

SYNTHESE DES ATELIERS DE TRAVAIL DE LA JOURNEE OPEN DATA DU 14 NOVEMBRE 2019



Séminaire et ateliers pour éprouver une méthodologie, des outils, pour identifier les jeux de données à ouvrir « en priorité » et des arguments généraux sur l'OpenData

29 NOVEMBRE 2019
VERSION 1.0

1. Journée Open Data du 14 novembre

1.1. Matinée sous forme d'un séminaire

Dix intervenants issus de la société civile ou de la communauté universitaire se sont relayés en salle de conférence de la Maison des Universités à Paris, pour présenter le sujet de l'Open Data dans son ensemble. Ce fût l'occasion de faire un focus sur le sujet de l'Open Data dans le contexte Enseignement Supérieur et Recherche avant de conclure sur une présentation des travaux de l'Amue en la matière. Les liens vers les supports et les vidéos de cette matinée se trouvent sur la page open data de l'Amue : data.amue.fr

1.2. Ateliers de travail l'après midi

Répartis en 2 groupes simultanés, une vingtaine de participants d'établissements ont contribué à cette réflexion : Eprouver une méthode et des outils pour identifier des jeux de données à ouvrir en priorité et ainsi apporter des éléments à une feuille de route Open Data d'un établissement.

1.3. Au-delà des ateliers, un travail individuel : la matrice de décision

Les participants à la journée ont été sollicités pour compléter une matrice de décision détaillée dans le dernier chapitre de ce document. Le travail de partage autour des résultats n'a pu être réalisé durant la journée, le riche travail produit méritant une analyse spécifique.



2. Zoom sur les ateliers de travail

Le protocole de travail de ces ateliers repose sur trois moments liés entre eux durant lesquels les participants inscrits sont invités à construire un outil d'aide à la décision.

Par soucis de promouvoir les échanges entre les participants, l'atelier se déroule dans deux lieux différents mais selon la même organisation, les animateurs, issus de la coopérative [Dataactivist](#) et de l'Amue ayant préparé en amont. La répartition entre les deux salles a été constituée avant l'évènement et respecte la pluralité des statuts des participants (VP, DSI, Chargé de mission Data, acteur de l'opendata des universités ou écoles)

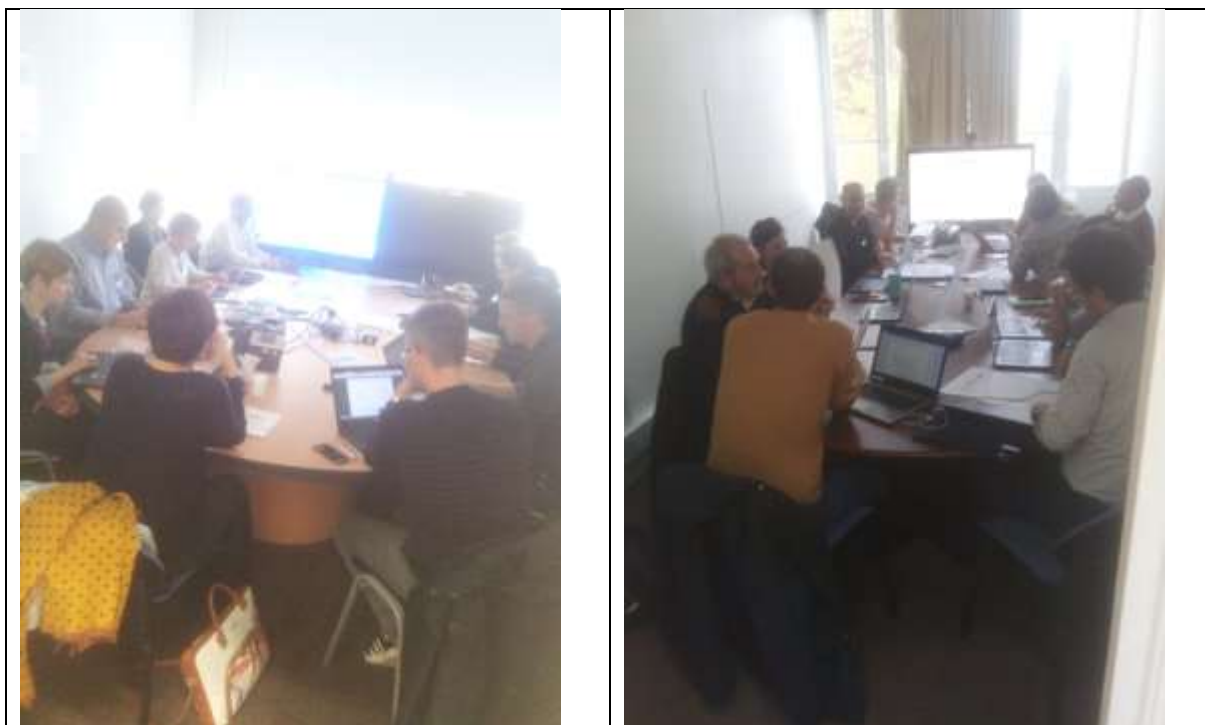


Figure 1 : Deux ateliers dans deux salles différentes

2.1. Cherchons les données

Le premier moment d'une durée de 30 minutes, intitulé idéation, permettait aux participants de proposer des jeux de données qui pourraient être ouverts dans leur entourage. Ils sont invités à renseigner une base de données en ligne (une par salle) en précisant la description du jeu de données (« *Décrire le jeu 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.* ») l'usage envisagé suite à l'ouverture de cette donnée et aussi la thématique (cf figure suivante).



Thématique

Select an option

Formation et vie de l'étudiant ✕ Recherche ✕

Conservation culturelle (bibliothèques) ✕ RH ✕ Finances ✕ Achats ✕

Juridique ✕ Gestion des missions ✕ Gestion du SI ✕ Gestion de la sécurité ✕

Immobilier ✕

Gestion de la documentation et de la connaissance (GED, archivage etc...) ✕

Données transverses et partagées ✕ Moyens généraux ✕ Pilotage ✕ Autre ✕

Figure 2 : Thématique de jeu de données

En fin, il est demandé si le jeu de donnée existe ou non dans l'université.

Données idéales	Thématique	Usage(s) envisagé(s)	Trouvée ?
occupation des salles de cours	Immobilier	optimiser l'allocation à diverses échelles de temps (insta...	✓
Taux Homme/Femme par services et structures	RH	Observatoire mixité, égalité	✓
Jeu de données produit actuellement : parcours d'étude par étudiant e...	Formation et vie de l'étu...	- connaître notre "production". - publication simplifiée p...	✓
Liste des bâtiments avec - coordonnées GPS -liste des laboratoires et s...		permettre aux étudiants, chercheurs et personnels de s...	✓
Personnels des laboratoires par établissement et corps	Recherche	Calcul des coûts complets de la recherche avec charges ...	✓
Occupation des salles de groupe dans les bibliothèques universitaires	Conservation culturelle (...)		✓
Formations ouvertes à la mobilité internationale	Formation et vie de l'étu...	Exploitation par les dispositifs européens (es, projets KA...	✓
parcours étudiants	Formation et vie de l'étu...	comprendre les itinéraires des étudiants pour - mieux ...	✓
Liste des personnels recherche par laboratoire	Recherche	utilisation RH, annuaire interne et externe calcul des ress...	✓
Nombre des étudiants par filière	Formation et vie de l'étu...		✓
Formations compatibles NCU (dont au moins une unité d'enseignement...	Formation et vie de l'étu...	Exploitation par les organismes de formation (inter-étab...	✓
Taux de réussite par Discipline	Formation et vie de l'étu...		✓
Suivi des étudiants (inscription, réussite, réorientation...) par domaine d...		faciliter l'orientation des étudiants et le suivi pédagogiq...	✓
Nombre de publications par laboratoire	Recherche	Visibilité des laboratoires	✓
Offre de formation accréditée	Formation et vie de l'étu...	Exploitation par des organismes type Onisep	✓
Nombre de mobilités par laboratoire	Recherche	Attractivité recherche et formation établissement	✓
(HAL) Nombre de publications en CC	Gestion de la documenta...	Campagne HCERES	✓
données cartographiques de l'établissement avec les information séma...	Gestion de la sécurité - imm	Comme les données sont hébergées sur OSM, il est pos...	✓
Taux d'étudiants en emploi après deux ans // diplômé	Formation et vie de l'étu...	Emploabilité répondant au marché de l'emploi	✓
(Macro-) Structure analytique	Pilotage	Benchmark / best practice	✓
Scolarité et notes (y compris pré-ESR), anonymisé et non-anonymisé, hi...	Formation et vie de l'étu...	- orientation et remédiation basées sur des statistiques ...	✓
Liste des thèses soutenues par laboratoire par ED et par directeur de t...	Recherche	Connaissance des axes de recherche d'un EC, d'une égal...	✓
Modules de formation continue	Formation et vie de l'étu...	Exploitation par les organismes de formation et d'emploi	✓
Nombre de publications en CC par laboratoires	Recherche	Campagne HCERES	✓
Origine des financements des contrats de recherche		Suivre l'évolution des partenariats de recherche et le rôl...	✓
Taux de personnes et étudiants utilisant quotidiennement les transports...	Autre	Permettre de voir les répercussion des aléas/retards de ...	✓
Nombre d'étudiants/Heure de début des cours et Nombre d'étudiants/...	Formation et vie de l'étu...	Lissage des fils d'attente au RL, fluidification des transp...	✓

Figure 3 : Extrait des résultats de la phase d'idéation

2.2. Qualifions les données

Après que les animateurs aient « toilé » les deux bases de données, notamment en cherchant les jeux de données similaires pour un usage similaire, un deuxième moment, intitulé évaluation, débute d'une durée de 45 minutes, permettant aux participants de qualifier davantage les jeux de



données toujours par groupe de trois, chacun qualifiant ses jeux de données renseignés précédemment.

<p>En combien de temps ce jeu peut-il être disponible pour être retravaillé ? 1 : des mois ou plus ; 2 : quelques semaines ; 3 : quelques jours ; 4 : immédiatement</p>	<p>Quel est le format source ? 1 : pdf ; 2 : format d'une application métier ; 3 : Excel (non tabulaire) ; 4 : CSV ou autre format lisible machine (JSON...)</p>	<p>Quel serait le travail de mise en qualité nécessaire avant l'ouverture de ce jeu de données ? 1 : travail très important nécessaire ; 2 : travail moyen nécessaire ; 3 : peu de travail nécessaire ; 4 : aucun travail nécessaire</p>
<p>Ce jeu de données présente-t-il une sensibilité ou des limites juridiques à sa réutilisation, par exemple au regard du RGPD ? 1 : aucune limitation ; 2 : de légères incertitudes nécessitent un éclaircissement juridique ; 3 : le jeu peut être ouvert à condition d'être retravaillé (anonymisé, par exemple) ; 4 : il est impossible de publier ce jeu de données.</p>	<p>Potentiel de création de service 1 : aucun intérêt 2 : intérêt à terme 3 : plusieurs services imaginables 4 : nombreux services imaginables</p>	<p>Potentiel de réutilisation interne 1 : aucun service interne ou partenaire 2 : un service interne ou partenaire 3 : plusieurs services internes ou partenaires 4 : tous les services internes ou partenaires</p>
<p>Opportunité de communication/transparence 1 : pas d'intérêt 2 : enjeu de communication/transparence existant 3 : fort enjeu 4 : essentiel pour l'établissement</p>		

Figure 4 : Evaluation du jeu de donnée

Thématique	En combien de temps ...	Quel est le format source ?	Quel serait le trava...	Sensibilité RGPD	Potentiel de créati...
Occupation des salles de g...	Conservation culturelle (bi...	★★★★	★★★	★	★★★★
Taux de réussite par Discop...	Formation et vie de l'étud...	★★	★★	★★★	★★★★
Formations ouvertes à la ...	Formation et vie de l'étud...	★★★★	★★	★	★★
occupation des salles de c...	Immobilier	★★	★★	★	★★
Taux hommes/femme par ...	RH	★★★★	★★★★	★	★★
Jeu de données produit ac...	Formation et vie de l'étud...	★★★★	★★★★	★★	★★★★
Nombre de publications p...	Recherche	★★	★★	★	★★
(HAL) Nombre de publicati...	Gestion de la documentati...	★★★★	★★	★	★★
Taux d'étudiants en emplo...	Formation et vie de l'étud...	★★	★★	★	★★★★
Les données de la recherche	Recherche	★	★	★★★★	★★★★
Liste des personnels reche...	Recherche	★★★★	★★	★★	★★
Liste des bâtiments avec ...		★★★★	★★	★★	★★
Taux d'usagers de l'établi...	Autre	★★	★★	★★	★★
Taux de personnes et étud...	Autre	★★	★★	★	★★
Personnels des laboratoir...	Recherche	★	★★	★★★★	★★
données cartographiques ...	Gestion de la sécurité, inem...	★★	★	★★	★★
Formations compatibles N...	Formation et vie de l'étud...	★	★★★★	★★	★★
Taux de personnes et étud...	Autre	★★	★★	★★	★★
Nombre de mobilités par L...	Recherche	★★★★	★★	★	★★
Nombre des étudiants par...	Formation et vie de l'étud...	★★★★	★★	★	★★
Origine des finacement d...		★★★★	★★★★	★★	★★
Liste des thèses soutenue...	Recherche	★★	★★★★	★★	★★
Offre de formations accrédi...	Formation et vie de l'étud...	★★★★	★★★★	★★	★★★★

Figure 5 : Extrait des résultats de la phase de qualification

2.3. Présentons la matrice produite par les ateliers

Le dernier moment, d'une trentaine de minutes, permet de rendre une visualisation en un graphe dont l'ordonnée est la faisabilité d'ouverture et l'abscisse la valeur d'usage.

Ces deux valeurs sont calculées à partir des différentes notes attribuées lors de la séquence d'évaluation.

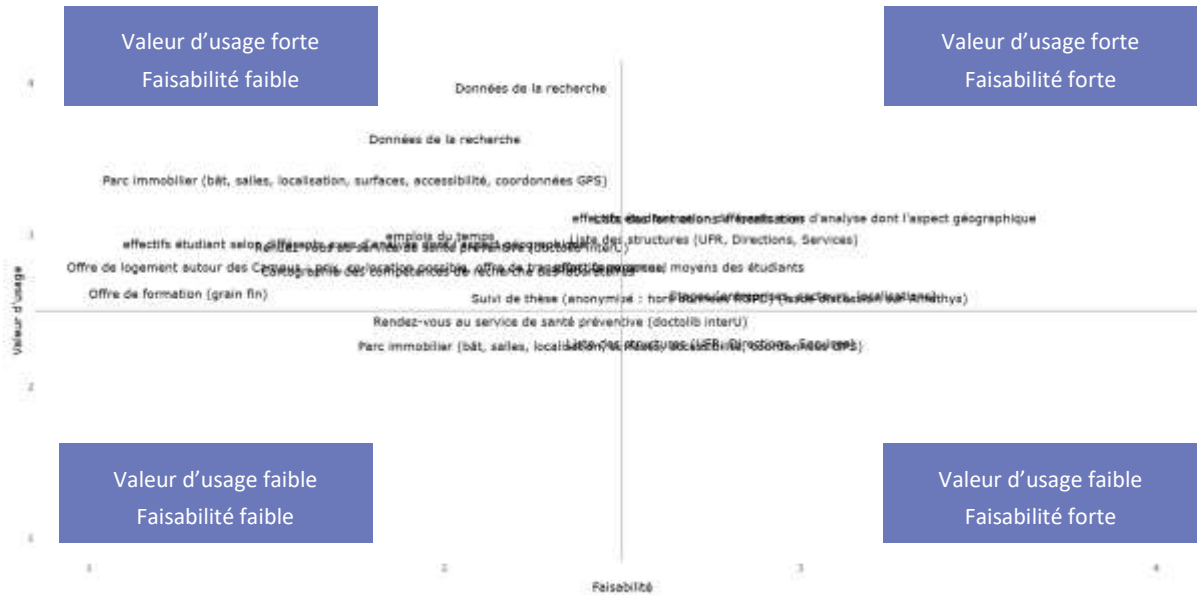


Figure 6: Exemple de matrice produite

2.4. Bilan des ateliers

2.4.1. Bilan sur la méthode et les outils utilisés

Appliquer cette démarche d'idéation puis de qualification des jeux de données a bien fonctionné. Les participants, à l'occasion de la synthèse, ont témoigné de leur satisfaction sur la méthode et l'intérêt de l'outil.

Quelques propositions d'amélioration ont été suggérées dans l'optique d'une généralisation de ce dispositif :

- Laisser à chaque établissement la possibilité de paramétrer la pondération des critères pour calculer les axes « faisabilité » et « valeur d'usage ».
- Améliorer l'outil de visualisation
- Doter ce dispositif d'une formation/information (webinaire, vidéo pédagogique,...) pour les futurs participants
- Optimiser l'ergonomie de l'outil

Ces différentes propositions pourront être traitées par le groupe de travail OpenData que l'Amue propose d'animer (prendre contact à numerique@amue.fr pour contribuer).

2.4.2. Première tendance

Avec toute la réserve que requiert la publication d'une tendance sur cet échantillon de participants, notons ces jeux de données qui apparaissent comme étant les plus positifs pour les deux axes d'analyse (valeur ajoutée et faisabilité) soit dans la zone haute à droite du cadran de résultat, soit une « valeur d'usage forte, faisabilité forte » :

- « Offre de formation accréditée » pour « Exploitation par des organismes type Onisep »
- « Liste des structures (UFR, Directions, Services) » pour « Usages internes et réutilisation SI »

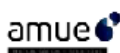


- « Effectifs étudiants selon différents axes d'analyse dont l'aspect géographique » pour « répondre aux fréquentes sollicitations par les partenaires de l'établissement pour l'aménagement du territoire »
- « Origine des financements des contrats de recherche » pour « Suivre l'évolution des partenariats de recherche et le rôle/poids des agences de financement »
- « Taux de réussite par Discipline par discipline sans usage identifié en séance
- « Taux d'étudiants en emploi après deux ans // diplômé » pour « Employabilité répondant au marché de l'emploi »

Cette première tendance mérite d'être confirmée ou infirmée par d'autres itérations de ce dispositif.

3. Matrice de décision

3.1. L'outil



Matrice de prise de décision

Quels sont les freins et les moteurs à l'ouverture des données ?

Prénom et nom :
Fonction :
Etablissement :

1. Lister les freins et les moteurs à l'ouverture
2. Noter chaque frein et chaque moteur de 1 (très faible) à 10 (très fort)

Freins	Moteurs
<i>Les avantages à ne pas ouvrir :</i>	<i>Les avantages, les bénéfices à ouvrir :</i>
<i>Les inconvénients, voire les risques à ouvrir :</i>	<i>Les risques à ne pas ouvrir :</i>
Poids total des freins	Poids total des moteurs



Il s'agit d'une matrice Freins/Moteurs appliquée au sujet Open Data qui a été soumise aux participants. Elle permet d'identifier et de noter un ensemble de freins, d'une part, et de moteurs, d'autre part. Au final, le cumul des poids, de part et d'autre, permet de se positionner positivement ou non sur l'ouverture des données. Il permet également de lister des freins pour lesquels un travail de sensibilisation explication peut être entrepris afin d'en diminuer l'importance.

3.2. Quelques utilisations de l'outil

3.2.1. Seul ou en collaboration en une itération

Compléter cette matrice pour une personne ou le faire en mode collaboratif par un groupe simultanément permet de donner une tendance positive ou négative à un instant T.

3.2.1. De manière itérative

Effectuer cette liste et mesure de résultats à deux moments donnés permet de mesurer l'évolution des avis, de la liste des arguments et leurs poids entre deux périodes. Par exemple, mener l'exercice en amont d'une réunion de présentation de l'Open Data puis faire « corriger » cette même matrice après : cela peut être un indicateur sur la pertinence de la réunion, qualité de réception des messages passés et montrer qu'une présentation sur le sujet peut faire lever certains freins ou certaines craintes (ou en ajouter).

C'est cet usage qui a été prévu pour la journée du 14 novembre mais faute de temps, n'a pu être mené.

3.3. Résultats et exemples

En très large majorité, la « balance » entre moteurs et freins sur les matrices données par les participants étaient à l'avantage des moteurs.

Ci-dessous quelques exemples d'arguments fournis par les participants. **NB** : les poids n'ont pas été repris dans cette compilation car ils sont sensibles à l'auteur de la notation.



Matrice de prise de décision Quels sont les freins et les moteurs à l'ouverture des données ?

QUELQUES EXEMPLES IDENTIFIES LE 14 NOVEMBRE
SANS LES POIDS

1. Lister les freins et les moteurs à l'ouverture
2. Noter chaque frein et chaque moteur de 1 (très faible) à 10 (très fort)

Freins	Moteurs
<p>Les avantages à ne pas ouvrir :</p> <ul style="list-style-type: none">- Economiser des ressources RH, matériel- ne pas confronter la culture du partage VS la culture de la propriété des données- Maitriser l'usage de nos données- « Pour vivre Heureux, vivons cachés »- Pas besoin de faire de tri entre données confidentielles/sensibles et les autres	<p>Les avantages, les bénéfices à ouvrir :</p> <ul style="list-style-type: none">- augmentation de la qualité des données- normalisation qui accroît l'interopérabilité des services publics- Créer de nouveaux services- transparence des pouvoirs publics, prérequis à la participation des citoyens- baisse des coûts- Nouvelles collaborations- Accroître la visibilité des projets de l'établissement et de l'expertise des chercheurs- Garantir la citation correcte des données- Plus de chapelles en interne, moins de silos.- Amélioration des processus transverses
<p>Les inconvénients, voire les risques à ouvrir :</p> <ul style="list-style-type: none">- risques liés à la vie privée- Risque juridique éventuel (réel ou non), données confidentielles- risque à la marchandisation des données de l'établissement- ne pas être en cohérence avec les données publiées sur le MESRI sur le même périmètre- Déontologie- RGPD (hors recherche, la plupart des données concernent des individus)- Risque de révéler des informations sensibles (locaux et terrorisme par ex).- Risque de comparaison et de perte d'estime	<p>Les risques à ne pas ouvrir :</p> <ul style="list-style-type: none">- Ne pas respecter une obligation légale, saisine de la CADA- Laisser circuler des données erronées- Non respect des exigences des financeurs- On rend moins de services- « Ne pas exister »
Poids total des freins	Poids total des moteurs