

# Ateliers Dat'accélère : méthodologie des ateliers

|   |          |
|---|----------|
| <b>Contexte : les données, un continent encore trop méconnu</b>           | <b>2</b> |
| La réutilisation des données : un parcours du combattant pour les novices | 2        |
| Des opportunités manquées pour les porteurs de projets                    | 3        |
| Des échanges trop rares entre producteurs et réutilisateurs de données    | 3        |
| <b>Objectifs et méthode des ateliers</b>                                  | <b>4</b> |
| <b>Contenu des ateliers</b>   | <b>5</b> |
| Phase 1 : "Wanted Data List", comprendre les besoins data du projet       | 5        |
| Phase 2 : la collecte et l'évaluation des données                         | 6        |
| Phase 3 : le plan de route data   | 7        |

## AVITEM-ODEON

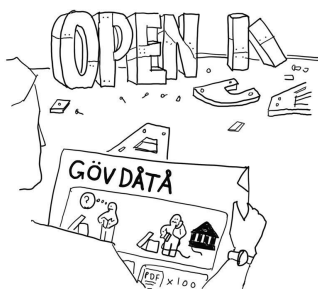
### “Ateliers Dat'accélère : méthodologie des ateliers”

# Contexte : les données, un continent encore trop méconnu

## La réutilisation des données : un parcours du combattant pour les novices

Notre expérience, à travers notamment l'organisation de la [Data Literacy Conference](#), nous a montré que la compréhension des données reste une compétence encore mal répandue dans les organisations. Souvent, nous entendons que les données sont arides, froides ou ennuyeuses, que “ce n'est pas pour moi” ou que c'est “trop technique”.

Quand des personnes novices dans le domaine dépassent cette première appréhension, elles se heurtent généralement à **trois obstacles** lorsqu'elles tentent de découvrir et d'exploiter les données disponibles.



Le premier est celui de la **découvrabilité** : à l'heure actuelle, les données sont très difficiles à trouver du fait de la faiblesse des moteurs de recherche des portails, de métadonnées incomplètes mais aussi de l'éparpillement des jeux de données sur différents portails ainsi que de l'absence d'harmonisation des données entre les différents acteurs publics. L'Open Knowledge Foundation dans le [rapport annuel de l'Open Data Index](#) a souligné cette difficulté : “la découvrabilité des données est un pré-requis pour que

l'open data réalise son potentiel et, à l'heure actuelle, la plupart des données sont très dures à trouver.”

Le deuxième obstacle majeur est celui de la **documentation** des données. Beaucoup de données sont accompagnées d'une description très réduite. Dans une étude portant sur les portails 12 villes majeures en France, une [étude de Dataactivist](#) a montré que la moitié des descriptions des jeux de données faisait moins de 180 caractères (soit un peu moins de la longueur d'un tweet) et 4% des jeux de données ont une description supérieure à 1000 caractères soit moins d'une demi-page. Dans ces conditions, il est très difficile pour un usager de comprendre ce que contient le jeu de données, ses conditions de production des données et ses limites.

Enfin, le troisième obstacle porte sur la **qualité** des données. Le rapport de [l'Open Data Barometer](#) signale ainsi que “les données des gouvernements sont souvent incomplètes, pas actualisées, de mauvaise qualité et fragmentaires” et que “de manière générale, la gestion et la publication des données ouvertes sont fragiles et sujettes à des

## AVITEM-ODEON

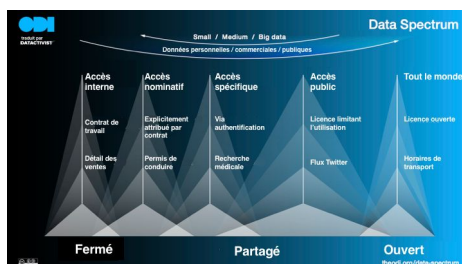
### “Ateliers Dat'accélère : méthodologie des ateliers”

erreurs multiples.” Cependant, la qualité ne peut être évaluée que par la confrontation entre des domaines de pratiques aux enjeux différents. En d'autres termes, la qualité des données ne dépend que de l'usage qui en est fait. Ce point rejoint le précédent : la documentation est indispensable pour que le public parvienne à saisir les usages précédents des données et les adapte à de nouveaux usages, souvent inconnus de celles et ceux qui les produisent.

Sans médiation, la réutilisation des données peut donc s'apparenter à un parcours du combattant. Le **mouvement de la data literacy** défend l'idée que la capacité de produire, comprendre et utiliser des données numériques devient une compétence essentielle, presque au même titre que lire, écrire ou compter. C'est dans cet esprit que nous souhaitons conduire les ateliers Data'accélère en accompagnant des porteurs de projet dans la compréhension du potentiel encore sous-exploité des données.

## Des opportunités manquées pour les porteurs de projets

Nombreux sont les porteurs de projet à passer à côté des données du fait des problèmes évoqués précédemment. Pourtant, les données peuvent potentiellement décupler l'impact d'un projet en offrant de nouvelles opportunités de création de services, en proposant de nouveaux éclairages sur les phénomènes étudiés ou en accélérant des processus métier peu optimisés.



Beaucoup de données sont encore inconnues des porteurs de projet alors qu'elles sont disponibles sur les portails open data. D'autres pourraient avoir un impact très fort sur les projets mais n'ont pas été mises à disposition par les acteurs qui les détiennent. Or, la loi permet de demander aux acteurs investis d'une mission de service public les données non-personnelles qu'elles produisent. La possibilité de demander ces données potentiellement ouvrables est encore relativement méconnue. Enfin, hors du spectre de l'open data, certaines données sont accessibles après la signature d'un partenariat ou l'acquittement d'une redevance.

Mobiliser ces données peut permettre aux porteurs de projets de bénéficier de facteurs de différenciation et d'avantages comparatifs décisifs.

## Des échanges trop rares entre producteurs et réutilisateurs de données

## AVITEM-ODEON

### “Ateliers Dat'accélère : méthodologie des ateliers”

Du côté des acteurs publics qui ouvrent des données dans le cadre de politiques d'open data, il est indispensable que les données mises à disposition trouvent un public qui les réutilise pour créer des services, les analyser, les visualiser ou encore mobiliser. Si les données ne trouvent pas un public, l'ouverture des données peut difficilement perdurer du fait du travail que génère la mise à disposition et l'amélioration continue des données. Il est donc indispensable pour les agents en charge d'un projet d'ouverture de données de montrer et de valoriser des cas d'usage.

Or, les espaces d'échange et de dialogue entre producteurs et réutilisateurs de données sont encore rares. Beaucoup de données sont ouvertes puis mises à jour sans que les potentiels usagers n'aient l'occasion d'exprimer leurs attentes, leurs besoins et de faire un retour sur les données mises à disposition. Dataactivist a montré dans une [étude](#) portant sur 16 villes en France, que seules 6 ont ouvert la possibilité d'ouvrir un fil de commentaires pour chaque jeu de données. De ce fait, il n'y a pas de possibilité d'entraide entre les usagers ou de mise en valeur des échanges qui peuvent avoir lieu entre la communauté et les réutilisateurs. Ces échanges sont pourtant essentiels pour fluidifier les frictions inévitables lors de la réutilisation de données produites dans un contexte inconnu.

Les ateliers devront ainsi permettre de favoriser la rencontre entre porteurs de projet et réutilisateurs afin d'aider à la découverte de données pertinentes, de prendre en compte les besoins exprimés et d'encourager à l'ouverture des données dont les usagers ont besoin.

## Objectifs et méthode des ateliers

Les ateliers Data'accélère dans le cadre du projet européen ODEON Interreg Med visent ainsi à répondre aux défis évoqués précédemment :

- Aider les porteurs de projet dans les trois domaines du programme (économies verte et bleue, industrie créative) à cerner le potentiel data de leur projet
- Démultiplier l'impact de projets en mobilisant les données ouvertes et en développant de nouvelles formes d'usage
- Favoriser la rencontre entre porteurs de projets et producteurs de données ouvertes pour améliorer la qualité des données
- Faire émerger 3 projets emblématiques qui seront accompagnés en détail dans la mise en œuvre du plan de route data élaboré à l'issue des ateliers.

Pour aider à la réutilisation des données, les expéditions de données désignent un ensemble de méthodologies pour comprendre le potentiel des données pour un projet

## AVITEM-ODEON

### “Ateliers Dat'accélère : méthodologie des ateliers”

donné. On peut citer par exemple le [data pipeline](#) de School of Data qui désigne sept grandes étapes d'un projet réutilisant des données. Toutes les méthodologies fonctionnent à peu près sur la logique d'un parcours constitué de grands jalons qui va de l'identification des besoins, au sourcing des données jusqu'au prototypage d'une solution.

Synthétisant et complétant ces expériences, nous proposons de nous appuyer sur le [parcours d'évaluation du potentiel data](#) composé par la FING dans le cadre de la campagne Infolab et enrichi par Simon Chignard. C'est une méthodologie française documentée, “open source” qui est déjà utilisée en entreprise auprès du groupe UP. Elle vise à évaluer le potentiel des données (notamment ouvertes) pour un projet en minimum une journée et peut être décomposé en deux ou trois sessions.

Cette méthodologie ne requiert aucune connaissance technique, est reproductible pouvant s'adapter à tout type de projet et débouche sur 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. Force de cette méthodologie, le plan d'action contient notamment une série de dix actions à réaliser maintenant, dans les trois prochains mois et au cours de la prochaine année.

## Contenu des ateliers

La méthodologie du parcours d'évaluation du potentiel data comporte trois phases principales. Selon les projets et les situations, ces phases pourront correspondre au découpage des trois ateliers ou s'étendre entre les différents ateliers. Les animateurs guideront les porteurs de projet selon le niveau d'avancement des groupes.

### Phase 1 : “Wanted Data List”, comprendre les besoins data du projet

Le premier atelier débute par un **décortilage des projets** pour comprendre leur fonctionnement actuel, leur objectif à court et moyen terme, les acteurs mobilisés et l'état des lieux de leur travail par rapport aux données.

Ensuite, les porteurs de projet devront remplir la “**Wanted Data List**” qui est la liste des données idéales qui seraient utiles ou que l'on pense en rapport avec notre sujet. C'est une manière de prendre pied avec le sujet. C'est un exercice simple au cours duquel les participants sont souvent très créatifs. Il est vraiment intéressant que cet exercice soit réalisé à plusieurs, si possible avec un public hétérogène.

## AVITEM-ODEON

### “Ateliers Dat'accélère : méthodologie des ateliers”

A ce stade, il convient de décrire les jeux de données de la manière la plus simple, sans a priori ni référence à un jeu ou un fournisseur de données particulier : c'est la nature des données qui nous intéresse. Par exemple :

- OK : sens de circulation des rues de Marseille.
- KO : SIG de la ville de Marseille (ça ne nous dit pas quelle donnée vous intéresse et ne permet donc pas de savoir si elle n'est pas plus accessible ailleurs).

## Phase 2 : la collecte et l'évaluation des données

Une fois les données idéales identifiées (Wanted Data List), il faut maintenant trouver de vrais jeux de données correspondants à travers une phase de **collecte des données**. Dans certains cas, la réponse est connue (par exemple : données INSEE) mais il est intéressant de trouver plusieurs sources pour une donnée, de manière à pouvoir les comparer en terme de facilité et coût d'usage, de complétude, de qualité des données, etc. -- ce travail sera réalisé à la phase suivante.

La collecte des données est réalisée par le porteur/l'équipe de projet accompagné par les participants individuels aux ateliers. La recherche de données peut être longue et frustrante, un travail de groupe est recommandé. Pour chaque donnée de la Wanted Data List, la méthodologie consiste à réaliser plusieurs types de recherches, des plus générales aux particulières (Wikipédia, moteurs de recherche, experts, portails open data, ressources de curation des données...). Nous fournirons une liste des principaux outils à consulter et accompagnerons les équipes dans les recherches des données. Chaque nouveau jeu de données doit être documenté en vue d'être évalué par la suite.

Dans **l'évaluation**, chaque jeu de données listé dans le catalogue fait l'objet d'une évaluation à travers la "Dataset check list". Cette évaluation produit deux notes sur une échelle de 0 à 4, permettant de positionner chaque jeu de données sur la cartographie des données. Il est intéressant de former des petits groupes pour se répartir l'évaluation des jeux : la cartographie finale, en revanche, gagne à être produite avec toute l'équipe projet, afin de partager les constats et interrogations de chacun, et ajuster le placement de chaque jeu de données.

Chaque donnée doit être notée de 1 à 4 pour sa **valeur**, de la moins critique à la plus critique pour la réussite du projet:

1. ce serait bien d'avoir cette donnée, mais elle n'est pas critique pour ce projet,
2. cette donnée permet d'augmenter le service rendu à l'utilisateur, mais ce n'est pas critique si elle n'est pas disponible,
3. cette donnée est essentielle pour la réussite du projet, cela va fortement réduire la valeur pour les utilisateurs si elle manque,
4. cette donnée est critique: ce serait très difficile de réaliser ce projet sans cette donnée.

...et de 1 à 4, selon sa **disponibilité**:

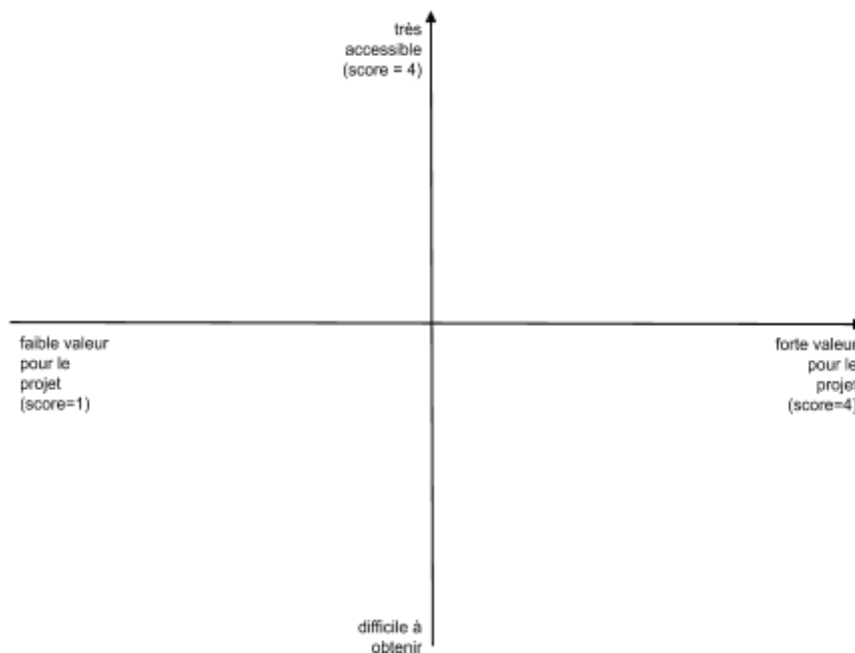
## AVITEM-ODEON

### “Ateliers Dat'accélère : méthodologie des ateliers”

1. cette 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 investissements supplémentaires,
3. cette donnée est déjà disponible, mais pas à un niveau de détails ou de qualité suffisants pour notre projet,
4. cette donnée est disponible, et son niveau de détails et sa qualité sont suffisants pour notre projet.

Au terme de cette évaluation, une cartographie des données permet de visualiser les données à mobiliser en priorité :

Cartographie des données pour votre projet  
Positionner chaque donnée selon ses scores de valeur et de disponibilité



## Phase 3 : le plan de route data

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

On recense 5 options en matière de **sourcing des données**:

1. Sécuriser : s'assurer qu'une donnée critique pour le projet ne sera pas dégradée ou que sa fourniture ne sera pas interrompue,

## AVITEM-ODEON

### “Ateliers Dat'accélère : méthodologie des ateliers”

2. Augmenter : investir dans l'amélioration d'une donnée, sa qualité ou son niveau de détails,
3. Collecter : mettre un place de collecte pour récupérer des données utiles pour le projet mais pas ou peu disponibles - on peut étudier l'intérêt du crowdsourcing,
4. Nouer des partenariats : établir des accords avec des tiers pour accéder à leurs données, ou faire de l'échange de donnée entre partenaires,
5. Acheter : identifier des sociétés (courtiers en données) qui pourraient vendre les données qui nous manquent.

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

| Donnée | Action (à expliciter en quelques mots) |
|--------|--|
|        |  |
|        |  |
|        |  |

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

- L'exactitude: les données sont-elles complètes et fidèles à la réalité qu'elles décrivent ?
- La fraîcheur: de quand datent les données ? à quelle fréquence sont-elles mises à jour ?
- La granularité: les donnée sont-elles disponibles au niveau individuel ou seulement de manière agrégée ?

Ensuite, les porteurs de projet sont invités à **identifier 10 actions à réaliser maintenant / dans les 3 prochains mois / au cours de la prochaine année.**