

**Données environnementales
des organismes de recherche :
quelles valorisations
économiques ?**





 **AllEnvi et
AllEnvi Solutions**

AllEnvi : définition, thèmes, missions

Raison d'être

Fédère les forces de la recherche publique pour programmer et coordonner la recherche environnementale française



- Alimentation
- Climat
- Eau
- Territoires

Missions

- Animer la communauté
- Participer aux temps forts internationaux
- Orienter l'action des décideurs
- Programmer la recherche
- Coordonner les infrastructures et la valorisation

AllEnvi Solutions



AllEnvi Solutions est au cœur d'un réseau de 20 000 chercheurs spécialistes de l'environnement.



Une expertise croisée intelligence économique & environnement : des consultants seniors, de formation scientifique & économique.



À l'interface entre recherche et entreprises, dédiée aux grands enjeux d'innovation durable.

Les grands enjeux d'innovation durable



Agriculture urbaine
Qualité de l'air # Végétalisation



Recyclage des métaux critiques
Réutilisation des eaux usées



Impact des entreprises
Données environnementales



Air intérieur # Air extérieur
Impact de l'exploitation minière



Chimie biosourcée # Ressources marines
Biologie de synthèse



Dessalement
Réutilisation des eaux usées



Protéines végétales
Agriculture urbaine



AgriTech # Capteurs connectés
Qualité de l'air



Impact du Tourisme
Circuits courts



Services climatiques



Dépollution des sols



Données environnementales



 **Analyse Stratégique Collective**

Analyse Stratégique Collective des valorisations économiques des données environnementales : un objectif commun

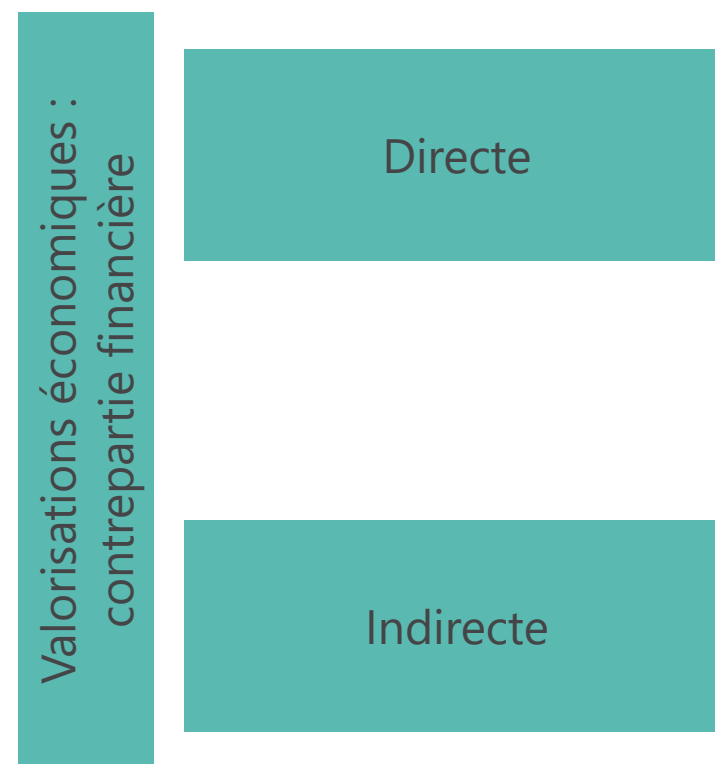
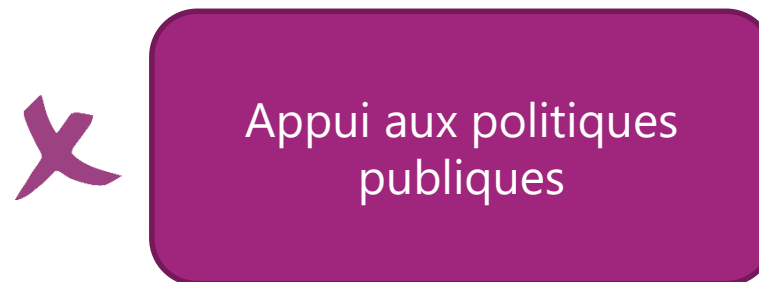
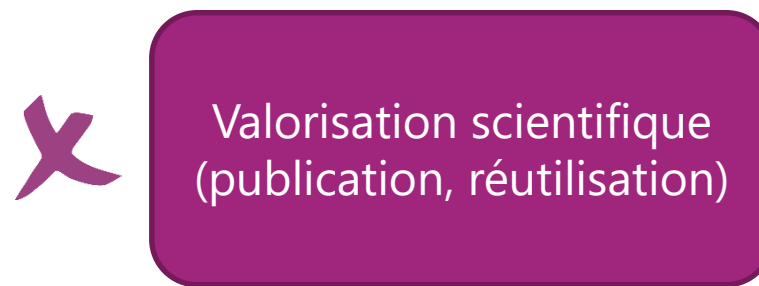
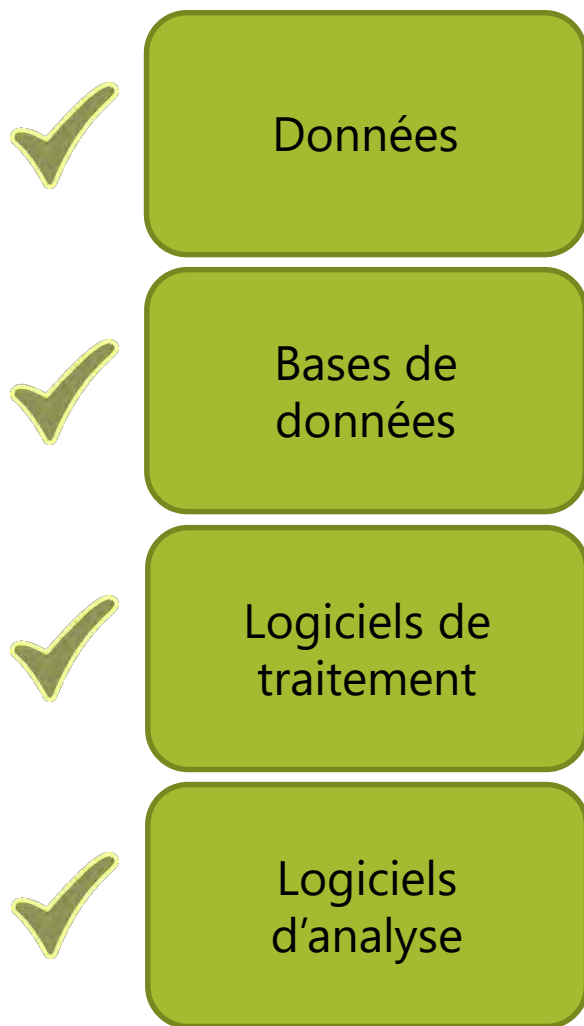
Évaluer la nature et l'ampleur de la valorisation économique des données environnementales produites par la recherche publique

L'Analyse Stratégique Collective (ASCo) vise à

- Partager les politiques de valorisation économique existant au sein des organismes
- Contextualiser le cadre réglementaire régissant l'usage des données publiques et les marchés concernés
- Faire connaître les modèles économiques viables pratiqués
- Contribuer ainsi à la conception de stratégies de valorisation ambitieuses



Périmètres de l'ASCo



Les données sont organisées dans des bases de données constituées *ad hoc*. Des logiciels permettent de les décrire, de les stocker, d'y accéder et de les traiter. Les données ne présenteraient que peu d'intérêt sans algorithme permettant de les interpréter et d'en extraire des informations utiles.

Cette ASCo concerne **les données, les bases de données, les logiciels de traitement et les logiciels d'analyse**.

Cette ASCo concerne les valorisations économiques de ces objets ; est donc hors périmètre :

- La valorisation scientifique (publication, réutilisation)
- L'utilisation de ces objets en appui aux politiques publiques.

La **valorisation économique** peut faire l'objet d'une contrepartie financière directe (rétribution financière) ou indirecte (mention / citation par exemple) qui participe à montrer l'impact économique des données ou logiciels fournis par la recherche.

Plan

Paysage des données environnementales

Production

- Types, chaîne de valeur, volume et croissance, rôle des pôles de données...

Contexte et tendances réglementaires

Grands enjeux

- Climat, biodiversité, santé, pollution, eau, air, risques naturels et industriels, ressources naturelles (sols, végétaux, minéraux, océan...)
- Données environnementales et santé
- Connexions avec la pression anthropique



Marchés et acteurs

L'économie des données environnementales

- Moyens de production et de transformation / d'enrichissement
- Stock et flux
- Sciences participatives

Acteurs

- Pôles de données
- Producteurs, diffuseurs, utilisateurs, intermédiaires... et leurs stratégies
- Entreprises
- Citoyens
- Géopolitique et souveraineté



Valorisation économique

Exemples emblématiques de valorisations économiques, directes et indirectes

Processus de création de valeur

Que permettent les nouveaux outils numériques ?

- API, Cloud, Traitement des données, Plateformes, Jumeaux numériques, IA...

Compétences nécessaires et « amorçage » de la valorisation

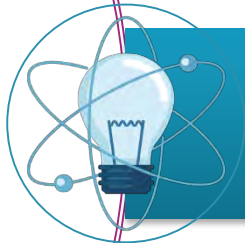


 **Contexte**

Introduction : vocabulaire



Jeux de données, Big data, smart data, open data, small data... ces concepts passent successivement sous l'éclairage médiatique. Les données sont souvent comparées au nouveau pétrole, au sens d'une matière première abondante, riche et transformable industriellement, pour donner des produits / services divers, nombreux, utiles et peu chers.



Les données de la recherche scientifique et académique s'extraitent progressivement de leurs supports traditionnels tels que les articles scientifiques, les rapports, les études et autres publications pour devenir des objets et informations à part entière, qui sont partagés librement sur le Web.



De même, les données publiques sont « ouvertes » (i. e. accessibles ; sous un format exploitable ; dans un délai raisonnable) pour être mises à disposition de tous les acteurs économiques.

Définitions : open data et open science

Open data (Source : Eduscol)

- « Une *donnée ouverte* est une information publique brute, qui a vocation à être librement accessible et réutilisable, [...], sans restriction [...]. »
- Ces jeux de données sont principalement mis à disposition par des organismes publics tels que des gouvernements, ministères, villes, régions, entreprises publiques etc.
- Il peut s'agir du cadastre, de la liste des organismes de formation, du recensement des équipements sportifs ou de l'emplacement des lampadaires... Les thématiques abordées sont larges et très nombreuses.
- L'open data n'est pas un phénomène nouveau et il continue à se développer (i.e. on accède à toujours plus de données), en France et partout dans le monde.

Open science (Source : OCDE)

- Les *données scientifiques*, aussi appelées « données de la recherche » ou research data, sont « des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus comme nécessaires pour valider les résultats de recherche.
- Ce terme ne s'applique pas aux éléments suivants : carnets de laboratoire, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris). »
- Le mouvement de l'open science / science ouverte vise à rendre les données de la recherche accessibles à tous.

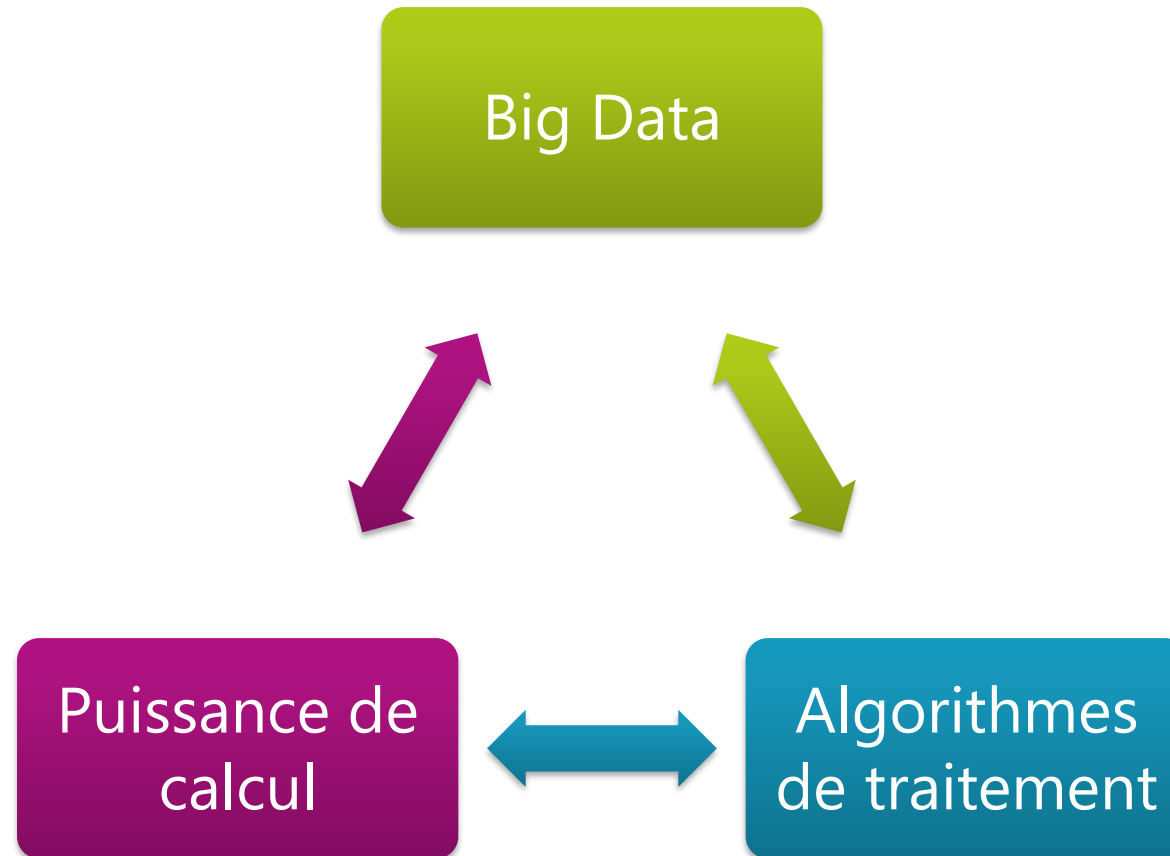
Définition des données environnementales générées par la recherche

Il serait très difficile de définir ce qu'est la « recherche environnementale », autrement dit ce qui en est et ce qui n'en est pas ; toute tentative de définition pourrait passer à côté de nombreuses données tout à fait pertinentes. En conséquence, la présente étude s'intéresse aux **données qui décrivent l'environnement produites dans le cadre de projets de recherche et/ou par les infrastructures de recherche des organismes d'AllEnvi, pour comprendre, prévoir et anticiper.**

Les données environnementales recouvrent toutes les données qui **permettent de caractériser les compartiments de l'environnement** suivant les valeurs de paramètres physiques, chimiques ou biologiques dans leurs évolutions temporelles. Elles sont de natures très variées. Pour l'OCDE il s'agit des données air et climat, biodiversité, déchets, eau, forêts, matériaux ou matière et politiques de l'environnement. Pour l'INERIS, l'inventaire des bases de données environnementales s'inscrit dans le Plan National Santé et Environnement.

Ces données sont produites dans le cadre de projets de recherche et/ou par les infrastructures de recherche des organismes d'AllEnvi. Les données produites par les infrastructures de recherche sont soumises aux orientations des organismes en matière d'open data ; ces orientations sont similaires d'une infrastructure à l'autre.

La transition numérique s'appuie sur 3 piliers





 **Enjeux pour la recherche**

Enjeux économiques pour la recherche

Investissement dans les données

- satellites : 1 G€ / an
- autres infrastructures nationales : investissement 1 G€ ; fonctionnement de 200 M€ /an et 1000 ETP mobilisés.

Au-delà de l'exploitation des données par la recherche, il est **nécessaire d'identifier des valorisations sectorielles (gratuites ou facturées)**. Les **conditions de mise à disposition des données, de leur exploitation et leur valorisation** sont actuellement en débat en raison de plusieurs facteurs internes et externes :

Pression budgétaire sur les infrastructures de recherche

Réglementation : gratuité d'accès aux données publiques

Évolution des pratiques de recherche : données de plus en plus nombreuses (services d'observation, big data) et outils de traitement de données qui évoluent

L'arrivée de **nouveaux acteurs économiques**, notamment issus des outils de communication et de la collecte et l'exploitation de masse de données

Malgré le développement de capteurs de plus en plus performants, l'acquisition de ces données reste une charge qui mobilise une part importante des moyens et des capacités de recherche.

Réduction des budgets : après le PIA, quelle maintenance ?

➔ L'enjeu économique global pour les acteurs de la recherche est de définir des stratégies, en proposant des offres compatibles avec les directives sur les données publiques et des modèles économiques viables.

Source : Sujets Transversaux Groupe Prospective Octobre 2014 ; analyse AllEnvi Solutions

Un cadre réglementaire instaure l'accès gratuit aux données

Concerne spécifiquement la recherche

2021

- [Politique des données, des algorithmes et des codes sources Feuille de route \(2021-2024\)](#)
- Plan National pour la Science Ouverte n°2

2018

- En France, le **Plan national pour la science ouverte** rend obligatoire l'accès ouvert aux publications et – autant que possible – les données issues de recherches financées par appels à projets sur fonds publics.

2017

- Le [programme européen H2020](#) oblige d'assurer le libre accès aux publications issues de la recherche qu'il finance et de partager les données et métadonnées nécessaires à la validation des publications.

2016

- La loi Lemaire, ou **loi Pour une République Numérique**, renforce les obligations d'ouverture des données publiques.

2015

- La **loi Valter** instaure la gratuité pour la réutilisation des informations contenues dans les documents administratifs du secteur public.

2007

- La directive européenne **Inspire** vise à faciliter l'accès à des informations spatiales interopérables utiles aux missions environnementales.

2003

- Le mouvement des données ouvertes est encadré au niveau européen par la **directive 2003/98/CE** (Informations du secteur public).

1998

- La **convention internationale Aarhus** est un accord international de « *démocratie environnementale* » sur l'accès à l'information, la participation du public au processus décisionnel et l'accès à la justice en matière d'environnement.

Le mouvement de l'Open Science est compatible avec la valorisation économique des données

- **Plus de valeur est générée quand les données sont en accès libre**
- Les **données** issues d'**une activité de recherche financée au moins pour moitié** par des dotations de l'État, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union Européenne **doivent être mises à disposition** (code de la recherche, [Art. L. 533-4](#))
- Pas de sanction en cas de manquement mais surveillance instituée
- Des courants idéologiques prônent l'ouverture en temps réel de tous les résultats de la recherche publique, librement et gratuitement sur Internet
- **Les institutions européennes tempèrent** et précisent : The underlying principle for the optimal reuse of research data should be: "**as open as possible, as closed as necessary**".

Une dilemme potentiel, engendré par des injonctions apparemment contradictoires

La « [loi pour une République numérique](#) » vise à donner un accès plus transparent aux administrés aux données les concernant, afin que les citoyens puissent exploiter ces données, pour générer de l'activité économique. Elle modifie marginalement le code de la recherche.

Les règles de publication des informations publiques sont précisées dans le code des relations entre le public et l'administration à [l'article L321-1 du chapitre sur l'étendue du droit de réutilisation des informations publiques](#).

La **valorisation économique** des données permet d'**augmenter la compétitivité des partenaires industriels** des organismes de recherche. Ces deux contraintes induisent un inconfort sur la politique à appliquer par les organismes de recherche, qui par défaut rendent accessibles les données et logiciels qu'ils produisent et en même temps donnent un avantage à des acteurs privés dans le cadre de collaborations public-privé.

L'établissement est concerné par le [dispositif de Protection du Patrimoine Scientifique et Technique de la Nation](#) (PPST) si *la captation induite ou le détournement des savoirs, savoir-faire et technologies développés ou mis en œuvre dans l'établissement peuvent porter préjudice de manière significative à la compétitivité, à celle des partenaires industriels ou à celle du pays.*

* Le PPST est un dispositif français, il ne s'agit pas de la retranscription nationale d'une directive européenne.

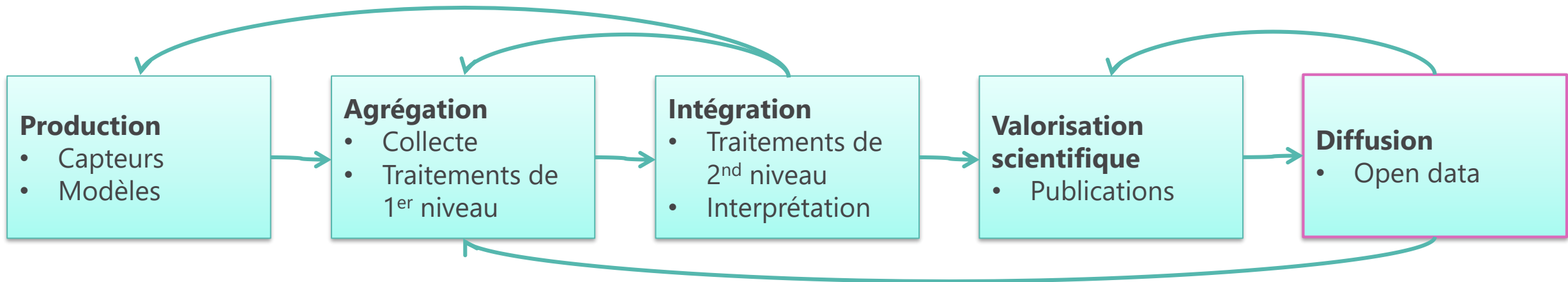
Sources : interview D. Margery, Inria ; Legifrance ; analyse ALLenvi Solutions



- **Chaînes de valeur des données environnementales**

Le processus des projets de recherche s'enrichit d'un maillon

Processus de recherche



Dans le cadre de **projets de recherche**, le processus suit des étapes séquentielles : les données sont générées en fonction du sujet. Puis, les données sont collectées, en intégrant le cas échéant des données issues de science participative ou de partenariats, nettoyées, traitées et interprétées, en faisant intervenir les expertises scientifiques nécessaires. La valorisation scientifique ultime est la communication sous forme de publication.

Dans ce processus d'acquisition de connaissances, l'étape d'intégration peut mettre en évidence la nécessité d'obtenir des données complémentaires ou de raffiner les traitements automatiques sur les données brutes ; les deux flèches issues de l'étape d'intégration illustrent ce fonctionnement.

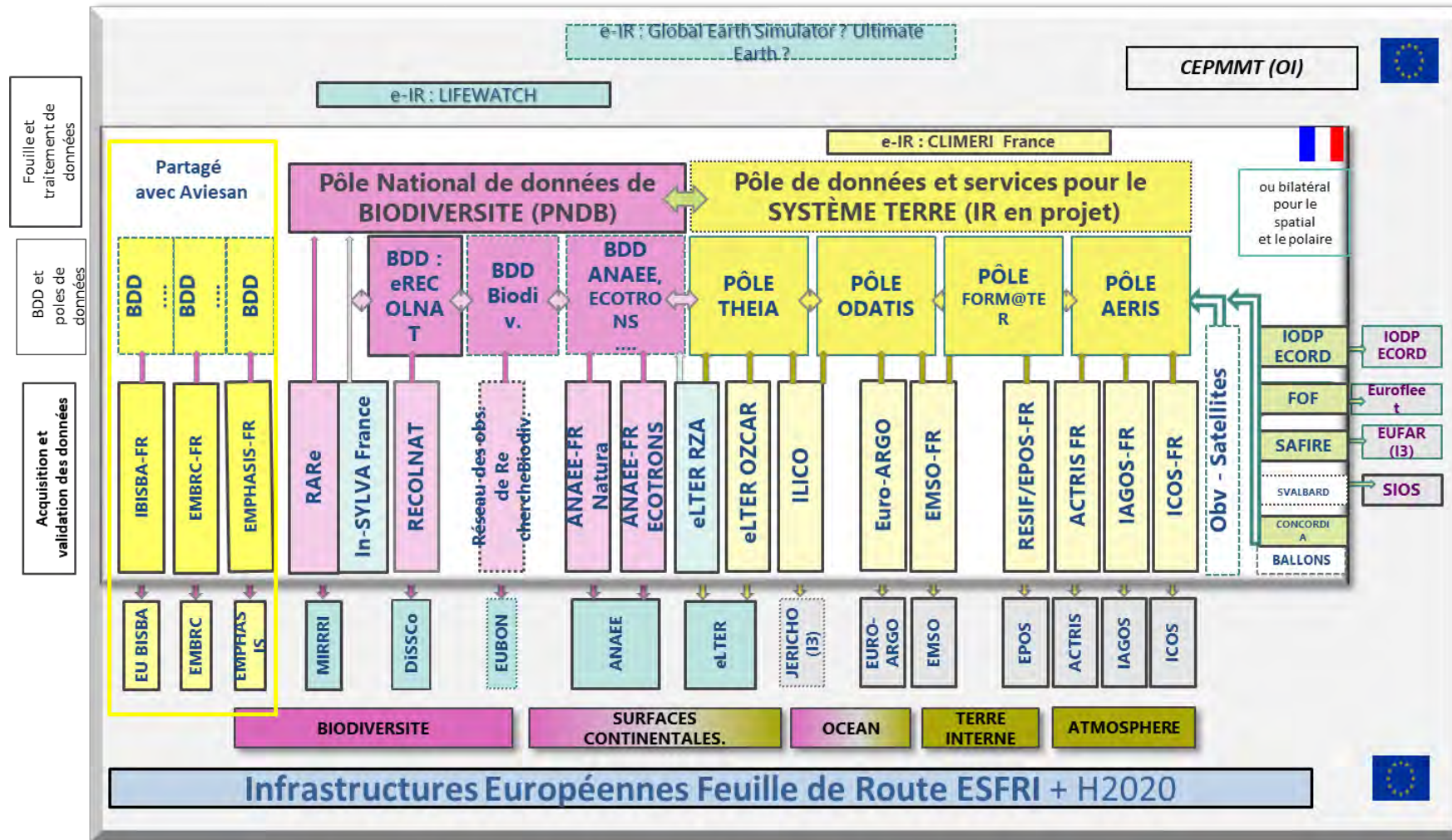
Quand le projet de recherche implique au moins un partenaire privé, la valorisation économique des données peut faire l'objet d'une répartition exclusive entre les acteurs, et seulement une partie des données est diffusée selon les modalités de l'open data.

Dans le cadre **d'infrastructures de recherche**, des données sont produites sur des sujets identifiés d'intérêt majeur (par exemple, la mesure de gaz à effet de serre), alimentant en continu des bases de données. Ces données peuvent être utilisées en tant que matière première pour des travaux de recherche, ou bien être agrégées à d'autres jeux de données. Elles sont aussi systématiquement directement mises à disposition.



- **Paysage des producteurs de données environnementales**

Infrastructures de recherche en environnement



Les infrastructures de recherche sur l'environnement produisent des données d'observation, expérimentation et modélisation, destinées d'abord à comprendre le fonctionnement et l'évolution du système terre et de ses différents compartiments. Il en existe aux échelles locales, nationales et européennes, sous différentes formes (localisée, distribuée, dématérialisée) et plus ou moins avancées, générant des **canaux multiples et imbriqués d'acquisition et de diffusion de données**. Pour la période 2019-2021, les infrastructures de la recherche publique française du domaine « Environnement et Système Terre » sont organisées selon la carte ci-contre.

Ce schéma global de l'organisation de la production de données environnementales par la recherche publique française peut être croisé selon trois critères : discipline scientifique – objet de l'institut – statut de l'institut de recherche. Les infrastructures qui hébergent ces données sont destinées d'abord à la réutilisation et au partage des données à des fins scientifiques, dans la démarche d'open science.

© E. Vergès

Source : MESRI

« Terre Vivante »

« Terre solide et fluide »

Quoi

- L'infrastructure de recherche Data Terra regroupe les données d'observation de la Terre, générées par la recherche française, pour apporter une réponse commune aux besoins des scientifiques, mais aussi des collectivités territoriales voire des industriels : « *La mission principale de Data Terra est de développer un dispositif global d'accès et de traitement de données, produits et services permettant d'observer, comprendre et prévoir de manière intégrée l'histoire, le fonctionnement et l'évolution du système Terre soumis aux changements globaux.* »

Comment

- « *L'ensemble de ces données multi-sources, produits et services sont accessibles au travers de portails unifiés et cohérents.* »
- Le [projet Gaïa-Data](#) vise à permettre et faciliter la **mise en œuvre de cas d'usage transversaux**, c'est-à-dire impliquant des moyens informatiques et des données hétérogènes. Il s'agit de proposer un accès transparent et continu à l'ensemble des données multi-sources du système Terre.

Sources : interview F. Huynh, G. Maudire ; site web Data Terra ; analyse AllEnvi Solutions

Pour qui

- « *Destiné principalement à la **communauté scientifique, ce dispositif s'adresse également aux acteurs publics et de l'innovation.*** »

Par qui

- *Plus de 160 ETP/an, mobilisant plus de 400 personnes issues des 26 organismes et universités partenaires*

Combien

- *40 M€ en coûts complets (2019)*
- *Plus de 1000 produits et services*
- *Plus de 15 000 utilisateurs*

Quand

- **Inscrite à la feuille de route nationale depuis sa création en 2016**

Sources : interview F. Huynh, G. Maudire ; site web Data Terra ; analyse AllEnvi Solutions

La mise à disposition des données : des politiques en structuration, encore peu appliquées

Les politiques de mise à disposition des données (dont les données environnementales) sont très variables d'un organisme à un autre, d'une plateforme à une autre :

- **maturité de la réflexion** sur la mise à disposition des données plus ou moins avancée
- **ouverture à l'Open data** plus ou moins appliquée
- **modèle économique** en cours de réflexion pour la plupart des infrastructures collaboratives
- **enjeux variables** : raison d'être ou simple produit de la recherche, volonté de monétiser ou non
- **valorisation sectorielle** plus ou moins ancienne et intéressante



AllEnvi pourrait-elle jouer le rôle de **vecteur de structuration** ?

Le modèle économique ? Quel objectif ? Juste payer la maintenance ? Réinvestir ? Pouvoir s'étendre ? Garder une certaine autonomie ? Activer l'innovation des usages ?

A-t-on idée de l'ampleur de coût que cela représentera à l'avenir ?

On pourrait aussi parler de « valeur économique des données » et d' « enjeux politiques ». Souvent c'est l'État qui détermine si l'on doit créer une base de données et comment le faire, et non les seuls organismes. Dès lors, on parlera de missions.

À quel moment il rentre dans les missions des organismes de collecter et stocker certaines données *ad vitam æternam* ?

Source : interviews chercheurs

Source :

Les dynamiques de partage des données de recherche et pour la recherche

- La maturité de la culture du partage des données dépend de la discipline scientifique.
- Tout dépend de la nature, de la quantité et du mode d'acquisition des données.
Exemple : astronomie et géologie vs biodiversité
- Les rapprochements de données de champs disciplinaires différents sur un même objet de recherche, sont particulièrement enrichissants pour la recherche environnementale nécessairement interdisciplinaire
- Les données peuvent désormais être acquises pour d'autres finalités par des services publics, des entreprises, des associations, voire des particuliers
- La multitude des moyens de partage des données transforme la relation science/société
- L'obtention de données de biodiversité repose plus particulièrement sur les sciences participatives

Source : Sujets Transversaux Groupe Prospective Octobre 2014 ; analyse AllEnvi Solutions

La valeur des données dépend de l'usage

La **valeur d'un jeu de données** provient à la fois des **métadonnées**, qui montrent l'organisation des paramètres d'intérêt pour une thématique donnée, et **des données elles-mêmes**.

Les données sont des actifs immatériels, leur valeur provient de leur utilisation. De plus, **la valeur des données provient en partie du flux, pas uniquement du stock**. Les données sont souvent qualifiées de « *nouvel or noir* » ; mais, contrairement au pétrole, les données peuvent se périmer.

Les données n'ont pas toutes de la valeur car **toutes ne sont pas utiles**. Par exemple, connaître la température du sol dans la forêt voisine n'est pas utile quand on connaît déjà la température du sol de son champ. Ou savoir qu'un agriculteur a semé telle quantité de telle variété, à telle heure tel jour, 3 heures avant qu'il ne pleuve 30 mm, quand cette pratique est classique, ordinaire et prévisible, constitue en fait une connaissance qui n'a que peu de valeur, aucun acteur économique ne souhaite payer pour l'acquérir.

En parallèle, **stocker les données a un coût**. La quantité de données produites croît très rapidement, plus rapidement que ne diminue le coût du stockage diminue, si bien que le coût du stockage est un frein à la conservation des données.

Source : Interviews

De très nombreux secteurs économiques exploitent et dépendent des données environnementales

Secteurs économiques exploitant les compartiments du système terre et dépendant des données associées (sous-sol, surfaces continentales, océan, climat, biodiversité) :



- Transport : terrestre, maritime, aérien ; infrastructures



- Eau : approvisionnement, risques d'inondation (pouvoirs publics, collectivités), sécheresses



- Santé-sécurité : qualité de l'air (collectivités, associations...), risques naturels (séismes, éboulements...)



- Assurances



- Appuis aux politiques publiques (bureau d'étude, ingénierie, conseil), aux investisseurs (BTP, mines et carrières, forêts...)



- Tourisme : sports d'hiver – enneigement et sécurisation des sites...



- Agriculture : semenciers, coopératives, négociants...



- Construction : aménageurs, promoteurs immobiliers...



- Énergie : transformation, transport, obtention


Une pré-étude de AllEnvi Solutions s'est intéressée aux besoins des acteurs socio-économiques du secteur agricole : startups, PME, grands groupes, instituts techniques, coopératives, agriculteur, organismes de conseil et développement agricole.

Les sociétés françaises qui travaillent déjà avec les données de la recherche peuvent indiquer des pistes de valorisation

Société	Type de données
Cybeletech	
Predict services	
Geomatys	Images spatiales
VisioTerra	Images spatiales SPOT ; données météorologiques, climatiques, biogéophysiques
AmpliSIM ?	Concentration en polluants atmosphériques, données météo (direction et force des vents, précipitations, température), ... ?
L'Avion Jaune ?	
GeographR ?	
La Telescop ?	
KisanHub	
Antea Group (GeoHyd)	

Travail en cours

- Champs complémentaires :
- Utilisation,
 - Métier,
 - Produits,
 - Marchés,
 - Flux entrants,
 - Technologie propriétaire,
 - Collaborations



- **Modèles économiques
pour les valorisations
économiques**

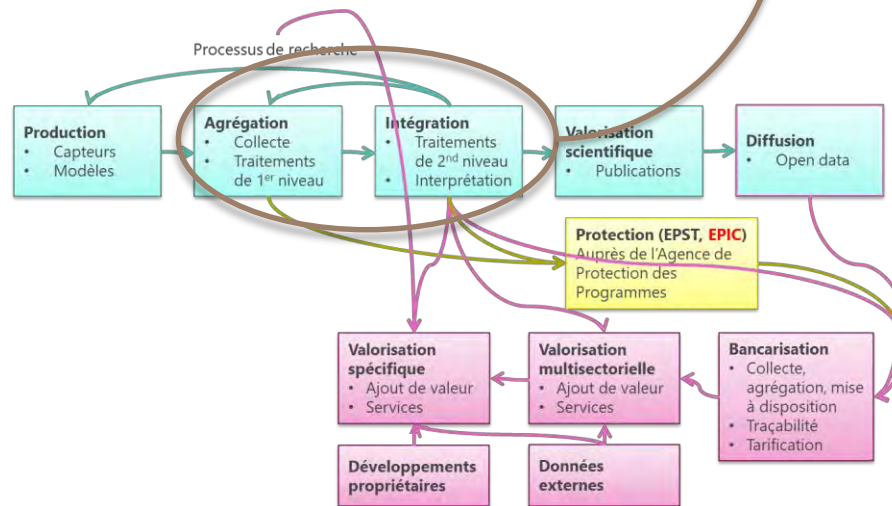
Valoriser de la valeur ajoutée fondée sur les données

Agrégation des données

Pas de recours à l'expertise

Caractéristiques

- la plus évidente à mettre en œuvre
- faible valeur ajoutée
- peu rémunératrice pour les organismes de recherche – mais impact socio-économique + enrichissement scientifique
- risque de position dominante



La vente de données n'est pas possible, car il est estimé que les organisations ont été financées pour les produire et donc que les données relèvent du régime de l'open data.

Ainsi, les données obtenues par les capteurs piézométriques du BRGM sur le niveau des nappes phréatiques permettent des prévisions de sécheresse ou d'inondations, prévisions actuellement vendues à Veolia et à des collectivités territoriales.

- « Évidence » : les données environnementales ne suffisent pas pour répondre aux questions des interlocuteurs, elles ne décrivent « que » l'environnement)
- position de marketplace, intermédiaire → marges sur volumes d'échanges
- Position dominante : incontournable comme Booking.com

Intermédiaires



Pour chaque développement d'une technologie, l'infrastructure précède les applications.

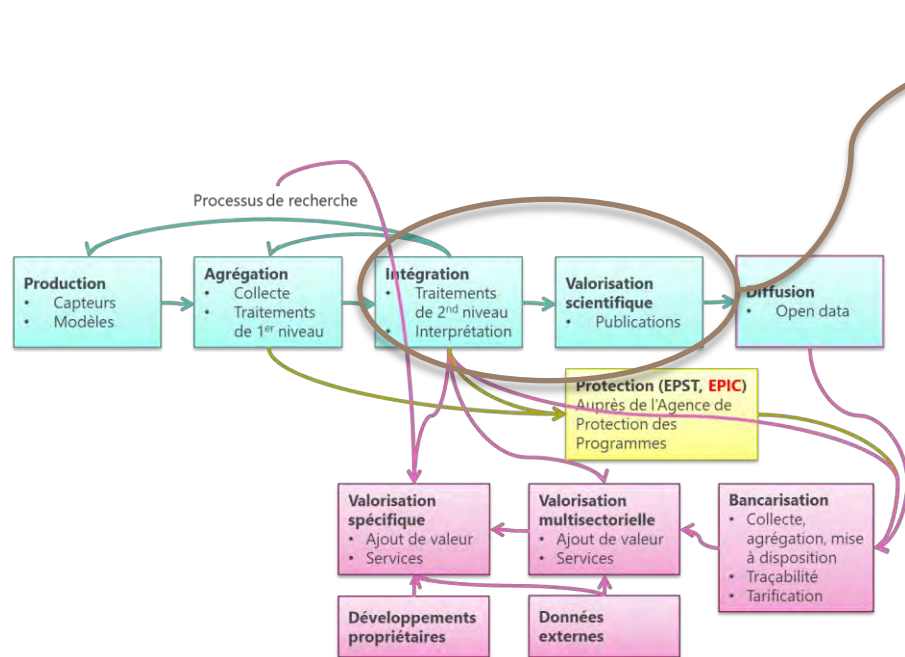
Intermédiaires : entreprises qui développent outils et infrastructures sur lesquels d'autres entreprises bâtissent des solutions.

Offres recensées :

- Une plateforme de cloud où des données semi-structurées sont interrogeables via des requêtes analytiques non formelles ;
- Une offre cloud de machine learning automatisé
- Une solution de préparation de jeux de données

Databricks, DataRobot et Scale AI
progressions de croissance extrêmement élevées + intros en bourse

Vendre de la valeur ajoutée fondée sur les données



- Codes sources et logiciels ne sont plus transférés
- Valorisation économique de la valeur ajoutée obtenue grâce à **l'interprétation et au traitement** des données : **services d'expertise, formation.**
- *IGNfab, CeremaLab, MeteoNet...*
 - hébergent des start-ups pendant plusieurs mois / années.
 - fournissent un accès à leurs jeux de données et aux compétences des chercheurs sur site tandis que les start-ups développent des solutions fondées sur des usages spécifiques des données.



 **Conclusion... à ce stade**

Points saillants



- Enjeux économiques pour la recherche
- Un contexte réglementaire clarifié
- La valorisation peut intervenir à plusieurs étapes de la chaîne de valeur
- Bases de données : une offre large, riche, en phase d'unification (Data Terra + FAIR), partiellement poussée par la contrainte économique, « concurrentielle » par rapport aux pure-players du numérique
- Logiciels : valorisation via expertise, formation
- Des cas d'usage extrêmement variés
- Des modèles économiques pour la valorisation en cours d'examen

Valo logiciels : – piste des consortiums

ASCo pilotée par



François Christiaens, AllEnvi Solutions – francois.christiaens@inrae.fr

Dominique Morin, BRGM – d.morin@brgm.fr