

Quelques définitions

1. OPEN DATA

L'expression « Open Data » désigne un ensemble de stratégies d'action fondées sur l'ouverture et le partage de données publiques (c'est-à-dire non personnelles ni protégées par un secret légal), mises en ligne dans des formats ouverts et structurés, et autorisant la réutilisation libre et gratuite par toute personne à d'autres fins que celles de la mission pour laquelle la donnée a été produite ou collectée.

Une stratégie d'Open Data peut être mise en œuvre par une organisation publique ou privée, pour différentes fins :

- stratégie de plateforme
- devoir d'information
- transparence du fonctionnement de l'organisation
- communication et modernisation des échanges internes d'information
- création de référentiels partagés avec ses clients / fournisseurs

Cela peut se faire selon différentes modalités (autoriser le libre téléchargement des données, plateforme dédiée, autoriser la consultation des bases à travers des « API », etc.). Ces modalités doivent respecter pour autant le principe de la plus grande liberté de réutilisation des données publiques, en évitant le plus possible les contraintes d'ordre technique, financier, juridique ou autre. Les données publiées en Open Data ont vocation à être **librement accessibles et réutilisables gratuitement par tous**.

Une licence, telle que la Licence Ouverte adoptée par la France en 2011, est par ailleurs associée à chaque jeu de données publiques diffusé en *open data*, afin de préciser les conditions de réutilisation accordées par les fournisseurs de données et d'apporter la sécurité juridique nécessaire aux producteurs et aux réutilisateurs de données publiques.

Le gouvernement français a engagé une politique d'open data service de trois objectifs majeurs :

- **Accroître la transparence et améliorer le fonctionnement démocratique** : en ouvrant les données publiques qu'il produit ou collecte, l'administration rend des comptes aux citoyens, met à leur disposition des informations essentielles d'intérêt général et ouvre le dialogue avec la société civile ;
- **Moderniser l'action publique** : en simplifiant le fonctionnement administratif et en diffusant la culture de la donnée au sein et au service des administrations, notamment pour un pilotage objectif et une plus grande efficacité des politiques publiques.
- **Soutenir l'innovation économique et sociale et stimuler la croissance** : en mettant des données publiques réutilisables à la disposition de tous (usagers, innovateurs, start-ups, entreprises, chercheurs, journalistes, ...), l'administration génère des externalités positives sur l'économie en permettant l'invention de services innovants (applications mobiles, sites Internet, visualisations données ou « datavisualisations », etc.), en stimulant l'économie numérique et en soutenant la croissance dans des secteurs d'avenir ;

Un des considérants de la directive 2013/37/UE du 26 juin 2013, sur les « Informations du secteur public » énonçait ainsi que autoriser la réutilisation « *apporte de la valeur ajoutée aux réutilisateurs, aux utilisateurs finaux, à la société dans son ensemble et, dans de nombreux cas, à l'organisme public lui-même, en favorisant la transparence et la responsabilité et en permettant le retour d'informations des réutilisateurs et des utilisateurs finals, ce qui permet à l'organisme du secteur public concerné d'améliorer la qualité des informations recueillies* ».

L'objectif de l'open data est ainsi d'encourager la créativité, de stimuler l'innovation et de favoriser la réutilisation la plus large possible des données publiques en se reposant sur l'intelligence collective et la volonté des citoyens de créer de nouveaux services innovants et utiles à tous.

2. L'OUVERTURE DES DONNEES PUBLIQUES

En France, la politique d'open data prend appui sur la loi de 1978 organisant l'accès aux documents administratifs (loi CADA) modifiée par la directive européenne Informations du secteur public (2003, 2013), même si elle dépasse parfois les seuls termes de cette loi.

✓ La loi CADA : accès aux documents administratifs et libre réutilisation des informations publiques contenues dans ces documents administratifs

La loi n°78-753 du 17 juillet 1978 (dite « loi CADA ») pose un régime légal applicable aux administrations concernées (cf. *infra*) qui se fonde sur deux droits pour tout individu :

- Le droit pour toute personne d'obtenir la communication des documents administratifs,
- le droit pour toute personne de réutiliser les informations publiques contenues dans ces documents administratifs.

Le socle juridique de l'ouverture et du partage des données publiques (« Open Data ») repose en effet sur ce régime légal fixé par la loi CADA qui reconnaît à toute personne le droit d'obtenir communication des documents détenus dans le cadre de sa mission de service public par une administration, tels que définis dans son article 1^{er} : « (...) *quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions et décisions.* (...) ».

La loi CADA définit donc les administrations concernées comme les administrations de l'Etat (centrales, déconcentrées, établissements publics administratifs), les collectivités territoriales, et les personnes de droit public ou de droit privé chargées d'une mission de service public.

Les informations devant être rendues publiques aux termes de cette loi (parfois appelées « données publiques ») sont toutes les informations (ou données) contenues dans un document administratif communicable (selon les critères fixés par la loi CADA) produit ou reçu par une administration dans le cadre de sa mission de service public, qui ne comporte pas de donnée à caractère personnel, qui ne fait pas l'objet d'un secret protégé par la loi ou qui n'est pas l'objet d'un droit de propriété intellectuelle appartenant à un tiers

En France, les administrations visées par la loi CADA ont un devoir d'ouverture et de partage des données publiques, fondé sur cette loi modifiée par la directive européenne sur les Informations du secteur public de 2003 ou Directive « PSI » (la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public a été transposée par l'ordonnance du 6 juin 2005 et le décret du 30 décembre 2005. Elle a fait l'objet

d'une révision par la directive du 26 juin 2013). Ces textes ont par ailleurs été précisés par plusieurs décrets et circulaires du Président de la République et du Premier ministre, notamment ceux du 26 mai 2011, du 17 mai 2012 et du 17 septembre 2013.

En 2003, la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public (« directive PSI ») – transposée par l'ordonnance du 6 juin 2005 et le décret n°2005-1755 du 30 décembre 2005 – a permis de compléter la loi « CADA », en y introduisant notamment un Chapitre II relatif à la réutilisation des informations publiques (*cf infra*). La circulaire du Premier ministre et le décret du 26 mai 2011 ont par ailleurs arrêté un principe de la réutilisation libre, facile et gratuite pour tous les citoyens s'agissant des données publiques de l'Etat.

- ✓ ***Ne sont pas des données devant être rendues publiques au sens de la loi CADA: les données à caractère personnel, les données contenues dans des documents administratifs non communicables, les données contenues dans un document protégé par un droit de propriété intellectuelle appartenant à un tiers***

Au regard des articles 6 et 10 de la loi CADA, **une information contenue dans un document administratif non communicable n'est pas réutilisable**, sauf si ce document a fait l'objet d'une diffusion publique.

Ainsi, les documents préparatoires et non définitifs de l'administration ne peuvent être communiqués aux citoyens et les informations qui y sont contenues ne peuvent être réutilisées. De même, les informations contenues dans des documents protégés par des secrets légaux (art. 6 loi CADA) ne sont pas réutilisables (secret de la défense nationale, secret des délibérations du Gouvernement, etc).

Les administrations ne sont pas tenues de communiquer les documents faisant par ailleurs l'objet d'une diffusion publique en vertu du 2e alinéa de l'article 2 de la loi du 17 juillet 1978, tel que modifié par la loi du 12 avril 2000. Le législateur a considéré que la diffusion publique de ces documents remplissait la fonction de communication dudit document à l'ensemble des citoyens. Pour caractériser la « **diffusion publique** », la CADA s'attache notamment à l'accessibilité du document tant en termes géographiques, techniques ou de prix. Depuis 2005, rompant avec une jurisprudence contraire, la CADA a ainsi considéré que des documents administratifs aisément accessibles sur internet, au moment où elle a à connaître de la demande de communication, font l'objet d'une diffusion publique. La CADA s'attache à l'accessibilité pratique du document. La CADA considère qu'une mise à disposition sur internet pendant une période limitée de temps ne suffit pour considérer que le document a fait l'objet d'une diffusion publique (conseil n° 20060837 du 16 février 2006). **Par conséquent, une information contenue dans un document administratif ayant fait l'objet d'une diffusion publique est réutilisable au regard des dispositions de la loi CADA.**

- ✓ ***Les données à caractère personnel ne sont pas des données destinées à être rendues publiques***

L'article 13 de la loi CADA rend possible la réutilisation des données publiques comportant des données à caractère personnel, soit parce que les personnes intéressées y ont apporté leur consentement préalable, soit parce que les données à caractère personnel dans les données publiques ont fait l'objet d'une anonymisation, ou à défaut parce qu'une disposition légale ou réglementaire la prescrit.

L'article 13 de la Loi « CADA » dispose en effet que : « **Les informations publiques comportant des données à caractère personnel peuvent faire l'objet d'une réutilisation** soit lorsque la personne intéressée y a consenti, soit si l'autorité détentrice est en mesure de les rendre anonymes ou, à défaut d'anonymisation, si une disposition législative ou réglementaire le permet ». En tout état de cause, cet article 13 précise que la réutilisation des données publiques comportant des données à caractère personnel demeure soumise aux prescriptions de la loi CNIL du 6 janvier 1978.

Pour des motifs légitimes de protection de la vie privée et de respect du secret statistique, **les informations statistiques doivent par ailleurs être publiées dans le respect du secret statistique**, tel que défini par la loi du 7 juin 1951, ainsi que de l'article 285 du Traité instituant la communauté européenne.

3. LES DONNEES A CARACTERE PERSONNEL ET LA PROTECTION DE LA VIE PRIVEE

On entend par donnée à caractère personnel « *toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres* » (article 2, Loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés). Ainsi, une donnée est considérée « à caractère personnel » dès lors qu'elle concerne des personnes physiques identifiées directement ou indirectement.

Il peut s'agir du nom d'une personne, d'une photo, du numéro de sécurité sociale, de téléphone, de compte bancaire, d'une adresse e-mail, d'une empreinte digitale, de l'ADN ...).

La loi CADA, qui organise le devoir de redevabilité de l'Etat quant aux données qu'il détient, limite drastiquement la diffusion des données à caractère personnel par les administrations.

La **loi n°78-17 relative à l'informatique, aux fichiers et aux libertés** du 6 janvier 1978 (plus connue sous le nom de **loi informatique et libertés** de 1978) régit aujourd'hui notamment la pratique du fichage, manuel ou informatique. Elle autorise ces réutilisations sous certaines conditions de finalité et de proportionnalité des données recueillies à cette finalité (et après accord de la CNIL) et introduit un droit d'information, un droit d'opposition, un droit d'accès et un droit de rectification de ces fichiers.

4. LES DONNEES DE SANTE

Il existe un grand nombre d'informations en rapport plus ou moins direct avec la santé. On peut distinguer d'abord les données individuelles (relatives initialement à des personnes, que celles-ci soient identifiables ou non) et les autres, par nature anonymes (qualité des eaux, taux de pollution, objectif national de dépenses d'assurance maladie, numéros clausus...).

Ces données sont mentionnées, dans la loi Informatique et libertés, comme des données appelant une vigilance particulière.

On peut (à titre d'éclairage, sans prétention ici à l'exhaustivité) les classer par leur forme, par leur objet, ou par leurs sources :

- ✓ par leur forme :
 - données brutes (dans un format qui facilite ou non leur réutilisation),
 - données extraites (échantillon de données brutes ou données rendues moins précises dans telles ou telles de leurs dimensions),

- données agrégées et statistiques plus ou moins élaborées (totaux, indicateurs, évolutions, comparaisons...) sous forme statique ou dynamique (tableau qui se met à jour).
- ✓ par leur objet :
 - Informations statistiques sur l'état de santé d'une population, la prévalence et l'incidence des maladies, les variations selon l'âge, le sexe, les catégories professionnelles et sociales, le revenu du ménage, le niveau d'études, certains caractères génétiques, la qualité de l'air ou de l'eau, d'autres facteurs liés à l'environnement, ou selon l'indice de masse corporelle, l'alimentation, la consommation d'alcool ou de tabac, les pratiques sexuelles etc.
 - Informations sur l'offre de soins : établissements et professionnels de santé, classés selon la nature de l'offre (spécialités, équipements, capacités...), l'activité, la distance, la disponibilité, la réputation et d'autres indicateurs de la qualité, les tarifs...
 - Informations sur la dépense de santé (consommation de soins et biens médicaux, indemnités journalières, prévention environnementale et alimentaire, formation des professions de santé, recherche dans le domaine de la santé, dépenses de santé liées au handicap et à la dépendance...) ou informations sur le financement de la dépense de santé (assurance maladie obligatoire, assurances complémentaires, ménages...).
 - Informations sur les professionnels de santé (démographie, activités, revenus...).
 - Informations sur les facteurs de l'état de santé (environnement, données sociales...).
 - etc.
- ✓ Par leurs sources :

L'annexe 4 du rapport Bras reproduit l'annexe du rapport du HCSP de 2009 sur les données de santé qui présentait une liste de sources. On pourrait ajouter à cette liste nécessairement incomplète les grandes cohortes mises en place depuis 2009, des recueils locaux ou régionaux, à commencer par les registres mais aussi les grands référentiels nationaux (Répertoire partagé des professionnels de santé et Fichier national des établissements sanitaires et sociaux). Une cartographie aussi complète que possible des sources de données sera présentée à la commission au cours d'une prochaine réunion.

Qu'appelle-t-on donnée personnelle de santé ?

En l'absence de définition légale, on considère généralement qu'une donnée de santé à caractère personnel est une donnée susceptible de révéler une maladie de la personne ou une prédisposition à la maladie. La définition contenue dans le projet de règlement européen sur la protection des données est plus étendue : « toute information relative à la santé physique ou mentale d'une personne, ou à la prestation de services de santé à cette personne » (article 4 du projet de règlement).

Ainsi le fait de consulter un psychiatre ou d'acheter des médicaments antidiabétiques et de faire réaliser certains examens biologiques est révélateur de l'état de santé d'une personne. Inversement, on peut s'interroger sur le caractère confidentiel de certaines informations de santé pourtant personnelles mais apparentes, telles que la taille ou le poids par exemple.

Qu'est-ce qu'un jeu de données réidentifiantes ?

On en reste à ce stade à une définition très générale, qui sera approfondie avec la présentation du rapport de Pierre-Louis Bras et une présentation, au cours d'une séance ultérieure, des travaux pilotés par la DREES sur les risques de ré-identification.

Les jeux de données réidentifiantes sont des ensembles de données supposées anonymes (ne comportant pas de nom, d'adresse, ni le NIR...) mais permettant d'identifier la personne dès lors que

l'on dispose par ailleurs d'informations sur cette personne qui figurent dans le jeu de données et qui, prises ensemble, lui sont propres (par exemple un parcours de soins, des dates d'hospitalisation, un code postal, un âge...). Si une seule personne, dans une base exhaustive, présente un ensemble de telles caractéristiques :

- elle peut être identifiée à coup sûr,
- mais seulement par un tiers (proche, employeur, assureur, journaliste...) disposant de l'information sur la personne et ayant accès aux données.

5. PERSPECTIVES POUR L'OPEN DATA APPLIQUEE AUX DONNEES DE SANTE

Appliquée au champ des données publiques de la santé, une stratégie d'Open Data concernerait l'ensemble des données publiques qui ont vocation à être mises à la disposition de tous et réutilisées, **au bénéfice de la santé publique, des patients et plus largement de l'information des usagers du service public de la santé**. Elles peuvent concerner l'offre de soins et son accessibilité, son coût et son efficacité, les médicaments et la pharmaco-vigilance, le bon usage des soins, l'épidémiologie, les parcours de soin, mais aussi des facteurs nutritionnels, environnementaux, etc.

Prioritairement, il **importerait d'ouvrir et de partager des données susceptibles de présenter un enjeu démocratique**. A cet effet, les séries complètes, les données permettant de construire des référentiels, les données fréquemment actualisées, les données géolocalisées ou encore les données portant sur la transparence de l'action publique sont particulièrement utiles pour répondre à l'attente des citoyens. Ces informations ont vocation à être présentées sous un format permettant leur traitement automatisé et leur réutilisation libre et gratuite, diffusées à une granularité la plus fine possible, dans le respect le cas échéant des lois en vigueur et notamment de la loi sur le secret statistique. Par ailleurs, elles ont vocation à s'appuyer sur des référentiels partagés et des nomenclatures décrites et publiées.