

Contribution à la consultation publique de la Commission européenne sur une Stratégie pour une union européenne des données

Arcep – Juillet 2025

AVERTISSEMENT : La version originale anglaise, disponible [ici](#), fait foi. Le présent document en est une traduction en français.

Les données sont au cœur du développement de l'IA. Au cours des cinq dernières années, les cas d'usage d'IA se sont considérablement développés, avec les modèles conversationnels en première ligne. Cette évolution a été rendue possible grâce à des jeux de données accessibles, tels que Common Crawl, et grâce à des investissements importants dans l'annotation, le nettoyage et le prétraitement des données par les acteurs de l'IA. Aujourd'hui, les grands modèles de langage (LLM) sont capables de résoudre des problèmes et des exercices de niveau universitaire, avec des performances qui s'améliorent chaque mois, tandis que les agents basés sur les LLM remplacent les assistants virtuels traditionnels. En parallèle, l'IA peut renforcer les cas d'usage industriels (par exemple, la détection d'anomalies, la découverte de molécules, ou encore l'aide au diagnostic médical). Pour que l'économie européenne puisse exploiter pleinement le potentiel de l'IA et atteindre les objectifs de la décennie numérique, l'accès à de grandes quantités de données de haute qualité est essentiel¹.

Ces orientations ont été exprimées par l'Arcep dans sa [réponse](#) à la consultation publique de la Commission européenne sur la concurrence dans le domaine de l'IA générative : outre la puissance de calcul du cloud, les talents et les capitaux, les données constituent un intrant essentiel pour le développement de l'IA.

À ce jour, l'Arcep estime que le potentiel des données européennes reste inexploité, car leur disponibilité pour l'utilisation et la réutilisation est encore assez faible : les écosystèmes de partage des données sont encore en cours d'émergence, tandis que certaines données restent entre les mains de quelques acteurs, sans contrôle efficace pour les entreprises et les particuliers.

La stratégie européenne pour les données de 2020 a jeté les bases d'un cadre réglementaire et d'investissement visant à remédier à ces problèmes et à contribuer à l'émergence d'un marché unique interconnecté des données. Avec l'entrée en application prochaine du règlement sur les données, qui complète la réglementation établie par le règlement sur la gouvernance des données (Data Governance Act, DGA) et les réglementations sectorielles, leur effet combiné sur les flux de données et l'économie des données devrait être visible à court et moyen terme, favorisant l'innovation et garantissant une concurrence loyale dans l'économie des données, en écho à l'approche réglementaire et à [l'ambition de l'Arcep](#).

Compte tenu des besoins croissants en données pour développer et tirer parti des avantages de l'IA, l'Arcep salue chaleureusement l'initiative de la Commission en faveur d'une nouvelle stratégie européenne pour une union des données, visant notamment à permettre une utilisation et une

¹ Comme souligné dans la [communication sur le thème «Accroître les jeunes pousses et l'innovation dans le domaine de l'intelligence artificielle digne de confiance»](#) de la Commission européenne ainsi que dans la [stratégie de l'UE en matière d'IA](#) (2018).

disponibilité accrues des données. Les principes déjà énoncés par le DGA et le règlement sur les données doivent être mis en œuvre : ils fournissent un cadre stable et prévisible pour les échanges de données. Sur la base de notre expérience en tant que régulateur du DGA, nous estimons que des ajustements mineurs pourraient être nécessaires pour garantir la viabilité économique des fournisseurs de services d'intermédiation de données. L'Arcep est également favorable à des mesures de simplification pour les entreprises, notamment par le biais d'une consolidation juridique, mais sans compromettre les objectifs d'une réglementation prévisible du partage des données.

1 Le cadre réglementaire de l'accès aux données et du partage des données, destiné à libérer les données, notamment pour l'intelligence artificielle, commence à produire ses effets

L'Arcep est l'autorité compétente pour l'application des dispositions relatives aux prestataires de services d'intermédiation de données (PSID) du DGA en France, et a été chargée de réguler les services cloud en prévision de l'entrée en application du règlement sur les données. Cela a permis à l'Arcep de recueillir des informations importantes auprès des écosystèmes du cloud et des données sur les obstacles qui subsistent en matière d'accès aux données et de partage des données.

Les acteurs ont signalé que des obstacles à un marché européen des données efficace et ouvert persistent encore aujourd'hui : premièrement, **la méfiance à l'égard des accords de partage des données et de potentielles utilisations illégitimes des données persiste** ; deuxièmement, **les droits d'accès aux données sont limités** ; troisièmement, **l'incertitude quant à la valeur des données** n'incite guère au partage de celles-ci. En outre, les préoccupations relatives à la souveraineté numérique et aux transferts internationaux de données peuvent également nuire à la confiance et, par conséquent, à la réutilisation des données.

La précédente stratégie pour les données, adoptée en 2020, visait à remédier à ces obstacles :

- avec le statut de PSID introduit par le DGA, qui vise à renforcer la confiance dans l'économie des données grâce à des acteurs neutres chargés du partage des données ;
- en fournissant un cadre de gouvernance et des outils techniques communs via les espaces de données, destinés à accélérer les échanges de données et à permettre l'émergence de solutions d'IA ;
- en imposant, dans le règlement sur les données, des obligations d'accès visant à accroître la disponibilité et la circulation des données, ainsi qu'une interopérabilité effective grâce à la normalisation.

La plupart des acteurs estiment que l'entrée en application imminente du règlement sur les données permettra à de nombreuses entreprises européennes d'accéder plus aisément à leurs données, de les partager et de les monétiser.

Premièrement, le règlement sur les données donnera plus de choix aux utilisateurs et permettra aux PME européennes de concurrencer les grands acteurs dans la fourniture de services connexes.

Deuxièmement, il protégera les entreprises européennes, y compris les PME et les start-ups, contre les pratiques contractuelles déloyales en matière de partage des données.

Troisièmement, il renforcera la confiance et l'interopérabilité au sein des espaces de données grâce à la normalisation.

Plus généralement, les obligations de partage de données ont encouragé et continueront d'encourager la circulation des données dans l'UE.

Les données du secteur public et les données sectorielles peuvent bénéficier de ces obligations : la directive sur les données ouvertes a accru la disponibilité des jeux de données à forte valeur ajoutée, les autorités locales et régionales chargées de la mobilité utilisent les réglementations en matière de données pour mettre à disposition les données nécessaires à l'organisation de services de transport intermodal et à des usages innovants, et le futur espace européen des données de santé permettra aux start-ups de développer des solutions innovantes dans le domaine de la santé. **Ces obligations de partage et ces droits d'accès vont élargir le champ des données accessibles et mettre en circulation de plus grands volumes de données de haute qualité. Par conséquent, l'Arcep estime que l'efficacité des droits d'accès aux données, en particulier ceux prévus par le règlement sur les données, sera cruciale pour entraîner des modèles d'IA innovants, y compris dans le secteur industriel.**

Ces dernières renforceront à leur tour la nécessité de disposer de cadres de gouvernance fiables, tels que les espaces de données, et d'intermédiaires neutres, notamment sous la forme de PSID. À cette fin, **l'Arcep soutient la création de nouveaux espaces de données et, plus largement, d'infrastructures de partage de données lorsque le besoin s'en fait sentir**, et reconnaît l'importance de **renforcer à la fois la gouvernance et l'interconnexion des espaces de données existants**. Ces deux éléments sont essentiels pour garantir l'efficacité de cette approche. À cet égard, l'Arcep salue les efforts déployés par la Commission européenne et les parties prenantes pour faire émerger des espaces de données européens.

Dans l'ensemble, l'Arcep estime que le cadre défini par la précédente stratégie européenne pour les données constitue une base solide, dont la mise en œuvre efficace peut libérer d'énormes quantités de données et garantir un environnement fiable pour le partage des données, renforçant ainsi la position de l'Europe dans le domaine de l'IA.

2 De nouvelles mesures pourraient être adoptées pour favoriser la circulation des données et garantir de nouveaux droits d'accès aux données, afin de libérer davantage de données pour l'IA et l'économie de l'UE.

Tirer parti du cadre du DGA et des PSID pour mettre en place des laboratoires de données pour l'IA.

Pour répondre aux besoins spécifiques des développeurs d'IA, une gouvernance des données fiable et interopérable, notamment grâce à des espaces de données interconnectés (par exemple grâce à des PSID interopérables), pourrait offrir un accès aux données de divers secteurs et promouvoir la confiance et la coopération en garantissant la sécurité ou l'intégrité des données partagées, ainsi que le respect des règles sectorielles. En outre, les besoins des développeurs d'IA vont souvent au-delà de l'accès aux données et englobent également le nettoyage, l'enrichissement, la pseudonymisation ou l'anonymisation des données, ou encore la production de données synthétiques. Sous réserve de quelques clarifications mineures du DGA, notamment sur la possibilité pour les PSID d'enrichir ou d'améliorer les jeux de données, ces structures de gouvernance des données pourraient constituer un modèle pour les laboratoires de données proposés par la Commission dans son plan d'action pour le continent de l'IA.

Encourager le développement d'interfaces afin de faciliter l'accès des PSID et des autres parties prenantes aux données.

Certains intermédiaires de données signalent qu'ils sont parfois confrontés à des difficultés techniques lorsqu'ils accèdent à des données pour le compte d'utilisateurs, par exemple lorsqu'ils exercent des droits à la portabilité, notamment en l'absence d'interfaces. Ces difficultés peuvent entraver la circulation des données et empêcher les intermédiaires de données de réaliser pleinement leur potentiel de manière neutre, sûre et fiable. À ce titre, pour les sources de données importantes,

qu'elles soient privées ou publiques, le développement et la mise à disposition d'interfaces d'accès aux données, telles que des API, pourraient être encouragés.

Étendre les droits de portabilité des services cloud existants aux logiciels sur site afin de libérer davantage de données.

Une faible portabilité peut empêcher les entreprises de développer ou d'utiliser de nouveaux services avec leurs données. En particulier, le secteur des logiciels traditionnels souffre encore parfois d'obstacles injustifiés au changement de fournisseur et au transfert de données. Certaines parties prenantes estiment que cela est préjudiciable, car dans de nombreux secteurs (médecine, agriculture, etc.), l'adoption du cloud est encore en cours, ou les professionnels peuvent préférer des solutions sur site. À l'instar des solutions cloud, les obstacles auxquels les clients sont confrontés lorsqu'ils tentent de changer de fournisseur ou de transférer leurs données peuvent être d'ordre tarifaire, contractuel ou technique. Avec l'application du règlement sur les données à tous les services cloud, le droit à la portabilité pourrait être étendu aux logiciels sur site, ce qui permettrait une plus grande circulation des données.

Aider les entreprises à évaluer la valeur de leurs données afin d'encourager plus avant le partage de données.

Dans le cas général, certaines entreprises déclarent investir principalement dans des données brutes, collectées ou achetées, afin de les transformer en jeux de données de haute qualité disponibles pour des usages internes ou pour l'entraînement de modèles d'IA. Cependant, l'évaluation de la valeur des jeux de données de haute qualité peut s'avérer difficile, en particulier pour les PME et les start-ups. Cela entraîne une réticence à vendre ou à partager les données, car la valeur créée grâce à la mise en commun de données provenant de diverses sources peut également s'avérer difficile à évaluer. En dehors de l'Union européenne, certaines mesures ont été prises pour inciter les entreprises à mieux évaluer la valeur de leurs données². Face à ces évolutions, le développement de telles solutions en Europe pourrait aider les entreprises à s'engager dans des projets innovants nécessitant l'utilisation conjointe de données dispersées ou la production de nouvelles données de haute qualité.

Garantir la contestabilité sur le marché de l'IA grâce à un accès FRAND aux données d'utilisation des grands acteurs.

L'Arcep estime que l'ouverture des marchés et la contestabilité sont essentielles pour contribuer d'une part à la disponibilité d'alternatives européennes pour la transition numérique, et d'autre part à l'autonomie stratégique numérique de l'UE. À cet égard, un accès inégal aux données liées au développement de l'IA générative pourrait constituer une nouvelle source de fermeture du marché. Les services d'IA à usage général peuvent être intégrés dans des moteurs de recherche ou utilisés comme applications autonomes. Ces services sont améliorés en continu grâce aux données et aux retours fournis par les utilisateurs : plus il y a de données, plus le service est performant. Cela renforce encore la concentration sur le marché des moteurs de recherche, et la même dynamique de concentration observée avec les moteurs de recherche pourrait se manifester sur le marché de l'IA, menant à une situation de « *winner takes all* ». Afin de garantir la contestabilité dans le secteur de l'IA à usage général, des conditions d'accès FRAND pourraient être implémentées pour éviter la concentration de ces données.

² Par exemple, en leur permettant d'inclure les données dans leurs bilans comptables, ou en désignant des tiers chargés d'évaluer leur valeur et les scénarios possibles pour répartir la valeur générée par le partage entre les contributeurs. Voir par exemple [La stratégie chinoise pour les données](#) – Institut d'études de sécurité de l'Union européenne.

3 La consolidation et la rationalisation de la réglementation pourraient améliorer la clarté réglementaire tout en maintenant la stabilité du cadre récent en matière de données

Compte tenu des nombreuses réglementations applicables aux données, qu'elles soient générales ou sectorielles, la consolidation des réglementations existantes pourrait améliorer la clarté réglementaire. Certaines réglementations, telles que le DGA et le règlement sur la libre circulation des données non personnelles (FFDR), sont moins connues, et les règles qu'elles contiennent pourraient gagner en visibilité grâce à une consolidation, par exemple au sein du règlement sur les données. Cela pourrait également réduire l'incertitude réglementaire, soulignée par certains acteurs, liée à l'interaction entre ces réglementations, et éviter le chevauchement des règles. Cette consolidation appelle également à une gouvernance claire et ambitieuse, comprenant des mécanismes de coopération, et l'EDIB pourrait voir son mandat et sa composition adaptés (par exemple via une gouvernance à deux niveaux, basée sur un niveau stratégique et un niveau technique pour les questions opérationnelles, et une liaison institutionnalisée avec le BEREC sur les questions relatives au cloud et à l'IA dans le contexte du règlement sur les données).

Forte de son expérience dans le secteur des télécommunications, l'Arcep estime que la stabilité réglementaire est essentielle, notamment pour stimuler les investissements. En matière de partage et d'accès aux données, le DGA est encore récent et difficile à évaluer en profondeur, tandis que le règlement sur les données entrera bientôt en application. Les effets combinés du DGA et du règlement sur les données ne se feront pas sentir avant un certain temps, tant pour le secteur privé que pour les organismes publics. Nous pensons donc que seules des adaptations prudentes et ciblées devraient être envisagées en ce qui concerne ces textes.

Compte tenu des défis posés par le partage et l'accès aux données pour l'IA, ainsi que des garanties nécessaires dans ce contexte, nous considérons que les objectifs sous-jacents du DGA, à savoir favoriser un partage fiable des données et éviter les conflits d'intérêts grâce à la conception et à la réglementation du marché, restent pertinents. À ce titre, **l'Arcep estime que le maintien du régime obligatoire pour les PSID pourrait s'avérer bénéfique**, notamment afin de donner plus de temps au marché du partage de données pour se développer, tout en maintenant un environnement fiable pour les échanges de données et en favorisant l'innovation. En outre, cela permettrait d'éviter les préoccupations soulevées par les PSID concernant une éventuelle distorsion de concurrence entre les intermédiaires de données réglementés et non réglementés qui pourrait résulter d'un régime volontaire. **Par ailleurs, les PSID labellisés pourraient être mieux démarqués des autres**, par exemple grâce à une meilleure publicité pour les PSID labellisés, afin de fournir un signal plus clair au marché.

En outre, afin de réduire la charge administrative, certaines obligations applicables aux PSID pourraient être alignées avec d'autres réglementations : la sécurité des données pourrait notamment être couverte par un cadre horizontal, tel que la directive NIS 2.

De plus, les obligations pourraient être clarifiées davantage afin de renforcer la sécurité juridique : les acteurs ont mentionné que le caractère trop restrictif du régime actuel occasionne des difficultés pour les PSID, qui peinent à émerger et à être économiquement viables. En particulier, le périmètre des services supplémentaires que les PSID pourraient fournir pourrait être clarifié, et inclure explicitement les services destinés à améliorer l'utilisabilité des données, comme l'analyse, la sélection, le prétraitement ou l'évaluation de la valeur des données, tant qu'ils sont menés sur des données destinées à être partagées via les services d'intermédiation, à la demande expresse du détenteur des données ; ou sur des données acquises par le biais de ces services, à la demande expresse de leur utilisateur. Cela pourrait favoriser en retour l'objectif de promotion de l'innovation du DGA, tout en maintenant l'accent réglementaire sur les potentiels conflits d'intérêts préjudiciables à la confiance des parties intermédiées.

Enfin, des bacs à sable réglementaires, mieux adaptés aux phases de test de marché, pourraient notamment soutenir les PME et les start-ups, étant donné que le développement de services innovants de partage de données peut comporter des risques importants difficiles à prévoir, tandis que les perspectives de marché et la viabilité économique sont incertaines.