

Données publiques et données personnelles : un « mariage pour tous » aux limites de la légalité

D'ici deux ans, soit à partir du 18 juillet 2015, les Etats membres de l'Union européenne doivent transposer la nouvelle directive « Réutilisation des informations du secteur public ». Mais des pays comme la France s'organisent déjà via Etalab, sous la « surveillance » de la Cnil.

Par Winston Maxwell, avocat associé Hogan Lovells LLP



L'Union européenne vient d'adopter une directive rendant obligatoire la libre réutilisation de données du secteur public. Cette directive modifie la directive de 2003 (1), qui avait déjà établi un cadre européen pour la réutilisation de données du secteur public, mais qui n'a pas rendu cette réutilisation obligatoire. La nouvelle directive « Réutilisation des informations du secteur public » du 26 juin 2013 (2) vient renforcer la politique « Open Data » de la Commission européenne telle qu'exprimée dans sa communication « Ouverture des données publiques » du 12 décembre 2011 (3).

L'objectif est double : d'une part, renforcer la transparence du gouvernement à l'égard du citoyen ; d'autre part, contribuer à l'innovation et à la croissance. Comme l'a constaté le rapport Colin & Collin sur la fiscalité numérique (4), les données massives détenues par les administrations et les grandes entreprises constituent une source de richesse considérable. Il ne s'agit pas seulement d'accorder un accès à quelques chercheurs triés sur le volet, mais plutôt de rendre les données accessibles à toute société qui en ferait la demande. La directive du 26 juin 2013 adopte cette philosophie car les Etats membres devront – à partir du 18 juillet 2015 – rendre des données accessibles pour une réutilisation à des fins commerciales ou non. Cette vision large du Open Data est déjà appliquée par le gouvernement français, via la mission Etalab à laquelle participent Google, Microsoft ou encore Orange dans le cadre de son programme Dataconnexions (voir encadré page suivante).

L'anonyme peut être ré-identifié

Une semaine après l'adoption de la nouvelle directive, le groupe « Article 29 » des « Cnil » européennes (« G29 ») a publié un avis mettant les Etats membres en garde contre une lecture trop extensive de la nouvelle directive (5). Il a en effet rappelé que la nouvelle directive sur la réutilisation des données du secteur public ne dérogeait pas aux règles sur la protection

des données personnelles : une communication de données dans le cadre d'une initiative Open Data ne peut être envisagée que si l'ensemble des conditions de protection des données personnelles sont réunies.

• **Anonymisation ou pas ?** La législation en matière de protection des données personnelles ne s'applique pas aux données anonymes. Mais attention à la fausse anonymisation ! De nombreuses études scientifiques montrent que des données en apparence anonymes peuvent faire l'objet d'une ré-identification en croisant les données avec de nombreuses autres sources, et en appliquant des algorithmes de probabilité statistique. De plus, la ré-identification de données anonymes est une science évolutive. Les données qui sont anonymes aujourd'hui ne le seront peut-être plus dans un an, en raison d'évolutions dans les techniques de ré-identification. Le G29 préconise d'effectuer des tests de ré-identification avant de déclarer une base de données anonyme, et donc hors du champ de la régulation des données personnelles.

• **Données non anonymes : forcément exclues du Open Data ?** Même si l'anonymisation totale est la situation préférée des autorités, une anonymisation partielle ou imparfaite peut être envisagée. Les chercheurs et sociétés innovantes auront souvent besoin d'accès à des données individualisées afin de créer des applications innovantes. Ces données individualisées – par exemple les temps de trajet des Franciliens dans le métro parisien – constituent des données à caractère personnel, même si l'identité réelle de l'individu reste cachée. Pour ces données partiellement anonymisées, le G29 préconise une série de mesures avant d'envisager leur ouverture au public.

A défaut de consentement explicite

D'abord, l'administration qui détient ces données doit conduire une étude d'impact en matière de données personnelles pour évaluer les risques pour les citoyens, et envisager des mesures pour réduire ces risques. Par exemple, même si une anonymisation n'est pas totale, l'analyse de risques pourrait conclure

Notes

(1) - Directive européenne 2003/98/CE du 17 novembre 2003 concernant la réutilisation des informations du secteur public.

(2) - Directive européenne 2013/37/UE du 26 juin 2013 modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public.

(3) - Communication de la Commission européenne du 12 décembre 2011 sur l'ouverture des données publiques : un moteur pour l'innovation, la croissance et une gouvernance transparente

(4) - Rapport de la « mission d'expertise de la fiscalité de l'économie numérique », de Pierre Collin et Nicolas Colin, 18 janvier 2013.

des données personnelles : une communication de données dans le cadre d'une initiative Open Data ne peut être envisagée que si l'ensemble des conditions de protection des données personnelles sont réunies.

que les mesures d'anonymisation appliquées sont suffisamment robustes pour écarter les risques les plus graves. Ensuite, il faut examiner la finalité recherchée dans la réutilisation des données. Certaines finalités seront « compatibles », et certaines ne le seront pas. Par exemple, une finalité purement scientifique visant à déceler des tendances générales découlant d'une série de données serait légitime. En revanche, la création de profils individuels dans le but de vendre de la publicité ciblée ne serait pas légitime selon le G29.

Data Privacy versus Open Data

Sur ce point, la philosophie des lois sur la protection des données personnelles (Data Privacy) rentre en conflit direct avec la philosophie Open Data. Cette dernière vise à permettre des expérimentations les plus diverses et inattendues possibles. Les plus grandes innovations et découvertes peuvent venir d'idées a priori farfelues. En données personnelles, en revanche, la finalité est extrêmement importante. La finalité d'une réutilisation de données doit être compatible avec la finalité pour laquelle les données ont été collectées à l'origine. Sinon, il faut recueillir le consentement explicite de la personne concernée – tâche en pratique impossible pour des masses importantes de données. Ainsi, la Cnil (6) et les autres autorités de données personnelles en Europe veilleront à ce que les finalités de réutilisation visent un objectif « noble » tel que la science, l'histoire ou les études statistiques. Des applications plus commerciales risquent d'être jugées moins « compatibles » avec la finalité d'origine.

- **Eviter les téléchargements massifs.** Le G29 préconise par ailleurs la mise en place de mécanismes afin d'éviter des transferts massifs de données. Par exemple, lorsqu'une administration publie le nom et la

date de naissance d'administrateurs de sociétés anonymes en France, ces fichiers doivent être consultables de manière à éviter le téléchargement massif de l'ensemble de la base de données. Le G29 préconise l'utilisation d'API, des interfaces de programmes d'applications, afin d'éviter que le contenu d'une base de données ne soit aspiré dans sa totalité.

- **Les clauses de licences.** Enfin, le G29 préconise l'utilisation de licences de réutilisation de données, afin d'obliger la personne qui obtient l'accès aux données publiques de faire une utilisation conforme à la loi sur la protection des données personnelles. Notamment, la personne doit s'engager à ne pas tenter de retrouver l'identité individuelle des utilisateurs, ni d'utiliser les données afin de créer des profils individuels incompatibles avec la loi sur la protection des données à caractère personnel.

- **Une question institutionnelle en France.** En l'absence d'anonymisation totale, l'application des principes d'Open Data sera difficile, particulièrement par rapport à la question délicate de l'utilisation « compatible » de données à caractère personnel. Actuellement, la loi a confié à la Commission d'accès aux documents administratifs (CADA) le soin d'évaluer la conformité de toute réutilisation d'informations publiques.

La Cnil possède une expertise évidente sur l'anonymisation et sur les études de risques en données personnelles. Actuellement, la loi française ne prévoit pas de passerelle officielle entre la CADA et la Cnil sur ces sujets. Les deux autorités administratives seront sans doute amenées à se rapprocher sur ces questions complexes.

Afin d'établir si une réflexion est nécessaire sur l'Open Data au regard de son champ de compétence, la Cnil a lancé en mars 2013 une consultation des acteurs publics et privés concernés (7). @

Notes

(5) - "Opinion on open data and public sector information re-use", adopté le 5 juin 2013 (Article 29 Data Protection Working Party), mais publié le 3 juillet 2013.

(6) - Commission nationale de l'informatique et des libertés (Cnil).

(7) - Dans le cadre de cette concertation, la Cnil a organisé le 9 juillet 2013 – avec le soutien d'Etalab – une journée d'échanges et de retours d'expériences avec les acteurs et experts du sujet

Focus

Google, Microsoft et Orange, partenaires d'Etalab au sien de Dataconnexions

En France, la politique d'ouverture en ligne des données publiques (« Open Data ») est pilotée par la mission Etalab, placée sous l'autorité du Premier ministre depuis février 2011. Et depuis octobre 2012, cette mission est rattachée directement au secrétaire général pour la modernisation de l'action publique. Etalab gère le portail unique interministériel Data.gouv.fr, lequel met « à disposition librement l'ensemble des informations publiques de l'Etat, de ses établissements publics et, si elles le souhaitent, des collectivités territoriales et des personnes de droit public ou de droit privé

chargées d'une mission de service public ». Cela représente actuellement 355.000 informations publiques gratuites et réutilisables.

Etalab a en outre rassemblé plus d'une trentaine d'acteurs de l'innovation en France au sein d'une communauté appelée Dataconnexions. On y retrouve comme partenaires « Industrie » Google, Microsoft, Orange ou encore Salesforce.com.

Objectif : organiser des concours en vue de récompenser des projets particulièrement prometteurs. @