

Biens communs de données

Foire aux questions

11 mai 2026

Stefaan Verhulst
Andrew J. Zahuranec
Hannah Chafetz
Leona Verdadero
Jennifer Hansen

Table of Contents

Introduction: Pourquoi les biens communs de données sont-ils si importants aujourd'hui ?	3
Questions	4
Qu'est-ce qu'un bien commun de données ?	4
D'où vient l'idée des biens communs ?	5
Comment les biens communs sont-ils gérés ?	6
Tous les biens communs (de données) respectent-ils pleinement les huit principes d'Elinor Ostrom ?	7
Qu'est-ce qui caractérise un bien commun de données, par opposition à un « commons » ?	7
En quoi les biens communs de données se distinguent-ils des « data collaboratives », des « trusts », des « spaces » et d'autres types d'initiatives ?	8
En quoi un bien commun de données se distingue-t-il des données ouvertes ?	9
Quels problèmes les biens communs de données visent-ils à résoudre ?	9
Quelles sont certaines des critiques formulées à l'encontre des biens communs de données ?	10
Pourquoi l'adhésion sociale est-elle essentielle pour les biens communs de données ?	12
Quel est le rôle de la gestion des données dans les biens communs de données ?	13
Existe-t-il déjà des biens communs de données ?	14
Quels sont les modèles de financement utilisés pour soutenir les biens communs de données ?	15
Pourquoi est-ce important pour l'intelligence artificielle ?	16
Pourquoi ce modèle est-il adapté aux groupes marginalisés ?	17
Lectures complémentaires	18



INTRODUCTION

Pourquoi les biens communs de données sont-ils si importants aujourd'hui ?

L'un des grands paradoxes de notre ère de la « datafication » est que nous vivons à la fois dans une abondance et une pénurie sans précédent. Alors même que les données jouent un rôle de plus en plus central dans notre capacité à promouvoir le bien commun, elles restent profondément – et peut-être de plus en plus – inaccessibles et sous le contrôle du secteur privé. En réponse à cela, on entend de plus en plus d'appels en faveur de biens communs de données (data commons) – des pools de données qui seraient (auto-)gérés par des communautés ou des entités distinctes agissant dans l'intérêt public. Ces pools pourraient alors être rendus accessibles et réutilisés pour le bien commun.

Les biens communs de données constituent une alternative à l'utilisation croissante de silos de données privatisés ou à la réutilisation extractive des ensembles de données ouverts. Ils permettent d'organiser les données en tant que [ressource partagée](#), gérée collectivement et mise au service de l'intérêt général. Bien que cette pratique en soit encore à ses débuts et que les données disponibles restent limitées, les biens communs de données pourraient servir d'infrastructure utile pour l'IA d'intérêt public en :

- ▶ Favoriser un accès équitable à des données de haute qualité
- ▶ Soutenir les systèmes d'IA s'appuyant sur des informations diversifiées et représentatives
- ▶ Mettre en place des mécanismes institutionnels pour une réutilisation responsable des données
- ▶ Instaurer la confiance grâce à la participation, à la transparence et à la responsabilité

Les biens communs de données ne se créent pas spontanément. Ils nécessitent une gouvernance réfléchie, une gestion rigoureuse et une acceptation sociale. Cette FAQ offre un aperçu concret de ce que sont les biens communs de données, de leur importance et de la manière de les concevoir de manière responsable.



Questions

Qu'est-ce qu'un bien commun de données ?

[Les biens communs de données](#) sont des écosystèmes de données gérés de manière collaborative, conçus pour mettre en commun et offrir un accès responsable (et réglementé) à des ensembles de données variés et de haute qualité provenant d'un ou de plusieurs secteurs, afin de permettre le développement et le déploiement d'applications d'IA générative visant à relever des défis d'intérêt public.

Cette approche se distingue en ce qu'elle considère les données comme une ressource partagée, régie par une communauté, plutôt que comme un bien relevant exclusivement de la propriété individuelle ou institutionnelle. Elle reconnaît en outre que les données ouvertes, ouvertes par défaut et souvent entièrement publiques, n'offrent pas à elles seules une solution complète. Dans un paysage des données en constante évolution, pour garantir que cette ouverture serve les objectifs d'intérêt public, il faut se pencher sur la manière dont la reconnaissance et les avantages sont répartis, en particulier lorsque les données accessibles au public peuvent être largement réutilisées sans que les communautés qui les gèrent n'en tirent de bénéfices clairs.

Dans ce modèle réinventé de biens communs de données, des communautés distinctes œuvrant dans l'intérêt public définissent les conditions selon lesquelles les données sont collectées, partagées et (ré)utilisées. L'accès est partagé, mais uniquement selon les conditions spécifiques fixées par des structures de gouvernance qui reflètent les priorités et les contraintes collectives. L'utilisation est souvent limitée à un usage précis, en accord avec les besoins et les attentes des communautés contributrices. De nombreux biens communs de données intègrent également des mécanismes de partage des bénéfices afin de garantir que ceux qui fournissent l'accès aux connaissances ou aux ressources profitent des avantages et que la participation ne renforce pas les inégalités existantes.

Au fond, les biens communs de données ne se résument pas à un simple ensemble de données ou à une plateforme de données. Il s'agit de dispositifs institutionnels qui définissent clairement qui détient l'autorité sur les données, qui peut y accéder, dans quelles conditions, à quelles fins, avec quelles garanties et quel contrôle, et quels sont les résultats attendus. Cela inclut des règles claires en matière d'accès, des mécanismes de contrôle et d'application, ainsi que des structures de participation qui vont au-delà de la simple contribution de données pour englober la gouvernance. Elles visent à faire évoluer la perception des données, non plus comme une ressource à exploiter, mais comme un bien géré collectivement pour le bien public.

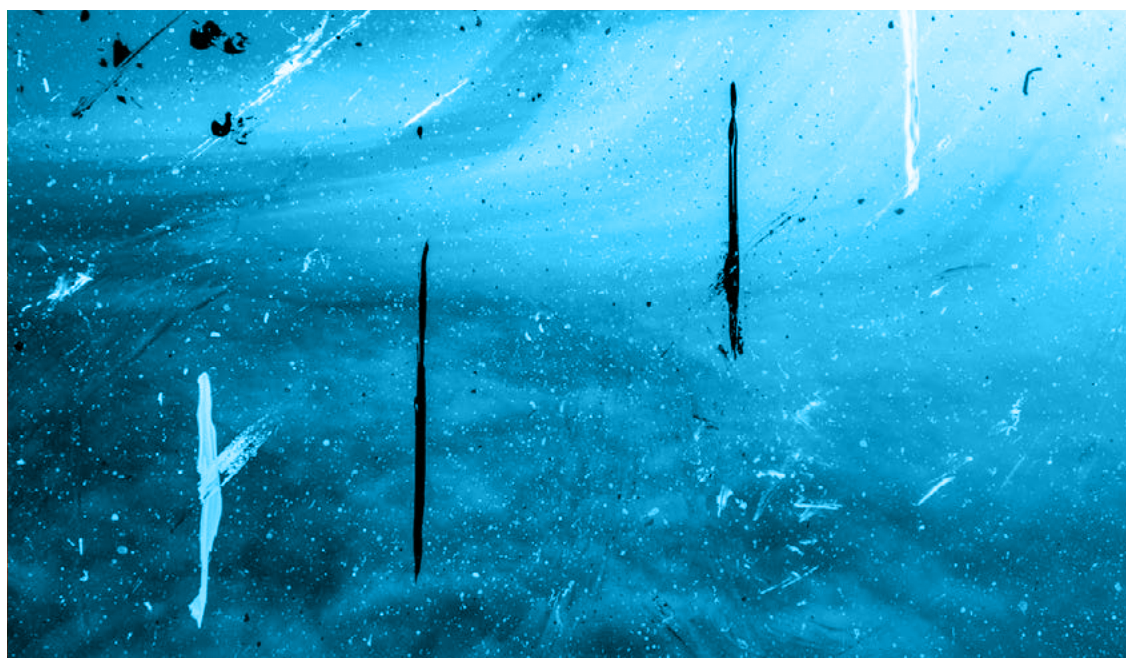
D'où vient l'idée des biens communs ?

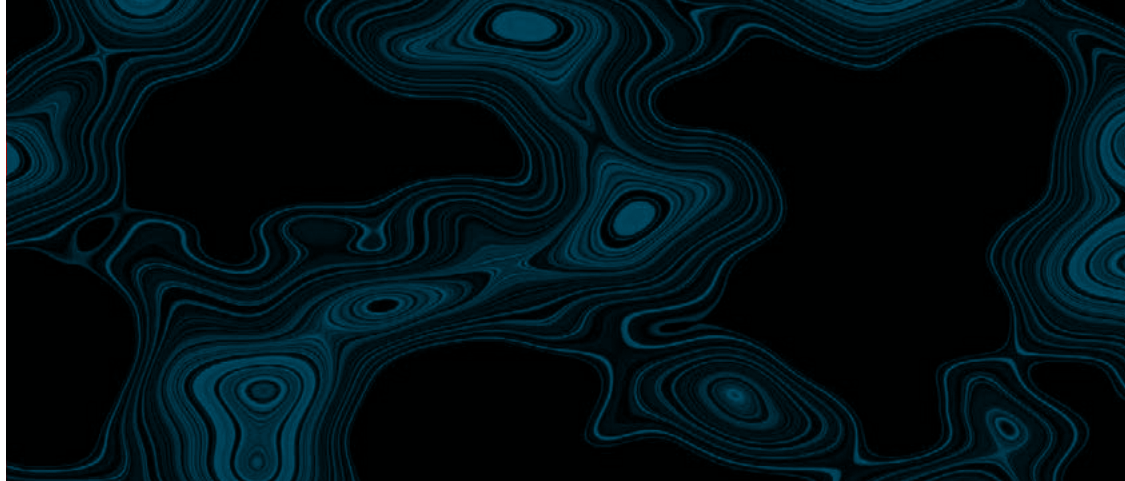
Un [bien commun](#) est une ressource partagée détenue collectivement par une communauté ou un groupe d'utilisateurs, régie par des règles qui en dictent l'utilisation. Les biens communs [incarnent généralement](#) les principes d'autonomie locale, de responsabilité mutuelle et de partage des bénéfices. Parmi les exemples typiques de biens communs, on peut citer les forêts, les pêcheries et les nappes phréatiques.

Le concept de « biens communs » existe depuis des siècles, mais il a été popularisé à l'époque moderne dans le cadre de l'ouvrage de Garrett Hardin intitulé « [La tragédie des biens communs](#) ». S'appuyant sur les arguments de l'économiste du XIXe siècle William Forester Lloyd, Hardin a soutenu que les individus agissant dans leur propre intérêt agissent à l'encontre du bien commun, et que, mus par leur intérêt personnel, ils sont pris au piège dans un comportement qui nuit au bien collectif.

Les travaux de Lloyd et Hardin, qui portaient notamment sur la surpopulation et la rareté des ressources, ont été largement réfutés par Elinor Ostrom, économiste et politologue lauréate du prix Nobel. Ses travaux, en particulier son ouvrage [Governing the Commons](#), démontrent que les êtres humains ne sont ni pris au piège ni impuissants, et que la « tragédie » liée à la gestion des ressources communes n'est pas inévitable, mais qu'il est possible de la surmonter grâce à la coopération, à la surveillance et au respect des règles.

Bien qu'Ostrom ait écrit sur la [connaissance](#) en tant que [ressource commune](#) dès le début des années 2000, le concept de « biens communs de données » n'est apparu qu'au cours des années 2010. Dans les domaines biomédical et de la recherche, les biens communs de données étaient considérés comme un moyen de [contrôler](#) les données scientifiques. En 2019, la directrice de la Bibliothèque nationale de médecine, Patti Brennan, a [décrit](#) comment les biens communs de données s'imposaient comme un moyen important de partage des données ainsi que comme un « ensemble de principes, de stratégies de gouvernance et d'utilités ». Une tentative récente d'appliquer ce concept aux données a été menée par la [Fondation Mozilla](#).





Comment les biens communs sont-ils gérés ?

Les biens communs peuvent prendre de nombreuses formes et être utilisés par les communautés pour gérer toutes sortes de ressources partagées. Nos recherches s'appuient sur les [*huit principes d'Elinor Ostrom pour la gestion des biens communs*](#). Les voici :

- ▶ **Des limites clairement définies :** l'identité du groupe et les limites de la ressource commune sont clairement définies.
- ▶ **Adéquation entre les besoins locaux et les conditions :** les membres de la communauté négocient un système qui récompense les membres pour leurs contributions. Les avantages disproportionnés doivent se mériter. Une inégalité injuste sape les efforts.
- ▶ **Mécanismes de prise de décision collective :** les membres du groupe établissent les règles régissant le bien commun et prennent leurs décisions par consensus. Même en cas de désaccords individuels, tous s'alignent sur les objectifs généraux du groupe.
- ▶ **Droits reconnus d'organisation :** les groupes disposent de leur propre autorité pour mener et superviser leurs propres affaires. Ils ne sont pas contraints par des règles imposées de l'extérieur qui ne peuvent s'adapter aux circonstances locales.
- ▶ **Surveillance efficace :** pour éviter le parasitisme ou l'exploitation active, le bien commun surveille les activités afin de s'assurer que tous les membres respectent les règles convenues. Idéalement, cette surveillance est peu coûteuse et permet d'identifier rapidement les violations des normes.
- ▶ **Sanctions progressives :** les violations des règles entraînent des conséquences de plus en plus sévères.
- ▶ **Résolution accessible des conflits :** le groupe s'est mis d'accord sur des mécanismes permettant de résoudre les litiges rapidement et d'une manière largement perçue comme équitable par le groupe.
- ▶ **Entreprises imbriquées structurées :** la responsabilité de la gestion de la ressource commune a été intégrée dans des niveaux imbriqués. La gestion des ressources est assurée par défaut au niveau le plus bas possible, mais peut être reliée à un système interconnecté plus vaste.

Tous les biens communs (de données) respectent-ils pleinement les huit principes d'Elinor Ostrom ?

Les huit principes d'Ostrom peuvent s'appliquer à divers secteurs. Cependant, ils restent souvent au stade de l'idéal. Dans la pratique, de nombreuses initiatives ne parviennent pas à respecter pleinement chacune de ces normes, notamment en ce qui concerne la prise de décision collective, l'autorité communautaire et les mécanismes d'application.

Plusieurs biens communs conçus pour les données fonctionnent comme des infrastructures de partage de données ou des plateformes collaboratives gérées par des gouvernements et d'autres institutions qui prétendent agir au nom du public et ne sont pas autonomes au sens où l'entendent les chercheurs. Elles peuvent permettre la participation à la contribution de données et offrir un accès structuré, mais de nombreux exemples existants ne vont pas jusqu'à accorder pleinement le contrôle aux communautés.

Qu'est-ce qui caractérise un bien commun de données, par opposition à un « commons » ?

Les biens communs de données présentent l'ensemble de caractéristiques suivantes, couramment mises en œuvre :

Objectif d'intérêt public : vise à fournir les données nécessaires pour résoudre des problèmes d'intérêt public et développer une IA d'intérêt public.

Gouvernance participative : dispose d'une structure de gouvernance offrant aux membres du bien commun des moyens concrets d'exercer leur influence sur l'utilisation de leurs données.

Mécanismes de responsabilité : assure un suivi continu de l'application des règles du bien commun. Dispose de systèmes permettant de gérer les litiges et de demander des comptes à ceux qui ne respectent pas les règles.

Avantages pour les contributeurs : Offre des avantages clairs aux parties prenantes qui ont contribué à la communauté. Cela peut inclure des avantages financiers, un accès anticipé aux outils d'IA ou aux recherches développés, la mention de leur contribution, des informations les concernant, une adhésion gratuite, l'accès à des infrastructures pour développer de nouveaux outils d'IA, etc.

Contrôles d'accès : comprend des règles clairement définies régissant comment, quand et par qui les ensembles de données peuvent être consultés et utilisés. Cela peut impliquer des cadres standardisés tels que les licences Creative Commons, ainsi que des modèles d'accès à plusieurs niveaux, restreints ou basés sur le consentement.

Financement et soutien : le patrimoine de données est souvent financé en tant que bien public, souvent avec le soutien du gouvernement, de la philanthropie ou via un modèle d'adhésion/coopératif.

En quoi les biens communs de données se distinguent-ils des « data collaboratives », des « trusts », des « spaces » et d'autres types d'initiatives ?

Le paysage actuel des données a vu apparaître [une multitude de modèles de gouvernance](#). Ces concepts sont souvent considérés comme des alternatives concurrentes, autant de voies différentes menant au même objectif. En réalité, un tel cadre occulte le fait que chaque modèle répond à des défis de gouvernance distincts et intègre des modèles différents en matière d'autorité, de participation, de répartition des valeurs et de gestion des risques. Sept approches se sont particulièrement imposées, chacune répondant à un défi de gouvernance spécifique auquel sont confrontés les partenariats de réutilisation des données :

- ▶ **Les intermédiaires de données**, qui jouent le rôle de courtiers afin de réduire les coûts de transaction et de faciliter des échanges fiables entre les détenteurs et les utilisateurs de données. Ils sont particulièrement utiles pour remédier aux **défaillances de coordination** en proposant des services de courtage et une réduction des coûts de transaction grâce à des règles négociées et à une intermédiation opérationnelle.
- ▶ **Les syndicats ou coalitions de données**, qui regroupent les participants afin de renforcer le pouvoir de négociation collective sur les marchés des données. Ils permettent de résoudre les **asymétries de négociation** en rééquilibrant les rapports de force et en définissant clairement les conditions d'utilisation des données.
- ▶ **Les fiducies de données**, qui délèguent le pouvoir décisionnel à des fiduciaires indépendants agissant au nom de bénéficiaires définis. Ces fiducies remédient aux **déficits de légitimité** en créant une gestion fiduciaire indépendante qui garantit la neutralité et des engagements crédibles.
- ▶ **Les coopératives de données**, qui intègrent directement la propriété démocratique et le contrôle des membres dans les dispositifs de gouvernance des données. Les coopératives remédient aux **inégalités de propriété** en permettant une propriété démocratique et en donnant aux membres la possibilité de redistribuer le contrôle et la valeur.
- ▶ **Les bacs à sable de données**, qui créent des environnements expérimentaux contrôlés pour tester les règles de gouvernance et les utilisations des données avant leur déploiement à grande échelle. Les bacs à sable de données permettent de remédier à l'**incertitude liée aux systèmes** en offrant un environnement permettant de tester la gouvernance, les règles, les mesures de protection et la viabilité institutionnelle.
- ▶ Les espaces de données, qui fournissent une infrastructure fédérée et des normes communes pour permettre l'interopérabilité entre les secteurs ou les juridictions. Les espaces de données contribuent à résoudre la complexité liée à la mise à l'échelle en proposant des systèmes de coordination fédérée et en favorisant l'interopérabilité entre les acteurs et les juridictions.

- ▶ Les biens communs de données, qui établissent des structures de gouvernance participative pour la gestion partagée d'ensembles de données mutualisés. Les biens communs de données répondent aux besoins de gouvernance collective en offrant des environnements propices à l'élaboration participative de règles et à la gestion partagée au bénéfice de tous.

Un bien commun de données peut coexister avec l'un de ces autres modèles ou en découler (par exemple, il peut évoluer à partir d'un « sandbox », être géré par un intermédiaire ou se transformer en « data space »), mais il s'en distingue.

En quoi un bien commun de données se distingue-t-il des données ouvertes ?

Le modèle des biens communs de données vise à relever le défi des asymétries d'accès (décrites ci-dessous), qui se manifestent lorsque des organisations rendent leurs données largement accessibles, ce qui conduit à leur collecte et à leur extraction par des tiers sans contrepartie.

Si les biens communs de données **permettent** de mettre les données à la disposition du grand public, leur structure permet à leurs organisateurs de déterminer quelles données ils mettent à disposition, à quelles fins et sous quelles conditions. Au lieu d'imposer un modèle « ouvert par défaut », les biens communs de données allient accessibilité et gouvernance.

Ils offrent des moyens structurés de déterminer et de fournir l'accès aux données — publiques, privées ou hybrides — tout en précisant les conditions d'utilisation, les obligations de partage des bénéficiaires, les mécanismes de contrôle et les systèmes d'attribution et de rémunération.

Ils cherchent à repenser l'ouverture à une époque où le contexte, l'acceptation sociale et le contrôle importent autant que la disponibilité.

9

Quels problèmes les biens communs de données visent-ils à résoudre ?

Les biens communs de données visent à remédier à plusieurs défaillances structurelles :

Asymétries d'accès : De nombreuses grandes organisations cherchent à utiliser les données ouvertes pour tirer profit des données d'autrui tout en limitant l'accès à leurs propres données propriétaires. Les biens communs de données peuvent mettre en place des mécanismes permettant aux communautés d'influencer qui accède à leurs données, comment et dans quel but.

Défaillances de coordination : Le partage de données avec un seul acteur ou de manière ponctuelle entraîne souvent des coûts de transaction élevés (par exemple, l'élaboration d'accords de partage de données sur mesure ou de normes). Les biens communs de données peuvent créer des processus plus standardisés pour le partage des données selon des règles cohérentes.

Manque de confiance : dans de nombreux accords de partage de données, les intérêts et les attentes du public ou des communautés contributrices ne sont pas pris en compte de manière significative. Cela sape l'acceptabilité sociale et limite la légitimité de toute (ré) utilisation des données. Les biens communs de données s'efforcent de placer ces préoccupations au centre de leurs préoccupations par le biais de la gouvernance, de la supervision et des conditions d'accès.

Sous-utilisation : l'un des problèmes qui entravent la réutilisation des données est que les ensembles de données les plus précieux ne sont pas utilisés pour le bien public, mais font l'objet de restrictions. Les biens communs de données peuvent élargir l'accès à un éventail plus large d'acteurs, à condition que ceux-ci s'engagent à respecter les normes définies par les organisateurs.

À cet égard, les biens communs de données peuvent ouvrir la voie à des relations de partage de données plus durables et plus équitables, au service de l'intérêt général. Ils peuvent répondre à certaines des préoccupations qui alimentent la pénurie de données et l'actuel « hiver des données ». Ils peuvent recentrer le débat sur l'IA afin de mettre l'accent sur l'IA d'intérêt public et la prise de décision.

Quelles sont certaines des critiques formulées à l'encontre des biens communs de données ?

10

Les critiques à l'égard des biens communs de données sont importantes — et nécessaires. Comme le soulignent notamment [Tommaso Fia et Gijs van Maanen](#), certains débats autour des biens communs de données recourent au discours de la « communauté » sans préciser suffisamment ce que ce terme recouvre réellement, qui le définit, qui en bénéficie et qui a la capacité d'y participer. Ils mettent également en garde contre le « commonswashing » et le « community-washing », où la rhétorique de la gouvernance collective est principalement utilisée à des fins de réputation sans redistribution significative du pouvoir ou des avantages.

Ces critiques doivent être prises au sérieux. Il existe en effet un risque que certaines initiatives de biens communs de données deviennent trop technocratiques, abstraites ou « solutionnistes » — traitant la gouvernance comme un problème d'architecture technique tout en ignorant les asymétries politiques, institutionnelles et économiques plus profondes. On observe également, dans certains débats politiques, une tendance à dresser de longues listes de modèles de gouvernance (voir ci-dessus) — trusts de données, coopératives de données, biens communs de données, espaces de données, intermédiaires — comme s'il s'agissait d'options interchangeables d'un menu, sans interroger suffisamment les valeurs, les conditions institutionnelles ou les structures de pouvoir qui les sous-tendent.

Dans le même temps, on observe également une tendance croissante chez certains critiques à adopter une sorte de purisme en matière de gouvernance : rejeter toute expérimentation en matière d'innovation institutionnelle à moins que des conditions parfaites d'équité,

de représentation ou de légitimité démocratique n'existent déjà. Si les préoccupations qui motivent cette prudence sont compréhensibles, une telle approche risque d'entraîner une paralysie précisément à un moment où les systèmes de marché et de régulation existants échouent manifestement à gouverner les données et l'IA de manière à produire des bénéfices sociétaux partagés.

La réalité est que bon nombre des préjudices actuels liés au développement extractif de l'IA et à l'accès asymétrique aux données sont eux-mêmes le résultat d'une innovation insuffisante en matière de gouvernance institutionnelle. Se contenter de s'appuyer sur les structures réglementaires existantes ou les incitations du marché ne suffira pas pour relever les défis liés à l'accès aux données d'intérêt public, au partage des bénéfices, à la gouvernance participative ou à l'action collective à l'ère de l'IA.

Les biens communs de données ne doivent donc pas être considérés comme une solution miracle ou un substitut à la réglementation. Ils ne doivent pas non plus être idéalisés comme étant intrinsèquement démocratiques ou équitables. Ils doivent plutôt être considérés comme un instrument de gouvernance parmi d'autres dans une boîte à outils institutionnelle plus large, aux côtés de la surveillance publique, des mécanismes fiduciaires, de la gouvernance participative et des infrastructures publiques.

Ce qui importe, ce n'est pas de savoir si une initiative se qualifie elle-même de « bien commun de données », mais si elle fait réellement progresser :

- ▶ une gestion collective de la gouvernance des données ;
- ▶ une répartition plus équitable des bénéfices ;
- ▶ la transparence et la responsabilité ;
- ▶ une prise de décision participative ;
- ▶ une gestion durable des ressources communes ; et
- ▶ des mécanismes visant à contrer les dynamiques d'exploitation.

Concrètement, cela implique d'être très précis sur plusieurs points :

- ▶ Cette communauté existe-t-elle réellement, au-delà des discours ?
- ▶ Qui définit les critères d'adhésion et les règles de gouvernance ?
- ▶ De quelles ressources et capacités les participants disposent-ils pour assurer une gouvernance efficace ?
- ▶ Existe-t-il des mécanismes permettant de remédier aux déséquilibres existants en matière de pouvoir et d'expertise ?
- ▶ Qui tire profit de cette collaboration ?
- ▶ Quelles sont les formes de responsabilité, de contestation et de sortie qui existent ?
- ▶ Les résultats obtenus améliorent-ils concrètement l'équité, l'accès ou la valeur publique ?

Il est important de noter que les réponses à ces questions varient en fonction du contexte. Un patrimoine bien commun de données fondé sur la souveraineté des peuples autochtones en matière de données, par exemple, soulève des enjeux de gouvernance fondamentalement différents de ceux d'un patrimoine bien commun de données sur la mobilité destiné à l'urbanisme ou d'un projet collaboratif de données de santé à des fins de recherche. Comme le soulignent Fia et van Maanen, les communautés se distinguent par leur histoire, leur identité, leurs liens sociaux, leurs capacités de gouvernance et leurs aspirations politiques.

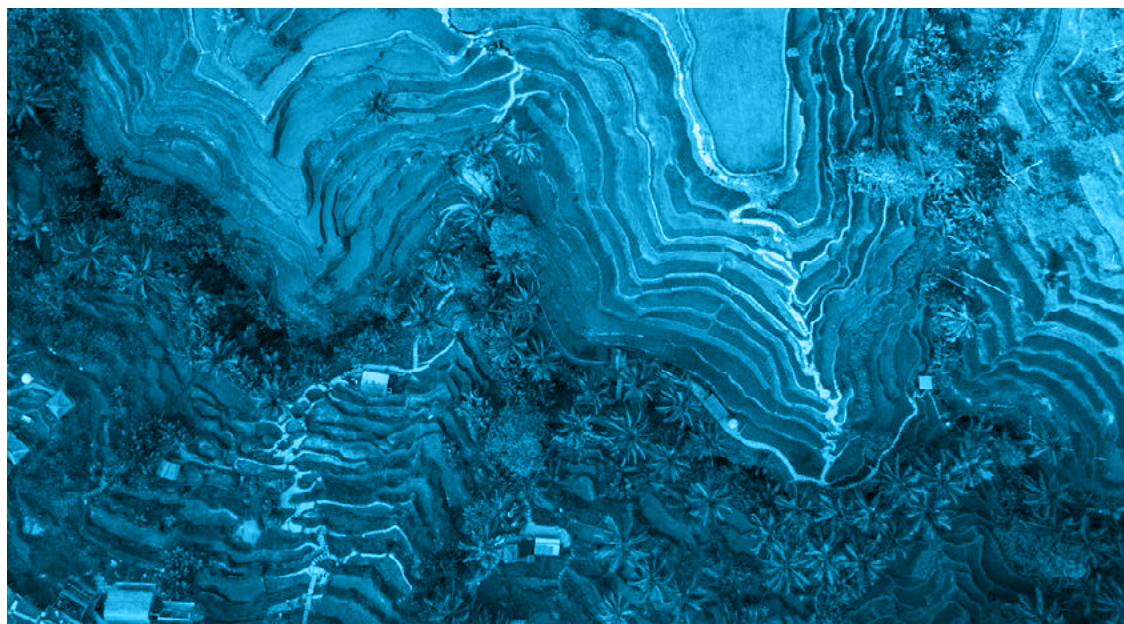
L'objectif n'est donc pas d'imposer un modèle de gouvernance universel, mais de développer des dispositifs de gouvernance adaptés au contexte, ancrés dans les institutions et capables de générer des avantages partagés légitimes. Dans de nombreux cas, cela nécessitera de combiner des approches fondées sur les communs avec une réglementation plus stricte, des institutions publiques, des fonctions de gestion des données et des mécanismes participatifs.

En fin de compte, la critique des biens communs de données ne devrait pas nous conduire à abandonner l'expérimentation de la gouvernance collective. Elle devrait nous pousser à concevoir ces systèmes de manière plus rigoureuse, plus transparente et plus honnête sur le plan politique. L'alternative – laisser la gouvernance des données entièrement aux mains des marchés ou d'une autorité étatique centralisée – a déjà montré ses limites.

Pourquoi l'adhésion sociale est-elle essentielle pour les biens communs de données ?

12

Une [licence sociale](#) est une acceptation implicite d'une activité par la société, fondée sur la confiance, la légitimité et la perception d'une adhésion aux valeurs de la communauté. Elle se distingue des licences de données ouvertes en ce qu'elle n'est pas de nature juridique, mais constitue plutôt un accord ou un plan élaboré grâce à l'engagement et à la collaboration des parties prenantes. Dans le contexte des données, elle traduit le consentement collectif en mettant l'accent sur l'approbation tacite des individus et des communautés concernant la collecte, l'utilisation et la réutilisation de leurs données.





Le concept de licence sociale reconnaît que la conformité juridique et le consentement sont nécessaires, mais pas suffisants. Alors que les régimes juridiques existants recherchent l'accord individuel au moment de la collecte des données, une licence sociale vise à créer un processus permettant aux communautés de façonner et de superviser la réutilisation continue des données.

Une licence sociale garantit que les données reflètent les intérêts collectifs. Cela permet ensuite de considérer l'utilisation des données comme légitime et de mettre en place des processus qui assurent une responsabilité continue. Au lieu d'un consentement ponctuel — un simple processus d'adhésion ou de désadhésion —, la licence sociale encourage les acteurs à penser à long terme, à réexaminer périodiquement les besoins et les attentes afin de s'assurer que toutes les activités sont toujours largement approuvées par les parties prenantes de la communauté.

Une licence sociale peut s'avérer cruciale lorsqu'il s'agit de gérer d'importantes disparités de pouvoir entre les acteurs, disparités qui créent une vulnérabilité ou un risque d'exploitation. Elle peut également être utile pour orienter les efforts non pas vers des besoins à court terme (par exemple, cet acteur approuve-t-il ce travail à l'instant présent ?), mais vers des intérêts à long terme (par exemple, que faut-il pour garantir une implication continue ?).

13

Quel est le rôle de la gestion des données dans les biens communs de données ?

Les gestionnaires de données sont des responsables de données mandatés par leurs organisations pour créer de la valeur publique grâce à des échanges de données intersectoriels. Ils constituent le pilier opérationnel des biens communs de données, contribuant ainsi à faciliter et à institutionnaliser ce dispositif.

En général, les gestionnaires de données interviennent pour faciliter l'accès aux données et négocier les conditions de leur échange. Ils veillent également au respect des règles et à l'utilisation éthique des données, aident à faire valoir les intérêts des différents fournisseurs et utilisateurs de données, et s'efforcent de maintenir la confiance entre les parties au fil du temps.

Existe-t-il déjà des biens communs de données ?

Les biens communs de données constituent une pratique émergente. Peu d'initiatives existantes répondent pleinement à la définition de systèmes gérés de manière collaborative, dans lesquels les communautés exercent un pouvoir sur les données. Au contraire, les exemples actuels couvrent un large éventail de modèles qui intègrent certaines caractéristiques propres aux biens communs, telles que l'accès partagé, la participation multipartite ou l'orientation vers l'intérêt public, sans pour autant mettre toujours en œuvre une gouvernance collective.

Les initiatives ci-dessous illustrent cette diversité :

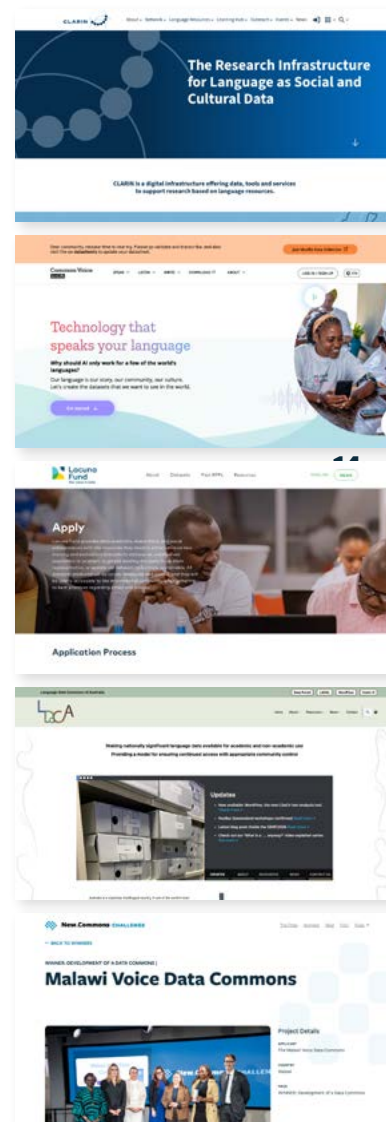
CLARIN : CLARIN fournit l'infrastructure nécessaire pour regrouper les données linguistiques provenant de diverses institutions européennes à des fins de recherche. La gouvernance est multipartite au niveau institutionnel, mais le pouvoir décisionnel est exercé par les États membres et les organisations plutôt que par l'ensemble de la communauté contributrice.

Common Voice : Common Voice est un ensemble de données vocales ouvertes issues du crowdsourcing qui peut être utilisé pour entraîner des applications vocales basées sur l'IA. L'initiative vise à élargir l'accès aux données vocales pour les langues non anglophones et d'autres groupes généralement sous-représentés dans les ensembles de données vocales. L'initiative est dirigée par la Fondation Mozilla.

Lacuna Fund : Le Lacuna Fund soutient le développement d'ensembles de données étiquetées sur les pays à revenu faible et intermédiaire pouvant être utilisés pour l'apprentissage automatique. L'équipe accepte les propositions de création de jeux de données dans quatre domaines : l'agriculture, les langues, la santé et le climat.

Language Data Commons of Australia : Le Language Data Commons of Australia est un partenariat entre l'Australian Research Data Commons, la School of Languages and Cultures de l'université du Queensland et d'autres partenaires (par exemple [First Languages Australia](#)) qui vise à mettre à disposition des données linguistiques australiennes à des fins « académiques et non académiques ».

Querido Diário : Open Knowledge Brasil cherche à développer des biens communs de données pour les journaux officiels municipaux. Sa plateforme Querido Diário vise à faciliter l'accès à ces ensembles de données et à aider à analyser les informations qu'ils contiennent. L'objectif est d'améliorer la prise de décision au niveau local au Brésil.



Malawi Voice Data Commons : le Malawi Voice Data Commons du Programme de recherche et d'éducation pour la paix de l'Université de New York (NYU), développé en collaboration avec Ushahidi, le PNUD et la Fondation Mozilla, permet aux habitants des zones rurales du Malawi de signaler des situations d'urgence dans leur langue maternelle, créant ainsi des ensembles de données multilingues et compatibles avec l'IA pour les interventions humanitaires et la préservation des langues. Le projet pilote se déroulera au Malawi et devrait ensuite être étendu à l'Afrique subsaharienne.



Quels sont les modèles de financement utilisés pour soutenir les biens communs de données ?

La recherche dans ce domaine en est encore à ses débuts. Nous avons toutefois identifié quelques modèles de financement récurrents :

Modèles pilotés par les pouvoirs publics : les biens communs de données sont soutenus par des organismes publics, généralement en partenariat avec d'autres entités. Ces modèles bénéficient d'engagements de financement à long terme.

Modèles philanthropiques : les biens communs de données sont soutenus par des donateurs issus du secteur privé et de la société civile. Ces communs associent les capitaux des donateurs à la participation communautaire, s'inscrivant ainsi dans le modèle de bénéfice indirect identifié dans la littérature : la valeur réside dans la réalisation de la mission et l'engagement de l'écosystème plutôt que dans la monétisation.

Modèles d'adhésion et coopératifs : les utilisateurs ou contributeurs (y compris les gouvernements) deviennent actionnaires ou membres, devenant ainsi copropriétaires des communs. Ces modèles intègrent la gouvernance et la durabilité par le biais de la propriété collective, faisant écho aux archétypes commerciaux du « freemium » et de la « lame de rasoir », où l'accès est libre mais où les services, outils ou privilèges de gouvernance supplémentaires nécessitent une contribution ou des frais.



Pourquoi est-ce important pour l'intelligence artificielle ?

Les biens communs de données peuvent s'avérer précieux à l'ère de l'intelligence artificielle. Ils constituent une infrastructure essentielle pour l'accès aux données et leur réutilisation, et permettent de garantir que cette utilisation soit conforme aux intérêts de la communauté. Nous décrivons cela dans notre [plan d'action](#), qui souligne que les biens communs de données peuvent être utiles au développement de l'IA dans la mesure où ils permettent :

- ▶ Rassembler les ensembles de données disparates et variés nécessaires au développement et au déploiement de l'IA ;
- ▶ Fournir l'infrastructure nécessaire pour normaliser les données dans des formats compatibles avec l'IA ;
- ▶ Réduire les coûts de normalisation et contribuer à éviter la duplication des efforts ;
- ▶ Accroître la diversité culturelle des ensembles de données disponibles ;
- ▶ Mettre en œuvre de nouveaux modèles de réutilisation des données pour le bien commun ;
- ▶ Mettre en place des mécanismes de traçabilité des données qui améliorent la transparence et la traçabilité au sein des systèmes d'IA ; et
- ▶ Aider les communautés de développeurs à créer des applications d'IA qui répondent à des besoins réels.

The Weaponisation of Openness? Toward a New Social Contract for Data in the AI Era

By Stefaan G. Verhulst

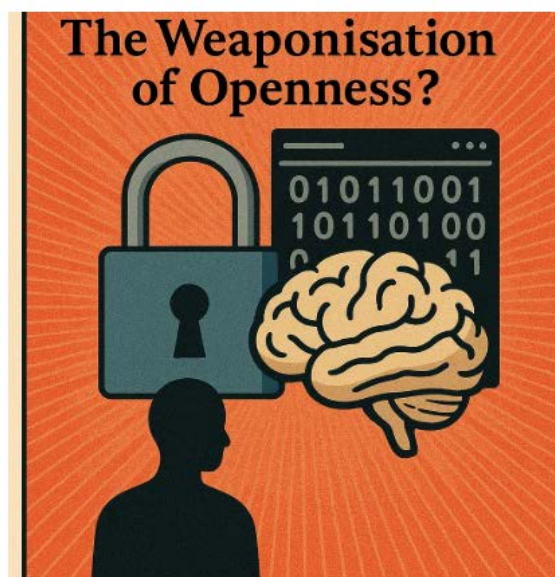


Stefaan G. Verhulst

Follow

5 min read · Oct 21, 2025

150



(Dall-E Created)

Pourquoi ce modèle est-il adapté aux groupes marginalisés ?

L'un des principaux problèmes liés au développement de l'IA et d'autres outils basés sur les données réside dans leur caractère potentiellement extractif. Les données mises à disposition pour servir une communauté sont collectées à grande échelle afin d'entraîner des systèmes propriétaires, sans que l'on se soucie guère de reconnaître ou de rémunérer ces communautés. Il en résulte que quelques grandes organisations s'enrichissent encore davantage, ce qui accentue les asymétries de pouvoir, tandis que les individus n'ont pratiquement aucun contrôle sur l'utilisation qui est faite des informations les concernant, laquelle va à l'encontre de leurs intérêts ou de leurs souhaits explicites.

Nous appelons ce problème « *l'instrumentalisation de l'ouverture* ». Il s'agit d'un problème bien connu de nombreuses communautés autochtones qui ont vu leurs savoirs traditionnels et leur biodiversité exploités par des laboratoires pharmaceutiques sans aucune reconnaissance ni contrepartie.

Les biens communs de données offrent un cadre permettant de concilier accessibilité et gouvernance. Ils peuvent fournir des moyens structurés d'accéder aux données — qu'elles soient publiques, privées ou hybrides — tout en précisant les conditions d'utilisation, les obligations en matière de partage des avantages, les mécanismes de contrôle et les systèmes d'attribution et de compensation. Cela permet de garantir que les données sont utilisées d'une manière conforme aux attentes, aux besoins et aux intérêts d'une communauté.

Cependant, ce n'est pas une panacée. Cela doit s'inscrire dans un environnement où la communauté est bien définie, dispose de ressources pouvant être mobilisées pour protéger ses intérêts, et où un consensus est possible. Les biens communs de données pourraient bénéficier d'un cadre réglementaire plus large, spécialement conçu pour préserver les intérêts collectifs (par exemple, des mécanismes conventionnels pour les communautés autochtones) face à des entités extractives plus puissantes.

Lectures complémentaires

Chafetz, Hannah, Andrew J. Zahuranec, and Stefaan Verhulst. 2025a. “10 Data Commons for Cultural Knowledge and Preservation.” *10 Data Commons for Cultural Knowledge and Preservation*, November 24. <https://opendatapolicylab.org//articles/blog-post-10-data-commons-for-cultural-knowledge-and-preservation/>.

Chafetz, Hannah, Andrew J. Zahuranec, and Stefaan Verhulst. 2025b. *A Blueprint to Unlock New Data Commons for AI*. The Open Data Policy Lab. <https://incubator.opendatapolicylab.org/files/data-commons-for-ai-blueprint.pdf>.

Chafetz, Hannah, Andrew J. Zahuranec, and Stefaan Verhulst. 2025c. “Appendix B: Taxonomy of Data Commons Use Cases for AI.” Brooklyn, New York, March. <https://incubator.opendatapolicylab.org/files/appendix%20b.pdf>.

NYU Tandon. 2025. “The New Commons Challenge Proves the Power of Data Collectives | NYU Tandon School of Engineering.” Brooklyn, New York, October 20. <https://engineering.nyu.edu/news/new-commons-challenge-proves-power-data-collectives>.

Verhulst, Stefaan, Hannah Chafetz, and Andrew Zahuranec. 2025. “Emerging Funding and Business Models for Data Commons: A Comparison.” *Open Data Policy Lab*, October 29. <https://opendatapolicylab.org//articles/blog-post-emerging-funding-and-business-models-for-data-commons-a-comparison/>.

Verhulst, Stefaan, Hannah Chafetz, and Andrew J. Zahuranec. 2024. “Data Commons: Under Threat by or The Solution for a Generative AI Era? Rethinking Data Access and Re-Use.” SSRN Scholarly Paper No. 4836354. Social Science Research Network, May 21. <https://doi.org/10.2139/ssrn.4836354>.

Verhulst, Stefaan, Andrew J. Zahuranec, and Adam Zable. 2025. “Participatory Approaches to Responsible Data Reuse and Establishing a Social License.” In *Global Public Goods Communication: Mapping Actors, Policies, and Narratives*, edited by Sónia Pedro Sebastião and Anne-Marie Cotton. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-90667-1_10.

Verhulst, Stefaan, Andrew J. Zahuranec, Adam Zable, and Peter Addo. 2025. “Reimagining Data Governance for AI: Operationalizing a Social License for Data Reuse.” SSRN Scholarly Paper No. 5255677. Social Science Research Network, April 28. <https://doi.org/10.2139/ssrn.5255677>.

Verhulst, Stefaan, Burton Davis, and Andrew Schroeder. 2025. “Data Commons: The Missing Infrastructure for Public Interest Artificial Intelligence.” *LinkedIn*, April 29. <https://www.linkedin.com/pulse/data-commons-missing-infrastructure-public-interest-verhulst-phd-k8eec/>.

Zahuranec, Andrew J., Hannah Chafetz, and Verhulst Stefaan. 2025. “Why Responsible Data Access Will Determine the Future of AI: The Increased Importance of Data Commons.” *Open Data Policy Lab*, February 8. <https://opendatapolicylab.org//articles/blog-post-why-responsible-data-access-will-determine-the-future-of-ai-the-increased-importance-of-data-commons/>.

Biens communs de données

Foire aux questions