DONNÉES ENVIRONNEMENTALES DU PÔLE MINISTÉRIEL :

Étude sur les usages, difficultés et besoins des utilisateurs, dans la recherche, la production et l'utilisation de données ouvertes.

Première phase de l'enquête
novembre 2022 - avril 2023

MÉL : martin.cahen@developpement-durable.gouv.fr- MAI 2023

DONNÉES ENVIRONNEMENTALES DU PÔLE MINISTÉRIEL :
Étude sur les usages, difficultés et besoins des utilisateurs,
dans la recherche, la production et l'utilisation de données ouvertes.
Enquête de novembre 2022 - avril 2023 (rapport détaillé)

Pilotée par

MINISTÈRES
TRANSITION ÉCOLOGIQUE
COHÉSION DES TERRITOIRES
TRANSITION ÉNERGÉTIQUE
MER
Liberté
Égalité
Fraternité

En collaboration avec

OPENDATA FRANCE

DOCUMENT

OBJET

Résultats détaillés et éléments d'analyse d'une enquête menée entre novembre 2022 à avril 2023 auprès d'utilisateurs professionnels des données, afin d'identifier leurs usages, difficultés et besoins, dans la recherche, la production et l'utilisation de données environnementales ouvertes soutenant les politiques publiques.

HISTORIQUE DES VERSIONS

V.0 2023.10.04 Rédaction initiale.

RÉDACTION

Valentin LABIT

Avec la participation de

Alexandra DOUINE

Matias GOMEZ

SOMMAIRE

1. CONTEXTE

1.1. Contexte de l'étude

1.2. La naissance d'Écosphères

1.3. Les fonctionnalités d'Écosphères

1.4. Les objectifs de l'étude

2. DÉROULÉ DE L'ÉTUDE

2.1. Méthode

Revue de l'existant
Le questionnaire
Les ateliers collectifs
Retours critiques

2.2. Population

3. PRINCIPAUX RÉSULTATS

3.1. Codage des données

3.2. Définition de la population - Qui sont les cibles ?

Types de structures
Famille de professions

3.3. Données recherchées - Quelles données recherchent les cibles ?

Modalités de recherche - Comment recherchent-elles les données ?
Usages de données - À quelles fins ?
Difficultés - Quelles sont les difficultés ?
Solutions proposées - Quelles sont leurs attentes et besoins ?
Types de solutions
Autres solutions pertinentes

3.4. Retour sur l'hypothèse - Est-ce que la nouvelle entrée par cas d'usage est adaptée ?

Avis sur l'hypothèse
Capacité perçue
Inconvénients de l'hypothèse

Retours de l’équipe

3.5. Constitution de la mesure d’impact

Difficultés et fonctionnalités
Difficultés, fonctionnalités, changements et changements

Points de vigilance et limites de l’étude

L’importance de la communauté
Le public cible
Les bouquets de données orientés

Conclusion

Annexes

Liste des inconvénients de l’hypothèse
Liste des principes de fonctionnement d’Écosphères

1. CONTEXTE

1.1. Contexte de l'étude

Dans le cadre de ses missions d'ouverture et de diffusion de la donnée, le pôle ministériel (désignant ci-après Ministère de la transition écologique et de la cohésion des territoires) entretient des catalogues nationaux et régionaux visant à mettre à la disposition, en accès libre ou restreint, les données produites. Cet engagement est souligné par le rapport Bothorel et encouragé au travers de la législation sur l'ouverture et le partage des données.

Dans un contexte de crise environnementale, l'ensemble de ces données revêtent effectivement un potentiel important pour soutenir les politiques publiques et faciliter les choix et initiatives des décideurs en matière de transition écologique.

Toutefois, une enquête de la Direction du Numérique du pôle ministériel (DNUM anciennement SNUM) menée en novembre 2020 auprès d'utilisateurs externes des données (bureaux d'études (BE), agences d'urbanisme (AU) organismes de recherches universitaires et collectivités) et des travaux conduits par OpenDataFrance dans la cadre de leur programme "Données et Transition", ont conclu que de nombreux obstacles rendent difficiles l'exploitation des données ouvertes.

L'objet de la présente étude, pilotée par l'Ecolab, laboratoire d'innovation du Commissariat Général au Développement Durable, réalisée en collaboration avec l'association OpenDataFrance, est d'explorer les usages actuels des utilisateurs pour tenter d'identifier les solutions qui pourraient favoriser l'appropriation massive de ces données au service de la transition environnementale.

1.2. La naissance d'Écosphères

Les difficultés précédemment évoquées ont conduit le Ministère de la Transition écologique et de la cohésion des territoires à développer Ecosphères, un guichet unique d'accès à la donnée du pôle ministériel. Mais des entretiens de terrains menés par l'Ecolab ont démontré que la possibilité d'accéder à l'ensemble de ces données au sein d'une seule et même plateforme ne suffisait pas à résoudre la problématique de découvrabilité des données.

> *"On ne souhaite pas avoir accès à 30000 jeux de données, on souhaite le ou les jeux de données dont j'ai besoin au moment où j'en ai besoin"*

Pour réussir à dépasser ces difficultés, le Ministère de la Transition écologique et de la cohésion des territoires a imaginé intégrer dans Écosphères un système d'entrée d'accès aux données par "cas d'usages".

En complément d'une entrée par territoires, thématiques et par producteurs de données, Ecosphères vise à faciliter l'identification et l'accès aux données ouvertes, mais aussi restreintes du pôle ministériel, via des bouquets de données prédéfinis et adaptés aux besoins des utilisateurs.

1.3. Les fonctionnalités d'Écosphères

Dans sa version actuelle, l'outil propose les fonctionnalités suivantes :

- Moissonner différentes bases de données pour centraliser l'accès à différents jeux de données.
- Rechercher et filtrer les données via le guichet

À l'avenir, Écosphères envisage les fonctionnalités suivantes :

- Proposer un mode de recherche par cas d'usages permettant d'accéder à des bouquets de données pré-constituées.
- Permettre la constitution et l'animation de ces bouquets
- S'adosser à Data.gouv pour une mise en commun des outils et des données de la plateforme

1.4. Les objectifs de l'étude

À ce stade, et avant d'implémenter cette nouvelle entrée, l'équipe d'Écosphères s'interroge sur les utilisateurs de ces données (leurs profils), leurs usages (à quelles fins ils recherchent les données) et de quelle manière ils les recherchent. L'équipe se questionne également sur les fonctionnalités (entrée par recherche libre, par territoire, thématique ou organisation) qui pourraient enrichir l'outil et être associées à cette nouvelle entrée par les usages.

Cette étude vise donc à répondre aux objectifs suivants :

- Catégoriser les utilisateurs d'Écosphères et définir leur périmètre d'action : définir les cibles, les données recherchées, comment et à quelles fins, les difficultés rencontrées, leurs attentes et leurs besoins.
- Vérifier la nouvelle hypothèse d'entrée par les usages : connaître ces usages et s'assurer que la recherche par ces derniers est adaptée aux besoins des utilisateurs.
- Cadrer Écosphères au niveau des fonctionnalités annexes à cette entrée par les usages.
- S'appuyer sur les conclusions de l'étude de la DNUM et les vérifier à plus grande échelle. Cette précédente étude illustre le parcours de recherche de la donnée en identifiant les irritants rencontrés par les utilisateurs. La population participante étant faible, seulement 22 répondants, notre travail vise à approfondir ce travail initial.

Elle s'inscrit également dans les objectifs plus généraux du programme "Données et Transition" mené par OpenDataFrance qui vise à rechercher les meilleures conditions pour mettre la donnée au service de la transition environnementale en poursuivant les objectifs suivants :

- Partager diagnostics, expériences et inspirations, liés au potentiel de transformation des données environnementales en faveur de la transition écologique pour les territoires.
- Identifier les bases de données environnementales, leurs contenus et conditions d'accès.
- Valoriser les cas d'usage où les données sont utilisées de façon exemplaire.
- Mettre à disposition des outils qui facilitent l'extraction et l'appropriation de ces données au niveau local en faveur de projets de transition.

2. DÉROULÉ DE L'ÉTUDE

2.1. Méthode

Notre méthode s'appuie sur les étapes suivantes :

Revue de l'existant

Pour construire notre étude, nous nous sommes appuyés sur les travaux réalisés en amont par l'équipe, mais également sur des études récentes. L'atelier et les entretiens réalisés par l'équipe ainsi que l'étude de la DINUM nous ont permis d'identifier les différentes difficultés que semblent rencontrer les utilisateurs dans leur recherche de données. Suite à l'analyse de nos premiers résultats, nous avons travaillé à définir notre propre typologie de difficultés.

Le questionnaire

Nous avons développé un questionnaire auto-administré et anonyme à remplir en ligne. Il présente 29 items dont une majorité sous la forme de questions ouvertes (20) afin de recueillir des informations plus riches et plus nuancées que dans le cas des questions fermées et de ne pas limiter l'expression des participants aux options préétablies d'un questionnaire à choix multiples. Cet outil nous a permis de toucher un public large et de recueillir un nombre important de données quantitatives. Il visait particulièrement à identifier les utilisateurs cibles d'Écosphères, à comprendre leurs usages, leurs difficultés et à définir le mode de recherche le plus adapté à leurs besoins.

Les ateliers collectifs

Nous avons ensuite réalisé deux ateliers collectifs. Ces ateliers nous ont permis de toucher un public plus restreint, aux profils différents, et d'obtenir des données qualitatives par le biais d'une

interaction directe où les participants étaient libres de partager leurs avis, leurs expériences et ressentis pour nous apporter des éléments de compréhension supplémentaires sur le sujet étudié. Ils ont suivi une méthodologie dite « en entonnoir » qui nous ont finalement permis de confronter les résultats du questionnaire à la réalité des participants, notamment en repartant de leurs difficultés, avant d'imaginer collectivement et d'approfondir des solutions adaptées à leurs besoins.

La matrice d'impact

L'équipe d'Écosphères a souhaité travailler à construire la mesure d'impact du projet. Au travers d'une méthode visant à redéfinir les cibles et objectifs de changement sous-tendu par la plateforme, nous avons travaillé à identifier des indicateurs de mesure permettant d'évaluer l'impact du dispositif.

Retours critiques

Pour la dernière phase de l'étude, nous avons souhaité capitaliser au maximum sur les retours des participants en réalisant un entretien directif approfondi à partir des difficultés recueillies lors des deux premières phases de l'étude. Cet entretien avait pour objectif de "confronter" l'équipe projet aux réserves et points bloquants des répondants vis-à-vis de l'hypothèse d'entrée par les usages intégrée à Écosphères. Les principales difficultés, incompréhensions ou critiques ont ainsi été traduites sous la forme de questions à laquelle l'équipe était invitée à répondre.

Quelques exemples de questions :

- Comment être sûr de récupérer toutes les données environnementales utiles ?
- Comment faire le lien entre politique publique et données associées ?
- Est-ce qu'il faut par conséquent utiliser des mots clés précis ?
- Comment ne pas exclure les utilisateurs qui ne connaissent pas les politiques publiques ?

2.2. Population

L'équipe projet d'Écosphères et l'association OpenDataFrance se sont chargées du recrutement pour les ateliers et de la diffusion du questionnaire en s'appuyant notamment sur leurs réseaux partenaires. Pour toucher un public aussi représentatif que possible, l'idée était de contacter des utilisateurs de données localisés sur l'ensemble du territoire, exploitant des données sur des échelles géographiques différentes (locales, départementales, régionales, voire nationales), travaillant dans des domaines métiers portés par les ministères (environnement, urbanisme, aménagement du territoire...) consommant des données de natures différentes (téléchargement, flux, cartes, métadonnées...).

Le questionnaire a été diffusé à deux listes de partenaires du MTECT sur lesquelles sont inscrits 226 DDT et 181 DREAL et a fait l'objet de partages sur des canaux conversationnels dédiés à la thématique des données ouvertes (slack dataforgood et Mattermost Beta gouv) et sur deux newsletters (data.gouv et interne CGDD). 270 mails de sollicitation ont également été envoyés aux membres de l'association OpenDataFrance entre le 21/02/2023 et le 03/05/2023.

La population participante est la suivante :

Questionnaire	Atelier
131 répondants	Niort 14
	Lyon 8

3. PRINCIPAUX RÉSULTATS

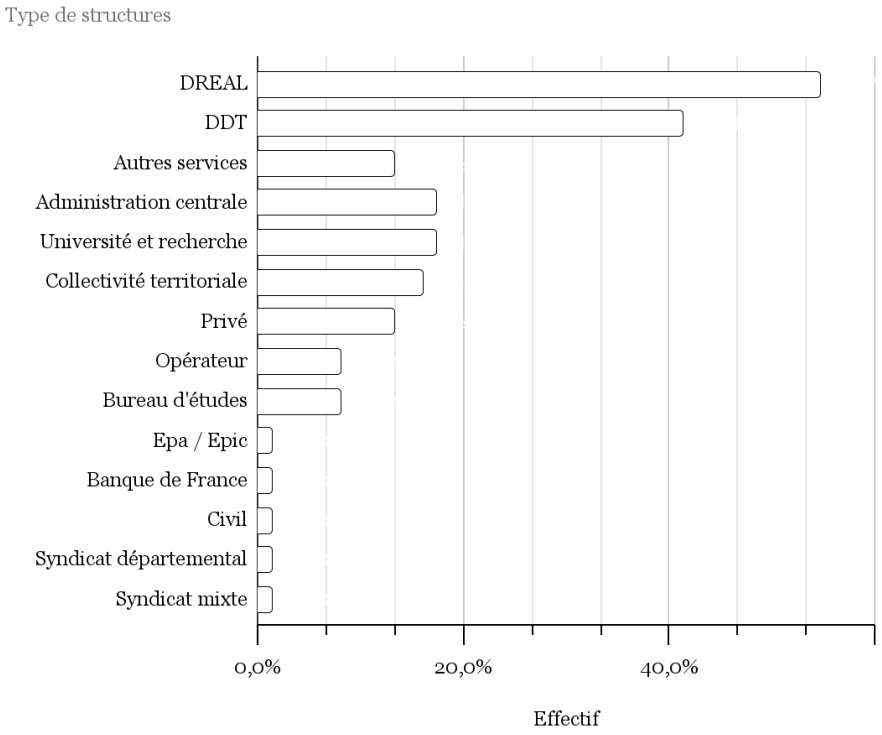
3.1. Codage des données

Les données issues des 20 questions ouvertes du questionnaire et recueillies lors des ateliers ont fait l'objet d'un codage selon la méthode dite **concordance ou fidélité inter-juges**. Le travail visant à les classer dans un ensemble de catégories a ainsi été réalisé de manière indépendante par trois "juges" différents, avant de vérifier dans un second temps l'existence d'un degré d'accord suffisamment élevé entre les jugements de ces différents examinateurs, La principale source d'erreur étant l'évaluateur, cette mesure de la fidélité permet de s'assurer de l'objectivité de la démarche et d'apporter un niveau de validité supplémentaire aux résultats.

3.2. Définition de la population - Qui sont les cibles ?

Types de structures

Les répondants sont en grande majorité issus de services déconcentrés de l'Etat. L'autre majorité est issue des administrations centrales, des universités et de la recherche et des collectivités territoriales. Cela s'explique par le fait que le questionnaire a été diffusé notamment à des usagers travaillant dans des domaines métiers portés par le pôle ministériel.

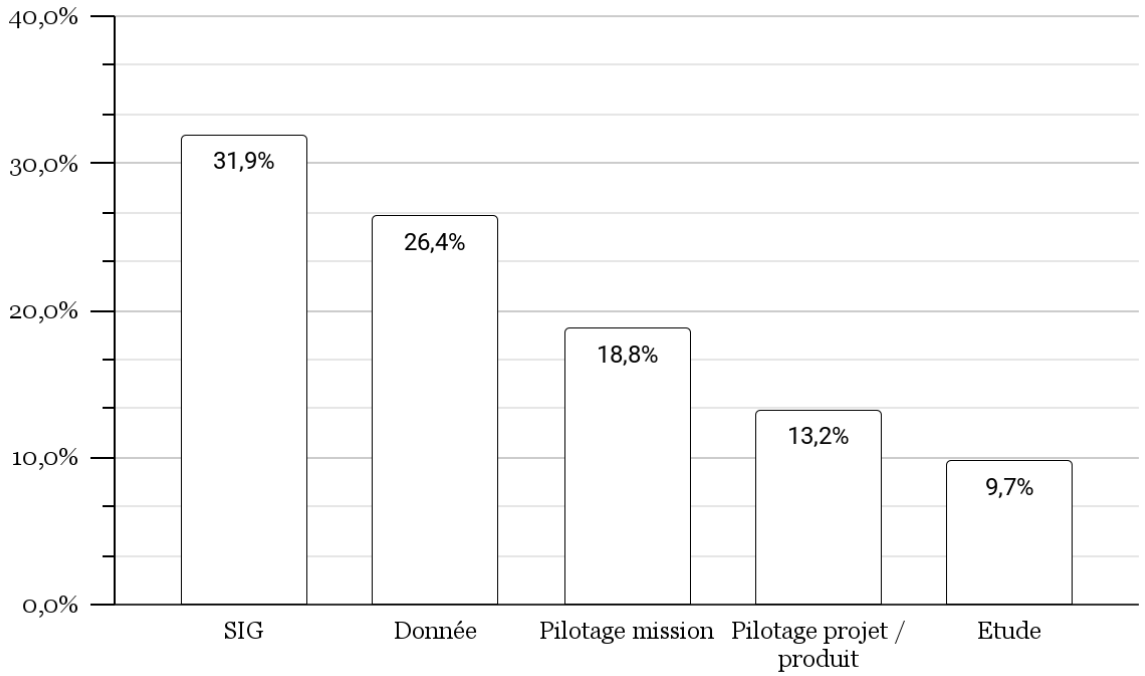


Famille de professions

Les familles de professions qui arrivent en 1e et 2e positions concernent respectivement le SIG et la donnée. Les répondants occupent donc principalement des postes relatifs au SIG, par exemple : *SIGiste, animation SIG, géomaticien* ou des postes relatifs à la donnée, par exemple : *gestionnaire de données, ingénieur de la donnée, data strategist*.

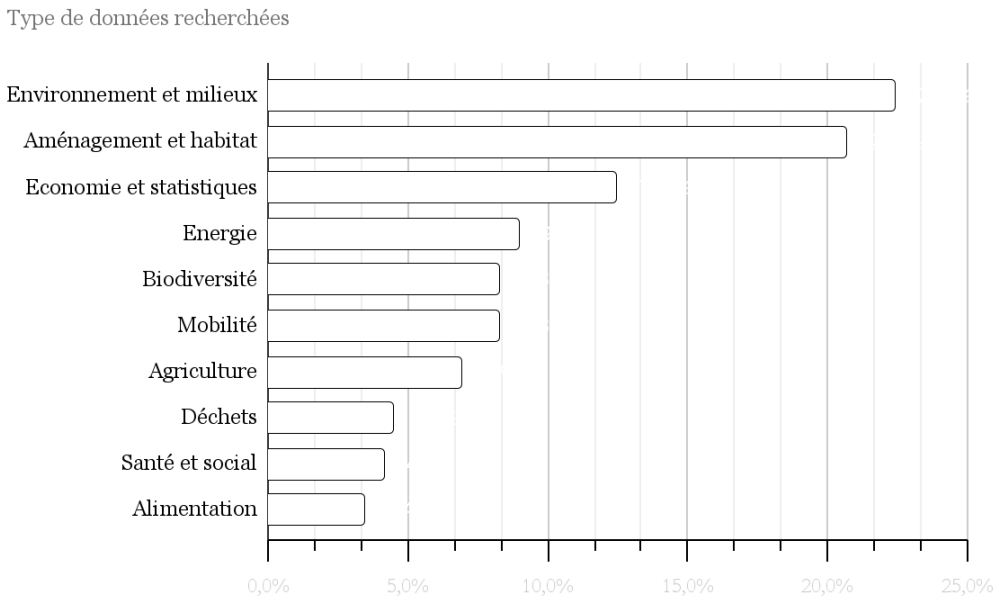
En 3e et 4e position, on retrouve des postes liés à du pilotage de mission ou du pilotage de projet/produit, tels que *chargé de mission, Responsable/chef de la cellule Évaluation et Statistiques, chef de projet, chef de produit, directeur de projet*.

Enfin, en dernière position, on retrouve des professions qui revêtent du domaine de l'étude tels que *chargé d'étude, directeur de recherche, professeur, chargé d'animation observation et prospective.*



3.3. Données recherchées - Quelles données recherchent les cibles ?

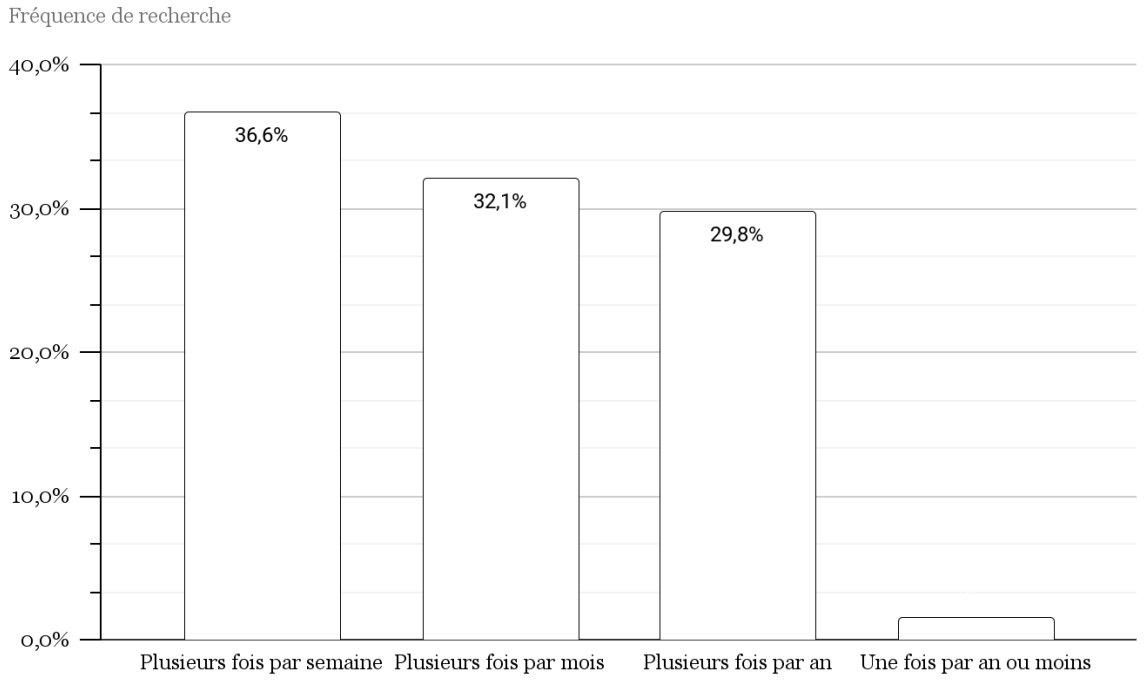
Les répondants recherchent des données qui peuvent être regroupées dans les différentes thématiques portées par les ministères. On observe, à partir de cette vision, que certains domaines sont particulièrement présents : environnement et milieux (eau, qualité de l'air, sol climat risques naturels, bruits et rayonnements), aménagement et habitat, économie et statistiques ou encore l'énergie, la biodiversité ou la mobilité. On peut également noter que de rares thèmes sont un peu plus en retrait : les déchets, la santé ou l'alimentation qui dépend du Ministère de l'Agriculture.



Modalités de recherche - Comment recherchent-elles les données ?

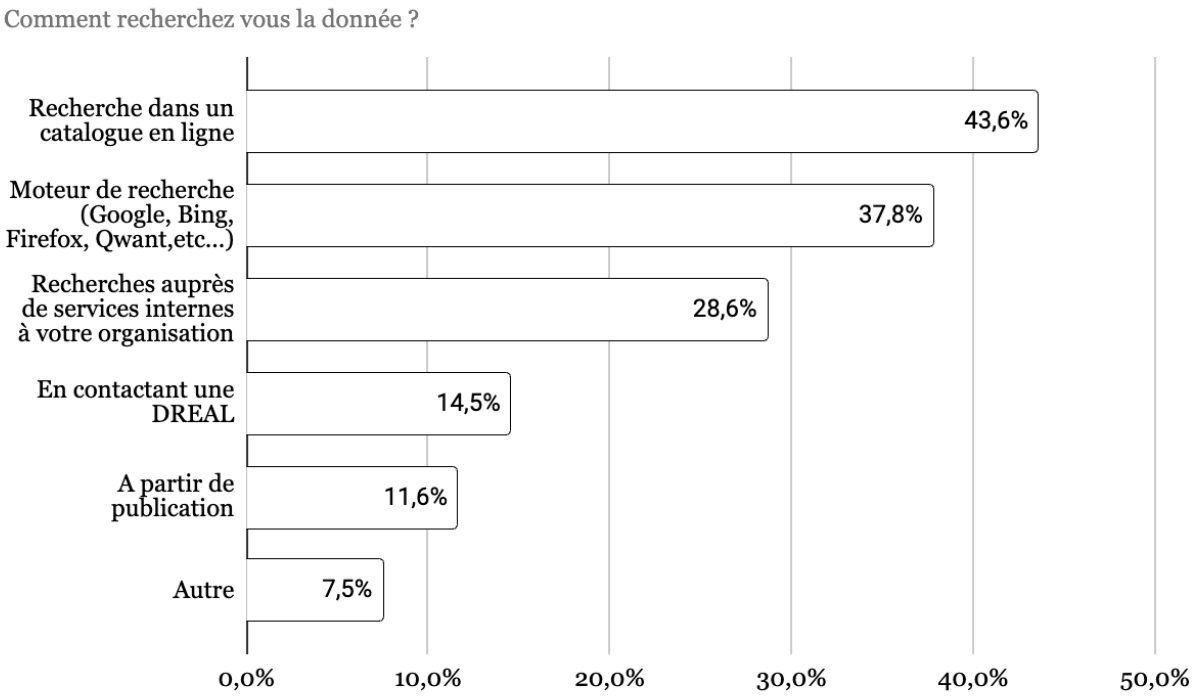
Fréquence de recherche

Les répondants de notre étude ont une fréquence de recherche assez régulière puisque la majorité d'entre eux recherchent de la donnée plusieurs fois par mois.



Modalité de recherche

Les répondants ont une façon de rechercher active puisqu'ils recherchent principalement eux-mêmes depuis des catalogues en ligne (30,3%) ou depuis un moteur de recherche (26,3%). L'utilisation des plateformes de données semble largement implantée dans les habitudes de travail des professionnels.



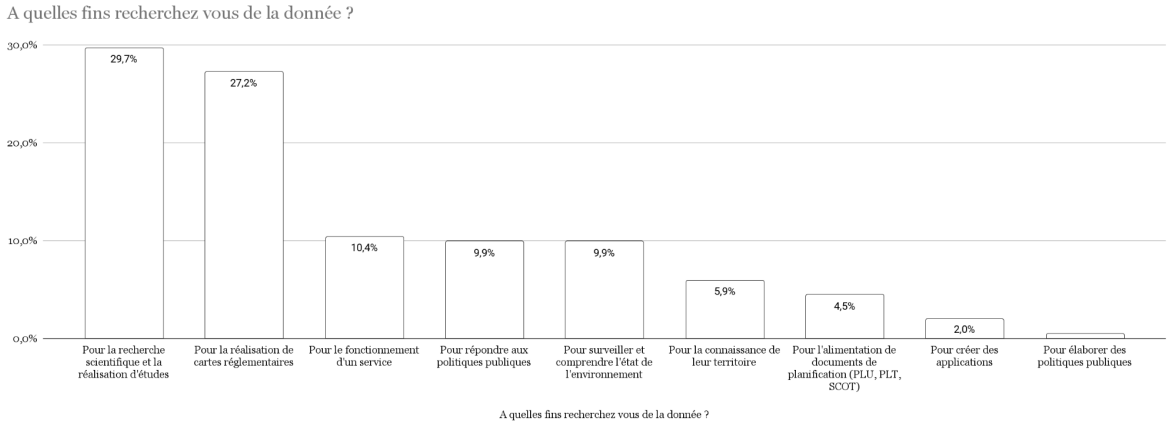
Usages de données - À quelles fins ?

Les cas de réutilisation des données sont classés par grandes thématiques et publiés sur la plateforme ouverte des données publiques françaises data.gouv.fr. Ici, le codage des réponses nous permet d'identifier que l'utilisation des données est dédiée majoritairement à la réalisation d'études et de cartes réglementaires. L'élaboration et la réponse à des politiques publiques ne sont pas directement exprimées comme un usage par les participants. Il est cependant important de signifier que certaines réponses et indices fournis dans les questions suivantes pourraient nous permettre d'inférer ce lien et de supposer que la finalité de ces usages est bien reliée au pilotage de politiques publiques.

Exemple :

"Pour identifier des enjeux tels que la biodiversité, la qualité de l'air et de l'eau, sur les paysages, la nature des sols, les pollutions, les zonages naturels..."

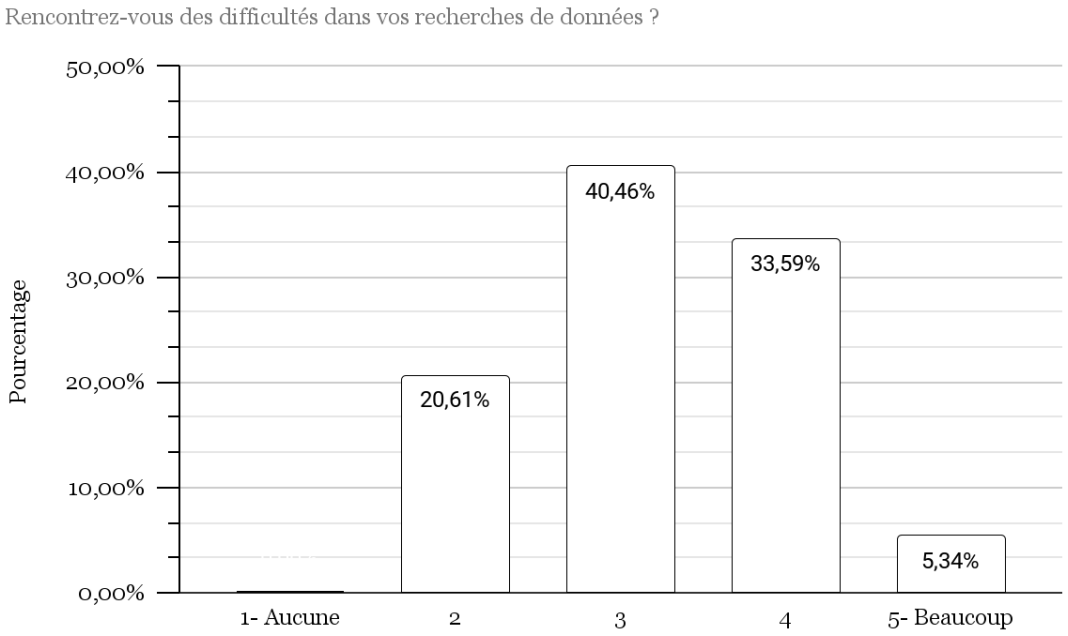
"Pour répondre à des commandes internes réalisées par nos services métiers"



Difficultés - Quelles sont les difficultés ?

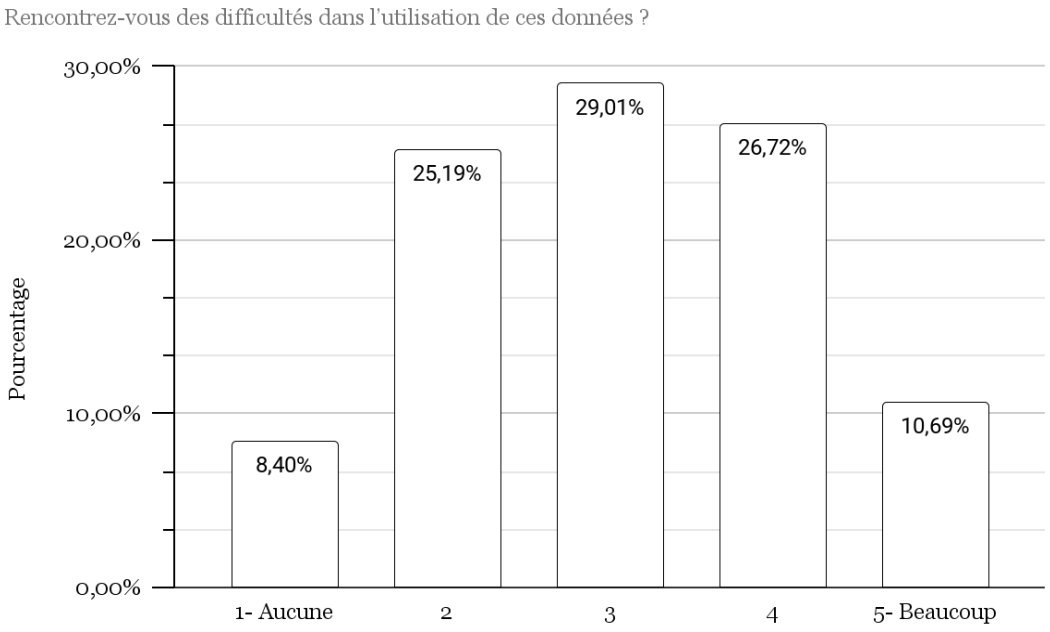
Difficultés dans la recherche

La recherche de données semble poser des difficultés pour les répondants. Aucun répondant ne déclare avoir eu aucune difficulté, la note moyenne atteint 3,2.



Difficultés dans l'utilisation

L'utilisation des données obtient des résultats très proches avec une moyenne de 3.1.



Types de difficultés

Les difficultés d'exploitation des données exprimées par les participants de l'étude s'articulent et se regroupent autour de 6 grandes thématiques qui semblent se suivre au cours du parcours des utilisateurs :

- **Acculturation** : difficultés pour les participants à comprendre et à intégrer les principes des données ouvertes dans les pratiques et les processus des organisations. Cela inclut un manque de connaissances techniques pour utiliser les données et une réticence de certains acteurs à partager des données en raison de préoccupations en matière de confidentialité ou de propriété intellectuelle.

- **Valeur d'usage** : certains participants disent ne pas avoir la capacité de savoir quelles données pourraient être utiles à leurs projets, ce qui affecte leur valeur d'usage. Les données peuvent être complexes et nécessiter une expertise technique pour une utilisation efficace. Il est donc important de fournir des outils pour aider ces utilisateurs.

- **Gouvernance** : il est souvent compliqué de déterminer qui est responsable en cas de mauvaise utilisation des données. Les répondants constatent également qu'il manque de stratégie claire en matière de gouvernance des données ouvertes, ce qui peut compliquer la gestion des données et la mise en place d'outils nécessaires pour leur réutilisation.

- **Découvrabilité** : réussir à trouver des données pertinentes et les bonnes sources. Cette difficulté d'accès a été exprimée notamment en lien avec la massification des données, la multiplication des portails open data, source de complexité, les jeux de données comme les plateformes étant disparates. Cela engendre des coûts de

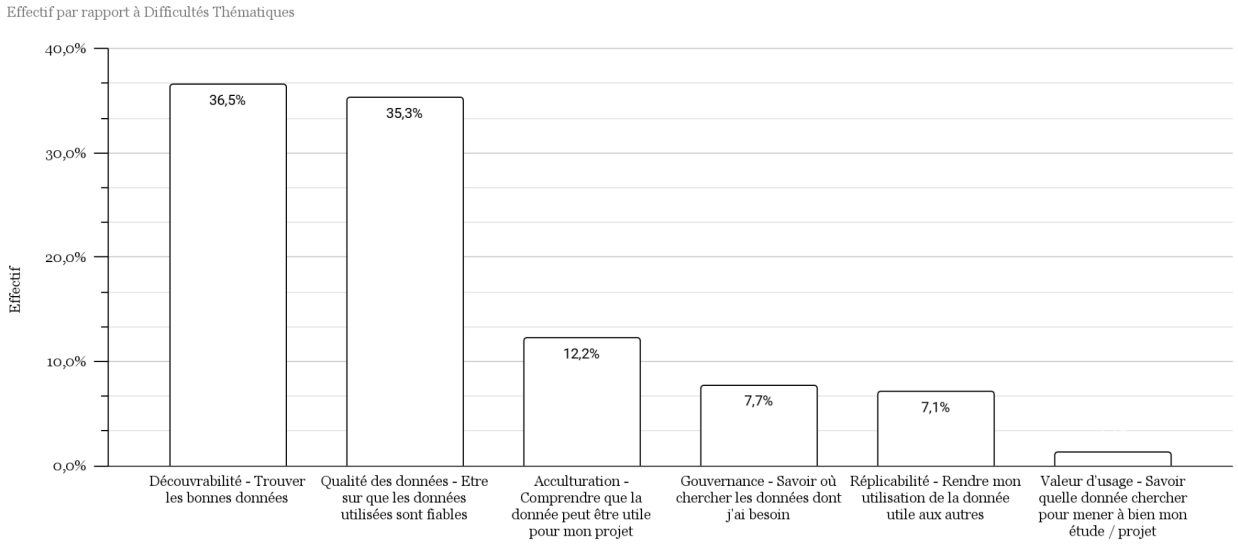
recherche et une dimension chronophage importante qui freine une potentielle réutilisation.

- **Qualité des données** : elle dépend de différents facteurs tels que la qualité de la documentation producteurs, les métadonnées associées et les systèmes automatisés de classification et d'indexation et de standardisation des données. Pour les participants, ouvrir les données signifie d'abord disposer de données homogènes, mises à jour et structurées, qui s'appuient sur une référence, un modèle, un standard partagé.
- **Réplicabilité** : pour les participants, les analyses et les travaux basés sur les données ouvertes, dont la qualité et la fiabilité varient, sont difficiles à reproduire, ce qui limite la confiance dans les conclusions tirées et la capacité à les réutiliser pour leurs décisions.

Afin de valider ce modèle de difficultés, nous l'avons soumise aux participants des ateliers.

Les différentes difficultés remontées par ces derniers correspondaient à cette organisation. Nous n'avons pas identifié de types de difficultés supplémentaires.

Codage thématique ci-dessous



Solutions proposées - Quelles sont leurs attentes et besoins ?

Après avoir identifié les difficultés que rencontrent les utilisateurs, nous nous sommes intéressés aux solutions que les répondants proposent. Nous avons récolté 212 propositions de solutions entre les ateliers et le questionnaire.

Les thématiques adressées par ces solutions ont été regroupées dans les mêmes six catégories de difficultés que nous avons définies.

Aspect important, la problématique de la gouvernance obtient le plus de solutions proposées.

Types de solutions

L'analyse des différentes solutions recueillies par le biais des questionnaires et des ateliers nous a permis d'identifier 11 typologies de solutions différentes.

La recherche augmentée (17,7%)

Cette catégorie regroupe les solutions proposant de simplifier la recherche de la donnée par un moteur de recherche augmenté fonctionnant via des mots clés précis ou une intelligence artificielle comprenant le langage naturel. Cette solution vise à proposer des résultats plus pertinents et adaptés à leurs besoins. Sa faisabilité est déclarée comme complexe au vu des développements nécessaires et de la nécessité de proposer des métadonnées complètes.

Exemples de solutions proposées

- *Avoir un moteur de recherche à la Google*
- *Recherche sur les mots clés, les thématiques, le producteur et le titre*
- *Un bon moteur de recherche (Chat GPT)*

La communauté (15,6%)

Se faire aider par un tiers le plus souvent expert de son sujet est à la fois une des pratiques, mais également une solution. Cette aide peut être apportée par une personne ou plusieurs à la fois. L'une des solutions approfondie en atelier est celle du steward qui *"connait le catalogue , l'utilisation des données et leur réutilisation pour orienter et conseiller les utilisateurs. Il accompagne et coach les producteurs pour la mise en œuvre des bonnes pratiques, spécification, documentation et standard."*

Exemples de solutions proposées

- *Avoir un responsable national qui recueille les problèmes, les partage sur internet et les résout*
- *Permettre à tous les utilisateurs de pouvoir compléter les métadonnées non disponibles dans un jeu de données et pouvoir rajouter des liens/infos sur l'origine du jeu de données s'il s'agit d'une simple "extraction" : édition collaborative*

- *Avoir du réseau ou de l'accompagnement*

Le guichet unique (15%)

Très en lien avec la recherche augmentée, les solutions de guichets uniques visent à centraliser les accès aux données via un seul et même portail. L'objectif est principalement d'éviter l'utilisation de différents outils. Data.Gouv est ciblée comme plateforme d'atterrissage de ce guichet. La complexité reste élevée pour les mêmes raisons de qualités des métadonnées.

Exemples de solutions

- *Une seule et même plateforme qui propose toutes les données (data.gouv)*
- *Arrêter le foisonnement des plates-formes de diffusion ou les obliger au moissonnage par une unique plateforme nationale*
- *Un site qui centralise toutes les bases de données avec dates de mise à jour*

La recherche par cas d'usage (11,6%)

Propositions de solutions semblables à celle d'Ecosphères, la recherche par cas d'usages propose un niveau de résultats simplifiés permettant de créer du lien entre les données à utiliser pour un même besoin. Les liens avec la notion de communauté et de valorisation sont assez étroits puisque ces cas d'usages proviennent pour certains d'utilisations passées. Certaines solutions mentionnent clairement l'utilisation des politiques publiques comme cas d'usages.

Exemple de solutions

- *Faciliter l'exploration des données par thématiques ou politique publique*
- *Le suivi d'une politique publique peut s'effectuer par l'analyse d'indicateurs clés*
- *Un moteur de recherche basé sur les objectifs de notre action, projet etc.*

Observation

Pourquoi les répondants ne font pas le lien dans la question des usages des données et la font-ils dans les solutions ?

L'hypothèse la plus probable de ce décalage est le décalage de niveaux de pensée dans les deux questions.

La première question "quels usages faites-vous des données et à quelles fins" obtient majoritairement des réponses concernant les livrables produits par la données : cartographies et études.

Lors de l'identification des solutions la question qui est posée est en réponse aux difficultés précédemment citées et intervient suite à un enchaînement de questions qui vient faire prendre du recul aux répondants. *Avez-vous d'autres idées qui pourraient simplifier la recherche et l'utilisation de données ?*

S'ajoute également notre incapacité à inférer une implication plus forte des politiques publiques dans les usages des répondants. Comme évoqué plus haut, nous ne sommes pas parvenus à identifier la finalité de ces livrables, car cela ne serait pas scientifiquement valable.

La valorisation (10,9%)

Dans une idée assez proche, la valorisation fait référence à donner à voir l'utilisation de données dans des cas de réutilisations précis. Ces solutions insistent davantage sur l'idée de partager son résultat sans forcément standardiser le travail mais plus de l'ouvrir dans la logique de transparence. Facile à mettre en œuvre, son exploitation par des tiers peut nécessiter de la modération.

Exemples de solutions

- *Publier des illustrations de réutilisation de la donnée*
- *Accéder à des réutilisation et des résultats validés*
- *Partager son utilisation via divers média*

L'indice de qualité (7,5%)

Dans la continuité de l'indice de qualité des métadonnées présent sur data.gouv, la solution proposée consiste à mettre en place un indicateur de qualité des données pour illustrer le niveau de fiabilité d'une donnée. La notation par une communauté ou un expert reste en débat, remettre cette responsabilité sur le producteur peut être problématique et dépendre de l'usage.

Exemples de solutions

- *Indicateur de qualité de la donnée : métadonnées complétées, actualisation fréquente, bonne précision spatiale*
- *Attribuer des indicateurs de qualité d'actualisation et de nombre d'utilisateurs (notation)*
- *Contrôle qualité*

La standardisation (6,8%)

Destinée à la fois à la conception des jeux de données, mais également à la métadonnée, cette solution vise à simplifier la recherche et l'utilisation via la définition de standards à respecter dans la conception et la diffusion d'un jeu de données.

Exemples de solutions

- *Définir par type de producteur un socle de données à diffuser conformément à un standard donné et dont les métadonnées seraient collectées dans un catalogue unique type data.gouv*
- *Utilisation de standard de données et métadonnées associées*
- *Exploiter les métadonnées, les standards et des données précises*

La documentation (5,4%)

Pareillement liée à la notion de standard, la documentation accompagnant les jeux de données est une piste d'amélioration proposée par les répondants. La documentation peut concerner aussi bien l'utilisation de la donnée que sa provenance.

Exemples de solutions

- *Associer à chaque jeu de données une documentation détaillée qui indique la manière dont ont été recueillies et retraitées les données, qui détaille chacune des colonnes (si l'on parle de données tabulaires), qui indique le périmètre temporel et géographique du jeu de données, l'année géographie utilisée (si l'on parle de données territoriales)...*
- *Avoir un descriptif des métadonnées bien renseignés*
- *Documenter la réutilisation*

L’appel à commun (4,1%)

Appuyé sur la notion de communauté, l'appel à commun regroupe des solutions proposant d’associer les ressources et moyens sur des objectifs partagés. Un exemple concret est de mettre en commun les appels à projets pour travailler à construire des exploitations communes de la donnée. Dans le cas d’Écosphères, cette notion d’appel à commun pour la réalisation de bouquets peut être une fonctionnalité intéressante à creuser.

Exemples de solutions

- *Coordonner toutes les bonnes initiatives qui existent dans la sphère publique et leur donner une cohérence globale pour faciliter l'accès aux usagers*
- *Passer de la réplicabilité à la massification / mutualisation des solutions entre territoires*

Les compétences (2,7%)

Les participants ont exprimé le besoin que davantage d'individus disposent de compétences dans l’utilisation des données. Notamment en formant des personnes autres que les CDO et la DSI sur son intérêt, les normes et protocoles associés et les enjeux de sécurité et d'éthique pour la protection des données sensibles. Pour remédier à cela, il apparait également nécessaire d’outiller les agents et de s’appuyer sur des cas d’usage associés aux enjeux nationaux de transition écologique qui pourraient permettre de convaincre différents acteurs sur l’utilité concrète de la donnée.

Exemples de solutions :

- *Formation et communication sur la donnée*
- *Ambassadeurs data dans chaque direction métier*
- *Créer des liens avec les savoirs académiques via les sciences participatives*

La transparence de la production (2,7%)

Améliorer la visibilité sur les modalités de production de la donnée et simplifier le lien avec les producteurs. Ceci afin d’identifier vers qui se tourner en cas de problème, mais également pour responsabiliser les producteurs de données en les rendant plus visibles.

Exemples de solutions

- *Rendre les 'chaînes de production' lisibles (ex. données produites localement > consolidées ou validées par X > transmise à un niveau régional ou autre (ex. district hydrographique) puis bancarisé au niveau national*
- *Identifier clairement les producteurs de données et les services de référence*
- *Expliquer le mode de production des données et leur limites via des métadonnées*

Autres solutions pertinentes

En parallèle de cette analyse, certaines propositions de solutions en particulier ont suscité notre intérêt. Elles sont pour la plupart peu voire pas développées, mais peuvent représenter des évolutions intéressantes pour Ecosphères.

DOI.ser les jeux de données (créer des numéros de série pour les jeux)

Le DOI (Digital Object Identifier) est un identifiant unique pour les ressources numériques (articles de périodique, livres, chapitres de livre, etc.). Il permet de repérer la version complète d'un document de différentes façons.

Afin de faciliter l'accès et la traçabilité, le fait de créer des numéros de série par jeux ou bouquets de données peut améliorer considérablement leur exploitation.

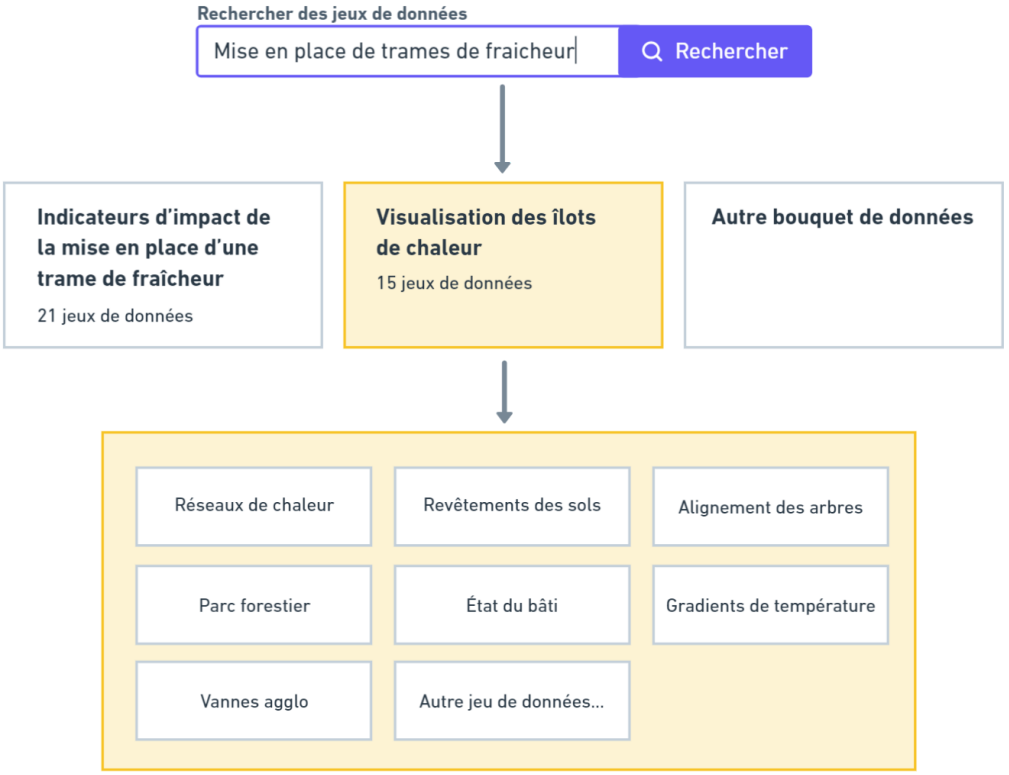
3.4. Retour sur l'hypothèse - Est-ce que la nouvelle entrée par cas d'usage est adaptée ?

Afin d'évaluer la valeur ajoutée d'Écosphères et sans avoir d'éléments concrets à faire tester aux participants, nous avons fait le choix d'illustrer le fonctionnement imaginé au travers d'une hypothèse.

Concrètement la question prenait la forme suivante :

> *Afin de faciliter la recherche de données, nous souhaitons étudier la mise en place d'un nouveau mode de recherche des données.*
>
> *Des travaux sont en cours pour définir des standards de métadonnées afin de faciliter leur utilisation.*
>
> *En parallèle, nous cherchons à améliorer la recherche de données et étudions la mise en place d'un nouveau mode de recherche.*

Rechercher les données par politiques publiques simplifie la recherche de données et leur utilisation.



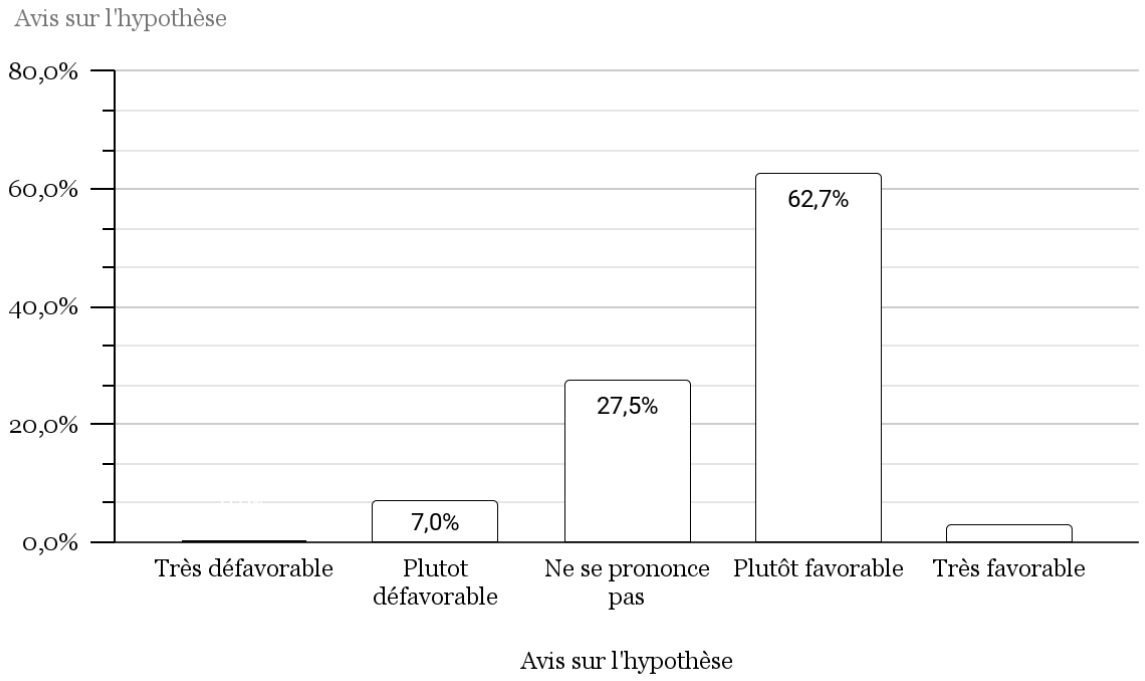
Suite à cette présentation, nous demandions aux répondants, ce qu'ils pensaient d'une telle hypothèse, les avantages et les inconvénients qu'ils identifient concernant cette dernière et s'ils pensaient être capables de constituer eux-mêmes des bouquets de données (si non, pourquoi).

Avis sur l'hypothèse

Ces résultats sont issus d'un recodage des réponses des participants. Comme nous le voyons, la majorité des répondants considère positivement l'hypothèse d'Écosphères. Il est important de prendre en compte le biais de désirabilité sociale qu'entraîne ce type de question. L'écart entre très favorable et favorable s'explique en partie par l'interrogation des répondants sur la faisabilité de l'hypothèse. Les résultats suivants viennent illustrer cet aspect.

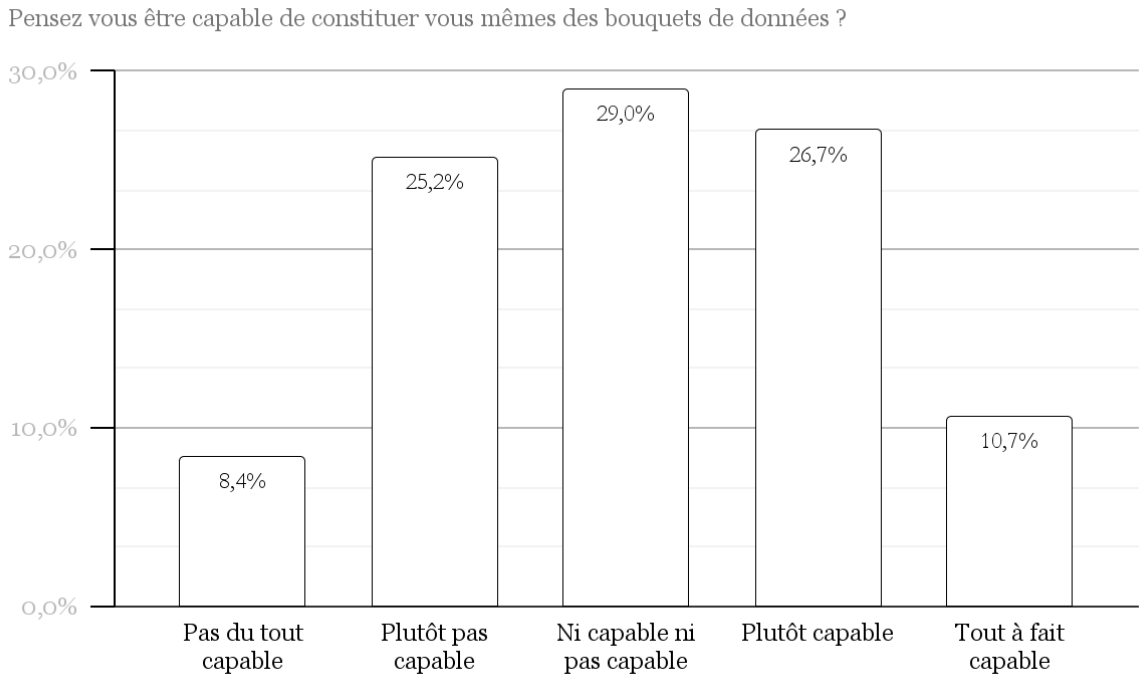
Désirabilité sociale

Le biais de désirabilité sociale est défini comme la tendance qu'ont les individus à vouloir se présenter sous un jour favorable (Hays, 1989) et à ne pas reporter des informations qui les feraient mal voir (Fisher, 1993).



Capacité perçue

Les résultats ci-dessous viennent montrer un manque de positionnement sur leur capacité à eux-mêmes constituer des jeux de données. Une fois de plus cette partie des résultats tend à nous montrer un flou pour les utilisateurs dans la création de lien entre jeux de données et politiques publiques.



Inconvénients de l’hypothèse

L’analyse de différentes questions nous a permis d’identifier une liste de 26 difficultés que les répondants identifient dans la réalisation et l’utilisation d’Écosphères.

Comme nous l’avions mentionné plus haut, c'est principalement la dimension de lien entre la politique publique et les données qui est mentionnée comme difficulté de mise en place de l’hypothèse.. La capacité à récupérer toutes les données utiles, la mise en œuvre concrète du lien ou encore les modalités de recherches sont les aspects perçus comme les plus problématiques pour les répondants.

Problèmes de l'hypothèse	**Effectifs**
Comment être sûr de récupérer toutes les données utiles ?	30
Comment faire le lien entre politique publique et donnée ?	28
Est-ce qu'il faut par conséquent utiliser des mots clés précis ?	19
Est-ce que c'est plus simple pour les utilisateurs en dehors des administrations qui ne connaissent pas les politiques publiques ?	18
Est-ce que ça ne risque pas de limiter la recherche ou la création d'indices basées sur plusieurs bases ?	14
Comment maintenir à jour les jeux de données et les noms des politiques publiques ?	14

Retours de l’équipe

Afin d’apporter une réponse aux différentes difficultés remontées par les répondants, nous avons conduit un entretien avec un membre de l’équipe projet Écosphères. Cet entretien a consisté en une présentation des résultats ci-dessus et une prise de notes des éléments de réponses apportés par l’équipe.

De cet entretien ressortent les éléments venant répondre tout ou partie aux interrogations identifiées dans l'étude quantitative. L'analyse de cet entretien nous a également permis de définir des principes forts d'Écosphères qui seront utiles à l'équipe lors des prochaines présentations ou dans l'évolution du projet. Ces principes sont affichés sous chacune des réponses ci-dessous.

Comment être sûr de récupérer toutes les données utiles ?

Écosphères ne conçoit pas la notion de bouquet comme clos à un instant, les bouquets évoluent et sont alimentés par une communauté.

◎ Écosphères propose des bouquets qui évoluent constamment.

Comment faire le lien entre politique publique et donnée ?

Ce lien peut se faire via différents moyens et mobilise différentes personnes. L'équipe envisage dans un premier temps de mobiliser des experts d'une politique publique pour constituer les premiers bouquets. De plus, certains bouquets peuvent être construits sur la base d'éléments présents dans des textes de lois, ou dans des études déjà réalisées. Une fois de plus, la logique communautaire de mise en commun devrait permettre de répondre.

◎ Écosphères met à disposition des bouquets de jeux de données conçus par des experts. (mais pas que)

Est ce qu'il faut par conséquent utiliser des mots clés précis ?

Non et d'ailleurs le terme de recherche par politique publique ne couvre pas vraiment l'idée complète, puisqu'il s'agit davantage de recherche par "cas d'usages".

◎ Écosphères permet une recherche par cas d'usages liés à des politiques publiques.

Est ce que ça ne risque pas de limiter la recherche ou la création d'indices basés sur plusieurs bases ?

La proposition de simplification de la recherche d'Ecosphères ne veut pas dire que la recherche classique de données par thématiques, producteurs, ou encore zone géographique ne sera pas possible. La plateforme proposera également un moteur de recherche "classique".

◎ Écosphères propose également un mode de recherche classique de données.

Comment maintenir à jour les jeux de données et le nom des politiques publiques ?

La mise à jour et, de façon plus large, l'évolution des bouquets sera réalisée par la communauté d'utilisateurs. Autre point important : Ecosphères envisage de ne centraliser que quelques centaines de bouquets au maximum. Les bouquets seront évolutifs, mais peu nombreux.

◎ Écosphères souhaite confier la mise à jour des bouquets à la communauté des utilisateurs.

Est ce que l'on parle de donnée d'élaboration, de suivi ou de mise en œuvre de la politique publique ?

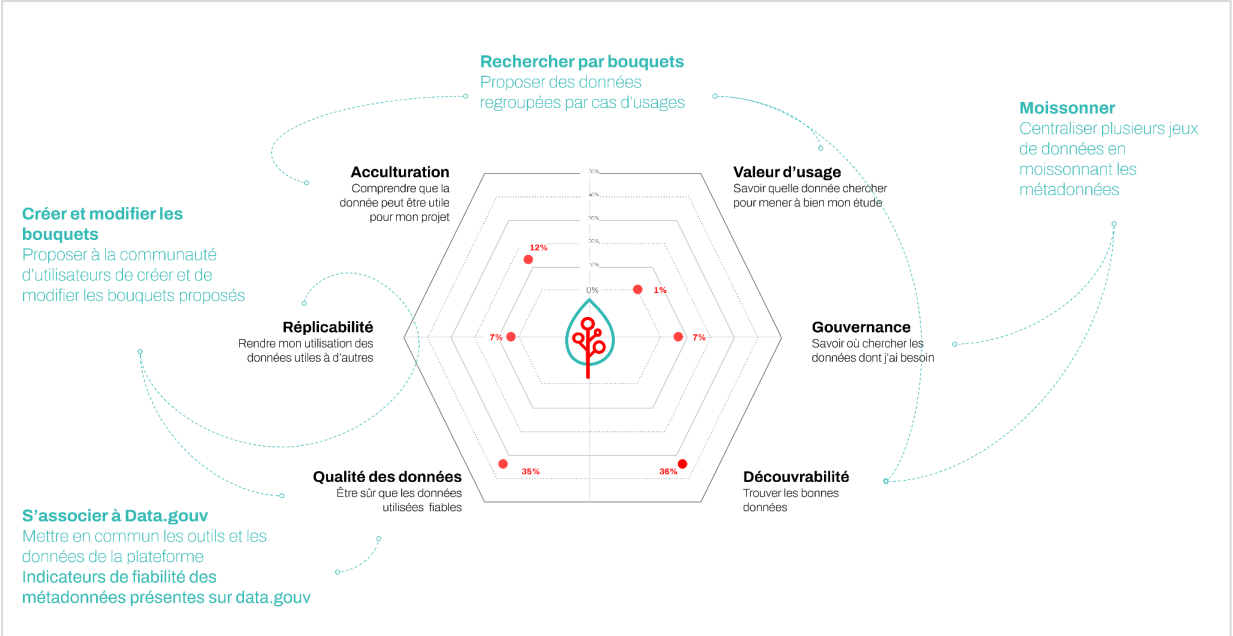
L'objectif est d'adresser plusieurs orientations pour différents bouquets comme "suivre la qualité de l'air", "comprendre les îlots de chaleur", "sensibiliser à la nuisance sonore" ou encore mettre en place un cas d'usage. Il ne s'agit encore que d'idées qui sont en train d'être travaillées via les entretiens experts.

◎ Écosphères souhaite proposer des bouquets orientés sur des cas d'usages spécifiques (étudier, mettre en place, etc...)

Les autres questions sont disponibles dans les annexes remises avec le bilan de l'étude, et nous permettent d'illustrer certains aspects d'écosphères utiles à sa définition.

3.5. Constitution de la mesure d'impact

Évoquée en début de projet comme un souhait de sujet à investir par l'équipe projet, la mesure d'impact a été travaillée dès le lancement de l'intervention. Nous avons réalisé une première analyse du projet, puis sollicité l'équipe pour un entretien visant à identifier les cibles pressenties, les difficultés que celles-ci rencontrent, les fonctionnalités d'Écosphères et les changements que sont censés apporter ces fonctionnalités. La figure ci-dessous résume comment les fonctionnalités du service ont été imaginées pour répondre aux principales difficultés des utilisateurs.



Difficultés et fonctionnalités

Ce premier entretien a permis de faire une première ébauche de ces éléments. Suite à notre étude, il est maintenant temps de la compléter. Des difficultés identifiées et des propositions de l'équipe, nous pouvons établir la matrice suivante

Difficultés	Fonctionnalités
Savoir où chercher les données dont j'ai besoin	Guichet unique
Trouver les bonnes données	Communauté qui identifie les données utiles par bouquet
Être sûr que les données utilisées sont fiables	/
Rendre mon utilisation de la donnée utile aux autres	Créer / modifier des bouquets
Comprendre que la donnée peut être utile pour mon projet	Avoir un aperçu de l'ensemble des données propres à un cas d'usages et une explication de l'utilité de chaque donnée dans le bouquet
Savoir quelle donnée chercher pour mener à bien mon étude / projet	Accéder à des données par cas d'usages

Difficultés, fonctionnalités, changements et indicateurs

La suite de la définition de la mesure d'impact prévoit de définir les changements que les fonctionnalités sont censées apporter.

À l'heure actuel le travail sur la construction de la mesure d'impact en reste à la définition des changements attendus.Pour aller plus loin il sera nécessaire d'avancer sur l'outil pour définir et positionner les indicateurs pertinents

Difficultés	Fonctionnalités	Changements	Indicateurs	Outils
Acculturation Comprendre que la donnée peut être utile pour mon projet	**Bouquets** Proposer des données regroupé par cas d'usages	Faciliter la mobilisation des données ouvertes pour un projet		
Valeur d'usage Savoir quelle donnée chercher pour mener à bien mon étude	**Rechercher par bouquets** Proposer des données regroupé par cas d'usages	Faciliter l''identification des jeux données à mobiliser pour mon projet		
Gouvernance Savoir où chercher les données dont j'ai besoin	**Moissonner** Centraliser plusieurs jeux de donnée en moissonnant les métadonnées	Faciliter l'identification de la plateforme à utiliser pour sa recherche		
Découvrabilité Trouver les bonnes données	**Rechercher par bouquets** Proposer des données regroupé par cas d'usages **Moissonner** Centraliser plusieurs jeux de donnée en moissonnant les métadonnées	Faciliter l'accès aux jeux de données à utiliser pour mon projet		
Qualité des données Être sûr que les données utilisées sont fiables	**Créer et modifier les bouquets** Proposer à la communauté d'utilisateurs de créer et de modifier les bouquets proposer **S'associer à Data.gouv** Mettre en commun les outils et les données de la plateforme Indicateurs de fiabilité présent sur data.gouv	Simplifier l'accès à de la donnée fiable		
Réplicabilité Rendre mon utilisation des données utiles à d'autres	**Créer et modifier les bouquets** Proposer à la communauté d'utilisateurs de créer et de modifier les bouquets proposer	Faciliter la mobilisation des données ouvertes pour un projet		

Points de vigilance et limites de l’étude

Suite à cette étude, nous identifions plusieurs points de vigilance pouvant impacter la réussite du projet.

L’importance de la communauté

L’équipe d’Écosphères conçoit la valeur ajoutée de la plateforme principalement sur les apports de sa communauté. Les utilisateurs auront de nombreux rôles :

- Chercher / utiliser des bouquets
- Créer des bouquets
    - Créer des bouquets socles
    - Créer des bouquets complémentaires
    - Documenter des bouquets
- Modifier / Mise à jour des bouquets
- Noter des bouquets

Ces actions sont nombreuses et augmentent la complexité de création de l’outil, mais également sa dépendance à sa communauté. Notre recommandation serait dans un premier temps de se concentrer sur la production de bouquets auprès d’experts qui seraient mis à disposition des utilisateurs. Puis de déployer la logique de communauté dans un second temps. Cela permettra d’éprouver la création de bouquets pour ensuite en imaginer l’interface utilisateur.

Le public cible

L’étude que nous venons de réaliser a majoritairement touché un public de producteur de données. Or l’outil s’adresse plutôt à un public de consommateur de données. Afin de poursuivre le travail de conception, il serait utile de travailler plus étroitement avec des personnes consommatrices de données ouvertes pour tester les fonctionnalités imaginées. De plus, se rapprocher de ces publics permettrait d’estimer le nombre de personnes concernées ou auprès de qui communiquer pour sensibiliser à l’utilisation de la donnée.

Les bouquets de données orientés

L'idée de l'équipe de proposer des bouquets de données orientés (comprendre, analyser, suivre) est un sujet à creuser du point de vue de la faisabilité. Les différences entre les bouquets peuvent être difficiles à définir, voire ne pas avoir d'utilité. La conception de l'interface et le développement des éléments nécessaires à cette fonctionnalité peuvent être nombreux et complexes à créer.

Conclusion

Dans le cadre de cette étude, notre équipe a entrepris d'interroger de nombreux réutilisateurs de données ouvertes travaillant dans des domaines métiers portés par le pôle ministériel.

L'objectif était de mieux connaître ces usagers, les formes d'usages des données, les difficultés qu'ils rencontrent et leurs besoins afin d'évaluer si le développement de l'outil Ecosphères répond de manière pertinente aux défis de la donnée auxquels ils sont confrontés quotidiennement.

Les principaux freins à l'exploitation des données ouvertes rencontrées par les participants interrogés peuvent être regroupés autour de six grandes thématiques : l'acculturation, la valeur d'usage, la gouvernance, la découvrabilité, la qualité et la réplicabilité des données.

Les fonctionnalités de l'outil Ecosphères semblent en phase avec ces difficultés et la majorité des participants interrogés sont favorables à l'hypothèse d'entrée par les usages sur laquelle il repose. Pour poursuivre la démarche, il est désormais essentiel d'identifier la meilleure approche pour traduire cette hypothèse, développer un prototype pour la tester opérationnellement et questionner l'utilisabilité pratique de la plateforme en observant les publics cibles dans un contexte réel d'utilisation. Cette seconde phase pourrait également s'accompagner d'entretiens d'experts susceptibles de nous aider à identifier les bouquets de données à créer en priorité.

Annexes

Liste des inconvénients de l'hypothèse

Problèmes de l'hypothèse	**Effectifs**
Comment être sûr de récupérer toutes les données utiles ?	30
Comment faire le lien entre politique publique et donnée ?	28
Est ce qu'il faut par conséquent utiliser des mots clés précis ?	19
Est-ce que c'est plus simple pour les utilisateurs en dehors des administrations qui ne connaissent pas les politiques publiques ?	18
Est ce que ça ne risque pas de limiter la recherche ou la création d'indices basés sur plusieurs bases ?	14
Comment maintenir à jour les jeux de données et le nom des politiques publiques ?	14
Est ce que ça représente du travail en plus de faire les liens ?	13
Comment ça marche pour les données qui concernent plusieurs politiques publiques ?	11
Comment être sûr que la métadonnée est renseignée ?	11
Comment connaître toutes les politiques publiques ?	10

Comment prendre en compte les données spécifiques d'un territoire ?	10
Est ce que cela ne va pas récupérer trop de données ?	8
Est-ce utile pour les professionnels ?	5
Est ce que cela permet de voir le cycle de production de la donnée ?	4
Est ce que ca n'existe pas déjà ?	4
Est ce que cela n'ajoute pas une arborescence de plus ?	4
Est ce que c'est facile à utiliser ?	2
Quels sont les liens avec Data Gouv ?	1
Est ce que ça adresse le problème de la multiplicité des catalogues ?	1
Est ce que l'on parle de donnée d'élaboration, de suivi ou de mise en œuvre de la politique publique ?	1
Est ce que l'on comprend ce que l'on trouve ?	1
Est ce qu'il y a un risque de s'éloigner de la recherche initiale ?	1
Est ce que cela adresse le problème de gouvernance et d'administration des données ?	1
Comment gérer le nombre de producteurs ?	1

Liste des principes de fonctionnement d'Écosphères

- Écosphères propose des bouquets qui évoluent constamment.
- Écosphères met à disposition des bouquets de jeux de données conçus par des experts. (mais pas que)
- Écosphères permet une recherche par cas d'usage lié à des politiques publiques.
- Écosphères propose également un mode de recherche classique de données.
- Écosphères souhaite confier la mise à jour des bouquets à la communauté des utilisateurs.
- Écosphères souhaite proposer des bouquets orientés sur des cas d'usages spécifiques (étudier, mettre en place, etc...)
- Écosphères propose de créer des bouquets sur la base de métadonnées précises à compléter.
- Écosphères souhaite créer des bouquets de données socles qui peuvent être complété par des bouquets spécifiques adaptés à certains territoires.
- Écosphères va proposer des bouquets à l'utilisation fortement documentés.
- Écosphères ne proposera que quelques centaines de bouquets de données.
- Écosphères moissonne des métadonnées et non des données.
- Écosphères s'adresse principalement aux agents des administrations publiques.
- Écosphères s'associe à DATA Gouv pour récupérer des données et certaines fonctionnalités.