



**Conseil national
de l'information statistique**

GROUPE DE TRAVAIL DU CNIS

**RAPPORT AU SECRÉTAIRE D'ÉTAT CHARGÉ DE L'INDUSTRIE,
DU NUMÉRIQUE ET DE L'INNOVATION**

L'ACCÈS DES CHERCHEURS AUX DONNÉES ADMINISTRATIVES

État des lieux et propositions d'actions

M. ANTOINE BOZIO (EHESS, PSE)

Directeur de l'Institut des politiques publiques (IPP)

M. PIERRE-YVES GEOFFARD (CNRS, EHESS, PSE)

Directeur de l'École d'Économie de Paris

- MARS 2017 -

*A la mémoire de Dominique Allain,
disparue trop tôt pour voir
l'aboutissement de ce travail
qui lui doit beaucoup.*

Sommaire

PRINCIPAUX ACRONYMES UTILISÉS	5
S YNTHÈSES ET PRINCIPALES RECOMMANDATIONS	7
INTRODUCTION	10
1. L'ACCÈS DES CHERCHEURS AUX DONNÉES ADMINISTRATIVES : ÉTAT DU DROIT	13
1.1. En principe, les données publiques sont librement communicables	14
1.2. Les données couvertes par un secret protégé par la loi ne sont communicables qu'à l'expiration des délais fixés par le code du patrimoine, sauf demande d'accès anticipé	16
1.2.1. Les secrets protégés par la loi	16
1.2.2. L'accès anticipé aux données protégées	19
1.2.2.1. Le régime de droit commun pour accéder de manière anticipée aux données protégées	20
1.2.2.2. Les régimes dérogatoires pour accéder de manière anticipée aux données protégées	21
1.2.2.2.1. S'agissant des données statistiques	21
1.2.2.2.2. S'agissant des données fiscales	23
1.2.2.2.3. S'agissant des données du système national des données de santé	23
1.2.2.2.4. S'agissant des données détenues par la Banque de France	24
1.3. Le traitement des données à caractère personnel est soumis à la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	26
1.3.1. Le champ d'application de la loi Cnil est particulièrement vaste	27
1.3.2. Règles et formalités préalables à tout traitement de données à caractère personnel	27
1.3.2.1. Les conditions de licéité des traitements de données à caractère personnel	27
1.3.2.2. Les formalités préalables au traitement de données à caractère personnel	28
1.3.2.3. Les opérations d'anonymisation	29
1.3.3. Le traitement de certaines données est soumis à un régime restrictif	30
1.3.3.1. Le traitement des données dites « sensibles »	30
1.3.3.2. Le traitement des données relatives aux infractions, condamnations et mesures de sûreté	31
1.4. La loi pour une République numérique entend favoriser l'accès des chercheurs aux données publiques	31
1.4.1. Un appariement des données facilité	32
1.4.2. L'ouverture de la saisine du comité du secret statistique	33
1.4.3. L'absence de sanctions pénales pour les administrations en cas de non-respect du secret professionnel par les chercheurs	34
2. COMPARAISONS INTERNATIONALES	35
2.1. Périmètre et méthodologie de la comparaison	35
2.2. Des États à « registres »	37
2.2.1. L'exemple du Danemark	37
2.2.2. L'exemple des Pays-Bas	38
2.3. Des systèmes statistiques non centralisés	38
2.3.1. L'exemple du Royaume-Uni	38
2.3.2. L'exemple des Etats-Unis	39
2.3.3. L'exemple de l'Allemagne	40
2.4. La situation de la France	40
3. AMÉLIORER L'ACCÈS DES CHERCHEURS AUX DONNÉES PUBLIQUES	42
3.1. Des dispositifs législatifs complexes et mal connus	42
3.1.1. Une large méconnaissance des dispositifs législatifs d'accès aux données	42
3.1.2. Des dispositifs législatifs et réglementaires qui peuvent encore être améliorés	43
3.1.3. Les conditions de réussite de la loi pour une République numérique	44
3.2. Améliorer les modalités pratiques d'accès aux données	45
3.2.1. Le besoin d'une instance d'une concertation unique	45
3.2.2. L'identification des données existantes est difficile	46
3.2.3. La diversité des procédures et des modalités d'accès aux données	47

3.2.4.	Favoriser les partenariats de recherche entre administrations productrices et chercheurs	48
3.3.	Développer un modèle économique pérenne de l'accès aux données administratives	49
3.3.1.	Les coûts associés à la préparation des données par les producteurs	49
3.3.1.1.	La suppression des identifiants individuels	50
3.3.1.2.	L'extraction de données de gestion	50
3.3.1.3.	La documentation des données administratives	51
3.3.1.4.	Quels financements pour les coûts de préparation des données?	51
3.3.2.	Les coûts associés à l'accès aux données	52
3.4.	Améliorer les pratiques d'archivage	54
3.4.1.	Archivage des données publiques : état du droit	54
3.4.1.1.	Définition et champ d'application	54
3.4.1.2.	Sélection, conservation et format des données à archiver	56
3.4.2.	Difficultés pratiques	58
3.4.2.1.	Etat des lieux des difficultés rencontrées	58
3.4.2.2.	Facteurs d'explication	59
3.4.2.3.	Pistes de réflexion	59
4.	CONCLUSION	61
	ANNEXES	62

PRINCIPAUX ACRONYMES UTILISÉS

ACOSS	Agence centrale des organismes de sécurité sociale
ADAMANT	Administration des Archives et de leurs Métadonnées aux Archives nationales dans le Temps
AGIRC	Association générale des institutions de retraite des cadres
ARRCO	Association pour le régime de retraite complémentaire des salariés
CADA	Commission d'accès aux documents administratifs
CASD	Centre d'accès sécurisé aux données
CE	Conseil d'État
CNAF	Caisse nationale des allocations familiales
CNAV	Caisse nationale de l'assurance vieillesse des travailleurs salariés
Cnis	Conseil national de l'information statistique
Cnil	Commission nationale de l'informatique et des libertés
CNNum	Conseil national du numérique
CNRS	Centre national de la recherche scientifique
CSNS	Code statistique non signifiant
CRPA	Code des relations entre le public et l'administration
DEPP	Direction des études, de la prospective et de la performance (ministère de l'éducation nationale)
DACS	Direction des affaires civiles et du sceau
DGFIP	Direction générale des finances publiques
DG Trésor	Direction générale du Trésor
DINSIC	Direction interministérielle du numérique et du système d'information et de communication de l'État
DPJJ	Direction de la protection judiciaire de la jeunesse
DUA	Délai d'utilisation administrative / durée d'utilisation administrative
Etalab	Service du Premier ministre, rattaché au SGMAP, chargé de l'ouverture des données publiques
EHESS	École des hautes études en sciences sociales
GENES	Groupe des écoles nationales d'économie et statistique
IDS	Institut des données de santé
INSEE	Institut national de la statistique et des études économiques
IPP	Institut des politiques publiques
LPF	Livre des procédures fiscales

MEN	Ministère de l'éducation nationale
MinEFI	Ministère de l'Économie, des finances et de l'industrie
NIR	Numéro d'inscription au répertoire national d'identification
Progedo	TGIR dont la mission est d'organiser l'appui à la collecte, la documentation, la préservation et la diffusion d'ensembles de données mobilisables par les méthodes quantitatives utiles aux sciences humaines et sociales
RNIPP	Répertoire national d'identification des personnes physiques
RGI	Référentiel général d'interopérabilité
SEDA	Standard d'échange de données pour l'archivage
SIRENE	Système d'identification du répertoire des établissements
SGMAP	Secrétariat général pour la modernisation de l'action publique
SIAF	Service interministériel des archives de France
SIARD	Software Independent Archiving of Relational Databases
SIES	Sous-direction des systèmes d'information et des études statistiques (ministère de l'éducation nationale)
SNIIRAM	Système national d'information inter-régimes de l'assurance maladie
SSA	Services statistiques académiques
TeraLab	Projet initié par l'Institut Mines-Télécom et le GENES visant à favoriser la recherche et l'innovation dans le domaine numérique du Big Data
TGIR	Très grande infrastructure de recherche
VITAM	Valeurs immatérielles transmises aux Archives pour mémoire

SYNTHÈSES ET PRINCIPALES RECOMMANDATIONS

Le groupe de travail a élaboré plusieurs recommandations visant à améliorer l'exploitation des données administratives à des fins de recherche scientifique. Ces recommandations peuvent se regrouper sous quatre grands thèmes : faire connaître le droit, favoriser la concertation entre producteurs et utilisateurs, développer un modèle économique viable, et améliorer les pratiques d'archivage des administrations.

Faire connaître le droit régissant l'accès aux données administratives

Le constat principal du groupe de travail est que le droit régissant l'accès aux données administratives est mal connu, aussi bien des chercheurs, potentiels utilisateurs, que des administrations productrices. L'enjeu est donc important de mieux faire connaître les dispositions législatives et réglementaires, issues de textes multiples, en proposant une synthèse adaptée au besoin de la recherche scientifique. Le rapport propose à la partie 1 une mise à plat de l'état du droit, qui pourrait servir de base à un document synthétique à diffuser largement.

Recommandation n°1 : élaborer et diffuser un vade-mecum synthétique à destination des administrations et des chercheurs, rappelant l'état du droit applicable, présentant les différentes procédures d'accès et les voies de recours à disposition des chercheurs pour contester une décision de refus d'accès.

Le constat général du rapport est que les différentes mesures législatives de ses dernières années ont levé l'essentiel des obstacles juridiques à l'accès aux données administratives à des fins de recherche. Le rapport relève néanmoins un certain nombre de cas pour lesquels le droit actuel pourrait être amélioré (possibilité de recours, cas du traitement des données judiciaires, etc.).

Recommandation n° 2 : le législateur pourrait rapprocher les procédures législatives d'accès et de traitement des données couvertes par un secret.

Recommandation n° 3 : inciter les administrations à faire pleinement usage de la faculté qui leur est ouverte par la loi pour une République numérique de saisir le comité du secret statistique.

Favoriser la concertation entre producteurs et utilisateurs

La conviction forte du groupe de travail est qu'il est possible d'améliorer substantiellement la pratique de l'accès des chercheurs aux données administratives en favorisant une plus grande concertation des différents acteurs. L'objectif est à la fois de faciliter l'accès des chercheurs par des procédures harmonisées et une meilleure information, et aussi de mieux valoriser auprès des administrations les travaux des chercheurs, afin que le caractère mutuellement bénéfique de l'accès aux données administratives devienne une évidence partagée par tous. L'instance de concertation, que le groupe de travail appelle de ses vœux, devrait être le moteur principal de ces échanges indispensables.

Recommandation n° 4 : mettre en place une instance de concertation unique, regroupant des représentants des chercheurs et des administrations, ainsi que les centres sécurisés d'accès aux données et les organes interministériels compétents en matière de systèmes d'information et d'accès aux données et à leur traitement.

Recommandation n° 5 : inciter, *a minima*, les administrations à enrichir le répertoire des informations publiques visé à l'article L. 322-6 du code des relations entre le public et l'administration et engager, en coordination avec les services producteurs de données, l'administration des archives et les chercheurs, un processus plus large de recensement des principales sources de données existantes.

Recommandation n° 6 : harmoniser les formulaires existants qui doivent être remplis par les chercheurs pour accéder aux données publiques.

Recommandation n° 7 : prévoir des procédures d'accès sécurisé qui rendent techniquement possible l'appariement des bases de données.

Recommandation n° 8 : encourager les partenariats entre chercheurs et administrations, en insistant sur le caractère mutuellement bénéfique d'un accès large aux données administratives à des fins de recherche.

Promouvoir un modèle économique viable de l'accès aux données administratives

L'accès aux données administratives permet d'exploiter des données existantes d'une très grande richesse. Il s'agit donc d'information dont le coût est en grande partie déjà payé par les administrations productrices. Pour autant, l'accès par les chercheurs à ces données révèle un coût économique certain : extraction, dé-identification et documentation par les administrations, coût des procédures d'accès, travail statistique pour transformer des données de gestion en objet de recherche, etc.

Le groupe de travail a souhaité clarifier l'ampleur de ces coûts et a proposé des pistes pour un modèle économique viable. L'idée principale est que des moyens publics doivent être alloués à la prise en charge de ces coûts, du côté des administrations pour la mise à disposition de ces données, et du côté de la recherche pour l'accès sécurisé à ces données. Dans tous les cas, les options permettant de limiter ces coûts pour la société doivent être privilégiées.

Recommandation n° 9 : prévoir dans les conventions d'objectif et de gestion des organismes sous tutelle des moyens dédiés aux travaux concourant à la mise à disposition de données administratives.

Recommandation n° 10 : encourager, afin d'en évaluer l'intérêt pour la recherche, une première exploration des données brutes, simplement dé-identifiées.

Recommandation n° 11 : faire un bilan de l'impact de l'ouverture des possibilités de saisine du Comité du secret statistique sur son fonctionnement et proposer les ajustements nécessaires (organisation pratique, moyens) et prévoir un budget de fonctionnement pour l'instance de concertation.

Recommandation n° 12 : les financements publics de la recherche doivent prendre en compte la nécessité de financer les coûts d'accès aux données administratives protégées par des secrets.

Améliorer les pratiques d'archivage des administrations.

Le groupe de travail a pu constater que l'archivage des données administratives était déficient en raison d'un manque de connaissances du droit par les administrations. Rappeler à toutes les administrations productrices que les dispositions de la Cnil n'imposent pas la destruction des données, mais leur archivage, devrait conduire à mieux conserver notre patrimoine pour les générations futures, dont les futurs historiens.

Recommandation n° 13 : sensibiliser les administrations sur la question de l'archivage pérenne des données qu'elles détiennent, en coordination avec l'administration des archives et conformément au code du patrimoine, notamment en leur rappelant que les données peuvent, en vertu de l'article 36 de la loi de 1978, être conservées à l'issue de la durée de conservation déclarée ou fixée par la Cnil à des fins scientifique, statistique ou historique.

Recommandation n° 14 : établir une coopération entre les archives et les centres sécurisés pour améliorer les modalités d'archivage des fichiers résultant des opérations d'appariements.

Recommandation n° 15 : associer davantage les chercheurs de différentes disciplines au processus de sélection des données qui méritent d'être archivées.

INTRODUCTION

La révolution numérique a transformé la façon dont les entreprises et les administrations fonctionnent au quotidien, et ce faisant elle a entraîné la production massive de données numériques. Ce qui est moins connu, c'est que l'exploitation de ces données à des fins de recherche a conduit à une révolution majeure dans les sciences humaines et sociales : la capacité d'évaluer l'impact des politiques publiques a été multipliée et des pans entiers de la connaissance ont été renouvelés par cet apport. Aujourd'hui, on peut affirmer sans exagération que les données administratives sont une des matières premières essentielles aux travaux empiriques les plus en pointes. A titre d'illustration, les données administratives utilisées dans les publications de la meilleure revue internationale en économie, l'*American Economic Review*, sont passées de 5% de l'ensemble des travaux empiriques publiés en 2008 à 26% en 2014¹.

Tous les pays n'ont pas saisi cette opportunité aussi rapidement. Les pays nordiques (Danemark, Suède et Norvège) ont très tôt ouvert leurs données administratives aux chercheurs, et une multiplication de travaux sur ces pays a eu lieu, de chercheurs locaux mais aussi internationaux, y-compris des meilleures universités américaines. En économie, par exemple, une part significative des travaux les plus cités ces dernières années, vient de travaux réalisés sur les données danoises, suédoises ou norvégiennes. Le constat était si frappant que des chercheurs américains, parmi les plus influents, ont jugé nécessaire d'envoyer une lettre ouverte à la *National Science Foundation* pour déplorer la perte du leadership américain – rien de moins – en raison des difficultés d'accès aux données administratives aux États-Unis², et qu'une commission a été chargée de réaliser un audit sur ces questions³.

La France, si elle n'a pas l'organisation centralisée des données des pays à registre comme les pays nordiques, possède néanmoins des données administratives nombreuses et de grande qualité. Elle a été pionnière dans l'utilisation à des fins de recherche de certaines de ces données, sous la houlette de l'Insee, qui cumulait la fonction de production de données et d'études. Néanmoins, cette avance initiale n'a pas été maintenue en raison des difficultés d'accès des chercheurs aux données. Dans le cadre d'un droit complexe, tiraillé entre la nécessaire protection de la vie privée, les secrets professionnels et les exigences nouvelles de l'*open data*, les accès des chercheurs aux bases de données administratives sont longtemps restés lettre morte.

Depuis quelques années, le législateur a amorcé un changement de cap visant explicitement à faciliter l'accès des chercheurs aux données administratives : le mouvement a commencé en 2008 par les données détenues par la Statistique publique – Insee et services statistiques ministériels –, en 2013 pour les données fiscales, en 2016 pour les données de santé, et enfin cette même année avec la loi pour une République numérique pour les autres données produites par des administrations et les institutions réalisant des services d'intérêt général. Ce mouvement d'ouverture a pu se réaliser en proposant des modes d'accès aux données extrêmement sécurisés, garantissant que les travaux de recherche sur ces larges bases de données ne puissent être utilisés à des fins d'identification des individus. Ce mouvement d'ouverture place aujourd'hui la France parmi les pays en pointe dans le domaine – c'est un

¹ Einav, L. et Levin, J. (2014), « Economics in the age of big data », Vol. 346, No. 6210, *Science*.

² Card, D., Chetty, R., Feldstein, M. et E. Saez (2010), « Expanding Access to Administrative Data for Research in the United States », NSF white paper.

³ *Commission on Evidence-Based Policymaking* (CEP) établie par le *Evidence-Based Policymaking Commission Act of 2016* (P.L. 114-140), 30 mars 2016.

constat du rapport –, même si l'organisation très morcelée des bases administratives en France ne permet pas d'égaliser la situation des pays nordiques.

Ces avancées doivent impérativement être maintenues et poursuivies. Les gains pour la société française sont nombreux, même s'ils ne sont pas toujours visibles : gains pour la connaissance avec des travaux de recherche, gains pour les administrations avec une exploitation de leurs données qui complètent leurs propres études, gains pour les décideurs qui disposeront d'évaluations de politiques publiques plus nombreuses, et enfin gains pour l'ensemble de la société, les meilleures connaissances de nos politiques publiques conduisant *in fine* à des interventions plus efficaces et à un débat démocratique plus solide.

Ce rapport, commandé par la Secrétaire d'État auprès du ministre de l'Économie et des Finances chargée du Numérique et de l'Innovation, Mme Axelle Lemaire, sous l'égide du Cnis, et de sa présidente Mme Yannick Moreau, est le produit d'un groupe de travail du Cnis qui s'est réuni intensément de septembre à décembre 2016, dans un calendrier serré, afin de faire un bilan, en forme de point d'étape, du mouvement d'ouverture des données administratives aux chercheurs.

Les conclusions du rapport mettent en évidence l'utilité d'un tel travail, dont on peut espérer qu'il fasse référence sur la question, et qu'il soit lu, par les chercheurs comme par les administrations.

Premièrement, le droit sur l'accès aux données administratives à des fins de recherche est complexe, mouvant, mais surtout très mal connu, des administrations productrices, comme des chercheurs. Si une des recommandations est la réalisation d'un vade-mecum des dispositifs juridiques en vigueur, une part importante du rapport est consacrée à la mise à plat de ce droit complexe, où le régime de droit commun est moins connu que les nombreux cas dérogatoires.

Deuxièmement, si le rapport fait le constat d'une absence large d'obstacles juridiques – avec quelques exceptions – à l'accès aux données par les chercheurs, les difficultés pratiques restent notables. Les dispositifs législatifs ayant été mis en place au cas par cas, selon le type de données et le secret particulier qui les protégeait, les procédures d'accès sont diverses, pas nécessairement unifiées. Le rapport préconise la création d'une instance de concertation impliquant les producteurs de données, les chercheurs, et les administrations compétentes pour la protection et l'accès aux données publiques (Archives, Cnil, administrateur des données, Cada). Cette instance, qui devrait faciliter via une interface interactive les retours d'expérience des chercheurs et les souhaits des administrations, a vocation à répondre au constat d'un manque général de coordination des différentes filières. Elle pourrait ainsi faciliter les ponts, que l'on souhaite développer, entre les mondes de la recherche et de l'administration.

Troisièmement, le rapport rappelle une évidence, parfois oubliée : l'accès à ces données administratives a un coût. Il y a le coût de support pour les instances (de concertation, le comité du secret), le coût d'accès via des dispositifs sécurisés et le coût de préparation de données anonymisées ou dé-identifiées. Le rapport étudie plusieurs pistes de modèle économique, tout en soulignant l'importance à limiter au maximum ces coûts pour éviter tout gaspillage de ressources limitées.

Enfin, le rapport n'oublie pas l'avenir, et nos futurs collègues historiens, en consacrant une partie à la question de l'archivage de ces données administratives. Là encore, la méconnaissance du droit semble conduire à des décisions dommageables pour la sauvegarde de notre patrimoine. Il semble qu'une confusion largement répandue dans les administrations existe sur les contraintes de destruction des données imposées par la Cnil, et l'obligation d'archivage de ces mêmes données. Le droit est là très clair : à l'issue de la période de conservation des données par les administrations définies par la Cnil, celles-ci doivent *archiver*,

et non détruire ces données. Un rappel large de ce principe mériterait d'être fait dans l'ensemble des administrations productrices.

1. L'ACCÈS DES CHERCHEURS AUX DONNÉES ADMINISTRATIVES : ÉTAT DU DROIT

Les données produites par les administrations publiques, y compris les données conservées sous un format numérique, peuvent être regardées comme des documents administratifs (cf. définition dans l'encadré 1) dont l'accès est régi par deux ensembles de dispositions législatives coordonnées :

- les premières sont issues de la loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal (dite loi CADA) qui sont désormais codifiées au sein du livre III du code des relations entre le public et l'administration (CRPA) ;
- les secondes sont issues de la loi n° 2008-696 du 15 juillet 2008 relative aux archives, codifiée au sein du code du patrimoine.

Encadré 1. Définitions

Archives

« *Les archives sont l'ensemble des documents, y compris les données, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité.* » (article L.211-1 du code du patrimoine).

La mention « y compris les données » a été ajoutée par l'article 59 de la loi n° 2016-925 du 7 juillet 2016 relative à la liberté de la création, à l'architecture et au patrimoine.

Base de données

« *On entend par base de données 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 par tout autre moyen* » (article L.112-3 du code de la propriété intellectuelle).

Document administratif

« *Sont considérés comme documents administratifs, au sens des titres Ier, III et IV du présent livre, 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'État, 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, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions, codes sources et décisions* » (article L. 300-2 du code des relations entre le public et l'administration).

Données à caractère personnel

« *Constitue une 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. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de*

permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne. [...] Constitue un fichier de données à caractère personnel tout ensemble structuré et stable de données à caractère personnel accessibles selon des critères déterminés » (article 2 de la loi n°78-17 du 6 janvier 1978).

Données sensibles

Les données sensibles sont les « *données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci* » ainsi que « *données à caractère personnel relatives aux infractions, condamnations et mesures de sûreté* » (articles 8 et 9 de la loi n° 78-17 du 6 janvier 1978).

Traitement

« *Constitue un traitement de données à caractère personnel toute opération ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé, et notamment la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, ainsi que le verrouillage, l'effacement ou la destruction* »(article 2 de la loi n°78-17 du 7 janvier 1978).

Anonymisation

Les opérations d'anonymisation s'entendent comme permettant de rendre une base de données non indirectement identifiante, c'est-à-dire qu'au-delà de la suppression des informations nominatives (nom, numéro identifiant, etc.), la base de données ne dispose pas d'informations permettant par leur croisement une identification indirecte des personnes. Une base ainsi anonymisée n'est plus une base de données à caractère personnel.

De-identification

Les opérations de de-identification sont définies, dans ce rapport, comme la suppression des données nominatives, permettant une identification directe. Une base de donnée dé-identifiée n'est donc pas forcément anonyme (au sens précisé ci-dessus).

1.1. En principe, les données publiques sont librement communicables

Il résulte des dispositions combinées des articles L. 311-1 du CRPA et L. 213-1 du code du patrimoine que les informations produites par les administrations visées à l'article L. 300-2 du CRPA sont librement communicables, sauf si elles contiennent des secrets protégés par la loi.

Les institutions productrices de données concernées par ces dispositions ne se limitent pas aux administrations visées à l'article L. 300-2 du CRPA, mais également les personnes privées dans l'exercice d'une mission de service public (régimes complémentaires de retraite, certaines associations, etc.)⁴.

⁴ Au sens de la doctrine de la Commission d'accès aux documents administratifs (CADA), telle que validée par le Conseil d'État. Les personnes morales de droit privé ne sont en effet pas exclues, de ce seul fait, du champ du livre III du code des relations entre le public et l'administration. Lorsqu'elles exercent une mission de service public, les documents qu'elles produisent ou reçoivent dans ce cadre sont susceptibles de revêtir un caractère administratif

Le demandeur a le choix des formes sous lesquelles s'effectue la communication. L'article L. 311-9 du CRPA prévoit quatre modes d'accès aux documents administratifs :

- la consultation gratuite sur place des documents, le demandeur pouvant alors être accompagné d'une tierce personne ;
- leur reproduction aux frais du demandeur et l'envoi, à son domicile ou à l'adresse postale indiquée ;
- l'envoi par courrier électronique et sans frais, lorsque le document est disponible sous forme numérique : l'administration est tenue de fournir une copie identique, tant du point de vue du support que du format, à celle ou à l'une de celles dont elle dispose ou est susceptible de disposer à l'issue d'une opération de transfert ou de reproduction courante ;
- depuis l'entrée en vigueur de la loi pour une République numérique, la publication des informations en ligne, à moins que les documents ne soient communicables qu'à l'intéressé en application de l'article L. 311-6 du CRPA.

Le principe du libre choix du mode d'accès souffre néanmoins plusieurs tempéraments qui ont vocation à concilier le droit d'accès avec le bon fonctionnement du service public.

- en premier lieu, le demandeur ne saurait exiger de l'administration qu'il saisisse qu'elle communique le document demandé sous une forme incompatible avec les moyens matériels dont elle dispose. Le droit d'accès s'exerce donc dans la limite des possibilités techniques de l'administration. En particulier, l'administration n'a pas l'obligation de numériser des documents disponibles uniquement sous format papier pour les faire parvenir par voie électronique, ni de se doter des moyens de graver des CD-Rom ou des DVD-Rom. Toutefois, lorsque les contraintes humaines et matérielles de l'administration portent sur les moyens de reprographie sous format papier, celle-ci est tenue de faire appel à un prestataire extérieur et de faire parvenir le devis correspondant au demandeur afin qu'il y donne suite, s'il le souhaite ;
- en deuxième lieu, l'état de conservation des documents peut justifier une restriction aux modalités d'accès. Il en va ainsi lorsque leur reproduction ou leur consultation risquerait de les endommager ;
- en troisième lieu, en cas de demande portant sur un volume de documents important, l'administration peut inviter le demandeur à venir les consulter sur place pour n'emporter copie que des pièces qu'il souhaite. Elle peut, à cette fin, élaborer un règlement de consultation précisant les modalités de cette voie d'accès (horaires, lieux, présence d'un agent...). Si le demandeur maintient son souhait de recevoir copie des documents, l'administration est fondée à en échelonner l'envoi dans le temps : elle doit alors en aviser l'intéressé et, dans la mesure du possible, convenir avec lui d'un échéancier de communication. L'administration doit s'efforcer de respecter le délai d'un mois qui lui est imparti pour satisfaire la demande, sauf si le volume des documents demandés y fait manifestement obstacle ;
- enfin, alors même qu'elle dispose des moyens humains et matériels nécessaires, l'administration peut choisir de recourir à un prestataire extérieur si elle estime que

s'ils présentent « un lien suffisamment direct avec la mission de service public » (CE, 17 avril 2013, *La Poste c. M. Bigi*, 342372).

la charge de travail induite par une demande risquerait de perturber le fonctionnement des services.

L'article R. 311-11 du CRPA prévoit en outre que l'administration peut exiger le paiement de frais correspondant au coût de reproduction et, le cas échéant, au coût de l'affranchissement postal⁵. En revanche, le coût du personnel mobilisé pour assurer la collecte, la reproduction et l'envoi des documents ne peut être mis à la charge du demandeur. Enfin, ces plafonds ne peuvent s'appliquer lorsque l'administration a fait appel à un prestataire extérieur en raison de moyens techniques limités.

Ces dispositions, conçues et bien adaptées pour des documents conservés sous format papier, ne permettent pas toujours de prendre en compte les spécificités des bases de données informatisées (en particulier leur anonymisation ou les extractions complexes, cf. partie 4.4.1).

1.2. Les données couvertes par un secret protégé par la loi ne sont communicables qu'à l'expiration des délais fixés par le code du patrimoine, sauf demande d'accès anticipé

1.2.1. Les secrets protégés par la loi

Certaines données produites par les administrations visées à l'article L. 300-2 du CRPA ne peuvent être communiquées, en application des dispositions des articles L. 311-5 et L. 311-6 du même code.

Aux termes du 1° de l'article L. 311-5, la communication de certains documents administratifs est, du fait de leur nature, proscrite⁶. L'on y trouve notamment les avis du Conseil d'État et des juridictions administratives, les documents émis ou reçus par la Cour des comptes et les chambres régionales des comptes dans le cadre de leur pouvoir de contrôle, les documents élaborés ou détenus par l'Autorité de la concurrence dans le cadre de l'exercice de ses pouvoirs d'enquête, d'instruction et de décision, les documents élaborés ou détenus par l'Autorité pour la transparence de la vie publique dans le cadre des missions prévues à l'article 20 de la loi n° 2013-907 du 11 octobre 2013.

La communication des documents doit en outre être refusée lorsqu'elle pourrait porter atteinte à l'exercice des activités régaliennes de l'État et à l'intérêt général. Le 2° de l'article L. 311-5 du CRPA énumère les secrets protégés par la loi⁷, dont les contours ont été précisés par la jurisprudence du Conseil d'État. Y sont en particulier visés : le secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif, le secret de la défense nationale, la conduite de la politique extérieure de la France, la sûreté de l'État, à la sécurité publique, à la sécurité des personnes ou à la sécurité des systèmes d'information des

⁵ Ce dernier inclut le coût du support et le coût d'amortissement et de fonctionnement du matériel utilisé, plafonné par l'arrêté du 1er octobre 2001 à 0,18 € par photocopie de format A4, en impression noir et blanc, 1,83 € pour une disquette et 2,75 € pour un cédérom ; le coût de reproduction des autres supports est fixé en fonction de leur prix exact et conformément aux principes posés par le décret du 30 décembre 2005.

⁶ Avant l'intervention de l'ordonnance du 29 avril 2009, ces documents étaient exclus du champ d'application de la loi du 17 juillet 1978. La CADA se déclarait incompétente pour connaître des demandes de communication de tels documents. La commission est donc désormais compétente mais elle émet un avis défavorable à la communication de ces documents.

⁷ Soulignons qu'un document ne peut, du seul fait de sa nature, faire l'objet d'un refus de communication sur le fondement d'un des alinéas du 2° de cet article. L'appréciation se fait au niveau du contenu des documents : il convient de rechercher si, eu égard au contenu des documents demandés, leur divulgation risque effectivement de porter atteinte à un secret protégé par ces dispositions (CE, 22 février 2013, *Fédération chrétienne des témoins de Jehovah France*, n°s 337987 et s.).

administrations, la monnaie et au crédit public, le déroulement des procédures engagées devant les juridictions ou d'opérations préliminaires à de telles procédures.

Enfin, l'article L. 311-6 dispose que seul l'intéressé peut obtenir communication des documents administratifs :

- dont la communication porterait atteinte à la protection de la vie privée, au secret médical et au secret en matière commerciale et industrielle ;
- portant une appréciation ou un jugement de valeur sur une personne physique, nommément désignée ou facilement identifiable ;
- ou faisant apparaître le comportement d'une personne, dès lors que la divulgation de ce comportement pourrait lui porter préjudice⁸.

Lorsqu'un document est intégralement concerné par l'un des secrets visés aux articles L. 311-5 et L. 311-6 du CRPA, il ne peut être communiqué. En revanche, l'administration ne peut refuser la communication d'un document au seul motif qu'il comporterait une mention couverte par l'un des secrets protégés par la loi. L'article L. 311-7 du code des relations entre le public et l'administration (ancien III de l'article 6 de la loi du 17 juillet 1978, dans sa rédaction résultant de l'ordonnance du 6 juin 2005) prévoit en effet que : « *Lorsque la demande porte sur un document comportant des mentions qui ne sont pas communicables en application du présent article mais qu'il est possible d'occulter ou de disjointer, le document est communiqué au demandeur après occultation ou disjonction de ces mentions* ». Cette communication partielle est donc subordonnée à deux conditions : le document doit être divisible, c'est-à-dire qu'il doit permettre de procéder en pratique à l'occultation ; l'occultation ne doit pas dénaturer le sens du document ni priver d'intérêt la communication⁹.

En sus des dispositions du CRPA, il existe un ensemble de dispositions législatives concernant le secret professionnel. Ces dispositions sont nombreuses¹⁰ et on peut citer les principales :

- *le secret statistique* : il est défini dans la loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques. Il interdit, pendant une durée de soixante-quinze ans et sauf dérogation, toute communication de données ayant trait à la vie personnelle et familiale, et plus généralement, aux faits et comportements d'ordre privé recueillis au moyen d'une enquête statistique.
- *le secret professionnel en matière fiscale* : il est défini à l'article L. 103 du livre des procédures fiscales qui soumet au secret professionnel en matière fiscale les informations recueillies par les services fiscaux à l'occasion des opérations d'établissement, de contrôle, de recouvrement ou de contentieux des impôts, droits,

⁸ En cas de recours pour excès de pouvoir contre le refus de l'administration de procéder à la communication d'un document administratif sur le fondement de ces dispositions, le juge administratif procède à un contrôle d'erreur de droit sur le point de savoir si la communication envisagée en l'espèce porte ou non atteinte à l'un des secrets protégés par la loi (cf. pour le secret des procédures juridictionnelles en cours : CE, 5 juin 2008, *SA Baudin Chateauneuf*, n° 309518; pour le secret de la vie privée : CE, 26 janvier 2011, *Mme Saffray*, n° 310270, pour la recherche des infractions fiscales : CE, 4 mai 2011, *Ministre du budget, des comptes publics, de la fonction publique et de la réforme de l'État c/M. Casanovas*, n° 328914).

⁹ Ainsi, un document comportant un très grand nombre de mentions couvertes par un secret et dont l'occultation s'avérerait particulièrement difficile pour l'administration doit être regardé comme non communicable (c CE, 26 mai 2014, *Communauté d'agglomération de Bayonne-Anglet-Biarritz*, n° 342339).

¹⁰ On peut retrouver sur le site de la CADA une présentation synthétique des différents textes concernant des secrets professionnels qui réduisent l'accès à certains documents administratifs (<http://www.cada.fr/les-dispositions-legislatives-nationales.6145.html>).

taxes et redevances prévus au code général des impôts de Paris. À titre d'exemple, les informations contenues dans le logiciel CEIL (Observatoire des évaluations immobilières locales) de la DGI sont protégées par ces dispositions, et ne peuvent être communiquées qu'après anonymisation (avis n° 20054655 du 5 janvier 2006, 20070921 du 8 mars 2007, et 20081202 du 20 mars 2008). C'est seulement lorsque les conclusions de documents fiscaux sont opposées à la personne qui en demande communication que l'administration ne peut valablement, en vertu de l'article 3 de la loi du 17 juillet 1978, lui opposer le secret en matière fiscale.

- *le secret professionnel en matière de prélèvements sociaux* : l'article L. 723-42 du code rural impose ainsi le secret aux personnes appelées à intervenir dans l'établissement des cotisations sociales. Il ne fait toutefois pas obstacle à la communication des documents nominatifs aux personnes qu'ils concernent (CE, 23 novembre 1990, Caisse MSA du Maine-et-Loire c/ Jonchère et avis n° 20071871 du 24 mai 2007).

Les mentions couvertes par un secret protégé par la loi deviennent toutefois communicables au terme des délais et dans les conditions fixés par les articles L. 213-1 et L. 213-2 du code du patrimoine¹¹. Le régime d'accès aux archives prend alors le relais des dispositions du code des relations entre le public et l'administration pour permettre la communication de données à l'expiration d'un certain délai, plus ou moins long selon leur sensibilité, sauf si ces données ont été détruites par le service qui les détient ou si elles relèvent de l'exception prévue au II de l'article L. 213-2 du code du patrimoine¹². Le tableau suivant présente les délais applicables pour les secrets relevant du code du patrimoine.

Délais de communicabilité prévus par le code du patrimoine

Secrets protégés par la loi	Délais prévus
Cas général	Immédiatement communicable
Délibérations du Gouvernement, relations extérieures, monnaie et crédit, recherche d'infractions fiscales et douanières, secret en matière industrielle et commerciale, et de statistique (sauf données d'ordre privé)	25 ans
Actes des juridictions administratives et financières, du Médiateur de la République, accréditation des établissements de santé, contrat de prestation de service pour une ou des personnes déterminées	25 ans
Secret médical	25 ans à compter de la date du décès de l'intéressé ou, si la date de décès n'est pas connue,

¹¹ CE, 8 avril 1994, *Ministère des affaires étrangères c/ Mme Jobez*, n° 96246.

¹² La loi du 15 juillet 2008 a introduit, au II de l'article L. 213-2 du code du patrimoine, la notion d'archives perpétuellement non communicables. Ne pourront donc jamais être consultées par des tiers les archives publiques dont la communication est susceptible d'entraîner la diffusion d'informations permettant de concevoir, fabriquer, utiliser ou localiser des armes nucléaires, biologiques, chimiques ou toutes autres armes ayant des effets directs ou indirects de destruction d'un niveau analogue.

Secrets protégés par la loi	Délais prévus
	120 ans après la naissance
Secret de la défense nationale, intérêts fondamentaux de l'État en matière de politique extérieure, sûreté de l'État, sécurité publique	50 ans
Vie privée, jugement de valeur sur une personne, ou document qui révèle un comportement dans des conditions dont la divulgation pourrait nuire à l'intéressé	50 ans
Documents relatifs à la construction ou au fonctionnement des établissements pénitentiaires	50 ans à compter de la désaffectation
Enquêtes des services de police judiciaire, dossiers des juridictions (sauf dispositions particulières aux jugements), actes notariés, actes de naissance et de mariage de l'état civil (à compter de la clôture des registres), secret statistique (données d'ordre privé)	75 ans à compter de la date du document ou du document le plus récent inclus dans le dossier, ou si le délai est plus bref, 25 ans à compter de la date du décès de l'intéressé
Documents d'enquêtes judiciaires ou des juridictions se rapportant à une personne mineure, décisions de justice révélant l'intimité de la vie sexuelle	100 ans à compter de la date du document ou du document le plus récent inclus dans le dossier ou, si le délai est plus bref, 25 ans à compter de la date du décès de l'intéressé
Documents couverts ou ayant été couverts par le secret de défense nationale et dont la communication peut porter atteinte à la sécurité de personnes identifiables (agents du renseignement)	100 ans à compter de la date du document ou du document le plus récent inclus dans le dossier ou, si le délai est plus bref, 25 ans à compter de la date du décès de l'intéressé
Documents comportant des informations permettant la fabrication, l'utilisation ou la localisation d'armes de destruction massive	Ni consultables, ni communicables, quel que soit le délai envisagé

1.2.2. L'accès anticipé aux données protégées

Avant l'expiration de ces délais, l'article L. 213-3 du code du patrimoine ouvre néanmoins la possibilité d'obtenir la communication par anticipation des documents énumérés à l'article L. 213-2, « dans la mesure où l'intérêt qui s'attache à la consultation de ces documents ne conduit pas à porter une atteinte excessive aux intérêts que la loi a entendu protéger ». Cette procédure dérogatoire obéit à des règles distinctes en fonction de la nature des données dont il est demandé la communication de façon anticipée.

Il existe un régime de droit commun, qui vaut pour la plupart des données contenues dans des documents administratifs, et trois régimes dérogatoires, pour lesquels la loi prévoit trois procédures d'accès par anticipation distinctes. Ces régimes dérogatoires concernent respectivement les données statistiques, les données couvertes par le secret professionnel en matière fiscale ainsi que les données médico-administratives contenues dans le système national des données de santé (SNDS).

1.2.2.1. Le régime de droit commun pour accéder de manière anticipée aux données protégées

Dans le cadre du régime d'accès par anticipation, l'autorisation est accordée ou refusée, dans un délai de deux mois à compter de l'enregistrement de la demande, par l'administration des archives du ministère de la culture, après accord du service producteur des documents. Seule la consultation des documents est alors autorisée, non la délivrance de copies. Par exemple, lorsqu'un chercheur fait une demande d'accès à des fichiers d'étrangers non librement communicables, l'autorisation de consultation suppose un accord préalable de la préfecture qui a produit ces documents. Quel que soit le sens de l'avis de l'autorité productrice, le refus de l'administration des archives doit être motivé, fût-ce par appropriation de la motivation de l'avis de l'autorité dont émanent les documents, dès lors que cette dernière a également été portée à la connaissance du demandeur¹³.

En cas de refus de communication d'un document, soit directement par l'administration productrice des documents, soit par les Archives de France lors d'une demande de dérogation, le chercheur a la possibilité de saisir la CADA. Pour rendre son avis, la commission tient compte par exemple de l'ancienneté du document et de la proximité de l'échéance du délai de libre communicabilité, de la sensibilité des informations qu'il contient mais également des motivations et de la qualité du demandeur ou encore de l'intérêt administratif ou familial, du degré de notoriété du document, etc. Elle prend par exemple en compte l'existence de personnes encore en vie dont la vie privée est évoquée dans les documents litigieux. Elle s'attache aussi à examiner si des demandes identiques ont été satisfaites par le passé ou si les médias ont pu se faire l'écho, de manière plus ou moins précise, des informations en cause.

Après avoir saisi la CADA, le chercheur est en droit de demander le réexamen de sa demande de communication. En cas de refus réitéré de la part de l'administration (soit l'administration productrice des données, soit l'administration des archives), il peut saisir le tribunal administratif pour contester la légalité de ce refus¹⁴ – le jugement du tribunal n'est susceptible que d'un recours en cassation devant le Conseil d'État. Le juge administratif exercera alors un contrôle d'erreur de droit sur la nature des intérêts mis en avant par l'administration pour refuser l'accès par dérogation aux archives publiques. Concrètement, ces intérêts rejoignent ceux qui sont énumérés à l'article 6 de la loi du 17 juillet 1978. Le juge administratif s'en tient en revanche à un contrôle d'erreur manifeste d'appréciation sur le point de savoir si la communication porterait une atteinte excessive aux intérêts protégés par la loi¹⁵.

¹³ CE, 29 juin 2011, *Mme Rouzaud*, n° 335072.

¹⁴ Plus précisément, l'article L. 342-1 du code des relations entre le public et l'administration prévoit que le recours devant la CADA constitue un préalable obligatoire à tout recours contentieux. Le fait que l'administration ait consulté la CADA avant de refuser la communication, dans le cadre d'une demande de conseil, ne dispense pas le requérant d'effectuer cette saisine préalablement à son action contentieuse. Un recours contentieux introduit devant le juge administratif en l'absence de recours devant la CADA est irrecevable. Toutefois, les avis de la CADA sont dépourvus de caractère contraignant. Ils ne constituent pas des décisions administratives faisant grief et ne sont donc pas susceptibles de faire l'objet d'un recours pour excès de pouvoir, pas plus que les courriers adressés au demandeur par le secrétariat de la Commission l'invitant à fournir certaines pièces à l'appui de son recours. Le chercheur débouté peut attaquer la nouvelle décision de refus qui naît deux mois à compter du jour où la CADA a enregistré sa demande d'avis, et quel que soit le sens de ce dernier. Ce point de départ du délai permet d'éviter toute manœuvre dilatoire de la part de l'administration. La Commission pouvant être saisie un mois après la demande initiale auprès de l'administration, c'est donc, au plus tard, au terme d'un délai total de trois mois à compter de cette demande que l'administré pourra déférer au juge administratif le refus implicite opposé à sa demande de communication.

¹⁵ CE, 29 juin 2011, *Mme Rouzaud*, n° 335072, *op. cit.*

Cette procédure de droit commun est peu utilisée par les chercheurs qui recherchent un accès à de larges bases de données statistiques¹⁶, dont l'accès comporte un risque de divulgation accidentelle ou intentionnelle de secrets (cf. 1.2.2.2.1.).

1.2.2.2. Les régimes dérogatoires pour accéder de manière anticipée aux données protégées

1.2.2.2.1. S'agissant des données statistiques

Jusqu'en 2008, la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques interdisait toute communication de « *renseignements individuels (...) ayant trait à la vie personnelle et familiale et d'une manière générale, aux faits et comportement d'ordre privé* ». Les chercheurs ne pouvaient donc pas accéder à des informations confidentielles issues des enquêtes ou des documents administratifs relatifs aux ménages.

En juillet 2008, la loi sur les Archives a modifié la loi de 1951 : elle a autorisé l'accès aux données confidentielles relatives aux individus et aux ménages sous réserve d'une « *décision de l'administration des archives, prise après avis du comité du secret statistique et relative à une demande effectuée à des fins de statistique publique ou de recherche scientifique ou historique* » (article 6 de la loi du 7 juin 1951). Cette procédure ne déroge pas complètement au régime de droit commun ; elle le complète par le passage pour avis devant le comité du secret statistique. Comme pour le régime de droit commun, la CADA est ensuite compétente, aux termes de l'article L. 342-1 du CRPA, pour connaître des éventuelles décisions de refus émises par l'administration des archives.

Pour ces données statistiques, comme pour les données fiscales (cf. *infra*), le comité du secret statistique doit émettre un avis avant l'intervention de la décision de l'autorité compétente. L'accès aux données repose, en amont de l'autorisation accordée, sur une pré-instruction avec le producteur des données et, en aval, sur le plan technique, par un accès sécurisé aux données.

L'ensemble de cette procédure repose sur la logique suivante : les chercheurs qui demandent l'accès à des bases de données personnelles ne sont pas intéressés par la révélation d'aspects de la vie privée de chacune des personnes concernées ; c'est l'exploitation de l'ensemble de l'information contenue dans la base qui les intéresse. Toutefois, l'accès à la base doit être suffisamment sécurisé pour éviter toute diffusion de données couvertes par un secret protégé par la loi, soit accidentelle soit intentionnelle. Ces précautions sont d'autant plus nécessaires que les informations contenues dans les enquêtes statistiques des services publics ou les bases de données administratives sont souvent sensibles.

Les étapes de la procédure sont les suivantes¹⁷ :

- **pré-instruction** : le porteur du projet contacte le service producteur des données. Il lui transmet le formulaire décrivant son projet, y compris la liste des membres du projet et des données confidentielles nécessaires pour le mener à bien. Les

¹⁶ La procédure de droit commun est utilisée *a contrario* par de nombreux chercheurs exploitant des données administratives sous format papier.

¹⁷ Ces étapes sont décrites sur le site du Cnis, dans la partie consacrée au comité du secret statistique à partir de laquelle les formulaires de demande sont téléchargeables (http://www.cnis.fr/cms/Accueil/activites/ trois_comites/Comite_du_secret_statistique/Les_procedures_dacces_au_x_donnees_confidentielles).

producteurs des données concernées étudient l'adéquation entre l'objet de la recherche et les bases de données demandées, et échangent si nécessaire avec le porteur du projet. Le cas échéant, ils le réorientent vers des bases de données non confidentielles si celles-ci suffisent pour la recherche envisagée. Si le contenu des bases de données disponibles correspond aux objectifs du projet, le porteur du projet transmet le formulaire finalisé au secrétariat du comité. Si le dossier est conforme, le comité du secret statistique est saisi ;

- **saisine du comité du secret statistique** : le comité du secret statistique émet un avis (favorable ou défavorable) pour chaque projet de recherche soumis¹⁸. Après que le comité du secret statistique a émis son avis ou ses recommandations, et sous réserve de l'accord de l'autorité dont émanent les documents, l'autorisation de communication est accordée par l'administration des archives¹⁹. En pratique, il est très rare que le Comité, le producteur ou les archives se prononcent contre la mise à disposition des données. En cas de décision favorable de l'administration des archives, dans le cas de données statistiques, ou du ministre chargé du budget, pour les données fiscales, l'habilitation obtenue rend personnellement responsable l'utilisateur des usages faits avec les données pour lesquelles l'accès a été autorisé. Cette responsabilité est pénale : les chercheurs sont astreints dans ce cadre au secret professionnel, sous les sanctions prévues aux articles 226-13 et 226-14 du code pénal ;
- **formalités auprès de la Cnil** : lorsque les chercheurs souhaitent accéder à des données à caractère personnel, le porteur du projet doit également réaliser les formalités préalables auprès de la Cnil pour les travaux envisagés. Un récépissé de déclaration (ou l'autorisation) doit alors être fourni avant que l'accès aux données soit effectif (cf. 1.3.) ;
- **accès sécurisé aux données** : l'accès aux données peut être assuré par le service producteur ou via un accès sécurisé distant, sous la forme de celui proposé par le Centre d'accès sécurisé aux données (CASD) du Groupe des écoles nationales de statistique et d'économie (GENES), qui permet aux chercheurs de traiter des données confidentielles dans des conditions de sécurité et de protection de la confidentialité élevées²⁰. Les coûts d'accès via le CASD sont facturés aux chercheurs.

¹⁸ Aux termes de l'article 17 du décret n° 2009-318 du 20 mars 2009 relatif au Conseil national de l'information statistique, au comité du secret statistique et au comité du label de la statistique publique, « *le comité du secret statistique émet son avis en prenant en compte la nature et l'intérêt des travaux pour l'exécution desquels la demande est formulée ainsi que la qualité de la personne ou de l'organisme présentant la demande et les garanties qu'il présente. Il vérifie que le volume des informations demandées n'est pas excessif par rapport aux travaux qui justifient leur communication et que celle-ci ne conduit pas à porter une atteinte excessive aux intérêts que la loi du 7 juin 1951 susvisée a entendu protéger. Dans le cas de demandes portant sur des faits et comportements d'ordre privé, il vérifie que cette demande est effectuée à des fins de statistique publique ou de recherche scientifique ou historique. Dans tous les cas, le comité du secret statistique détermine les conditions dans lesquelles ces renseignements individuels pourront être portés à la connaissance du demandeur.* »

¹⁹ Cf. article 17 du Décret n° 2009-318 du 20 mars 2009 modifié relatif au Conseil national de l'information statistique, au comité du secret statistique et au comité du label de la statistique publique.

²⁰ Depuis les années 2000, des pays comme le Danemark, les Pays-Bas et le Royaume Uni ont mis en place des solutions d'accès sécurisé à distance pour les chercheurs. Ces solutions reposent sur l'installation sur l'ordinateur du chercheur de logiciels particuliers d'accès distant (« *remote access* » en anglais). L'idée sous-jacente est de transformer les centres physiques en centres d'accès virtuels grâce aux technologies du numérique. Ces solutions, qui s'appuient sur un poste de travail entièrement maîtrisé par l'utilisateur, sont cependant apparues assez peu sécurisées et complexes à mettre en œuvre, notamment à cause des problèmes de compatibilité, de conflits

1.2.2.2. S'agissant des données fiscales

L'accès aux informations fiscales relève également d'un régime dérogatoire et autonome, qui est déterminé par le III de l'article L. 135 D du livre des procédures fiscales (issu de l'article 104 de la loi n° 2013-660 du 22 juillet 2013). L'accès aux informations fiscales à des fins de recherche avant l'expiration des délais de libre communicabilité est ainsi décidé par le ministre chargé du budget, après avis favorable du comité du secret statistique.

Une éventuelle décision de refus du ministre chargé du budget peut-elle faire l'objet d'un recours préalable devant la CADA, comme c'est le cas pour la décision des Archives concernant les données statistiques ? La réponse à cette question semble être négative. En effet, la commission dispose d'une compétence générale en matière de refus de communication, sur le fondement de l'article L. 342-1 du CRPA, que viennent compléter les dispositions de l'article L. 342-2. Ces dernières recensent l'ensemble des régimes spécifiques dont la CADA a à connaître. Si les dispositions du 11° du A de cet article visent le a et le b de l'article L. 104 et les articles L. 106, L. 111 et L. 135 B du livre des procédures fiscales, elles ne renvoient pas à l'article L. 135 D du livre des procédures fiscales (LPF). En l'absence de ce renvoi et dès lors que le CRPA vise les autres procédures d'accès que contient le livre des procédures fiscales, il est difficile de soutenir que la CADA est compétente sur le fondement des seules dispositions générales de l'article L. 342-1 de ce même code.

Pourtant, le recours préalable devant la CADA est utile : dans la grande majorité des cas, il permet d'éviter, très précisément, le recours au juge en rappelant à l'administration ses obligations en matière de droit d'accès ou en précisant les contours de son refus de communication. L'on peut donc regretter que le législateur ne soit pas venu modifier les dispositions du 11° du A de l'article L. 342-2 du CRPA pour y viser également l'article L. 135 D du LPF (cf. recommandation n°2 *infra*).

L'accès à ces données fiscales s'effectue nécessairement, après avoir accompli le cas échéant, selon la nature des données, les formalités nécessaires auprès de la Cnil, au moyen du CASD (article R. 135 D-1 du LPF).

1.2.2.3. S'agissant des données du système national des données de santé

La loi n° 2016-41 du 26 janvier 2016 relative à la modernisation de notre système de santé est venue modifier le code de la santé publique afin de créer le système national des données de santé, qui rassemble et met à disposition des données médico-administratives à caractère personnel relatives à la santé (article L. 1461-1 du code²¹). Ces données sont accessibles dans deux cas : en vue de l'accomplissement des missions des services de l'État, des établissements publics ou des organismes chargés d'une mission de service public, d'une

d'installation et de maintenance. Ces contraintes pratiques, qui se révèlent dirimantes lorsque l'on ne dispose que de moyens restreints, ont poussé le CASD à concevoir un équipement spécifique permettant de répondre au besoin des chercheurs en évitant les inconvénients précédents : la SD-Box.

²¹ Cet article dispose plus précisément que :

« Le système national des données de santé rassemble et met à disposition :

1° Les données issues des systèmes d'information mentionnés à l'article L. 6113-7 du présent code ;

2° Les données du système national d'information interrégimes de l'assurance maladie mentionné à l'article L. 161-28-1 du code de la sécurité sociale ;

3° Les données sur les causes de décès mentionnées à l'article L. 2223-42 du code général des collectivités territoriales ;

4° Les données médico-sociales du système d'information mentionné à l'article L. 247-2 du code de l'action sociale et des familles ;

5° Un échantillon représentatif des données de remboursement par bénéficiaire transmises par des organismes d'assurance maladie complémentaire et défini en concertation avec leurs représentants. (...) ».

part, et à des fins de recherche, d'étude ou d'évaluation répondant à un motif d'intérêt public, d'autre part (article L. 1461-3).

La procédure d'accès à ces données est imbriquée dans la procédure d'autorisation de traitement délivrée par la Commission nationale informatique et libertés (Cnil). L'accès est mis en œuvre par la CNAMTS (qui est responsable du traitement SNDS) lorsqu'elle est destinataire de l'autorisation délivrée par la Cnil au chercheur qui souhaite traiter ces données. Les voies de recours contre ces décisions sont celles du droit commun du régime d'autorisations délivrées par la Cnil : le chercheur peut les contester directement devant le Conseil d'État, qui statue en premier et dernier ressort.

Pour délivrer une autorisation de traitement, aux termes de l'article 54 de la loi informatique et libertés, la CNIL s'appuie sur le comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé, qui donne un « *avis sur la méthodologie retenue, sur la nécessité du recours à des données à caractère personnel, sur la pertinence de celles-ci par rapport à la finalité du traitement et, s'il y a lieu, sur la qualité scientifique du projet* » et éventuellement de l'Institut national des données de santé²², qui se prononce sur le caractère d'intérêt public.

1.2.2.2.4. S'agissant des données détenues par la Banque de France

Les données détenues par la Banque de France doivent faire l'objet d'un examen à part entière²³ : elles sont en effet protégées par le droit national et le droit européen. L'ensemble de ces données est couvert par le secret professionnel applicable aux agents de la Banque mais il convient de distinguer les catégories de données suivantes :

- en ce qui concerne les données statistiques : toutes ces données sont également couvertes par le secret en matière statistique mais certaines données, appelées données confidentielles (ISC) font l'objet d'une protection supplémentaire. Elles sont définies par les textes européens et nationaux comme des données permettant l'identification, directe ou indirecte, d'unités statistiques, ce qui a pour effet de divulguer des informations individuelles. Pour déterminer si une unité statistique est identifiable, il est tenu compte de tous les moyens appropriés qui pourraient raisonnablement être utilisés par un tiers pour identifier l'unité statistique ;
- le fonctionnement du système européen des banques centrales (SEBC) implique également la gestion de données qui ne sont pas de nature statistique. Les règles d'accès à ces informations non statistiques établies sont fixées dans une décision de la BCE du 4 mars 2004 relative à l'accès du public aux documents de la BCE (BCE/2004/3), dont il ne sera pas question ici.

Les règles d'accès aux ISC sont fixées par des textes européens, lorsque ces données sont détenues par le SEBC²⁴, et par le droit français lorsqu'elles sont détenues par la Banque de France.

²² Le groupement d'intérêt public dénommé Institut national des données de santé (INDS) est constitué entre l'État et les organismes assurant une représentation des usagers du système de santé, des producteurs et des utilisateurs des données, a été prévu par la loi. Il est notamment chargé de veiller aux conditions générales de mise à disposition des données (cf. art. L. 1462-1 du code de la santé publique).

²³ Voir en particulier le rapport du Cnis, Capelle-Blancard, Bellando, R. et G., Lacroix, R (2015), « L'accès aux données bancaires et financières : une mission de service public ».

²⁴ Aux termes de l'article 1^{er} du règlement 2533/98, les ISC sont définies comme « des informations statistiques que les agents déclarants sont tenus de fournir qui sont nécessaires à l'accomplissement des missions du SEBC et

En ce qui concerne les ISC du SEBC, leurs modalités de collecte, de gestion et d'accès sont régies par le règlement (CE) n°2533/98 modifié du Conseil du 23 novembre 1998. Ces informations statistiques prennent la forme de données collectées auprès d'institutions financières ou d'entreprises dans le cadre notamment de l'élaboration des statistiques monétaires, de la balance des paiements et de la position extérieure de la France.

- La transmission au sein du SEBC de ces données a lieu « *dans la mesure et au niveau de détail nécessaire à l'accomplissement des missions du SEBC visées dans le traité ou des missions relevant du domaine de la surveillance prudentielle confiées aux membres du SEBC, pour autant que cette transmission soit nécessaire au développement, à la production ou à la diffusion efficaces des statistiques (...) ou pour accroître leur qualité* ».
- Toutefois, certaines données peuvent être transmises, à titre dérogatoire, à des organes de recherche scientifiques (art. 8,§1,c du règlement (CE) n°2533/98 modifié). Les chercheurs doivent donc soumettre une demande d'accès aux données par l'intermédiaire de l'organe de recherche auquel ils se rattachent. L'article 8.1.c) précité prévoit non seulement le consentement explicite préalable de l'autorité qui, le cas échéant, a initialement fourni les informations à la BdF mais également que les informations ainsi transmises « ne permettent pas une identification directe ».
- La Banque de France offre l'accès effectif aux données confidentielles du SEBC qu'elle détient dans le cadre d'une salle sécurisée (*Open Data Room*), après dé-identification et, le cas échéant, déletion d'autres informations. Un accord préalable du « Comité d'examen des demandes d'accès aux données de la Banque de France » est requis, ce comité, composé de représentants de la Banque et de personnalités du monde académique, s'assurant seulement que le motif de la demande est bien un projet de recherche scientifique à des fins de publication.

Les ISC provenant de collectes hors SEBC²⁵ ou relèvent d'autres informations que la Banque de France collecte dans le cadre de ses missions autres que statistiques et hors SEBC qui sont donc régies par des textes s'appliquant à ces missions et déterminant les modalités d'échange d'informations confidentielles (statistiques économiques et financières qui inclut les enquêtes de conjoncture, les statistiques d'entreprise ou des systèmes et moyens de paiements mais également des données personnelles sur le surendettement, etc.). Ce type de données a pris une importance grandissante au fur et à mesure des missions confiées par le législateur à la Banque de France. On pourra citer, à titre d'exemples récents, les données collectées au titre

qui permettent d'identifier les agents déclarants ou toute autre personne morale ou physique, entité ou succursale, que ce soit directement par leur nom, leur adresse ou par un code d'identification officiel qui leur a été attribué, ou indirectement par déduction, ce qui a pour effet de divulguer des informations d'ordre individuel. Afin de déterminer si un agent déclarant ou toute autre personne morale ou physique, entité ou succursale est identifiable, il est tenu compte de tous les moyens qui pourraient raisonnablement être utilisés par un tiers pour identifier ledit agent déclarant ou la personne morale ou physique, l'entité ou la succursale. »

²⁵ Trois de ces collectes sont, pour des raisons historiques, labellisées par le CNIS dans le cadre du programme statistique annuel défini par arrêté du ministre.

des observatoires des délais de paiement²⁶, de l'épargne réglementée²⁷, de l'inclusion bancaire²⁸, du suivi du financement des entreprises²⁹.

- Sauf informations publiques (par exemple les informations reçues du registre du commerce et des sociétés), les ISC ne sont pas communicables, sur le fondement de l'article L. 311-5 du CRPA qui permet de refuser l'accès à des documents administratifs dont la communication porterait atteinte à la monnaie et au crédit public).
- Toutefois, l'accès à ces données reste possible à titre dérogatoire selon différentes modalités :
 - o pour les données non statistiques, le Service Interministériel des Archives de France peut, après recueil de l'avis de la Banque de France, autoriser la consultation de ces données par anticipation sur le fondement de l'article L. 213-3 I. du code du patrimoine (régime de droit commun) ;
 - o pour les informations confidentielles qui sont couvertes par la loi de 1951, le Service interministériel des Archives de France est compétent, après avis du comité du secret statistique, sur le fondement de la loi de 1951 et de l'article L. 213-3 du code du patrimoine, pour en autoriser l'accès à titre dérogatoire (dispositif « données statistiques » décrit au 1.2.2.2.1). L'accès des chercheurs à ces données, lorsqu'il a été autorisé, est également assuré dans le cadre de la salle sécurisée de la Banque de France ;
 - o pour les informations confidentielles qui ne sont pas couvertes par la loi de 1951, l'accès aux données individuelles anonymisées pour les chercheurs est assuré dans le cadre de l'« Open Data Room » de la Banque de France après accord préalable du « Comité d'examen des demandes d'accès aux données de la Banque de France » comme indiqué plus haut.

1.3. Le traitement des données à caractère personnel est soumis à la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés

Au-delà de l'accès aux données, tout traitement de données à caractère personnel qui sera réalisé par un chercheur avec les informations contenues dans les documents dont il obtient communication est soumis à la loi de 1978 relative à l'informatique, aux fichiers et aux libertés (1.3.1). Le traitement doit donc faire l'objet d'une démarche auprès de la Cnil, sous forme de déclaration ou, le cas échéant, de demande d'autorisation (1.3.2). Certaines données font par ailleurs, en raison de leur nature, l'objet d'un régime spécifique qui entend limiter les possibilités de traitement (1.3.3).

²⁶ Afin de vérifier la bonne application des règles en matière de délais de paiement instituées par la loi de modernisation de l'économie de 2008.

²⁷ L'Observatoire de l'épargne réglementée, créé par la loi de modernisation de l'économie, est « chargé de suivre la mise en œuvre de la généralisation de la distribution du livret A, notamment son impact sur l'épargne des ménages, sur le financement du logement social et sur le développement de l'accessibilité bancaire » (Art. L. 221-9 du CMF).

²⁸ Aux termes de l'article L. 312-1-1 B du code monétaire et financier, créé par la loi n°2013-672 du 26 juillet 2013, « il est créé, auprès de la Banque de France, un observatoire de l'inclusion bancaire chargé de collecter des informations sur l'accès aux services bancaires des personnes physiques n'agissant pas pour des besoins professionnels, sur l'usage que ces personnes font de ces services bancaires et sur les initiatives des établissements de crédit en la matière.

²⁹ Article L. 141-6 du code monétaire et financier.

1.3.1. *Le champ d'application de la loi Cnil est particulièrement vaste*

Outre que la loi s'applique aussi bien aux traitements du secteur public qu'à ceux du secteur privé, son article 2 retient une définition large de la notion de «traitement de données à caractère personnel», puisqu'il s'agit aussi bien de traitements automatisés (mécaniques, puis informatiques et enfin numériques) que de traitements non automatisés (fichiers papiers par exemple).

De même, alors que la loi de 1978, dans sa version antérieure à la loi du 6 août 2004, mentionnait les «informations nominatives», c'est désormais la notion de «donnée à caractère personnel », plus large, qui est employée. Elle désigne en effet tout élément permettant directement ou indirectement d'identifier une personne (identité, coordonnées, numéros identifiants, mais aussi des données telles que la voix, l'image ou toute «empreinte biométrique»).

Quant aux traitements, ils doivent être conformes à la loi «quel que soit le procédé technique utilisé», ce qui rend ces dispositions législatives applicables à toute nouvelle technologie numérique.

Ainsi, toute opération, quelle qu'elle soit, que le chercheur souhaitera effectuer à partir des données publiques auxquelles il a pu accéder, sera, si de telles données sont directement ou indirectement identifiantes, soumise aux principes et règles fixés par la loi du 6 janvier 1978.

1.3.2. *Règles et formalités préalables à tout traitement de données à caractère personnel*

Le traitement par les chercheurs de données à caractère personnel doit s'effectuer dans les conditions et en accomplissant les formalités auprès de la Cnil prévues par la loi informatique et libertés.

1.3.2.1. Les conditions de licéité des traitements de données à caractère personnel

La loi du 6 janvier 1978 impose tout d'abord aux responsables de traitement de veiller à ce que les opérations qu'ils envisagent de mettre en œuvre, et ce quelle qu'en soit leur forme ou leur finalité, respectent une liste de conditions cumulatives expressément énumérées par les dispositions de cette loi.

Ainsi :

- une opération de traitement ne peut porter que sur des données personnelles « *collectées et traitées de manière loyale et licite* » (1° de l'article 6) ;
- cette collecte doit être effectuée « *pour des finalités déterminées, explicites et légitimes* » sans que les données en cause ne puissent être ultérieurement utilisées de manière incompatible avec ces finalités, sauf à ce que le traitement ultérieur s'effectue à des fins statistiques, scientifiques ou historiques (2° de l'article 6) ;
- les données doivent être « *adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont collectées et de leurs traitements ultérieurs* » (3° de l'article 6) ;
- dans cette même logique, les données personnelles doivent être « *exactes, complètes et, si nécessaire, mises à jour* » par des mesures appropriées (4° de l'article 6) ;

- la loi impose enfin que la conservation de ces données n'excède pas la durée nécessaire à l'accomplissement des finalités pour lesquelles elles ont été collectées et traitées (5° de l'article 6).

1.3.2.2. Les formalités préalables au traitement de données à caractère personnel

Outre ces conditions, la loi informatique et libertés a instauré un régime de formalités préalables. Il peut s'agir, selon les cas, d'un régime déclaratif ou d'un régime d'autorisation par la Cnil ou par un acte réglementaire pris après avis de la Cnil. Dans tous les cas, il s'agit bien de formalités «préalables» à la mise en œuvre du traitement, lequel ne peut légalement être mis en œuvre tant que le responsable n'a pas, selon les cas, reçu un récépissé ou une autorisation de la Cnil³⁰.

Le régime déclaratif couvre aujourd'hui la grande majorité des traitements de données à caractère personnel.

Alors que les principes de la loi s'appliquent à l'ensemble des traitements, le régime déclaratif ne s'impose qu'aux traitements automatisés de données à caractère personnel, ceux non automatisés étant expressément exclus du champ de ces dispositions par l'article 22.

La déclaration peut intervenir par voie postale ou en ligne (ce qui représente la majorité des saisies aujourd'hui), et fait l'objet de la délivrance d'un récépissé «sans délai» (actuellement en moins de quatre jours). La délivrance de ce récépissé, comme pour tout régime ordinaire de déclaration, ne vaut pas reconnaissance de la conformité effective du traitement à la loi, mais implique simplement que l'obligation de «porter à connaissance» de manière complète a été respectée, même si, dans la pratique, la Cnil s'assure généralement que le régime de la déclaration est bien le régime applicable au traitement en question.

La déclaration doit obligatoirement comporter les informations qui conditionnent l'appréciation de la licéité du traitement, mentionnées à l'article 30, notamment l'identité du responsable de traitement, la finalité du traitement, l'identification des données collectées, les destinataires et la durée de conservation.

Le régime d'autorisation des traitements de données à caractère personnel est triple, en fonction de la nature des données traitées et, le cas échéant, de l'identité du responsable de traitement. Les dispositions qui les organisent (articles 25 à 27) doivent, pour une bonne compréhension, être rapprochées de celles des articles 8 et 9, qui énumèrent les données sensibles par nature (cf. *infra*).

Un premier ensemble de traitements est soumis à l'autorisation préalable et exclusive de la Cnil (article 25). Il s'agit majoritairement de traitements particulièrement sensibles en raisons des données qu'ils traitent : données génétiques, infractions, appréciations sur les difficultés sociales des personnes, numéro d'inscription des personnes au Répertoire national d'identification des personnes physiques ou consultation de ce répertoire sans inclure ce numéro d'inscription, données biométriques nécessaires au contrôle de l'identité des personnes. Sont également concernés les traitements qui portent sur des données sensibles au sens de l'article 8, lorsque le traitement est justifié par un intérêt public, que les données font l'objet d'une anonymisation à bref délai ou qu'il s'agit d'un traitement statistique mis en œuvre par l'INSEE.

Les autres traitements concernés le sont à raison de leur objet ou de leur finalité: il s'agit soit de ceux qui sont susceptibles, du fait de leur nature, de leur portée ou de leur finalités d'exclure des personnes du bénéfice d'un droit, d'une prestation ou d'un contrat en l'absence de toute

³⁰ CE 19 juillet, 2010, *M. Fristot et autres*, n°s 317182 et 323441.

disposition législative ou réglementaire; soit des traitements ayant pour objet des interconnexions entre fichiers qui répondent à des intérêts publics ou à des finalités différents.

Au-delà de ce «socle» du régime d'autorisation, certains traitements ne peuvent être autorisés que par un décret en conseil d'État pris après avis motivé et publié de la Cnil, compte tenu, d'une part, de l'identité du responsable de traitement et, d'autre part, des finalités et données traitées (articles 26 et 27). Il en va ainsi par exemple des traitements de données qui portent sur le NIR, même si, nous le verrons, la loi pour une République numérique a substantiellement allégé les formalités préalables en la matière afin de faciliter le traitement de ces données à des fins de recherche (cf. *infra*).

La déclaration ou l'autorisation traduisent l'engagement du responsable de traitement à respecter l'ensemble des prescriptions légales et des conditions dont il fait état. Il engage ainsi, non seulement sa responsabilité au sens de la loi informatique et libertés, mais également, le cas échéant, sa responsabilité pénale. En outre, les informations transmises dans ce cadre constituent un outil de transparence important pour le citoyen, et une information précieuse pour la Cnil, notamment en cas de plainte ou de contrôle.

1.3.2.3. Les opérations d'anonymisation

Pour satisfaire aux exigences posées par l'article L. 322-2 du CRPA, qui subordonne la réutilisation³¹ d'informations publiques comportant des données à caractère personnel au respect des dispositions de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, un travail de traitement des données par l'administration peut s'avérer nécessaire afin de mettre à disposition des données qui ne sont plus soumises au secret.

L'article R. 322-3 de ce code prévoit à cet égard que « *lorsque la réutilisation n'est possible qu'après anonymisation des données à caractère personnel, l'autorité détentrice y procède sous réserve que cette opération n'entraîne pas des efforts disproportionnés* ».

L'article L. 311-7 du CRPA, consacrant une jurisprudence constante du Conseil d'État³², prévoit, nous l'avons vu, que « *lorsque la demande porte sur un document comportant des mentions qui ne sont pas communicables (...) mais qu'il est possible d'occulter ou de disjoindre, le document est communiqué au demandeur après occultation ou disjonction de ces mentions* ». Le principe est donc, s'agissant des données à caractère personnel, celui de l'obligation pour l'administration de rendre ces données anonymes.

Si, en l'état du droit, les réutilisateurs de données ne peuvent procéder eux-mêmes à cette anonymisation et sont donc tributaires des capacités techniques de l'administration, il est néanmoins toujours possible pour l'administration de faire supporter par le réutilisateur le coût de ces opérations d'anonymisation, qu'elle y ait procédé elle-même ou qu'elle ait recouru, au cas où elle ne disposerait pas des moyens techniques nécessaires à la réalisation de ces opérations, à un prestataire de service extérieur. En effet, bien que le principe demeure celui de la gratuité de la réutilisation des informations publiques, l'article L. 324-1 du CRPA prévoit la possibilité pour les administrations qui sont tenues de couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service public³³ d'exiger le

³¹ L'article L. 321-1 du CRPA définit la notion de réutilisation de manière très large comme l'utilisation « *à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus* ». Dans ces conditions, et compte tenu des progrès techniques, la distinction est de plus en plus compliquée entre l'accès et la réutilisation des données.

³² CE, 16 juin 1989, *Office public d'HLM de la ville de Paris*, n° 83476.

³³ Depuis la parution du décret du 29 novembre 2016, seuls l'IGN, Météo-France et le SHOM peuvent percevoir des redevances (hors secteur des bibliothèques, musées et archives).

versement d'une redevance de réutilisation, notamment pour couvrir le coût des traitements permettant d'anonymiser des informations, à condition que « *le produit total du montant de cette redevance, évalué sur une période comptable appropriée, ne dépasse pas le montant total des coûts liés à la collecte, à la production, à la mise à la disposition du public ou à la diffusion de leurs informations publiques* » (article L. 324-1 du CRPA).

1.3.3. Le traitement de certaines données est soumis à un régime restrictif

La loi Cnil prévoit que certaines informations doivent, en raison de leur nature, faire l'objet d'une protection renforcée : il s'agit, d'une part, des données qualifiées de sensibles, d'autre part, des données relatives aux infractions, aux condamnations et aux mesures de sûreté.

1.3.3.1. Le traitement des données dites « sensibles »

Le traitement des données qualifiées par la loi de « sensibles » - car concernant directement l'intimité des individus - est en principe prohibé. Ainsi, l'article 8 de la loi du 6 janvier 1978 prévoit qu'« *il est interdit de collecter ou de traiter des données faisant apparaître directement ou indirectement, les origines raciales ou ethniques, les opinions philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé, ou à la vie sexuelle de celles-ci* ».

Malgré le caractère éminemment intime des "données sensibles", le législateur a tout de même souhaité pouvoir déroger à ce principe pour des raisons pratiques, « *dans la mesure où la finalité du traitement l'exige* » (II de l'article 8). Outre l'hypothèse des traitements aux fins de journalisme ou d'expression littéraire et artistique, la loi du 6 janvier 1978 prévoit à cet effet pas moins de dix cas dans lesquels il est possible de déroger à l'interdiction de traiter des données sensibles.

On relèvera, parmi les dérogations susceptibles d'intéresser les chercheurs :

- s'agissant des données de santé, l'autorisation des traitements nécessaires à la recherche et aux études ainsi qu'aux évaluations dans le domaine de la santé (art. 8, II, 8°) ;
- l'autorisation des traitements statistiques réalisés par l'INSEE ou l'un des services statistiques ministériels ; ces traitements doivent faire l'objet d'une autorisation de la Cnil, après avis du CNIS (art. 8, II, 7°) ;
- l'autorisation des traitements qui portent sur des données sensibles « *appelées à faire l'objet à bref délai d'un procédé d'anonymisation* », le procédé d'anonymisation devant, au préalable, avoir été reconnu conforme par la Cnil (art. 8, III) ;
- l'autorisation des traitements justifiés par l'intérêt public. Un tel intérêt public doit néanmoins être dûment constaté aux termes d'une autorisation délivrée par la Cnil (art. 8, IV).

1.3.3.2. Le traitement des données relatives aux infractions, condamnations et mesures de sûreté

Bien que ne portant pas à proprement parler sur des données "sensibles", les traitements relatifs aux infractions, aux condamnations et aux mesures de sûreté sont eux aussi entourés de garanties spécifiques³⁴.

Ainsi, l'article 9 de la loi du 6 janvier 1978 prévoit que : « *Les traitements de données à caractère personnel relatives aux infractions, condamnations et mesures de sûreté ne peuvent être mis en œuvre que par : / 1° Les juridictions, les autorités publiques et les personnes morales gérant un service public, agissant dans le cadre de leurs attributions légales ; / 2° Les auxiliaires de justice, pour les stricts besoins de l'exercice des missions qui leur sont confiées par la loi ; (...) 4° Les personnes morales mentionnées aux articles L. 321-1 et L. 331-1 du code de la propriété intellectuelle, agissant au titre des droits dont elles assurent la gestion ou pour le compte des victimes d'atteintes aux droits prévus aux livres Ier, II et III du même code aux fins d'assurer la défense de ces droits* ».

Le traitement de telles données emporte ainsi une double conséquence :

- en termes de formalités préalables, de tels traitements sont soumis au régime de l'autorisation de la Cnil, conformément à l'article 25-3°, sous réserve de l'application des régimes prévus par les articles 26 et 27 ;
- en termes de responsabilité de la mise en œuvre du traitement, seuls les responsables limitativement énumérés peuvent être autorisés à traiter ces données, les chercheurs n'entrant dans aucune de ces catégories.

Quant au champ d'application matériel, si les notions d'infraction ou de condamnation ne soulèvent aucune difficulté, celle de « mesure de sûreté » est un peu plus délicate. Les délibérations successives de la Cnil font apparaître que sont regardées comme telles toutes les mesures qui, prises en vertu d'un impératif d'ordre public, ont pour effet de restreindre la liberté, notamment d'aller et venir, de la personne.

Dans ces conditions, l'accès des chercheurs à ce type de données, s'il est possible et soumis aux conditions du droit commun, risque de présenter un intérêt somme toute limité, dans la mesure où aucun traitement de ces données ne peut être effectué par un chercheur.

Ainsi, en l'état du droit, la seule solution à disposition du chercheur pour travailler sur de telles données est sans doute de se mettre à disposition d'un service du ministère de la justice.

1.4. La loi pour une République numérique entend favoriser l'accès des chercheurs aux données publiques

Issue d'un projet de loi présenté en Conseil des ministres le 9 décembre 2015, après une phase de relecture publique ayant suscité plus de 8 500 contributions et près de 150 000 votes, la loi n° 2016-1321 pour une République numérique a été définitivement adoptée le 28 septembre 2016 et est entrée en vigueur le 9 octobre 2016.

Cette loi vise à favoriser l'ouverture et la circulation des données et du savoir, à garantir un environnement numérique ouvert et respectueux de la vie privée des internautes et à faciliter l'accès des citoyens au numérique.

³⁴ Il convient ici de rappeler que seul le traitement de ces données est soumis à un régime spécifique ; leur accès, en revanche, s'effectue selon les modalités du droit commun définies par le CRPA.

Deux dispositifs, issus des articles 38 et 36 qui figurent au chapitre II (« Économie du savoir») du titre premier (« La circulation des données et du savoir ») de la loi, intéressent plus particulièrement l'accès des chercheurs aux données publiques.

1.4.1. Un appariement des données facilité

Certains projets de recherche scientifique nécessitent d'apparier des sources de données entre elles, c'est-à-dire de mettre en correspondance des données se rapportant aux mêmes phénomènes du monde réel que l'on souhaite analyser mais provenant de jeux de données différents. C'est le cas par exemple lorsqu'il s'agit d'étudier les liens entre les revenus salariaux et les revenus de remplacement ou encore les liens entre les trajectoires scolaires et les trajectoires professionnelles ultérieures des individus, etc.

L'un des identifiants privilégiés pour apparier les informations correspondant à un même individu est le Numéro d'Inscription au Répertoire national d'Identification (NIR), plus communément appelé numéro de sécurité sociale. Il s'agit d'un identifiant individuel unique de très bonne qualité que l'on retrouve dans de nombreuses bases de données administratives.

Or, l'article 27 la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés prévoit que le recueil du NIR, notamment pour effectuer un appariement, doit être autorisé par un décret en Conseil d'État, après avis motivé et publié de la Commission nationale de l'informatique et des libertés (Cnil).

Dans la pratique, cette exigence s'est souvent révélée insurmontable pour les organismes de recherche, rares étant ceux qui ont pu obtenir qu'un ministre prenne l'initiative de porter un décret en Conseil d'État pour permettre un appariement dans le cadre d'un projet de recherche. Cette difficulté a le plus souvent conduit les chercheurs à renoncer à effectuer certains appariements ou à les effectuer de manière dégradée, en choisissant d'apparier les fichiers à partir de variables moins sûres que le NIR.

C'est dans ce contexte que l'article 38 de la loi pour une République numérique est venu modifier les articles 22 et 25 de la loi du 6 janvier 1978 afin de créer une nouvelle procédure spécifique d'accès à certaines données publiques à des fins statistiques ou de recherche publique.

Ce nouvel article substitue ainsi, en cas de demande d'accès à des données comprenant le numéro de sécurité sociale (NIR), un nouveau régime de déclaration à la Cnil (pour les travaux de la statistique publique) ou d'autorisation de la Cnil (pour les projets de la recherche publique) à l'ancien régime d'autorisation par un décret en Conseil d'État.

Désormais, les appariements effectués par la recherche scientifique publique se feront à partir d'une clé d'appariement non signifiante (CSNS), obtenue par une opération cryptographique réalisée sur le NIR. La clé associée à l'opération cryptographique sera spécifique à chaque projet de recherche, les données concernées contenant ce code ne pouvant être utilisées en dehors du projet de recherche.

À condition que les délais d'obtention de l'autorisation de la Cnil ne s'avèrent pas prohibitifs, cette mesure devrait permettre de réaliser des appariements de fichiers jusqu'alors impossibles en pratique, d'optimiser les coûts de réalisation des appariements et, au final, un meilleur usage des données par la communauté des chercheurs permettant des analyses plus fines et objectives.

1.4.2. *L'ouverture de la saisine du comité du secret statistique*

Le second dispositif, issu d'un amendement présenté par le Gouvernement en première lecture devant le Sénat, est inscrit à l'article 36 de la loi. Il vise à compléter la procédure d'accès anticipé aux archives (article L. 213-3 du code du patrimoine, cf. *supra*) afin de prendre en compte le cas des grandes bases de données utilisées à des fins de recherche ou d'étude présentant un intérêt public.

Il est en effet apparu que certaines administrations, telles que la Caisse nationale des allocations familiales (CNAF) ou l'Agence centrale des organismes de sécurité sociale (ACOSS), rencontraient des difficultés pour développer l'accès des chercheurs à leurs données en raison de l'incertitude du cadre juridique dans lequel pouvait s'organiser un tel accès, les sources gérées par ces administrations ne relevant pas de la compétence du Comité du secret statistique.

En effet, d'après la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, le comité du secret statistique était compétent pour les enquêtes statistiques soumises au visa préalable du ministre chargé de l'économie et du ministre à la compétence duquel ressortissent les intéressés (article 2) ou pour des données obtenues par l'Insee ou les services statistiques ministériels via l'article 7bis. Les données administratives détenues par les administrations de Sécurité sociale, non obtenues par la voie de l'article 7 bis, par exemple ne font pas partie de cet ensemble de la Statistique publique, et n'entrent donc pas dans le champ du comité du secret statistique.

Afin d'étendre la compétence du comité du secret statistique, l'article 36 de la loi pour une République numérique modifie l'article L. 311-8 du code des relations entre le public et l'administration (CRPA), qui prévoit désormais que lorsqu'une demande d'accès anticipé « *porte sur une base de données et vise à effectuer des traitements à des fins de recherche ou d'étude présentant un caractère d'intérêt public, l'administration détenant la base de données ou l'administration des archives peut demander l'avis du comité du secret statistique institué par l'article 6 bis de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques* ». Le comité pourra recommander le recours à « *une procédure d'accès sécurisé aux données présentant les garanties appropriées* », son avis devant tenir compte i) « *des enjeux attachés aux secrets protégés par la loi, notamment la protection de la vie privée et la protection du secret industriel et commercial* » ainsi que ii) « *de la nature et de la finalité des travaux pour l'exécution desquels la demande d'accès est formulée* ».

Ce passage par le comité du secret statistique permettra de sécuriser les administrations afin de les inciter à autoriser l'accès aux bases de données qu'elles détiennent dans des conditions présentant toutes les garanties requises. Cet accès sécurisé pourra être assuré par le producteur ou être fourni via des services tels que le CASD (centre d'accès sécurisé aux données) géré par le Groupe des Écoles Nationales d'Économie et Statistique (GENES)³⁵ ou l'ODR (open data room) géré par la Banque de France.

³⁵ Signalons ici que des applications développées par les services publics d'archives pourraient également faciliter la mise à disposition de données archivées. Cf. programme interministériel VITAM (Valeurs Immatérielles Transmises aux Archives pour Mémoire) piloté par la direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC) en partenariat avec les Archives de France.

1.4.3. L'absence de sanctions pénales pour les administrations en cas de non-respect du secret professionnel par les chercheurs

Par ailleurs, certaines administrations (CNAF et ACOSS) estimaient que l'article 226-13 du code pénal, qui réprime les atteintes au secret professionnel, avait vocation à s'appliquer en la matière et craignaient, compte tenu de la nature des données dont elles disposaient, que leur transmission fût constitutive d'un délit.

Afin de sécuriser les administrations productrices de ces bases de données, l'article 36 de la loi pour une République numérique complète l'article L. 213-3 du code du patrimoine afin de confirmer, de manière explicite, que l'article 226-13 du code pénal relatif aux sanctions pénales en cas de non-respect du secret professionnel, n'est pas applicable aux procédures d'ouverture anticipée des archives publiques. Cette précision lève toute ambiguïté et sera de nature à inciter les administrations concernées à faciliter l'accès aux données qu'elles détiennent.

2. COMPARAISONS INTERNATIONALES

La lettre de mission mandatant les présidents du groupe de travail souligne l'importance des comparaisons internationales en matière d'accès aux données publiques. De fait, la question de l'ouverture des données administratives à la recherche est, dans de nombreux pays, au premier plan des débats relatifs à l'*Open Data* et à l'évaluation des politiques publiques.

2.1. Périmètre et méthodologie de la comparaison

Pour réaliser cet état des lieux des bonnes pratiques au niveau international, cinq pays ont été retenus : l'Allemagne, le Danemark, les Pays-Bas, le Royaume-Uni et les États-Unis. Ces pays sont souvent cités en exemple par les chercheurs pour l'ouverture des données à la recherche.

Deux des pays étudiés, le Danemark et les Pays-Bas, font partie des pays dit « à registres de population », dont le système statistique repose assez largement sur les données administratives.

Dans les autres pays étudiés, la mobilisation de ce type de données est une préoccupation plus récente, qui a adopté les contours d'un mouvement général de la statistique publique, coordonné, en ce qui concerne l'Union européenne, par Eurostat dans le cadre du Système statistique européen (SSE).

Cet échantillon a semblé suffisamment varié aux membres du groupe de travail pour situer le dispositif existant en France et tirer des enseignements utiles en matière de bonnes et de mauvaises pratiques.

Au-delà de l'étude de ces cinq pays, ce panorama prend notamment en compte³⁶ :

- les évolutions juridiques et les travaux qui les ont précédées en matière d'accès aux données au sein de l'Union européenne, en particulier concernant le règlement européen sur la protection des données personnelles, le règlement pour l'accès des chercheurs aux données Eurostat dans le cadre du Système Statistique Européen ou encore l'encadrement juridique des données du réseau des banques centrales nationales³⁷, coordonné par la Banque centrale européenne ;
- les études et recommandations de l'OCDE dans ce domaine³⁸ ;
- les travaux menés dans le cadre de projets européens, d'une part, dans le domaine de l'accès national et transnational aux données confidentielles de la statistique publique et

³⁶ Cet état des lieux s'est en particulier appuyé sur les travaux et les discussions qui ont été menés dans le cadre du projet européen *Data without Boundaries* (DwB) 2011-2016 dans le cadre du 7^{ème} programme cadre ainsi que ceux qui se sont tenus au sein de l'*Expert group, du Microdata Access* à l'OCDE en 2011-2012, de l'*OECD Global Science Forum expert group on Research Infrastructures for the Social Sciences*, en 2012, du *Working Group on Statistical Confidentiality (WGSC)* du Système statistique européen à Eurostat qui ont tous mobilisé nombre d'acteurs, tant du côté des détenteurs des données que des utilisateurs. Des compléments d'information ont par ailleurs été recherchés pour chacun des pays avec notamment des échanges avec Tanvi Desai de l'ADRN et du UKDS, Lars Vilhuber de Cornell University, Ruben Dood de CBS et Maurice Brandt de DESTATIS.

³⁷ Cf. G. Capelle-Blanchard, R. Bellando et R. Lacroix, *L'accès aux données bancaires et financières : une mission de service public*, rapport du groupe de travail du CNIS, 2015

³⁸ Cf. les travaux de l'*Expert Group for International Collaboration on Microdata Access*, 2014 et, dans le domaine de l'accès aux données santé, le rapport *Health Data Governance, Privacy, Monitoring and Research*, 2015

administrative (FP7 *Data without Boundaries*) et, d'autre part, dans le domaine des données de santé (FP7 *Euroreach*, 2010-2013).

Quelques précautions méthodologiques méritent enfin d'être rappelées.

En premier lieu, la notion de données publiques ne renvoie pas nécessairement au même périmètre dans les pays considérés.

Traditionnellement³⁹, les données administratives sont identifiées à partir des critères suivants : i) le service collecteur n'est pas le service statistique auquel ces données peuvent ensuite être fournies pour un usage statistique, les données collectées le sont donc initialement pour un objectif non statistique, ii) il existe une couverture exhaustive de la population visée ; iii) le contrôle des méthodes de collecte et des traitements demeure entre les mains de l'agence administrative.

Toutefois, une définition plus large inclut les sources privées⁴⁰, ce qui permet de prendre en compte le transfert, dans certains pays, de la création et de la gestion de certains registres administratifs à des structures privées ou semi-privées. C'est cette définition large qu'a retenu le groupe de travail, pour tenter d'obtenir des comparaisons pertinentes.

En deuxième lieu, il importe de tenir compte de l'organisation institutionnelle et administrative du pays considéré.

Dans un État centralisé, une grande part des bases administratives est gérée par l'administration centrale, ce qui n'est pas le cas pour des États fédéraux ou dans lesquels les entités territoriales ont une forte autonomie.

En troisième lieu, le poids de l'histoire explique certaines caractéristiques des pays étudiés.

Ainsi, l'existence de « registres de population municipaux » constitués vers le milieu du XIX^{ème} siècle (liés à l'obligation de s'enregistrer en cas d'installation ou de déménagement) dans une vingtaine de pays de l'Union Européenne, a donné naissance dans nombre de cas à des registres nationaux sur lesquels se sont appuyés les recensements. Il s'agit également du point de départ d'opérations d'appariements avec d'autres registres nationaux (logement, sécurité sociale) qui ont pu conduire, comme au Danemark, à la création d'un système statistique intégré⁴¹. Dans ces pays, dans la mesure où ces données administratives sont plus difficiles à anonymiser que des données issues d'enquêtes, les difficultés liées à l'accès, par des chercheurs, à des données très détaillées ou confidentielles ont été appréhendées beaucoup plus tôt que dans les pays dépourvus de registres.

L'histoire nationale a également une incidence sur l'existence d'identifiants, très déterminant pour les possibilités d'appariement entre bases administratives. A l'opposé du modèle du Danemark et des Pays-Bas, il existe en Allemagne, depuis la fin de la seconde guerre mondiale, une grande méfiance à l'égard des appariements.

En dernier lieu, la comparaison doit être attentive aux problèmes de terminologie. La notion de données couvertes par un secret, de données confidentielles ou encore de données

³⁹ Cf. not. Gordon Brackstone of Statistics Canada in « *Statistical Issues of Administrative Data: Issues and Challenges* », 1987.

⁴⁰ Cf. la définition des données administratives donnée par la Conférence des statisticiens européens (CES, Conference of European Statisticians) dans sa *Terminology on Statistical Metadata* : « *Data collected by sources external to statistical offices* ».

⁴¹ L'institut danois de statistiques, très dépendant des sources administratives, centralise l'ensemble des données et l'accès des chercheurs aux données de la statistique publique est, dès la production de celles-ci, lié à l'accès aux données administratives.

personnelles ne recouvre pas le même sens juridique selon les pays concernés. Il en va de même pour qualifier les modalités d'accès sécurisé à des bases de données publiques.

2.2. Des États à « registres »

2.2.1. L'exemple du Danemark

Le Danemark est le pays le plus souvent donné en exemple en Europe du fait de l'ancienneté et de l'ampleur de l'ouverture aux données publiques qui s'est très tôt accompagnée de la mise en place d'un accès sécurisé à distance par l'Institut national de statistique. La présence de registres dans la production de la statistique publique, accrue avec la création en 1968 d'un numéro unique d'identification, utilisé pour le recensement et les travaux d'appariements, ont facilité la mise en place d'un cadre normatif unifié et relativement simple d'accès pour l'ensemble des données couvertes par des secrets protégés par la loi.

Dans ce pays, l'Institut national de statistique (*Statistics Denmark*) joue dès lors un rôle central dans l'accès sécurisé et unique des chercheurs à un ensemble très large de données, avec une procédure unifiée. Cet accès a été ouvert en 1986, d'abord sur site puis, en 2000, en ligne. Il a été étendu aux chercheurs danois en poste à l'étranger depuis 2003 et, en 2007, aux chercheurs étrangers non-résidents dans les mêmes conditions.

Cette ouverture s'est accompagnée d'une modification du cadre normatif. En 1985, le *Public Administration Act* a permis à une administration d'imposer le secret professionnel à d'autres personnes. Cette disposition, complétée en 1987, charge *Statistics Denmark* de fournir l'accès aux données pour les chercheurs.

En parallèle, des négociations ont été engagées à partir de 2001 avec le ministère de la recherche pour mettre en place un modèle économique reposant sur un financement de l'accès par le budget de la recherche. Le centre sécurisé de *Statistics Denmark* effectue ainsi à la demande des chercheurs des appariements entre bases de données. Les appariements peuvent être demandés pour un très grand nombre de registres et bases administratives danoises détenues par le gouvernement, les organismes qui en dépendent ou des institutions extérieures. Y sont incluses, après autorisation de l'autorité en charge de la protection des données personnelles danoise, les données détenues par l'autorité danoise de santé (*Danish Health and Medicines Authority*).

Pour accéder à ces données, les chercheurs doivent justifier d'une appartenance à un environnement de recherche « pré-approuvé » par *Statistics Denmark*, soit une liste d'universités et centres de recherche publics agréée en collaboration avec le ministère en charge de la recherche. Dans le secteur privé, les institutions disposant d'un environnement de recherche permanent sont éligibles de même que les sociétés de conseil (effectuant notamment des analyses à la demande d'une autorité gouvernementale) ou d'autres entreprises (avec des critères sur les compétences et l'expérience en matière de recherche). L'accès n'est en revanche pas autorisé pour les médias.

Les utilisateurs disposent donc d'un point d'entrée unique, permettant un accès très large aux bases de données publiques. Cette ouverture repose sur une politique forte et constante de l'Institut national de statistique, en coopération avec le monde de la recherche, et s'appuie sur un modèle économique qui sollicite l'utilisateur final.

2.2.2. *L'exemple des Pays-Bas*

Les Pays-Bas se caractérisent également par l'ancienneté et l'importance de leur politique d'accès des chercheurs aux données publiques.

Cet accès est organisé par l'institut national de statistique (CBS) : il existe une offre de fichiers anonymisés pour la recherche et un accès sécurisé, sur site ou à distance, pour des données confidentielles.

CBS ne recueille ni ne stocke les données : il effectue les appariements nécessaires à la statistique publique, à partir desquels sont également produits des fichiers sécurisés pour la recherche ou des appariements à la demande. Certaines institutions détiennent des bases administratives qui ne sont pas utilisées pour les besoins de la statistique publique mais elles peuvent transférer ces bases à CBS qui peut, avec l'accord de ces institutions, les mettre à disposition des chercheurs dans son centre sécurisé. C'est le cas, notamment, pour les données de santé.

Cette centralisation se traduit également par une procédure d'habilitation des chercheurs unifiée et centralisée par CBS. L'habilitation se fait de façon simplifiée pour les universités et d'autres institutions ayant une fonction de recherche ou de planification : la liste de ces organismes est établie et mise à jour par un comité externe indépendant. Les demandes émanant d'autres utilisateurs finaux sont examinées par CBS. Une fois l'habilitation obtenue, CBS n'exerce un contrôle que sur l'anonymisation des résultats sortants.

En outre, l'ensemble des parties prenantes, dont les chercheurs, se réunit régulièrement au sein d'un comité de concertation, dont CBS assure le secrétariat.

Cette politique de large accès aux données publiques est favorisée, en amont, par une politique d'archivage à moyen et long terme. Les données sont normalement conservées pour une durée de 5 ans mais les chercheurs peuvent demander un archivage plus long. Par ailleurs, tout fichier qui a été utilisé dans le cadre du service sécurisé pour la recherche est archivé de façon pérenne pour les générations futures. Là encore, l'archivage des bases de données est assuré par CBS.

Enfin et à la différence du modèle économique danois, il n'y a pas de financement national par la recherche pour l'organisation de cet accès aux données. Les données elles-mêmes ne sont pas payantes, seul le service spécifique rendu à la recherche est facturé à l'utilisateur final.

2.3. Des systèmes statistiques non centralisés

Au Royaume-Uni, aux États-Unis et en Allemagne, l'accès des chercheurs aux données administratives se heurte à de plus grandes difficultés en raison d'une organisation statistique non centralisée et d'une moindre culture en matière de collecte et d'ouverture de données, dont certaines peuvent être couvertes par un secret protégé par la loi. Les États-Unis et le Royaume-Uni partagent toutefois des préoccupations liées à la recherche et ont souhaité, depuis plusieurs décennies, favoriser la constitution de grandes banques de données pour les sciences sociales.

2.3.1. *L'exemple du Royaume-Uni*

Au Royaume-Uni, l'institut de statistique britannique, l'*Office for National Statistics* (ONS), n'est pas en mesure de coordonner l'accès pour les chercheurs à ces bases mais, en collaboration avec le *Research Council (RCUK)* en charge des sciences sociales, il mène une

politique active d'ouverture des données à la recherche et de concertation avec les différentes parties prenantes.

C'est ainsi qu'a été créé le réseau d'accès sécurisé aux données administratives, *Administrative Data Research Centres (ADNR)*, qui s'appuie sur un centre d'accès dans chaque *nation* du Royaume-Uni et dispose d'un comité de pilotage qui permet une concertation régulière de l'ensemble des acteurs. L'habilitation des projets de recherche est désormais harmonisée, avec une procédure unique et des critères identiques. L'accès est ouvert aux chercheurs des institutions figurant sur une liste approuvée par le *RCUK* et sur laquelle figurent les universités ainsi que des institutions gouvernementales ou des organismes à but non lucratif au sein desquels travaillent des chercheurs. Une réflexion est en cours pour étendre cet accès au secteur des organismes à but lucratif lorsqu'ils sont en charge de travaux commandés par le gouvernement. La procédure d'habilitation est identique à celle utilisée par l'*ONS* pour l'accès des chercheurs aux données de la statistique publique.

La création de l'*ADNR* s'est accompagnée d'un financement important du *RCUK* pour la mise en place du réseau des centres sécurisés mais également pour les producteurs de données administratives afin de faciliter la préparation des fichiers et leur documentation.

Le Royaume-Uni se distingue également par la grande attention portée à la communication et la participation des citoyens sur les questions d'accès aux données individuelles, avec notamment des panels de citoyens siégeant dans certaines instances.

2.3.2. *L'exemple des États-Unis*

La dispersion des bases de données est encore bien plus importante aux États-Unis, en raison de la structure fédérale de ce pays. Chaque État fédéré dispose d'une forte autonomie tant en matière de production de données qu'en matière de réglementation et de procédure pour l'accès à ces données.

En outre, y compris au niveau fédéral, l'on dénombre de nombreuses agences, ce qui atteste d'une certaine réticence à la centralisation en matière de données et aux échanges entre administrations. La part laissée au secteur privé pour nombre de services (éducation, santé etc.) est également importante et implique une grande variété des modes d'accès aux données.

Le tableau général est difficile à saisir dans son ensemble compte tenu de la pluralité des cadres et des acteurs.

Le *Census Bureau* offre certes depuis longtemps des accès aux données confidentielles de la statistique publique aux chercheurs, avec un réseau de *Federal Statistical Research Data Centres (RDC)* qui s'est progressivement étendu. Chaque RDC dessert un périmètre d'utilisateurs limités en principe aux universités impliquées dans le financement du centre, entraînant des difficultés d'accès pour les chercheurs ne relevant pas de ces universités.

Le périmètre des données accessible s'est également élargi à des données de différentes instances fédérales avec lequel le *Census Bureau* négocie pour l'ensemble du réseau, mais sans implication à ce niveau d'instances de la recherche. Il n'existe du reste pas de cadre général de concertation entre la statistique publique fédérale et la recherche.

La mobilisation des données administratives tant pour la statistique publique que pour la recherche fait actuellement l'objet de plusieurs initiatives. Ainsi, un workshop a été organisé avec les chercheurs à Washington les 29 et 30 novembre 2016 par le *Public-Academic Research Colloquium* (« *Leveraging Administrative Data for Social Policy Innovation* ») avec un intérêt fort pour les expériences de pays européens en la matière. Au même moment, un

questionnaire (également adressé à des centres de recherches à l'étranger) vient d'être lancé par l'Université de Pennsylvanie sur l'accès à ce type de données pour essayer de faire avancer les choses aux États-Unis. Au niveau fédéral enfin, le Département du Commerce vient de mettre en place des auditions devant une *Commission on Evidence Based Policymaking* qui ambitionne notamment de dresser un état des lieux des difficultés et de proposer des recommandations sur la mobilisation des données administratives tant pour la statistique publique que pour la recherche. La Commission a manifesté de l'intérêt pour les évolutions en cours en France dans le domaine.

2.3.3. L'exemple de l'Allemagne

L'Allemagne dispose également d'une organisation fédérale, les *Länder* bénéficiant d'une certaine autonomie dans la production de nombreuses bases de données - leur accord est ainsi nécessaire pour autoriser, au niveau fédéral et *via* l'office fédéral de la statistique (*DESTATIS*), l'accès aux données.

La mobilisation des sources administratives pour la statistique fédérale devrait prendre de l'ampleur avec la nouvelle loi sur la statistique fédérale qui comporte notamment un article consacré aux données administratives. Il y est indiqué qu'avant toute collecte de données au niveau fédéral, *DESTATIS* vérifie si ces informations peuvent être couvertes par des données administratives existantes de qualité suffisante. Le processus est très contrôlé avec une première étape de transmission aux offices statistiques des métadonnées, puis un échantillon de données anonymisées et enfin si la qualité est jugée suffisante, le transfert des données dans le cadre de la loi statistique mais sous réserve cependant des autres lois existantes pouvant couvrir spécifiquement ces bases. A terme, cela pourrait favoriser un accès plus facile pour la recherche *via* l'office fédéral de la statistique.

L'accès aux données couvertes par les secrets légaux et professionnels, notamment celles de la statistique publique, est resté en Allemagne longtemps très difficile, voire impossible. Depuis quelques années, un processus d'ouverture est à l'œuvre, sous l'impulsion du *German Data Forum*. Cette instance, créée au sein du ministère en charge de la recherche (sur le modèle du Comité de concertation pour les sciences humaines et sociales qui avait été créé en France en 2001 et qui a été supprimé depuis en 2015), fonctionne depuis une dizaine d'années, rassemblant régulièrement producteurs de données (notamment *DESTATIS* et les ministères) et représentants des chercheurs pour discuter de l'organisation de l'accès aux données. Cette instance est à l'origine de la création d'un réseau de centres d'accès aux données, financé par le ministère de la recherche sur proposition du *German Data Forum*, pour les uns donnant des accès sécurisés à des données confidentielles dont des données gouvernementales, pour d'autres favorisant le partage de données académiques. Des propositions ont été faites pour avancer dans l'harmonisation des procédures d'habilitation et de leurs critères (très strictement restreints au monde académique), mais des difficultés demeurent en raison de la multiplication des centres sécurisés et de la faible synergie existante entre ces derniers.

2.4. La situation de la France

Aux termes de cette comparaison, quelques constats peuvent être dressés, qui permettent d'appréhender de manière plus précise la position de la France dans le paysage international en matière d'accès aux données publiques.

L'accès des chercheurs aux bases administratives est facilité lorsque le système statistique est lui-même fortement ancré dans ces bases, comme c'est le cas dans des pays historiquement à « registres de population ». Le cadre juridique est généralement plus unifié et met l'institut national de statistique en position centrale pour organiser l'accès pour les chercheurs et effectuer également des appariements à leur demande. L'existence d'un point unique d'information comme d'accès avec une procédure d'habilitation unique facilite également les choses pour les chercheurs.

A l'opposé, des systèmes statistiques mobilisant peu les sources administratives et reposant sur une production décentralisée de ces sources rendent d'autant plus complexe l'accès des chercheurs à ces données que le cadre institutionnel est lui-même décentralisé comme c'est le cas dans des systèmes fédéraux ou à forte autonomie régionale. Dans de telles configurations, les exemples précédents montrent l'intérêt de mettre en place des instances de concertation permanentes pour favoriser un point d'accès unique à l'information pour les chercheurs, harmoniser les procédures et les critères et créer des synergies permettant une économie de moyens. A la différence des pays où l'institut statistique national est en position centrale pour organiser l'accès, ces contextes se traduisent le plus souvent par la multiplication des accès sécurisés, source de difficultés supplémentaires lorsque les chercheurs ont besoin de travailler sur des sources hébergées dans des centres différents.

La France se situe à cet égard dans une position intermédiaire. Très en retard jusqu'au début des années 1990 en ce qui concerne l'ouverture à la recherche des données de la statistique publique et, plus généralement, le partage des données, elle se trouve aujourd'hui plutôt en pointe sur le plan juridique pour l'accès aux données confidentielles et sensibles, y compris administratives, après les changements successifs intervenus dans les cadres juridiques et réglementaires.

L'Insee a été un acteur essentiel pour soutenir cette ouverture, dans la continuité de la coopération entamée avec la banque de données pour la recherche, le Réseau Quetelet, depuis plus de 20 ans. Les données ainsi ouvertes à la recherche sont aujourd'hui très nombreuses et extrêmement riches, attirant également des chercheurs de nombreux pays.

Le retard dans la création d'un centre permettant un accès sécurisé aux données confidentielles a également été comblé avec la mise en place du CASD, appuyé sur une technologie innovante, rapidement monté en puissance, avec une croissance continue des sources disponibles tant du côté de la statistique publique que maintenant des données administratives.

Des faiblesses demeurent cependant au regard des avancées d'autres pays et des nouveaux enjeux qui se dessinent. Confrontés longtemps à des difficultés d'accès aux données, les chercheurs français restent bien moins utilisateurs des données que dans d'autres pays, notamment ceux considérés plus haut, en particulier dans les autres disciplines que l'économie.

Ce sont ces difficultés qu'il convient désormais d'examiner afin d'être en mesure de formuler quelques pistes d'amélioration en la matière.

3. AMÉLIORER L'ACCÈS DES CHERCHEURS AUX DONNÉES PUBLIQUES

Si l'état du droit en matière d'accès aux données publiques, en dépit de sa relative complexité, est de nature à permettre un accès assez large des chercheurs aux données détenues par les administrations, en pratique, des difficultés demeurent, sur lesquelles il importe de revenir avant de formuler des recommandations en la matière.

3.1. Des dispositifs législatifs complexes et mal connus

Si les dispositifs législatifs actuels permettent, en droit, un large accès des chercheurs aux données publiques, deux obstacles en limitent la portée pratique : i) les dispositifs législatifs sont mal connus, aussi bien des chercheurs que des administrations productrices ; ii) la complexité et la diversité des régimes d'accès aux données couverts par un secret se traduit par des procédures lourdes, et parfois dirimantes.

3.1.1. Une large méconnaissance des dispositifs législatifs d'accès aux données

Le groupe de travail a pu constater, durant ses travaux, que l'état du droit sur l'accès aux données publiques par des chercheurs – y compris avant les avancées législatives de la loi pour une République numérique – n'était que très imparfaitement connu ou compris des chercheurs comme des administrations productrices. Les lois applicables répondent en effet à des objectifs différents (gestion des archives, protection des données personnelles, droit d'accès aux documents administratifs...), ce qui contribue à rendre la maîtrise de l'ensemble des textes et de leur articulation complexe pour un non-spécialiste. Les évolutions constantes du droit renforcent cette difficulté.

Les principes de droit commun, la nature des secrets couverts et les dispositifs dérogatoires (cf. supra) devraient être largement disséminés afin de faciliter la connaissance des démarches à entreprendre pour l'obtention de données dans des délais raisonnables. En parallèle, une explicitation du droit adaptée aux besoins des spécialistes des données permettra aux producteurs de connaître le corpus juridique de référence régissant la mise à disposition et la conservation de données à des fins statistiques ou de recherche.

Une grande partie des réticences des administrations à autoriser l'accès des chercheurs aux données dont elles disposent tient à cette complexité, et à l'insécurité juridique qu'elle entraîne pour les administrations. À cet égard, on peut considérer comme une avancée notable les clarifications effectuées par la loi pour une République numérique s'agissant de la non-application de l'article 226-13 du code pénal dans le cadre des procédures d'ouverture anticipée des archives publiques.

De manière générale, il est important d'encourager les actions visant à faire connaître le cadre juridique applicable. A cet égard, le présent rapport pourra constituer une base utile en vue de l'élaboration d'un vade-mecum à destination des administrations et des chercheurs souhaitant accéder aux données publique (cf. recommandation n° 1).

En outre, le groupe de travail a constaté que les voies de recours précontentieuses et contentieuses pour contester le refus d'une administration de rendre ses données accessibles étaient le plus souvent inconnues des chercheurs⁴². Cette méconnaissance des voies de droit

⁴² Par exemple, en ce qui concerne les données statistiques, pour lesquelles la CADA est compétente, on ne dénombre qu'un seul avis rendu, en 2011, concernant l'accès, par dérogation aux délais fixés par l'article L.213-2

susceptibles d'être empruntées pour contester une décision de refus et tenter d'obtenir les données utiles pour mener à bien un projet de recherche s'explique, au moins en partie, par le peu de vulgarisation des dispositifs existants. Elle constitue, en aval et indépendamment de la complexité des procédures d'accès par anticipation, un facteur de blocage dans l'accès des chercheurs aux données publiques, auquel il pourrait être remédié à court terme par des mesures à fins de vulgarisation, simples à mettre en œuvre.

Recommandation n°1 : élaborer et diffuser un vade-mecum synthétique à destination des administrations et des chercheurs, rappelant l'état du droit applicable, présentant les différentes procédures d'accès et les voies de recours à disposition des chercheurs pour contester une décision de refus d'accès. L'élaboration de ce vade-mecum pourrait être confiée à l'instance de concertation (cf. recommandation n°4).

3.1.2. *Des dispositifs législatifs et réglementaires qui peuvent encore être améliorés*

Les normes législatives en vigueur prévoient donc une grande diversité de régimes de communication par anticipation pour les données couvertes par un secret protégé par la loi. Les procédures dérogatoires, qui viennent se superposer au régime de droit commun, ne sont pas harmonisées : l'administration compétente pour statuer sur une demande diffère pour chaque type de données et l'augmentation du nombre des autorités chargées d'émettre un avis au cours d'une procédure de demande d'accès par anticipation allonge le temps de traitement de ces demandes.

Ce facteur ne facilite pas les démarches des chercheurs et peut, dans de nombreux cas, les décourager. Cela peut notamment alourdir les demandes nécessitant l'accès à des données relevant de filières différentes. Il s'agit là également d'un élément de complexité pour les administrations elles-mêmes, qui peuvent être incitées à refuser tout accès aux données qu'elles produisent, faute d'être en mesure d'identifier, d'une part, les éventuelles données protégées par un secret et, d'autre part, le régime de communication applicable.

Au-delà de la recommandation n°1 qui vise, dans un délai court, à diffuser l'état du droit, il pourrait être envisageable, à plus long terme, que le législateur propose un rapprochement des procédures législatives et réglementaires d'accès aux données couvertes par un secret. L'objectif serait de proposer un cadre législatif plus cohérent, sans pour autant remettre en question les dispositions spécifiques à certaines données du fait de leur nature ou de leur sensibilité.

Les points suivants ont été ainsi relevés :

- L'absence de recours possible à la CADA dans le cas de l'accès aux données fiscales (cf. 1.2.2.2.2) ;
- L'absence d'un référentiel de sécurité commun pour les données sensibles : le décret d'application de l'accès aux données fiscales⁴³ prévoit un accès uniquement par le CASD, tandis que l'accès aux données du SNDS est déterminé par un référentiel de

du code du patrimoine, des documents conservés dans le service des archives départementales de l'Ain et portant une enquête statistique ayant eu lieu en 1988 dans une exploitation agricole. La demande émanait en l'espèce non pas d'un chercheur mais du fils de l'exploitant agricole concerné.

⁴³ Décret n° 2014-950 du 21 août 2014 relatif à l'accès aux informations fiscales à des fins de recherche scientifique prévu par l'article L.135 D du livre des procédures fiscales.

- sécurité, et les données de la Banque de France est permis par le biais du centre sécurisé de la Banque ;
- Le traitement des données judiciaires n'est pas actuellement possible, alors même que l'accès à ces données est bien couvert par les dispositions de la loi pour une République numérique.

Recommandation n° 2 : Le législateur pourrait rapprocher les procédures législatives d'accès et de traitement des données couvertes par un secret.

3.1.3. *Les conditions de réussite de la loi pour une République numérique*

La nouvelle disposition législative issue de la loi Pour une République numérique devrait contribuer à lever les incertitudes juridiques et favoriser ainsi un meilleur accès des chercheurs aux données publiques. Pour que cette disposition législative, au-delà de la clarification qu'elle apporte, soit une réussite pratique, il importe néanmoins que les administrations utilisent effectivement la faculté qui leur est ouverte de saisir le comité du secret statistique.

Rappelons en effet qu'une telle saisine est purement facultative. A cet égard, les conditions de saisine, et notamment le critère de l'intérêt public, ne semblent pas de nature à constituer, en tant que tels, des obstacles à la saisine du comité. Il est en effet peu envisageable que le comité du secret statistique, dont la compétence *rationae materiae* est désormais extrêmement étendue puisqu'elle porte sur l'ensemble des bases de données, refuse d'éclairer une administration au seul motif que la demande initiale ne revêtirait pas un caractère d'intérêt public.

La question est davantage celle de savoir si, en pratique, et bien qu'elles n'en aient nullement l'obligation, les administrations feront effectivement usage de cette possibilité de saisine du comité. Certaines administrations pourraient en effet hésiter à saisir le comité, par crainte d'alourdir inutilement la procédure d'accès pour elles-mêmes et pour le chercheur ou, de manière générale, par crainte de faire intervenir un organisme tiers dans une question qu'elles estiment relever de leur seule compétence. La saisine du comité étant facultative, les administrations pourraient ainsi choisir de mettre en place leur propre système d'habilitation et d'accès sécurisé.

Pourtant, les administrations ont tout à gagner, afin de fixer au plus vite des standards en matière de modalités d'habilitation et d'accès sécurisé aux données par des chercheurs, à faire usage de cette faculté de saisine du comité du secret statistique. Celui-ci, doté d'une réelle expertise en la matière, dont les administrations ne disposent pas nécessairement, sera en effet en mesure de lever d'éventuels doutes sur les difficultés d'ordre juridique et de recommander des modalités d'accès sécurisé aux données concernées. Pour les administrations concernées, cette possibilité de saisine du comité du secret statistique, réservée aux données soumises au secret, devra néanmoins être articulée de manière efficace avec leur organisation en matière de mise à disposition de données (et notamment de données non soumises au secret), pour leur permettre d'avoir une approche cohérente et une vue globale sur les utilisations de leurs données et les recherches associées⁴⁴.

⁴⁴ Les questions que soulève la saisine du comité, notamment en termes de moyens, sont traitées dans la partie 4.

Recommandation n° 3 : inciter les administrations à faire pleinement usage de la faculté qui leur est ouverte par la loi pour une République numérique de saisir le comité du secret statistique.

3.2. Améliorer les modalités pratiques d'accès aux données

Malgré les progrès récents du cadre législatif permettant l'accès aux données administratives à des fins de recherche, il reste des difficultés pratiques de cet accès pour les chercheurs, sur lesquels des progrès peuvent être fait à moyen terme.

3.2.1. Le besoin d'une instance d'une concertation unique

Dans un contexte de grande dispersion des services producteurs et de modalités d'accès à leurs données, le groupe de travail a fait le constat de la nécessité d'une instance de concertation unique permettant une meilleure coordination entre les différentes institutions en charge de l'accès aux données par les chercheurs (Archives, Cnil, Cnis, Administrateur général des données, et ministère de la recherche), les administrations productrices et les chercheurs.

Les tentatives de coordination et d'unification des procédures qui ont été récemment mises en œuvre, notamment via le réseau Quetelet, sont louables mais insuffisantes. Peu formalisées, ces initiatives peinent à produire les synergies nécessaires pour répondre aux besoins des chercheurs. Il n'existe pas non plus de structure de concertation permanente entre l'ensemble des acteurs, à l'instar de la structure de gouvernance de l'*Administrative Data Research Centre* au Royaume-Uni, du *German Data Forum* en Allemagne ou du comité de coopération aux Pays-Bas⁴⁵, le Comité de concertation pour les données en sciences humaines et sociales (CCDSHS) créé en 2001, ayant été supprimé en 2015 par le ministère de la recherche.

Cette instance de concertation unique aurait vocation à centraliser les informations sur les modalités d'accès aux différentes sources (cf. recommandation 1 sur le vade-mecum sur les questions juridiques), à centraliser l'information sur les bases de données existantes, à recueillir les retours des administrations productrices et des chercheurs et à organiser la concertation entre les différentes parties prenantes afin d'améliorer le fonctionnement pratique de l'accès aux données, selon l'ensemble des filières (cf. recommandations 5, 6 et 7).

Cette instance, outre son caractère de concertation, aurait vocation à entretenir un site web 2.0, un portail numérique interactif, permettant de diffuser largement les informations juridiques et pratiques sur les accès aux données, ainsi que sur les travaux des chercheurs réalisés à partir de ces données.

Cette mission se rapproche en grande partie des objectifs et des pratiques du Cnis pour ce qui concerne les données statistiques. En raison de son expérience dans ce domaine, le Cnis aurait vocation à participer activement à une telle instance. Dans le même temps, les données administratives couvrant un champ nettement plus large que la Statistique publique, confier une telle instance au Cnis aurait pour effet d'en modifier la nature, l'éloignant par définition de la Statistique publique, et élargissant son rôle à l'ensemble des données publiques. On pourrait citer de la même façon le comité du secret, qui pourrait jouer ce rôle d'instance de coordination, mais dont la mission devrait évoluer de façon similaire. Par ailleurs, si le financement du Cnis a naturellement vocation à être pris en charge par l'Insee, le financement d'une instance visant à faciliter l'exploitation des données administratives provenant de

⁴⁵ Cf. la partie 3 du rapport sur les comparaisons internationales.

l'ensemble des administrations publiques, et à des fins de recherche scientifique, invite à réfléchir à une solution institutionnelle, et un mode de financement, reflétant ces objectifs spécifiques (incluant par exemple le ministère en charge de l'Enseignement supérieur et de la Recherche). Enfin, le rôle des Archives dans le processus d'autorisation des accès aux données, et de suivi de l'archivage des données sources, invite à bien articuler son positionnement au sein de cette instance de concertation.

Recommandation n° 4 : mettre en place une instance de concertation unique, regroupant des représentants des chercheurs et des administrations, ainsi que les centres sécurisés d'accès aux données et les organes interministériels compétents en matière de systèmes d'information et d'accès aux données et à leur traitement.

3.2.2. *L'identification des données existantes est difficile*

Le premier problème auquel sont confrontés les chercheurs est celui de l'identification et de la localisation des données disponibles. Il est difficile de savoir quelles sont les données existantes et, lorsque celles-ci sont connues des chercheurs, quelle administration solliciter pour en demander l'accès.

Si premier alinéa de l'article L. 322-6 du code des relations entre le public et l'administration, qui est venu codifier l'article 17 de la loi du 17 juillet 1978 sur le droit d'accès aux documents administratifs, dispose que : « *Les administrations qui produisent ou détiennent des informations publiques tiennent à la disposition des usagers un répertoire des principaux documents dans lesquels ces informations figurent* », les dispositions de l'article R. 322-7 du CRPA (ancien 36 du décret du 30 décembre 2005) ne confèrent pas un caractère exhaustif à ce répertoire. Chaque administration dispose ainsi d'une marge d'appréciation pour constituer ce répertoire.

Certaines démarches ont été récemment mises en œuvre par le Réseau Quetelet⁴⁶ et le CASD⁴⁷ afin d'établir des listes de données soumises à une procédure d'accès sécurisé ou non et recensant les millésimes disponibles, mais ces démarches ne suffisent pas à couvrir le champ très vaste des données potentiellement disponibles.

L'objectif que l'instance de concertation (cf. recommandation 4) pourrait se donner est d'entamer un travail de recensement des données existantes, sans viser à dresser une liste complète des documents administratifs existants. Le travail de recensement pourrait être orienté, de façon prioritaire, sur l'identification des bases de données qui sont le plus susceptibles de présenter un intérêt pour les réutilisateurs (au premier rang desquels les chercheurs). Ce travail de recensement pourrait ainsi être disponible sur le site web de l'instance de coordination, et se voir enrichi au fur et à mesure que le recensement est réalisé.

Recommandation n° 5 : inciter, *a minima*, les administrations à enrichir le répertoire des informations publiques visé à l'article L. 322-6 du code des relations entre le public et l'administration et engager, en coordination avec les services producteurs de données, l'administration des archives et les chercheurs, un processus plus large de recensement des principales sources de données existantes.

⁴⁶ Composante de Progedo, en charge des bases de données détaillées anonymes. <http://www.reseau-quetelet.cnrs.fr>.

⁴⁷ Le portail du CASD dresse ainsi la liste des bases de données dont il dispose : <https://casd.eu/fr/donnees>.

3.2.3. La diversité des procédures et des modalités d'accès aux données

Lorsque les données ont été identifiées, de nouvelles difficultés surgissent pour les projets de recherche mobilisant les données de plusieurs services producteurs, y compris dans le cas où ces services appartiennent au même ministère. La diversité des producteurs peut se traduire par une multiplicité de procédures et formulaires d'habilitation et à des modes d'accès différents avec lesquels le chercheur doit à chaque fois se familiariser.

Par exemple, afin d'estimer l'impact du système socio-fiscal pour le retour à l'emploi des personnes disposant des plus bas revenus, il est nécessaire de mobiliser des sources statistiques (DADS), des sources fiscales (déclaration de revenu), des données détenues par Pôle Emploi concernant les allocataires chômage, ainsi que des données de la CNAF (échantillon national des allocataires, RSA, APL, etc.). En l'état actuel du droit, trois dispositifs juridiques distincts encadrent actuellement l'accès à ces différentes sources, ce qui peut supposer trois démarches et modes d'accès différents.

Dans le cas particulier des données fiscales, s'ajoute d'ailleurs un inconvénient lié à la multiplication des formulaires exigés : un chercheur souhaitant accéder à ce type de données devra par exemple signer un engagement de confidentialité pour le comité du secret et un autre pour le service producteur, c'est-à-dire la DGFIP, alors qu'il s'agit du même projet et des mêmes données.

Recommandation n° 6 : harmoniser les formulaires existants qui doivent être remplis par les chercheurs pour accéder aux données publiques.

Ce constat vaut également pour les données en matière d'éducation, qui permet d'illustrer le cas dans lequel l'ensemble des services producteurs relèvent du même ministère. En effet, le ministère de l'Éducation nationale dispose de deux services ministériels⁴⁸, dont l'activité est au demeurant totalement autonome, auxquels viennent s'ajouter les services statistiques académiques et d'autres services du ministère, qui disposent de bases de données très riches, notamment au sein la direction générale des ressources humaines. Lorsque les travaux des chercheurs nécessitent de croiser les fichiers conservés par ces différents services producteurs (par exemple, pour des projets de recherche concernant la transition entre le lycée et l'enseignement supérieur ou la mobilité des enseignants), les chercheurs doivent réaliser autant de démarches que de bases de données auxquelles ils souhaitent accéder, alors qu'il s'agit de la même administration.

L'appariement des données de gestion détenues des services producteurs différents se heurte donc, en l'état actuel des modalités d'accès, à l'impossibilité de réunir ces services au sein d'une instance unique, afin de permettre un accès sécurisé plus rapide aux bases de données pertinentes pour un projet de recherche et selon une procédure unique. En outre, dans certains champs de recherche, il n'existe pas à ce jour de procédure de cryptage unique pour certains identifiants personnels (par exemple, concernant l'Éducation nationale, la DEPP, le SIES et les rectorats ont chacun leur propre algorithme de cryptage pour l'identifiant élève, dit INE).

De la même manière, lorsqu'un chercheur souhaite évaluer la portée économique des aides à l'innovation pour les entreprises, il lui est nécessaire d'utiliser conjointement des données Banque de France (bases entreprise), des données fiscales (liasses fiscales) et des

⁴⁸ Il s'agit de la DEPP pour les données portant sur le premier et le second degré et du SIES pour les données de l'enseignement supérieur.

données liées aux subventions elles-mêmes (BPI France notamment). Il lui faudra donc s'adresser respectivement à la Banque de France, qui n'autorise l'accès à ses données que via son propre centre sécurisé, à BPI France, qui rend ses données accessibles via le CASD et ainsi qu'au ministère des finances pour les données fiscales, procédure qui fait intervenir le comité du secret statistique en amont et le CASD en aval.

Au-delà de la difficulté liée au temps nécessaire à la réalisation de ces démarches, l'exploitation des données ne sera possible que si les services producteurs acceptent que l'accès aux données soit conjoint, ce qui n'est possible actuellement que par l'usage du CASD. Des solutions efficaces devront être trouvées pour permettre le rapprochement des données, en évitant d'augmenter la charge des producteurs et la duplication non nécessaire de données ou de centres sécurisés d'accès : il semble plus efficace d'organiser des transferts ponctuels de données entre dispositifs d'accès sécurisés (sous réserve que chacun vérifie les garanties minimales exigées par le producteur) que de demander aux producteurs de multiplier le nombre des centres avec lesquels ils contractualisent et dans lesquels ils déposent directement leurs données.

Recommandation n° 7 : prévoir des procédures d'accès sécurisé qui rendent techniquement possible l'appariement des bases de données.

3.2.4. Favoriser les partenariats de recherche entre administrations productrices et chercheurs

En dernier lieu, des obstacles d'ordre davantage culturels pourraient également expliquer les réticences de certaines administrations à autoriser l'accès des chercheurs aux données qu'elles détiennent.

Rappelons tout d'abord que la conservation et la mise à disposition des données à des fins de recherche ne figurent pas, à proprement parler, parmi les missions de service public des administrations qui détiennent ces données. Elles peuvent de ce fait rarement consacrer des moyens conséquents à ces activités non prioritaires.

Ensuite, il existe aujourd'hui un sentiment ancré de patrimonialité des administrations sur « leurs » données, qui n'est évidemment pas sans effet sur la politique d'ouverture des données publiques. Livrer ses données c'est s'exposer à la critique, en abandonner la maîtrise et l'expertise est souvent perçue comme une perte de pouvoir. Ainsi, comme le relevait le Conseil national du numérique, « *basculer d'une culture de la détention de l'information à une culture de l'ouverture et de la diffusion* »⁴⁹ correspond à un bouleversement auquel l'ensemble des administrations ne sont pas nécessairement préparées.

Sans doute un tel changement culturel ne se décrète pas. Il peut néanmoins être favorisé par certaines actions visant à instaurer un climat de confiance entre administrations et acteurs de la recherche et à faire prendre conscience que l'accès aux données publique s'avèrera le plus souvent mutuellement bénéfique : les chercheurs optimiseront leurs résultats, les administrations renforceront leur connaissance et l'évaluation de la politique publique dont elles sont responsables.

Ainsi, la mise en place de partenariats de recherche (autour de cofinancements par exemple entre l'Agence nationale de la recherche (ANR) et des administrations productrices de

⁴⁹ Conseil national du numérique, « Ambition numérique. Pour une politique française et européenne de la transition numérique », Rapport remis au Premier ministre, juin 2015.

données) peut être un levier fort pour des travaux de recherche conjointement menés par des administrations et des équipes de chercheurs. Les administrations peuvent dans ce cadre inciter à favoriser des thèmes de recherche qui ont au cœur de leur mission, et pour lesquels elles peinent souvent à attirer la recherche scientifique. A l'inverse, les chercheurs peuvent bénéficier de l'expertise propre des administrations sur leurs données (leur mode de production, leurs limites, etc.).

Afin d'instaurer un climat de confiance entre administrations et monde de la recherche, il convient également d'assurer un bon retour des publications et travaux réalisés à partir des données mises à dispositions. A cet égard, il convient de veiller à ce que les chercheurs adressent aux administrations productrices des données les résultats de leur recherche (qui peuvent être aussi diffusés par le biais du site web de l'instance de concertation).

Recommandation n° 8 : encourager les partenariats entre chercheurs et administrations, en insistant sur le caractère mutuellement bénéfique d'un accès large aux données administratives à des fins de recherche.

3.3. Développer un modèle économique pérenne de l'accès aux données administratives

L'utilisation des données administratives à des fins de recherche implique des coûts, de la préparation de ces données par les producteurs, à l'exploitation de ces données par les chercheurs, en passant par les procédures d'accès. Ces coûts doivent être identifiés, afin de déterminer qui soit les supporter, et surtout pour développer un modèle économique pérenne le plus efficace possible pour l'ensemble de la société, c'est-à-dire en limitant le plus possible les coûts inutiles.

3.3.1. Les coûts associés à la préparation des données par les producteurs

La mise à disposition de données administratives à des fins de recherche requiert en premier lieu que les producteurs de ces données mènent des travaux pour concevoir, produire, documenter, puis mettre à disposition des bases de données répondant aux besoins de la recherche. Parfois, ils disposent déjà de telles bases documentées, notamment dans le cas où ils mettent déjà leurs données à disposition d'autres utilisateurs publics. Dans ce cas, des coûts supplémentaires sont à prévoir pour supprimer les identifiants individuels⁵⁰ (cf. 3.3.1.1.) et pour les diffuser aux chercheurs de manière sécurisée (contractualisation, modes d'accès sécurisés, etc.).

Quand il n'existe pas de bases de données déjà prêtes pour une mise à disposition, des moyens plus importants sont parfois nécessaires pour concevoir puis extraire ces données des systèmes de gestion. Le coût peut être particulièrement élevé quand les systèmes d'informations concernés ne sont pas conçus dès l'origine pour la constitution de bases de données pour des usages statistiques ou recherche, ou quand il est difficile de répondre de manière simple à la diversité des besoins (cf. 3.3.1.2).

⁵⁰ Aux termes de l'article R. 322-3 du CRPA : « Lorsque la réutilisation n'est possible qu'après anonymisation des données à caractère personnel, l'autorité détentrice y procède sous réserve que cette opération n'entraîne pas des efforts disproportionnés ».

3.3.1.1. La suppression des identifiants individuels

La charge que constitue le travail préalable d'anonymisation (cf. 1.3.2.3) pour rendre les données librement communicables peut conduire les administrations à préférer donner accès à une base non anonymisée, simplement dé-identifiée. C'est d'ailleurs souvent la seule manière de répondre au besoin du chercheur, dans le cas où des données librement communicables ne seraient pas suffisamment détaillées pour lui.

La suppression des identifiants individuels nécessite des opérations simples de préparation des données, dont le coût ne peut être supporté que par les administrations productrices.

3.3.1.2. L'extraction de données de gestion

Outre cette exigence de dé-identification des données sollicitées, les administrations peuvent parfois refuser l'accès à des données qu'elles détiennent au motif que ces dernières, afin d'être exploitables par la recherche scientifique, nécessitent un traitement spécifique ou ne constituent pas des documents achevés. Une telle difficulté concerne essentiellement les données dites de gestion (ou brutes)⁵¹, c'est-à-dire celles qui sont recueillies au jour le jour par les administrations et qui n'ont pas vocation à faire l'objet par l'administration d'une opération d'agrégation⁵² ou d'extraction complexe⁵³.

Il est vrai que le droit d'accès ne saurait s'exercer à l'égard d'un document dont l'existence n'est pas établie⁵⁴ et que la loi du 17 juillet 1978 ne fait pas obligation aux autorités administratives d'élaborer un document à la demande d'une personne⁵⁵. Dès lors, l'administration n'est tenue, lorsqu'elle est saisie d'une demande tendant à la communication d'un dossier qui n'existe pas en tant que tel, ni « *de faire des recherches en vue de collecter l'ensemble des documents éventuellement détenus* »⁵⁶, ni d'établir un document en vue de procurer les renseignements ou l'information souhaités⁵⁷. La communication peut donc être refusée s'agissant de documents à confectionner, par le biais d'un traitement quelconque, de recherches, d'une synthèse, d'une analyse, etc.⁵⁸

Il n'en demeure pas moins que les autorités administratives sont tenues de communiquer les documents qui peuvent être obtenus par simple extraction des bases de données dont elles disposent. Tout dépend alors des caractéristiques de cette base (format, fonctionnalités, etc.), des termes de la demande et des efforts que l'Administration devrait déployer pour la

⁵¹ Rappelons que ces données de gestion, si elles ne constituent pas des données statistiques au sens de la loi du 7 juin 1951, entrent néanmoins dans le champ de l'article L. 213-3 du code du patrimoine et peuvent dès lors faire l'objet d'une demande d'accès anticipé.

⁵² Les données de gestion ne contiennent pas toutes les variables utiles pour des analyses. En particulier, certaines variables peuvent n'être calculées qu'en tant que de besoin (par exemple pour le calcul du droit d'un assuré à sa demande), à partir d'un ensemble d'autres variables et de la législation du moment. La variable ainsi calculée n'est en général pas enregistrée.

⁵³ Sur le plan technique, les bases de données de gestion sont souvent organisées en vue d'optimiser leurs performances en production (par exemple, beaucoup de petites tables avec peu d'information pour avoir des temps de réponse courts), ce qui complexifie les extractions (puis requiert des retraitements avant exploitation). Les données de gestion ne contiennent pas toutes les variables utiles pour des analyses : certains calculs ne sont en effet effectués qu'à la volée, en cas de consultation d'un dossier par le gestionnaire, et ne sont pas stockés.

⁵⁴ Cf. not. CE, 11 décembre 2006, *Ministre des affaires étrangères c/ M. Laurent*, n° 279113

⁵⁵ Cf. not. CE, 22 mai 1995, *Association de défense des animaux victimes d'ignominies ou de désaffection*, n° 152393.

⁵⁶ CE, 27 septembre 1985, *Ordre des avocats du barreau de Lyon*, n° 56543.

⁵⁷ Avis CADA, 8 janvier 1987, *Thomas*, 5^{ème} rapport page 109 - CE, 30 janvier 1995, *Ministre d'État, ministre de l'éducation nationale. c/ Mme Guigue*, n° 128797.

⁵⁸ CE, 9 mars 1983, *Association « SOS Défense »*, n° 43438.

satisfaire⁵⁹. De manière générale, les demandes pourront d'autant plus facilement être satisfaites que les bases de données et systèmes d'informations auront été conçus pour y répondre. Sur ces aspects, les évolutions s'inscriront dans le moyen-long terme, et devront tirer parti des créations ou rénovations des systèmes d'information.

3.3.1.3. La documentation des données administratives

Les bases de gestion posent également la question de leur documentation. Elles résultent souvent de législations, règles de gestions, systèmes d'information qui évoluent en permanence, ce qui peut rendre difficile la bonne compréhension de leur contenu, d'autant que la documentation métier vise plus souvent à décrire les règles et procédures de gestion actuelles que leur historique. Par ailleurs, si la documentation renseigne sur le contenu « théorique » des données, elle permet souvent difficilement d'appréhender leur qualité réelle, en particulier pour les informations qui ne sont pas essentielles pour la gestion (ou qui ne seront utilisées en gestion que plus tardivement, après fiabilisation)⁶⁰.

3.3.1.4. Quels financements pour les coûts de préparation des données?

Dans la mesure où la mise à disposition de données à des fins de recherche n'est en général pas une mission centrale des administrations, les moyens pouvant être mobilisés pour cette mise à disposition souffrent de la concurrence d'autres priorités. Cette problématique concerne notamment les organismes de sécurité sociale, dont les missions premières consistent à gérer des risques sociaux, à partir de financements dédiés (notamment sous forme de cotisations sociales). La question des moyens est plus prégnante quand la mise à disposition de données requiert d'importants travaux préalables pour définir, constituer et documenter des bases de données diffusables, et le cas échéant les anonymiser.

Dans ce contexte, il est souhaitable que les conventions d'objectif et de gestion (COG)⁶¹ des caisses de sécurité sociale prévoient des moyens supplémentaires à consacrer au développement de la mise à disposition de données et préciser les attentes associées⁶². Ces attentes dépendront évidemment de la nature et du contenu des données détenues par les caisses. Elles sont particulièrement marquées pour la mise à disposition des données médico-administratives du système national des données de santé. D'ailleurs, la COG 2014-17 de la Cnamts prévoyait déjà de « *poursuivre l'enrichissement des données du SNIIRAM et mettre en œuvre les décisions des pouvoirs publics relatives à l'ouverture des données de santé* » et indiquait les travaux à mener en priorité⁶³.

Dans un souci d'efficacité collective, ces attentes devront aussi dépendre de la manière dont ces données sont mises à disposition par ailleurs (notamment par la statistique publique) et de leur intérêt compte tenu de l'ensemble des sources existantes. Cela peut par exemple amener à envisager prioritairement la diffusion de bases regroupant déjà des données de tous les

⁵⁹ Ainsi, l'administration peut refuser de communiquer des données dont l'extraction suppose des requêtes informatiques complexes ou une succession de requêtes particulières qui diffèrent de l'usage courant pour lequel ce fichier a été créé, cf. avis CADA, n° 20140860 du 5 juin 2014, à propos de la base SNIIRAM.

⁶⁰ Ainsi, l'absence d'une donnée signifie-t-elle que la personne n'est pas concernée ou que l'information n'est pas encore disponible, voire qu'elle n'est plus disponible (données purgées quand elles ne sont plus utiles en gestion) ? Comment faire la part entre "vraies évolutions" et évolutions découlant uniquement de changements dans la gestion ?

⁶¹ Les COG sont des documents contractuels conclus entre l'État et les caisses nationales des principaux régimes de sécurité sociale. Elles formalisent la délégation de gestion du service public de la sécurité sociale aux organismes gestionnaires. Ces COG diffèrent selon chaque branche ou régime en fonction des axes stratégiques qui lui sont propres.

⁶² Cela concerne également le développement de l'offre open data.

⁶³ Cf. fiche 14 de la convention d'objectif et de gestion de la Cnamts <http://www.securite-sociale.fr/IMG/pdf/cogetatcnamts2014-2017-2.pdf>.

régimes d'une branche de la sécurité sociale, plutôt que de solliciter chaque régime indépendamment.

Recommandation n° 9 : Prévoir dans les conventions d'objectif et de gestion des organismes sous tutelle des moyens dédiés aux travaux concourant à la mise à disposition de données administratives.

Enfin, s'agissant de la question de la tarification de l'accès à cette catégorie de données, comme le souligne une étude conduite par l'université de Strasbourg sur les informations du secteur public⁶⁴, « *si l'objectif des politiques publiques est de diffuser des informations brutes, difficiles à comprendre et à utiliser pour le non initié alors il est optimal que ces informations soient gratuites* ». En effet, faire payer l'accès à ce type de données peut s'avérer économiquement sous-optimal compte tenu du faible consentement à payer de la majorité des potentiels utilisateurs : lorsque le potentiel de valeur ou l'intérêt scientifique des données ne peut être perçu, le montant de la redevance constitue uniquement un frein à la réutilisation.

Afin de faciliter l'accès des chercheurs à des données de gestion, sans pour autant faire reposer sur l'administration une charge de travail démesurée, il convient de favoriser l'accès « exploratoire » des chercheurs à des données sous forme « brute », avec simple suppression des identifiants individuels (cf. supra).

Un tel accès, qui devra naturellement s'effectuer dans des conditions de sécurité appropriées et conformément au cadre législatif et réglementaire en vigueur, permettrait aux chercheurs d'évaluer l'utilité des données en cause pour la recherche et, dès lors, l'opportunité d'effectuer un travail d'agrégation, de mise en forme, et de documentation.

Si les données en cause s'avèrent, au final, dépourvues d'intérêt scientifique, aucun travail de documentation ne sera réalisé. Si, au contraire, les données méritent d'être exploitées à des fins de recherche scientifique, le chercheur pourra effectuer un travail de traitement et de documentation de ces données, en partenariat, le cas échéant, avec l'administration productrice. Ce travail fera nécessairement l'objet d'une validation par l'administration.

Au final, un tel accès exploratoire permettrait ainsi d'exploiter, à des fins de recherche, des données brutes, non documentées, en limitant la charge de travail de l'administration et, le cas échéant, en permettant de documenter de telles données, dans une logique mutuellement bénéfique.

Recommandation n° 10 : encourager, afin d'en évaluer l'intérêt pour la recherche, une première exploration des données brutes, simplement dé-identifiées.

3.3.2. *Les coûts associés à l'accès aux données*

Pour le comité du secret statistique, des saisines fréquentes de la part des administrations pourraient rapidement poser des questions de moyens. Il sera donc nécessaire de trouver l'organisation la plus à même de limiter la charge de l'ensemble des intervenants (producteur, chercheur, comité du secret statistique), en particulier pour les demandes les plus classiques.

⁶⁴ J. PENIN (coord.) et autres, *La valorisation des informations du secteur public (ISP) : un modèle économique de tarification optimale*, université de Strasbourg et CNRS (UMR 7522), décembre 2010.

L'article 36 de la loi pour une république numérique prévoit une possibilité de saisine supplémentaire du comité du secret statistique, ce qui va dans le sens d'un alourdissement de la charge pour ce comité et son secrétariat. Il est cependant trop tôt pour évaluer la charge supplémentaire associée. Celle-ci dépendra notamment du nombre d'administrations sollicitant le comité et du type de sollicitations (demandes standards ou non, sollicitations systématiques ou non...), et il y aura très certainement une période importante de montée en charge. Un bilan de ces demandes supplémentaires devra être réalisé, pour évaluer la charge supplémentaire qu'elles représentent et analyser les apports du passage au comité. Ces éléments devront permettre de faire évoluer le cas échéant l'organisation du comité (simplifications ou adaptation des procédures pratiques, moyens supplémentaires...).

La structure de coordination prévue par la recommandation n°4 devra également bénéficier du support (secrétariat, site web, etc.) et des budgets nécessaires à son fonctionnement. Cette structure ayant pour vocation d'organiser périodiquement une concertation, ou des travaux ponctuels, il n'est pas envisagé qu'elle dispose d'un personnel ou de moyens qui lui soient entièrement dédiés, mais elle doit cependant avoir des moyens réservés au sein de l'institution à laquelle elle sera rattachée.

Recommandation n° 11 : Faire un bilan de l'impact de l'ouverture des possibilités de saisine du Comité du secret statistique sur son fonctionnement et proposer les ajustements nécessaires (organisation pratique, moyens) et prévoir un budget de fonctionnement pour l'instance de concertation.

Un autre coût de l'accès aux données tient aux procédures d'accès sécurisé distants (comme le CASD) qui impliquent des moyens techniques et humains dans la garantie des protocoles de sécurisation des accès. Ces coûts d'accès distants sont souvent très inférieurs au coût d'accès *in situ* chez le producteur des données ou aux options de type centre physiques sécurisés (*secured data lab*) au sein des centres de recherche ou universités, expérimentés dans certains pays étrangers. Les accès sécurisés distants permettent en effet une utilisation conjointe des mêmes données par de nombreux chercheurs, qui peuvent être localisés dans les centres de recherche ou universités partout en France, voire même dans l'Union européenne⁶⁵.

De nombreux chercheurs ont en revanche critiqué l'importance des coûts exigés pour accéder à distance et de façon sécurisée à certaines données, plaidant pour leur prise en charge *via* une subvention publique. Il convient néanmoins de souligner que les coûts actuels du CASD restent nettement inférieurs à ce qui est pratiqué dans la plupart des pays européens ayant mis en place des dispositifs similaires. Demeure également la question du mode de financement de la subvention publique : la prise en charge doit-elle se faire au niveau de l'utilisateur final (le chercheur et donc à inclure dans les financements de projets) ou bien plus en amont au sein du budget du ministère chargé de la recherche avec des subventions allouées pour l'organisation de l'accès ?

Dans le cadre de la loi pour une République numérique, qui ouvre potentiellement l'accès des données administratives à des acteurs hors de la sphère de la recherche publique, il a semblé préférable au groupe de travail que les budgets d'accès aux données soient allouées directement aux chercheurs, qui demeurent les mieux à même à identifier l'intérêt pour la recherche scientifique des coûts d'accès à ces données.

⁶⁵ L'accès sécurisé distant permet ainsi à un moindre coût l'accès aux données administratives aux chercheurs ou universitaires non parisiens, facilitant ainsi une plus grande égalité d'accès à ces données sur le territoire tout en respectant des conditions strictes d'accès.

Recommandation n° 12 : les financements publics de la recherche doivent prendre en compte la nécessité de financer les coûts d'accès aux données administratives protégées par des secrets.

3.4. Améliorer les pratiques d'archivage

Il est essentiel, pour la qualité de la recherche scientifique, de pouvoir étudier les phénomènes sur une longue durée et d'avoir ainsi accès non seulement aux données les plus récentes mais également, le cas échéant, à des données relativement anciennes. A cet égard, l'archivage des données produites aujourd'hui constitue un enjeu essentiel pour la recherche scientifique future, et il est souhaitable que les chercheurs de demain soient en mesure d'utiliser les grandes bases de données actuellement produites.

Or, la nature particulière des grandes bases de données, produites à flux continu et en quantité croissante, renouvelle au moins en partie la question de leur archivage et pose des problèmes spécifiques sur lesquels il convient de revenir. L'ouverture à la recherche des données publiques conduit dès lors à s'intéresser aux conditions dans lesquelles ces données sont archivées par les administrations qui les produisent.

3.4.1. Archivage des données publiques : état du droit

3.4.1.1. Définition et champ d'application

Tout document administratif, au sens du livre III du CRPA, est également une archive publique au sens du Livre II du code du patrimoine. Ainsi, l'article L. 211-1 du code du patrimoine définit les archives comme « *l'ensemble des documents, y compris les données, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité* ».

Les archives publiques forment néanmoins un ensemble plus vaste que les documents administratifs dans la mesure où elles comprennent des documents exclus du champ d'application du livre III du CRPA, notamment les documents juridictionnels.

Lorsque les documents en cause portent sur des données à caractère personnel, tout traitement déclaré à la Cnil ou autorisé par elle doit indiquer leur durée de conservation. La loi de 1978 relative à l'informatique, aux fichiers et aux libertés ne fixe aucune durée de traitement, mais simplement un principe de conservation pendant la durée nécessaire à l'accomplissement de ses finalités (article 6 de la loi de 1978). C'est donc à l'administration responsable du traitement, sous le contrôle de la Cnil, voire à la Cnil elle-même, de définir les durées pertinentes, qui sont évidemment très variables selon les hypothèses (de quelques secondes à plusieurs années).

Classification réglementaire des archives publiques et délai d'utilité administrative

Les articles R. 212-10 à R. 212-12 du code du patrimoine distinguent les archives courantes, les archives intermédiaires et les archives définitives.

Les archives n'arrivent à l'âge définitif qu'au bout d'un certain délai (DUA), dont la fixation relève de la responsabilité commune du service producteur des données et de l'administration des archives.

Sont considérés comme **archives courantes** les documents qui sont d'utilisation habituelle pour l'activité des services, établissements et organismes qui les ont produits ou reçus. La durée de conservation de ces archives, très variable, est au moins de cinq ans. Ces archives, qui sont par définition conservées par les services producteurs, peuvent également être déposées auprès d'un service public d'archives. La question de la destruction de cette catégorie d'archives, sauf cas exceptionnels, ne se pose pas réellement, dans la mesure où elles sont le plus souvent intégrées au système d'information du service qui les détient.

Sont considérés comme **archives intermédiaires** les documents qui ont cessé d'être considérés comme archives courantes et ceux qui ne peuvent encore, en raison de leur intérêt administratif, faire l'objet de tri et d'élimination conformément à l'article R. 212-14 du code du patrimoine. Ces archives, qui correspondent à des fichiers encore « vivants » et susceptibles d'intéresser les chercheurs car portant sur une longue période, sont le plus souvent conservés par l'administration productrice, sous réserve des hypothèses où, portant sur des données à caractère personnel, le délai de conservation déclaré ou fixé par la Cnil a expiré et où l'administration concernée, ignorant la possibilité ouverte par l'article 36 de la loi de 1978 (cf. *infra*), estime ne pas avoir d'autre choix que de les éliminer.

Sont considérés comme **archives définitives** les documents qui ont subi les tris et éliminations définis aux articles R. 212-13 et R. 212-14 du code du patrimoine et qui sont à conserver sans limitation de durée. L'existence de ces documents, qui constituent des sources susceptibles d'être utilisées par la recherche scientifique de demain, dépend étroitement de la rigueur des pratiques administratives d'archivage.

3.4.1.2.Sélection, conservation et format des données à archiver

La sélection des documents à archiver ainsi que leur conservation relèvent des administrations productrices de ces données, sous le contrôle de la personne chargée du contrôle scientifique et technique de l'État sur les archives (article R. 212-10 à R. 212-12 du code du patrimoine).

Organisation du réseau des Archives

L'article L. 211-2 du code du patrimoine dispose que : « *La conservation des archives est organisée dans l'intérêt public tant pour les besoins de la gestion et de la justification des droits des personnes physiques ou morales, publiques ou privées, que pour la documentation historique de la recherche* ».

Les documents qui constituent des archives au sens du code du patrimoine sont conservés dans de nombreuses institutions et organismes, sur lesquels (sauf s'agissant des archives des ministères des Affaires étrangères et de la Défense) le Service interministériel des archives de France (SIAF) exerce un contrôle scientifique et technique.

Les Archives nationales sont constituées de trois services à compétence nationale :

- i) les sites, généralistes, de Paris, Fontainebleau et Pierrefitte-sur-Seine ;
- ii) les Archives nationales d'Outre-mer, situées à Aix-en-Provence ;
- iii) les Archives nationales du monde du travail, situées à Roubaix.

Les collectivités territoriales et les groupements de collectivités sont propriétaires de leurs archives, dont ils assurent eux-mêmes la conservation et la mise en valeur (articles L. 212-6 et L. 212-6-1 du code du patrimoine). Ils peuvent, dans certaines conditions fixées par la loi, confier cette tâche au service d'archives d'une autre collectivité territoriale. L'article 212-11 prévoit en outre que les archives anciennes des communes de moins de 2000 habitants doivent être déposées aux Archives départementales, sauf dérogation accordée par le préfet de département.

Un certain nombre d'organismes publics sont autorisés, à titre dérogatoire (article L. 212-4 du code du patrimoine), à conserver eux-mêmes leurs archives définitives, à condition de les traiter et de les communiquer au public conformément aux règles en vigueur dans les services d'archives publics.

Au sein de l'administration centrale de chaque ministère, à l'exception des Affaires étrangères et de la Défense, sont implantées des missions des Archives avec à leur tête un agent du Service interministériel des archives de France. Les ministères mettent à leur disposition du personnel, des locaux et des moyens de fonctionnement. Pour certaines administrations, la mission est également chargée de la gestion des archives intermédiaires (pré-archivage). C'est par l'intermédiaire de ces missions que les archives définitives des organes centraux de l'État sont versées aux Archives nationales.

La première étape consiste, pour les administrations productrices, à sélectionner, à l'expiration de leur période d'utilisation courante, les documents qui méritent d'être conservés en raison de l'utilité administrative ou de l'intérêt historique ou scientifique qu'ils présentent (article L. 212-2 du CRPA).

Cette sélection des documents qui ont vocation à être conservés se fait toujours en concertation avec l'administration des archives.

En vertu des articles 36 de cette loi⁶⁶ et L. 212-3 du code du patrimoine⁶⁷, une telle concertation a également lieu afin de sélectionner les données à caractère personnel collectées dans le cadre de traitements régis par la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Ces dispositions, trop souvent ignorées par les administrations, prévoient ainsi la possibilité de verser aux archives, après l'expiration de la période de conservation déclarée ou fixée par la Cnil, des données à caractère personnel.

L'élimination des documents est faite sous le contrôle scientifique et technique du service interministériel des archives de France, dont le visa est indispensable avant toute élimination (article R. 212-14 du code du patrimoine).

La seconde étape porte sur la conservation proprement dite des documents ainsi sélectionnés.

L'article L. 212-4 du code du patrimoine prévoit ainsi que ces documents sont en principe versés dans un service public d'archives. Ces services sont définis au titre Ier du livre II de la partie réglementaire du code du patrimoine : ainsi, les archives des ministères sont versées aux Archives nationales, à l'exception des ministères des Affaires étrangères et de la Défense qui disposent de leurs services propres. Précisons que les dispositions de l'article L. 212-4 du code du patrimoine ne s'appliquent pas aux collectivités territoriales, qui disposent de leurs propres archives.

Par dérogation, l'article L. 212-4 du code du patrimoine permet aux administrations productrices de données de conserver elles-mêmes les données qu'elles produisent, si ces administrations présentent des conditions satisfaisantes de conservation, de sécurité, de communication et d'accès des documents. Une telle dérogation est, en vertu de l'article R. 212-12 du code du patrimoine, « *subordonnée à la signature d'une convention entre l'administration des archives et le service ou l'organisme intéressé, qui prévoit les conditions de gestion, de conservation et de communication au public des archives, les prescriptions scientifiques et techniques du service interministériel des archives de France de la direction générale des patrimoines qui s'y appliquent et l'emploi d'une personne responsable qualifiée en archivistique* ».

Se pose en dernier lieu la question du format sous lequel les données doivent être conservées. Afin de faciliter la collecte et la pérennisation des données archivées, les Archives

⁶⁶ Aux termes de l'article 36 de la loi du 6 janvier 1978 : « *Les données à caractère personnel ne peuvent être conservées au-delà de la durée prévue au 5° de l'article 6 qu'en vue d'être traitées à des fins historiques, statistiques ou scientifiques ; le choix des données ainsi conservées est opéré dans les conditions prévues à l'article L. 212-3 du code du patrimoine. / Les traitements dont la finalité se limite à assurer la conservation à long terme de documents d'archives dans le cadre du livre II du même code sont dispensés des formalités préalables à la mise en œuvre des traitements prévues au chapitre IV de la présente loi. / Il peut être procédé à un traitement ayant des finalités autres que celles mentionnées au premier alinéa : / -soit avec l'accord exprès de la personne concernée ou en vertu de ses directives, formulées dans les conditions définies à l'article 40-1 ; / -soit avec l'autorisation de la Commission nationale de l'informatique et des libertés ; / -soit dans les conditions prévues au 8° du II et au IV de l'article 8 s'agissant de données mentionnées au I de ce même article. »*

⁶⁷ Aux termes de l'article L. 212-3 du code du patrimoine : « *Lorsque les archives publiques comportent des données à caractère personnel collectées dans le cadre de traitements régis par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, ces données font l'objet, à l'expiration de la durée prévue au 5° de l'article 6 de ladite loi, d'une sélection pour déterminer les données destinées à être conservées et celles, dépourvues d'utilité administrative ou d'intérêt scientifique, statistique ou historique, destinées à être éliminées. / Les catégories de données destinées à l'élimination ainsi que les conditions de cette élimination sont fixées par accord entre l'autorité qui a produit ou reçu ces données et l'administration des archives* ».

de France préconisent de suivre les recommandations du référentiel général d'interopérabilité (RGI) édité par le secrétariat général pour la modernisation de l'action publique (SGMAP)⁶⁸. Un certain nombre de formats sont ainsi recommandés pour archiver les bases de données dont, notamment, le format SIARD (*Software Independent Archiving of Relational Databases*), mis au point par les Archives fédérales suisses et qui rend possible la conservation du modèle conceptuel de la base et des données qu'elle contient, ainsi que le format d'échanges SEDA (Standard d'Échange de Données pour l'Archivage), développé par les Archives de France, qui permet de décrire à la fois les transactions entre les systèmes et la forme des messages échangés pour le transport des archives et de leurs métadonnées.

3.4.2. Difficultés pratiques

L'importance croissante prise par les grandes bases de données, dont l'intérêt pour la recherche scientifique n'est plus à démontrer, impose une réflexion renouvelée sur les critères de sélection des documents pertinents ainsi que sur les modalités pratiques de leur conservation.

3.4.2.1. État des lieux des difficultés rencontrées

En 2015, les Archives nationales ont réalisé un panorama des versements d'archives électroniques effectués entre 1983 et 2015. Durant cette période, les Archives nationales ont ainsi collecté 253 applications informatiques, 94 114 910 fichiers de bureautique générés par des traitements de texte, 94 382 photos numériques ou enregistrements sonores et audio numériques. L'analyse plus fine de cet inventaire permet de dresser un certain nombre de constats.

En premier lieu, il apparaît que les dépôts d'enquêtes, s'ils restent significatifs, représentent une part plus faible des archives collectées à partir de la fin des années 2000. Ce constat concerne les enquêtes réalisées par l'INED ou l'INSEE, mais également celles réalisées par les différents ministères.

En deuxième lieu, les données dites de gestion, qui, nous l'avons vu, sont susceptibles de présenter un grand intérêt pour la recherche scientifique, semblent faire rarement l'objet d'un versement aux Archives nationales. Cette situation est sans doute appelée à évoluer avec la mise en place du projet ADAMANT (Administration des Archives et de leurs Métadonnées aux Archives nationales dans le Temps), qui s'inscrit dans le cadre du programme VITAM (Valeurs Immatérielles Transférées aux Archives pour Mémoire), et devrait permettre à terme d'améliorer la collecte des archives définitives électroniques et notamment les grandes bases de données de gestion.

Enfin, des incertitudes demeurent quant au fait de savoir si les fichiers archivés l'ont été avec leurs identifiants directs (noms, prénoms, adresse, NIR, SIRET notamment) ou bien si ces informations ont été détruites (destruction qui peut avoir été demandée par la Cnil lors de la phase de réalisation des formalités préalables). La conservation de tels identifiants auxquels les chercheurs n'ont pas accès et n'ont pas besoin d'avoir accès dans le cadre des accès sécurisés, peut permettre des appariements. Si certains de ces appariements peuvent être effectués (via un tiers de confiance) à la demande de la statistique publique ou des chercheurs durant la période

⁶⁸ Dans sa dernière version, ce référentiel est annexé à l'arrêté du 20 avril 2016 portant approbation du référentiel général d'interopérabilité, JORF n° 0095 du 22 avril 2016.

d'utilité administrative, facilités par le nouveau cadre juridique issu de la loi pour une République numérique, d'autres pourraient être effectués dans le futur si les identifiants étaient conservés en vue d'une utilisation pour la recherche. Cette question devrait faire l'objet d'une réflexion spécifique tant en amont qu'au moment du dépôt des données pour l'archivage pérenne.

3.4.2.2.Facteurs d'explication

Deux raisons principales permettent d'expliquer ces difficultés.

D'une part, un certain nombre d'administrations ne sont pas nécessairement au fait de leurs obligations en matière d'archivage des données. Ces administrations sont alors conduites à conserver des données en interne ou à les éliminer sans consultation préalable du service des archives, voire sans s'interroger sur leur potentielle utilité pour la recherche scientifique (cette dernière hypothèse concerne essentiellement les données de gestion, peu ou pas documentées, qui sont pourtant susceptibles de présenter un intérêt pour la recherche scientifique). En outre, l'archivage est réalisé suivant des modalités techniques permettant une conservation longue mais rendant les données difficiles à exploiter avec des traitements de masse (ex : microfilms).

D'autre part, la destruction de certaines données peut s'expliquer par une lecture parcellaire des exigences posées par la loi de 1978 relative à l'informatique, aux fichiers et aux libertés. En effet, les administrations font-elles trop rarement usage de la possibilité qu'elles ont, en vertu des articles 36 de la loi de 1978 et L. 212-3 du code du patrimoine, de verser aux Archives des données à caractère personnel qui présentent un intérêt scientifique, statistique ou historique.

3.4.2.3.Pistes de réflexion

Cet état des lieux conduit à envisager un certain nombre de pistes afin d'améliorer les pratiques d'archivage des grandes bases de données.

En tout premier lieu, il convient de sensibiliser les administrations sur l'importance de l'archivage pérenne des données qu'elles détiennent. Il s'agit de favoriser la prise en compte de la problématique de l'archivage le plus en amont possible, notamment au stade de la demande d'autorisation Cnil lorsqu'il s'agit de bases de données à caractère personnel.

Recommandation n° 13 : sensibiliser les administrations sur la question de l'archivage pérenne des données qu'elles détiennent, en coordination avec l'administration des archives et conformément au code du patrimoine, notamment en leur rappelant que les données peuvent, en vertu de l'article 36 de la loi de 1978, être conservées à l'issue de la durée de conservation déclarée ou fixée par la Cnil à des fins scientifique, statistique ou historique.

Ensuite, les caractéristiques des grandes bases de données posent avec une acuité nouvelle la question des critères de sélection pour l'archivage de ces données.

Ainsi, s'agissant des bases de données encore vivantes, notamment les archives qualifiées par le code du patrimoine d'intermédiaires, la collecte des données consolidées peut s'effectuer à périodes régulières, par versements automatiques. Il pourrait en aller ainsi des bases administratives dont nombre sont alimentées de façon récurrente par de nouvelles

informations, riches de potentialités dans la mesure où elles permettent de disposer de données individuelles longitudinales qui intéressent particulièrement la recherche.

En outre, l'ouverture plus large à la recherche scientifique, *via* des accès sécurisés, de ces grandes bases de données va être à la source de nombreux sous-fichiers (avec dé-identification, dans certains cas extractions et éventuellement la création de fichiers anonymisés). Les chercheurs pourront également, nous l'avons vu, contribuer à enrichir les métadonnées des fichiers sur lesquels ils auront travaillé. Se pose dès lors la question des modalités d'archivage de ces fichiers, afin notamment d'éviter de devoir réitérer les opérations de dé-identification et/ ou d'extraction pour des projets de recherche ultérieurs et de pallier, le cas échéant, la destruction des fichiers sources chez le producteur.

Recommandation n° 14 : établir une coopération entre les archives et les centres sécurisés pour améliorer les modalités d'archivage des fichiers résultant des opérations d'appariements.

Il est également souhaitable, afin d'aider les administrations à identifier l'intérêt scientifique des données qu'elles détiennent, d'associer davantage le monde de la recherche au processus de sélection, en particulier les utilisateurs des données concernées. Un exemple intéressant est fourni par le cas du ministère de la justice où des chercheurs ont été associés à la sélection de la commission des archives de la justice, et conformément au cadre méthodologique pour l'évaluation, la sélection et l'échantillonnage des archives publiques⁶⁹.

Recommandation n° 15 : associer davantage les chercheurs de différentes disciplines au processus de sélection des données qui méritent d'être archivées.

Enfin, il importe de généraliser, dans le cadre de cette coordination, les opérations de recensement, incluant une veille permanente des fichiers existants et de leur utilisation. A cet égard, la méthode MOZART mise en œuvre par le ministère de la culture et de la communication, qui a permis de tenir davantage compte du cycle de vie des données et de renforcer le dialogue avec les experts archivistes, ainsi que l'effort de recensement actuellement conduite par la direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC) visant à cartographier les bases de données qui méritent d'être conservées, constituent des démarches qu'il convient d'encourager (cf. recommandation n° 5).

⁶⁹ Cf. <http://www.gouvernement.fr/cadre-methodologique-archives-france>.

4. CONCLUSION

Ce rapport a permis de faire un large tour d'horizon de la situation actuelle sur l'accès aux données administratives par les chercheurs. La remise à plat du droit s'est avérée essentielle à la bonne compréhension des situations diverses, selon les données en cause. Le constat général est celui d'un progrès notable de la France dans ce domaine depuis quelques années, constat qui mérite d'être souligné dans un contexte où les comparaisons internationales dans le domaine de la recherche sont loin d'être unanimement favorable à la France.

Ce constat positif ne doit pas faire oublier qu'il reste du chemin à parcourir pour exploiter au mieux ces données à des fins de recherche, et au profit de l'ensemble de la société française. Les obstacles tiennent essentiellement au morcellement de nos administrations publiques, et au manque de coordination des différents acteurs autour de la production de données administratives de qualité.

C'est pourquoi l'instance de concertation que nous appelons de nos vœux sera un rouage essentiel dans les progrès pratiques qui pourront être fait à l'avenir. Nous pourrions ainsi espérer que si un état des lieux devrait être fait dans 10 ans sur la situation de l'accès aux données administratives par les chercheurs, celui-ci puisse constater les progrès réalisés depuis.

ANNEXES

A. Lettre de mission



SECRETARIAT D'ETAT CHARGE DU NUMERIQUE

LA SECRETAIRE D'ETAT

Paris, le

Nos Réf. : Dem-C/2016/42514

Monsieur le Directeur,

Depuis une dizaine d'années, l'accès des chercheurs aux données des administrations publiques s'est fortement développé dans deux directions :

- d'une part, la publication en ligne de données anonymisées et agrégées qui peuvent ainsi être accessibles à toute personne (« open data ») ;

- d'autre part, des voies d'accès réservées aux chercheurs pour certaines données protégées par des secrets, qui ne peuvent donc être accessibles qu'à certaines personnes pour des finalités précises et sous des conditions élevées de confidentialité et de sécurité.

L'accès des chercheurs aux données publiques a été rendu possible par trois lois successives :

- pour les données détenues par le service statistique public, l'article 25 de la loi du 15 juillet 2008 relative aux archives ;

- pour les données fiscales, l'article 104 de la loi du 22 juillet 2013 sur l'enseignement supérieur et la recherche ;

- pour les données de santé, l'article 193 de la loi du 26 janvier 2016 de modernisation de notre système de santé.

L'accès des chercheurs aux données bancaires a, par ailleurs, été récemment réorganisé par la Banque de France.

L'accès des chercheurs à ces données confidentielles est subordonné à l'avis d'un comité d'expertise indépendant : pour les données statistiques et fiscales, il s'agit du comité du secret statistique prévu à l'article 6 bis de la loi du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. L'accès aux données statistiques et fiscales a également été facilité au plan technique par la création en 2011 du Centre d'accès sécurisé aux données (CASD) du Groupement des écoles nationales d'économie et de statistique. Pour les données de santé, un nouveau comité d'expertise ad hoc est prévu par la loi du 26 janvier 2016 ; la Banque de France a, de son côté, mis en place un comité qui examine les demandes.

Monsieur Pierre-Yves GEOFFARD
Directeur
PSE-Ecole d'Economie de Paris
48 boulevard Jourdan
75014 Paris

.../...



139 rue de Bercy - Télédéc 181 - 75572 Paris cedex 12

Le projet de loi pour une République numérique prévoit de compléter ces différentes dispositions sectorielles par une disposition similaire à celle qui existe pour les données statistiques et pour les données fiscales, mais de portée transversale : son actuel article 18 bis AB prévoit une procédure d'accès sécurisé aux données publiques à des fins de recherche ou d'étude présentant un caractère d'intérêt public, qui comprend la faculté de consulter le comité du secret statistique.

Dans ce contexte, je souhaite vous confier une mission pour préparer la mise en œuvre de ce nouveau dispositif. Cette mission prendra la forme d'un groupe de travail du Conseil national de l'information statistique (CNIS) consacré à l'accès sécurisé aux données publiques, dont je souhaite, comme nous en sommes convenus avec le CNIS, que vous exerciez la co-présidence et que vous me remettiez les conclusions finales.

Votre mission aura pour objectifs de préciser les conditions de réussite du nouveau dispositif prévu par le projet de loi pour une République numérique :

- au plan juridique, elle pourra nourrir la préparation du décret d'application du texte ;

- en termes d'organisation administrative, elle pourra préciser la bonne articulation à rechercher entre les administrations productrices des données, le comité du secret statistique et l'administration des archives. Elle pourra également recommander des évolutions quant au fonctionnement du comité du secret statistique, dont la réforme devrait à terme augmenter l'activité ;

- au plan pratique, elle pourra recenser les besoins des chercheurs et orienter le nouveau dispositif afin qu'il soit le plus facile d'usage pour la communauté scientifique ;

- elle pourra explorer les modalités d'un suivi annuel de l'accès des chercheurs aux données confidentielles par les différents réseaux d'accès organisés, afin que chacun connaisse les pratiques des autres, puisse s'inspirer des meilleures pratiques et que les difficultés éventuelles soient bien analysées ;

- elle s'attachera enfin à mener un travail de comparaison internationale, en étudiant la situation des pays les plus avancés en matière d'accès sécurisé des chercheurs aux données publiques.

Votre mission portera une attention particulière aux données détenues par les organismes du champ de la protection sociale qui sont incluses dans le champ du nouveau dispositif (caisses de sécurité sociale, organismes de retraites complémentaires, complémentaires santé, etc.) Elle prendra également en considération les données détenues par les collectivités locales.

Pour préparer ses recommandations, la mission pourra s'appuyer, d'une part, sur un premier bilan des deux réformes législatives précitées portant sur l'accès aux données fiscales ou détenues par le service statistique public et, d'autre part, sur les travaux menés actuellement par le ministère de la santé pour la mise en œuvre de la loi du 26 janvier 2016 et la création d'une voie d'accès spécifique aux données de santé.

.../...

Votre mission pourra s'appuyer sur la Direction interministérielle du numérique et du système d'information et de communication de l'Etat ainsi que sur le Service interministériel des Archives de France. Elle mènera ses travaux en lien étroit avec le Comité du secret statistique et son président. Elle pourra solliciter le réseau des missions économiques de la Direction générale du Trésor afin d'établir des éléments de comparaison internationale.

Je souhaite que vous me remettiez ses conclusions pour le 1^{er} janvier 2017, après un point d'étape au 15 octobre 2016.

Je vous prie de croire, Monsieur le Directeur, à l'assurance de ma considération distinguée.



Axelle LEMAIRE

B. Mandat du groupe de travail du Cnis



Mandat du groupe de travail

L'accès des chercheurs aux données selon les différentes dérogations et filières

Depuis une dizaine d'années en France, de nouvelles dispositions législatives ont ouvert l'accès des chercheurs, selon des filières différentes, à des catégories de données couvertes par des secrets professionnels légaux, notamment :

- pour les données détenues par le service statistique public, l'article 25 de la loi du 15 juillet 2008 relative aux archives¹ ;
- pour les données fiscales, l'article 104 de la loi du 22 juillet 2013 sur l'enseignement supérieur et la recherche ;
- pour les données de santé, l'article 193 de la loi du 26 janvier 2016 de modernisation de notre système de santé.

Par ailleurs, le projet de loi pour une République numérique prévoit une procédure d'accès sécurisé aux données publiques de portée transversale complétant ces dispositions (article 18bis AB²). Afin que le dispositif soit adapté aux demandes des chercheurs, il est utile d'engager une réflexion sur la mise en œuvre de cet article et, de manière plus générale, sur l'articulation des différentes filières et le suivi de leur fonctionnement.

L'ouverture des données étant un enjeu national et européen, cette réflexion tiendra compte de la réglementation et de l'expérience des autres pays - notamment les plus avancés dans l'ouverture des données aux chercheurs - afin d'en tirer si possible un guide de bonnes pratiques.

Conformément à ses orientations générales, le Cnis souhaite contribuer à faciliter, selon des modalités adaptées, l'accès des différents utilisateurs et notamment des chercheurs aux données de la statistique publique et plus largement aux données susceptibles d'être utilisées pour la recherche.

La secrétaire d'État au numérique ayant décidé de confier à Pierre-Yves Geoffard, directeur de l'École d'économie de Paris et à Antoine Bozio, directeur de l'Institut des politiques publiques, une mission qui poursuit les mêmes objectifs, il a été convenu qu'elle prenne la forme d'un groupe de travail du Cnis.

Ce groupe de travail aura pour missions de :

- participer à la concertation sur les textes réglementaires liés à l'article 18bis AB du projet de loi pour une République numérique (cf. article en annexe) ;
- rechercher de premiers critères permettant de caractériser les "traitements à des fins de recherche et d'étude présentant un caractère d'intérêt public", dans le cadre de l'article 18bis AB et de l'article L213-3 du code du patrimoine ;
- étudier les conditions techniques et de moyens permettant aux producteurs de données et au comité du secret statistique de répondre dans de bonnes conditions aux demandes d'accès. Cette réflexion pourra prévoir des préconisations de court terme et proposer des perspectives pour le moyen terme ;
- étudier les voies et moyens permettant de faire un bilan annuel des demandes d'accès dans les différentes filières et proposer les informations de suivi nécessaires pour établir ce bilan ;
- formuler des recommandations propres à encourager les échanges et l'articulation entre les différentes filières, à mutualiser les modèles de documentation et de supports fournis et à harmoniser au mieux les procédures mises en œuvre dans ces filières ;
- faire un état des lieux des bonnes pratiques au niveau international et des possibilités d'accès transnational au sein de l'espace européen.

Le groupe de travail rassemblera des représentants des chercheurs ayant une expérience des demandes d'accès dans différents secteurs, des producteurs de données de la statistique publique et les instances chargées d'expertiser les demandes des chercheurs et de leur donner accès aux données. Il pourra consulter et auditionner les organismes produisant des fichiers de gestion tels que des directions de ministère, des établissements publics (Pôle emploi, etc.) ou des caisses de sécurité sociale. Il veillera à échanger aussi souvent que nécessaire avec Jean Gaeremynck, président du comité du secret statistique.

Le groupe de travail débutera ses travaux en septembre 2016 et présentera son rapport ou, si certains points doivent faire l'objet d'un travail complémentaire, un état de ses travaux et de ses recommandations lors de la réunion du bureau de décembre 2016. Un point d'étape est prévu mi-octobre.

Son rapport sera remis au ministre de l'économie et des finances et à la secrétaire d'État au numérique et à l'innovation.

¹ L'article s'applique aux données concernant les ménages, celles concernant les entreprises relèvent d'une législation datant de 1984.

² Numérotation provisoire

RÉPUBLIQUE FRANÇAISE

Secrétariat général du Cnis : Timbre H030 - 18, Bd. A. Pinard 75675 PARIS Cedex 14 – Tél. : 01 41 17 52 62 – Fax : 01 41 17 55 41 – secretariat-general@cnis.fr – www.cnis.fr

C. Composition du groupe de travail

Présidents

M. Pierre-Yves GEOFFARD, professeur d'économie, directeur de l'École d'économie de Paris

M. Antoine BOZIO, maître de conférences à l'École des hautes études en sciences sociales (EHESS), directeur de l'Institut des politiques publiques

Rapporteurs

M. Clément MALVERTI, auditeur au Conseil d'État

Mme Manon PERRIERE, auditrice au Conseil d'État

Mme Pascale BREUIL, Directrice des statistiques, prospective et recherche, Caisse nationale d'assurance vieillesse (Cnav)

Membres

Mme Ketty ATTAL-TOUBERT, chef de la division "Exploitation des fichiers administratifs sur l'emploi et les revenus", Insee

Mme Muriel BARLET, adjointe à la sous-directrice de l'Observation de la santé et de l'assurance maladie, Direction de la recherche, des études, de l'évaluation et des statistiques (Drees), ministère des affaires sociales et de la santé

M. Jean-Charles BEDAGUE, Chef du « bureau des études et des partenariats scientifiques », Service interministériel des Archives de France (SIAF)

Mme Pascale BREUIL, Directrice des statistiques, prospective et recherche, Caisse nationale d'assurance vieillesse (Cnav)

M. Gunther CAPELLE-BLANCARD, professeur d'économie à l'Université Paris I Panthéon-Sorbonne

M. Martin DANTANT, Direction des affaires juridiques, Ministère de l'économie et des finances

M. Kamel GADOUCHE, directeur du Centre d'accès sécurisé à distance aux données (CASD)

M. Cyrille HAGNERE, agence centrale des organismes de sécurité sociale (Acoss)

M. Michel ISNARD, secrétaire du comité du secret statistique

M. Renaud LACROIX, directeur de l'ingénierie et de la maîtrise d'ouvrage statistiques, Banque de France

M. Jean MAÏA, Directeur des affaires juridiques, Ministère de l'économie et des finances

M. Jeanne MALLET, bureau des études et des partenariats scientifiques, SIAF

M. Thomas MENAND, chargé de mission au pôle juridique du service projets de la direction interministérielle pour la modernisation de l'action publique, secrétariat général pour la modernisation de l'action publique (SGMAP)

M. Bruno RICARD, sous-directeur de la communication et de la valorisation des archives, SIAF

Mme Valérie SERVICE TSETOU-LEBON, Direction des affaires juridiques, Ministère de l'économie et des finances

Mme Roxane SILBERMAN, directrice de recherche émérite au CNRS

M. Périca SUCEVIC, conseiller juridique du chef du service projets de la direction interministérielle pour la modernisation de l'action publique, secrétariat général pour la modernisation de l'action publique (SGMAP)

D. Personnes auditionnées, consultées ou rencontrées

Mme Chantal Cases, présidente de l'Institut des données de santé (IDS)

Mme Caroline Nisand, directrice adjointe à la Direction des affaires criminelles et des grâces au Ministère de la justice,

Mme Françoise Banat-Berger, directrice des archives nationales

Mme Marie Laperdrix, directrice du centre des archives économiques et financières, mission des archives au ministère de l'Économie et des Finances

M. Godefroy Beauvallet, vice-président du Conseil national du numérique (CNNum)

Mme Anne-Sophie Taillandier, directrice de la plateforme TeraLab, Institut Mines-Télécom

M. Pascal Buléon, directeur de recherche au CNRS, président de Progedo

Philippe Casella, directeur du développement de la recherche à l'EHESS

M. Bernard Tapie, directeur des statistiques de la Cnaf

M. Yves Denéchère, professeur d'histoire contemporaine à l'université d'Angers

M. Julien Grenet, directeur adjoint de l'Institut des Politiques publiques (IPP), professeur associé à l'École d'Économie de Paris (PSE)

La Cnil a également été sollicitée pour une audition (sans suite).

E. Organismes consultés via un questionnaire

Association générale des institutions de retraite des cadres (AGIRC)

Association pour le régime de retraite complémentaire des salariés (ARRCO)

Caisse des dépôts et consignation, direction des retraites et de la solidarité

Direction des affaires civiles et du sceau (DACS)

Direction générale des finances publiques (DGFIP), sous-direction CE1, comptabilité de l'État

Direction générale des finances publiques (DGFIP), sous-direction GF3, missions foncières, fiscalité du patrimoine et statistiques

Direction générale des finances publiques (DGFIP), service des retraites de l'État (SRE)

Direction de la protection judiciaire de la jeunesse (DPJJ)

F. Bibliographie sélective

D. MANDELKERN et B. du MARAIS, *Diffusion des données publiques et révolution numérique*, Commissariat général du plan, La Documentation française, 1999

B. ORY-LAVOLLEE, *La diffusion numérique du patrimoine, dimension de la politique culturelle*, rapport à la ministre de la culture et de la communication, 2002

J. PENIN (coord.) et autres, *La valorisation des informations du secteur public (ISP) : un modèle économique de tarification optimale*, université de Strasbourg et CNRS, 2010

R. LACOMBE, P.-H. BERTIN, F. VAUGLIN et A. VIEILLEFOSSE, *Pour une politique ambitieuse des données publiques. Les données publiques au service de l'innovation et de la transparence*, rapport au ministre de l'industrie, de l'énergie et de l'économie numérique, 2011

S. CHIGNARD, *L'Open data, comprendre l'ouverture des données publiques*, FYP éd., 2012

CONSEIL D'ETAT, *Le patrimoine immatériel des personnes publiques*, actes du colloque organisé par le Conseil d'État le 16 mars 2012 à l'École nationale d'administration, La Documentation française, 2013

M. A. TROJETTE, *Ouverture des données publiques. Les exceptions au principe de gratuité sont-elles toutes légitimes ?*, rapport au Premier ministre, 2013

CONSEIL D'ETAT, *Le numérique et les droits fondamentaux*, Étude annuelle, La Documentation française, 2014

C. DUCHESNE, L. CYTERMANN, L. VACHEY, M. MOREL et T. AUREAU, *Rapport relatif aux données d'intérêt général*, rapport pour le ministre de l'économie, de l'industrie et du numérique, 2015

G. CAPELLE-BLANCARD, R. BELLANDO et R. LACROIX, *L'accès aux données bancaires et financières : une mission de service public*, rapport du groupe de travail du CNIS, 2015

L. VACHEY, C. DUCHESNE, M. MEYER, L. RUAT, G. TIROT, C. OLSINA et M. PERRIERE, *Les données d'intérêt général. Phase 2*, rapport pour le ministre de l'économie, de l'industrie et du numérique, 2016