

Paris, le 30 mars 2017 - N° 024/H030

Groupe de travail du Cnis

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

Relevé de décisions de la 5^{ème} réunion du 16 novembre 2016

Étaient présents à cette cinquième réunion du groupe de travail :

Présidents :

- Pierre-Yves Geoffard, professeur d'économie, directeur de l'École d'économie de Paris
- Antoine Bozio, professeur d'économie, directeur de l'Institut des politiques publiques

Rapporteurs :

- Pascale Breuil, Directrice des statistiques, prospective et recherche, Caisse nationale d'assurance vieillesse (Cnav)
- Clément Malverti, auditeur au Conseil d'État
- Manon Perrière, auditrice au Conseil d'État

Membres :

- 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
- Jean-Charles Bédague, chef du « bureau des études et des partenariats scientifiques », Service interministériel des Archives de France (SIAF)
- Élise Coudin, chef de la division "Salaires et revenus d'activité", Insee (en remplacement de Kitty Attal-Toubert)
- Martin Dantant, Direction des affaires juridiques, Ministère de l'économie et des finances
- Kamel Gadouche, directeur du Centre d'accès sécurisé à distance aux données (CASD)
- Cyrille Hagneré, agence centrale des organismes de sécurité sociale (Acosse)
- Michel Isnard, secrétaire du comité du secret statistique
- Valérie Tsétou-Lebon, Direction des affaires juridiques, Ministère de l'économie et des finances
- Bruno Ricard, sous-directeur de la communication et de la valorisation des archives, SIAF
- Roxane Silberman, CNRS

Membres excusés

- Gunther Capelle-Blancard, professeur d'économie à l'Université Paris I Panthéon-Sorbonne
- Renaud Lacroix, directeur de l'ingénierie et de la maîtrise d'ouvrage statistiques, Banque de France
- Jean Maïa, Directeur des affaires juridiques, Ministère de l'économie et des finances
- Jeanne Mallet, bureau des études et des partenariats scientifiques, SIAF
- Thomas Ménand, 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)
- 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)

Secrétariat général du Cnis : Dominique Allain, secrétaire générale adjointe du Cnis (excusée)

1. Cette réunion a débuté par l'audition de Mme Françoise Banat-Berger, directrice des archives nationales, et de Mme Marie Laperdrix, cheffe du centre des archives économiques et financières du Ministère de l'Économie et des Finances.

Mme Banat-Berger a présenté les missions des archives nationales.

Les archives nationales sont un service à compétence nationale depuis 2007. Elles ont pour mission de conserver et de communiquer les archives des services centraux de l'État (et des opérateurs pour lesquelles elles sont responsables de l'archivage) à l'exception, pour des raisons historiques, des archives relevant du ministère des affaires étrangères et du ministère de la défense. Elles gèrent des archives publiques mais également des archives privées. Leur système d'information a été profondément rénové. Une salle des inventaires virtuelle a été mise à disposition sur internet, avec notamment des archives numérisées. Au départ, les archives numériques étaient surtout constituées de bases de données structurées (bases de l'Insee et de l'Ined principalement). Les nouvelles archives numériques sont principalement constituées d'archives audiovisuelles ou mails. Jusqu'en 2009, la législation ne permettait pas l'accès aux bases de données numériques contenant des informations personnelles. L'accès aux bases de données intéresse un public très divers (chercheurs, citoyens à la recherche de leurs droits, journalistes, magistrats, généalogistes...).

L'archivage électronique est porté dans les programmes de travail de l'État depuis 2006. Des travaux ont permis de définir les standards d'archivages de données, en lien avec la Caisse des dépôts puis la Dinsic (direction interministérielle du numérique et du système d'information et de communication de l'État). Une méthodologie d'archivage a été définie avec l'Insee. Cela a permis de mettre en place un système de récupération de données extraites de grandes bases statistiques et a nécessité un investissement fort avec l'Insee sur la documentation de ces données. Depuis les années 70, ont été mises en place des missions des archives au sein des ministères, qui sont en relation avec les producteurs.

Un programme d'investissement d'avenir (programme Vitam) a été lancé, porté par trois ministères (culture, défense, affaires étrangères). Vitam est un logiciel libre, qui facilitera l'archivage, et qui sera interfacé avec le système d'information des archives (projet de refonte ADAMANT : Administration Des Archives et de leurs Métadonnées aux Archives nationales dans le Temps). La sélection des archives numériques à conserver pose cependant des questions spécifiques, ne serait-ce qu'en raison de leur volume.

Mme Laperdrix est responsable du centre des archives économiques et financières du Ministère de l'Économie et des Finances, dont elle présente les spécificités. Ce service intervient sur la gestion du cycle de vie des données. Le public demande des consultations simples, mais il y a aussi une forte demande de personnes qui veulent retraiter les données. La gestion du cycle des données est maintenant gérée dans la plupart des applications informatiques du Ministère de l'Économie et des Finances. Quand cela n'est pas prévu initialement, le coût d'adaptation est élevé. Se pose la question de l'obsolescence numérique du support de données, ainsi que la question de l'authenticité des données (quand il y a plusieurs versions notamment). Il est en effet important que les données ne soient pas altérées. Des données peuvent être versées sur un archivage numérique tout en restant disponibles pour le producteur.

2. Cette réunion s'est poursuivie par une audition sur TeraLab, représenté par M. Godefroy Beauvallet, Président du conseil national du numérique et Mme Anne-Sophie Taillandier, Directrice de la plateforme TeraLab, Institut Mines-Télécom.

Auparavant, les projets big data reposant sur des open data publiques pouvaient être distingués de ceux reposant sur des données très sensibles. Il y a maintenant une diversification des cas d'usage du big data, pouvant faire appel à ces deux types de données. Cela pose la question de la réidentification par rapprochement avec les données externes. Dans ce contexte, la notion d'i-Space se développe au niveau européen : un i-Space est un espace de confiance souverain orienté recherche permettant d'accueillir des données tout en maîtrisant l'ouverture. TeraLab est l'i-Space français, accueillant des données publiques ou privées. Il se compose du Centre d'Accès Sécurisé aux Données (CASD) pour les données soumises à secret et de l'institut Mines Télécom pour les autres.

3. Le projet Progedo a ensuite été présenté par M. Pascal Buléon, directeur de recherche au CNRS, président de Progedo, et M. Philippe Casella, Directeur du développement de la recherche à l'Ehess.

Progedo (PROduction et GESTion des DONnées) est une très grande infrastructure de recherche (TGir) chargée d'impulser et structurer une politique publique des données pour la recherche en sciences sociales en sciences humaines et sociales. L'objectif central de la politique scientifique de Progedo est le développement de la culture des données dans les communautés de recherche en sciences humaines et sociales en France, notamment auprès des étudiants et jeunes universitaires. Progedo est centré sur des données d'enquête. La France est très bien placée en matière de production de données, mais doit développer leur usage. Ce projet capitalise sur les travaux du centre Quételet, en y ajoutant un axe visant à développer l'usage des données. Progedo est impliqué dans des projets européens comme le CESSDA (Consortium of European Social Science Data Archives) ou dans des ERIC (European Research Infrastructure Consortium) consacrés à de grandes enquêtes européennes. Il s'agit d'obtenir des investissements financiers sur le long terme pour financer les enquêtes.

Le conseil scientifique de Progedo examine des questions proches de celles que se pose le groupe. Il traite notamment des conditions d'accès aux données, avec l'idée d'un continuum de conditions d'accès en fonction du type de données. Cela permet notamment pour développer la culture des données. Cela soulève la question de l'anonymisation des données.

4. Le groupe a ensuite auditionné M. Bernard Tapie, Directeur des Statistiques, des Études et de la Recherche à la Cnaf.

La Cnaf a rencontré des difficultés pour développer l'accès des chercheurs à ses données, en raison d'incertitudes sur le cadre juridique (risque de rupture du secret professionnel). L'article 36 règle désormais cette question, et va permettre à la Cnaf de mettre ses données à disposition. Un conseil scientifique sera créé au sein de la Cnaf, qui s'inspirera des critères du comité du secret statistique. La mise à disposition des données a un coût pour le producteur. Ce coût est élevé s'il faut concevoir une offre de données à mettre à disposition, documenter les données, présenter les données, accompagner les chercheurs. Il y a donc un risque d'éviction par rapport à d'autres besoins éventuellement prioritaires.

5. En raison de l'absence (excusée) de M. Renaud Lacroix, le point prévu à l'ordre du jour sur l'accès aux données du secteur bancaire et financier n'a pas été traité.

6. Enfin, le groupe de travail a discuté du contenu du futur rapport.