

Consultation de l'Arcom sur l'accès aux données des plateformes en ligne pour la recherche

Contribution de Numeum

Propos liminaire

Numeum, né en 2021 de la fusion de Syntec Numérique et de TECH IN France, est l'organisation professionnelle de l'écosystème numérique en France. Nous représentons les entreprises de services du numérique (ESN), les éditeurs de logiciels, les plateformes et les sociétés d'Ingénierie et de Conseil en Technologies (ICT). Nos 2 300 entreprises membres totalisent 85 % du chiffre d'affaires total du secteur numérique en France, soit 46 milliards d'euros. 50 % de nos membres sont installés en région, et nous comptons 50 % de startups et de PME.

De manière générale, **Numeum soutient toutes les initiatives visant à rapprocher le monde de la recherche et les acteurs économiques.** Nous pensons en effet qu'il est nécessaire de créer davantage de ponts entre le monde de la recherche, qui est majoritairement publique et fondamentale, et celui de l'entreprise, opérationnelle et orientée *business*. Il s'agit de l'un des principaux défis à relever pour accompagner l'essor des entreprises technologiques françaises, tout en s'inscrivant dans un cadre qui soit à la fois protecteur des données, qu'elles soient personnelles ou relèvent du secret des affaires, et qui favorise l'innovation partenariale.

La contribution de Numeum à la consultation de l'Arcom sur l'accès aux données des plateformes en ligne pour la recherche est pour nous l'occasion de partager les positions de nos membres et de relayer leur **volonté de coopérer** de la manière la plus efficace et pragmatique qui soit en matière de recherche, **au service d'un numérique sûr et de confiance. Nous remercions donc l'Autorité de nous en donner ici la possibilité.**

Sur les politiques de partage des données mises en œuvre par les acteurs

Question A.5 : avez-vous établi une politique de partage de vos données avec des tiers à des fins de recherche ?

Les données jouent aujourd'hui un rôle clé dans l'économie et la société et elles constituent un actif stratégique et économique pour les entreprises. Leur traitement est donc majeur pour leur développement, en particulier sur des thématiques d'innovation et d'amélioration continue de leurs services. A ce titre, les relations avec la recherche sont clés car l'innovation est guidée par l'utilisation et l'analyse de données. Cela permet en effet de créer ou d'améliorer de nouveaux produits, processus et méthodes organisationnelles. Les initiatives volontaires de

partage de données peuvent différer d'un acteur à l'autre en raison d'une multitude de facteurs qui ont trait à leur taille, la structuration de leurs activités, leur budget, etc. **Nos adhérents se rejoignent sur deux éléments essentiels : la nécessité de protéger les données personnelles des utilisateurs, en respectant leur consentement, et l'impératif de respect du secret des affaires.**

Sur la gouvernance

Question B.1 : doit-on définir et éventuellement limiter en amont les types d'acteurs pouvant bénéficier d'un accès à des données : chercheurs, journalistes, ONGs, think tanks, société civile, etc. ?

Il nous semble absolument essentiel de définir avant tout quel est le type de recherche qui est visé. Dans le cadre d'un dispositif avec un tiers de confiance, il serait ainsi important que ce dernier puisse certifier le bien-fondé de la démarche de recherche et de ses initiateurs. Dès lors, des critères clairs, transparents et qui auraient été co-construits avec toutes les parties prenantes semblent être la solution la plus à même de répondre à ces enjeux (voir également nos réponses aux questions B.3, C.2 et C.5).

Par ailleurs, il faut souligner que **si tout un continuum d'acteurs participe activement à l'augmentation de la connaissance générale d'un phénomène donné, ceux-ci ne répondent pas individuellement aux mêmes exigences déontologiques et n'ont pas nécessairement la même appréhension de certains aspects inhérents à la réception, au traitement, et à la restitution des données.** Ainsi, sans la connaissance de la maîtrise du cycle de vie de la donnée qui permet de comprendre à quel moment celle-ci est la plus exploitable ni les compétences techniques qui vont de pair, il semble difficile de prévoir un accès large à des données dont le caractère sensible a déjà été souligné. Dès lors, il serait pertinent d'imaginer que les accès qui pourraient être envisagés soient le fruit de discussions multipartites et débouchent sur des approches pragmatiques afin d'être les plus pertinentes possibles.

Question B.3 : selon vous, un modèle d'accès à des données via la formulation des demandes d'accès à un tiers de confiance est-il pertinent ?

Il faut tout d'abord souligner que l'intervention d'un tiers de confiance n'est pas prévue dans le Digital Services Act (DSA) et son article 31 auxquels la consultation de l'Arcom fait référence. Cet article 31 du DSA organise l'accès aux données nécessaires au suivi et à l'évaluation de la conformité des très grandes plateformes pour le *Digital Services Coordinator* du pays d'établissement ainsi que la détection, l'identification et la compréhension par des chercheurs préalablement approuvés des risques systémiques, et des moyens mis en œuvre par les opérateurs pour atténuer ces risques.

De fait, il apparaît inopportun de prévoir des mécanismes nationaux d'organisation de l'accès aux données des plateformes en ligne pour la recherche. En effet, les entreprises dans le champ du DSA ont par nature des activités qui dépassent le seul cadre national. Dès lors, il apparaît essentiel qu'il existe une harmonisation au niveau européen afin d'éviter la fragmentation de notre marché unique. Numeum attire donc l'attention de l'Arcom sur la nécessité que l'article 31, qui bénéficierait de précisions sur sa mise en œuvre, puisse déjà

produire ses effets, avant qu'une évaluation puisse ensuite être faite sur les éventuels compléments qu'il conviendrait d'y apporter.

Si un mécanisme de tiers de confiance au niveau européen devait aboutir, il est important que celui-ci bénéficie de toutes les garanties nécessaires en matière d'indépendance, notamment via la participation de toutes les parties prenantes concernées au premier rang duquel se situent la recherche et les plateformes. Un tel organe reste en tout cas essentiel au fonctionnement équilibré en matière de demande, d'accès et d'analyse des données visées. Ce tiers devrait également avoir pour rôle d'interpréter les obligations légales existantes, d'évaluer les besoins du monde académique et d'élaborer des standards adéquats de protection de la vie privée et de sécurité.

Par suite, une adhésion obligatoire à ce tiers de confiance ne peut, par parallélisme des formes, qu'être le reflet des obligations légales d'accès aux données. Ce point est particulièrement crucial afin d'assurer une sécurité juridique dans le cadre de données à la sensibilité forte, que celle-ci soit liée à leur caractère personnel ou au secret des affaires.

Sur la construction des projets scientifiques

Question C.2 : qui définirait le contour des projets de recherche et leur rattachement à une ou plusieurs missions d'intérêt général et présidant à l'identification des données auquel l'accès serait nécessaire ?

Question C.5 : dans quelle mesure le caractère plus ou moins contraignant des obligations d'ouverture de leurs données pour les plateformes impliquerait-il leur présence dans les comités d'évaluation ?

Les projets de recherche qui pourraient bénéficier d'un accès à certaines données des plateformes en ligne devraient pouvoir permettre tout autant de **faire avancer la connaissance collective de sujets majeurs comme la lutte contre les contenus préjudiciables**, mais également de **permettre aux entreprises de prendre des mesures afin d'améliorer la confiance et la sûreté de leurs services**.

Comme indiqué plus haut, Numeum se positionne en faveur d'un accès à des données dans un cadre juridique clair. Le DSA pose déjà un tel cadre, de même que le Code de bonnes pratiques contre la désinformation. A titre d'exemple, ce Code publié par la Commission le 16 juin 2022 dans sa version renforcée comporte un chapitre dédié au monde académique¹. En faisant référence à la proposition de l'EDMO d'un Code de conduite en matière d'accès aux données des plateformes, le Code prévoit notamment des engagements en matière de :

- Publication de données pour les chercheurs dans une forme anonymisée et agrégée (engagement n°26) ;
- Accès à des chercheurs approuvés via un tiers indépendant dont le développement et le financement sera multipartite dans le cadre d'une réelle co-construction, et avec la mission d'évaluer les projets de recherche et leurs auteurs (engagement n°27). Il faut relever ici que

¹ Commission européenne, « The Strengthened Code of Practice on Disinformation 2022 », *European Commission*, 16 juin 2022. URL : <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>

la mesure n°27-3 implique déjà le développement de projets pilotes de partage de données avec des chercheurs approuvés sans attendre la mise en place de ce tiers.

Aussi, il est nécessaire que les plateformes en ligne puissent continuer à être parties prenantes de cette définition des projets de recherche, avec des garanties suffisantes en termes de transparence comme de confidentialité pour créer un véritable cercle vertueux.

Dans un contexte d'identification des données pour un projet de recherche donné (question C.2), il serait judicieux de prévoir que celles sélectionnées puissent être réutilisées pour d'autres projets. La répliquabilité de la recherche, en plus d'être un impératif dans le monde académique afin de permettre de croiser et renforcer ainsi les analyses, doit être un guide afin d'assurer l'efficacité technique comme financière des plateformes dans la production des données adéquates, en particulier pour les plus petites d'entre elles. Ainsi, les ensembles de données aux usages multiples, rattachés à une ou plusieurs missions d'intérêt général par un tiers indépendant et externe tenant compte des risques d'atteinte à la vie privée et au secret des affaires, seront les plus à même de bénéficier à l'ensemble de la communauté scientifique plutôt qu'à un nombre restreint d'acteurs académiques.

Enfin, une attention concernant la simplicité et la transparence du processus qui pourrait être mis en place, tant pour les chercheurs que pour les entreprises, est essentielle afin de garantir sa pertinence. Si l'ouverture des données des plateformes en ligne à des fins de recherche était soumise à l'intervention d'un tiers de confiance, il serait par exemple souhaitable de mettre en place un protocole d'accès transparent afin que les chercheurs puissent avoir connaissance du nombre de requêtes reçues et du nombre de requêtes refusées, ainsi que les motifs de refus.

Question C.9 : quelles doivent être les précautions à prendre en ce qui concerne la publication des études menées, par exemple eu égard à la sensibilité des données qui auraient été exploitées ?

Nous attirons l'attention de l'Arcom sur le fait que **cette question adresse un public spécifique de la recherche** : celui des chercheurs travaillant dans le cadre de publications scientifiques établies, avec des codes de déontologie spécifiques et partagés par l'ensemble du monde académique. Dans le cadre d'une potentielle ouverture à d'autres acteurs du continuum de la recherche, ces procédés stricts de publication diffèrent voire n'ont pas d'équivalents. **Il est donc essentiel de réfléchir à une harmonisation par le haut des standards de recherche et de publication des travaux** si cette piste devait aboutir.

Par ailleurs, Numeum comprend et soutient l'objectif final visé aussi bien par l'article 31 du DSA que par cette consultation de développer un environnement numérique sûr et de confiance. Il est donc important de s'assurer que la publication des résultats de recherche, si elle est importante, ne peut être la seule manière de répondre à cet objectif lorsqu'il s'agit de problématiques comme la lutte contre les manipulations de l'information ou contre les contenus haineux en ligne. Le développement collaboratif de normes en matière de rapports de transparence et de jeux de données agrégés, répondant aux nécessités de protection de la vie privée et du secret des affaires, est également un moyen de répondre à cet objectif que nous partageons.

Sur la protection des données et les considérations techniques

Question D.2 : les accès aux données peuvent-ils et doivent-ils faire l'objet d'une co-construction à part égale entre acteurs de la gouvernance, chercheurs et plateformes sur le modèle du CASD de l'Insee ?

En premier lieu, il est impératif d'assurer que les modes d'accès et de stockage des jeux de données soient hautement sécurisés. La co-construction de la gouvernance entre monde académique et plateformes nous semble ici pouvoir jouer un rôle essentiel. Ce type de modèle a prouvé son efficacité et donne des résultats concrets pour toutes les parties impliquées.

Un exemple avec lequel un parallèle peut être fait est celui du supercalculateur Jean Zay². Fruit d'une collaboration entre la recherche académique notamment en intelligence artificielle et un constructeur mondial, Hewlett Packard Entreprise, il a été inauguré le 24 janvier 2020 à Paris-Saclay, au sein de l'Institut du développement et des ressources en informatique scientifique (IDRIS). Ce supercalculateur a permis de doubler la puissance de calcul française et s'inscrit dans le plan AI for Humanity lancé par la France en 2018. La co-conception du supercalculateur ainsi que l'appui technique des équipes d'Hewlett Packard Entreprise, avec la recherche publique, sont ici des facteurs-clés de succès de ce projet.

Sur la faisabilité de l'accès et les mécanismes d'incitations

Question E.3 : comment mettre en place des incitations efficaces et équilibrées pour que les plateformes s'inscrivent dans des dynamiques d'ouverture des données ? Comment intégrer ces acteurs dans le dispositif de manière cohérente et favoriser les bonnes pratiques ?

Les récentes initiatives européennes, notamment le DSA et le Code de bonnes pratiques contre la désinformation que nous avons accueillis très favorablement, nous semblent également ici apporter des réponses concrètes.

Alors que le marché unique européen est le plus pertinent pour les entreprises, il nous semble nécessaire que toute initiative complémentaire au niveau français soit faite en bonne **coordination avec les instances européennes** afin qu'elle soit la plus efficace possible.

Il nous semble également important de pouvoir raisonner avec le **recul suffisant lié à la mise en œuvre de ces textes**, afin d'évaluer leurs effets et d'assurer une bonne identification des éventuelles problématiques qu'elles n'auraient pas résolu, avant d'imaginer d'autres incitations et dispositifs.

² CNRS, « Supercalculateur Jean Zay : les défis de la co-construction », *CNRS.fr*, 23 janvier 2020. URL : <https://www.cnrs.fr/fr/cnrsinfo/supercalculateur-jean-zay-les-defis-de-la-co-construction>