



GOUVERNEMENT

Liberté
Égalité
Fraternité

MISSION BOTHOREL

Pour une politique publique de la donnée

DECEMBRE 2020

1. Les données et les codes sources : de quoi parle-t-on ?

1.1. Données, bases de données et métadonnées

On trouve une définition juridique de la **donnée** dans un arrêté du 22 décembre 1981, comme « *représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement* »⁴. Une **base de données**, quant à elle, est définie par le droit européen comme un « *recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique et individuellement accessibles par des moyens électroniques ou d'une autre manière* »⁵. La donnée, et *a fortiori* l'**open data**, ne se limite donc pas au champ des documents administratifs défini par la loi CADA de 1978.

Les données sont habituellement traitées afin d'en extraire de l'information au sein d'une base de données ; on parle alors de données « structurées ». Toutefois, l'évolution des techniques d'analyse de données a permis de traiter un éventail plus large des données telles les données non structurées (par exemple les données générées par le passage sur une page internet).

À toute donnée comme à toute base de données peuvent être associées des **métadonnées**, c'est-à-dire des données qui peuvent indiquer la date, le lieu, la méthode de collecte ou toute autre information relative aux données elles-mêmes.

Plusieurs organisations productrices de données font la distinction entre **données « brutes »** et **données « travaillées »** (correspondant par exemple à des statistiques calculées ou estimées à partir des bases de données). Si d'un strict point de vue juridique il n'y a pas lieu d'opérer une telle distinction, elle peut fonder des stratégies d'ouverture différentes.

1.2. Les codes sources, à distinguer des algorithmes

Un code source peut être défini comme un ensemble d'instructions exécutables par un ordinateur. Un code source se différencie d'un algorithme, qui est défini par la CNIL comme « *la description d'une suite finie et non ambiguë d'étapes (ou d'instructions) permettant d'obtenir un résultat à partir d'éléments fournis en entrée* »⁶, par le fait que l'algorithme n'est pas nécessairement informatisé. De manière simplifiée, l'algorithme est une recette de cuisine, et le code sa réalisation concrète : ouvrir l'algorithme est un premier acte de transparence, mais rien ne garantit que la recette a été réellement suivie. C'est en cela que l'ouverture du code est un gage de transparence supplémentaire, en permettant de voir comment le processus est réellement mis en œuvre. L'application ou le programme, enfin, est le produit de cette recette, prêt à être utilisé.

Un « algorithme public » est une procédure administrative dont tout ou partie est informatisée et qui intervient dans un processus de décision pour les citoyens. Un algorithme de répartition des places en crèches avec l'intervention humaine d'un comité d'attribution, combinée au tri automatisé d'une machine qui permet de décider qui en bénéficie, constitue un exemple.

⁴ Vocabulaire des technologies de l'information et de la communication (TIC), Délégation générale à la langue française et aux langues de France, 2017.

⁵ Directive 96/9/CE du Parlement européen et du Conseil du 11 mars 1996 concernant la protection juridique des bases de données.

⁶ Rapport de la CNIL de décembre 2017, *Comment permettre à l'homme de garder la main ?*

Recommandations

Recommandations transversales

Recommandation n° 1 : Initier un débat public sur les conditions de la confiance dans le numérique, permettant de définir les principes fondamentaux de sécurité et de transparence qui doivent s'imposer à la puissance publique

Recommandation n° 2 : Associer la société civile, par les consultations citoyennes et le Forum du Partenariat pour un gouvernement ouvert, à l'identification des jeux de données et des codes sources à ouvrir

Recommandation n° 3 : Conduire une évaluation de l'impact économique, social et scientifique de l'ouverture et du partage des données et des codes sources

Portage de la politique

Recommandation n° 4 : Assurer un portage politique au niveau du Premier ministre des enjeux de la donnée et des codes source. Inscrire à l'ordre du jour des comités interministériels présidés par le Premier ministre le suivi et la mise en œuvre de cette politique. Édicter une circulaire établissant les principes (gouvernance, missions et responsables dans les administrations, interopérabilité, qualité, guides juridiques)

Mise en œuvre de l'ouverture des données et des codes sources

Recommandation n° 5 : Nommer un administrateur général de la donnée, des algorithmes et des codes sources (AGDAC), missionné par le Premier ministre, auprès du DINUM, ayant pour mission à temps plein de piloter la stratégie nationale d'ouverture de la donnée et des codes sources, en s'appuyant sur les administrateurs ministériels des données, des algorithmes et des codes source (AMDAC)

Recommandation n° 6 : Structurer le pilotage et le suivi de la politique d'ouverture des données et des codes sources au niveau interministériel (indicateurs de performance, insertion dans les études d'impact des projets de loi)

Recommandation n° 7 : Engager la puissance publique sur la voie d'une participation plus active aux communs numériques

Recommandation n° 8 : Créer un « Open Source Program Office » (OSPO) ou une mission logiciels libres au sein de TECH.GOUV, chargée d'aider l'administration à ouvrir et à réutiliser les codes sources publics, d'identifier les enjeux de mutualisation et de créer des liens avec les communautés open source existantes et d'accompagner les talents français dans ce domaine

Recommandation n° 9 : Élargir et renforcer la fonction d'administrateur ministériel des données, des algorithmes et des codes sources (AMDAC) :

- en redéfinissant leurs missions dans une fiche de poste type
- en dotant les AMDAC d'une lettre de mission signée par les ministres concernés après consultation des directions générales et de la DINUM
- en s'assurant que l'AMDAC a des moyens d'intervention suffisants
- en systématisant des formations conjointes entre AMDAC et délégués à la protection des données

Recommandation n° 10 : Confier à l'ANCT une mission d'accompagnement des collectivités territoriales dans la publication des données et des codes sources *via* des programmes cofinancés entre État et régions

Recommandation n° 11 : Prendre davantage en compte les démarches d'open source et d'open data pour le rayonnement de la recherche française dans les évaluations et le financement des projets

Droit et régulation

Recommandation n° 12 : Faire évoluer le droit d'accès aux documents administratifs pour renforcer l'effectivité de la loi en confiant un pouvoir de sanction à la CADA en cas de non-respect des dispositions du CRPA relatives à la communication et à la publication des données et documents et pour alléger l'activité de la CADA sur les saisines simples, et pour fluidifier la gestion des dossiers récurrents devant la CADA

Recommandation n° 13 : Évaluer les besoins en ressources humaines de la CNIL pour renforcer son rôle de conseil et d'accompagnement et assortir l'augmentation des moyens correspondant d'un suivi au travers d'indicateurs de performance sur la satisfaction des usagers (dans le cadre du PLF)

Recommandation n° 14 : Prévoir dans les collèges de la CNIL et de la CADA deux personnalités qualifiées compétentes, l'une en matière de sécurité des systèmes d'information et l'autre sur les nouveaux usages de la donnée

Recommandation n° 15 : Associer l'ANSSI à la mise en œuvre de la politique d'ouverture des données et des codes sources afin d'assurer que cette politique n'entre pas en contradiction avec les impératifs de sécurité des systèmes d'information :

- prévoir que la CADA et la CNIL puissent saisir l'ANSSI pour avis quand il y a un doute sérieux en matière de sécurité des systèmes d'information ;
- prévoir la possibilité, pour l'AGDAC de solliciter l'ANSSI pour un audit de bibliothèques et de logiciels libres sensibles

Recommandation n° 16 : Vérifier que la loi garantit l'ouverture de toutes les données de services publics mis en œuvre par des acteurs privés (professions réglementées de la justice notamment)

Acculturation et politique RH

Recommandation n° 17 : Développer une politique de formation de la fonction publique plus ambitieuse sur les enjeux du numérique (obligation de formation des cadres dirigeants aux enjeux du numérique, séminaires de cadres dirigeants, offre de formation pour tous les niveaux hiérarchiques, plans de formation ministériels complémentaires à l'offre interministérielle, modules dans l'ensemble des cursus de formation de la fonction publique)

Recommandation n° 18 : Poursuivre les travaux relatifs à la gestion des emplois et des compétences du numérique et structurer dans la formation initiale une filière technique de la fonction publique pour les métiers experts du numérique, en créant des parcours pour les corps techniques et en pérennisant en CDI les agents contractuels apportant des compétences non disponibles dans les corps existants

Recommandation n° 19 : Diversifier les parcours des administrateurs et des attachés de l'INSEE dans l'ensemble des administrations, au-delà des services statistiques ministériels, et valoriser le travail et la carrière des agents choisissant ces parcours

Recommandation n° 20 : Accroître l'attractivité de l'État pour les métiers du numérique en tension (rendre le référentiel de rémunération obligatoire, développer la communication auprès des formations spécialisées)

Recommandation n° 21 : Passer à l'échelle et inscrire dans la durée le programme d'entrepreneurs d'intérêt général

Recommandation n° 22 : Proposer une offre de formation dédiée aux élus sur les enjeux de la donnée et des codes sources dans les politiques publiques

Qualité de la donnée

Recommandation n° 23 : Créer un label de service producteur de la donnée pour reconnaître les efforts investis dans la donnée, par exemple dans le cadre du service public de la donnée

Recommandation n° 24 : Définir et mettre en œuvre une politique interministérielle d'interopérabilité et de qualité de la donnée (démarches de standardisation, label FAIR, doctrine sur les métadonnées, catalogage)

Recommandation n° 25 : Encourager les écosystèmes à définir des principes de gouvernance de la qualité, en désignant un référent qualité et en créant des communautés de réutilisation avec participation active des producteurs de la donnée

Infrastructures, partage et accès sécurisé

Recommandation n° 26 : Orienter les investissements du plan de relance vers les infrastructures favorables à la circulation de la donnée (appels à projets de la DINUM et appels à projets sectoriels)

Recommandation n° 27 : Encourager la création de « hubs » sectoriels ou intersectoriels, selon des modalités adaptées à chaque secteur, et dans des conditions assurant leur interopérabilité

Recommandation n° 28 : Créer un dispositif de bac à sable expérimental permettant à la CNIL de déroger aux textes existants pour autoriser la réutilisation de données personnelles dans des jeux d'apprentissage d'algorithmes d'intelligence artificielle, et leur conservation pour une durée plus longue que celle autorisée lors de leur collecte initiale

Recommandation n° 29 : Mettre en œuvre les dispositifs techniques permettant d'utiliser la procédure d'appariement de fichiers sur la base du code statistique non signifiant à des fins de statistique publique et de recherche scientifique et historique

Recommandation n° 30 : Améliorer la prise en charge des demandes des chercheurs, en associant les AMDAC et les SSM (délai de réponse obligatoire, création d'un recours, recours à la consultation du comité du secret statistique à titre facultatif)

Données d'intérêt général

Recommandation n° 31 : Privilégier une approche incitative et concertée, le recours à d'éventuels dispositifs coercitifs devant être dûment justifié et faire l'objet d'une évaluation préalable

Recommandation n° 32 : Sécuriser le cadre juridique du partage volontaire de données d'intérêt général concernant l'utilisation des données à caractère personnel (par un guide de conformité de la CNIL) et l'application du droit d'accès et de réutilisation applicable aux données du secteur privé reçues par les administrations

Recommandation n° 33 : Encourager les initiatives de portabilité citoyenne des données au service de l'intérêt général, notamment par l'organisation de campagnes de mobilisation citoyenne

Utilisation par le secteur public de données issues du secteur privé (B2G)

Recommandation n° 34 : Clarifier le régime juridique de la réquisition pour permettre à la puissance publique d'accéder à des données du secteur privé en cas de motif impérieux d'intérêt général et d'urgence

Recommandation n° 35 : Confier au réseau de l'AGDAC et des AMDAC une mission de facilitation et de médiation de l'accès et de l'utilisation des données du secteur privé par le secteur public (B2G), en lien avec la direction générale des entreprises (DGE)

Recommandation n° 36 : Garantir l'effectivité des dispositions relatives aux données d'intérêt général de la loi pour une République numérique qui rencontrent des difficultés d'application :

- en matière de données détenues par les concessionnaires et délégataires du service public (clausiers types pour les acteurs publics)
- en matière d'utilisation des données privées à des fins statistiques, étudier l'opportunité d'élargir l'article 19 à certains services fondés sur les données

Partage de données entre acteurs privés (B2B)

Recommandation n° 37 : Développer le partage de données privées au service d'intérêts partagés (B2B) au sein des comités stratégiques de filières, dans les appels à projets publics (PIA), et en soutenant les initiatives associatives et privées

SOCIÉTÉ

Sélections

Partage

f

✉

🔄

Parcoursup : le Défenseur des droits enquête sur des soupçons de discrimination

Estimant que la plate-forme avait défavorisé des candidats issus de quartiers populaires, des particuliers et des élus avaient saisi l'institution.

Par Camille Stromboni

Publié le 24 août 2018 à 11h30 - Mis à jour le 24 août 2018 à 12h52 - 🕒 Lecture 2 min.

🔒 Article réservé aux abonnés

Parcoursup a-t-il été discriminant envers certains lycéens ? C'est l'une des interrogations auxquelles le Défenseur des droits a décidé de répondre en ouvrant une instruction sur le fonctionnement de la nouvelle plate-forme d'admission dans l'enseignement supérieur, à la suite des réclamations reçues. « J'entends traiter ces questions de manière prioritaire », écrit même Jacques Toubon, dans un courrier en date du 8 août adressé à Stéphane Troussel, président du conseil départemental de Seine-Saint-Denis.

Lire aussi | 📄 [Parcoursup : le gouvernement au pied du mur à l'approche de la rentrée](#)

L'élue socialiste a saisi l'institution en juillet, s'inquiétant de « l'opacité » des « algorithmes locaux » mis en place dans les établissements pour classer les dossiers des 812 000 candidats aux études, tout en pointant son sentiment d'une « discrimination supplémentaire à l'égard des jeunes de quartiers populaires ». « J'espère que nous pourrions enfin avoir les réponses aux questions que nous posons depuis le mois de mai », se félicite le président du département le plus jeune et défavorisé de France métropolitaine, qui a interpellé la ministre de l'enseignement supérieur, Frédérique Vidal, au printemps.

Au lendemain des premières réponses des formations aux candidats, le 22 mai, certains lycéens, enseignants et syndicats ont accusé les nouvelles règles d'avoir pu défavoriser celles et ceux qui ne viendraient pas du bon lycée ou du bon territoire. Notamment les jeunes de banlieue parisienne, qui s'estimaient également évincés des filières de Paris intra-muros. Des discriminations alors démenties du côté du gouvernement.

Absence de transparence

Avec la réforme de la loi orientation et réussite des étudiants, les universités – comme les filières sélectives le faisaient déjà – ont procédé, pour la première fois, au classement des dossiers des candidats. Les critères employés par chaque établissement ont été multiples : bulletins de notes, lettre de motivation, CV, avis du conseil de classe de terminale, activités extrascolaires... Mais les établissements n'ont pas eu à rendre public le détail de ces processus, mis en place en quelques semaines et restés sous le sceau du « secret des délibérations ».

Lire aussi | 📄 [Les lycéens de banlieue et les embûches de Parcoursup](#)

Édition du jour

Daté du mercredi 16 mars



[Lire le journal numérique](#)

[Les éditions précédentes](#)