



UNIVERSITÉ PARIS II PANTHÉON - ASSAS

Écosystèmes de données : une exploration des acteurs et des modèles économiques qu'ils pratiquent

*Vers une topologie de référence des modèles de revenus et de tarification dans les
écosystèmes de données*

Mackenzie Himmelbauer

29 août 2021

Université Paris II- Master 2
Management Stratégique et Entrepreneuriat 2020-2021
Mémoire de maîtrise
Visions SAS



Table des matières	
Remerciements	4
Résumé	5
I. Introduction :	6
II. Écosystèmes	7
A. Introduction	7
B. Définition d'un écosystème	8
C. Catégorisation des écosystèmes	8
1. orchestrateurs publics versus orchestrateurs privés	9
2.Écosystèmes centralisés et décentralisés	10
D. Acteurs des écosystèmes	11
1.Contributeurs de données brutes ou non traitées	12
2.Processeurs	12
3.Connecteurs/orchestrateurs	13
4.Prestataires de services	13
E. Quelles formes peuvent prendre les écosystèmes ?	14
1.Place du marché	14
2.Écosystèmes de données privés :	15
3.Échanges de données ouvertes et data lakes	16
F. Visions Ecosystèmes	18
1.Visions Ecosystèmes	19
2.Entreprises privées	19
3.Agences gouvernementales	20
4.Universités et organismes de formation	21
III. Modèles commerciaux : Modèles de revenus et de tarification	21
A.Introduction	21
B.Modèles de revenus	22
1.Fournisseurs de données brutes	23
2.Processeurs de données	25
3.Connecteurs	28
4.Nouveaux fournisseurs de produits et de services	34
5.Conclusion	36
C. Modèles de tarification	36
1.Transactions sur demande unique	37
2.Transactions de paquets de volume	37

3.Transactions temporelles	38
4.Autres modèles de prix observés dans les publications	39
D. Les échanges non monétaires	40
IV. Conclusion	44
A. Conclusion sur les écosystèmes :	44
B. Conclusion Section sur les modèles d'entreprise :	46
C.Où vont les écosystèmes de données ?	47
Références	49

Remerciements

Je tiens à remercier sincèrement tous les participants qui ont consacré du temps, malgré leurs emplois du temps chargés, pour partager avec moi leurs points de vue et leurs connaissances professionnelles sur les écosystèmes de données. Il a été très intéressant d'apprendre de vous sur les différentes industries et les façons dont les écosystèmes de données évoluent.

Je remercie tout particulièrement l'équipe Visions d'avoir consacré du temps et des ressources à ce projet et d'avoir veillé à ce que je dispose de toutes les informations et de tout le soutien nécessaire à la réalisation de ce projet. J'ai hâte de transformer cette recherche en quelque chose d'utile pour les écosystèmes de données à un niveau plus pratique. J'ai hâte de voir comment ces découvertes vont évoluer.

Résumé

Les écosystèmes de données sont un nouveau phénomène dans lequel des données sont échangées entre divers acteurs, tant publics que privés, et à des fins variées. Chaque type d'échange de données s'accompagne d'échanges de valeurs entre les acteurs en échange de ces données. Comme il existe une grande variété d'acteurs dans les écosystèmes de données, les modalités de ces échanges varient considérablement. Il existe de nombreuses recherches sur les modèles économiques en général et un petit nombre de publications sur les modèles économiques dans les marchés de données. Cette thèse de master vise à examiner les modèles économiques (modèles de revenus et modèles de tarification) qui se produisent dans les écosystèmes de données d'une manière plus holistique.

Une stratégie multi-méthodes a inclus une approche de recherche qualitative qui a donné lieu à sept entretiens semi-directifs, à l'analyse de plus de dix publications universitaires et études de cas et à l'examen de plus de trente publications en ligne. L'analyse inductive et l'expérience de l'industrie apportée par les participants ont permis de découvrir d'autres modèles de revenus et de tarification qui n'étaient pas encore exposés dans les publications universitaires.

Cette recherche caractérise les modèles commerciaux dans les écosystèmes de données comme des éléments en constante évolution. Les écosystèmes de données doivent d'abord être présentés et catégorisés, en contenant les principaux acteurs et les typologies. Les modèles économiques de ces écosystèmes de données connaissent un développement continu, c'est pourquoi ils ont été examinés dans des sections complémentaires. Ces sections comprennent les modèles de revenus, les modèles de tarification et les transactions non monétaires qui peuvent se produire dans les écosystèmes.

La stratégie de recherche, les archétypes et les modèles proposés peuvent orienter les recherches ultérieures sur le sujet.

Mots clés : Écosystème de données, échanges de données, modèles commerciaux dans les échanges de données, modèles de revenus, modèles de tarification, économie de l'écosystème de données.

Que sont les écosystèmes de données et les modèles économiques qui y sont pratiqués ?

I. Introduction :

Les données sont clairement la ressource la plus précieuse au monde. Avec l'évanescence des smartphones, des plateformes et autres technologies qui traquent nos moindres mouvements, les sources de données se sont multipliées par milliers, et donc la quantité de données créées aussi. Au cours de l'année écoulée, nous avons généré plus de données personnelles que jamais dans l'histoire. Cela se traduit par une pléthore d'opportunités, de nouvelles technologies et de nouveaux services, de technologies révolutionnaires et d'avancées incroyables pour l'humanité.

Chaque année, nous utilisons et produisons davantage d'applications, de dispositifs et de produits qui consomment et traitent nos données personnelles. Avec la pandémie mondiale de COVID-19 en 2020, ces chiffres sont montés en flèche car le citoyen moyen était confiné chez lui et passait beaucoup plus de temps sur des applications, tant pour le travail que pour le divertissement. Cela signifie que de plus en plus d'organisations partagent des données, entre secteurs et entre industries. Cela donne naissance à un nouveau modèle de partage des données, appelé "écosystèmes de données". Au sein de ces écosystèmes de données, les informations sont vendues, échangées, transformées et analysées. Divers acteurs apportent leurs produits, services et données dans ces écosystèmes. Avec l'évolution de ces écosystèmes, les modèles commerciaux et tarifaires pour ces échanges de données évoluent également.

Afin de mieux comprendre ces écosystèmes et les propositions de valeur qui peuvent en découler, l'objectif de ce document est de répondre à la question suivante : Quels types de modèles économiques peuvent être utilisés dans les écosystèmes de données ?

Pour tenter de répondre à cette question, 3 sections ont été élaborées grâce à 7 entretiens semi-directifs, et une multitude de sources académiques et non-académiques.

Dans la première section, nous allons explorer les écosystèmes de données afin de mieux comprendre les propositions de valeur qui y sont faites. Nous explorerons la définition d'un écosystème, les façons de le catégoriser, les quatre principaux archétypes d'acteurs qui le composent et les formes qu'il peut prendre. Il y aura également une section spécifique dédiée à [Visions](#), une entreprise française qui travaille à construire et à créer de la valeur à partir de ces écosystèmes autour des données de compétences, pour laquelle je travaille actuellement.

Dans une seconde partie, nous aborderons les différents types d'échanges de valeur possibles au sein de ces écosystèmes. Dans cette étude, nous avons défini les composantes clés d'un modèle économique : son modèle de revenus et son modèle de tarification. Ainsi, cette section

est divisée en une partie pour chaque sujet. Comme les modèles d'entreprise sont souvent construits sur mesure en fonction de l'entreprise et des clients qu'elle sert, ce document a disséqué les modèles d'entreprise et a présenté les modèles de revenus et les modèles de tarification dans des sections distinctes. Ce faisant, je vise à fournir une grande variété d'exemples et de contenus afin que les personnes travaillant dans le secteur puissent combiner les modèles de revenus et de prix afin de construire des business modèles personnalisés basés sur leurs activités spécifiques. Dans la première sous-section, nous examinerons les modèles de revenus présentés, classés en quatre archétypes principaux exerçant ces modèles d'affaires. Dans cette section, des questions importantes concernant ces modèles économiques seront soulevées, telles que : Qui utilise ces services/produits ? Quelle valeur cela apporte-t-il aux entreprises ? Qui orchestre ce type d'échanges ? Peu après, une section sera consacrée à une variété de modèles de tarification couramment pratiqués dans les écosystèmes de données. En plus de ces modèles de tarification, nous explorerons également une section tournant autour des échanges non monétaires possibles entre les acteurs.

Par souci de concision et pour transmettre la meilleure compréhension de ces modèles, tous les modèles tournant autour des revenus des services publicitaires sont brièvement évoqués mais ne sont pas exposés en détail. Ces modèles sont des modèles très particuliers, incroyablement riches et comportant souvent des échanges, des infrastructures et des composants en temps réel très spécifiques et complexes. Les modèles de revenus tournant autour de la publicité sont le plus souvent pratiqués par les GAFAM (Google, Apple, Facebook, Amazon et Microsoft) et sont des modèles qui ont mis des années à se développer. J'ai décidé d'exclure ce sujet dans ce mémoire de master, car il pourrait constituer un sujet de thèse entièrement distinct.

Dans une dernière section, nous concluons ce mémoire de master en présentant la portée de ce projet et les évolutions prévues dans ces écosystèmes au fur et à mesure de leur évolution.

Avant de nous plonger dans la complexité des modèles économiques pratiqués dans les écosystèmes de données, nous devons d'abord les comprendre et en connaître les principaux composants. C'est pourquoi, dans la prochaine section, nous allons explorer les écosystèmes et leurs compositions.

II. Écosystèmes

A. Introduction

Les écosystèmes sont des organismes complexes, aux multiples facettes, qui englobent un ensemble d'acteurs et de types d'échanges. Afin de mieux comprendre ce qu'est un écosystème de données et comment il fonctionne, je présente les principales caractéristiques d'un écosystème. Dans une section suivante, les différents acteurs d'un écosystème sont détaillés. Peu après, je donne des exemples détaillés de types d'écosystèmes qui existent aujourd'hui sur le marché.

Afin de mieux comprendre ce qu'est un écosystème, nous allons d'abord examiner la manière dont il est défini.

B. Définition d'un écosystème

Bien qu'il existe une littérature abondante sur les écosystèmes de données, nous constatons souvent un manque de vocabulaire commun.

Nardi et O'Day, donnent une définition large de ce qu'est un écosystème, " un système de personnes, de pratiques, de valeurs et de technologies dans un environnement local particulier ". Plus précisément, " les écosystèmes sont constitués de composants en interaction, relativement bien connectés et présentant des interdépendances substantielles. Les composantes spécifiques varient d'un écosystème à l'autre (Martin, Sébastien & Turki, Slim & Renault, Samuel. (2017))."

Décrits par d'autres structures plus orientées vers les entreprises, les écosystèmes de données sont "une combinaison d'infrastructures et d'applications d'entreprise (...) utilisées pour agréger et analyser des informations" ("What Is a Data Ecosystem ?", 2021). Un écosystème de données peut également être caractérisé comme "une plate-forme qui combine des données provenant de nombreux fournisseurs et crée de la valeur par l'utilisation des données traitées" (Abdulla, Ahmed, et al., 2021).

À partir de ces ressources, nous comprenons que les écosystèmes de données sont des environnements dans lesquels deux acteurs ou plus, qu'ils soient publics, privés ou une combinaison des deux, se réunissent pour échanger des données ou des services de diverses manières.

Avant d'explorer les modèles économiques pratiqués au sein d'un écosystème, il est important de pouvoir classer les types d'écosystèmes que nous observons. Afin de mieux distinguer les écosystèmes de données, je propose dans une prochaine section un aperçu de deux façons de classer les écosystèmes en deux catégories distinctes : écosystèmes publics ou privés et écosystèmes centralisés ou décentralisés.

C. Catégorisation des écosystèmes

Les écosystèmes et les modèles économiques qui y sont pratiqués sont influencés par le type d'écosystème qu'ils constituent. Bien que les écosystèmes de données contiennent une variété de participants (publics et privés), deux critères distincts nous aident à déterminer le type d'écosystème que nous observons. La première distinction que nous pouvons faire dans les écosystèmes est celle entre les écosystèmes organisés par le public et ceux organisés par le privé. La deuxième façon de catégoriser les écosystèmes est de savoir s'ils sont centralisés (ce qui signifie que toutes les données sont téléchargées vers une plateforme spécifique) ou

décentralisés (les données restent dans des plateformes partenaires et des routes API permettent à d'autres parties d'accéder à leurs données).

Dans cette première sous-section, nous allons explorer la différence entre les écosystèmes à orchestration publique et ceux à orchestration privée.

1. Orchestrateurs publics versus orchestrateurs privés

Les modèles économiques varient fortement entre les orchestrateurs publics et privés, c'est pourquoi il est important de pouvoir classer les écosystèmes par catégories d'orchestrateurs publics ou privés.

Écosystèmes orchestrés par le public

Dans les écosystèmes à gestion publique, une entité gouvernementale, qu'elle soit à petite échelle (gouvernement régional) ou à grande échelle (par exemple, la Commission européenne ou le gouvernement national), gère la création, l'organisation et la gouvernance de l'écosystème.

Un exemple d'écosystème orchestré par le public est un échange en [Corée entre KT Corporation \(Korea Telecom\) et KCDC](#) (Korea Centers for Disease Control and Prevention) pour atténuer l'épidémie de la maladie MERS (Middle East respiratory syndrome) (Boral, Austin, et al., 2019). Afin d'endiguer une épidémie de MERS en 2020, le gouvernement coréen a orchestré un écosystème de données à réponse rapide entre le KCDC et le géant des télécommunications KT. Une publication "[Data Collaboration for the Common Good](#)" du WEF (Forum économique mondial) détaille les paramètres de cet écosystème : "Le KCDC a aidé les autorités gouvernementales à coordonner leurs activités lors de l'élaboration d'une législation essentielle pour répondre aux préoccupations du secteur privé en matière de confidentialité des données et de protection des consommateurs. Plus précisément, la législation prévoyait la destruction de toutes les données de mobilité utilisées au cours de la collaboration après une période déterminée. Cette directive a permis de gagner la confiance des dirigeants et des abonnés de KT, tout en améliorant la transparence de la collaboration en matière de données. L'effort de KT pour améliorer la santé publique via sa plateforme de préparation aux épidémies a encouragé la participation d'autres grands opérateurs mobiles à participer à l'initiative" (Boral, Austin, et al., 2019).

Les écosystèmes orchestrés par les gouvernements, tels que ceux-ci, ont souvent des objectifs très spécifiques pour leur création. Ces écosystèmes ont également des contrats, des mesures et des accords spécifiques sur la manière exacte dont les données seront collectées, utilisées et souvent détruites après la collaboration. Les modèles économiques des écosystèmes orchestrés par le gouvernement sont définis par un contrat et contiennent des propositions commerciales et tarifaires spécifiques entre les acteurs, que nous examinerons dans les sections suivantes.

Des écosystèmes orchestrés par le secteur privé

Les écosystèmes organisés par le secteur privé sont les types d'écosystèmes les plus courants et présentent des modèles de revenus et de prix couramment pratiqués, qui seront présentés plus loin dans cette étude. Les écosystèmes à organisation privée sont ceux qui sont organisés par une entreprise privée.

Un exemple d'écosystème orchestré par le secteur privé est mentionné dans l'article de Marcus Roth et de ses collègues intitulé "[Four ways to accelerate the creation of data ecosystems](#)", dans lequel "un fournisseur de services numériques et une chaîne de supermarchés se sont associés pour résoudre un problème récurrent pour les entreprises de produits de consommation emballés (CPG) : comprendre comment la publicité affecte les achats. Chaque partenaire recueille de grandes quantités de données - le fournisseur de services numériques collecte des données sur la publicité et l'audience à partir de millions de comptes, et la chaîne de supermarchés recueille des données sur les achats à partir de millions de clients. Combinées, ces sources de données peuvent être utilisées pour découvrir le lien entre la publicité et les achats - un défi important pour les spécialistes du marketing des produits de grande consommation et les plateformes de publicité numérique. L'écosystème crée des avantages pour les deux partenaires" (Aaser, Mohammed, et al., 2020).

Dans ces écosystèmes orchestrés par le secteur privé, nous observons une grande variété de modèles de revenus et de tarification. Afin de mieux comprendre l'étendue de ces écosystèmes, nous définissons plus précisément la différence entre les écosystèmes centralisés et décentralisés dans la sous-section suivante.

2.Écosystèmes centralisés et décentralisés

Les écosystèmes centralisés et décentralisés permettent d'appliquer différents types de modèles commerciaux et tarifaires. C'est pourquoi nous allons décrire dans cette section la différence entre un écosystème centralisé et décentralisé.

Centralisé :

Dans une place de marché de données centralisée, les données sont proposées via un emplacement centralisé (Fruhworth, Michael & Rachinger, Michael & Prlja, Emina., 2020). Cela signifie que les données peuvent être téléchargées dans une variété de formats vers un point centralisé tel qu'une place de marché, un lac de données ou tout autre type d'infrastructure disponible.

Un exemple d'écosystème de données centralisé serait la [plateforme d'échange de données Dawex](#). Dans ce modèle, Dawex fait office de place de marché centralisée pour que les acteurs privés puissent fournir, acheter et échanger des données provenant de diverses sources. Le

modèle centralisé est le modèle le plus couramment pratiqué dans les écosystèmes de données aujourd'hui.

Décentralisé :

Le modèle décentralisé est un modèle plus récent de l'écosystème des données créé dans un effort d'innovation, en particulier lors de l'utilisation de données personnelles (données traçables à un individu). Dans un marché de données décentralisé, "les actifs de données restent chez le fournisseur de données, en utilisant par exemple une blockchain" (Fruhworth, Michael & Rachinger, Michael & Prlja, Emina., 2020). Cela signifie que des infrastructures techniques telles que des API ou des technologies blockchain sont mises en place pour construire des voies permettant aux bases de données de communiquer et de transmettre des données. Ce type d'écosystème décentralisé nécessite beaucoup d'innovation en matière de modèle économique.

Un exemple de ce type d'écosystème décentralisé est [Visions dans le projet Grand Calais](#), dans lequel une variété d'acteurs, tant publics que privés, se connectent à une API leur permettant d'envoyer et de demander des données de tous types de sources par le biais de l'infrastructure commune. Outre des modèles commerciaux cohérents, les règles de gouvernance de ces types d'écosystèmes sont extrêmement importantes pour en assurer la stabilité.

Si les types d'orchestrateurs et la nature de l'écosystème peuvent avoir un impact sur les types de modèles de revenus et de tarification choisis dans ces écosystèmes, il est également important de comprendre les acteurs qui participent à ces écosystèmes, car ils ont eux aussi un impact sur la façon dont les échanges sont monétisés.

D. Acteurs des écosystèmes

Chaque type d'acteur de l'écosystème a tendance à exercer certains modèles de revenus et à appliquer les modèles de tarification correspondants. C'est pourquoi cette section est consacrée à la description des types d'acteurs de l'écosystème avant de se pencher sur les archétypes de modèles commerciaux les plus couramment pratiqués.

Pour des raisons de brièveté et de cohérence, j'ai réparti les participants aux écosystèmes en 4 catégories d'acteurs parmi les plus courantes dans les écosystèmes. Les modèles de revenus les plus couramment exposés correspondant à chacun de ces acteurs seront présentés dans une section ultérieure intitulée "Modèles commerciaux : Modèles de revenus et de tarification" dans la sous-section "Modèles de revenus".

Afin de fournir une topologie des acteurs les plus courants dans les écosystèmes, j'ai divisé ces catégories en 4 types d'acteurs.

- Contributeurs de données brutes ou non traitées
- Processeurs de données
- Connecteurs/Orchestrateurs
- Prestataires de services

Cette présentation des acteurs peut être mise en parallèle avec le type de chronologie dans laquelle une entreprise moyenne interagit avec un écosystème : l'entreprise achète des données brutes ou non traitées, elle fait appel à un processeur de données pour transformer ses données en une ressource plus précieuse, elle peut ensuite utiliser un connecteur pour interconnecter, fournir ou échanger des données supplémentaires dans un écosystème, et si nécessaire, elle peut faire appel à un fournisseur de services pour transformer certaines de ses informations en informations exploitables ou autres services pertinents.

En outre, nous commençons notre exploration des acteurs de l'écosystème des données par les contributeurs de données brutes ou non traitées.

1. Contributeurs de données brutes ou non traitées

Les données brutes sont "les données qui sont collectées à partir d'une source, dans leur état initial. Elles n'ont pas encore été traitées - ou nettoyées, organisées et présentées visuellement" (Hiter, Shelby, 2021). Ces acteurs de l'écosystème fournissent des données brutes ou non traitées. Ces contributeurs peuvent être des entités publiques ou privées, selon la forme que prend l'écosystème (ces formes d'écosystèmes seront abordées dans la section suivante). Les contributeurs de données brutes peuvent fournir des données personnelles ou non personnelles. Ces acteurs peuvent également être des entreprises de toutes tailles qui souhaitent apporter leurs données à l'écosystème, que ce soit en autorisant le transfert de leurs données depuis leur cloud ou leur base de données (décentralisé) ou en fournissant des ensembles entiers de données brutes pour que d'autres acteurs puissent les interroger ou y accéder directement sur une plateforme (centralisé). Les ensembles de données brutes, même s'ils ont beaucoup moins de valeur que les ensembles de données traitées, sont souvent importants pour que les entreprises puissent développer de nouveaux services et avoir des idées sur le comportement des consommateurs. Ces contributeurs peuvent être des fournisseurs de données, c'est-à-dire qu'ils vendent leurs propres données, ou des courtiers en données, qui vendent les données d'autres entreprises. Alors que les modèles de revenus pour ce type de fournisseur de données sont assez simples, les contributeurs de données non traitées peuvent parfois avoir des difficultés à fixer le prix de leurs données car elles n'ont pas encore été nettoyées ou transformées. Nous aborderons les modèles de tarification couramment utilisés dans les sections suivantes.

Les entreprises qui achètent des données non traitées ont souvent besoin de services qui traitent les données pour qu'elles puissent être utilisées de diverses manières. C'est pourquoi nous allons définir dans la section suivante le rôle des processeurs de données dans un écosystème.

2. Processeurs

Les processeurs de données sont les acteurs qui fournissent des services de nettoyage, de transformation, d'agrégation, d'exploration et d'exploitation des données afin d'en accroître la valeur. Des sociétés comme [Palantir](#) et le [service Aladdin de Blackrock](#) sont classées dans ce document comme des processeurs de données. De plus amples informations sur les spécificités

de ces entreprises et les services qu'elles fournissent seront mentionnées dans la section suivante "Modèles commerciaux : Revenus et tarification" dans la sous-section "Modèles de revenus".

Le traitement des données fait partie intégrante des écosystèmes de données. Lorsqu'une entreprise dispose d'une grande quantité de données traitées, elle peut, à l'aide d'orchestrateurs/connecteurs, les partager et les monétiser avec d'autres organisations. Dans la prochaine section, nous verrons comment ces connecteurs/orchestrateurs sont définis et quel rôle ils jouent dans les écosystèmes de données.

3.Connecteurs/orchestrateurs

Les connecteurs ont plusieurs noms et plusieurs rôles dans les écosystèmes de données. Ces acteurs fournissent soit un lieu où les échanges de données peuvent avoir lieu et/ou l'infrastructure technique pour d'autres entreprises qui souhaitent le faire. Par exemple, la plateforme [Harbr](#) produit l'infrastructure permettant aux entreprises de créer leurs propres échanges de données centralisés, orchestrés de manière privée. D'autres connecteurs fournissent des API ou des exigences communes pour les échanges de données décentralisés. Une sous-catégorie de ces acteurs sont les facilitateurs, qui fournissent des technologies et des services supplémentaires permettant aux organisations de partager ou d'échanger des informations.

Dans un entretien avec Markus Spiekermann, celui-ci mentionne que "le composant connecteur peut être considéré comme une passerelle vers chaque participant de l'écosystème, et ce qu'il fait, c'est qu'il dispose d'API et d'interfaces pour échanger les données" (Spiekermann, M., 2021). L'idée que les connecteurs peuvent ouvrir des écosystèmes entiers présente de nombreuses opportunités et défis lors de l'identification de modèles commerciaux spécifiques.

En outre, les modèles de revenus et les modèles de tarification des connecteurs peuvent varier considérablement en fonction du type d'écosystème dans lequel ils se trouvent. Il peut s'agir de modèles de frais d'écosystème ou de modèles de source ouverte. Les connecteurs permettent la formation d'écosystèmes de données qui peuvent fournir et partager des quantités massives d'informations. Ces quantités massives d'informations donnent lieu à une sous-économie entière dans les écosystèmes de données, l'économie des fournisseurs de services.

4.Prestataires de services

Une pléthore de nouveaux types de services et de sous-services évoluent chaque jour dans les écosystèmes de données. Un seul fournisseur de services peut proposer une grande variété de services qui peuvent être exécutés une fois que les données sont traitées et/ou connectées. Certains fournisseurs de services pratiquent un ou plusieurs modèles économiques différents en fonction des services qu'ils fournissent et des produits qu'ils génèrent à partir de ces services.

Si les fournisseurs de services sont nombreux, il est également important de comprendre comment ces écosystèmes peuvent éventuellement être formés. Nous avons maintenant exploré la signification d'un écosystème, les distinctions entre les types d'écosystèmes et les acteurs qui jouent les rôles les plus importants dans les écosystèmes. Nous sommes maintenant prêts à découvrir les formes que peuvent prendre ces écosystèmes de données.

E. Quelles formes peuvent prendre les écosystèmes ?

Nous avons maintenant catégorisé les écosystèmes et pouvons commencer à voir comment tous ces éléments s'assemblent et les formes que peuvent prendre les écosystèmes de données.

Les écosystèmes de données peuvent prendre diverses formes et chaque forme d'écosystème implique des modèles de revenus et de tarification qui sont spécifiques aux besoins de leurs utilisateurs. Afin de fournir une vue d'ensemble des types d'écosystèmes de données les plus courants, ce document les divise en trois catégories principales :

- Les marketplaces
- Les circuits de données privés
- Les données ouvertes et « data lakes »

Une section supplémentaire détaillant les écosystèmes [Visions](#) a été ajoutée afin de présenter un nouveau modèle d'écosystème qui évolue actuellement dans l'Union européenne.

Afin de mieux comprendre les formes les plus courantes de ces écosystèmes, nous commençons par explorer le type d'écosystème le plus répandu : la marketplace des données.

1. Les marketplaces

Les marketplaces de données sont les types d'échanges de données les plus courants et, en tant que telles, elles contiennent la plus grande variété de modèles commerciaux et tarifaires. Les marketplaces de données sont le plus souvent organisées par le secteur privé et sont par définition des plateformes centralisées. Ces marketplaces peuvent s'adresser à des acheteurs publics ou privés, bien que ce dernier cas soit le plus fréquent. Les marketplaces sont principalement composées de fournisseurs de données brutes et de processeurs qui vendent les données traitées. Pour partager des données, les participants à l'écosystème utilisent généralement la technologie de connexion de la place de marché. Dans certains cas, cependant, une place de marché peut acheter son infrastructure auprès de ces orchestrateurs.

Pour commencer, nous explorons la définition d'une place de marché de données comme "des plateformes électroniques qui facilitent l'échange de données [...] Un écosystème marketplace se compose de fournisseurs de données, d'acheteurs de données, de fournisseurs de services tiers et d'un propriétaire de la marketplace" (Fruhwith, Michael & Rachinger, Michael & Prlja, Emina., 2020). Les marketplaces comportent souvent des systèmes de notation dans lesquels la qualité des données et les fournisseurs de données sont évalués. Ces types d'écosystèmes

peuvent parfois comporter des mécanismes d'évaluation de la qualité ou des processus de sélection pour les nouvelles sources qui souhaitent vendre ou échanger leurs données.

De nombreuses recherches ont été menées pour cataloguer les marketplaces de données les plus importantes. Markus Spiekerman, dans son [article](#) d'Interconomics, présente une liste de 14 marketplaces de données généralistes bien connues, dont certaines incluent : [Dawex](#) et [Data Broker](#), et [Qlik](#). D'autres marketplaces peuvent être beaucoup plus spécifiques à un secteur, comme [Farmobile](#) qui fournit des données agricoles provenant des agriculteurs, ou spécifiques à un certain type de données qu'elles fournissent, comme les marketplaces de données IOT comme [Terbine](#).

Les marketplaces de données, bien que de plus en plus connues et accessibles, sont confrontées à quelques défis. Le prix des données peut être difficile à évaluer en fonction du temps et des ressources nécessaires pour transformer les données en quelque chose d'utilisable. Les modèles de tarification dans ce type d'écosystème peuvent varier considérablement et seront examinés plus loin dans cette étude. Cela dit, les deux modèles de tarification les plus courants sont les ensembles de données à prix fixe ou les ensembles de données mis aux enchères. Les données sur les marchés de données peuvent également être achetées en totalité, louées pour une période limitée ou facturées pour l'interrogation. Les marketplaces peuvent également offrir des services supplémentaires pour le nettoyage et la transformation des ensembles de données, ainsi que d'autres utilitaires. Ces services seront étudiés dans une phase ultérieure du présent document.

Bien que les marketplaces figurent parmi les types d'écosystèmes de données les plus importants et les plus courants, les écosystèmes de données privés constituent l'autre type d'écosystèmes de données majeurs. Nous allons voir à quoi ressemblent les écosystèmes de données privés dans la section suivante.

2.Écosystèmes de données privés :

Alors que les marketplaces de données donnent accès aux données d'une multitude d'entreprises pour des échanges monétaires, certains écosystèmes de données sont formés entre deux ou plusieurs entreprises à des fins spécifiques. Les écosystèmes de données privés sont le plus souvent orchestrés par des entreprises privées et peuvent être des plateformes centralisées ou décentralisées. Ces écosystèmes de données privés s'adressent le plus souvent à des entreprises privées, mais peuvent inclure des acteurs publics ou des données provenant de sources publiques. Les écosystèmes de données privés sont le plus souvent peuplés par des entreprises disposant de données IoT et non personnelles qu'elles souhaitent échanger avec d'autres entreprises afin d'améliorer leurs processus, leurs produits ou de mieux comprendre certains marchés. Ces écosystèmes sont orchestrés par les entreprises, soit avec leurs propres infrastructures, soit en achetant des produits auprès de connecteurs qui leur permettent de partager des données.

Ces échanges peuvent avoir lieu **en interne**, au sein d'une grande organisation où les différentes branches peuvent échanger des données (gratuitement), ou en **externe** et peuvent être partagés entre plusieurs entreprises.

Écosystèmes internes :

Dans le cas des écosystèmes de données privées internes, les grandes entreprises créent des bibliothèques de données dans lesquelles toutes les branches de l'entreprise peuvent trouver des informations cataloguées. Ces données sont souvent classées et catégorisées afin d'en optimiser l'utilisation. Total Energies, par exemple, a donné accès aux données des quatre branches de l'entreprise dans le cadre de son [projet DARE](#).

Écosystèmes externes :

Dans le cas des circuits de données privés externes, des accords sont négociés entre les entreprises qui permettent soit des échanges ponctuels d'ensembles de données spécifiques, soit la libre circulation des données, en fonction de certains paramètres et délais. Ces processus sont généralement négociés, un contrat est rédigé avec un cas d'utilisation spécifique et un POC (Proof of Concept) et un modèle commercial prédéterminé pour une période de temps spécifique. Cette pratique est normalement menée avec des données non personnelles, comme les données d'usine ou les données IoT, et ces contrats sont renégociés fréquemment.

Ces types d'écosystèmes se caractérisent souvent par : des obligations contractuelles, des mesures spécifiques/indicateurs clés de performance (ICP) définis dans le contrat, un calendrier et des paramètres spécifiques, ainsi qu'une période d'essai suivie de renouvellements contractuels continus. Dans ces écosystèmes, les prix sont généralement négociés lors du processus de rédaction du contrat et sont progressivement réévalués à mesure que la valeur et l'utilité des données deviennent plus claires. Il est difficile d'évaluer avec précision l'analyse coûts-avantages de ce type d'échange de données, car les données doivent souvent être nettoyées, transformées et analysées pour que l'une des entreprises partenaires puisse en tirer profit.

Dans l'ensemble, les écosystèmes de données privées présentent plusieurs complexités qui ont un impact sur les types de modèles économiques qui y sont pratiqués. De telles restrictions sont beaucoup moins fréquentes dans les échanges de données ouvertes et les data lakes que nous verrons dans la section suivante.

3.Échanges de données ouvertes et « data lakes »

Les échanges de données ouvertes sont le plus souvent facilités par les gouvernements et lancés afin de stimuler l'innovation et la croissance d'une économie ou de répondre à un problème spécifique. Les échanges de données ouvertes sont le plus souvent orchestrés par le public et constituent des plateformes centralisées. Ces écosystèmes de données ouvertes et ces data lakes fournissent le plus souvent des informations publiques que d'autres services peuvent utiliser. Les échanges de données ouvertes et les data lakes sont principalement alimentés par des

données non personnelles ou des données personnelles anonymisées. Si les échanges de données ouvertes et les data lakes sont unis par le fait qu'ils sont souvent animés par des entités publiques, il est important de comprendre la différence entre ces deux catégories.

Données ouvertes

Les initiatives de données ouvertes sont presque toujours des ensembles de données gratuites, rendues accessibles à tous les types de services et de fournisseurs de services, publics ou privés. Cette démarche vise à promouvoir l'innovation, la création de nouveaux services et, dans certains cas, à stimuler également les initiatives gouvernementales. Dans un entretien avec Martin de Saulles, maître de conférences en innovation et marketing à l'université de Brighton, il mentionne que les gouvernements monétisent parfois leurs données, comme les données de prévision météorologique par exemple, et que les bénéfices de cette opération sont reversés au secteur public (De Saulles, M., 2021).

Il existe de nombreux autres exemples d'écosystèmes de données ouvertes. L'un d'eux est le service national de santé du Royaume-Uni, qui "est probablement le plus grand ensemble de données sur la santé au monde". Bien que cette base de données nationale fournisse une grande quantité d'informations éducatives et de recherche générées par le système de soins de santé et qu'elle soit anonymisée, Martin De Saulles explique que même si "elle est anonymisée, (...) il n'est souvent pas si difficile de désanonymiser les données et de déterminer réellement d'où viennent ces données et à qui elles se rapportent" (De Saulles, M., 2021).

Un autre exemple est le projet Smart Région du Conseil Régional d'Île-de-France, qui permet un large accès à des informations publiques de toutes sortes. L'Île-de-France a [mis à disposition des données publiques](#) sur une grande variété de sujets tels que des données administratives, des données environnementales, des données sur la construction et bien plus encore. C'est une tendance que nous observons également dans de nombreuses autres administrations, tant au niveau local que national, dans de nombreux pays européens.

Bien que les données ouvertes soient un type d'écosystème en constante évolution, nous notons également la présentation d'un autre type d'écosystème plus orienté vers la recherche, les data lakes.

Data lakes :

Les data lakes constituent un type d'écosystème intéressant dans lequel les données des secteurs public et privé sont rassemblées à des fins scientifiques ou d'apprentissage de l'IA/des algorithmes. "Les données volumineuses sont souvent stockées dans un lac de données (...). Les data lakes peuvent prendre en charge divers types de données et sont généralement basés sur des clusters Hadoop, des services de stockage d'objets en nuage, des bases de données NoSQL ou d'autres plateformes de données volumineuses" (Botelho, Bridget et Stephen Bigelow, 2021). Dans les data lakes, de grandes quantités de données sont anonymisées, apportées par un grand nombre d'organisations publiques ou privées et partagées. Cela se fait normalement à des fins de recherche, permettant de tester et de former scientifiquement des

outils d'IA et d'apprentissage automatique. Ils sont le plus souvent gérés par le gouvernement ou des initiatives financées par le gouvernement pour améliorer une industrie ou promouvoir la recherche scientifique/l'innovation.

Si les data lakes sont le plus souvent gérés publiquement, on trouve quelques exemples de data lakes privés, comme celui d'[Amazon Web Services](#). Un autre exemple a été évoqué lors de mon entretien avec Martin De Saulles lorsqu'il a mentionné une autre entreprise, AOL, et ses activités de promotion de la recherche : "AOL a mis des millions de requêtes de recherche à la disposition de la recherche universitaire il y a des années", mais de grands risques ont été exposés car "si vous saviez comment analyser (les informations dans le lac de données) pour retracer les recherches individuelles", [cela représentait un problème de sécurité majeur](#). Malheureusement, le danger du croisement des données, qui permet aux entreprises de déduire des informations sensibles ou personnelles à partir d'informations supposées anonymes, est connu sous le nom d'"effet mosaïque" et constitue un risque dans les écosystèmes de données.

Dans l'ensemble, le fait de comprendre les marchés de données, les échanges de données privés et les échanges de données publics permet d'avoir une vision holistique des types d'écosystèmes de données. Afin de démontrer un type unique d'écosystème décentralisé, orchestré par le secteur privé, qui permet aux données publiques de circuler, la section suivante est consacrée à la présentation de [Visions](#), l'entreprise pour laquelle je travaille actuellement, et de la création d'un nouveau modèle de collaboration public/privé.

Si nous avons abordé toutes sortes de types d'échanges de données, la section suivante présente un exemple concret d'un nouveau type d'écosystème décentralisé.

F. Visions Ecosystèmes

Afin de démontrer un nouveau type d'écosystème de données en cours de construction, la prochaine section sera consacrée à [Visions](#)¹, et à son travail de mise en place d'un écosystème décentralisé d'enseignement des compétences.

Visions est une entreprise qui travaille actuellement à la mise en place d'une infrastructure, d'outils et de mécanismes de gouvernance pour relever ces défis européens actuels en matière de données. Les écosystèmes de données facilités par Visions sont considérés comme étant privés et sont des plateformes décentralisées. Les écosystèmes de visions fournissent des informations publiques et privées qui peuvent être partagées grâce à une infrastructure API commune. Ces écosystèmes sont peuplés de structures publiques et privées qui font circuler des données personnelles.

Visions est un connecteur/orchestrator qui s'efforce de fournir non seulement des solutions techniques, telles qu'une infrastructure d'API décentralisée, mais qui travaille également en

¹ L'auteur de cette étude, Mackenzie Himmelbauer, travaille actuellement pour cette entreprise grâce au programme d'alternance proposé par Assas.

collaboration avec divers fournisseurs de compétences à la rédaction des normes en matière de partage et de gouvernance des données. La principale technologie de Visions est Visions Trust, l'API et le tableau de bord connecté qui permet à l'individu de donner son autorisation concernant chaque type d'échange de données pour chaque plateforme qui souhaite partager ou recevoir ses données.

Afin de construire des outils pertinents pour une telle circulation de données qui peuvent ensuite être adaptés au niveau européen, Visions a été financé pour construire plusieurs exemples d'écosystèmes de données dans lesquels les données peuvent librement circuler entre une variété de plates-formes.

1. Visions Écosystèmes

[Visions](#) construit plusieurs types d'écosystèmes dans lesquels des plateformes de différents types sont interconnectées afin de faciliter la mobilisation des données sur les compétences personnelles. Dans ces "écosystèmes" où les données peuvent circuler librement, un groupe restreint d'acteurs développent et interconnectent leurs plateformes pour fournir des profils actualisés en temps réel (mis à jour en fonction des évolutions sur les sites d'emploi ou de formation) et des informations supplémentaires sur la personne.

Bien que ces écosystèmes puissent avoir une variété de types d'acteurs interconnectés en leur sein, il y a le plus souvent 5 types d'acteurs tissés dans ces écosystèmes :

- Employeurs
- Organisations publiques (gouvernements régionaux)
- Plateformes Edtech et Jobtech
- Organismes de formation
- Universités

Ces groupes, grâce à une infrastructure API commune, des règles de gouvernance spécifiques et des consentements personnels, permettent aux organisations de partager des données entre elles. Cette libre circulation des données profite différemment à chaque acteur. L'infrastructure commune leur permet de partager et de recevoir des informations personnelles afin de mieux comprendre l'individu, de proposer des offres ciblées et de créer de meilleurs produits et services. De leur côté, les étudiants, les demandeurs d'emploi et les employés sont également en mesure de tenir leur profil à jour en temps réel et d'avoir une meilleure connaissance de leurs compétences et de leurs capacités.

Les écosystèmes de Visions comptent trois types d'acteurs : les entreprises privées, les organismes gouvernementaux et les universités/institutions de formation. Nous commençons par étudier les entreprises privées dans les écosystèmes de Visions :

2. Entreprises privées

Les entreprises privées dans les écosystèmes Visions prennent généralement quelques formes :

- Edtechs - services en ligne qui contribuent à l'analyse ou au perfectionnement de compétences pouvant être liées à l'éducation ou ayant une valeur éducative.
- JobTechs - services en ligne qui aident les employés ou les demandeurs d'emploi à mieux évaluer et formuler leurs compétences et à pouvoir les proposer aux employeurs.

D'autres services éducatifs ou axés sur les compétences peuvent également être trouvés dans les écosystèmes, sous la forme de plateformes d'évaluation des compétences, de plateformes d'orientation, de plateformes de mise en correspondance des compétences, etc. Elles peuvent utiliser ces données pour :

- Créer de meilleurs services pour les individus
- Recevoir des informations d'autres sociétés qui leur permettent d'avoir une expérience utilisateur plus unique ou de fournir de meilleurs services
- Résoudre les problèmes qu'ils peuvent rencontrer en raison d'un manque d'informations ou de services connectés (recherche de nouveaux utilisateurs, fourniture de contenu pertinent, etc.)

3. Agences gouvernementales

[Visions](#) travaille avec le gouvernement français à deux niveaux : national et régional. Les projets nationaux font appel à des opérateurs de données comme Visions lors de la mise en place d'une infrastructure nationale pour la collecte et la circulation des données. Grâce à ses technologies de gestion des autorisations, Visions permet au gouvernement d'avoir une "séparation des pouvoirs" lorsqu'il collecte et utilise les données des citoyens à des fins diverses. Visions est également intégré dans des projets régionaux qui impliquent les gouvernements régionaux en France pour être un fournisseur de données ainsi que l'utilisation des données dans leurs plates-formes créées par le gouvernement pour l'orientation, l'éducation et d'autres aspects. Visions est en mesure d'intégrer ces acteurs dans ses écosystèmes afin de fournir des données et de recevoir des données et des informations en temps réel sur ses électeurs.

Si Visions collabore directement avec les gouvernements nationaux et régionaux français, il est également engagé avec d'autres institutions françaises et européennes. Par exemple, l'agence française pour l'emploi "Pôle Emploi" est un acteur majeur dans la réception et le partage des données sur les compétences. En permettant à une personne de télécharger de manière transparente l'intégralité de ses données sur le lycée, l'université et les compétences trouvées sur diverses plateformes, un conseiller en emploi peut facilement proposer des opportunités pertinentes et réduire considérablement la durée de chômage d'une personne. D'autres organisations publiques ou gouvernementales peuvent également être incluses de la même manière à des fins différentes.

Si les acteurs énumérés précédemment sont les types d'acteurs les plus importants dans les écosystèmes, ces services sont entrelacés grâce à une variété d'acteurs secondaires souvent appelés "facilitateurs". Il s'agit de partenaires qui fournissent des services tels que la connectivité, l'analyse d'ensembles de données complexes et la cartographie des données qui

permettent à ces types de circulation de fonctionner. Les acteurs de ce type, les "enablers" peuvent être ceux tels que mindmatcher, HeadAI, etc.

4. Universités et organismes de formation

Les établissements d'enseignement ont un rôle unique à jouer dans ces écosystèmes. Les universités, les écoles secondaires et les établissements de formation recueillent tous des informations très précieuses sur les étudiants à mesure que leurs connaissances et leurs compétences évoluent. Ces institutions sont toutes à la recherche de services permettant à leurs étudiants d'avoir un meilleur accès aux informations sur leurs compétences et de pouvoir les fournir aux employeurs pour avoir de meilleures opportunités. Les besoins spécifiques des universités ont conduit à la mise en place d'écosystèmes plus petits, en circuit fermé, comprenant un nombre réduit d'acteurs qui répondent à un point sensible identifié par l'université. Un exemple de cela est un ensemble de réorientation dans lequel 3 Edtechs se sont réunies pour construire un processus semi-automatisé pour les étudiants qui ne souhaitent plus poursuivre une voie académique, et qui souhaitent se diriger vers un autre domaine d'études. La combinaison de ces outils, en communication avec les offres de cours et de cours professionnels en constante évolution, permet aux conseillers d'étudiants de conseiller les étudiants de manière plus précise et plus efficace sur leurs changements de majeure ou de mineure.

En conclusion, il est important de pouvoir catégoriser ces types d'écosystèmes, qu'ils soient orchestrés de manière publique ou privée, centralisés ou décentralisés, ou catégorisés en tant que place de marché, écosystème privé, écosystème de données ouvertes ou écosystème holistique décentralisé de type Visions. La façon dont ces organisations sont classées a un impact important sur les modèles économiques qu'elles sont susceptibles de pratiquer. Maintenant que nous disposons de méthodes de classification et de typologies clairement définies, nous pouvons examiner les types d'échanges de valeur qui se produisent dans ces écosystèmes.

III. Modèles commerciaux : Modèles de revenus et de tarification

A. Introduction

Maintenant que nous avons défini et catégorisé en détail ce qui constitue un écosystème, nous pouvons commencer à voir comment toutes les classifications, tous les acteurs et tous les types d'écosystèmes sont entrelacés. Toutes les interactions dans les écosystèmes de données entre les acteurs et la clientèle qu'ils servent impliquent une certaine forme de revenus et de modèles de tarification. Dans cette section, nous démontrons les modèles commerciaux qui maintiennent les écosystèmes de données ensemble en discernant les modèles de revenus et les modèles de tarification pour chaque archétype majeur présenté dans les écosystèmes.

Dans la section suivante, nous allons nous plonger dans le cœur de cet exercice académique, qui est d'explorer les modèles d'affaires dans les écosystèmes de données. Ce chapitre vise à répondre à l'énoncé du problème au début de cette thèse en fournissant une topologie organisée des modèles économiques que l'on peut trouver dans les écosystèmes de données.

Par définition, un modèle économique contient les types de produits et services qu'une entreprise peut vendre pour générer des revenus et les modèles de tarification utilisés pour générer ces revenus. C'est pourquoi nous avons divisé le chapitre sur les modèles économiques en deux sections principales : une section détaillant et classant les principaux modèles de revenus analysés par archétype d'acteur de l'écosystème (producteur de données brutes, processeur, connecteur/orchestracteur et fournisseurs de services) et une section axée sur la tarification et la monétisation des modèles pratiqués.

Une séparation délibérée a été faite entre les modèles de revenus et les modèles de tarification associés à ces types d'acteurs. La raison de cette séparation est que pour chaque type de modèle de revenus, il existe une multitude de modèles de tarification utilisés pour chaque acteur. Dans un souci de clarté et de brièveté, les modèles de tarification sont mentionnés séparément et il y est fait référence une fois que les principaux modèles de revenus ont été identifiés.

Pour poursuivre cette étude des échanges de valeurs dans les écosystèmes de données, nous nous plongerons dans la sous-section consacrée au modèle de revenus.

B. Modèles de revenus

Les modèles économiques sont des entités complexes, comportant souvent plusieurs parties mobiles et des modèles de revenus simultanés appliqués les uns sur les autres. J'ai choisi de catégoriser ces modèles en fonction du type d'acteur qui les pratique, comme cela a été présenté précédemment pour les " acteurs " des écosystèmes de données. Ainsi, la composition de cette sous-section est la suivante :

- Modèles de revenus pratiqués par les organisations qui vendent des ensembles de données brutes ou non traitées
- Modèles pratiqués par les entités qui traitent, nettoient ou transforment les données
- Modèles généralement utilisés par les connecteurs ou les acteurs des écosystèmes qui connectent les services ou orchestrent les échanges.
- Des modèles de revenus axés sur l'offre de produits et de services innovants.

Pour donner une perspective aux complexités multicouches entourant les modèles de revenus dans les écosystèmes de données, il est important de noter que la majorité de ces modèles de revenus peuvent être utilisés simultanément par le même acteur de l'écosystème de données.

Des services complémentaires peuvent être proposés par chacun des acteurs précités. Pour illustrer cela à l'aide d'un exemple : la plateforme [Harbr](#), dont le rôle principal est de fournir

une infrastructure en tant que service (services de connecteur), propose également des services supplémentaires payants pour affiner ou traiter les données (services de traitement des données). Dans le but de simplifier ces complexités, ce document abordera le modèle économique principal associé à chaque acteur.

Bien qu'il existe un nombre illimité d'autres modèles de revenus qui sont créés ou pratiqués chaque jour, les modèles présentés dans les sous-sections suivantes sont ceux que l'on retrouve le plus souvent dans la pratique et dans la littérature.

En outre, nous commencerons notre exposé des modèles de revenus par les acteurs les plus rudimentaires de l'écosystème. Nous commencerons par découvrir les modèles de revenus pratiqués par les fournisseurs de données brutes.

1. Fournisseurs de données brutes

Les fournisseurs de données brutes sont nombreux et peuvent être des acteurs publics ou privés dans tout type d'écosystème (place de marché, lac de données, échange de données privé). Les ventes ou les échanges de données brutes sont le plus souvent observés sur des marketplaces de données organisées par le secteur privé.

Avant d'entrer dans les détails de ce type de fournisseur, nous allons d'abord définir ce que sont les données brutes. Il existe des milliers, voire des millions, de types de données brutes générées chaque milliseconde. Les données brutes peuvent être écrites ou tapées manuellement, enregistrées ou saisies automatiquement par une machine. Vous pouvez trouver des données brutes à différents endroits, notamment dans des bases de données, des fichiers, des feuilles de calcul, et même sur des appareils sources, comme un appareil photo. Les données de ce type peuvent provenir de diverses sources et industries. Parmi les types de données brutes les plus couramment échangées figurent les données de localisation et les données des capteurs IoT.

Maintenant que nous comprenons les données brutes, nous pouvons mieux comprendre pourquoi les entreprises les achètent. La plupart du temps, les achats de données brutes sont le fait d'entreprises du secteur privé. Les gouvernements peuvent acheter des données brutes dans un but précis, mais la majeure partie des achats de données brutes reste le fait d'entreprises privées.

Les données brutes sont, en moyenne, des données beaucoup moins chères à acquérir car elles n'ont pas été traitées ou filtrées. Au premier coup d'œil, les données de localisation brutes semblent attrayantes : elles sont moins chères à l'achat que les données traitées, et il y en a beaucoup plus à exploiter" ("The True Cost of Raw Location Data", 2020). "Les entreprises utilisent ces données pour créer de nouveaux produits et services, mieux comprendre et identifier les segments de clientèle, et avoir plus d'informations sur le comportement des clients" ("The Different Ways Businesses Can Use Raw Data", 2021).

Un exemple spécifique de données brutes qui peuvent être fournies est celui des données de localisation. "Les données de localisation sont normalement générées à partir de données mobiles, collectées à partir de signaux mobiles, de GPS, de balises et de WiFi. Ces informations peuvent être incroyablement utiles pour comprendre le comportement des consommateurs ou le marketing basé sur la localisation et donnent accès à un grand nombre de points de données" ("Raw Location Data : Best Datasets & Providers", 2021).

En outre, nous explorons les données brutes en tant que service et quelques exemples de tels fournisseurs.

Fourniture de données brutes : (Données brutes en tant que service)

Nous commencerons par définir ce qu'est le service de données brutes en tant que service. Dans l'un de ses articles, Justin Lokitz décrit le concept de Data as a Service (DaaS) : " DaaS s'articule autour d'une proposition de valeur pour la fourniture de grandes quantités de (...) données avec l'idée que le travail du client consiste à trouver des réponses ou à développer des solutions pour ses clients. Dans ce cas, les clients peuvent être des fournisseurs de solutions cherchant à utiliser (...) des données brutes pour améliorer leurs propres offres (c'est-à-dire la proposition de valeur) ou même des développeurs souhaitant développer des applications de niche pour répondre aux problèmes des consommateurs" (Lokitz, 2021).

Les entreprises dont la principale source de revenus provient des données brutes peuvent être connues comme des producteurs de données (fournissant leurs propres données à partir de leurs propres sources) ou des courtiers en données, rassemblant des ensembles de données brutes pour la revente. Les entreprises qui contribuent aux données brutes dans les écosystèmes utilisent rarement les données brutes comme principale source de revenus. Martin de Saullès l'a présenté au cours de notre conversation, le fait que "pas mal d'entreprises penseront (qu'elles) pourraient obtenir un peu de revenus supplémentaires en vendant ces données qu'elles ont en réserve" (De Saullès, M., 2021). Ce type d'information est généralement utilisé pour fournir des flux de revenus supplémentaires. Bien que ce type de contribution des données brutes d'une entreprise comporte quelques risques concurrentiels et inconvénients potentiels, il s'agit d'une pratique très courante sur les marketplaces de données et les écosystèmes de données privées, que l'on appelle aussi contribution d'"ensembles de données de base" (Wolfert, S., 2021).

Les producteurs ou courtiers de données brutes tels que [Lifesite](#), [Locationscloud](#) ou [start.io](#) peuvent fournir des ensembles volumineux de données brutes de localisation en échange d'échanges monétaires, ce qui rend les modèles de revenus des producteurs de données brutes les plus simples.

Le plus souvent, une fois que les entreprises ont acheté des données brutes, elles doivent les traiter en interne grâce à des scientifiques et des spécialistes des données compétents qu'elles ont embauchés. Cependant, pour celles qui n'ont pas accès à ces ressources, les entreprises

peuvent engager des data processors pour ces services. C'est pourquoi, dans la prochaine section, nous allons examiner à quoi ressemblent les modèles de revenus des data processors.

2. Processeurs de données

Les processeurs de données sont les organes vitaux d'un écosystème de données. Ils traitent, affinent, agrègent et nettoient les données afin qu'elles puissent fournir des résultats durables aux entreprises. Certaines entreprises auront des équipes de traitement internes, et d'autres entreprises réalisent leurs revenus en vendant des services, des logiciels ou des infrastructures de traitement des données. Les processeurs peuvent également retravailler ces données pour des clients directs, en les vendant directement ou en les vendant à un courtier en données qui les revendra ensuite.

Les sous-traitants sont nombreux, et sont à 99 % des entreprises du secteur privé. Les gouvernements peuvent payer des frais de traitement afin d'encourager les initiatives, ou engager des associations à but non lucratif qui agissent en tant que processeurs afin de faciliter certaines initiatives. Les sous-traitants font partie d'un écosystème de données plus large et sont des éléments nécessaires et complémentaires aux marketplaces et aux échanges de données privés. Les processeurs de données sont nécessaires dans les écosystèmes de données tant publics que privés.

Dans le cadre de cette thèse, j'ai identifié deux principaux types de responsables du traitement des données, les explorateurs et les nettoyeurs de données.

L'exploration et l'exploitation des données en tant que service :

Les explorateurs et les mineurs de données ont la capacité technique d'agréger une pléthore de sources de données et de tirer des enseignements pertinents de ces ensembles massifs de données. L'exploration et l'exploitation des données ne se limitent pas à fournir des informations aux entreprises. Il s'agit d'une forme d'art consistant à combiner plusieurs sources de données complexes et à les transformer en processus ou en informations exploitables qui peuvent être utilisés de manière stratégique. Dans cette sous-section, nous examinons de plus près l'exploration et l'extraction de données.

Avec une grande variété d'ensembles de données, l'exploration des données est la première étape nécessaire pour tirer des enseignements des données. "L'exploration des données est l'étape initiale de l'analyse des données, au cours de laquelle les utilisateurs explorent un grand ensemble de données de manière non structurée pour découvrir les modèles, les caractéristiques et les points d'intérêt initiaux" ("What Is Data Exploration ?", 2021). Pendant la phase d'exploration des données, les données peuvent être visualisées et vendues. L'objectif de l'exploration des données est d'identifier des tendances plus larges qui permettront à une entreprise de déterminer les tendances à étudier plus en détail et d'éliminer les données non pertinentes.

Dans la section suivante, nous allons explorer ce qu'implique l'exploration de données. Comme l'exploration de données, l'exploration de données est l'utilisation de logiciels spécialement développés pour passer au crible des lots massifs de données brutes réparties sur plusieurs sources différentes, trouver des modèles et développer des idées stratégiques. Un exemple, mentionné par Sarb Mann fournit un exemple d'exploration de données comme " un développeur individuel construit un produit de données en récoltant des données géologiques librement disponibles qui peuvent être très précieuses pour les entreprises pétrolières et gazières " (Mann, 2018).

Après avoir mieux compris ce qu'est l'exploration et l'exploitation des données, nous examinons pourquoi cela est utile aux entreprises. L'exploration et l'exploitation des données sont essentielles aux entreprises pour leur croissance et leur développement. La combinaison de plusieurs sources de données permet aux entreprises de mieux comprendre les marchés, les produits, le comportement des consommateurs et bien plus encore. Pour y parvenir, il faut toutefois disposer d'une infrastructure hautement technique, d'IA formées, de mécanismes de cloud computing et d'apprentissage automatique calibrés à des fins spécifiques. Ces types de ressources et de *savoir-faire* ne sont souvent pas disponibles en interne dans les entreprises. Sarb Mann attire notre attention sur les difficultés que rencontrent les entreprises pour effectuer de l'exploration de données en interne : "Il ne serait pas facile de trouver des données précieuses à partir d'un réservoir massif de données acquises auprès de diverses sources. L'exploration exige des efforts considérables et il y aura une opportunité pour les acteurs et les fournisseurs de services qui peuvent choisir un domaine ou un ou plusieurs segments et développer des compétences " (Mann, 2018). Même certaines des plus grandes entreprises ou organisations du monde recherchent d'autres organisations pour les aider à relever leurs plus grands défis en matière de données.

Pour cette raison, les entreprises d'exploration et d'exploitation des données sont incroyablement lucratives et sont souvent pratiquées par des mastodontes de l'industrie des données. [Palantir](#) est un exemple bien connu dans le secteur de l'exploration et de l'exploitation des données. Avec des clients tels que l'armée et la marine américaines, Airbus, IBM et Amazon, Palantir garantit une sécurité de niveau militaire et des technologies capables de combiner de vastes ensembles de données et de fournir des informations immédiatement exploitables. Par exemple, dans l'un des cas d'utilisation des technologies de Palantir, [Gotham](#), une simulation d'une menace de sécurité nationale maritime mobilise 5 sources de données complexes différentes. Dans ce cas d'utilisation, l'imagerie satellite est utilisée pour détecter une attaque potentielle, les modèles d'apprentissage automatique tirent des données des routes nautiques potentielles et calculent l'heure d'arrivée, la base de données des avions automatisés des États-Unis est activée pour trouver et identifier le navire, tout cela pendant que les options codées de déploiement humain sont analysées à partir d'une base de données. Cette opération est rendue possible grâce aux services d'exploration et de fouille de données créés par Palantir. Le modèle de revenus de Palantir est intéressant dans la mesure où il repose sur la construction d'un projet pilote à perte, afin d'avoir une preuve de concept validée pour les clients à un prix

raisonnable. Une fois que cette phase est validée, Palantir s'étend verticalement au sein d'une organisation et, enfin, fait passer les projets à une plus grande échelle, ce qui garantit des bénéfices très importants pour l'entreprise.

Un autre excellent exemple d'exploration et d'exploitation de données est un outil complet dans le secteur financier. Cet outil d'exploration de données, [Aladin](#), a été construit pour faciliter les processus d'investissement et la prise de décision en fournissant des outils qui interconnectent virtuellement plusieurs milliers de ressources financières. Ce service, qui s'adresse principalement aux investisseurs institutionnels et aux gestionnaires de patrimoine, combine des informations sur les marchés publics et privés afin d'apporter de la clarté et de mettre en évidence un certain nombre d'opportunités d'investissement. Le modèle de revenus d'Aladdin consiste à construire et à affiner l'outil Aladdin et à fournir régulièrement des informations et des suggestions d'investissement aux clients.

Les modèles de tarification dans le domaine de l'exploration et de l'exploitation des données varient considérablement et sont souvent adaptés aux projets ou objectifs spécifiques d'une entreprise. On peut observer des modèles de tarification

Alors que l'exploration et l'exploitation des données sont des secteurs tape-à-l'œil et bien connus des grands acteurs, le nettoyage des données est quelque chose de beaucoup plus rudimentaire mais d'incroyablement important pour les entreprises, surtout lorsqu'elles traitent des données brutes.

Le nettoyage des données en tant que service :

Le nettoyage des données implique de prendre les données brutes, d'éliminer les erreurs et de modifier les formats afin qu'elles puissent être facilement vendues, manipulées par l'IA ou l'apprentissage automatique ou plus lisibles pour d'autres sources exploitant les informations. Dans sa publication LinkedIn, Sarb Mann note que " le nettoyage des données est un aspect important de la gestion et de la transformation des données, et (la) phase de nettoyage est étroitement intégrée aux autres phases (Mann, 2018) ".

Les entreprises se livrent à une bataille constante contre les données : combien de temps ou d'investissement me faudra-t-il pour collecter/acheter et utiliser les données par rapport au retour sur investissement de ces données. Nicolaus Henke, associé principal de McKinsey & Company à Londres, indique que pour les entreprises, "le problème fondamental... est que, d'un point de vue économique, c'est comme une tragédie des biens communs dans la plupart des écosystèmes. Les efforts et les avantages du partage des données sont normalement asymétriques. Par exemple, un détaillant et un opérateur téléphonique. La société de télécommunications dispose de nombreuses données transactionnelles de localisation, de nombreuses données très utiles pour les détaillants, que ces derniers pourraient utiliser à des fins promotionnelles, mais pour les rendre utiles au détaillant, la société de télécommunications a besoin de beaucoup de travail. Alors, est-ce que vous passez du temps à nettoyer les données

ou vous passez du temps essentiellement à vous concentrer sur ce qu'elles contiennent ? " (Henke, N., 2021). C'est clairement un problème auquel de nombreuses entreprises sont confrontées et le nettoyage des données est un service très précieux pour les entreprises qui n'ont pas le temps et les ressources pour le faire. Cela étant dit, pour les entreprises à plus petite échelle, le nettoyage des données devient de plus en plus banal. "Avec les technologies d'aujourd'hui, le modèle commercial du nettoyage en tant que service n'est plus réservé aux grandes organisations, note Sarb Mann, ", dans le nouveau paradigme, les entrepreneurs peuvent volontairement nettoyer et conserver des ensembles de données et les vendre à des marchands et des courtiers de données ayant des exigences uniques." Donc, maintenant que nous comprenons ce qu'est le nettoyage de données et à quoi il ressemble, explorons quelques entreprises qui fournissent le nettoyage de données.

Les services de nettoyage de données sont souvent des logiciels en tant que service mis à disposition par une variété d'acteurs. et les modèles de revenus diffèrent avec chaque service. Des entreprises comme [Trifacta](#), [TibcoClarity](#), [Openrefine](#) et [DemandTools](#) sont des types de services logiciels avec des modèles de revenus communs de vente de services de nettoyage de données. Nous examinons les modèles de tarification de ces entreprises dans une section ultérieure.

Dans une application très intéressante de la vie réelle, nous avons également vu le nettoyage des données comme un échange de valeur possible ou un échange de marchandises dans lequel certaines bases de données accepteront de partager leurs données avec un autre acteur en échange de leur nettoyage. C'est ce qu'a démontré une interview de Sjaak Wolfert dans laquelle Farmobile propose une sorte d'**échange de marchandises** : les données IoT des agriculteurs sont collectées par des boîtiers que les agriculteurs fixent sur leurs tracteurs, les données IoT sont agrégées et en échange, elles sont nettoyées par Farmobile. Ces données nettoyées peuvent ensuite être revendues et les agriculteurs reçoivent une petite commission sur la vente, en plus de rapports curatés pour les agriculteurs sur leurs données. Ce cas illustre la diversité des types de modèles de tarification, mais constitue une application réelle fascinante de ce type de modèle commercial (Wolfert, S., 2021).

Bien que les processeurs de données soient de plus en plus importants avec la collecte de plus en plus de données brutes chaque année, les connecteurs le sont tout autant. Les connecteurs sont la colle qui maintient les écosystèmes ensemble en fournissant des produits et des services permettant aux acteurs de tous types de s'interconnecter. Dans la section suivante, nous explorons certains des modèles commerciaux des connecteurs.

3.Connecteurs

Les connecteurs sont le corps de l'écosystème de données. Ils mettent en place une infrastructure technique et des outils pour interconnecter les membres de l'écosystème afin que les données puissent être échangées. Les connecteurs peuvent prendre de nombreuses formes différentes : intermédiaires de données, fournisseurs de logiciels, développeurs d'outils. À

l'heure actuelle, les connecteurs/orchestrateurs sont presque exclusivement des entités privées par opposition aux entités gouvernementales. Les connecteurs font également partie de l'écosystème de données au sens large et sont des éléments nécessaires et complémentaires à la construction de marketplaces ou de tout autre type d'écosystème. Les modèles commerciaux des connecteurs étant les plus complexes à déterminer, nous les classerons en trois catégories : orchestration en tant que service, infrastructure en tant que service et facilitateurs. Avant d'explorer les modèles, nous allons d'abord définir ce qu'est un connecteur et qui utilise ces services.

Nous commençons par définir ce qu'est un connecteur. " La fonction de connecteur peut être définie de plusieurs façons : leader, architecte, instigateur, stratège, facilitateur ou encore simulateur. La principale spécificité de la fonction de stimulateur [ou connecteur] est qu'elle implique de penser et d'influencer l'écosystème " (Martin, Sébastien & Turki, Slim & Renault, Samuel., 2017). Les connecteurs construisent des architectures techniques, des infrastructures et des services habilitants par-dessus. Les connecteurs/orchestrateurs ont aussi généralement des responsabilités en matière de gouvernance ou d'application des règles au sein des écosystèmes qu'ils créent. Ils fournissent généralement des services supplémentaires en plus de leur architecture.

En termes simples, sans connecteurs/orchestrateurs, les écosystèmes et toute l'économie qui les entoure n'existeraient pas. Les connecteurs permettent aux entreprises de partager, d'acheter et de travailler avec des données. La mise en place de ces connexions permet à l'ensemble de l'économie des écosystèmes de données d'exister, et donc à un grand nombre d'entreprises, grandes et petites.

Dans cette prochaine section, nous explorerons l'un des modèles commerciaux les plus courants pour ces connecteurs, l'orchestration en tant que service.

Orchestration d'écosystèmes en tant que service :

Les orchestrateurs fournissent l'infrastructure technique pour faciliter l'interconnexion des services, gèrent l'écosystème et ses composants afin de garantir la valeur pour chacun de ses participants. Les orchestrateurs sont responsables de la mise en place de l'écosystème, des principes de gouvernance et de l'application des règles communément établies par le groupe. L'article du BCG intitulé "[Where is Data Sharing Headed](#)" indique que les orchestrateurs "sont les organisateurs et les gestionnaires de l'écosystème. Ils établissent les règles de gouvernance et de captation de la valeur, fournissent la plate-forme pour le partage des données et l'innovation, coordonnent les activités des participants et fournissent un canal pour les produits et services des contributeurs" (Russo, Massimo, et Tian Feng, 2021).

D'autres exemples d'orchestrateurs peuvent être trouvés à travers une variété de projets de cas d'utilisation. Par exemple, dans le projet [Boost 4.0](#), "dirigera la construction de l'espace européen des données industrielles pour améliorer la compétitivité de l'industrie 4.0 et guidera l'industrie manufacturière européenne dans l'introduction du Big Data dans l'usine, en

fournissant au secteur industriel les outils nécessaires pour obtenir le maximum de bénéfices du Big Data" (Boost 4.0, 2021). Dans cet espace de données, l'orchestrateur ou le coordinateur dans cette situation est l'association [Innovalia](#). Si Innovalia est l'organisateur de ce cas d'utilisation, elle ne fournit pas l'infrastructure technique que ce projet utilise pour faire circuler les données. Il s'agit d'un exemple d'orchestrateur en tant que service, dans lequel le modèle de revenus est un modèle de financement, les fonds étant émis par la Commission européenne. Ce type d'activité d'orchestration est fréquemment financé par des initiatives gouvernementales ou d'autres mécanismes de financement.

Des écosystèmes décentralisés :

Les organisations orchestratrices sont beaucoup plus courantes dans les écosystèmes décentralisés où les données restent dans les bases de données des producteurs. Dans les écosystèmes décentralisés, sur la base des infrastructures d'autorisation, les parties souhaitant échanger des informations demandent et interrogent des données directement à partir des bases de données des autres. Les orchestrateurs décentralisés sont un nouveau type de modèle qui évolue actuellement en Europe. Les entreprises qui fournissent ce type de services sont [Vastuu](#), [Visions](#), [OneCub](#). La majorité des orchestrateurs d'écosystèmes décentralisés fonctionnent selon le principe de la séparation des pouvoirs, ce qui signifie qu'ils n'ont pas accès aux données qui circulent dans l'écosystème, mais qu'ils fournissent l'infrastructure technique et rassemblent les acteurs concernés pour faire circuler les données.

Les écosystèmes centralisés :

Pour les écosystèmes centralisés "Il existe de nombreux orchestrateurs bien connus qui gèrent des écosystèmes basés sur des plateformes dans les domaines des médias sociaux, du commerce électronique, des transports, de la banque et même de l'exploitation minière. Il s'agit notamment des leaders technologiques Google, Amazon, Facebook et Apple, ainsi que d'entreprises plus anciennes, telles que Maersk et Cisco. Et il y a de nombreux leaders émergents parmi les nouvelles startups, notamment la plateforme de crédit MoneyPark, basée en Suisse, et GlobalSpec, une plateforme d'information basée aux États-Unis qui sert les communautés de l'ingénierie, de l'industrie et de la technique" (Dexheimer, Maximilian, et al., 2020). Le modèle de revenus de l'orchestrateur centralisé en tant que service ressemble souvent à la typologie des marketplaces, mais il est également présent dans les échanges privés, les échanges publics et les échanges publics/privés.

L'orchestration globale en tant que service, bien que les modèles de financement soient légèrement différents, va de pair avec le modèle d'infrastructure en tant que service que nous présentons dans la section suivante.

L'infrastructure en tant que service :

L'infrastructure en tant que service est un nouveau type de modèle commercial dans lequel ces organisations orchestratrices fournissent une infrastructure ou une technologie commune qui relie une multitude d'entreprises, de services ou de fournisseurs de données pour permettre les

échanges de données. Cette infrastructure peut prendre différentes formes, comme la fourniture d'API pour favoriser la circulation des données ou la fourniture de l'infrastructure technique pour que d'autres entreprises puissent développer leurs propres écosystèmes ou marketplaces de partage de données. "Ce modèle économique est choisi par des entreprises agissant comme des intermédiaires qui facilitent l'accès aux (...) ressources par des développeurs ou des scientifiques à but lucratif" (Ferro, Enrico & Osella, Michele, 2013). Les clients de l'infrastructure en tant que service peuvent être variés ; les orchestrateurs peuvent être clients de l'infrastructure en tant que service et peuvent également fournir l'infrastructure en tant que service, ce que nous verrons dans la section suivante. Les études du BCG indiquent que "les orchestrateurs tirent généralement parti des solutions de plateforme et d'infrastructure en tant que service des hyperscalers", ce qui signifie que les plateformes d'orchestration sont généralement des fournisseurs de solutions d'infrastructure en tant que service pour des organisations aussi grandes qu'Amazon ou Microsoft.

Les entreprises de toutes tailles achètent des infrastructures en tant que service. "Les grandes entreprises se contentent souvent d'utiliser le logiciel comme un niveau de service. Elles veulent avoir un fournisseur de cloud qui a le logiciel mais qui gère aussi toute l'infrastructure pour elles". Au contraire, "les petites et moyennes entreprises, elles ont leurs propres petits réseaux, donc vous avez un ou deux gars techniques dans leur entreprise, et ils veulent juste vouloir utiliser leur propre infrastructure parce que les services d'infrastructure coûtent trop cher, mais ensuite ils ont besoin du logiciel" fourni par les entreprises IaaS (« Information as a service ») (Spiekermann, M. 2021).

On peut se demander quelle valeur offre réellement l'infrastructure en tant que service. Pour simplifier un peu les choses, elle se divise en deux composantes : l'infrastructure construite pour une entreprise ou un groupe et les outils créés pour que les entreprises puissent construire leur propre infrastructure sur mesure.

Pour les orchestrateurs/connecteurs qui fournissent une infrastructure générale à un groupe, une étude de McKinsey révèle que "cet archétype construit une infrastructure de base et une plateforme technologique sur laquelle d'autres entreprises établissent leur activité dans l'écosystème. Les plateformes de gestion des données et les fournisseurs d'infrastructures de paiement sont des exemples de ces entreprises" (Abdulla, Ahmed, et al., 2021). Dans une explication d'un atelier Google "Open Data on the Web", le groupe conclut que l'une des principales propositions de valeur que ce modèle commercial d'infrastructure en tant que service propose "repose sur une offre initiale attrayante, peu coûteuse ou gratuite (...) qui encourage les achats futurs continus d'articles ou de services de suivi (...) qui sont généralement des consommables caractérisés par une courbe de demande inélastique et des marges élevées" (Ferro, Enrico & Osella, Michele., 2013). Un exemple de ce type de service est le connecteur international d'espace de données. Cette solution "est le composant technique central pour un échange de données sécurisé et de confiance. Le connecteur envoie vos données directement au destinataire depuis votre appareil ou votre base de données dans un espace de données certifié et de confiance, de sorte que le fournisseur de données d'origine garde toujours le

contrôle sur les données et fixe les conditions de leur utilisation. Le connecteur utilise une technologie qui place vos données à l'intérieur d'une sorte de "conteneur" virtuel, ce qui garantit qu'elles ne sont utilisées que dans les conditions convenues par les parties concernées " (*IDS Components, 2021*).

Construisez votre propre infrastructure :

L'infrastructure permettant de connecter les entreprises peut être incroyablement complexe à construire en interne. Même au sein des grandes entreprises, la création d'une infrastructure pour le partage des données entre les succursales peut être écrasante. C'est pourquoi il existe un fournisseur de services qui crée une telle infrastructure prête à l'emploi. Un exemple spécifique d'un fournisseur d'infrastructure en tant que service est la société [Harbr](#). Il s'agit d'une société privée qui fournit le logiciel (Software as a Service ou SaaS) permettant de rejoindre ou de créer des échanges de données d'entreprise ou commerciaux personnalisés. Dans un entretien avec Markus Spiekermann, celui-ci explique que l'infrastructure en tant que service repose sur l'existence de "produits et d'un bon service pour construire de tels écosystèmes." L'infrastructure en tant que service doit soutenir tous les "participants, et ensuite essayer de construire des offres de services, des descriptions, et avoir des frais d'abonnement que l'opérateur collecterait afin de répondre aux besoins de l'écosystème" (Spiekermann, M., 2021).

Si l'infrastructure en tant que service, qu'il s'agisse d'une version prête à l'emploi ou d'une version à construire, est le composant central de cet archétype de modèle commercial de connecteur, il est important de se rappeler qu'une fois l'écosystème construit, un certain nombre de fonctionnalités et de services supplémentaires peuvent être proposés. C'est pourquoi les écosystèmes contiennent également des Enablers et des business models de type Enabler.

Les facilitateurs et les services qu'ils fournissent :

Outre ces organisations d'infrastructure en tant que service, les facilitateurs constituent un autre modèle de revenu intéressant dans les services écosystémiques. Une publication explorant les modèles commerciaux dans les écosystèmes de données explique ce modèle en ces termes : "l'enabler agissant comme intermédiaire fournit aux développeurs un accès plus facile aux ressources (...) (...) et les ensembles de données sont ensuite catalogués à l'aide de métadonnées, harmonisés en termes de formats et exposés par le biais d'API, ce qui facilite la récupération dynamique des données de manière significative" (Ferro, Enrico & Osella, Michele., 2013).

Les facilitateurs peuvent revêtir de nombreuses formes différentes : intermédiaires de données, spécialistes de la gouvernance et des normes, technologies de cartographie des données, etc. Ces entreprises "fournissent à l'écosystème des infrastructures et des outils, tels que la connectivité, la sécurité ou les ressources informatiques. Plusieurs facilitateurs seront nécessaires pour soutenir les différents modèles" (Russo, Massimo et Tian Feng, 2021).

Les intermédiaires comme facilitateurs

Par exemple, les intermédiaires de données sont des facilitateurs ajoutés aux écosystèmes. Ils assurent la gestion des permissions pour les données personnelles et proposent un tableau de bord de gestion des permissions afin que les individus puissent gérer le partage de leurs données. [Visions](#) est également un exemple d'intermédiaire de données, qui fournit un mécanisme permettant aux individus de donner leur autorisation pour que leurs données soient partagées entre les fournisseurs de services. Cette autorisation peut être révoquée ultérieurement sur un tableau de bord personnalisé. Les intermédiaires de données pratiquent souvent des modèles de tarification tels que des frais de fonctionnement annuels basés sur le nombre de personnes pour lesquelles ils gèrent leurs données.

Gaia-X est également un catalyseur qui travaille à l'élaboration de normes communes pour le partage des données au niveau européen. "GAIA-X est un projet initié par l'Europe pour l'Europe et au-delà. Son objectif est de développer des exigences communes pour une infrastructure de données européenne. Par conséquent, l'ouverture, la transparence et la possibilité de se connecter à d'autres pays européens sont au cœur de GAIA-X" (*GAIA-X - Home, 2021*). Pour ce faire, des normes sont établies pour définir les métadonnées et les descriptions des contributeurs de données afin que les données puissent être facilement suivies et cataloguées lorsqu'elles sont utilisées. L'objectif global est de permettre aux écosystèmes transnationaux de fonctionner de manière transparente et plus efficace. Gaia-X est une organisation à but non lucratif que de nombreux acteurs de tous horizons soutiennent pour permettre les échanges de données et les principes de gouvernance.

La cartographie des données est un autre problème majeur des écosystèmes de données que les entreprises doivent résoudre. Les fournisseurs de données se réunissent avec des données dans une variété de formats qui ont tous différents types de données correspondants. Par exemple, sur LinkedIn, les diplômes d'un utilisateur sont répertoriés sous "éducation" alors que sur une autre plateforme comme JobTeaser, ils sont répertoriés sous "expériences et éducation". Si l'on devait automatiser le processus de mise à jour de chacune des plateformes à chaque fois qu'un étudiant obtient son diplôme de licence ou de master, il serait important de relier chacun des points de données correspondants sur chaque site. Des entreprises spécialisées dans le mappage automatisé des données existent et développent des outils permettant de détecter et d'automatiser ce mappage des données. L'un de ces outils est [Mindmatcher](#), qui met en correspondance les données relatives aux études et aux compétences. Ce modèle commercial est une sorte de "matching as a service" et permet de fixer des modèles de prix, d'évaluer l'étendue du travail et de fixer un prix fixe pour le travail de cartographie à effectuer. De même, un autre type d'Enabler sont les Enablers qui harmonisent et cataloguent les données afin que d'autres serveurs, services et développeurs puissent avoir un accès plus facile aux données.

Globalement, dans un écosystème, il est essentiel de disposer de ces orchestrateurs/connecteurs et des nombreux services qu'ils fournissent et des sous-services qu'ils permettent. Tout comme les sous-services que les connecteurs rendent possibles, une multitude de nouveaux produits et fournisseurs de services apparaissent chaque jour avec la croissance de ce modèle d'écosystème

de partage des données. Nous allons explorer quelques-uns des archétypes les plus courants de ces fournisseurs de produits et de services dans la section suivante.

4. Nouveaux fournisseurs de produits et de services

Dans cette dernière section, nous explorons une variété de produits et services différents qui sont rendus possibles grâce aux écosystèmes de données. Ces produits et services peuvent être considérés comme une sorte de sous-ensemble de services qui ont évolué à partir de ces écosystèmes de partage de données.

Dans la section suivante, nous présentons deux types de services primaires que nous observons au sein des écosystèmes de données : l'optimisation en tant que service et les connaissances en tant que service.

Optimisation en tant que service :

L'optimisation dans les entreprises est un facteur clé de succès car l'optimisation des données peut permettre à une entreprise de réaliser des économies importantes. C'est pourquoi plusieurs fournisseurs de services se sont concentrés sur l'optimisation de leurs données et de leurs ensembles de données afin d'aider une entreprise à fonctionner plus efficacement.

Optimisation interne : Ce type d'optimisation en tant que service peut provenir du service de conseil interne d'une entreprise ou peut être un consultant externe qui intervient pour fournir une efficacité opérationnelle technique. "Cet archétype intègre verticalement les données au sein de l'entreprise et de la chaîne de valeur au sens large afin de réaliser des gains d'efficacité opérationnelle. Un exemple est un écosystème qui intègre les données des entités d'une chaîne d'approvisionnement pour offrir une plus grande transparence et des capacités de gestion " (Abdulla, Ahmed, et al. , 2021).

Optimisation de la chaîne d'approvisionnement : Un autre exemple de ce type d'optimisation consiste à utiliser une combinaison d'ensembles de données externes à une organisation et internes à l'organisation afin de mieux comprendre et de fournir de la valeur à l'ensemble de la chaîne d'approvisionnement. Cet archétype intègre verticalement les données au sein de l'entreprise et de la chaîne de valeur au sens large afin de réaliser des gains d'efficacité opérationnelle. Il s'agit par exemple d'un écosystème qui intègre les données des entités d'une chaîne d'approvisionnement pour offrir une plus grande transparence et des capacités de gestion (Abdulla, Ahmed, et al. , 2021). Sjaak Wolfert le démontre à travers son exemple de la chaîne d'approvisionnement des producteurs de lait. "Un autre modèle intéressant est la création nette de valeur (...) dans lequel vous partagez les données entre les différents acteurs de la chaîne d'approvisionnement. J'ai donc toujours l'exemple d'une production de lait, où vous avez des agriculteurs qui produisent du lait pour le transformateur de lait, mais vous avez besoin de nourriture pour cela. (...) Donc, fondamentalement, les semences, la nourriture, il faut aussi des équipements, comme des robots, etc. Maintenant, toutes ces entreprises sont tous ces acteurs

que j'ai mentionnés, ils ne sont pas fondamentalement concurrents les uns des autres. Mais si elles pouvaient partager leurs données, elles pourraient en tirer une valeur ajoutée, car, par exemple, le fournisseur de semences livre des aliments à un agriculteur, mais il ne dispose d'aucune donnée sur la production de lait et sa qualité. Il n'a aucune idée des ressources génétiques utilisées et cela serait très intéressant pour un fournisseur de semences. Vous savez que c'est parce qu'alors vous pouvez peut-être optimiser vos intrants alimentaires pour l'agriculteur et vendre de meilleurs aliments qui sont plus adaptés à ses besoins" (Wolfert, S., 2021). Dans cet exemple, le professeur Wolfert fait la démonstration d'un modèle spécifique appelé création de valeur nette, dans lequel les connaissances et les gains d'efficacité obtenus grâce à ces échanges constituent le mode de paiement de ce type d'échange.

Insights as a Service :

"Insights" est un mot à la mode dans le monde des données. Les types d'informations qui en sont tirées sont multiples et ont été classées en quelques sous-catégories.

L'information en tant que service :

L'information en tant que service est une notion présentée dans l'article de Justin Lokitz intitulé "*Exploring Big data business models & the winning value propositions behind them*". Il décrit ce modèle d'information en tant que service dans lequel " les clients ne veulent pas ou ne disposent pas des ressources nécessaires pour traiter et analyser les données. Ils sont plutôt disposés à échanger de la valeur contre une analyse effectuée par des parties de confiance. Le modèle commercial IaaS consiste à transformer les données en informations pour les clients qui ont besoin de quelque chose - et qui sont prêts à payer pour quelque chose - de plus adapté. Le travail à accomplir par le client consiste davantage à tirer ses propres conclusions, voire à "vendre" une idée basée sur certaines informations" (Lokitz, 2021). Dans ce modèle, ceux qui possèdent les données ne savent pas nécessairement quels types de tendances ou d'informations ils recherchent et une entreprise fournira des informations sur certains marchés ou produits sur lesquels l'entreprise pourra ensuite agir pour lancer de nouveaux produits ou services sur différents segments.

Des réponses en tant que service :

Alors que dans le cas de l'information en tant que service, les clients paient pour obtenir des informations générales sur une quantité de données, les réponses en tant que service sont une situation dans laquelle une entreprise peut fournir "des réponses de haut niveau à des questions spécifiques basées sur des ensembles de données" (Lokitz, 2021). Un exemple serait une entreprise qui souhaite planifier son prochain projet de construction et qui souhaite utiliser des données de localisation achetées sur un réseau d'échange de données et les combiner avec des données géospatiales qui contiennent des informations sur les parcelles disponibles à l'achat. Les fournisseurs de services Answers as a service vont prendre des quantités massives de ces ensembles de données et fournir des réponses pour ce client spécifique et proposer 3 emplacements idéaux. Ce n'est qu'un exemple de ce qu'une entreprise comme celle-ci est capable de faire. [Navteq](#), racheté par Nokia en 2007, était un fournisseur de données

géospatiales qui pouvait combiner un certain nombre de ses ensembles de données pour vendre des informations packagées basées sur des données de localisation.

Dans l'ensemble, ce bref passage sur les produits et services intéressants qui ont évolué à partir des écosystèmes de données varie considérablement en termes de modèles de tarification, mais chacun représente des industries lucratives lorsqu'il est correctement exploité.

5. Conclusion

Comme nous l'avons vu tout au long de cette section, les écosystèmes de données tissent un réseau complexe de flux de revenus. Comme il s'agit d'une industrie naissante, nous voyons souvent différents types d'acteurs (orchestrateurs/connecteurs ou processeurs) fournir des services supplémentaires et pratiquer des modèles commerciaux composés. Il est déjà assez compliqué de les suivre, sans parler des modèles de tarification spécifiques appliqués en plus. Afin d'approfondir les modèles économiques, les modèles de tarification ont été délibérément séparés. Nous allons maintenant passer à l'examen des principaux archétypes de modèles de tarification et de leurs composants.

C. Modèles de tarification

Des millions de transactions ont lieu chaque seconde dans les écosystèmes de données et pour chaque type de transaction, il existe des modèles de tarification. Chaque type d'acteur et de service peut pratiquer un ou plusieurs de ces types de transactions et ainsi multiplier de manière exponentielle les types de modèles de tarification basés sur les services ou les produits qu'ils vendent dans leur modèle économique. Comme l'objectif de ce document est de présenter une variété de modèles d'affaires qui sont utilisés dans les écosystèmes de données et pour que cela soit la base de référence pour les professionnels qui souhaitent naviguer au mieux dans les écosystèmes de données et en tirer profit, il m'a semblé essentiel de mentionner les différents modèles de tarification qui sont apparus régulièrement dans la recherche et qui semblent avoir lieu dans les écosystèmes de données.

Comme il existe de nombreux types de modèles de tarification pratiqués simultanément par les fournisseurs de services de données, ces opérations ont été divisées en 4 sections :

1. Les transactions à demande unique et les méthodes de tarification généralement associées
2. Paquets de volume
3. Transactions temporelles
4. Autres méthodes de tarification couramment pratiquées dans les écosystèmes de données.

1. Transactions sur demande unique

Dans les transactions à demande unique, les échanges sont effectués ponctuellement. Un exemple de transaction à demande unique est l'achat d'un ensemble de données par l'intermédiaire d'une place de marché, comme sur la place de marché de données IoT [Terbine](#), où les entreprises peuvent acheter et échanger des données IoT. Les fournisseurs de données téléchargent de grands ensembles de données et les parties intéressées les achètent pour une somme d'argent fixe (similaire au modèle de tarification d'amazon).

Les transactions à demande unique peuvent être incroyablement utiles pour les entreprises, qui souhaitent ne payer que pour ce dont elles ont exactement besoin. Globalement, cela permet à ceux qui utilisent ce service de payer au fur et à mesure, ce qui le rend souvent pratique et abordable (Firth, Robert, 2021). Cela peut permettre aux petites entreprises disposant de budgets plus modestes d'avoir accès à des ensembles de données spécifiques dont elles ont besoin pour améliorer leurs produits et services.

Les modèles de tarification associés peuvent inclure des méthodes :

- Frais uniques et accès illimité : Également connus sous le nom de frais uniques ou frais forfaitaires
- Pay-per-use
- Paiement à l'unité
- Basé sur des requêtes
- Prix proposé par le client²
- Frais d'inscription
- Frais de transaction
- Ventes à la commission

Les modèles de tarification des producteurs de données brutes sont principalement des transactions à demande unique. En voici quelques exemples : Les actifs de données de Lifesight sont tarifés sur la base de frais d'abonnement mensuels allant de 2 000 \$ / mois à 60 000 \$ / an. Locationscloud propose des forfaits personnalisés pour des besoins spécifiques en matière de données de localisation.

Alors que ces modèles de tarification sont facilement applicables à des ensembles de données fixes qui peuvent être achetés en quelques clics, nous explorons dans la section suivante les types de modèles de tarification qui sont engagés avec de grands volumes de données.

2. Transactions par de volume

Les transactions en volume sont celles dans lesquelles de plus grands volumes de données sont échangés ou achetés, ou les entreprises paient pour avoir accès à de grands volumes de données

² les clients proposent le prix qu'ils sont prêts à payer pour le produit ou le service et les fournisseurs peuvent accepter ou refuser la proposition.

pendant des périodes déterminées. "Les entreprises utilisent le big data dans leurs systèmes pour améliorer leurs opérations, fournir un meilleur service à la clientèle, créer des campagnes de marketing personnalisées et prendre d'autres mesures qui, au final, peuvent augmenter les revenus et les bénéfices. Les entreprises qui l'utilisent efficacement détiennent un avantage concurrentiel potentiel sur celles qui ne le font pas, car elles sont en mesure de prendre des décisions commerciales plus rapides et plus éclairées" (Botelho, Bridget et Stephen Bigelow, 2021).

Les méthodes de tarification associées comprennent :

- Tarification par étapes³
- Frais de service
- Frais de stockage

Les transactions en volume peuvent varier considérablement en fonction du type d'écosystème et de la manière dont les données sont échangées. Certains des mêmes modèles de tarification par requête unique peuvent également se retrouver dans les modèles de tarification au volume, tels que le paiement au volume et le paiement à la requête. Une autre façon de catégoriser les modèles de tarification est basée sur les transactions liées au temps.

3. Transactions temporelles

"La tarification basée sur le temps est une méthode de tarification basée sur le moment où l'article ou le service est acheté" ("Time Based Pricing", 2021). Les méthodes de tarification basées sur le temps permettent aux entreprises de facturer les clients en fonction d'une durée spécifique. Les entreprises pratiquent ce type de modèle de tarification car il les aide à "capitaliser sur la valeur composée des relations avec les clients". Cela signifie que tant que les clients voient continuellement la valeur que l'entreprise leur apporte, ils continueront à payer pour elle " (Campbell, Patrick, 2019).

Les méthodes de tarification qui y sont associées sont les suivantes :

- Frais uniques et accès illimité
- Freemium-limité
- Freemium-illimité
- Frais d'abonnement
- Frais de forfait
- Tarification progressive
- Honoraires horaires

Modèles de processeurs de données : La plateforme Aladdin de Blackrock, mentionnée en tant que processeur de données (archétype du mineur/explorateur de données), utilise des méthodes de tarification basées sur le temps, avec un abonnement de base de type freemium

³ la tarification en escalier, qui signifie qu'à partir d'un certain niveau d'utilisation, le plan doit être mis à niveau : <https://www.billflow.io/pricing-models/step-pricing>.

offrant des informations de base, l'accès à plus d'informations et d'outils augmentant avec l'abonnement.

Ce type de modèle de tarification est de plus en plus populaire dans les communautés d'échange de données. Ce modèle repose sur le fait que les données fournissent une valeur cohérente et constante aux consommateurs.

Bien que les archétypes les plus courants de modèles de tarification aient été présentés dans les sections susmentionnées, d'autres modèles intéressants sont apparus tout au long de mes recherches de thèse et sont publiés ci-dessous dans une section plus généraliste.

4. Autres modèles de prix observés dans les publications

D'autres types de modèles de tarification ont été exposés dans les ressources mentionnées dans ma bibliographie. Ces modèles de prix peuvent être utilisés en conjonction avec les autres modèles de prix mentionnés précédemment. Il s'agit notamment de :

- Frais de prototypes pour la construction de POC puis la mise à l'échelle.⁴
- Tarification par projet
- Frais forfaitaires (incroyablement courant)
- Subventionnement croisé⁵
- Frais fixes
- Licences propriétaires
- Licence pour les sources ouvertes
- Double licence⁶
- Redevances
- Participation aux bénéfices
- Commodities exchange⁷
- Micro-services fournis en échange de redevances sur les écosystèmes

Les producteurs de données brutes démontrent quelques-unes de ces transactions : Les modèles de tarification des données de Start.io sont les suivants : achats ponctuels, licences mensuelles, licences annuelles et modèles de tarification basés sur l'utilisation.

⁴ pratiqué par Palantir

⁵ La pratique consistant à utiliser les bénéfices générés par un produit ou un service pour en soutenir un autre fourni par la même entité opérationnelle. - <https://stats.oecd.org/glossary/detail.asp?ID=4968>

⁶ *La double licence fait généralement référence à l'offre d'un logiciel sous une licence propriétaire et une licence open source.* - <https://www.synopsys.com/blogs/software-security/software-licensing-decisions-consider-dual-licensing/>

⁷ les produits de base entre les entreprises sont échangés, par exemple des données sont échangées pour des rapports et des informations qui peuvent être produits à partir de ces données.

Les services de nettoyage de données sont souvent des logiciels en tant que service mis à disposition par une variété d'acteurs et les modèles de revenus diffèrent avec chaque service. Des entreprises comme [Trifacta](#) et [TibcoClarity](#) proposent des technologies sans code et des services d'abonnement. D'autres, comme [Openrefine](#), offrent leur technologie en open source (modèle gratuit) et d'autres, comme [DemandTools](#), proposent des frais de licence annuels par utilisateur. Une fois encore, ces types de modèles de tarification seront étudiés dans un autre chapitre.

Les modèles de tarification pour les connecteurs/orchestrateurs sont propres à chaque orchestrateur et peuvent être présentés de différentes manières, telles que des frais par requête (le service demandeur est facturé par le service fournisseur), des frais mensuels (le service demandeur paie des frais mensuels) ou des frais d'adhésion tout compris (toutes les parties de l'écosystème paient les frais d'écosystème sur la base d'un abonnement et ont accès aux données des outils). L'IDSA est une association qui fournit des outils gratuits à code source ouvert à ses membres, en particulier aux orchestrateurs. Les frais d'adhésion à l'IDSA vont de 35 000 € pour les entreprises souhaitant adhérer, à 1 000 € pour les startups et les institutions non commerciales. IDSA n'est pas seulement un connecteur technique mais rassemble des projets lucratifs d'échange de données grâce à son riche réseau, et fournit l'accès et l'assistance pour postuler à des financements (*IDSA-MembershipFeeRegulations, 2021*).

En conclusion, il existe une variété de modèles de tarification qui peuvent être rattachés à chacun des modèles d'entreprise susmentionnés. Chaque entreprise ou prestataire de services appliquera un ou plusieurs de ces modèles de tarification à ses produits ou services.

Bien que ces modèles de tarification soient les plus couramment pratiqués dans les communautés des écosystèmes de données, nous avons identifié une autre proposition de valeur importante qui apparaît dans nos écosystèmes. Très peu de recherches académiques ont été menées sur cette section suivante, et elle est presque entièrement basée sur le travail de terrain et les meilleures pratiques dans les écosystèmes qui fonctionnent. Ces échanges sont des échanges non monétaires et seront démontrés dans la section suivante.

D. Les échanges non monétaires

Afin d'élargir la portée de l'analyse des modèles économiques et des modèles de tarification des écosystèmes, j'ai choisi d'inclure également une section détaillant quelques échanges non monétaires qui peuvent avoir lieu au sein des services écosystémiques. Certains de ces fournisseurs de services pratiquent également des échanges non monétaires, souvent pour rendre leurs offres de services plus attractives, ou pour permettre aux organisations à but non lucratif d'avoir accès aux données et aux ressources. Cette pratique consistant à proposer des échanges non monétaires vient le plus souvent en complément des échanges monétaires mais peut dans certains cas se substituer à ces derniers.

Dans certaines publications, ces échanges non monétaires sont qualifiés d'échanges de valeurs "douces", certains qualifient les types d'échanges de formels et d'autres de doux (ODI-Data-Ecosystem-Mapping, 2019).

Dans cette section, nous examinons certains de ces échanges non monétaires par catégorie. L'Open Data Institute ou ODI classe 6 types spécifiques d'échanges non monétaires qui peuvent être proposés dans les écosystèmes de données : idées, connaissances, retour d'information, conseils, réseau et politique. En plus de cela, j'ai contribué : les échanges de données contre des données et les échanges de marchandises. Peu après, j'explique brièvement d'autres soft-exchanges que nous avons identifiés en fonction des différents types d'acteurs divisés en 2 sous-sections : les entreprises privées de compétences et les acteurs publics tels que les gouvernements régionaux ou nationaux.

Nous commençons cette section en explorant les 8 échanges non-monétaires les plus courants identifiés jusqu'à présent.

Des idées :

Dans l'un des exemples précédents mentionnés tout au long de cette thèse, [Farmobile, une](#) plateforme d'agrégation de données agricoles, échange des rapports détaillés et des aperçus basés sur les données que les agriculteurs participants ont fournies. Les agriculteurs peuvent ensuite utiliser les informations générées pour adopter des pratiques agricoles plus efficaces et mieux se préparer aux besoins du marché (en plantant les cultures appropriées). D'autres idées peuvent être vues dans le [projet agricole Join Data](#), dans lequel les agriculteurs fournissent leurs données brutes et l'association nettoie les données afin d'extraire des informations sur l'industrie et de les transmettre à ses agriculteurs. Par exemple, un agriculteur déclare : "La valeur des données ne doit pas nécessairement être de l'argent, mais peut ajouter de la valeur à votre entreprise. En partageant les données, j'obtiens des informations meilleures, plus rapides et plus précises. Je ne veux pas, par exemple, recevoir le message suivant : "Peut-être qu'une de ces dix vaches est malade". Non, je veux savoir "qu'une seule vache est malade". Au final, cela se traduit par moins de travail, moins de stress, moins d'utilisation d'antibiotiques et moins d'inquiétude. Je veux aussi savoir exactement à quel moment de la journée une vache est en chaleur. Il y a 20 ans, c'était un grand progrès de savoir qu'une vache était en chaleur au cours des dernières 48 heures et qu'il fallait à nouveau faire attention dans 3 semaines, maintenant cela ne suffit plus " (JoinData, 2019).

Retour d'information :

Le retour d'information est un élément de plus en plus précieux pour les solutions de toutes tailles. À mesure que les plateformes évoluent et que de nouvelles technologies sont lancées, un écosystème est bien placé pour fournir un retour d'information. Le retour d'information peut prendre de nombreuses formes : qualité des données, efficacité de la plateforme. Chez Visions, l'un de nos principaux facteurs d'attractivité est le fait que nous pouvons fournir régulièrement

des commentaires d'étudiants-testeurs sur nos plateformes partenaires. Une grande majorité d'entreprises n'ont pas le temps ni les ressources nécessaires pour lancer un programme d'essai afin d'améliorer leurs solutions, et le fait de disposer d'un retour régulier est donc très précieux.

Conseils et connaissances :

Les connecteurs d'écosystèmes ont souvent un point de vue global que les autres parties prenantes de l'écosystème n'ont pas forcément. Ces écosystèmes peuvent fournir des conseils ciblés sur l'amélioration des opérations, l'évaluation des tendances à venir ou l'amélioration des solutions sur la base du retour d'information que les autres plateformes reçoivent de leurs utilisateurs et/ou que les autres plateformes doivent se donner entre elles. La connaissance et la compréhension des tendances qui se dessinent dans un écosystème sont des éléments qu'un orchestrateur de données ou un collectif d'acteurs d'un écosystème peuvent se fournir mutuellement.

Réseau :

Les écosystèmes et les interconnexions offrent de grandes synergies. Rejoindre un écosystème de données est un moyen fantastique de travailler avec des services complémentaires. Comme un écosystème est une plaque tournante avec plusieurs nœuds de connexion, les nouveaux partenariats et les synergies sont de grandes propositions de valeur des écosystèmes de données, permettant à chaque nœud de travailler ensemble dans un environnement contrôlé. En outre, le fait de connaître différentes sources et d'être souvent connecté à celles-ci facilite les collaborations entre les parties et peut déboucher sur de nouveaux cas d'utilisation et des projets de recherche et développement.

Politique :

La gouvernance et les politiques de groupe sont incroyablement importantes dans les écosystèmes de données. Elles garantissent des collaborations protégées entre les parties et peuvent également avoir des implications importantes pour les nouvelles politiques européennes ou nationales établies autour des réglementations sur le partage des données.

Données pour données :

Un phénomène qui existe presque exclusivement dans les applications pratiques plutôt que dans les publications universitaires est l'échange de données contre des données. Ce type d'échange a été mentionné par Christoph Mertens, membre de l'International Data Spaces Association (IDSA) qui travaille sur des cas d'utilisation avec de grandes entreprises dans divers secteurs. Il mentionne qu'ils "observent également des choses comme (les entreprises qui contribuent) cet ensemble de données spécifique et (reçoivent) un autre ensemble de données de votre côté, et ont vraiment un échange de données" (Mertens, C., 2021). Bien qu'il y ait de nombreux défis à relever pour évaluer la valeur des données dans ce genre d'échanges, c'est quelque chose que nous voyons fréquemment dans les écosystèmes de [Visions](#).

Bourse des matières premières :

Au cours de l'entretien avec Sjaak Wolfert, la notion de "bourses de marchandises" est évoquée, dans laquelle les données peuvent être échangées contre d'autres produits. Quelques propositions de tels produits peuvent être les suivantes :

- Données brutes
- Données traitées
- Rapports et documents
- Biens physiques
- Services
- Certificats

En conclusion, une pléthore d'échanges de valeur peuvent être proposés dans les écosystèmes de données, mais ils seront toujours adaptés à des situations spécifiques. Dans la prochaine sous-section, nous observons quelques échanges non-monétaires adaptés à des acteurs spécifiques.

Échanges spécifiques aux acteurs :

Dans cette section, nous démontrons des échanges non monétaires spécifiques à certains archétypes de participants à l'écosystème : entreprises privées, gouvernements régionaux et nationaux et autres acteurs publics. La section suivante a été créée à partir des échanges avec chacun de ces acteurs dans le cadre de la création d'écosystèmes de données de compétences décentralisées en France.

Les entreprises privées :

Le problème majeur pour les acteurs privés, les petites et grandes entreprises, est le plus souvent lié à la génération de revenus. Cela implique que les échanges non-monétaires possibles sont étroitement liés à la génération de flux de revenus supplémentaires ou à la conquête de nouvelles parts de marché.

Par exemple, ce que nous considérons généralement comme une proposition de valeur pertinente pour les Edtech (entreprises de technologies éducatives) qui rejoignent nos écosystèmes est le suivant :

1. générer des perspectives supplémentaires pour l'université ou l'école
2. la recherche de nouveaux étudiants ou professionnels pour rejoindre leur plateforme et donc l'accès à de nouveaux groupes de personnes à cibler.
3. la participation aux appels d'offres de financement des gouvernements locaux/régionaux
4. participation à des appels d'offres européens internationaux
5. tester/réagir à leurs plateformes et nouveaux produits et améliorer leurs services
6. la mise à jour des informations sur l'utilisateur et la mise à jour en temps réel des informations pertinentes provenant d'autres plateformes

Un autre exemple d'acteur privé sont les plateformes de recherche d'emploi :

1. plus de trafic sur leurs plateformes
2. des mises à jour en temps réel sur leurs plates-formes pour les tenir au courant des formations ou des programmes universitaires achevés, des certifications
3. des informations en temps réel sur le moment où une personne est employée
4. de nouvelles possibilités d'emploi au sein d'autres entreprises

Gouvernements régionaux et nationaux :

Ces gouvernements locaux ou régionaux sont souvent à la recherche d'échanges non-monétaires portant sur des informations statistiques et la promotion d'initiatives gouvernementales. Certains échanges non-monétaires ont été proposés au cours de notre travail avec les gouvernements régionaux :

1. l'interconnexion des différentes plateformes gouvernementales et des acteurs Edtech/locaux pour faciliter certaines initiatives⁸
2. des statistiques de placement permettant au gouvernement d'évaluer l'efficacité de ses programmes.
3. des informations leur permettant d'identifier et de résoudre facilement les problèmes et les tendances à venir.

En conclusion, il est possible de fournir des échanges de valeur douce à chaque type d'acteur afin de promouvoir la collaboration et la construction de nouveaux projets.

IV. Conclusion

Dans l'ensemble, cette étude nous a permis d'identifier les principaux composants des écosystèmes de données, les modèles de revenus et de tarification des principaux acteurs des écosystèmes, et la manière dont ils se réunissent pour échanger des informations et de la valeur. Nous concluons ce document par des réflexions finales sur les écosystèmes de données et sur les types d'évolution des modèles économiques que nous observons. Peu après, nous tirons les conclusions finales de ce mémoire de maîtrise en étendant la portée de ce projet et en décrivant les résultats que nous espérons obtenir de ce projet.

A. Conclusion sur les écosystèmes :

Afin de comprendre l'avenir des écosystèmes de données, il a été important de comprendre la manière dont nous définissons les écosystèmes de données, leurs composants et leurs typologies.

L'avenir que nous entrevoyons pour les écosystèmes publics/privés est le suivant : à un certain moment, certaines de ces grandes marketplaces de données pourraient devenir des biens

8

publics, car elles commenceront à contenir des quantités substantielles d'informations personnelles des citoyens. Les informations échangées dans ces écosystèmes ont déjà une valeur incroyable. Prenez les informations personnelles qui circulent dans l'écosystème de Google. Que se passerait-il si un seul gouvernement acquérait Google et toutes ses informations ? Des économies entières se déplaceraient-elles ? Certaines industries seraient-elles éliminées ?

L'avenir des écosystèmes centralisés/décentralisés est un débat assez fascinant, la communauté européenne se concentrant fortement sur l'avenir des écosystèmes de données. La tendance actuelle sur le marché des données de compétences est la "centralisation". Le problème avec ce modèle est que, lorsqu'un mastodonte fait cela (comme un GAFAM), tous les services se rassemblent autour du même acteur pour produire un grand catalogue de services à la manière d'une place de marché organisée. C'est ce que l'on observe avec le modèle de l'Apple Store - où Apple a créé un lieu centralisé où tous les services peuvent apparaître et être achetés. Cependant, lorsque de nombreux acteurs se disputent le même modèle centralisé, le marché se fragmente en plusieurs acteurs fournissant les mêmes services et se faisant concurrence. C'est ce qui se passe avec le marché des données sur les compétences ; de nombreux acteurs se disputent le titre de fournisseur de services centralisés et fragmentent le marché en 10 ou 15 méga-sites qui n'offrent pas toujours des services de qualité.

Ce problème n'est pas seulement propre au secteur privé. Le gouvernement français, en particulier, est prêt à innover et à fournir de nouveaux hubs de services basés sur ses propres innovations internes et de nouveaux types d'applications gouvernementales pour l'apprentissage, l'éducation ou la découverte de compétences. Le problème, c'est qu'en créant une plateforme centralisée avec leurs propres versions des technologies de l'éducation, ils mettent hors-jeu des acteurs plus petits qui fournissent en fait des services bien meilleurs que ceux que le gouvernement est capable de fournir avec ses ressources limitées.

Sachant cela, je prédis une nouvelle vague à venir, qui est la prolifération du modèle de plateforme décentralisée. Les modèles de décentralisation permettent à d'innombrables acteurs d'être interconnectés - permettant ainsi à tous les acteurs de participer et de prospérer et laissant le consommateur choisir où il veut partager ou obtenir ses données plutôt qu'une méga-plateforme gagnante. Dans ce modèle décentralisé, nous pouvons relier les initiatives des plateformes gouvernementales aux acteurs privés qu'elles affectent. Pour ce faire, nous leur permettons d'avoir accès aux données générées par les petits acteurs qui aident ensuite le gouvernement à améliorer son offre globale ou ses outils sans mettre personne en faillite. La nouvelle législation et les incitations de la Commission européenne à créer des bases de données sectorielles soutiennent ce type de modèle de décentralisation en finançant la création de principes de gouvernance et d'une infrastructure technique API.

Les acteurs des écosystèmes évolueront aussi radicalement avec les écosystèmes. Avec l'évolution de la technologie et des possibilités d'échange, je prédis qu'il y aura une prolifération

de types d'acteurs, et de services que chaque acteur peut proposer que je détaillerai pour chaque acteur dans la conclusion du business model.

Au contraire, je pense que les types d'échanges de données resteront assez simples. J'estime que les marketplaces vont commencer à fusionner en de grands conglomérats, et qu'elles seront très probablement détenues et exploitées par de grandes entreprises comme Amazon ou Microsoft.

B. Conclusion Section sur les modèles d'entreprise :

Les modèles commerciaux en général sont des entités complexes qui évoluent et se complexifient avec le temps. À mesure que les écosystèmes et leurs échanges se complexifient, il en va de même pour les types et le nombre de modèles de revenus qui y sont pratiqués. J'estime que le paysage des modèles de revenus sera complètement différent dans 5 ans, 10 ans et encore plus par la suite.

En outre, à mesure que nous créons davantage d'outils, d'applications et de plateformes qui suivent nos données, nous multiplions la quantité de données brutes existantes. Plus il y a de données brutes, plus les entreprises peuvent apprendre de leurs consommateurs et des habitudes des gens en général. Les fournisseurs de données brutes resteront extrêmement pertinents pour un grand nombre d'entreprises. La façon dont je vois cette industrie évoluer est que de plus en plus d'entreprises développeront des outils ou des mécanismes pour collecter leurs propres données brutes et seront donc peut-être moins nombreuses à payer pour ces services de collecte de données brutes.

Au contraire, les processeurs de données (mineurs, explorateurs, facilitateurs) gagnent en importance. Maintenant que nous disposons de quantités massives de données provenant d'un grand nombre de sources, comment pouvons-nous les transformer en informations exploitables ? Les grandes entreprises semblent produire de plus en plus leurs propres types de processeurs de données internes, mais cette industrie du traitement a encore une énorme expansion devant elle et est souvent dominée par de grandes entreprises comme Palantir qui ont une avance technologique et commerciale sur leurs concurrents.

Tout comme les processeurs, l'industrie des connecteurs est naissante mais se développe chaque jour. Avec la croissance estimée de ce modèle de décentralisation, comme mentionné dans la section précédente, les connecteurs deviennent également de plus en plus importants. Les connecteurs continueront à interconnecter les services, à créer des lieux d'échanges et des services de connectivité supplémentaires. Ce marché va connaître une croissance exponentielle au cours des prochaines années, car les entreprises produisent de plus en plus de données et souhaitent les monétiser avec d'autres entreprises.

En outre, le type d'industrie qui évolue le plus rapidement est celui *des services et des fournisseurs de services*. Nous commençons à voir des fournisseurs de services qui innovent et s'appuient sur d'autres services. Prenons l'exemple de ce service qui a été construit à partir d'un service original, qui fournissait des informations. Nous verrons ce type d'évolution par ricochet se produire continuellement à mesure que les écosystèmes et leurs services évolueront.

L'ensemble des écosystèmes de données et tous leurs composants évoluent à la vitesse de la lumière. C'est le moment idéal pour travailler dans un écosystème de données et faire partie de la sous-économie des écosystèmes de données. L'opportunité d'innover et d'être un entrepreneur ne peut être ignorée. L'idéal est de créer des services ou des composants qui se connectent ou participent à ce type d'écosystèmes. Vous avez des clients qui disposent d'importantes sommes d'argent, des gouvernements qui financeront vos prototypes, et vous n'avez besoin que d'une expertise technique/écosystémique initiale, qui est relativement simple - car le modèle d'écosystème n'en est qu'à ses débuts.

Les modèles de tarification continueront également d'évoluer à mesure que des services plus complexes seront formulés et exécutés pour des clients payants. Je suis également convaincu que les "échanges non-monétaires", tels qu'ils sont présentés dans ce document, évolueront et seront utilisés de plus en plus régulièrement pour ajouter de la valeur aux offres payantes ou pour permettre aux gouvernements de collaborer davantage avec le secteur privé.

C. Où vont les écosystèmes de données ?

Après avoir appris à connaître les écosystèmes et les modèles d'échange de valeur qui peuvent se produire au sein des écosystèmes de données, nous pouvons maintenant nous poser une question très importante : où vont les écosystèmes de données ?

Les écosystèmes de données sont naissants. Les associations incluses dans cette thèse, telles que l'IDSA (International Data Spaces Association), ont fourni des outils de cartographie qui nous permettent de voir les secteurs et les types d'écosystèmes qu'elles ont contribué à développer via leur [Data Space Radar](#). Il existe une variété d'espaces de données différents qui se construisent en reliant différents secteurs et pays en Europe. Lorsque j'ai demandé à Christophe Mertens, responsable de l'adoption à l'IDSA, comment il voyait l'avenir des écosystèmes de données, il a expliqué que sa "vision est vraiment que la compréhension des espaces de données va se développer, parmi les gens. Et d'après ce que j'observe actuellement en Europe, la vision protectionniste que beaucoup de gens adoptent dans différents domaines, j'espère vraiment que nous serons en mesure de surmonter cela avec l'aide de solutions techniques" et "si vous adoptez cette vision protectionniste et dites que nous voulons une solution européenne, et que nous sommes en concurrence avec le reste du monde, plutôt que de coopérer ou de collaborer, je pense que cela échouera". Je pense que cela va échouer". Son "espoir est que nous ayons des solutions internationales qui rassemblent des entreprises internationales, des chaînes d'approvisionnement (afin que) les données internationales

puissent circuler librement, indépendamment des frontières, si elles respectent, bien sûr, les lois et les règles" (Mertens, C., 2021).

J'ai beaucoup apprécié d'avoir la possibilité d'explorer ce sujet à plusieurs niveaux. J'ai pu le faire grâce à l'accès à un large éventail de publications universitaires ainsi qu'à l'expérience concrète de professionnels travaillant dans le secteur. Au cours de mes entretiens, j'ai pu poser plusieurs questions sur les écosystèmes de données, en plus des simples modèles économiques. Des questions ont été posées sur une variété de sujets tels que :

- qui paie les coûts de transformation dans les écosystèmes de données ?
- risques/avantages pour les orchestrateurs de données dans les écosystèmes ?
- risques/avantages pour les donneurs de données dans les écosystèmes ?
- mécanismes pour faire respecter le fair-play dans les écosystèmes de données ?
- des mécanismes spécifiques pour garantir la valeur pour tous les participants à l'écosystème ?

Comme ces sujets peuvent être considérés comme un peu hors sujet pour cette thèse sur les modèles économiques dans les écosystèmes de données, je les ai exclus de la thèse de master officielle. Cela étant dit, l'accès aux recherches effectuées sur ces sujets sera accordé à tous les participants aux entretiens et à toute autre partie intéressée par cette recherche. En outre, tous les participants aux entretiens ont eu accès à une liste d'articles annotés et de publications universitaires qui ont été pris en compte dans le cadre de ce projet de recherche.

Ce projet de recherche ne se terminera pas avec la soumission de cette thèse de masters. Une version modifiée de cette thèse sera publiée par mon employeur Visions en vue de notre demande de fonds auprès de la Commission européenne pour approfondir la recherche sur les modèles d'entreprise. D'autres grandes organisations telles que MyData Operators, ANewGovernance et la SkillsAlliance seront appelées à contribuer à la publication de cette forme "livre blanc" de ma thèse. Les recherches effectuées dans le cadre de ce projet seront également utilisées pour mettre en place un groupe de travail avec le ministère français de l'éducation nationale, qui construira une variante d'un espace de données sur les compétences pour les jeunes Français dans les prochaines années. Je prévois également de publier une série d'articles basés sur les différents sujets industriels présentés dans ce document, avec des contributions de spécialistes du secteur.

En conclusion, je partage la vision de Christoph Mertens et j'espère que, grâce à la recherche sur les modèles commerciaux et de gouvernance, à la mise en place d'infrastructures techniques et à des projets de collaboration, nous serons en mesure de combler les écarts entre le partage des données aux niveaux régional, national et international. Cela nous permettra d'avoir de meilleures technologies, une plus grande transparence et soutiendra l'évolution globale de l'économie de l'écosystème des données dans son ensemble.

Références

Aaser, Mohammed, et al. *"Four Ways to Accelerate Creating Data Ecosystems | McKinsey"*. McKinsey & Company, 23 nov. 2020, <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/four-ways-to-accelerate-the-creation-of-data-ecosystems>.

Abdulla, Ahmed, et al. *"Data Ecosystems Made Simple | McKinsey & Company"*.

McKinsey,

8 mars 2021, <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/tch-forward/data-ecosystems-made-simple>

Attard, Judie & Orlandi, Fabrizio & Auer, Sören. (2016). Réseaux de valorisation des données : Enabling a Nouvel écosystème de données. 453-456. 10.1109/WI.2016.0073.

Boost 4.0 | Big Data pour les usines. <https://boost40.eu/>. Consulté le 29 août 2021.

Boral, Austin, et al. *Data Collaboration for the Common Good Trust and Innovation Through Public-Privé. Insight Report, Forum économique mondial, 2019.*

Botelho, Bridget, et Stephen Bigelow. "Qu'est-ce que le Big Data et pourquoi est-il important ?" TargetTech, mai 2021, <https://searchdatamanagement.techtarget.com/definition/big-data>.

Campbell, Patrick. Modèles d'affaires par abonnement : Un guide des modèles d'abonnement pour 2019.

<https://www.priceintelligently.com/blog/subscription-business-model>. Accédé le 29 Août 2021.

Casadesus-Masanell, R., & Zhu, F. (2013). Business Model Innovation and Competitive Imitation : The Case of Sponsor-Based Business Models. *Strategic Management Journal*, 34(4), 464-482. Récupéré le 2 août 2021, de <http://www.jstor.org/stable/23471019>

De Saulles, M. (2021, 27 juillet). Communication personnelle [Interview].

Dexheimer, Maximilian, et al. "Four Strategies to Orchestrate a Digital Ecosystem". BCG, 4 Sept. 2020,

<https://www.bcg.com/fr-fr/publications/2020/four-strategies-to-orchestrate-digital-ecosystem>.

Ferro, Enrico & Osella, Michele. Atelier "Open Data on the Web" ; avril 2013, Google Campus, Shoreditch, Londres,
https://www.w3.org/2013/04/odw/odw13_submission_27.pdf

Firth, Robert. "Le modèle économique du paiement à l'usage expliqué par LinkedIn". LinkedIn,
<https://www.linkedin.com/pulse/pay-per-use-business-model-explained-robert-frith/>.
Consulté le 29 août 2021.

Fricke, Samuel et Maksimov, Yuliyana. (2017). Fixation des prix des produits de données dans les données
Marketplaces.10.1007/978-3-319-69191-6_4.

Fruhworth, Michael & Rachinger, Michael & Prlja, Emina. (2020). À la découverte de l'entreprise
Modèles de marketplaces de données. 10.24251/HICSS.2020.704.

GAIA-X - Accueil.

Consulté le 29 août 2021.

Ghezzi, Antonio & Georgiades, Michael & Reichl, Peter & Le Sauze, Nicolas & Di Cairano-Gilfedder, Carla & Mangiaracina, Riccardo. (2013). Générer de l'innovation
Modèles commerciaux d'interconnexion pour l'Internet du futur. Info. 15. 40-65.
10.1108/info-12-2012-0054.

HECK, ERIC VANBIG DONNÉES ET BOULEVERSEMENTS DES MODÈLES D'ENTREPRISE. Revista de
Administraçãode Empresas [en ligne]. 2019, v. 59, n. 6 [consulté le 6 août 2021] , p. 430-432. Available from: <<https://doi.org/10.1590/S0034-759020190608>>. Epub 10 Jan 2020. ISSN 2178-938X. <https://doi.org/10.1590/S0034-759020190608>.

Henke, N. (2021, 4 août). Communication personnelle [Interview].

Hiter, Shelby. "Qu'est-ce qu'une donnée brute ? | Datamation | Utilisations des données non traitées". Datamation, 4
Mai 2021, <https://www.datamation.com/big-data/raw-data/>.

Composants IDS | Espaces internationaux de données.

<https://internationaldataspaces.org/use/ids-components/>. Consulté le 29 août 2021.

IDSA-Réglementation des frais d'adhésion-2021-01.Pdf.

JoinData. <https://join-data.nl/?lang=en>. Consulté le 29 août 2021.

Lokitz, J. (2021, 22 juillet). Communication personnelle [Interview].

Lokitz, Justin. Exploration des modèles commerciaux du Big Data - Business Models Inc. <https://www.businessmodelsinc.com/big-data-business-models/>. Consulté le 29 août 2021.

Mann, Sarb. "L'échange et la place de marché des données, un nouveau modèle économique en devenir". LinkedIn, 26 mars 2018, <https://www.linkedin.com/pulse/data-bazaar-new-business-model-making-sarab-s-mann-mba/>.

Martin, Sébastien & Turki, Slim & Renault, Samuel. (2017). Les écosystèmes de données ouvertes :

Présentation de la fonction de stimulateur. 6ème conférence internationale sur le gouvernement électronique et la perspective des systèmes d'information EGOVIS 2017. https://www.researchgate.net/publication/317278860_Open_Data_Ecosystems_Introducing_the_Stimulator_Function/citation/download

Mertens, C. (2021, 24 août). Communication personnelle [Interview].

ODI-Data-Ecosystem-Mapping---Print-at-Home-Guide-A4---2019-06-26.Pdf.

Données brutes de localisation : Les meilleurs ensembles de données et fournisseurs 2021 | Datarade.

<https://datarade.ai/data-categories/raw-location-data>. Consulté le 29 août 2021.

Redman, T. (2021, 23 juillet). Communication personnelle [Interview].

Russo, Massimo, et Tian Feng. "Où va le partage des données ?" France FR, 7 février 2021, <https://www.bcg.com/fr-fr/publications/2021/broad-data-sharing-models>.

Spiekermann, M. (2021, 28 juillet). Communication personnelle [Interview].

Spiekermann, Markus. "Data Marketplaces : Trends and Monetisation of Data Goods". Interéconomie, vol. 2019, no. 4, 2019, p. 208-16, <https://www.intereconomics.eu/>.

" Les différentes façons dont les entreprises peuvent utiliser les données brutes ". CoFounder, 3 février 2021,

<https://www.cofmag.com/2021/02/different-way-use-raw-data/>.

"*Le véritable coût des données brutes de localisation*". Gravy Analytics, 1er septembre 2020, <https://gravyanalytics.com/blog/the-true-cost-of-raw-location-data/>.

"Time Based Pricing". Réseau Topio, <https://www.topionetworks.com/topics/time-based-pricing-50aa0ea13ae4b249e60179dd>. Consulté le 29 août 2021.

Wessels, B., Finn, R., Wadhwa, K., Sveinsdottir, T., Bigagli, L., Nativi, S. et Noorman, M. (2017). Big données, les données ouvertes et le secteur commercial. Dans *Open Data and the Knowledge Society* (pp. 159-174). Amsterdam : Amsterdam University Press. Consulté le 2 août 2021 à l'adresse <http://www.jstor.org/stable/j.ctt1pk3jhq.13>.

"Qu'est-ce qu'un écosystème de données ?" | Cognizant. <https://www.cognizant.com/us/en/glossary/data-ecosystem>. Consulté le 29 août 2021.

"Qu'est-ce que l'exploration de données ?" Sisense, <https://www.sisense.com/glossary/data-exploration/>. Consulté le 29 août 2021.

Wolfert, S. (2021, 12 août). Communication personnelle [Interview].