



Where ICT meets Social &
Economic Development

Atelier Qualité des jeux de données

- Khaled Sallemi (UAE)
- Rim Garnaoui (UAE)
- François Daoust (SBC4D)
- Ibrahim Elghandour (BM)





Heure	Contenu
09:00 - 09:30	Cadrage de l'atelier
09:30 - 10:00	Mesurer la qualité des jeux de données
10:00 - 10:45	Présentation des référentiels de données existants par l'INS
10:45 - 12:30	Discussions sur les éléments du décret
12:30 - 13:00	Debriefing & Résumé





1. Fournir un cadre juridique de réutilisation de l'information absent de la loi d'accès à l'information
2. Fournir un cadre juridique spécifique à l'Open Data
3. Fournir un cadre juridique à la mise en œuvre de l'initiative Open Data
4. Homogénéiser les actions et les approches de l'administration dans son ensemble





1. Développement d'un premier draft (2 mois)
 1. Sensibilisation
 2. Formation d'un groupe de travail et atelier techniques
 3. Première version
 4. Commentaires et version finale
2. Organisation d'une consultation nationale (1 mois)
 - Une conférence de présentation
 - Une période de collecte des commentaires
 - Une version finale
3. Adoption officielle du décret
4. Formation et sensibilisation de l'administration et des acteurs non gouvernementaux





1. Identifier les sections du décret

- Quels éléments sont nécessaires et non-contradictaires avec les législations actuelles?

2. Identifier et choisir les options en fonction de leurs impacts

3. Opter pour un niveau de détails

- Continuum entre un concept et une application spécifique: Continuum entre la définition d'un cadre et une certaine liberté vs une homogénéité dans l'implémentation





- Gouvernance: Quelle autorité et quelle gouvernance pour l'initiative Open Data?
 - Quel plan de suivi et évaluation?
- Acteurs: Quels organismes sont concernés et quelles sont leur obligations?
- Données: Quelles données doivent être publiées?
- Publication: Comment doivent être publiées les données?
- Réutilisation: Comment peuvent être réutilisées les données publiées?
 - Les licences de réutilisation
- Réutilisation: Comment favoriser la réutilisation et L'implication des acteurs non-gouvernementaux?





Recouvrir un ensemble large de critères:

- L'utilisation de référentiels de données (présentation de l'INS à la fin de cette section)
- Choix du/des formats de données de publication
- Gestion des versions, de l'historique et des erreurs
- Documentation / Langue(s) de publication
- Persistance des références
- Note sur une échelle donnée (5 ★, certificat ODI)
- Niveau de désagrégation





- **But:**

- Chacun peut chercher trouver ces données en ligne
- Chacun peut récupérer ces données (directement ou via une API) dans un format:
 - lisible par l'homme
 - **exploitable par un ordinateur**



CSV, JSON, XML, HTML, RDF, GeoTIFF, GeoJSON, ShapeFile, etc.

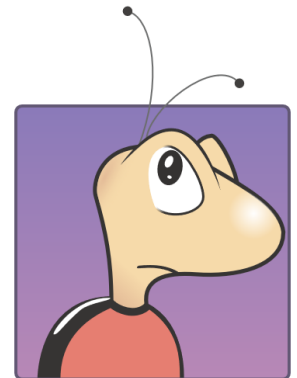


PDF, DOC, PNG, GIF, JPEG, etc.

Gestion des versions, historique, erreurs

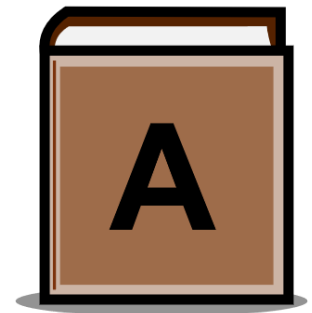


- Suivi des rapports d'erreurs remontées par les utilisateurs
- Identification des versions d'un jeu de données
- Maintenance des jeux de données anciens





- Description des jeux de données
- Description des métadonnées
- Description des référentiels de données existants
- Description des critères d'évaluation de la qualité
- Description des processus de publication et de maintenance
- ...





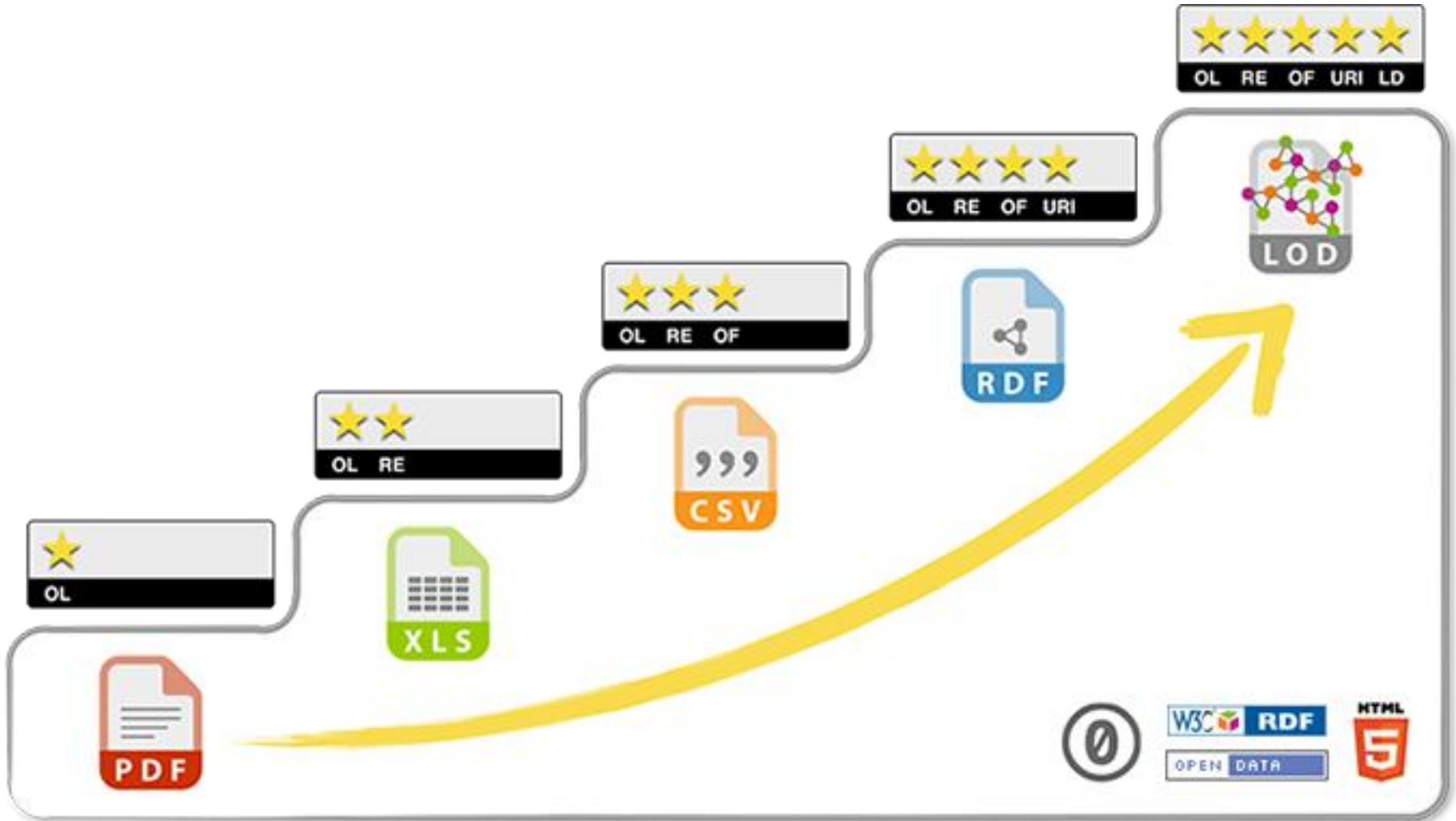
`http://{base}/{type}[/{domain}][/{sector}][/{resource}]`

- L'utilisation d'**URIs** est à la base du Web
- L'utilisation d'adresses **HTTP** permet de chercher la ressource
- Les identifiants doivent être **pérennes** et **logiques**, la recherche d'informations en sera grandement facilitée, et les liens entre les données renforcés et stables.
- Prévoir une gestion des **versions** de chaque document!

Bonnes pratiques:

<http://philarcher.org/diary/2013/uripersistence/>

Evaluer la qualité: L'échelle 5 ★



Source: <http://5stardata.info/>

Evaluer la maturité: L'échelle 5 ★



- ★
 - Publiez vos données sur le Web avec une licence ouverte (peu importe le format)
- ★★
 - Publiez vos données dans un format structuré (par exemple Excel au lieu d'une image tirée d'un scanner)
- ★★★
 - Utilisez un format ouvert (par exemple CSV au lieu d'Excel)
- ★★★★
 - Utilisez des URIs pour chaque objet, afin que n'importe qui puisse cibler vos données
- ★★★★★
 - Liez vos données avec d'autres données pour enrichir le contexte

Evaluer la maturité: Certificats Open Data ODI



- **BRONZE:**

Données publiées avec une license ouverte, sans restrictions, libres d'accès et légalement ré-utilisables.

- **SILVER:**

Niveau bronze, plus données publiées dans un format exploitable par un ordinateur, données fiables, support possible de l'éditeur via un canal de communication dédié.

- **GOLD:**

Niveau silver, plus données publiées dans un format standard et ouvert, avec des garanties de mises à jour régulières et de support élargi, une bonne documentation, et une déclaration de droits exploitable par un ordinateur.

- **PLATINUM:**

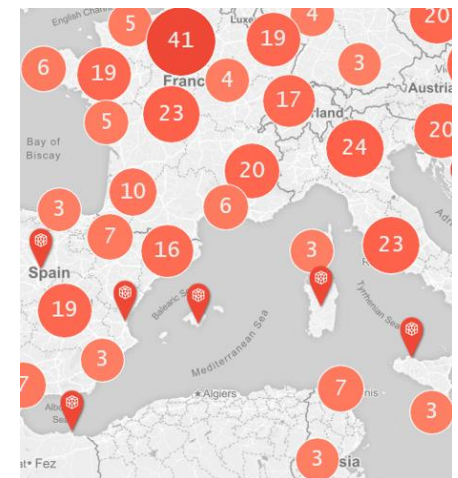
Niveau Gold, plus informations de provenances exploitables par un ordinateur, identifiants uniques, existence d'une équipe de communications chez l'éditeur.

<https://certificates.theodi.org/en/>





- Beaucoup de données géolocalisées
- Un bon niveau de désagrégation permet:
 - de conserver une taille acceptable pour certains jeux de données
 - de faciliter l'émergence d'applications locales
 - de faciliter le « mashup » de données dans ces applications





- Référentiels de données existants
- Autres référentiels possibles?





- Adoption de référentiels de données communs
- Adoption d'une échelle d'évaluation de la qualité, et d'un niveau cible
- Adoption d'une liste de langues de publication
- Adoption d'un schéma de persistance
- Adoption d'un niveau de désagrégation géographique minimum





- Adoption d'une échelle de qualité?
 - Si oui, laquelle/lesquelles? Niveau minimum?
- Prescrire l'utilisation de référentiels existants et futurs?
- Mentionner les formats de données recommandés?
 - Prescrire leur utilisation?
 - Multiplicité des formats?
 - Liste dans un futur décret?
- Prescrire une ou plusieurs langues de publication?
- Prescrire un niveau de désagrégation géographique minimum (Imada)?
- Prescrire la gestion des rapports d'erreurs?
- Prescrire la persistance des URLs?
- Prescrire un niveau de documentation des jeux de données dans une ou plusieurs langues?





Récapitulatif

THANK

YOU!