qui?

ROD, DSI, GdD, Prestataires SI

Maintenir:

- Corriger les erreurs
- Mettre à jour

Qui?

ROD, Gestionnaires & Producteurs de data

Session 4: Inventaire, classification et référentiels

Approche stratégique

La bonne approche, pour le bon contexte

Une administration peut avoir en gestion une quantité importante de data, +/- accessible, +/- à jour, +/- aisément publiable. Il est essentiel d'adopter une approche adaptée à son contexte:

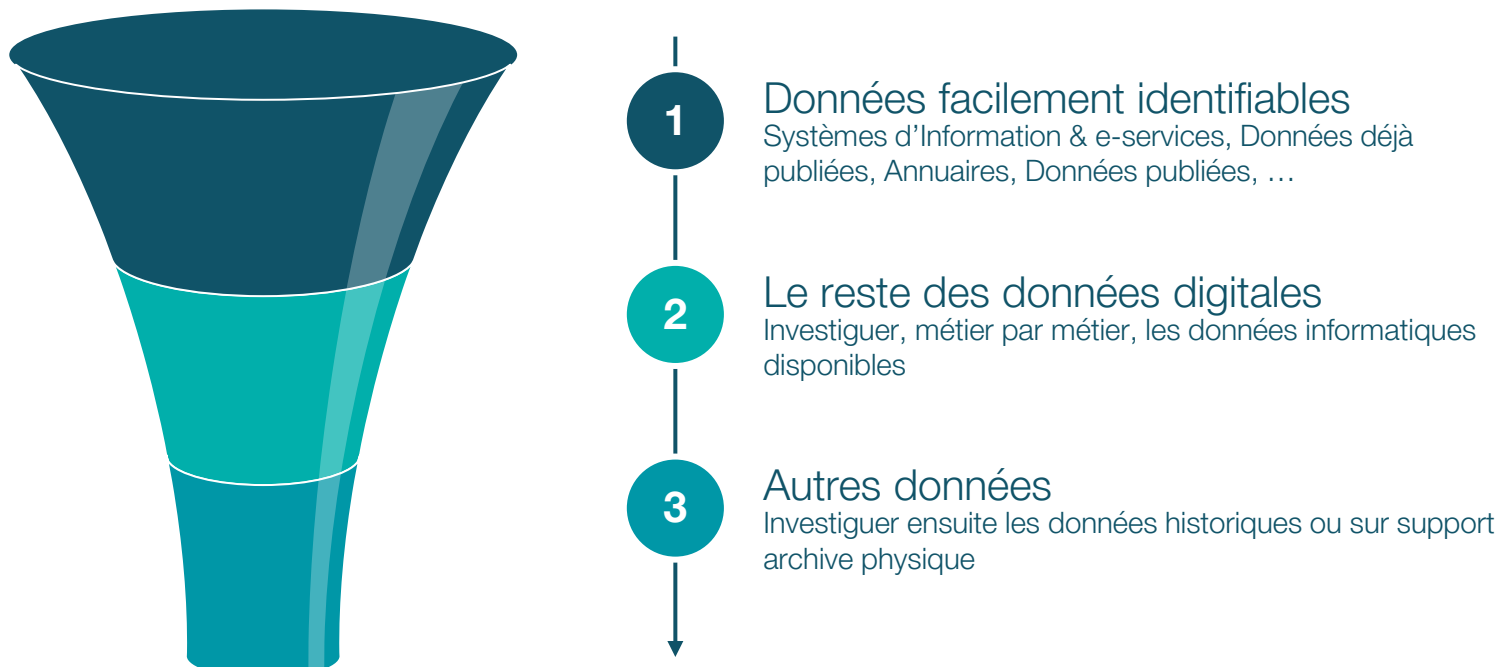


Session 4: Inventaire, classification et référentiels

Priorisation

La priorisation au centre de la démarche

L'approche recommandée, le plus souvent, est l'approche pragmatique. Elle consistera à balayer les sources de données progressivement, selon leur facilité d'accès et de disponibilité



Session 4: Inventaire, classification et référentiels

Classification

Identifier les données pour déterminer les traitements adéquats

Ce processus permet de définir la classification de chaque data en terme de droit d'accès & traitements à effectuer

Identifier les données

A- Publiables directement

B- Publiables après traitement

C- Nécessitent des habilitations spécifiques

Au Maroc, la classification est définie par 3 lois:

Loi 05-20 relative à la cybersécurité

Loi 31-13 relative au droit d'accès à l'information

Loi 09-08 relative à la protection des données à caractère personnel

Identifier les traitements

A appliquer afin que les données de type B soient publiables

Session 4: Inventaire, classification et référentiels

Métadonnées

! = obligatoire ou recommandé

Les métadonnées sont les informations descriptives du jeu de données. Elles sont essentielles pour assurer une organisation cohérente des données à court et long terme

Informations sur le jeu de données

Nom/Titre	!
Nom du jeu de données	
Description	!
Brève description du jeu de données	
Secteur	
Secteur concerné par les données	
Mot clé / Thème	!
Illustration du jeu de données	
Date de création	!
Date de 1ère collecte	
Date de fin de support	
Fin de vie de la collecte de données	
Date de dernière mise à jour	
Dernière maj du jeu de données	
Date de publication	
Date ou le jeu de données a été publié	
URL	
Lien vers le jeu de données	

Informations sur le producteur de données

Organisation en charge du jeu de données	!
Structure publique responsable	
Nom & Coordonnées du responsable métier	!
Responsable administratif du jeu de données	
Nom & Coordonnées du responsable technique	!
Responsable technique du jeu de données	

Informations sur les données

Couverture géographique	!
Portée géographique des données	
Niveau de désagrégation	
Géographique: National, Régional, Provincial, Communal, ..	!
Autre: Tranches d'âge, CSP, ..	
Couverture temporelle	!
Portée temporelle des données	
Couverture historique	
Profondeur	
Langues disponibles	!
Arabe, Amazigh, Anglais, Français, ...	
Format de stockage	!
Format de fichier	
Méthode d'extraction du SI / export	
Emplacement de stockage	
Emplacement physique	
Processus d'extraction	
Automatique ou non	
Descriptif des données	
Description des colonnes, référentiels, ..	
Schéma de validation	
Méthode d'automatisation de la validation	
Lisible par machine	
Oui, Non, N/A	
Taille des données	
Taille en Mo	
Nombre d'enregistrements	
Nombre de lignes	

Informations sur la collecte

Périodicité de collecte	!
Quotidien, hebdomadaire, mensuel, annuel, ..	
Date de disponibilité des données	
Après collecte	
Méthode de collecte	
Manière dont est collectée la donnée	

Informations sur la classification

Classification Loi 05-20	!
Très secret / Secret / Confidentiel Diffusion restreinte / Non classifié	
Niveau d'accès loi 31-13	!
Public / Public après traitement/ Accès restreint	
Statut de validation du niveau d'accès	!
Validée par le responsable de structure?	
Licences de réutilisation	!
Quelle licence associée	

Ouverture du jeu de données

Traitements avant publication	
Documentation des traitements à appliquer (anonymisation, transformation, transcodage, ..)	
Qualité des données	
Selon le scoring proposé	
Accessibilité des données	
Selon le scoring proposé	
Valeur du jeu de données	
Selon le scoring proposé	

Session 4: Inventaire, classification et référentiels

Valorisation

Mesurer la valeur de la data d'un point de vue Open Data

L'objectif ici est de focaliser les efforts vers des données à haute valeur ajoutée, pour maximiser leur impact. Cela peut se faire via un système de scoring de la data

Critères de qualité

Données lisibles par une machine Score 0 ou 3
Données existantes sous format ouvert Score 0 ou 3
Données mises à jour suivant la périodicité Score 0 ou 3
Données désagrégées Score 0 à 4
Dictionnaire des données est fourni et est complet Score 0 ou 3
Existence d'un schéma de validation Score 0 ou 2
Données utilisant les standards techniques Score 0 ou 1
Données utilisant des référentiels Score 0 ou 1

Critères facilité de publication

Les données sont disponibles sous format digital Score 0 à 3
Le système d'information offre des fonctionnalités d'extraction automatique Score 0 à 2
Les fonctionnalités d'extraction permettent d'obtenir des données structurées lisibles par une machine Score 0 ou 1
Les fonctionnalités d'extraction permettent d'obtenir des données répondant aux standards techniques Score 0 ou 1
Les fonctionnalités d'extraction permettent de générer les métadonnées importantes (dictionnaire, schéma de données...) Score 0 ou 1
Les données ne nécessitent pas de traitement d'anonymisation Score 0 ou 2

Critères de valeur de la data

Les données sont utilisées au sein de la structure Score 0 à 4
Les données sont utilisées/ demandées au sein de l'Administration Score 0-2-5-8
Les données sont utilisées/demandées par des acteurs non gouvernementaux Score 0 ou 2



Le portail Open Data propose un guide méthodologique pour l'élaboration de l'inventaire

[Source \(ADD-Guide méthodologique pour l'inventaire\)](#)

Séance 4:

Inventaire, classification et référentiels

Section 4.2:

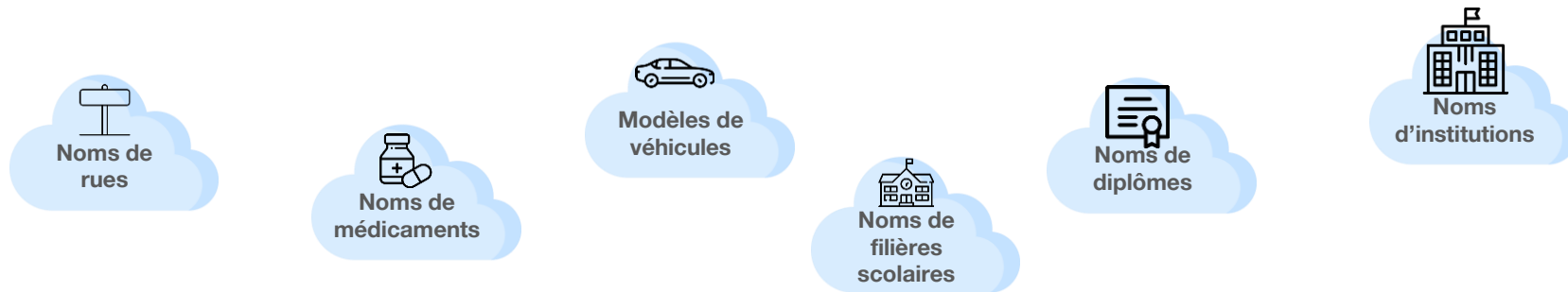
Référentiels de données communs

Session 4: Inventaire, classification et référentiels

Référentiels de données communs: Pourquoi?

Le référentiel comme besoin fondamental

Certaines types de données sont fréquemment utilisées pour sécuriser la qualité et standardisation de l'information.



Données 'connues'

Données qui ne sont pas des variables, mais plutôt figées dans le temps, malgré leur évolution.

ex. Noms de rues, Codes Postaux

Données régulières

Utilisées régulièrement et couramment dans de nombreux domaines.

ex. dans le cas des rues et Codes postaux, utilisé pour adresser un courrier, pour souscrire à un service public, ...

Inputs différents

Saisis dans les SI de façons différentes (structure, orthographe,)

ex. Avenue des FAR, Av. des FAR, Avenue des Forces Armées Royales, ..

Session 4: Inventaire, classification et référentiels

Avantage d'utiliser un référentiel commun: Exemple

Cas de la Classification Internationale des Maladies (CIM)

La Classification internationale des maladies (CIM) est un référentiel standardisé par l'OMS pour coder, analyser et comparer les maladies, troubles et causes de décès à l'échelle mondiale.

L'analyse des causes de décès est utilisée pour détecter les tendances sanitaires, allouer les ressources. Elles sont également transmises à l'OMS (Organisation Mondiale de la Santé) pour des analyses globales et par pays.

Classification Internationale des Maladies (CIM-11) contenant 85 000 codes

ROYAUME DU MAROC
FORCES ARMÉES ROYALES
HOPITAL MILITAIRE
MOUTAY ISMAIL DES F.A.R.
MEKNES

CONSTATATION DE DECES

Décès survenu le 17/03/2021 à 15 h 15 mn

Prénom M. NOUNA Nom KACIMI cin 1173000

Age 68 ans Sexe F Nationalité Marocaine

Domicile Rue Mohamed VI 10114 Hy. B1, Marrakech

Lieu du décès (2) Hopital Militaire Moutay Ismail Meknes

Cause du décès Choc cardiogénique

A Meknes, le 17/03/2021
Le Médecin traitant

Plusieurs libellés pouvant désigner la même maladie

"Diabète juvénile insulino-dépendant"

"Hyperglycémie chronique nécessitant de l'insuline"

"Insuffisance pancréatique insulino-dépendante"

"Maladie métabolique chronique (diabète sucré type 1)"

Elles désignent toutes la maladie

"Diabète sucré de type 1-code E10 dans le référentiel CIM-10"

Diag	Intitulé CIM10
E10	Diabète sucré de type 1
E100	Diabète sucré de type 1, avec coma
E101	Diabète sucré de type 1, avec acidocétose
E102	Diabète sucré de type 1, avec complications rénales
E103	Diabète sucré de type 1, avec complications oculaires
E104	Diabète sucré de type 1, avec complications neurologiques
E105	Diabète sucré de type 1, avec complications vasculaires périphériques
E106	Diabète sucré de type 1, avec autres complications précisées
E107	Diabète sucré de type 1, avec complications multiples
E108	Diabète sucré de type 1, avec complications non précisées
E109	Diabète sucré de type 1, sans complication

Session 4: Inventaire, classification et référentiels

Avantages d'utiliser un référentiel commun

Données pouvant faire l'objet de référentiels

Ci dessous des exemples de données généralement fournies sous forme de référentiel commun dans le monde



Adresses

Noms de Rues
Codes Postaux
Villes & Communes



Identifiants d'entités

Identifiant Commun des Entreprises
Répertoire des services publics
Liste des tribunaux
Établissements hôteliers, ...



Libellés

Produits (Distribution)
Maladies & causes de décès
Motif de contravention



Economie

Nomenclature Marocaine des
Activités (NMA)
Nomenclature douanière



Transport

Noms des gares
Noms des aéroports
Noms des ports
Noms des arrêts (transport public)



Education

Filières académiques
Etablissements scolaires

Session 4: Inventaire, classification et référentiels

Avantages d'utiliser un référentiel commun

Le double avantage du référentiel

L'utilisation de référentiels présente un double avantage: il permet de contrôler et enrichir la donnée

enrichissement de données

contrôle de données

CODE	NOM	DCI	DOSAGE1	UNITE_DOS/FORME	PRESENTATION	PPV	PH	PRIX_BR	PRINCIPAUX
6118001220668	URO / EAU POUR IRRIGATION	EAU POUR PREPARATION INJECTABLE	2000	ML	SOLUTION POUR IRRIGATION	35,00	-	35,00	0%
61180010118220	ELOXATINE 5 MG/ML	OXAUPLATINE	200	MG	SOLUTION A DILUER POUR PERFUSION	2 882,00	2 555,00	2 882,00	P 70%
6118001081646	ELOXATINE 5 MG/ML	OXAUPLATINE	200	MG	SOLUTION A DILUER POUR PERFUSION	2 882,00	2 555,00	2 882,00	P 70%
6118001182459	VIVALAN	VILOXAZINE	100	MG	COMPRIME PELLICULE	60,70	40,10	60,70	P 0%
6118000120155	ZADRYL	CETRIZINE	1	MG	SOLUTION BUVALE	31,90	19,90	31,90	G 70%
6118000241256	MYNAZOL	FLUCONAZOLE	50	MG	GELULE	87,00	54,20	87,00	G 70%
6118000091974	ALFACEFAL 125 MG/5 ML	CEFACLOR	1,5	G	POUDRE POUR SUSPENSION BUVALE	45,00	29,80	45,00	G 0%
6118000041344	INDCAN 50 MG	FLUCONAZOLE	50	MG	GELULE	35,00	21,80	35,00	G 70%
6118000100270	MEDIATOR	BENFLUOREX	150	MG	COMPRIME ENROBE	60,00	39,70	60,00	P 0%
6118000071068	GASTROLIBER	LANSOPRAZOLE	30	MG	GELULE GASTRO-RESISTANTE	75,00	46,70	75,00	G 70%
6118000180142	CLOFENE	DICLOFENAC	25	MG	COMPRIME ENROBE GASTRO-RESISTANT	20,50	12,80	20,00	P 70%
6118001170456	ZELDOX	ZIPRASIDONE	40	MG	GELULE	319,00	211,00	318,00	P 70%
6118001170333	TRIFLUCAN	FLUCONAZOLE	50	MG	SOLUTION POUR PERFUSION	144,30	89,60	144,10	P 70%
6118000031079	AVT	ACIDE ACETYL SALICYLIQUE / THIAMINE (VITAMINE B1)	600 / 5	MG / MG	SUPPOSITOIRE	10,50	6,90	10,50	G 0%
6118000150121	KLONOPIN	CLONAZEPAM	2	MG	COMPRIME QUADRISECALE	41,20	25,70	41,20	P 70%
6118000032182	HYDROXO 5000	HYDROXOCOBALAMINE	5000	µg	SOLUTION INJECTABLE	58,50	38,70	58,50	P 0%
6118000140160	DISPAMIX	AMOXICILLINE	500	MG	COMPRIME DISPENSIBLE	34,60	21,50	34,60	G 70%
6118000180418	MUSARIL	TETRAZEPAM	50	MG	COMPRIME ENROBE	51,80	34,30	51,80	G 0%
6118000100225	PRETERAX	INDAPAMIDE / PERINDOPRIL	0,625 / 2	MG / MG	COMPRIME SECABLE	239,00	-	239,00	P 70%
6118001080304	CLAFORAN	CEFTAXIME	1	G	POUDRE POUR SOLUTION INJECTABLE	69,00	43,00	69,00	P 70%
6118000080497	RHINOFEBRAL	PARACETAMOL / CHLORPHENAMINE	240 / 3,2	MG / MG	GELULE	14,30	9,50	14,30	P 0%
6118000091547	RIFASONE	RIFAMYCINE / PREDNISOLONE	0,25 / 1	G / G	POMADE	12,00	-	12,00	G 0%
6118000240426	FIBRON	AMBIROXOL	3	%	SOLUTION BUVALE	26,70	26,70	38,95	G 0%
6118000160168	AMOXIL	AMOXICILLINE	500	MG	POUDRE POUR SUSPENSION BUVALE	39,90	24,30	39,90	P 70%
6118000240150	BENCLAMID	GUBENCLAMIDE	5	MG	COMPRIME	36,20	22,60	36,20	G 70%
6118000250357	UNASYN	AMPCILLINE / SULBACTAM	1 / 0,5	G / G	POUDRE POUR SOLUTION INJECTABLE	42,00	26,20	42,00	P 70%
6118000060451	LASILX	FUROSEMIDE	20	MG	SOLUTION INJECTABLE	7,40	4,60	5,60	P 70%
6118018020155	ILATOR	SALBUTAMOL	2,5	MG	SOLUTION POUR INHALATION PAR NEBULISSEUR	41,20	25,70	41,20	G 70%
6118001101597	ILATOR	SALBUTAMOL	2,5	MG	SOLUTION POUR INHALATION PAR NEBULISSEUR	41,20	25,70	41,20	G 70%
6118010024320	FORMORTAL	FORMOCORTAL	0,05	%	COLLYRE	17,80	-	17,80	G 0%
6118000080275	GENTAGAM 120	GENTAMICINE	120	MG	SOLUTION INJECTABLE	15,80	9,80	15,80	G 70%
6118001070763	UMILINE PROFIL 30 PEN	INSULINE HUMAINE	100	UI	SUSPENSION INJECTABLE INTERMEDIAIRE	104,00	65,00	104,00	P 70%
6118000060567	BIPRODOGYL	SPIRAMYCINE / METRONIDAZOLE	1,5 / 250	MG / MG	COMPRIME PELLICULE	102,00	63,50	79,90	P 70%
6118000041207	GLEM	OLMEPRIDE	20	MG	COMPRIME	25,00	25,00	40,00	G 70%
6118000120475	SALAZOPYRIN	SULFASALAZINE	500	MG	COMPRIME ENROBE GASTRO-RESISTANT	141,70	88,30	141,70	P 70%
6118001070688	XANTHILUM LP	THEOPHYLINE	300	MG	GELULE LP	51,50	32,20	51,50	G 70%
6118000050414	AIRLX	CETRIZINE	1	MG	SOLUTION BUVALE	31,90	19,90	31,90	G 70%
6118000120384	GYNOMYK	BUTOCONAZOLE	100	MG	OVULE	62,20	41,10	62,20	P 0%
6118001050642	AVASTIN	BEVACIZUMAB	100	MG	SOLUTION A DILUER POUR PERFUSION	3 963,00	3 688,00	3 963,00	P 70%
6118001220159	FLUMAX 6000 UI ANTI-XA 0,6 ML	ENOXAPARINE	6000	UI	SOLUTION INJECTABLE	1 245,00	841,0	1 245,00	G 70%
6118001182725	PARAPLATINE	CARBOPLATINE	50	MG	SOLUTION POUR PERFUSION	835,30	552,50	835,30	P 70%
6118001110210	TRIFAX IV	CETIRIZINE	100	MG	POUDRE POUR SOLUTION INJECTABLE	61,00	38,00	61,00	G 70%

Source (Référentiel des médicament - Portail National)

Session 4: Inventaire, classification et référentiels

Avantages d'utiliser un référentiel commun

Le référentiel comme base solide pour la Data

La publication de référentiels communs permet d'accroître la valeur fournie par la data



Une meilleure compréhension

Avec une base commune d'appellation et de description des éléments

ex. dans le cas du référentiel des diplômes, il permet d'unifier la façon dont les recrutements & évaluations sont réalisés



Une meilleure réutilisation

Avec un enrichissement de données, augmentant ainsi la valeur des données

ex. dans le cas d'un référentiel des entreprises publié, l'information ICE permet d'enrichir avec la raison sociale, adresse, activité, ...



Une meilleure interopérabilité

Rendre l'information exploitable par tous les acteurs de l'écosystème

ex. un fichier produit par une administration est directement exploitable et intégrable dans les SI, avec un effort d'intégration minimum

Session 4: Inventaire, classification et référentiels

Exemples de référentiels disponibles

Disponibles sur le Portail National (data.gov.ma)



Codes postaux



Villes & Villages



Administrations
publiques



Etablissements
de santé



Secteurs d'Activité
économique

[Source](#)

Disponibles sur le Portail HCP (hcp.ma)



Nomenclature
Marocaine des
Activités (NMA)



Nomenclature
des diplômes



Codes
géographiques

[Source](#)

Note: l'ADD est entrain d'élaborer les référentiels des données communs prioritaires en concertation avec les acteurs concernés

Exemples de référentiels universels



Transport

**Codes des aéroports
IATA**

Codes standardisés à trois lettres attribués aux aéroports du monde entier

92

[Source \(IATA.org\)](#)



Economie

**Code des devises
(ISO 4217)**

Codification des devises mondiales en trois lettres (ex. : USD, EUR, MAD).

[Source \(IBAN.org\)](#)



Agriculture

Ontologie géopolitique

Référentiel pour les noms officiels des pays, codes et régions géographiques utilisés dans les statistiques mondiales agricoles.

[Source \(ISO.org\)](#)



Science

**Nomenclature des
produits chimiques**

Base de données internationale pour l'identification des substances chimiques,

[Source \(CAS.org\)](#)



Commerce

**Code HS (Harmonized
system)**

Codification des marchandises pour les échanges internationaux, utilisée par la plupart des pays

[Source \(trade.gov\)](#)

Session 4: Inventaire, classification et référentiels

Formats et Consistance des référentiels de données disponibles

Formats de référentiels courants

Les référentiels sont la plupart du temps dans des formats similaires aux jeux de données, avec les mêmes critères techniques



Fichiers plats (CSV, TSV)
exemple: Annuaire des agences de voyage sur data.gov.ma



Fichiers structurés (XML, JSON)



API ou Webservices

Fichiers⁽⁴⁾Réutilisations et API⁽⁵⁾Discussions⁽⁴⁾Ressources communautaires⁽⁰⁾Informations

3 FICHIERS PRINCIPAUX

CSV

Mis à jour le 3 décembre 2024 — csv — 4K

↓ CSV

Json

Mis à jour le 2 décembre 2024 — json — 966

↓ JSON

XLS

Mis à jour le 18 mars 2016 — xls — 2K

↓ XLS

1 API

API

Mis à jour il y a un jour — 750

↓ FICHIER



Fichiers propriétaires



Fichiers difficilement exploitables (PDF par ex.)
non recommandé

Session 4: Inventaire, classification et référentiels

Démarche d'élaboration de référentiels

Pour quelle raison produire un référentiel de données

Un organisme peut être amené à produire ses propres référentiels, pour des besoins internes ou pour standardiser l'utilisation de ses données à l'extérieur.



Pour ses besoins internes

Pour assurer une saisie et gestion cohérente de ses propres données

Exemple:

Une commune peut créer un référentiel des établissements publics gérés, pour avoir un enregistrement des noms harmonieux à travers les différents rapports, publications, ...
ex. "Ecole Primaire Al Fath", "Ecole Al Fath"



Pour standardiser pour les tiers

Pour assurer que ses données utilisées par des acteurs tiers le soient de façon harmonieuse

Exemple:

Une ville peut créer un référentiel des événements culturels avec des informations pertinentes. Ces données peuvent être utilisées par des médias locaux, agences de tourisme, ..

Séance 5: Evaluation de la qualité des données & Bonnes pratiques

13h30-14h30

Séance 5:

Evaluation de la qualité des données & Bonnes pratiques

Section 5.1:

Evaluer la qualité des données

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Qualité de la data

Qualité de la Data = Qualité?

Durant la 2e Guerre Mondiale

Les bombardiers des Alliés revenaient de mission très endommagés par des tirs ennemis.

Une équipe de scientifiques et statisticiens (Statistical Research Group) a été sollicitée pour trouver une solution et renforcer / blinder les zones qui doivent l'être.

Ils ont collecté des données sur les zones d'impact (figure)

Quelles zones devaient être blindées?

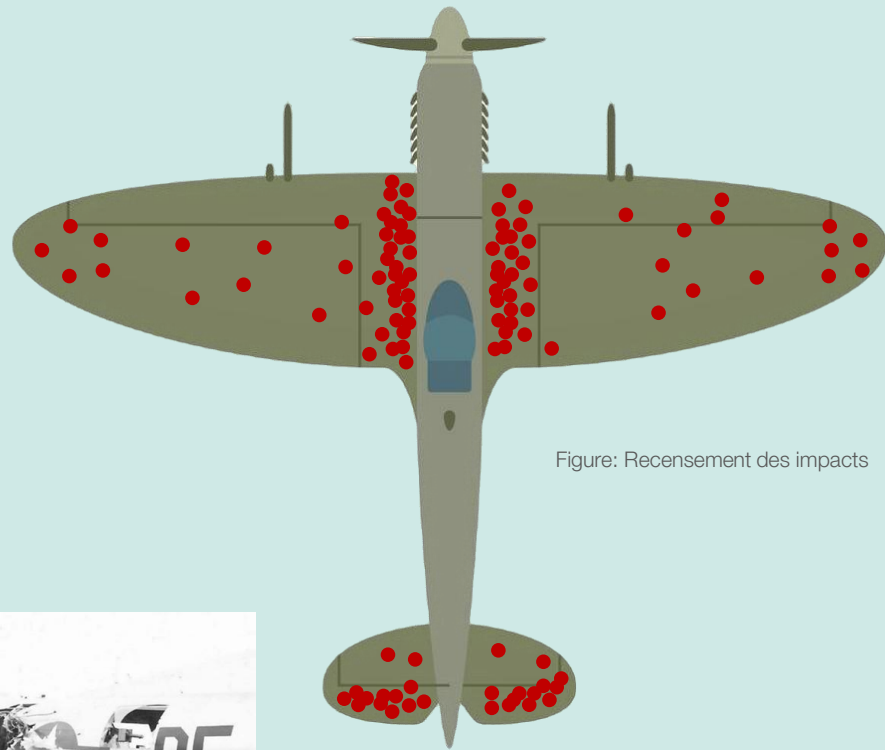


Figure: Recensement des impacts

Intuitivement les zones impactées doivent être renforcées

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Qualité de la data



Abraham Wald

Mathématicien et membre de
l'équipe Statistical Research
Group
1902-1950

Le biais du survivant

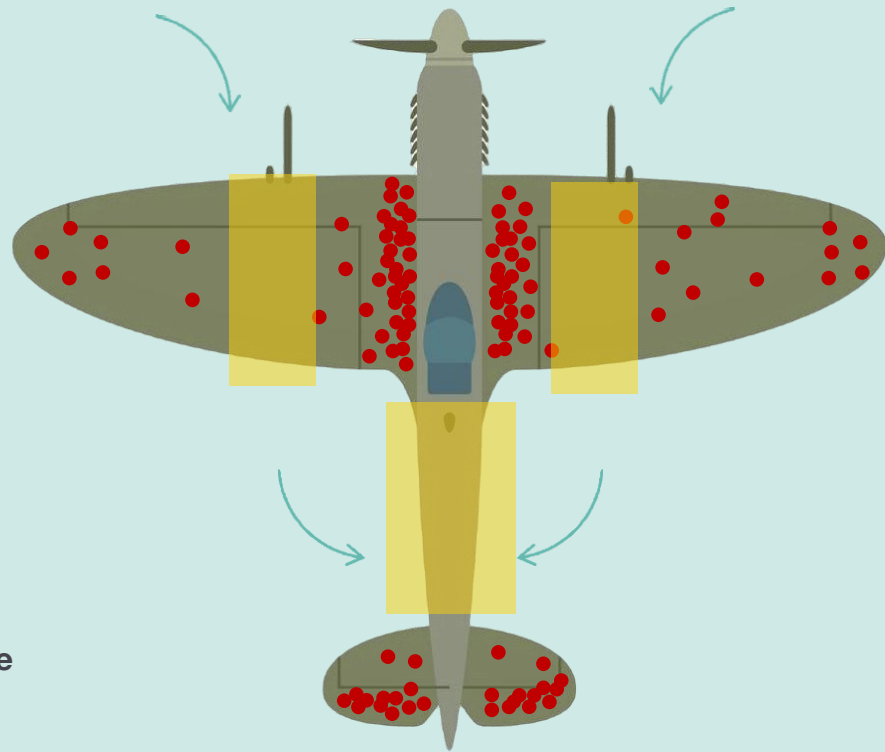
Son raisonnement ?

Les impacts de balles sur les avions « survivants » (d'où l'expression) **indiquent en réalité les points forts** :

les zones sur lesquelles un avion pouvait être touché et continuer malgré tout à voler.

En l'occurrence, celles qui précisément n'avaient **pas besoin de renforcement.**

Le blindage devait donc être posé aux **endroits sans impact.**



[Source \(Le biais du survivant - l'ADN Data\)](#)

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Les différents aspects de la Qualité

Maximiser l'impact de la donnée publiée

La qualité de la data ne se limite pas à la DATA elle-même, mais également à son contexte, la compréhension de sa portée et son exploitabilité.



Clarté de la data

La donnée fournie doit documentée avec des métadonnées claires pour comprendre l'usage et les limites des données.

Faciliter l'accès par API et mises à jour en temps réel



Interopérabilité

Le format de fichier est un des piliers de la qualité des données ouvertes pour permettre une interopérabilité, et donc une réutilisation efficace et permettant l'innovation



Fiabilité des données

Des données exactes et fiables sont la base d'un traitement de la data. Il est également essentiel que la data soit à jour et non obsolète

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Frameworks et méthodes d'évaluation

Différents frameworks, pour différents contextes

Parmi les multitudes de méthodologies d'évaluation, chacune va se concentrer sur un aspect ou dimension particulière. Tout en étant des méthodologies différentes, elles visent toutes à promouvoir une donnée ouverte, selon les bons standards, et avec la bonne implication organisationnelle et technologique de l'Administration Publique

Exemples de framework



Modèle des 5 étoiles (Tim Berners-Lee)

Évalue la qualité technique des données ouvertes.

1 étoile : Disponibilité en ligne, 5 étoiles : Interopérabilité et liens avec d'autres données.

Critère: Format & Interopérabilité des données

[Source \(5starsdata\)](#)



ODIN (Open Data Inventory)

Mesure la couverture et l'opérabilité des données ouvertes des statistiques nationales (ex. démographie, économie).

Critères : Disponibilité, accessibilité et qualité.

[Source \(ODIN: Open Data Inventory\)](#)



ODRA (Open Data Readiness Assessment)

Évalue la préparation institutionnelle d'un pays ou d'une organisation à mettre en œuvre une stratégie open data.

Critères : Cadre juridique, infrastructure, engagement des parties prenantes.

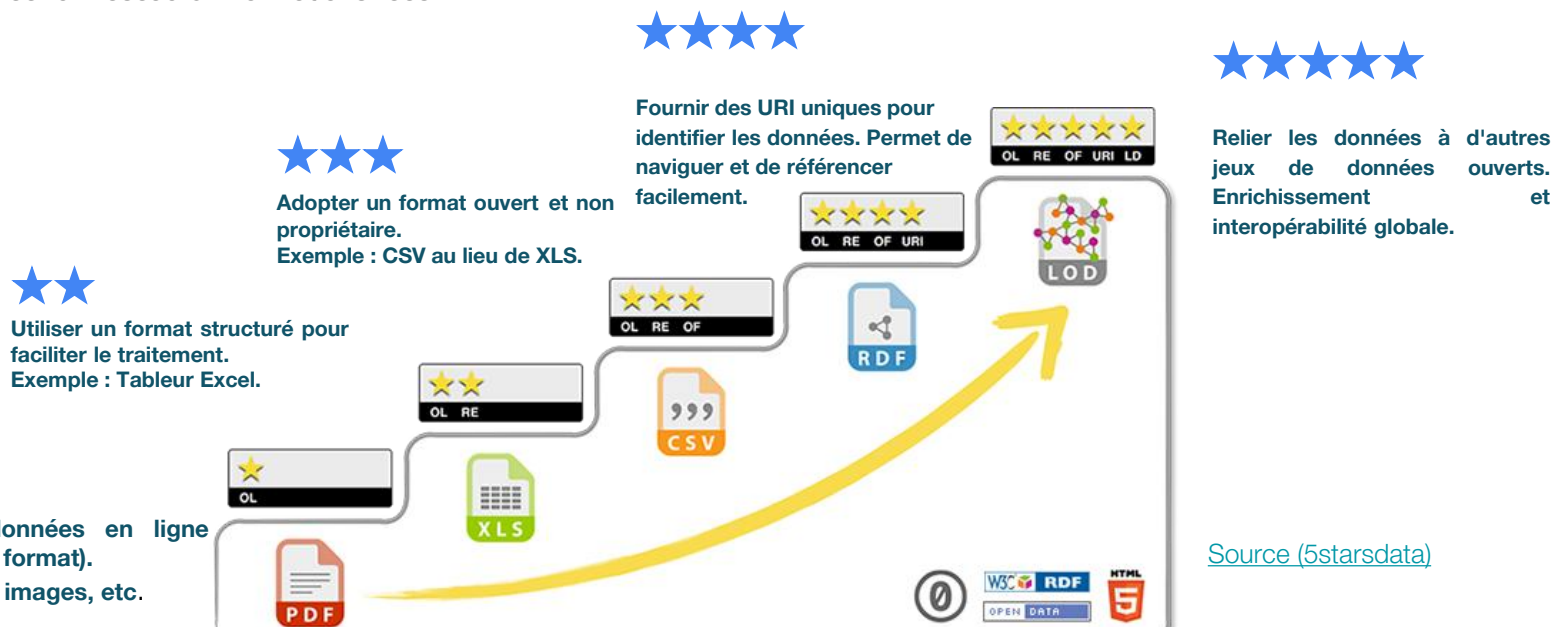
[Source \(Banque Mondiale\)](#)

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Évaluer l'interopérabilité

Le modèle 5 étoiles

Le modèle 5 étoiles de Tim Berners-Lee évalue la qualité des données ouvertes en promouvant leur accessibilité en ligne, leur structuration, l'utilisation de formats non-propriétaires, leur interopérabilité via des standards ouverts, et leur connexion à d'autres données pour créer un réseau d'informations liées.



Session 5: Evaluation de la qualité des données & Bonnes pratiques

Évaluer l'interopérabilité

Quelques exemples

The image displays two screenshots from the Casa Tramway website and app. The left screenshot shows the 'Lignes et Horaires' (Routes and Timetables) page with a navigation menu and a download button for 'Télécharger TM hiver VF (1).pdf'. The right screenshot shows a detailed timetable for line T3, listing departure times for 'De Hay Al Wahda' and 'De Casa Port' in both directions. Below the timetable is a map of the tramway network in Casablanca.

Publication des données

Actuel : Les horaires et itinéraires des bus et tramways sont souvent disponibles en PDF sur les sites des opérateurs (comme Casa Tram ou Alsa). ★

Limite : Ces fichiers sont difficiles à exploiter et n'offrent pas de possibilités d'analyse automatisée.

Amélioration: Utiliser le format GTFS (General Transit Feed Specification), un standard ouvert utilisé pour structurer les données de transport public. ★★

→ **Atteinte du 3 étoiles**

Autre amélioration: Interconnecter les data à d'autres jeux de données (ex. météo, événements affectant le trafic, points d'intérêt) ★★★★★

→ **Atteinte du 5 étoiles**

Source (CasaTramway)

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Évaluer l'accessibilité de la data et son impact

La méthodologie ODIN (Open Data Inventory)

Pour compléter le modèle TBL, l'Open Data Inventory (ODIN) propose une démarche plus globale de mesure de la qualité de la donnée ouverte, en incorporant des notions d'accessibilité, actualisation, etc..

La démarche ODIN se base sur

- La présence d'un format lisible par machine
- La data en format ouvert (non propriétaire)
- La présence de méta données exhaustives
- La présence de plusieurs canaux (fichier, API, ..)
- La présence d'une licence de réutilisation proprement documentée (Conditions d'utilisation)



L'Observatoire des Données Ouvertes (Open Data Watch) a, lors de sa session de 2022, classé le Maroc à la 25ème place sur 193 en matière d'Indice des Données Ouvertes (ODIN)

Dimension	Critères	Poids (%)	Description
Couverture des données	Critères spécifiques par secteur (démographie et population, économie, environnement et infrastructure)	60	Disponibilité des données Complétude temporelle (profondeur) Granularité (niveau de détail) Fréquence de mise à jour
	Ouverture des données	10	Données disponibles en ligne, gratuites, accessibles sans barrière Disponible sous API / Temps réel
Format	Format	10	Données publiées dans des formats ouverts et non-propriétaires (CSV, JSON, etc.).
	Documentation	10	Métadonnées claires sur la source, la portée, les limites et l'année de publication.
	Licence ouverte	5	Licence explicite autorisant la réutilisation des données.
	Mises à jour régulières	5	Données actualisées à des intervalles réguliers et définis.

[Source \(Classement ODIN\)](#)

[Source \(ODIN: Open Data Inventory\)](#)

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Autres méthodes d'évaluation de la data ouverte

L'Open Data Readiness Assessment (ODRA)

L'**Open Data Readiness Assessment (ODRA)** est une méthodologie développée par la Banque mondiale pour évaluer la préparation d'un pays ou d'une organisation à l'adoption de l'Open Data (adapté pour l'échelle du Pays ou d'une Institution étatique). Elle analyse huit dimensions essentielles pour identifier les opportunités et obstacles, et proposer des recommandations adaptées.

Les 8 dimensions de l'ODRA:



Leadership

Volonté politique et engagement des dirigeants à promouvoir l'Open Data.



Politiques et réglementation

Analyse des cadres législatifs et politiques pour soutenir l'ouverture des données.



Institutions et capacités

Capacité des institutions à gérer et publier des données ouvertes.



Données

Disponibilité, qualité et accessibilité des données déjà existantes.



Demandes des utilisateurs

Identification des besoins et attentes des utilisateurs finaux des données.



Technologies

Infrastructure technique pour stocker, gérer et partager des données ouvertes.



Engagement citoyen

Participation des citoyens et implication dans les initiatives Open Data.



Écosystème

Collaboration entre les parties prenantes : gouvernement, entreprises, société civile, et autres acteurs.

[Source \(Banque Mondiale\)](#)

[Source \(Banque Mondiale-Guide\)](#)

Diagnostic : Collecte d'informations à travers des entretiens, des ateliers et l'analyse documentaire.

Recommandations : Actions concrètes pour remédier aux lacunes identifiées et maximiser les opportunités.

Plan d'action : Feuille de route adaptée aux priorités et ressources disponibles.

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Autres méthodes d'évaluation de la data ouverte

Le self-scoring de Data.gouv.fr

Le score de qualité des métadonnées introduit par data.gouv.fr vise à aider les utilisateurs à identifier les jeux de données ayant des métadonnées complètes et fiables.

Critère	Description
Description des données	La description des données est de qualité (la description du jeu de données est suffisamment longue).
Mise à jour	La fréquence de mise à jour est renseignée et respectée.
Licence	La licence est renseignée et ouverte.
Métadonnées des ressources	Présence d'au moins une ressource avec un format ouvert déclaré.
Couverture spatiale	La couverture et la granularité spatiale est renseignée.
Couverture temporelle	La couverture temporelle des données est renseignée.



[Source \(Open Data France\)](#)

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Évaluer la fiabilité de la donnée

Une fois qu'on a la data (courante et historique)

Une Administration publique gère un volume important de data: opérationnelle, métier, technique, administrative, financière, etc.

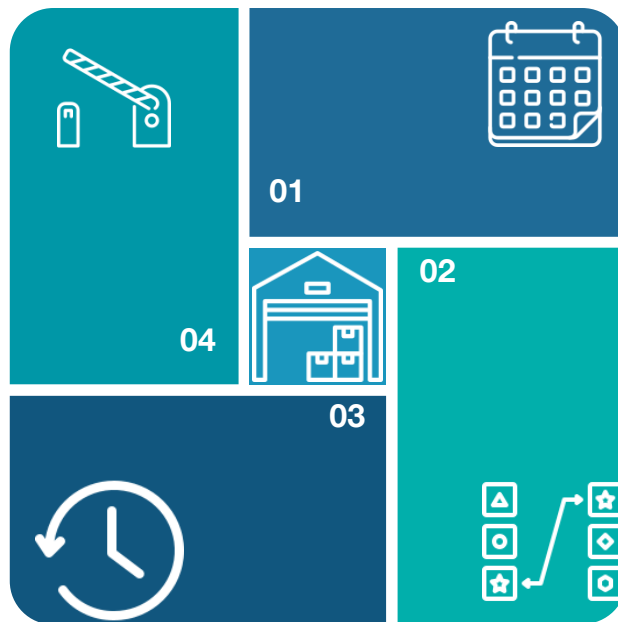
Data contrôlée à la source

Data provenant de formulaires web bien conçus,
champs auto contrôlés dans les SI, ...
Taux de qualité correct



Historique et autres data

Provenant différentes années, familles métiers, etc.;
dont la data n'a pas pu être contrôlée à la source
Taux de qualité inconnu



Tranches incomplètes

Tranches d'années dans l'historique ou
Départements particuliers avec des zones dont la
qualité est incertaine (donnée non collectée,
changement organisationnel..)



Tranches différentes

Avec des règles de gestion ou contextes différents
(donnée collectée d'une autre façon, avec une autre
structure, ...)



Session 5: Evaluation de la qualité des données & Bonnes pratiques

Évaluer la fiabilité de la donnée

Mesurer la qualité de la data stockée grâce à la norme ISO 2859

Malgré toutes les mesures mises en place, il est essentiel de pouvoir mesurer la qualité de la data disponible. Cela est d'autant plus vrai qu'une majorité de la data chez les institutions remonte à une profondeur historique importante dont la qualité n'est pas forcément connue.



Sample size Code Letter		Acceptance Quality Limit (AQL) in %																	
		0.010	0.025	0.040	0.060	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40
A	2																		
B	3																		
C	5																		
D	8																		
E	13																		
F	20																		
G	32																		
H	50																		
I	80																		
J	125																		
K	200																		
L	315																		
M	500																		
N	800																		
P	1250																		
Q	2000																		
R	3150																		

* Limited to AQL of 10%

Source ISO2859

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Fiabiliser les données

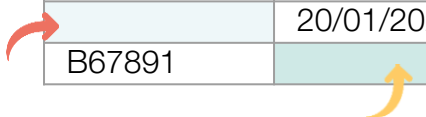
Bien définir ce qu'est une donnée correcte

La fiabilisation du patrimoine Data d'une institution est une opération titanesque. Afin d'avoir des résultats concrets et atteignables, il est essentiel d'adopter une approche stratégique, visant cibler les données que l'on souhaite fiabiliser, ainsi qu'une méthode industrialisée pour ce faire.

Complétude (présence)

- Une donnée obligatoire manquante n'a pas le même impact qu'une donnée facultative
- En tenir compte dans la notation

CIN	Date	Montant
A12345	20/01/2020	5000.00
	20/01/2020	5000.00
B67891		2500.00



Poids de la qualité

- Toutes les erreurs ne se valent pas
- En tenir compte dans la notation (pondération par exemple ou score)

Date: 20/01/2020 ↔ Date: 01/01/2020

CIN: A12345 ↔ CIN: A123

Session 5: Evaluation de la qualité des données & Bonnes pratiques

Fiabiliser les données

Adopter une approche stratégique pour la fiabilisation

La fiabilisation du patrimoine Data d'une institution est une opération titanesque. Afin d'avoir des résultats concrets et atteignables, il est essentiel d'adopter une approche stratégique, visant cibler les données que l'on souhaite fiabiliser, ainsi qu'une méthode industrialisée pour ce faire.



Prioriser par valeur

- Identifier les familles de données qui génèrent le plus de valeur (généralement les données métier)
- Prioriser les opérations de contrôle et fiabilisation sur les populations de données sélectionnées

Cas d'exemple:

Les Banques ont souvent recours à cette approche sur différentes lignes métier: Garanties & Crédits, Comptes, Recouvrement & Juridique, ...



Prioriser par besoins

- Prioriser selon les besoins émanant de l'intérieur ou extérieur
- Cela peut concerner des types de processus ou des tranches d'années, ou des régions

Cas d'exemple:

Il est possible de mettre en place une cellule de contrôle qui fiabilise les données avant publication, en continu
exemple: Fiabilisation de l'inventaire physique (très courant)



Prioriser par règles

- Utiliser des règles de gestion pour détecter les populations de données les plus susceptibles d'avoir des erreurs
- Extraire et contrôler les données n'ayant pas respecté les règles de gestion

Cas d'exemple:

En 2013, l'OMPIC a développé une solution de fiabilisation industrielle et paramétrable des données, visant à fiabiliser sur plusieurs années les données du Registre de Commerce. Les campagnes ont été proposées selon différents critères (CA, Région, Secteur) ou règles métier (Incohérence de données financières, incohérences métier, ...)



Fiabilisation industrielle

- Il peut être nécessaire de développer des **outils** de fiabilisation et de traitement industriel du contrôle (suivi, traçabilité, automatisation)
- Également possible de sous-traiter le traitement de la data

Pour terminer

Synthèse de la journée

Les grandes lignes

Les éléments vu au cours de l'atelier peuvent se résumer de la façon suivante



La data comme nouveau pétrole

Comme le pétrole, elle n'a de valeur que dans son potentiel, après transformation et création de valeur



L'Open Data comme vecteur de progrès

Dans le monde, l'Open Data a contribué à la croissance, progrès, innovation, inclusion sociale, ...



Le Maroc a lancé un chantier Open Data

Le cadre légal et opérationnel est en place. L'ADD encourage à publier sur le portail (Rôle central des ROD/GdD)



La Région peut tirer profit de l'Open Data

Transparence, Innovation, Optimisation, Croissance économique, etc.



L'Open Data a des standards techniques

Formats de fichiers, structuration des données, anonymisation, mise à jour régulières



Les référentiels de données: une nécessité

L'utilisation et le partage de référentiels de données communs est vital pour garantir la qualité



Les instruments de mesure de la qualité

La qualité de la data ne se limite pas à la data elle-même, mais à sa clarté et interopérabilité



Tirer profit des retours d'expérience

Les défis fréquents sont techniques ou organisationnels, avec une liste de bonnes pratiques

Questionnaire D'évaluation



Merci pour votre attention et pour votre
mobilisation

Data.gov.ma