

D4D  
HUB

AFRICAN UNION EUROPEAN UNION

DIGITAL FOR DEVELOPMENT HUB

Union  
Africaine



#ADD

Agence de Développement du Digital

# Open data: Cadres techniques

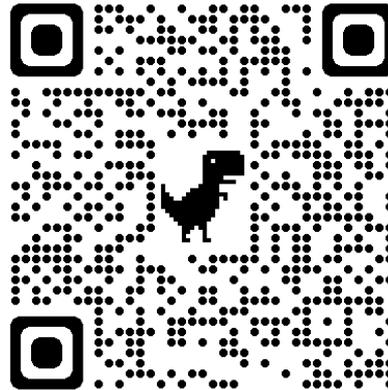
14/06/2023, Tanger



Soit, insérez le **#2466906** sur [slido.com](https://www.slido.com)

Soit, surfez a [t.ly/Xef6](https://t.ly/Xef6)

Quelles normes utilisez-vous ou envisagez-vous d'utiliser pour la publication de données ouvertes ?



# Contenu

- Standards techniques de l'Open Data (international)
- Normes de l'Open Data
- Référentiels de l'Open Data

# Introduction

Cette séance examine de plus près ce que sont réellement les données. Une compréhension de base des technologies de l'information est nécessaire pour comprendre comment publier des données ouvertes d'une manière accessible et utile aux réutilisateurs.

Cette séance ne s'adresse pas à des experts, mais donne un aperçu concret des concepts de base que les informaticiens traitent au quotidien. Une compréhension de haut niveau de ces concepts plutôt techniques permet à la direction de prendre des décisions efficaces concernant la gestion des données (ouvertes).

# Standards techniques de open data

## 1. Open Data est lisible par machine

Une condition importante pour pouvoir parler de données ouvertes est que ces données soient « lisibles par machine ». Dans cette série, nous discutons brièvement de la façon dont les ordinateurs stockent les données et de ce que signifie "lisible par machine".

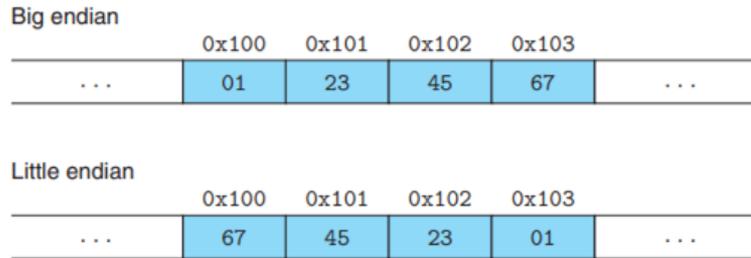
Notez que la condition est en fait "**librement** lisible par machine", ce qui signifie que les données doivent être lisibles sans logiciel propriétaire spécialisé.

En règle générale, si vous pouvez ouvrir un fichier dans Microsoft Notepad et le comprendre, il est librement lisible par machine. Par contre, il est toujours utile de savoir comment les données et le texte sont stockés.

# Standards techniques de open data

## 1. Open Data est lisible par machine

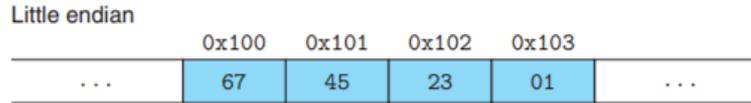
Toutes les données sont stockées en binaire, c'est-à-dire sous la forme d'une série de bits (0 ou 1). Mais comment pouvons-nous savoir ce que signifie ce flux d'informations ? La réponse est tout simplement de bons accords. Le plus simple exemple est le **direction** par laquelle une machine lit un byte, comparable à la façon qu'un homme lis des phrases du gauche à droit (Français) ou du droit à gauche (Arabe) :



# Standards techniques de open data

## 1. Open Data est lisible par machine

En effet, toutes les machines basées sur un architecture Intel x86 sont little-endian, donc on peut supposer que presque toutes les ordinateurs personnelles, sont little-endian. Donc il suffit de vous assurer que l'ordinateur avec lequel on traite les données pour la publication est comme ça, (c'est le cas 99% du temps) même pour des fichiers en Arabe.

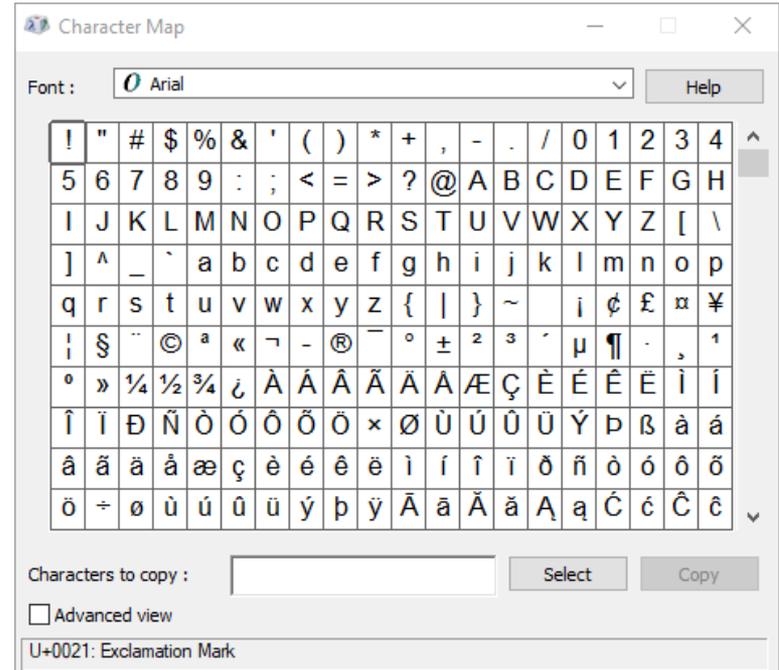


# Standards techniques de open data

## 1. Open Data est lisible par machine

Comme pour travailler avec différentes langues, nous devons nous mettre d'accord sur le nombre de caractères différents à utiliser et sur le nombre de caractères représentant un mot. Ici, nous pouvons être brefs : le "langage" le plus courant à ce niveau est **Unicode**, qui est utilisé dans le monde entier et peut décrire des caractères dans différentes langues humaines.

Pour voir le code "unicode" pour chaque caractère, par appuyer sur la touche windows et en tapant "charmap". Le code unicode se trouve en bas.



# Standards techniques de open data

## 1. Open Data est lisible par machine

Puis, on doit décider comment on va exprimer tous ces caractères en bytes. En bref, le codage le plus couramment utilisé est **UTF-8**, où chaque caractère est stocké dans 1 ou plusieurs octets (byte), selon la fréquence à laquelle il se produit. Pour enregistrer du texte dans le Notepad en Unicode, allez dans "Fichier -> Enregistrer sous..." et choisissez UniCode sous "Encodage :", appuyez sur "Enregistrer". De cette façon, vous serez sûr que tout le monde pourra ouvrir vos données.

CHARACTER	CODE POINT	UTF-8 BINARY ENCODING
A	U+0041	01000001
ज	U+0683	11011010 10000011
ଝ	U+0C9A	11100000 10110010 10011010
判	U+2070E	11110000 10100000 10011100 10001110
😄	U+1F601	11110000 10011111 10011000 10000001

# Standards techniques de open data

## amusant à savoir

Dans Microsoft Word et de nombreux autres programmes Windows, vous pouvez taper la valeur Unicode, puis appuyer sur Alt-X, et le symbole apparaîtra :

1F923 - Alt-X donne 

1F92A - Alt-X donne 

1F92F - Alt-X donne 

# Standards techniques de open data

## 1. Open Data est lisible par machine

En générale, deux types de données lisibles par machine peuvent être distingués :

- Données **non structurées**, telles que des fichiers texte ou une image. Les données sont lisibles, mais aucune distinction ne peut être faite entre les pièces ou composants individuels.
- Données **structurées**, où une structure est ajoutée aux données, par exemple dans une base de données.

# Standards techniques de open data

## 2. Des données structurées

Pour comprendre comment les données structurées sont stockées et échangées, il est important de comprendre les structures derrière ces données. Nous considérons souvent les données comme une liste plate (une « chaîne ») d'informations, mais il existe plusieurs façons de relier les parties de la liste. Afin de stocker et d'échanger de telles relations, nous devons chercher des moyens de formaliser également la structure des parties les unes avec les autres.

**Une bonne compréhension des structures de données vous permettra d'estimer dans quel format vous échangez le mieux les données** (ouvertes).

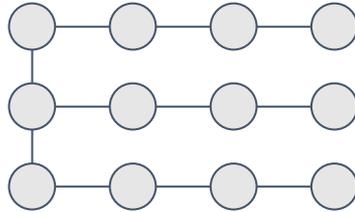
Dans la diapositive suivante, nous donnons un aperçu des structures de données les plus importantes

# Standards techniques de open data

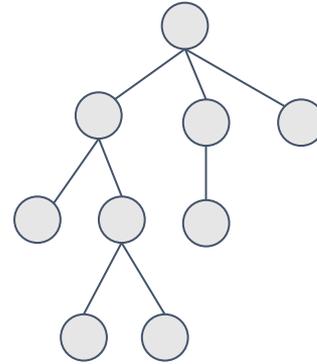
## 2. Des données structurées



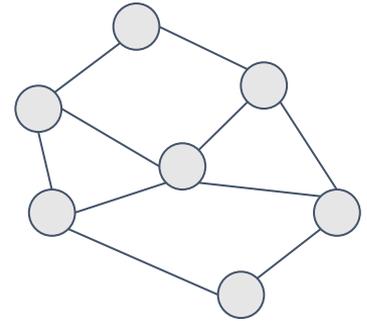
structure liste



structure tableau



arborescence



structure graphique

# Standards techniques de open data

## 2. Des données structurées

Un « format » est un accord qui indique comment une série de données doit être lue, et donc souvent quel type de programme doit ou peut être utilisé. Par exemple, vous connaissez probablement le format « JPG » pour les images. Pour les données structurées, le format le plus courant est simplement « texte ». Pour convertir des structures de données en texte, une convention supplémentaire est nécessaire sur la structure du texte, nous appelons cela des sérialisations. Les sérialisations courantes sont :

- CSV ("valeurs séparées par des virgules") pour les listes
- XML ou JSON pour les arborescences
- RDF pour les graphes (voir plus loin sous « données liées »)

# Normes de open data

## 1. Formats ouvertes

Afin de pouvoir interpréter les données, il est donc important que des accords aient été conclus sur la structure des données. Nous appelons ces accords formats et standards. On parle **de formats ou des normes ouvertes** lorsque la norme est documentée publiquement, de sorte que n'importe qui peut développer une application pour lire dans un certain type de données. Exemples des formats:

FORMAT PROPRIÉTAIRE	FORMAT OUVERTE
DOCX	ODT (Open Document Text)
XLSX	CSV (Comma Separated Values)
RTF	TXT (Plain Text)
PPTX	ODP (Open Document Presentation)
PDF	HTML (Hyper Text Markup Language)

# Normes de open data

## 2. Normes ouvertes

Un format ne détermine que la structure d'un jeu de données, tandis qu'une "norme" contient souvent plus d'informations sur le contenu et le type de données. Par exemple, les progiciels sectoriels (par exemple pour les données géographiques) sont souvent plus complexes. De nouveau, une distinction peut être faite entre les normes ouvertes et fermées.

En général, une norme est gérée par un "organisme de normalisation" tel que:

- l'Organisation internationale de normalisation (ISO)
- Commission électrotechnique internationale (IEC)
- Union internationale des télécommunications (ITU)

# Normes de open data

## 2. Normes ouvertes

Par contre, tout le monde ne peut pas faire partie de telles organisations. Pour parler d'un standard ouvert, cela doit être le cas. Des exemples d'organisations de normes ouvertes sont:

- Internet Engineering Task Force (IETF)
- World Wide Web Consortium (W3C)
- Open Geospatial Consortium (OGC)

Cette organismes étaient responsables pour quelques de nos plus importantes normes aujourd'hui, comme le Internet Protocol (IP), le Hypertext Transfer Protocol (HTTP) et le Web Feature Service (WFS)

Il n'est donc pas surprenant que de nombreux programmes de données ouvertes soient basés sur ces normes.

# Normes de open data

## 3. Normes ouvertes de protocoles d'échange

Notamment, ces trois normes IP, HTTP et WFS sont surtout utilisées pour standardiser **la communication** entre des systèmes. Aujourd'hui, cela est de plus en plus important, car l'évolution vers un **Internet des objets (IoT)** introduit de nouvelles façons pour les appareils de communiquer. Voici quelques normes pertinentes supplémentaires :

- **MQTT** (Message Queuing Telemetry Transport) : protocole de messagerie léger conçu pour une communication efficace entre les appareils IoT et les serveurs, idéal pour les environnements contraints.
- **LoRaWAN** (Long Range Wide Area Network) : protocole de réseau étendu à faible consommation d'énergie conçu pour la communication à longue portée dans les applications IoT, en particulier pour les scénarios à faible coût et à faible bande passante.
- **NGSI-LD** est un modèle d'information et une API pour publier, interroger et s'abonner à des informations contextuelles. Il est destiné à faciliter l'échange ouvert et le partage d'informations structurées entre les différentes parties prenantes et est géré par la fondation FIWARE

# Normes de open data

## 4. Normes ouvertes de données

Outre la standardisation des protocoles de communication, il est tout aussi important de standardiser l'information elle-même. Ces normes sont souvent appelées "modèles de données". Celles sont spécifiques à certains domaines, par exemple :

Mobilité	GTFS (General Transit Feed Specification) : format pour les horaires des transports en commun et les informations géographiques (développé par Google)
	OpenStreetMap : un projet de cartographie collaborative qui fournit des données géospatiales ouvertes pour divers modes de transport.
	SAE J2735 : Une norme pour les communications de véhicule à véhicule (V2V) et de véhicule à infrastructure (V2I) dans les systèmes de transport intelligents (ITS).
	DATEX II, développée par le Comité européen de normalisation (CEN) et est largement utilisée en Europe pour le partage d'informations de trafic en temps réel et statiques. Il prend en charge différents types de données liées à la gestion du trafic

# Normes de open data

## 4. Normes ouvertes de données

Environnement	OGC Sensor Web Enablement (SWE): A suite of standards for exchanging sensor data and observations in environmental monitoring.
	OGC Web Coverage Service (WCS): A standard for accessing and exchanging multi-dimensional raster data, such as satellite imagery and climate models.
	OGC WaterML: A standard for encoding hydrological and water-related data, including water observations, forecasts, and quality information.
	Darwin Core: A biodiversity data standard that provides a set of terms for describing and sharing species occurrence and other biodiversity-related data.

# Normes de open data

## 4. Normes ouvertes de données

Urbanisme	CityGML: An open data model and XML-based format for representing 3D urban models, including buildings, roads, and vegetation.
	OGC Planning Land Administration Domain Working Group (PLA DWG): A group working on standards related to land administration and urban planning.
	INSPIRE (Infrastructure for Spatial Information in the European Community): A framework of standards for sharing geospatial data across Europe, including data relevant to urban planning.

# Normes de open data

## 4. Normes ouvertes de données

Veillez noter qu'il ne s'agit pas d'une liste exhaustive et qu'il peut y avoir d'autres normes ouvertes dans chaque domaine de données. De plus, il peut y avoir des chevauchements entre ces normes, et il est essentiel d'examiner attentivement celles qui s'appliquent le mieux à votre cas d'utilisation.

Un bon point de référence pour rechercher les normes applicables et les modèles de données est:

Smart Data Models.

# Référentiels de open data

## 1. Master Data Management ou MDM

Certains jeux de données ou bases de données sont très importants car ils sont référencés par de nombreux autres systèmes et sources de données. Il est donc important de s'assurer que ces données ne sont pas dupliquées et bien gérées.

Nous appelons ces sources de données Master Data ou des référentiels. Il peut s'agir par exemple de concepts souvent utilisés tels que "Personne", "Service", "Emplacement", etc.

Le Master Data Management est une partie intégrale du "Data Management".

# Référentiels de open data

## 1. Master Data Management

Les principes de la gestion des données sont très clairement définis dans le "Corps de connaissances en gestion des données" (DMBoK). Ils sont:

1. Qualité des données
2. Gouvernance des données
3. Intégration des données
4. Sécurité des données
- 5. Gestion des données de référence (MDM)**
6. Entreposage de données (data warehousing)
7. Opérations de base de données
8. Cycle de vie des données

# Référentiels de open data

## 2. Data Management

Les principes de la gestion des données sont très clairement définis dans le "Corps de connaissances en gestion des données" (DMBoK). Ils sont:

1. Qualité des données
2. Gouvernance des données
3. Intégration des données
4. Sécurité des données
5. **Gestion des données de référence (MDM)**
6. Entreposage de données (data warehousing)
7. Opérations de base de données
8. Cycle de vie des données



# Référentiels de open data

## 3. Master Data management dans un contexte gouvernemental

Le Master Data Management (MDM) dans un contexte gouvernemental peut présenter plusieurs défis spécifiques en raison de la complexité et de la sensibilité des données impliquées. Voici quelques-uns des défis courants liés au MDM dans un contexte gouvernemental :

- Hétérogénéité des sources de données
- Volume et diversité des données
- **Collaboration interorganisationnelle**
  - => **créer des référentiels au niveau (supra)national et regionales**
- Sécurité et confidentialité des données
- Cadre réglementaire et conformité
- Évolution des politiques et des exigences

# Référentiels de open data

## 3. Master Data management dans un contexte gouvernemental, exemple Européenne

En Europe, l'agence en charge de la création des registres de données de référence est l'office des publications. À l'heure actuelle, ils ont 11 listes d'autorités nommées certifiées.

- Corporate bodies Name Authority List
- Countries Name Authority List
- Currencies Name Authority List
- File types Name Authority List
- Languages Name Authority List
- Multilingual Name Authority List
- Places Name Authority List
- Interinstitutional procedures Name Authority List
- Resource types Name Authority List
- Roles Name Authority List
- Treaties Name Authority List

# Merci !



Contact :  
[au-eu@d4dhub.eu](mailto:au-eu@d4dhub.eu)

Le D4D Hub UA-UE est financé par l'Union européenne et :



Le projet est mis en œuvre conjointement par :

