

4 mois avec des experts data pour transformer votre projet

Evénement de lancement

Marseille, 26 septembre 2019



Projet cofinancé par le Fonds européen de développement régional





Programme de l'après midi

- Présentation du projet ODEON
- Philosophie du projet et des ateliers
- Pitchs et pitchs inversés
- Présentation de la méthodologie des ateliers
- Ateliers
- Synthèse et annonce de la suite

Présentation du projet ODEON



Résumé du projet

Titre du projet	Open data for European open innovation Données ouvertes pour l'innovation ouverte européenne
Acronyme	ODEON
Programme	Interreg-Med
Durée	Février 2018-Juillet 2020
Chef de file	Région Vénétie (Italie)
9 partenaires de 7 pays: Croatie, Espagne, France, Grèce, Italie, Monténégro, Slovénie	CCI Montenegro, Agence croate pour les PME, l'innovation et l'investissement (HR) Technology Park de Ljubljana (SL) Société grecque des logiciels libres et ouverts (GR) Société aragonaise de gestion agro environnementale Agence des villes et territoires méditerranéens durables (FR) CCI Padoue (IT) Fondation Knowledge Innovation Market Barcelona(ES) Région de Crète (GR)





OBJECTIF

Améliorer la qualité et l'accès aux données ouvertes mises à disposition par les institutions publiques et soutenir la croissance et l'innovation des entreprises/porteurs de projet qui développent de nouveaux services ou activités à partir de ces données.





ACTIVITÉS PRINCIPALES

AU NIVEAU NATIONAL

Mettre en place des «hubs» pour accompagner entreprises/porteurs de projet dans le développement de services innovants à partir des données ouvertes. Domaines:

- économie verte
- économie bleue
- industrie créative

AU NIVEAU INTERNATIONAL

Créer un réseau méditerranéen composé d'entrepreneurs et d'experts pour:

- promouvoir la valeur commerciale et sociale des données
- internationaliser les projets
- mettre en commun les méthodologies et les approches





EN FRANCE

Focus sur l'adéquation entre données produites et besoins des utilisateurs => faciliter la rencontre et le dialogue entre producteurs et utilisateurs

QUOI

1 évènement de lancement

3 ateliers de co-construction

pour faire émerger/identifier des projets innovants dans chaque filière (économie verte, économie bleue, industrie créative)

Un accompagnement personnalisé pour les projets identifiés

QUAND

26 septembre 2019

Octobre-décembre 2019

Novembre-janvier 2019 · - · - · - ·



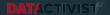




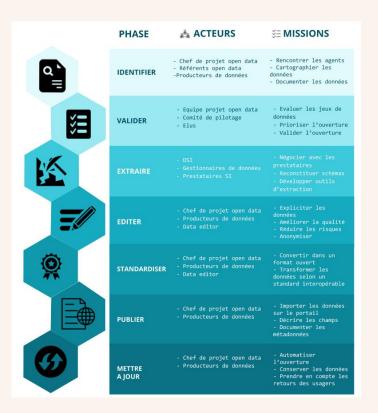
En savoir plus sur le projet: https://odeon.interreg-med.eu/



Philosophie des ateliers



Ouvrir des données, un travail justifié par la demande





Open Government Data Principles

Government data shall be considered open if it is made public in a way that complies with the principles below:

All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.

Data is as collected at the source, with the highest possible level of argularity, not in aggregate or modified forms,

Data is made available as quickly as necessary to preserve the value of the data.

Data is available to the widest range of users for the widest range of

5. Machine processable

Data is reasonably structured to allow automated processing.

6. Non-discriminatory Data is available to anyone, with no requirement of registration.

Data is available in a format over which no entity has exclusive cont

Data is not subject to any copyright, patent, trademark or trade secr





Des politiques d'open data de l'offre

En théorie

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

En pratique

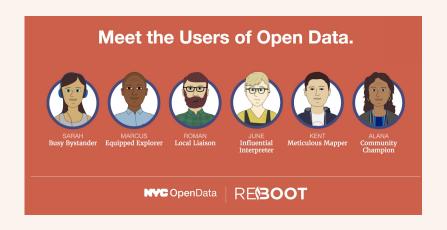
L'administration choisit quelles données ouvrir, comment les ouvrir, quand et ce qu'elles contiennent



Et souvent, des données inutilisées



Et pour les réutilisateurs, un parcours fait de frictions



Des données difficiles à trouver, pas ouvertes partout

Granularité temporelle et spatiale souvent trop faible

Manque de documentation

Des données pas mises à jour

Des portails qui ciblent les développeurs



Le problème récurrent de la qualité



"Les données des gouvernements sont souvent incomplètes, pas 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."

<u>opendatabarometer.org/4thedition/report/?lang=f</u>

<u>r</u>

DATACTIVIST

La question de la découvrabilité

"Data findability is a major challenge. We have data portals and registries, but government agencies under one national government still publish data in different ways and different locations. Moreover, they have different protocols for license and formats(...) Data findability is a prerequisite for open data to fulfill its potential and currently most data is very hard to find."



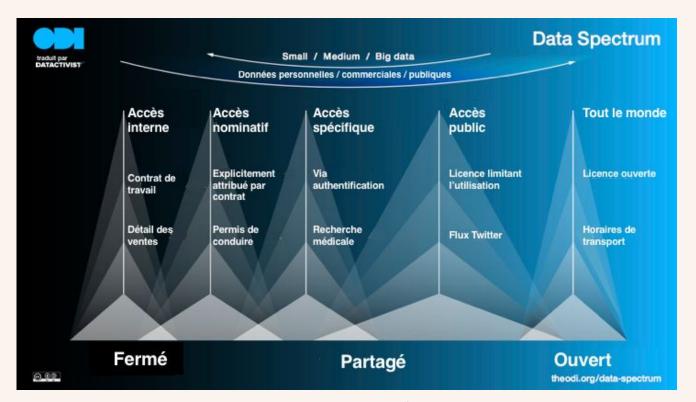
index.okfn.org/insights/

Publish with a purpose, ouvrir pour l'impact

Au cours des dernières années, il est devenu évident que l'ouverture des données exige une approche pratique. Après une première ruée vers « toutes les données brutes maintenant », la communauté de l'open data se demande maintenant « pour quoi faire ? » Les portails de données ouverts sont faciles et bon marché mais, pour être bien fait, l'open data demande du temps, des ressources et une compréhension claire de ce que les données doivent accomplir. Si les gouvernements croient vraiment à l'open data, ils doivent :

- Identifier les défis que les données gouvernementales peuvent aider à résoudre, et travailler en collaboration avec les communautés qui peuvent aider à produire un impact.
- S'engager à publier des jeux de données clés qui sont essentiels pour relever les défis et qui profiteront aux communautés.
- Suivre et adapter: étudiez l'impact de la publication des jeux de données, puis apprenez et adaptez-vous pour maximiser les bénéfices de l'ouverture des données.

Ne pas se limiter aux données publiques ouvertes



Pitchs et pitchs inversés

DATACTIVIST

Méthodologie des ateliers

DATACTIVIST

"La data, c'est compliqué!..."

- "On ne sais pas ce qui existe"
- "La recherche de données c'est vraiment mal fichu"
- "À vrai dire, on ne sait même pas ce qu'on cherche"
- "C'est ch<bip>.... la data : c'est aride, froid, ennuyeux. Les data c'est pour Fred de la compta"
- "À moins de bac +12 en informatique, trop de jeux sont trop compliqués à utiliser : formats ésotériques, fichiers trop volumineux, données mal documentées, etc."
- "Ne parlons même pas de la qualité!"

La littératie de données : ses outils et méthodes

- La littératie de données (data literacy) : les données requièrent un apprentissage au même titre que savoir lire et compter
- 7 ans de travaux novateurs sur la façon d'aborder les données, pour les débutants comme pour les confirmés :
 - plusieurs dizaines de méthodologies et de formats d'apprentissage
 - o un effort mené au 4 coins du globe (Australie, États-Unis, Grande-Bretagne, Allemagne, France, etc.)
- Des outils et méthodes qui rationnalisent la découverte de la révolution des données
- Retrouvez plus d'une dizaine de méthod sur le site de la communauté des infolabs : https://infolabs.io/ressources-cles



Exemple en séance : l'interview de la donnée

- Une méthodologie de découverte d'un jeu de données, sous forme de jeu de rôle : https://infolabs.io/interview-de-la-donnee
- Une personne ou une petite équipe joue le rôle d'un jeu de donnée qui est interviewé par un enquêteur (ou une autre équipe). Ce dernier cherche à connaître le plus de choses d'elle en un minimum de temps.
- Question types : qui es-tu ? comment t'appelles-tu ? qui sont tes parents ? que fais-tu dans la vie ? ...
- Qu'est-ce que cette méthodo change ? => désinhiber les publics nontechniques et amener les "sachants" à des réponses compréhensibles par tous
- Une méthodo idéale pour comprendre et documenter un jeu de données

 DATACTIVIST

50 millions de data, et moi, et moi, et moi?...

Ok, j'ai compris que les données peuvent transformer mon projet, comment on fait ?

Les expéditions de données, un ensemble de méthodologies pour comprendre le potentiel des données pour un projet donné :

- plusieurs exemples à travers le monde
- tous fonctionnent à peu près sur la logique d'un parcours constitué de grands jalons :
 - identification des besoins
 - sourcing
 - prototypage



Dat'accélère > le parcours d'évaluation data

- Une méthodologie française documentée, "open source", déjà utilisée en entreprise (Groupe UP)
- **Objectif** : évaluer le potentiel des données (notamment ouvertes) pour un projet
- Temporalité : une journée, décomposable en deux demi-journées
 - format d'animation modulaire
 - o ne requiert aucune connaissance technique
 - reproductible
- **Résultat** : le "plan de route données", un plan d'action permettant d'intégrer au projet les données utiles selon deux critères : le fort impact sur le projet et l'accessibilité des données.

Parcours d'évaluation du potentiel données1

Partie 1 - Définir le cas d'usage

Α.	Commencer par definir ce que vous voulez proposer en terme de projet voire de fonctionnalités et de services

B. Identifi	B. Identifier les acteurs et les organisations impliquées				

Les parties en	jaune sont à compléter.	Ce	document est	en lecture soule	Pour	débuter le vér	re.

Parcours d'évaluation du potentiel données, méthodologie sous Licence libre Creative Commons CC-8Y, par	

Partie 3 - La collecte de données

A. Rechercher les données

Donnée ⁴	Producteur / source de la donnée

¶ Il est souvent plus facile d'identifier le producteur d'une donnée avant de trouver la donnée elle-même. Par exemple, si vous cherchez "la protabilité de plus dans une

Quelques portails majeurs où trouver des données :

https://data.gouv.fr/
 https://www.europeandataportel.eu

Ajoutez autant de lignes que sous le souhaitez.

Partie 2 - La liste des données idéales ("Wanted Data List")

Lister toutes les données potentiellement utiles pour votre projet (ne vous limitez pas !)

Pour identifier des domées, penser : données internes / données données domées publiques / données domées, penser : données internes / données données publiques / données privilees, données structurées / données non-structurées, passé / présent / futur, données froides / données chaudes, homme / machine, temps /

Données ^a	

Ajoutez autant de lignes que vous le souhaitez.

Partie 4 - Évaluation de la valeur et de la disponibilité

Noter chaque donnée de 1 à 4, selon sa valeur, de la moins critique à la plus critique pour la

- feussite de votre projet :

 1 ce serait bien d'avoir cette donnée, mais elle riest pas cifique pour ce projet.
- cette donnée permet d'augmenter le service rendu à l'utilisateur, mas ce n'est pas critique si elle riest pas disponible,
- 3 cette donnée est essentielle pour la réussite du projet, cela va fortement réduire la véeur pour les utilisateurs si elle manque.
 4 cette donnée est crifique, co serait très difficile de réaliser ce projet sans cette donnée.
- Noter chaque donnée de 1 à 4, selon sa disponibilité : 1 cete donnée ne peut pas être facilement obtenue : elle n'existe pas sur le marché, elle est trop coûteuse pour notre projet, ou il y a des difficultés juridiques ou techniques pour
- collecter cette donnée (par exemple le respect de la vie privée).

 2 il est possible d'acquérir cette donnée, mais cela demande du temps et des
- investis sements supplémentaires, 3 cette donnée est déjà disponible, mais pas à un niveau de défails ou de qualité
- sufficients nour notre projet
- 4 cette donnée est disponible, et son niveau de détaits et sa qualité sont suffisants pour notre orpiet.

Cartographie des données pour votre projet Postanne drague donnée selan ses soores de salaur et de dispositifiés



Placer ici votre carte perceptuelle. Yous pouvez, à votre convenance :

Il a réaliser à la main, la prendre en proto et coller lo la photo :

Ia réaliser à l'aire main, la prendre en proto et coller lo la photo :

Interes haves au suite combandari nel 20 Xicquel XVIII

Elle na parte bos on d'étre partette mas tablis.

Partie 5 : le plan de route de vos données

A. A quoi sert le plan de route data ?

La cartographie des données réalisée à l'étape précédente est un outil pour préparer le plan de route données de votre projet.

On recense 5 options en matière de sourcing des données :

- Sécuriser : s'assurer qu'une donnée cétique pour le projet ne sem pas dégradée ou que sa fournitus ne sera pas interrompue.
- 2. Augmenter : investir dans l'amélioration d'une donnée, sa qualité ou son niveau de
- Collecter: metire en place une collecte pour récupérer des domées utiles pour le projet mais pas ou peu disponibles on peut étudier l'intérêt du crowdsourcing.
- Nouer des partenariats : établir des accords avec des tiers pour accéder à leuis données, ou faire de l'échange de donnée entre partenaires.
 Achete: identifier des sociéées courtiers en données) qui pourraient vendre les

En commençant par le cadran en haut à gauche de la cartographie (forte valeur pour le projet / données disponibles), déterminer la meilleure option possible pour chaque type de

Pour les données les plus importantes, il peut aussi être important d'évaluer leur qualité dans les dimensions suivantes:

L'exactitude: les données sons-elles complètes et fidèles à la réalité qu'elles

- La fraicheur : de quand datent les données ? à quelle fréquence sont elles mises à
- La granularité : les donnée sont-elles disponibles au niveau individuel ou seulement
- de manière agrégée ?

Action (à expliciter en quelques mots)

B. Identifier 10 actions à réaliser maintenant / dans les 3 prochains mois / au cours de la prochaine armée.





DATACTIVIST

Un avant goût des ateliers : la data wish list

- La "data wish list" ou "liste des données idéales" permet de lister, sur une mode créatif, toutes les données qui pourraient être utiles à un projet
- (incluse dans le parcours d'évaluation)
- une manière de prendre pied avec les data
- un exercice simple au cours duquel les participants sont souvent très créatifs
- intéressant que cet exercice soit réalisé à plusieurs, si possible avec un public hétérogène
- Cette méthodologie est open source, publiée et documentée : <u>https://infolabs.io/liste-donnees-ideales</u>



Data wish list > un exemple partiel de résultat

Données²

Le sujet ici : un service de partage de vélo entre particuliers, AirBnBike

Partie 2 - La liste des données idéales ("Wanted Data List")

Lister toutes les données potentiellement utiles pour votre projet (ne vous limitez pas !)
☐ Une astuce pour identifier des données, penser: données internes / données
externes, données publiques / données privées, données structurées / données non
structurées, passé / présent / futur, données froides / données chaudes, homme /
machine, temps / espace,

Géo-localisation des vélos en location
Infrastructures pour le vélo
Identification des vélos
Tarification pour chaque vélo
Taille /type/marque/vitesses/électrique/usage (vtt, urbain)/équipement (gourde, compteur, etc.)/casque/âge vélo/
itinéraire
Dispo (oui/non, heure limite, retour,)
Calories perdues dans chaque trajet
Historique de chaque trajet

Loueurs/locataire (évaluation, âge, coordonnées identités, moyens de paiement,)

Ateliers

DATACTIVIST

Synthèse