

L'ouverture des données publiques culturelles en pratique

Ministère de la Culture

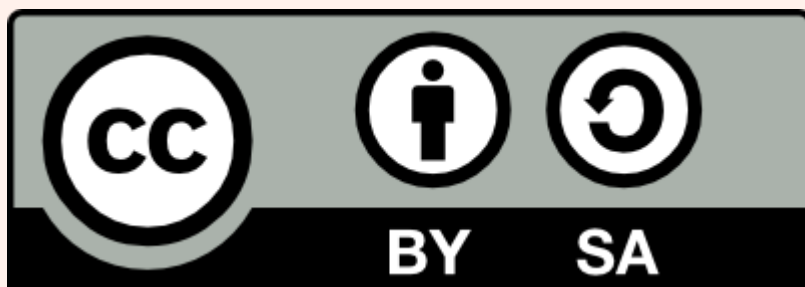
Dataactivist - Anne-Laure Donzel

Avril 2023

Ces slides en ligne : <http://dataactivist.coop/ministere-culture>

Sources : <https://github.com/dataactivist/ministere-culture>

Les productions de Dataactivist sont librement réutilisables selon les termes de la licence [Creative Commons 4.0 BY-SA](#).



Programme de la formation

Open Data : l'ouverture des données publiques culturelles **en pratique**

- Concevoir une stratégie open data
- Identifier des données à ouvrir
- Les formats de données
- Mettre en qualité des données
- Publier des données culturelles et produire des métadonnées de qualité
- Animer un projet d'open data et inciter à la réutilisation de données
- Gérer la relation avec les (ré)utilisateurs

Aujourd'hui, vous allez apprendre à :

- Concevoir une stratégie Open Data
- Comprendre la qualité nécessaire pour publier des données
- Préparer des données et s'initier au travail nécessaire à la publication
- Évaluer des opportunités d'ouverture de données
- Animer et valoriser la réutilisation

1. Quelques rappels pour commencer : L'essentiel de l'open data en 15 minutes

Avant d'approfondir la pratique, voici un **bref retour sur les bases**

- Comment définissez-vous l'open data ?
- Quels sont les enjeux de l'open data ?
- Quelles données peuvent être ouvertes dans le domaine de la culture ?

Une ou plusieurs définitions ?

- **L'open data** se définit comme : "des données devant être accessibles gratuitement, dans un format utilisable et modifiable pour servir tout objectif." D'après la définition du **savoir ouvert**, **l'open definition**
- **L'open content** va concerner l'ouverture des contenus culturels en eux-mêmes

Par exemple : l'open data va consister en l'ouverture des inventaires d'oeuvres et l'open content en l'ouverture des images des oeuvres. Dans les deux cas :

Ouverture = mise à disposition + possibilité de réutilisation

Les enjeux

- La transparence de la vie publique et la participation citoyenne à l'action publique
- L'efficacité de l'action publique : amélioration de la qualité des services publics
- Le développement de l'attractivité des territoires
- Le soutien à l'innovation économique et sociale
- Une opportunité en matière de démocratisation culturelle et de transmission des savoirs

Culture : principales catégories de données **concernées par l'ouverture**

- Les **données statistiques** des institutions culturelles : fréquentation, entrées, liste des équipements
- Les **bases de données des collections** : métadonnées et plus largement les données associées aux œuvres (localisation, description, auteur, œuvres associées, prix, propriétaire)
- Les **contenus numérisés** (ou nativement numériques): images des tableaux, archives, photos, sons, vidéos
- Les **listes de structures** : musées, monuments, cinéma, etc.
- Les **référentiels ou nomenclatures** : vocabulaires contrôlés, thesaurii etc.
- Les données de type **agenda** : événements, programmations
- Les **données financières** : financement, ressources, budgets annuels, subventions

Important : pour la mise à disposition des œuvres (numérisées ou nativement numériques), il faut tenir compte du droit de propriété intellectuelle avant l'ouverture.

**Derrière chaque fichier se cache
une découverte**

1 - Concevoir une stratégie open data

Réussir sa stratégie d'ouverture de données

- Au-delà de l'obligation légale : **poser les objectifs** poursuivis par la stratégie
- Se donner les moyens de connaître la **demande de données** pour y répondre
- Travailler à la **qualité des données**, garante d'usages réels
- Mener le projet d'ouverture comme un **projet de modernisation**

Surmonter les **difficultés**

- Problèmes **organisationnels** : ouvrir des données prend du temps et vient en plus de nombreuses tâches car cela n'a pas été pensé comme une politique publique
- Problèmes **techniques** : vouloir ouvrir ses données révèle souvent des problèmes de qualité ou d'interopérabilité
- Problèmes **juridiques** : les données à ouvrir sont-elles protégées par des droits spécifiques ?

Les politiques open data "de l'offre"

En théorie

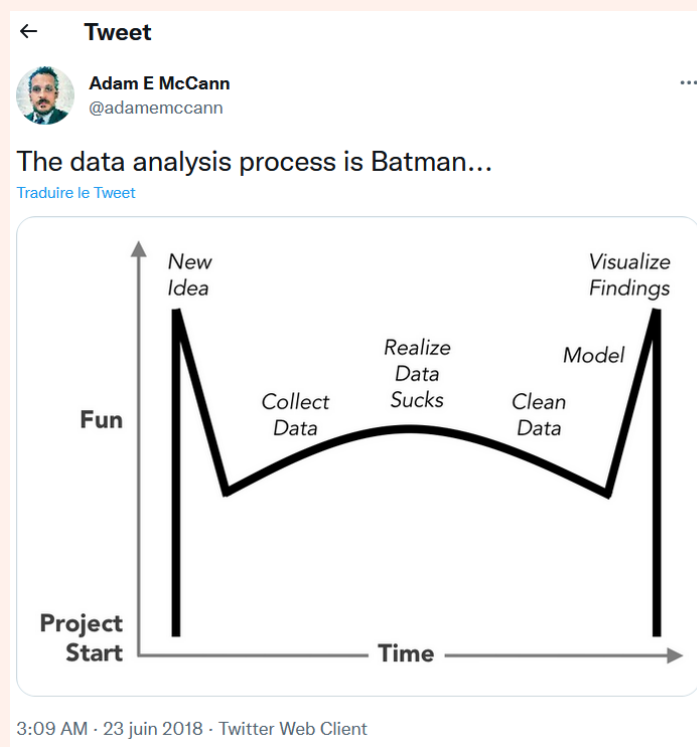
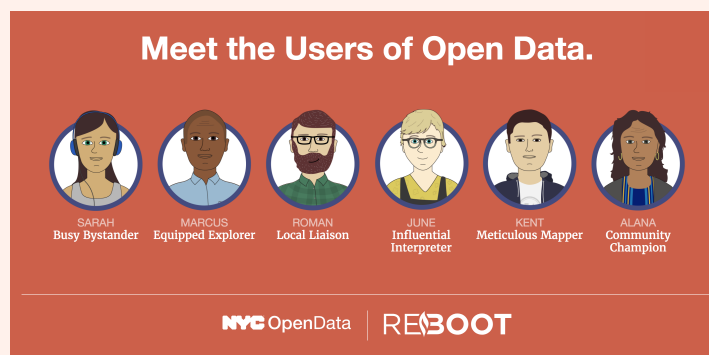
Toutes les données sont
ouvertes sous leur
forme brute dès leur
production

En pratique

L'administration choisit
quelles données ouvrir,
comment les ouvrir,
quand les ouvrir et ce
qu'elles contiennent

Les conséquences pour les réutilisateurs :

Un parcours fait de frictions et de déceptions



2 - Les étapes de l'ouverture des données

Comment ouvrir ses données : **les différentes étapes** pratiques à suivre

- **Étape 1** : Sélectionner les données
- **Étape 2** : Préparer les données
- **Étape 3** : Diffuser les données
- **Étape 4** : Animer et valoriser les réutilisations

Etape 1 : Sélectionner les données

**Quelles données,
ouvrables ou non, avez-
vous dans vos services ?**

Étape 1 : Sélectionner les données

1.1 Effectuer un état des lieux permettant de faire **l'inventaire des données existantes**. Cette opération permet également d'identifier d'éventuelles contraintes à l'ouverture et ainsi d'envisager les actions concrètes à mener.

Une cartographie de données a pour objectif de recenser les données issues d'un ou de plusieurs système d'information afin d'améliorer la connaissance d'un patrimoine de données potentiel.

MODÈLE DE TABLEAU
POUR UNE CARTOGRAPHIE
DES DONNÉES

Direction	Service	Nom de la base	Responsable de la base	Supérieur hiérarchique	Descriptif de la base	Statut actuel (ouvert ou non)	Impératif d'ouverture (du point de vue de la loi)	Perspectives d'ouverture et calendrier

Exercice de cartographie de données

Étape 1 : Sélectionner les données à ouvrir en priorité

1.2 Cette sélection se fait en fonction des objectifs et des moyens de votre structure.

Il y a plusieurs possibilités : se conformer à une stratégie formalisée, suivre les demandes des communautés de réutilisateurs...

2 **Évaluer les données**

Nom du jeu de données :
Évaluez le jeu de données par rapport à son utilisation dans le cadre de votre stratégie.

☒ **coche verte** ☐ **coche rouge** **X**

Importance du jeu de données dans la stratégie: *notez le rang de 1 à 3*

Bilan de l'évaluation générale : ★★★★★

Travail à accomplir pour utiliser les données : ☠☠☠☠☠

Adéquation des données à la stratégie
Les données correspondent-elles vraiment à notre besoin ?
Les données répondent aux objectifs de votre stratégie :
☐ les données sont brutes (ce ne sont pas des données calculées) ☐
☐ la fréquence des données est suffisante ☐
☐ la spatialité des données est suffisante ☐

Fiabilité des données
Peut-on vraiment s'appuyer sur ces données ?
La méthode de collecte des données apporte des gages de qualité :
☐ contrôle qualité ☐
☐ échantillon suffisant ☐
☐ échantillon diversifié ☐
☐ absence de données déclaratives ☐

Exploitabilité des données
Les données sont-elles vraiment accessibles et exploitables ?
Il est simple d'accéder aux données parce que :
☐ nous sommes producteurs des données ☐
☐ un requête d'extraction est possible depuis nos outils SI ☐
Le format des données simplifie leur traitement

Ouverture des données
Et si on ouvrait ces données ?
Y a-t-il des freins à l'ouverture :
☐ le jeu de données ne contient pas de données à caractère personnel ou il est déjà anonymisé ☐
☐ pas de sensibilité politique notable ☐
Les externalités prévisibles d'une

Exercice d'évaluation des données et de priorisation

Etape 1 : les **outils** de la sélection de données

- Le modèle de cartographie du Ministère - P.24 du **Guide pratique du Ministère**
- Orientations stratégiques de l'Etablissement ou prioriser ses ouvertures suivant les demandes et les opportunités
- Grille d'analyse

Etape 2 : Préparer les données

**C'est quoi la qualité des
données selon vous ?**

Etape 2 : l'importance de la **qualité des données**

Une remarque récurrente :

Les données publiques sont souvent incomplètes, non actualisées, de mauvaise qualité et fragmentaires.

Dans la plupart des cas, les catalogues ou portails de données ouvertes sont alimentés manuellement, conséquence d'une approche informelle de la gestion des données.

De plus, les procédures, les calendriers et les responsabilités des institutions publiques chargées de leur gestion manquent souvent de clarté. Ainsi, de manière générale, la gestion et la publication des données ouvertes sont fragiles et sujettes à des erreurs multiples.

Source : **Rapport mondial du 4e Open Data Barometer**

Etape 2 : Qualité et qualité

Qualité interne des données :

- Harmonisation
- Complétude
- Niveau de granularité
- Types
- Structure interne

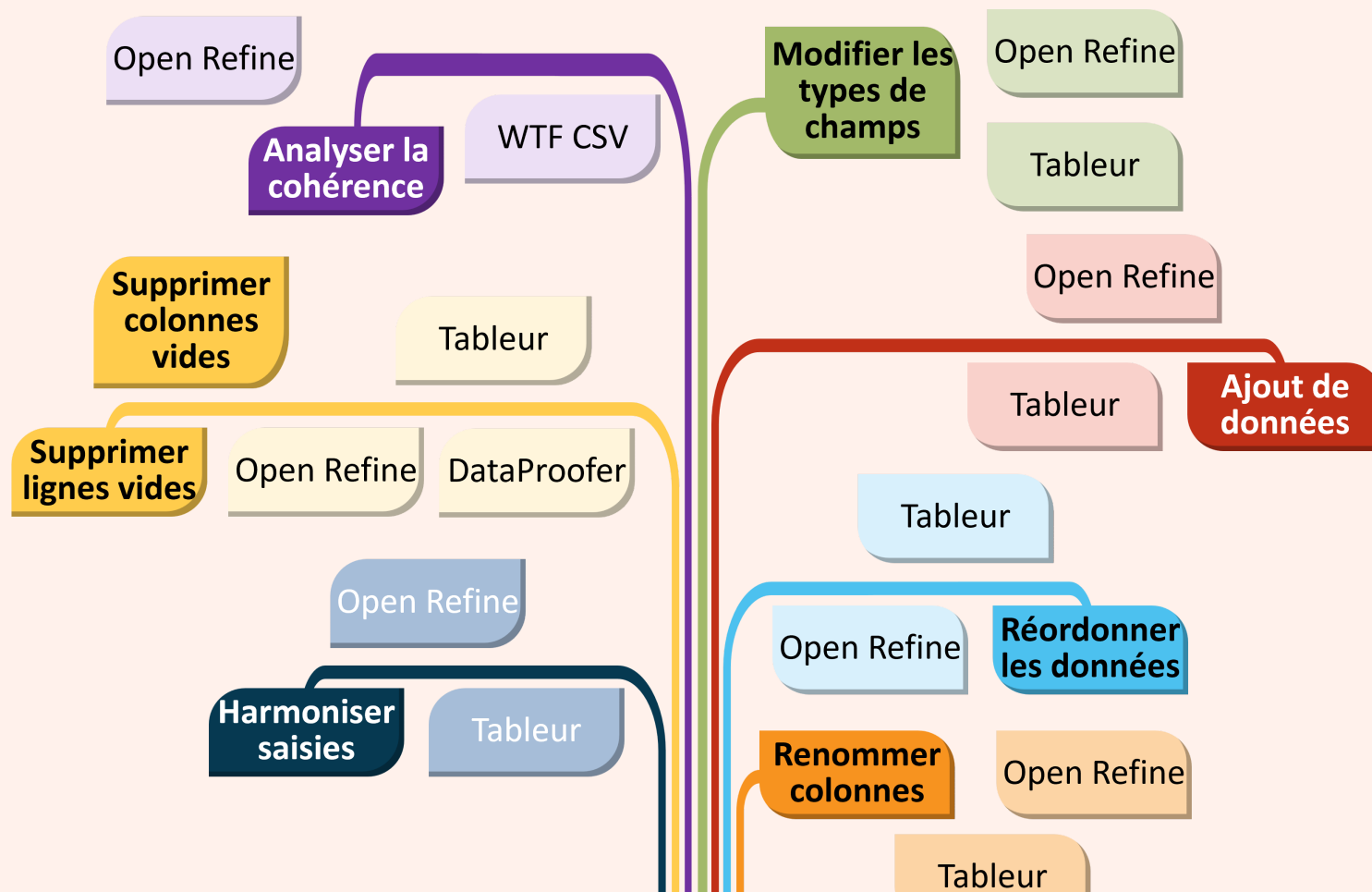
Nous y reviendrons

Qualité externe des données

- Format
- Standardisation
- Description

Exercice : analyser la qualité d'un jeu de données

Etape 2 : Les critères de la **mise en qualité**



OpenRefine : le meilleur ami de la communauté GLAM

OpenRefine est un outil gratuit de traitement de données en masse

Il se télécharge sur l'ordinateur et permet de faire des opérations de traitement de données sur de gros fichiers, qu'Excel ne peut pas traiter.

The screenshot displays the OpenRefine web interface. At the top, the Google Refine logo is visible, along with the text 'movies' and a 'Permalink' button. On the right, there are buttons for 'Open...', 'Export', and 'Help'. Below the header, a 'Facet / Filter' tab is active, showing 'Undo / Redo 24'. A sidebar on the left contains a section titled 'Using facets and filters' with instructions and a link to 'Watch these screencasts'. The main area shows a table with 44 rows. The table has columns for movie titles, directors, release years, and ratings. The first few rows are: 2. The Terminator (James Cameron, 1984, 60029358, 4), 3. Aliens (James Cameron, 1986, 60029358, 4), 4. How to Train Your Dragon (Chris Sanders, 2010-03-18, 70109893, 4.5), 5. (Dean DeBlois), 6. Ocean's Eleven (Steven Soderbergh, 2001-12-07, 60021783, 3), 7. A Beautiful Mind (Ron Howard, 2001-12-21, 60021793, 3.5), 8. Apollo 13 (Ron Howard, 1995-06-30, 262866, 3), 9. Away From Her (Sarah Polley, 2006, 70055883, 3.5), 10. Amadeus (Miloš Forman, 1984, 247351, 3), 11. Babel (Alejandro González Iñárritu, 2006-10-27, 70045866, 3), 12. Batman Begins (Christopher Nolan, 2005-06-15, 70021642, 4).

Row	Movie Title	Director	Year	Score	Rating
2.	The Terminator	James Cameron	1984	60029358	4
3.	Aliens	James Cameron	1986	60029358	4
4.	How to Train Your Dragon	Chris Sanders	2010-03-18	70109893	4.5
5.		Dean DeBlois			
6.	Ocean's Eleven	Steven Soderbergh	2001-12-07	60021783	3
7.	A Beautiful Mind	Ron Howard	2001-12-21	60021793	3.5
8.	Apollo 13	Ron Howard	1995-06-30	262866	3
9.	Away From Her	Sarah Polley	2006	70055883	3.5
10.	Amadeus	Miloš Forman	1984	247351	3
11.	Babel	Alejandro González Iñárritu	2006-10-27	70045866	3
12.	Batman Begins	Christopher Nolan	2005-06-15	70021642	4

Open refine

Démo de l'outil à partir des données des documents d'urbanisme

Etape 2 : les **outils** de la qualité des données

- L'outil "couteau suisse" : Open Refine, à défaut un tableur
- Les outils "coup de pousse" : DataProofer, WTF CSV
- Une checklist sur la qualité des données

Étape 3 : Diffuser les données

Étape 3 : **Diffuser** les données

3.1 Choisir le bon format

3.2 Bien documenter ses données pour rendre ses données découvrables

3.3 Choisir la bonne licence

3.4 Utiliser la bonne plateforme

Étape 3 : Diffuser les données

Etape 3.1 : retour sur la qualité externe des données

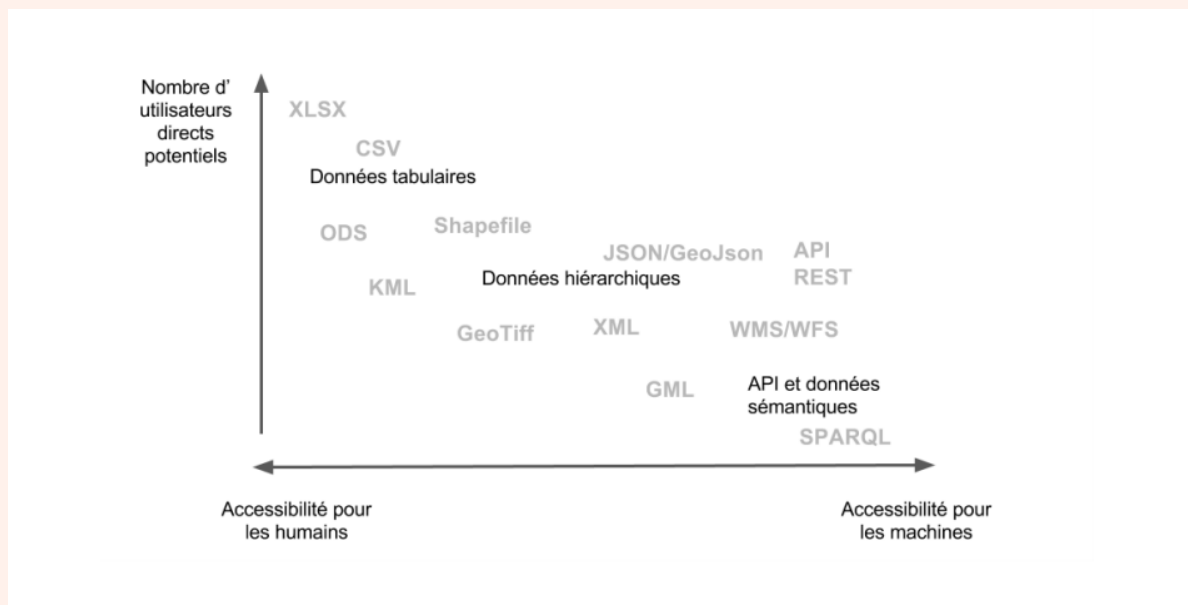
Les données doivent être téléchargeables dans un « standard ouvert », c'est à dire dans un format ouvert aisément réutilisable et exploitable par un système de traitement automatisé et pérenne. Les formats ouverts favorisent l'interopérabilité : les données pourront être réutilisées indépendamment du logiciel qui a été utilisé pour les créer.

Voici quelques exemples de formats « ouverts » conseillés pour la diffusion des jeux de données en ligne :

- **Texte brut** : ASCII (.txt)
- **Texte formaté** : TeX (.tex), OpenDocument Text (.odt), Hypertext Markup Language (.htm ou .html), XHTML (.xhtml), Feuilles de style en cascade (.css)
- **Images** : png (.png), jpeg (.jpg ou .jpeg)
- **Données brutes** : CSV (.csv) / sans extension, JSON (.json), XML (.xml)
- **Données géographiques** : KML (.kml), SHP (.shp)

Étape 3 : Diffuser les données

Etape 3.1 : Accessibilité des formats des données pour les humains et pour les machines



Source : Françoise Paquienséguy, Valentyna Dymytrova. Livrable 1.2 Analyse de portails métropolitains de données ouvertes à l'échelle internationale. Rapport de recherche 1.2, Equipe d'accueil lyonnaise en Sciences de l'information et de la communication. 2017. [hal-01449348](https://hal.archives-ouvertes.fr/hal-01449348)

Étape 3 : **Diffuser** les données

Attention : réalisez un fichier .csv de qualité

C'est quoi un bon fichier ?

- Il ne possède pas de cellules fusionnées
- Il ne possède pas de mise en page (couleur, gras...)
- Il ne contient pas de formules, macro etc...
- Il ne contient pas de totaux / sous-totaux

Étape 3 : Diffuser les données

3.2 Documenter les données

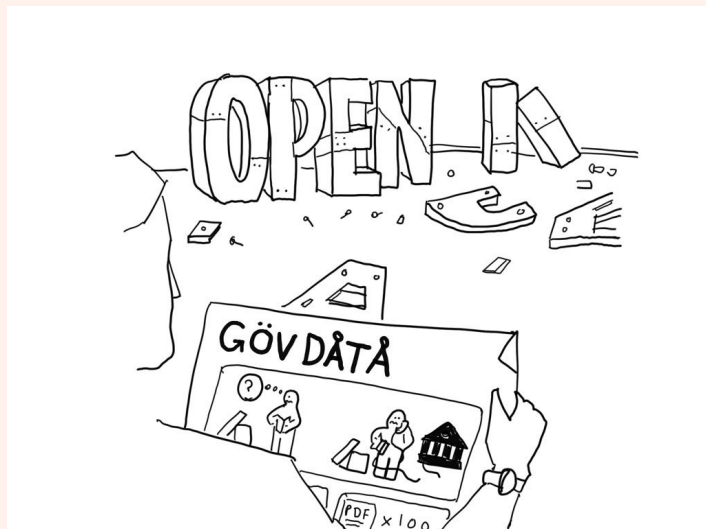
Des données mal décrites vont être introuvables et donc, elles ne seront pas utilisées.



The screenshot shows the data.gouv.fr website. At the top left is the French Republic logo with the text "LIBERTÉ ÉGALITÉ FRATERNITÉ". Next to it is the "data.gouv.fr" logo. A navigation bar contains links: "Données" (underlined), "Réutilisations", "Organisations", "Actualités", "À propos", and "Nous contacter". Below the navigation bar is a breadcrumb trail: "Accueil > Jeux de données > C2020-SBEP-075". The main title "C2020-SBEP-075" is displayed in large bold letters. Below it, the text "Mis à jour le 18 mai 2020 — License Not Specified" is shown. At the bottom, there are three tabs: "Producteur", "Métadonnées" (which is selected and underlined), and "Actions".

Étape 3 : Diffuser les données

3.2 La découvrabilité : un enjeu crucial



"La découvrabilité des données est un défi majeur. Nous avons des portails et des registres de données, mais les organismes gouvernementaux qui relèvent d'un seul gouvernement national publient toujours les données de différentes façons et à différents endroits (...).

La découvrabilité des données est une condition préalable pour que les données ouvertes réalisent leur potentiel. Or, actuellement la plupart des données sont très difficiles à trouver."

Source : Classement mondial de l'Open Data

Étape 3 : Diffuser les données

3.2 Rendre les données découvrables

"Les métadonnées sont de l'information structurée qui décrit, explique, localise ou facilite autrement l'obtention, l'utilisation ou la gestion d'une ressource d'information. Les métadonnées sont souvent appelées données des données ou information sur l'information."

Source : National Information Standards Organization

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Les Métadonnées descriptives décrivent une ressource à des fins de découverte et d'identification.

Exercice de description

Étape 3 : Diffuser les données

3.4 Utiliser la bonne plateforme de publication

Le portail data.culture.gouv.fr

Le portail des données ouvertes du ministère de la Culture met à disposition les données de l'administration centrale, d'une DRAC ou d'un SCN, référence les données des établissements publics et moissonne automatiquement ces données sur data.gouv.fr, pour pouvoir les exposer sur les deux plateformes à la fois.

Pour toute information sur la mise à disposition de nouveaux jeux de données, contacter le département de l'innovation numérique (SG/SCPCI) via l'onglet « Contact » de la plateforme data.culture.gouv.fr.

Étape 3 : Diffuser les données

3.4 Utiliser la bonne plateforme de publication

Le portail data.gouv.fr

Pour les établissements publics, la mise à disposition des données est à effectuer sur le site data.gouv.fr.

Afin d'améliorer la visibilité des données culturelles, le ministère référence également les données des établissements sur la plateforme ministérielle data.culture.gouv.fr.

Cela implique la création d'une page dédiée à l'établissement, puis la publication des jeux de données conformément [aux modalités précisées par Etalab](#)

Le ministère mettra en place les mécanismes nécessaires pour que toutes les données publiées de cette façon soient ensuite automatiquement visibles sur data.culture.gouv.fr.

Etape 3 : les **outils** de la diffusion de données

- Le **guide open data** du Ministère de la Culture offre de nombreux éléments sur la diffusion
- Sur les conditions juridiques de **réutilisation** et sur les **droits détenus par des tiers**
- Les plateformes : **data culture** et **data gouv** et notamment les **pages de tutoriel**

Animer et valoriser les réutilisations

Étape 4 : **Animer et valoriser les réutilisations**

En interne à votre structure

- Utiliser des données ouvertes, notamment les vôtres

En externe à votre structure

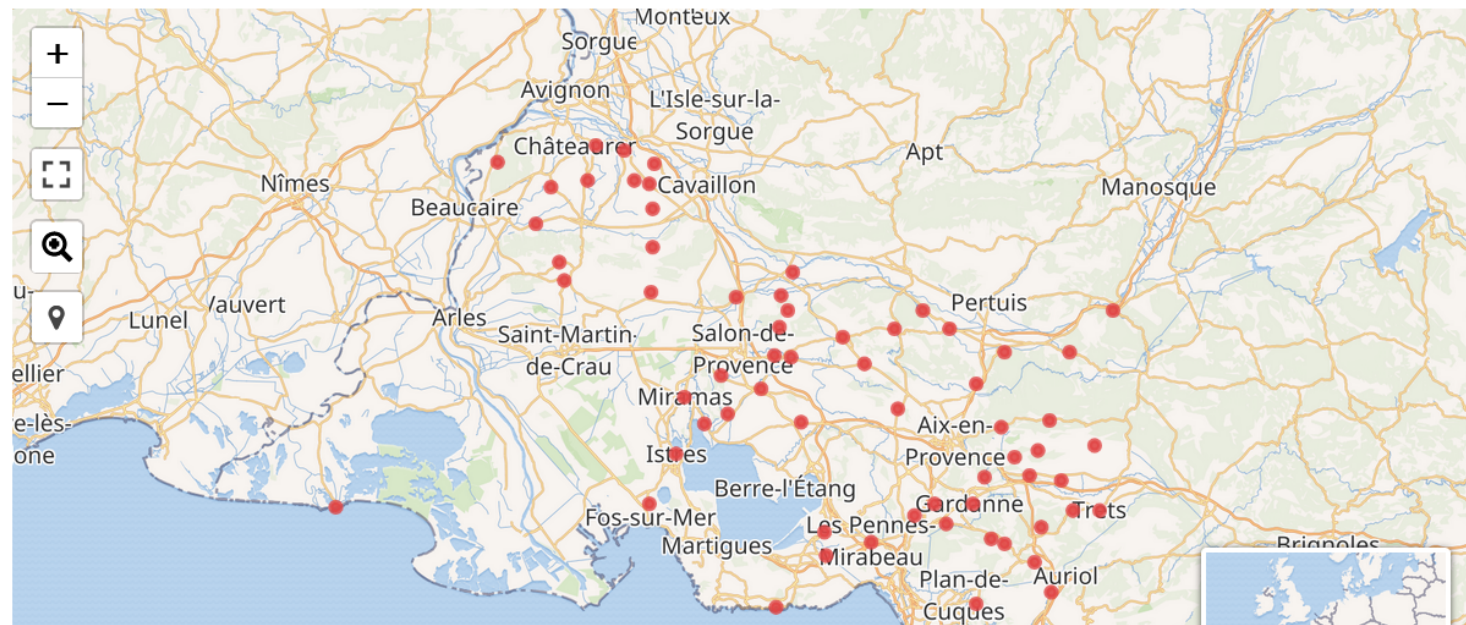
- Animer une communauté de réutilisateurs
- Nouer des partenariats
- Favoriser les réutilisations
- Valoriser les réutilisations

Étape 4 : Animer et valoriser les réutilisations

Intégrer la réutilisation de données à votre établissement

Carte des AD13 qui utilise les données de Wikidata pour présenter les **fonds communaux conservés**

CARTE DES COMMUNES AYANT DÉPOSÉ LEUR FONDS



Exercice : Créer votre première datavisualisation avec les données de data.culture.gouv.fr

À partir de l'onglet « Carte » de data.culture.gouv.fr, nous allons apprendre à créer une carte avec plusieurs jeux de données, ou à ajouter plusieurs calques et filtres.

Susciter les usages externes

- Editorialiser les données brutes en proposant des représentations visuelles facilitant leur lecture (graphiques, cartographies, etc.), par exemple la carte des institutions culturelles partenaires du site Histoire des Arts, les tendances clés sur la fréquentation des musées.
- Exemple de la Région Centre sur sa plateforme open data

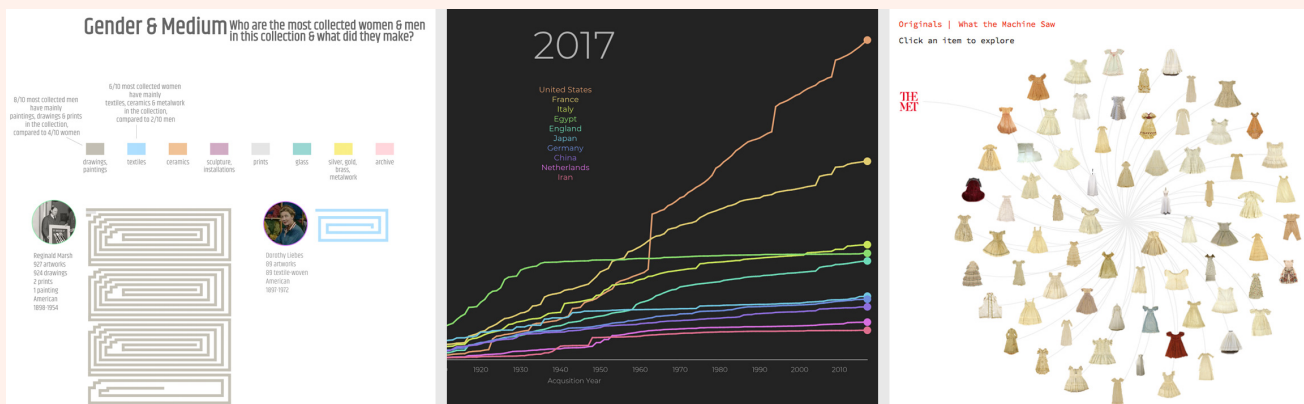
Organiser des événements

- Communiquer pour promouvoir et faire connaître les données ouvertes auprès des communautés des utilisateurs et réutilisateurs potentiels lors d'événements (rencontre, hackathon, concours, appel à projets, etc.).
- Hackathon des Archives Nationales
- Hackathon de la BNF, 3 éditions depuis 2016. Un exemple de production

Susciter les réutilisations avec **un programme de partenariats**

Exemple : **impliquer une école de graphisme à réaliser des visualisations**

- Analyse de la **représentation du genre** dans la collection d'art moderne et contemporain du Met
- Une **analyse de la forme des œuvres d'art au Met's Costume Institute**, utilisant l'apprentissage automatique pour découvrir de nouvelles relations entre les objets au-delà des métadonnées disponibles.
- Une évaluation de l'histoire de la collecte au Met et de l'évolution des pratiques de collecte au fil de l'histoire du Met



Château de Versailles et Wikimedia : un partenariat de valorisation de données

- 1er partenariat entre Wikimedia France et un musée Français
- Wikipedia est la 2e source d'information des internautes sur Versailles après le site internet du château
- Versailles s'engage dans la voie d'enrichissement de communs numérique, à travers les technologies libres et participatives

En février 2011, dans le cadre du partenariat, un wikipédien est venu en résidence 6 mois à Versailles. Son but était :

- D'améliorer les contenus de Wikipedia autour de Versailles,
- D'animer la communauté autour du sujet Versailles.
- D'expliquer, de rencontrer les équipes de Versailles, de trouver des opportunités de collaboration entre conservateurs et wikipédiens.

Un bilan positif du partenariat

- Un enrichissement spectaculaire des contenus : les articles concernant le château de Versailles sont passés de 70 à 380
- Idem pour les photos : plus de 2000 photos versées sur WikiCommons
- La qualité des articles et des photos a été améliorée
- Retombées médiatiques importantes

Exemple : le [plafond de la galerie des batailles](#) a remporté le concours Wiki Loves Monuments en 2011 (concours photo mondial visant à valoriser les monuments historiques).

Au-delà, cela a permis de changer les mentalités en interne : Wikipedia est maintenant regardé avec bienveillance par les équipes scientifiques du château.

Animer et acculturer une communauté de réutilisateurs

- Communiquer sur la démarche, mobiliser et interagir avec les réutilisateurs (associations, entreprises, acteurs de l'enseignement et de la recherche, médias et journalistes, citoyens) pour s'assurer que les données ouvertes correspondent à leurs attentes et soient effectivement réutilisées et valorisées.
- Améliorer les données produites, compléter ou produire de nouvelles données en s'appuyant sur la contribution et l'engagement citoyen (crowdsourcing).

Etape 4 : les outils de la valorisation de données

Des datavisualisations simples et rapides:

- Directement dans data culture
- Avec des outils spécifiques :
Datawrapper, Rawgraphs
- Contenus interactifs storymap

Des cartes simples et rapides :

- Umap
- Carto
- Khartis

Et des outils plus experts : Rstudio, Wikidata Query, QGis...

Questions pour **terminer**

- Y a-t-il une offre et une demande pour les données publiques ouvertes au sein de votre service, direction ou établissement ? Si oui, qui fournit quoi à qui ?
- D'après vous, quels sont les obstacles principaux pour la publication de données publiques ouvertes au sein de votre service, direction ou établissement ?

Résumé : les **facteurs clés de succès** d'une stratégie open data

Obtenir un soutien au plus haut niveau

- Pour mettre en oeuvre les moyens et les ressources à déployer

Configurer l'organisation pour l'ouverture

- Organiser la gouvernance, les circuits de validation

Faciliter la découverte et l'utilisation des données

- Faire attention au référencement, à la qualité des métadonnées

Rester à l'écoute et interagir avec les usagers

- Prévoir des boucles de participation

Merci !