



## Position du GFII préparant la réunion avec la RPUE *Décembre 2021*

### 1) Rappel des positions structurantes du GFII

Depuis de nombreuses années, le GFII contribue aux travaux et aux réflexions en matière d'ouverture et d'exploitation des données du secteur public, comme privé. Les contributions du GFII sont fondées sur l'expérience opérationnelle de ses membres, et à plusieurs occasions, ses remarques ont été confirmées lors des débats ultérieurs.

Nos axes principaux sont les suivants :

- **Toutes les données ne se valent pas.** Les données sont des actifs numériques essentiels, tant pour les États, que les entreprises ou les collectivités territoriales. Même si les données publiques sont dites « de libre parcours », il n'en demeure pas moins que certaines, dans des domaines stratégiques, ne peuvent pas être communiquées de manière identique et aveugle pour toute réutilisation, pas forcément bienveillante.
- **Toutes les données n'ont pas un potentiel économique identique.** Rien de commun entre les horaires des piscines, la localisation des places de parking et le fichier SIRENE. Intrinsèquement, les données recèlent des potentialités différentes. Or, les entreprises, qu'il s'agisse de start-up ou d'entreprises existant de longue date, ne peuvent développer des projets et de nouveaux services basés sur les données sans que la pérennité de la fourniture de données ne soit confirmée par le producteur et que les qualités essentielles des données soient maintenues, voire améliorées.
- **Les besoins des acteurs économiques sont distincts de ceux du grand public** et requièrent une qualité de données supérieure. Dans ce cadre professionnel, la « data » devient vitale, car elle permet à l'entreprise ou à l'administration, une meilleure connaissance de son environnement, de ses concurrents, de ses clients/usagers et, surtout, permet le développement de nouveaux services à valeur ajoutée, gages d'un développement numérique. Pour y parvenir, il est capital que la donnée soit de qualité. Comment définir la qualité ? C'est une donnée fiable, tenue à jour, exhaustive et dont le producteur garantit la permanence de la qualité, la maintenance et la pérennité de sa

fourniture. Sans qualité, pas de développements d'une IA fiable (risque de biais), pas de nouveaux outils d'aide à la décision.

- **L'ouverture des données génère des coûts très conséquents et croissants, et la qualité de service doit être financée** : tout fournisseur/détenteur de données, qu'il soit public ou privé, supporte des coûts de « mise au propre » de ses données conditionnant leur mise à disposition (élaboration de documentation technique, standardisation des données, mise en œuvre d'une architecture technique à même d'éventuellement devoir supporter un nombre important de requêtes etc.).

Pour un acteur public, cette exigence en termes de qualité de service a un coût qui ne peut pas entièrement reposer sur l'impôt. Cette problématique – de ce que nous comprenons – constitue une des difficultés à assurer l'ouverture des données publiques. De même, la problématique du financement du fournisseur/détenteur de données privées doit aussi être anticipée et bien prise en compte (notamment par le futur « Data Act ») ;

- **« Détenteurs » et « réutilisateurs » de données doivent dialoguer** : la relation « client – fournisseur », qui s'installe de facto entre un fournisseur/détenteur et un réutilisateur de données, nécessite une discussion et une compréhension réciproque. Ce besoin de dialogue est trop souvent sous-estimé, notamment dans le cas de l'ouverture des données publiques, alors qu'il constitue, pour le producteur, un observatoire idéal et vertueux lui permettant d'améliorer les données existantes ou la création de nouveaux jeux de données répondant à une demande identifiée. En revanche, l'animation de ces communautés suppose, là encore, le recrutement de nouvelles ressources.
- **La politique d'ouverture et de réutilisation des données doit être conduite en fonction des cas d'usage** : tous les réutilisateurs n'ont pas les mêmes besoins. La réussite de l'encadrement juridique et de la régulation implique de se fonder sur les cas d'usage, en évitant de partir du principe que tous les besoins sont identiques.

Sur ce sujet, les remarques du GFII en réponse à la publication du rapport du CESE français « Pour une politique publique de la donnée, des algorithmes et des codes source » publié en décembre 2020 sont également valables au niveau européen<sup>1</sup>.

---

<sup>1</sup> Voir document joint « *Les 11 propositions du GFII dans le cadre de la mission Bothorel sur la politique de la donnée et des codes sources* »

## 2) Le GFII soutient la politique de la Commission européenne en matière de données et d'intelligence artificielle (IA)

Bien que sa valeur financière soit difficilement estimable, la donnée reste un actif immatériel des acteurs européens, qu'ils soient publics ou privés, qu'il est donc nécessaire de protéger. Le GFII est par conséquent favorable au développement d'un **marché unique des données** pour favoriser l'ouverture et la réutilisation des données à l'échelle de l'Union européenne. Cette approche nécessairement « économique » permet de garantir aux fournisseurs/détenteurs de données de pouvoir tarifier l'accès et la réutilisation, voire la rediffusion de leurs données. Cette même logique s'applique bien sûr également aux services développés par les réutilisateurs à partir de ces mêmes données, qu'ils auront su agréger et « faire parler ».

Une tarification des données offre en effet plusieurs avantages. Premièrement, elle **facilite l'ouverture des données** en ne la conditionnant pas à l'existence d'un budget dédié, entre autres celui des collectivités publiques et organismes liés (surtout dans le contexte de tension budgétaire actuel). Deuxièmement, elle permet de **mieux répondre aux besoins potentiels de protection des données** (anonymisation ; infrastructures ; compétences humaines, etc.). Dernièrement, elle permet de **ne pas conditionner l'ouverture et le partage des données aux seuls fournisseurs/détenteurs de données qui disposent d'un budget dédié** à cet objectif. Autrement dit, la tarification permet d'impliquer un plus grand nombre de fournisseurs/détenteurs de données, notamment au sein du secteur public.

*L'exemple de l'évolution du droit français sur ce sujet, voire, récemment, de la directive sur la réutilisation des informations du secteur public de 2019 démontre clairement les lacunes d'une gratuité généralisée de la réutilisation, notamment en termes de qualité et disponibilité insuffisantes des données.*

Bien sûr, le GFII n'est d'aucune manière opposé à l'idée d'une **tarification préférentielle**, voire une gratuité, pour certains types de réutilisation/réutilisateurs (start up ; PME ; organismes de recherche, etc.).

Enfin, le GFII salue l'adoption par la Commission européenne d'une **approche globale et souveraine** en matière de données et d'IA (technologies cloud ; IA ; cybersécurité, etc.).

Cependant, le GFII souligne également certaines lacunes ou risques :

- **La superposition des textes et des thématiques** traitées (par exemple, la question du transfert de données vers les pays tiers semble être encadrée à la fois par le RGPD, dans le « Data Governance Act » et le prochain « Data Act »). Il existe donc le risque d'une perte de lisibilité du cadre juridique.

- **L'adaptation, ou la flexibilité, insuffisante du droit au domaine numérique** : le numérique est une matière qui évolue très rapidement et exige une certaine flexibilité du cadre juridique, en se fondant entre autres sur de réels cas d'usage.
- **La difficile compréhension du cadre juridique par les opérationnels du numérique** : la réglementation doit pouvoir être comprise et maîtrisée par tous les acteurs du numérique, qu'ils soient juristes, compliance officers, DPO ou non. Dans le cas contraire, il existe le risque qu'en pratique, cette réglementation ne soit pas appliquée ou mal. L'efficacité de ces règles implique également la rédaction de définitions traduisant suffisamment la pratique, tout en permettant une flexibilité et une adaptation nécessaires pour soutenir l'innovation. Si le projet d' « Artificial Intelligence Act » amorce cette évolution (ce que le GFII encourage), ses dispositions telles que celles relatives à la gouvernance des données (article 10) doivent encore être précisées pour les raisons explicitées précédemment.
- **Une confusion récurrente entre réutilisateur de données publiques ou privées et utilisateur / client final**  
 Un utilisateur / client final de « la donnée » peut être réutilisateur s'il procède **lui-même** aux requêtes, téléchargement et traitements. Mais, bien souvent, le réutilisateur est le « go between » entre le producteur de données (ou l'intermédiaire de données) et l'utilisateur / client final.  
 Il lui revient donc d'accéder aux données, de les traiter, de les « faire parler » et de les diffuser (si autorisé) au bénéfice de ses utilisateurs / clients finaux.  
 Le besoin du réutilisateur naît donc des besoins de ses utilisateurs / clients finaux, sources d'innovation s'il en est.
- **L'absence de vision globale, et suffisamment détaillée, sur les standards de données.**  
 Pour le secteur public, cette question est traitée avec le cadre d'interopérabilité européen. Mais il existe un besoin d'élargir cette action à l'ensemble des acteurs européens, publics comme privés. De notre compréhension, ce sujet devrait être un des points centraux de la construction des espaces européens de données. Le GFII sera d'ailleurs attentif à cette construction. En effet, l'intégration, ou non, de certains acteurs dans un espace de données particulier sera un enjeu important.
- **L'omniprésence du cadre RGPD dans les activités commerciales et non commerciales.**  
 Le GFII ne conteste pas le bien-fondé du cadre RGPD, mais estime nécessaire d'adopter une nouvelle approche et de redéfinir la proportionnalité d'intervention pour laisser plus de marge de manœuvre aux acteurs privés comme publics dans leurs activités. Ainsi soutient-il le souhait exprimé *mezzo voce* d'ouvrir le débat sur les difficultés d'application du RGPD dans le cadre de l'ouverture, réutilisation et diffusion des données publiques et privées, ainsi que dans le cadre de leur exploitation par l'IA.

En effet, le RGPD s'applique très largement, et de ce fait, est parfois en incohérence avec d'autres réglementations. Il nous semble, par exemple, impératif de clarifier et de différencier les différents cas de figure que sont :

- les données personnelles relevant de la « sphère privée » et au sens de « consommateur » : « toute personne physique qui agit à des fins qui n'entrent **pas** dans le cadre de son activité commerciale, industrielle, artisanale, libérale ou agricole (définition issue de l'article liminaire du code de la consommation français, qui s'avère proche de celle que l'on retrouve dans la directive (CE) n° 2008/48 « concernant les contrats de crédit aux consommateurs ») ;
- les données personnelles librement accessibles au public sur internet ;
- les données personnelles de publicité légale destinées à l'information des tiers mises à disposition en open data ;
- les données personnelles relevant de la « sphère professionnelle ».

Certaines de ces données étant *de facto* identiques, ce sont le contexte et/ou la finalité de leur collecte initiale et/ou de leur réutilisation qui permettent de faire la différence.

Or, ce contexte et cette finalité ne sont pas efficacement pris en compte dans le RGPD ni dans certaines licences de réutilisation des données.

Les données sur les entreprises, mentionnées comme jeux de données de forte valeur dans la directive dite PSI III, illustrent très bien cet enjeu. D'un côté, les entrepreneurs individuels sont considérés comme des entreprises, ce qui, en France, justifie leur inscription au « répertoire SIRENE<sup>2</sup> ». Ils doivent par ailleurs s'immatriculer au RCS ou au répertoire des Métiers, voire les deux (et à compter du 1<sup>er</sup> janvier 2023 au « Registre National des Entreprises<sup>3</sup> » ) et sont soumis à une obligation de publicité légale en vue d'assurer la transparence de la vie économique, raison pour laquelle les données les concernant sont mises à disposition en open data. De l'autre, les entrepreneurs individuels, lorsque les données **concernant leur entreprise** sont réutilisées, sont considérés comme des « sujets de données » au sens du RGPD.

La problématique est identique s'agissant des données des dirigeants statutaires, qui sont soumises à publicité légale (au Kbis, bien sûr, mais aussi dans le Registre National du Commerce et des Sociétés ou au BODACC, accessibles en open data) et sont protégées par le RGPD quand elles sont réutilisées, y compris pour une finalité en lien direct avec celle de la collecte initiale par les sources publiques.

Cette mise en cohérence des textes entre eux doit être effectuée tout au long de la chaîne de traitement de la donnée, soit du producteur de données publiques à

---

<sup>2</sup> Système Informatisé du Répertoire National des Entreprises et des Établissements

<sup>3</sup> Par application de l'ordonnance du 15 septembre 2021, prise sur habilitation de la loi n° 2019-486 du 22 mai 2019 relative à la croissance et la transformation des entreprises

l'utilisateur final de la donnée, qui, d'ailleurs, peut être assujéti à diverses réglementations exigeant l'identification de ses tiers.

- **Le nécessaire consentement des entreprises et leur indemnisation en cas de mise à disposition de leur données dans le cadre du « BtoG » (futur Data Act)**

La crise du Covid19 a clairement montré l'intérêt pour l'action publique au sens large de pouvoir s'appuyer sur des données de source privée ou « données d'intérêt général ».

Or, les données des sources privées représentent pour leur producteur un patrimoine numérique essentiel et un atout économique indispensable. Là encore, le GFII encourage à travailler sur la base de cas d'usage et surtout pas sur la base de grands principes généraux. C'est de la multiplicité des cas d'usage et des retours d'expérience que des principes généraux communs pourront éventuellement être tirés. Le sujet de l'indemnisation devra être exploré, d'autant qu'il faudra certainement faire des développements pour répondre au besoin de l'État et des collectivités publiques au sens large.

Exactement comme les acteurs publics doivent faire des développements pour permettre une réutilisation efficace de leurs données. Rendre le « BtoG » obligatoire serait contre-productif ; il conviendra donc d'être incitatif et non de procéder de manière coercitive. Ainsi, la mise à disposition de données des données du secteur privé au bénéfice de la puissance publique ne pourra à priori fonctionner que sur la base du volontariat des acteurs économiques privés et leur indemnisation.