

CONSEIL NATIONAL DE L'INFORMATION STATISTIQUE

Division Environnement Juridique de la Statistique

RAPPORT DU GROUPE DE TRAVAIL

LE SECRET STATISTIQUE CONCERNANT LES ENTREPRISES :

SITUATION 2000 ET PERSPECTIVES D'EVOLUTION

Président : M. Yvonick RENARD, MEDEF

Rapporteurs : M. Raoul DEPOUTOT, INSEE
M. Gérard LANG, INSEE

JANVIER 2002

Conseil National de l'Information Statistique

LE SECRET STATISTIQUE CONCERNANT LES ENTREPRISES

SOMMAIRE

Page

I : - Propositions du groupe de travail	7
A - Informer	7
B - Assouplir	9
C - Anticiper	11
II : - Liste des annexes du rapport du groupe de travail	13
III : - Références	15
0 – Introduction	17
1 – Collecte de données individuelles par les personnes publiques	17
1.1. Légalité et légitimité des collectes publiques de données individuelles	17
(i) CNIS/CERFA	
(ii) Autres collectes publiques	
1.2. Collecte de données individuelles concernant les personnes physiques	18
CNIL	
1.3. Transparence et secret dans l'administration	18
1.3.1. CADA	
1.3.2. Archives	
1.3.3. Droit de communication fiscal et douanier	
2 – Le secret statistique	19
2.0. Nature du secret statistique	19
2.0.1. Secret et obligation	
2.0.2. Secret et déontologie	
2.1. Secret professionnel	19
2.1.1. Code pénal	
2.1.2. Statut général des fonctionnaires	
2.2. Secret statistique	20
2.2.1. Personnel/Institutionnel	
2.2.2. Collecté/Mobilisé	
2.2.3. Individuel/Agrégé	
2.2.4. Personnes physiques/Entreprises	
2.3. Secret statistique concernant les entreprises	21
2.3.1. Article 6	
2.3.2. Jurisprudence COCOES	
2.3.3. Rapport du BIPE	
2.3.4. Assouplissement CNS	
2.3.5. Finalités strictement interdites	
3 – Le comité du secret statistique concernant les entreprises	22
3.1. Composition et présidence	22
3.1.1. 1984	
3.1.2. 1997	
3.1.3. 1995	
3.2. Champ de compétence	22
3.2.1. Article 22 du décret	
3.2.2. Visa de l'article 2 de la loi	
3.2.3. Collectes CERFA	
3.2.4. Entreprises	
3.2.5. Exploitations agricoles	
3.2.6. Etablissements publics nationaux	
3.2.7. Collectivités territoriales et établissements publics locaux	
3.3. Jurisprudence générale	24
3.3.1. Règles applicables à la diffusion des données	
3.3.2. Activités secondaires	
3.3.3. Avis du CNIS du 3 juillet 1986	
3.3.4. Projet d'avis du CNIS du 10 mars 1998	
3.3.5. Accord entre le SES et la FNPC	
3.4. Habilitations d'accès à des données individuelles	25
3.4.1. Permanentes : services statistiques ministériels	
3.4.2. Temporaires : Fichiers déjà constitués ; Finalité spécifique en faveur d'une équipe (pseudo personne morale) et en son sein liste de personnes physiques fixée	

4 – Evolution de l’environnement juridique	26
4.1. Droit communautaire et européen	26
4.1.1. Règlement de 1990	
4.1.2. Directive 1995	
4.1.3. Règlement de 1997	
4.1.4. Recommandation de 1997	
4.1.5. Règlement du 18 décembre 2000	
4.1.6. Projet de règlement Eurostat	
4.2. Droit français	28
4.2.1. Transposition	
4.2.2. Loi DCRA et conciliation des lois CNIL, CADA et Archives	
4.2.3. Comité de concertation pour les données en sciences humaines et sociales	
4.2.4. Projet de modification de la loi Archives	
5 – Améliorations envisagées au niveau des règles de fonctionnement du Comité	29
5.1. La loi du 7 juin 1951	29
5.2. Le décret	29
5.2.1. Composition du comité	
5.2.2. Création d'un bureau permanent	
5.2.3. Procédure écrite	
5.3. La pratique	30
5.3.1. Périodicité des réunions	
5.3.2. Avis ; Recommandations ; Opinions	
5.3.3. Participation des demandeurs à l'ensemble de la réunion	
5.3.4. Compte-rendu des réunions et accès aux décisions	
6 – Améliorations envisagées au niveau de la jurisprudence du comité	31
6.1. Dérégations générales	31
6.1.1. Paragraphe 2 de l'article 13 du règlement de 1997 : connu publiquement	
6.1.2. Données publiquement accessibles et agrégats statistiques	
6.1.3. Données publiquement accessibles et secret des données individuelles	
6.1.4. Possibilité de suppression du droit d'opposition à la diffusion	
6.1.5. Collectes et variables spécifiques et comité du label	
6.1.6. Interprétation de l'avis du 3 juillet 1986	
6.1.7. Effectifs et effet sur SIRENE	
6.1.8. Activités secondaires et effet sur SIRENE	
6.1.9. Chiffre d'affaires : modifier les tranches, passage à l'Euro, SIRENE	
6.1.10. Part du chiffre d'affaires à l'exportation et SIRENE	
6.1.11. Indication d'une activité de recherche et SIRENE	
6.1.12. Enquête LIFI : mettre en œuvre la recommandation du comité du secret du 10 mars 1998	
6.1.13. Indication de l'appartenance à un groupe : afficher un code binaire ; SIRENE	
6.2. Habilitations permanentes	34
6.2.1. Cas des SSM	
6.2.2. Autres cas	
6.3. Habilitations temporaires	34
6.3.1. Extension de finalité d'une habilitation	
6.3.2. Fichiers non disponibles	
6.3.3. Délai : notamment en vue du projet de règlement d'Eurostat	
6.4. Enseignement et recherche	35
6.4.0. Information des chercheurs	
6.4.1. Représentant de la recherche au Comité du secret statistique concernant les entreprises	
6.4.2. Enseignement (cas général)	
6.4.3. Enseignement et ancienneté des données dont le format est particulièrement original	
6.4.4. "Arbitre " pour les publications scientifiques	
6.4.5. Comité d'éthique	
6.4.6. Remplacement d'enquêtes par des sources administratives	
6.5. Information statistique et régulation de la concurrence	36

7 – Méthodes pour protéger les données confidentielles	37
7.1. Protection de la confidentialité statistique lors de la diffusion de données agrégées	37
7.1.1. La suppression pure et simple	
7.1.2. La réduction de finesse des ventilations	
7.1.3. Les arrondis	
7.1.4. Les perturbations aléatoires	
7.2. Protection de la confidentialité statistique lors de la diffusion de données individuelles.....	38
7.2.1. L'anonymisation	
7.2.2. La réduction de finesse	
7.2.3. La suppression	
7.2.4. La perturbation	
7.2.5. La permutation	
7.2.6. La micro-agrégation	
7.3. Logiciels de protection des données confidentielles.....	40
8 – Conclusion	41
Annexes (1 à 27)	43

I : PROPOSITIONS DU GROUPE DE TRAVAIL

A - INFORMER

Proposition A1 (Introduction, point 6.4.0. et Conclusion)

Mieux faire connaître, notamment au sein de la communauté scientifique, l'existence du Comité du secret statistique concernant les entreprises et les possibilités d'accès aux données individuelles collectées par la statistique publique.

Mieux faire connaître les résultats des analyses rendues possibles, en application des avis favorables du comité du secret statistique concernant les entreprises, en archivant systématiquement les retours émanant des organismes habilités.

Proposition A2 (Points 5.2.1., 6.4.1. et 6.4.5.)

Introduire un représentant porteur des considérations éthiques de la communauté scientifique, désigné après avis du Comité de concertation pour les données en sciences humaines et sociales, dans la composition du Comité du secret statistique concernant les entreprises.

Proposition A3 (Point 5.3.2.)

Qualifier d' "avis du Comité" le résultat de l'examen de demandes d'accès aux données issues de collectes statistiques relevant de la loi de 1951, qu'elles soient obligatoires ou non.

Réserver la qualification de "recommandation du Comité" au résultat de l'examen des demandes d'accès à des données issues de collectes administratives non statistiques.

Proposition A4 (Point 5.3.4.)

Publier sur le site internet du CNIS un relevé des décisions d'habilitation faisant suite aux réunions du Comité du secret statistique concernant les entreprises.

Proposition A5 (Points 6.4.4. et 6.4.5.)

Tenir compte, dans l'examen des dossiers par le Comité du secret statistique concernant les entreprises, de :

- l'égalité des chances d'accès des chercheurs aux données ;
- la vérification ultérieure par un arbitre des analyses faites par un chercheur habilité.

B - ASSOULIR**Proposition B1 (Points 5.2.3. et 6.3.3.)**

Examiner au point de vue juridique la possibilité d'une procédure exceptionnelle écrite pour l'examen par le Comité du secret statistique concernant les entreprises de certaines questions qui exigent une réponse rapide, l'avis favorable ne pouvant être accordée qu'à l'unanimité en l'absence de délibération.

Cette possibilité pourrait couvrir, d'une part, les demandes d'accès des chercheurs aux données confidentielles transmises en urgence à Eurostat et, d'autre part, les cas où il s'agit seulement d'étendre l'habilitation concernant un dossier déjà délibéré pour permettre l'accès à une source d'une d'année qui avait été demandée mais n'était pas réputée disponible au moment de la délibération.

Proposition B2 (Point 6.1.2.)

Autoriser la diffusion de données agrégées dans lesquelles une entreprise a une place prépondérante (moins de trois unités ou forte prédominance d'une unité) dans les deux cas suivants :

- lorsque l'entreprise concernée a donné explicitement son accord pour ce faire au service enquêteur ;

- lorsque les données fournies par l'entreprise, dans sa réponse à l'enquête, sont identiques aux résultats publiés par l'entreprise du fait des directives comptables communautaires ou d'obligations concernant le secteur public, ou sont remplacées par ceux-ci.

Proposition B3 (Point 6.1.8.)

Autoriser la mise à jour dans SIRENE, en référence à la NAF, non seulement du code APE mais également de la liste des activités [principale et secondaire(s) significative(s)] des entreprises, à partir de collectes statistiques telles que l'EAE ou les enquêtes de branche.

Proposition B4 (Point 6.1.9.)

Modifier la stratification des tranches de chiffre d'affaires dont la diffusion est autorisée et libeller ces tranches en euro en tenant compte du règlement communautaire de 1993.

Proposition B5 (Point 6.1.12.)

Mettre en œuvre l'avis du Comité du secret statistique du 10 mai 1998 concernant l'accès restreint aux données de l'enquête "Liaisons Financières" en tenant à la disposition des entreprises-têtes le demandant expressément l'organigramme de leur "groupe statistique".

Proposition B6 (Point 6.4.6.)

Engager une réflexion sur les moyens permettant de maintenir la possibilité d'accès, lorsque des données antérieurement accessibles en tant que données statistiques voient leur statut transformé en données administratives.

C - ANTICIPER**Proposition C1 (Introduction et conclusion)**

Charger le secrétariat du Comité du secret d'une mission de veille permanente sur la réglementation et les pratiques en vigueur dans les Etats membres de l'Union européenne et de l'OCDE concernant la confidentialité des données statistiques relatives aux entreprises et l'accessibilité des données individuelles correspondantes.

Proposition C2 (Point 4.2.4.)

Agir avec détermination pour maintenir un délai de protection indispensable (25 ans minimum au lieu de 30 ans actuellement) dans la loi sur les archives, en ce qui concerne les données individuelles d'ordre économique ou financier collectées par la statistique publique.

Proposition C3 (Point 6.1.5.)

Etendre la possibilité de rendre diffusables (soit auprès du public général, soit auprès des publics restreints) certaines variables spécifiques collectées par des enquêtes statistiques particulières auprès d'entreprises spécifiques.

A titre expérimental, cette possibilité ne serait examinée par le Comité du secret statistique concernant les entreprises qu'au vu d'une proposition argumentée des configurations "entreprises" ou "agriculture" du Comité du label.

Proposition C4 (Point 6.4.2.)

Faire produire par les services enquêteurs, notamment à des fins didactiques, des fichiers de données individuelles présentant toutes les garanties d'anonymisation nécessaires.

Encourager les travaux de recherche sur les procédés d'anonymisation.

II : LISTE DES ANNEXES

- Annexe 1 : Mandat du groupe de travail
- Annexe 2 : Composition du groupe de travail
- Annexe 3 : Calendrier des réunions du groupe de travail
- Annexe 4 : Documentation du groupe de travail
- Annexe 5 : Loi n° 51-711 du 7 juin 1951 modifiée sur la coordination, l'obligation et le secret en matière de statistiques
- Annexe 6 : Décret n° 84-628 du 17 juillet 1984 modifié fixant les attributions, la composition et le fonctionnement du CNIS et portant application de la loi n° 51-711 du 7 juin 1951
- Annexe 7 : Arrêté de 1997 relatif au fonctionnement du comité du label
- Annexe 8 : Extrait du Code Pénal
- Annexe 9 : Règlement du Conseil du 11 juin 1990 relatif à la transmission à Eurostat d'informations statistiques couvertes par le secret
- Annexe 10 : Règlement du Conseil du 22 juillet 1993 relatif à la coordination communautaire du développement des répertoires d'entreprises utilisés à des fins statistiques
- Annexe 11 : Règlement du Conseil du 17 février 1997 sur la statistique communautaire
- Annexe 12 : Projet de règlement de la Commission de février 2001 sur l'accès des chercheurs aux données confidentielles transmises à Eurostat par les Etats membres
- Annexe 13 : Résolution sur les principes fondamentaux de la statistique officielle (adoptée par le Conseil Economique et Social de l'ONU sur proposition de sa Commission de statistique)
- Annexe 14 : Note du directeur général de l'INSEE du 13 juin 1980 sur l'assouplissement des règles d'application du secret statistique concernant les entreprises
- Annexe 15 : Avis de l'Assemblée plénière du CNIS du 3 juillet 1986 relatif à la diffusion de certaines données
- Annexe 16 : Décision du 13 septembre 1989 relative à la mise à jour du code APE (Activité Principale Exercée) de SIRENE à partir des résultats des enquêtes annuelles d'entreprise
- Annexe 17 : Projet d'avis du CNIS sur l'enquête Liaisons financières (LIFI) adopté le 10 mars 1998 par le comité du secret
- Annexe 18 : Liste des textes pertinents en matière d'éthique et de déontologie concernant la statistique et les études en dérivant
- Annexe 19 : Lettre du président du comité du secret au directeur des archives de France
- Annexe 20 : Extrait (chapitre III) du projet de loi sur la société de l'information
- Annexe 21 : Décret n° 2001-139 du 12 février 2001 portant création du comité de concertation pour les données en sciences humaines et sociales
- Annexe 22 : Dossier FILOCOM
- Annexe 23 : Dossier CADA concernant la collecte statistique sur le prix de l'eau effectuée auprès des communes par l'IFEN
- Annexe 24 : Proposition d'explicitation des pratiques légitimes en matière statistique
- Annexe 25 : Secret statistique et "secret de Polichinelle"
- Annexe 26 : Du bon usage de la statistique dans la régulation de la concurrence
- Annexe 27 : Tableaux relatifs aux décisions d'habilitations prises après avis du comité du secret statistique concernant les entreprises

III : REFERENCES

- 0 Le Secret
Que sais-je ? n° 3244
Y.-H. BONELLO
Presses universitaires de France (Janvier 1998)
- 1 La Transparence et le Secret
Commission des Etudes du Conseil d'Etat
in Rapport annuel du Conseil d'Etat pour 1995
Etudes et documents n° 47 (1996) La Documentation française
- 2 Le secret statistique : possibilités d'assouplissement
Bureau d'informations et de prévisions économiques (BIPE) – Mars 1976
- 3 L'utilisation des sources administratives
en démographie, sociologie et statistique sociale
INED – Paris – Septembre 2000
- 4 Règles de secret statistique applicables à la diffusion
et à la cession des données et textes de référence
A.-M. JAMES et G. LANG
Note n° D9802 bis (juillet 1998) de la DCSRI – INSEE - Paris
- 5 Loi n° 2000-321 du 12 avril 2000 relative aux droits des citoyens
dans leurs relations avec les administrations
JORF, Lois et décrets du 13 avril 2000
- 6 La double nature de SIRENE et sa diffusion (Version 2)
G. LANG
Note n° 987/D110 du 22 mai 2000 – INSEE - Paris
- 7 La diffusion des données des grands fichiers détenus par l'INSEE
G. LANG, F. CAZIN et C. MADINIER
Note n° 9507 (avril 1995) de la DCSRI – INSEE - Paris
- 8 Secret statistique et impossibilité d'identification indirecte :
Déontologie, Droit et Dérogation
G. LANG
Note n° 42/D110 du 8 mars 2000 – INSEE - Paris
- 9 Transposition en droit français de la directive européenne n° 95/46/CE
du 24 octobre 1995 : Les attentes des statisticiens
Rapport du CNIS n° 55, janvier 2000

- 10 Les sciences sociales et leurs données
Rapport au ministre de l'éducation nationale, de la recherche
et de la technologie, par R. SILBERMAN
La Documentation française – Paris – juin 1999
- 11 Diffusion des données publiques et révolution numérique
Rapport (MANDELKERN) rédigé pour le Commissariat Général du Plan (CGP)
par l'atelier "Des moyens nouveaux au service de la diffusion des données publiques" dans le
cadre de la mission "L'Etat et les technologies de l'information et de la communication" confiée
à M. LASSERRE au sein du CGP
B. du MARAIS et alii – Septembre 1999
- 12 Données personnelles et société de l'information
Transposition de droit français de la directive n° 95/46
Rapport au Premier ministre de G. BRAIBANT
La documentation française – Paris - 1998
- 13 Droit et Statistique
Numéro spécial n° 77 (mars 1996) du courrier des Statistiques
- 14 Le secret et les statistiques d'entreprises dans la communauté européenne
Rapport à l'office statistique des communautés européennes
G. ADER – Février 1976
- 15 Statistical confidentiality in the 15 member states of the european union
A critical comparison
G. ALS – Luxembourg
Third International Seminar on Statistical Confidentiality
Bled 1996 - Slovenia
- 16 Manual on disclosure control methods
Statistical document, Thème 9 (Recherche et développement), Série E (Méthodes)
EUROSTAT – Luxembourg - 1996
- 17 Statistical data protection
Proceedings of the conference
Lisboa, 25 to 27 March 1998
EUROSTAT – Luxembourg - 1999
- 18 La statistique
Rapport sur les sciences et la technologie n° 8
Académie des Sciences – Paris - 2000

0 - INTRODUCTION

Le droit français en vigueur (article 6 de la loi du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques ; articles 21 et 22 du décret n° 84-628 modifié du 17 juillet 1984) a mis en place un dispositif à la fois souple et rigoureux régissant le secret statistique concernant les entreprises, notamment en créant au sein du CNIS un comité spécialisé pour l'étude de ces questions. Ce dispositif permet, grâce à une procédure classique mais pleine de nuances, de procéder à l'étude de ces questions au sein d'un comité où l'ensemble des parties concernées peuvent s'exprimer par l'intermédiaire de leurs représentants qualifiés. Ce dispositif, spécifique au droit français et envié par nos partenaires au sein de l'Union européenne, est considéré comme globalement très satisfaisant.

Pour satisfaisant qu'il soit, ce dispositif demeure méconnu d'un grand nombre de ses utilisateurs potentiels (notamment du monde de la recherche) qui ignorent à la fois les possibilités ouvertes, l'existence et la pratique du comité du secret statistique concernant les entreprises. Il n'en demeure pas moins que l'équilibre ainsi instauré entre la transparence et le secret est certainement perfectible, et aussi (notamment en référence à l'évolution de l'environnement juridique) ajustable à certaines situations particulières ; les coûts de transaction (dans le langage des économistes) pourraient également être abaissés. C'est donc l'objet du présent rapport, en application du mandat confié par le bureau du CNIS au groupe de travail qui l'a rédigé, de proposer diverses adaptations de ce dispositif, après l'avoir examiné en détail et en avoir rappelé les nombreux avantages.

1 – COLLECTE DE DONNEES INDIVIDUELLES POUR LES PERSONNES PUBLIQUES DU DROIT ADMINISTRATIF

1.1. Légitimité et légalité des collectes publiques de données individuelles

En raison des pouvoirs de nature régalienne dont elles sont dotées par les lois en vigueur, les personnes publiques de droit administratif (et notamment en leur sein l'Etat et les collectivités territoriales) disposent, par nature, de capacités susceptibles de porter atteinte aux libertés individuelles.

C'est notamment le cas lorsqu'elles collectent auprès des acteurs de la société, qu'il s'agisse de personnes physiques (les citoyens) ou de personnes morales (comme les entreprises), de manière systématique des données individuelles relatives à ces acteurs.

Ainsi, en considération du fait qu'une telle activité porte toujours en germe un risque d'atteinte aux libertés individuelles, toute collecte de ce type doit être strictement encadrée. Deux catégories de collecte disposent dans cette ligne d'une assise juridique claire :

- les collectes statistiques publiques, inscrites annuellement au programme arrêté par le ministre chargé de l'économie, après avis du CNIS. Les questionnaires statistiques, qui supportent ces collectes faites en application de la loi du 7 juin 1951, portent le visa du CNIS qu'elles soient déclarées obligatoires ou simplement déclarées d'intérêt général ;
- les collectes administratives, faites en application d'une loi, d'un décret ou d'un arrêté qui les prévoit explicitement pour l'application du droit dans un domaine particulier ont pour support un formulaire administratif et portent le visa du CERFA (Centre d'Enregistrement et de Décision des Formulaires Administratifs).

Toute autre collecte d'une administration, sans être strictement illégale, ne dispose que d'un fondement juridique contestable et donc d'une légitimité douteuse.

1.2. Collecte de données individuelles concernant les personnes physiques

Les personnes physiques sont considérées comme particulièrement vulnérables par rapport à l'emploi des données individuelles les concernant, notamment lorsque ces données sont rassemblées en fichiers aptes au traitement informatique (et spécialement lorsque le NIR, numéro d'identification du répertoire national des personnes physiques figure dans ce fichier). La loi du 6 janvier 1978 modifiée sur l'informatique, les fichiers et les libertés a donc créé la CNIL (Commission nationale de l'informatique et des libertés), autorité administrative indépendante à l'avis de laquelle est subordonnée la mise en œuvre par l'administration de tout traitement de données individuelles réputé indirectement nominatif.

Les entreprises individuelles entrent donc dans le champ de compétence de la CNIL.

1.3. Transparence et secret dans l'administration

Plusieurs lois ont organisé, autour des années 1980, la transparence de l'administration française jugée opaque et régalienne. Outre la création du Médiateur de la République, la loi sur la motivation des actes administratifs et la création de la CNIL, deux autres lois ont été prises dans ce cadre et ont un lien direct avec la statistique publique.

1.3.1. Le titre I^{er} de la loi n° 78-753 du 17 juillet 1978 (aujourd'hui modifié) a institué le droit d'accès des citoyens aux documents administratifs, et crée la CADA (Commission d'Accès aux Documents Administratifs) chargée de garantir cette liberté d'accès. L'article 6 de cette loi prévoit toutefois que les administrations peuvent refuser l'accès aux documents administratifs protégés par les "secrets protégés par la loi". De ce fait (et la jurisprudence récente de la CADA le confirme, voir l'annexe 23), les données collectées dans le cadre de la loi du 7 juin 1951, qui sont protégées par le secret statistique, ne sont pas soumises au droit d'accès des documents administratifs.

1.3.2. La loi n° 79-18 du 3 janvier 1979 (modifiée) sur les archives a, quant à elle, directement modifié la loi de 1951 par son article 6 en précisant d'une part que "les recensements et enquêtes statistiques ont le caractère d'archives publiques" et qu'un délai de protection de "100 ans suivant la date de réalisation du recensement ou de l'enquête" s'applique aux "questionnaires ayant trait à la vie personnelle et familiale".

Au surplus, il résulte de la combinaison de l'article 7, 4° et du troisième alinéa de l'article 8 de la loi du 3 janvier 1979 que la protection de 100 ans ne comporte aucune dérogation possible, ce qui est exceptionnel (et d'ailleurs jugé exorbitant par les archivistes).

Les autres données collectées par la statistique publique (dont celles concernant les entreprises) bénéficient de la protection de base donnée par la loi de 1979, qui est d'une durée de 30 ans.

1.3.3. Un autre principe général de transparence (interne à l'administration) est le droit de communication dont disposent les administrations fiscales (art. L83 du livre des procédures fiscales) et douanières (art. L64A du code des douanes) qui organise un droit d'accès généralisé à tout document détenu par tout organe administratif en faveur des Impôts et des Douanes. Les seuls renseignements détenus par l'administration auxquels ce droit de communication ne peut s'appliquer sont ceux qui sont collectés en application de la loi du 7 juin 1951, donc protégés par le secret statistique (point 2 de l'article L64A du code des douanes, article L84 du livre des procédures fiscales).

2 – LE SECRET STATISTIQUE

2.0. Nature du secret statistique

2.0.1. Il est tentant de penser que le secret statistique est défini par la loi du 7 juin 1951, et notamment qu'il constitue la contrepartie de l'obligation de réponse aux enquêtes statistiques publiques.

En fait, la loi de 1951 ne définit pas le secret statistique, son titre se référant exactement au "secret en matière de statistiques". D'ailleurs depuis la réinterprétation de la loi de 1951, effectuée en accord avec le Conseil d'Etat à l'occasion de l'examen du décret n° 97-497 du 10 octobre 1997 qui a créé le comité du label (en insérant dans le décret du 17 juillet 1984 un article 12.1), le programme annuel des enquêtes statistiques publiques comprend d'une part des collectes obligatoires (et pour lesquelles le défaut de réponse peut entraîner l'émission d'une amende administrative, par décision du ministre prise après avis du comité du contentieux du CNIS) et d'autre part des collectes non obligatoires, qualifiées d'intérêt général par le comité du label. Un arrêté du 12 décembre 1997 a ensuite précisé les modalités du fonctionnement du comité du label des enquêtes statistiques. Il serait bien sûr absurde de priver les réponses aux enquêtes non obligatoires de la protection des données par le secret statistique.

2.0.2. Ce serait même le moyen le plus efficace d'abaisser massivement le taux de réponse aux enquêtes. En effet, fondamentalement, la protection par le secret statistique des données individuelles collectées pour les besoins de la statistique publique est une garantie de nature déontologique offerte aux répondants. La nature même de l'activité statistique exige de traiter des données individuelles, qui ne peuvent être obtenues qu'auprès des individus concernés eux-mêmes, ces données n'étant qu'une matière première permettant de fabriquer les agrégats qui sont l'objet de la production statistique et n'ayant vocation à être vues que par des statisticiens professionnels, et uniquement en vue de la production statistique, à l'exclusion de toute utilisation individuelle de ces données. C'est donc, avant tout, un engagement institutionnel de la statistique publique de protéger, par le secret statistique, les données confiées de tout détournement de finalité. C'est la contrepartie qui permet de maintenir des taux de réponse aux enquêtes suffisamment élevés pour assurer leur représentativité ainsi que la fiabilité et la qualité des analyses et des études qui en découlent.

2.1. Secret professionnel

2.1.1. En droit, le secret statistique est un concept du droit pénal. C'est l'application aux agents de la statistique publique de la notion de secret professionnel, définie par l'article L226-13 du code pénal.

Cet article fait partie de la section 4 ("De l'atteinte au secret") du chapitre VI ("Des atteintes à la personnalité") du Titre II ("Des atteintes à la personne humaine") du Livre II ("Des crimes et délits contre les personnes") du Code Pénal et dit :

"La révélation d'une information à caractère secret par une personne qui en est dépositaire soit par état ou par profession, soit en raison d'une fonction ou d'une mission temporaire, est punie d'un an d'emprisonnement et de 100 000 F. d'amende".

Ainsi c'est le 3^e alinéa de l'article 6 de la loi du 7 juin 1951 qui astreint les agents de la statistique publique au secret professionnel dans l'exercice de leur profession et fonde la protection des données collectées en application de la loi par le "secret statistique", de même que des textes particuliers à ces professions fondent le secret médical, le secret des notaires et des avocats, voire le secret professionnel du prêtre pour la confession. L'article L226-4 du code pénal ne prévoit que très peu de cas d'exception au secret professionnel.

2.1.2. Il faut d'ailleurs ajouter que l'article 26 du statut général des fonctionnaires dit :

"Les fonctionnaires sont tenus au secret professionnel dans le cadre de règles instituées dans le code pénal.

Les fonctionnaires doivent faire preuve de discrétion professionnelle pour tous les faits, informations ou documents dont ils ont connaissance dans l'exercice ou à l'occasion de l'exercice de leurs fonctions.....".

Ainsi la loi de 1951 ne fait que renforcer cette obligation des fonctionnaires de la statistique publique, tout en l'étendant aux non titulaires, aux agents des organismes professionnels et des sous-traitants pouvant intervenir dans les activités de la statistique publique.

2.2. Secret statistique

2.2.1. Même si le code pénal a récemment introduit la responsabilité pénale des personnes morales, le secret professionnel s'applique essentiellement aux personnes physiques qui, en tant qu'agent individuel de la statistique publique, sont susceptibles de divulguer les informations confidentielles. Cependant, nous avons déjà dit que la nature d'impératif déontologique du secret statistique, son aspect central dans la relation de confiance à établir entre enquêteur et enquêté, en fait un engagement très fort de la statistique publique au niveau institutionnel.

Le code de procédure pénale introduit les seules dérogations générales à l'obligation de secret statistique, qui sont de principe en droit français, et concernent les pouvoirs des officiers de police judiciaire commis par un juge d'instruction dans le cadre d'une instruction pénale.

2.2.2. Il va de soi (mais ceci est explicité par l'alinéa 2 de l'article 7 bis de la loi de 1951) que le secret statistique s'applique aussi bien aux données directement collectées par la statistique publique sur des questionnaires statistiques qu'aux données collectées initialement par des formulaires administratifs et mobilisées par elle en vue d'une finalité statistique en application de l'article 7 bis de la loi.

2.2.3. De même, il va de soi que sont protégées par le secret statistique non seulement les données individuelles mobilisées, mais aussi les données finement agrégées à partir desquelles on pourrait remonter aux données individuelles.

En outre, les conditions de publication de données agrégées résultant d'une collecte administrative mobilisée par la statistique publique doivent tenir compte, d'une part, des règles du secret statistique et, d'autre part, des règles pouvant résulter du secret (fiscal, ...) applicable au statut initial de ces données administratives.

2.2.4. Les deux premiers alinéas de l'article 6 instaurent une distinction très importante pour l'application du secret statistique. Le premier alinéa instaure un secret absolu s'agissant des "renseignements individuels ayant trait à la vie personnelle et familiale et, d'une manière générale, aux faits et comportements d'ordre privé", d'une durée de 100 ans. Ainsi, en pratique, les données individuelles relatives aux personnes physiques collectées par les enquêtes statistiques auprès des ménages sont soumises à un secret statistique absolu. Au surplus ces collectes entrent dans le champ de la loi "Informatique et Libertés" et les traitements correspondants sont soumis à l'avis de la CNIL. Cette circonstance permet d'ailleurs d'interpréter la loi de 1951 (qui date de l'époque "tout-papier") à la lumière des fichiers informatiques, en comprenant que "donnée individuelle" signifie désormais "donnée individuelle nominative, même indirectement, au sens de la CNIL" (C'est-à-dire que l'on ne peut considérer que les données que contient un fichier sont non nominatives que lorsqu'on en a éliminé suffisamment de "données identifiantes", spécialement de localisation en général, pour affirmer le caractère "non nominatif, même indirectement" de celles qui restent).

Ajoutons que la CNIL considère également que le concept de données indirectement nominatives embrasse, au-delà des données individuelles, celui des données agrégées portant sur de trop petits effectifs ou permettant l'établissement de "profils de groupe" permettant de stigmatiser certains sous-

groupes d'une population. Toutefois la proximité des concepts de données statistiques confidentielles au sens du CNIS et de données indirectement nominatives au sens de la CNIL ne conduit pas nécessairement aux mêmes conclusions concrètes. C'est ainsi que l'examen par la CNIL des conditions de diffusion des résultats du recensement général de la population de 1990 a révélé un véritable conflit culturel avec les statisticiens, qui n'a été finalement résolu que par une décision du Conseil d'Etat du 7 octobre 1998.

L'alinéa 2 de l'article 6 est beaucoup moins absolu et laisse une assez large place à l'interprétation et à la jurisprudence en disant :

"Les renseignements individuels d'ordre économique ou financier ne peuvent en aucun cas être utilisés à des fins de contrôle fiscal ou de répression économique".

Il appartient clairement au CNIS de fixer les règles de secret statistique s'appliquant à ce type de collecte qui, en pratique, se confond avec la statistique publique "d'entreprise".

2.3. Secret statistique concernant les entreprises

2.3.1. Dès la publication du décret n° 52-1059 du 15 septembre 1952, pris pour l'application de la loi du 7 juin 1951 et relatif notamment au comité de coordination des enquêtes statistiques (COCOES, premier ancêtre du CNIS) et à son comité du contentieux, il est dit (article 15) :

"Dans la limite où l'article 6, deuxième alinéa, de la loi du 7 juin 1951 en autorise la communication, les renseignements individuels visés audit article ne peuvent être communiqués que par décision concertée du ministre dont relève le service enquêteur et du ministre dont relève l'INSEE.

Les administrations bénéficiaires de ces communications ne peuvent elles-mêmes les communiquer à quiconque".

Une ouverture (restreinte aux administrations, si l'on s'en rapporte au dernier alinéa) est ainsi ménagée, à l'intérieur du secret statistique concernant les entreprises, et donne lieu au développement d'une jurisprudence.

2.3.2. Celle-ci concerne aussi les données finement agrégées, qui font l'objet de "règles pratiques de diffusion élaborées le 7 juillet 1960 par le COCOES" (règle des 4 unités ; règle des 85%).

2.3.3. Assez rapidement ensuite, on convient qu'il faut élargir les possibilités jurisprudentielles ouvertes par l'article 6 de la loi et le BIPE produit en mars 1976, à la demande du Conseil national de la statistique (CNS, successeur du COCOES, mis en place par le décret n° 72-1104 du 8 décembre 1972) un rapport intitulé : "le secret statistique : possibilités d'assouplissement", qui conduit à la décision du 13 juin 1980, suivant l'avis du CNS réuni en séance plénière le 6 juin 1980, "d'assouplissement des règles d'application du secret statistique concernant les entreprises" (règle les 3 unités au lieu de 4...).

2.3.4. L'étape suivante est franchie, avec l'accord de l'ensemble des parties intéressées : le décret n° 84-628 du 17 juillet 1984, qui met en place le Conseil national de l'information statistique (CNIS, successeur du CNS), crée en son sein (articles 21 et 22) un "comité du secret statistique concernant les entreprises, chargé d'étudier les problèmes que pose l'application des règles relatives à celui-ci".

2.3.5. Dans le langage actuellement en vigueur, la mission du comité du secret statistique concernant les entreprises consiste donc essentiellement à veiller à ce que les données confidentielles collectées auprès des entreprises par la statistique publique en application de la loi de 1951 ne puissent être utilisées :

- pour une finalité de contrôle fiscal ;
- pour une finalité de contrôle administratif par une autorité de régulation ;
- pour fausser la concurrence économique, dans le cadre national, européen ou mondial.

Ainsi, depuis sa mise en place en 1985, c'est désormais le comité du secret statistique qui est chargé de veiller à l'application de l'alinéa 2 de l'article 6 de la loi du 7 juin 1951.

3 – LE COMITE DU SECRET STATISTIQUE CONCERNANT LES ENTREPRISES

3.1. Composition et présidence

3.1.1. L'article 21 du décret de 1984 a initialement fixé la composition du comité du secret statistique à onze membres :

- le président, membre du Conseil d'Etat désigné par son vice-président ;
- quatre représentants de l'administration, dont un représentant de l'INSEE et un représentant du ministre de la justice d'une part, et d'autre part pour chaque problème examiné un représentant du ministre compétent et un représentant du service enquêteur concerné ;
- quatre représentants qualifiés des entreprises, désignés par leurs organes professionnels ;
- un représentant des organisations syndicales de salariés ;
- un représentant des utilisateurs régionaux et locaux de l'information statistique, désigné par le ministre de l'intérieur.

3.1.2. L'article 9 du décret n° 97-497 du 10 octobre 1997 a ensuite ajouté à la composition du comité du secret, désormais formé de quatorze membres, les membres du CNIS désignés respectivement par :

- l'Assemblée nationale ;
- le Sénat ;
- le Conseil économique et social.

Les présidents successifs du comité ont été jusqu'ici : M. QUERENET ONFROY de BREVILLE, M. le général LERCHE, M. FAURE, M. ARNAULT de LA MENARDIERE.

3.1.3. A la suite d'une suggestion de M. QUERENET, contenue dans sa lettre de démission, l'article 3 du décret n° 95-105 du 31 janvier 1995 a précisé que : "En cas d'absence ou d'empêchement du président, le comité du secret statistique tient séance sous la présidence du représentant du garde des sceaux, ministre de la justice", disposition fort utile.

3.2. Champ de compétence

3.2.1. L'article 22 du décret n° 84-628 dispose :

"Dans la mesure où l'article 6 de la loi précitée du 7 juin 1951 n'en interdit pas la communication, les renseignements individuels visés au 2° alinéa dudit article ne peuvent être communiqués que par décision conjointe du ministre dont relève le service enquêteur et du ministre dont relève l'INSEE, prise après avis du comité du secret statistique concernant les entreprises créé par l'article 21 du présent décret.

Ces communications ne peuvent être faites qu'à des bénéficiaires qui s'engagent à ne les communiquer à quiconque".

Cet article donne donc au comité du secret statistique concernant les entreprises compétence pour donner son avis sur l'ensemble des questions liées à l'interprétation et à l'exécution du 2° alinéa de l'article 6 de la loi du 7 juin 1951. Notons que le dernier alinéa de l'article 22 élargit considérablement la liste des "bénéficiaires" en ne les restreignant plus à la catégorie des "administrations bénéficiaires" comme le faisait l'article 15 du décret n° 52-109 du 15 septembre 1952 relatif au COCOES, repris à l'identique par l'article 15 du décret n° 72-1104 du 8 décembre 1972 relatif au CNS.

3.2.2. Une stricte lecture des textes en vigueur, résultant de la mention "du visa prévu à l'article 2" dans le 2^e alinéa de l'article 6 de la loi de 1951 pourrait conduire à la conclusion que la compétence du comité ne s'applique vraiment qu'aux questionnaires obligatoires. De fait, le comité distingue l'examen des questions relatives à des collectes obligatoires (qui donnent lieu à des avis) de l'examen des questions relatives à des collections facultatives (qui donnent lieu à des recommandations). Pour les raisons exposées au deuxième paragraphe du point 2.0. ci-dessus, liées à la réinterprétation de la loi de 1951 intervenue en 1997, il paraît indispensable de confirmer aujourd'hui l'entière compétence du comité du secret pour l'ensemble des collectes effectuées en application de la loi du 7 juin 1951, obligatoires ou non. Il paraîtrait, en effet, particulièrement inopportun de ne pas donner compétence au comité du secret s'agissant de la jurisprudence concernant le secret des données collectées auprès des entreprises dans le cadre des enquêtes statistiques publiques non obligatoires de la loi de 1951.

3.2.3. Le comité ne s'interdit d'ailleurs pas d'émettre des recommandations concernant des collectes n'appartenant pas (ou plus) à la loi de 1951 lorsque la demande qui lui est adressée comprend de telles données. C'est notamment le cas pour certaines collectes ayant le statut CERFA mais qui sont effectuées avec la collaboration de la statistique publique, comme l'ESE, les DMMO voire les DADS.

Par contre, le comité du secret statistique n'aurait aucune légitimité pour se risquer au-delà des limites de sa compétence et se prononcer sur des demandes d'accès relatives à des données administratives n'ayant jamais eu le statut statistique. Les demandes relatives à cette catégorie de données doivent être adressés directement à l'administration responsable de leur collecte.

En cas de refus de l'administration concernée, le demandeur peut se retourner vers la CADA pour qu'elle se prononce sur la légitimité de sa demande d'accès. Une réponse de la CADA favorable au demandeur n'oblige pas l'administration à céder ses données, mais offre au demandeur un argument d'un grand poids pour se retourner vers le contentieux administratif.

3.2.4. L'apparition du "concernant les entreprises" figurant dans la dénomination du comité du secret statistique conduit à se demander quel est son effet. Il n'existe pas, à vrai dire, de définition du concept "d'entreprise" dans le droit français. Le répertoire national des entreprises et de leurs établissements (SIRENE) qui, à sa création par le décret n° 73-314 du 14 mars 1973 ne comprenait initialement quasi-uniquement que des personnes physiques (entrepreneurs individuels) et des personnes morales de droit privé, a vu ensuite son champ étendu, par le décret n° 83-121 du 17 février 1983, à l'ensemble des personnes morales du droit public (dont les collectivités territoriales), ainsi qu'aux institutions et services de l'Etat.

3.2.5. Il est en tous cas clair aujourd'hui que la statistique publique considère les exploitations agricoles (même individuelles) comme des entreprises, tant pour le comité du secret que pour le comité contentieux.

3.2.6. Le cas des personnes morales de droit public est plus délicat. Il paraît difficile de les qualifier véritablement d'entreprises, mais elles ont de plus en plus de compétences et d'activités rentrant dans le champ de l'économie. On pourrait donc fort légitimement soutenir que lorsqu'une collecte statistique publique s'intéresse à des renseignements d'ordre économique ou financier relatifs à ces personnes morales (effectifs salarié, activité de gestionnaire d'infrastructure, prix de l'eau, existence d'équipements,...), l'alinéa 2 de l'article 6 de la loi de 1951 est applicable et que cela entraîne la compétence du comité du secret statistique concernant les entreprises. Tel est certainement le cas pour les établissements publics nationaux à caractère industriel ou commercial, qui sont de véritables entreprises inscrites au registre du commerce et des sociétés. C'est déjà plus discutable lorsqu'il s'agit d'établissements publics nationaux à caractère administratif, pour leurs activités ayant un caractère industriel et commercial, comme par exemple les chambres de commerce et d'industrie en tant que gestionnaires d'infrastructure portuaires et aéroportuaires.

3.2.7. Pour ce qui concerne les données relatives aux collectivités territoriales et aux établissements publics locaux, le comité du secret statistique concernant les entreprises s'est explicitement déclaré incompétent lorsqu'il a été saisi d'un dossier de demande d'accès à une collecte statistique sur les services d'eau et d'assainissement effectuée par l'IFEN et le SCEES auprès des communes, collecte pour laquelle la CADA s'appuyant sur son caractère statistique avait refusé une autre demande d'accès, en tant que données administratives couvertes par le secret.

Le groupe de travail est favorable à la prise en compte de l'avis d'un certain nombre d'experts, pour éclairer les décisions ministérielles relatives aux dérogations au secret statistique concernant les données économiques et financières collectées par la statistique publique auprès des collectivités territoriales et des établissements publics locaux prises en application de l'alinéa 2 de l'article 6 de la loi du 7 juin 1951. Il considère qu'une telle consultation doit être spécifique à chaque cas particulier et ne devrait pas déboucher sur la création d'un comité permanent.

3.3. Jurisprudence générale

3.3.1. La jurisprudence générale comprend d'abord des règles applicables à la diffusion des données. Les deux règles en vigueur :

- règle des trois unités ;
- règle des 85 % ;

ont été fixées initialement après avis du COCOES, puis modifiées (passage de 4 à 3 unités) après avis du CNS (formation système productif). Elles ont été explicitement rappelées dans le compte rendu de la première réunion du comité du secret statistique concernant les entreprises du vendredi 29 mars 1985.

3.3.2. Ce même compte rendu rappelle également la libre diffusion, "sauf opposition des entreprises concernées", des activités secondaires des entreprises, adoptée après l'avis du CNS dans sa séance plénière du 6 juin 1980.

3.3.3. L'assemblée plénière du 3 juillet 1986 du CNIS a ensuite adopté un avis concernant la possibilité de diffuser (sauf opposition des entreprises concernées) les [quatre] variables suivantes :

- effectifs global d'une entreprise et des établissements (une fois par an) ;
- catégorie d'importance du chiffre d'affaires ;
- catégorie d'importance de la part du chiffre d'affaires réalisée à l'exportation ;
- indicateur de l'exercice d'une activité de recherche.

Pour la mise en œuvre de cet avis, qui a été publié au bulletin officiel de l'administration Centrale de l'Economie, des Finances et de la Privatisation (fascicule 86-2), le CNIS a précisé le 31 décembre 1986 que le refus ou l'absence de refus d'une entreprise vaut pour les données correspondantes des années antérieures.

3.3.4. Le comité du secret statistique concernant les entreprises a adopté définitivement, lors de sa réunion du 10 mars 1998, un avis concernant la possibilité d'une diffusion restreinte aux intéressés de données (contour d'un groupe, en faveur de l'entreprise "tête de groupe" ; appartenance de l'entreprise à un groupe, en faveur de l'entreprise concernée et sauf opposition de la "tête de groupe") collectées par l'INSEE dans l'enquête "Liaisons financières (LIFI)".

L'assemblée plénière du CNIS n'a pas encore repris cet avis, dont la mise en œuvre n'est donc pas encore effective, du fait également de la suppression du bulletin officiel de l'administration centrale du MINEFI, qui aurait vocation pour la publier.

3.3.5. Le comité du secret statistique a approuvé un accord entre le ministère de l'Équipement et la Fédération nationale des promoteurs-constructeurs (FNPC) concernant la possibilité d'un accès à certaines données individuelles collectées par l'enquête sur la commercialisation des logements neufs (ECLN). Le comité a admis que les dossiers de demandes d'accès correspondants remplis par des organismes locaux tels que les agences d'urbanismes pouvaient être présentés directement par le service enquêteur lorsqu'il s'agit de renouveler annuellement ces habilitations.

3.4. Habilitations d'accès à des données individuelles

3.4.1. En application du principe de coordination contenu dans le titre de la loi du 7 janvier 1951, le COCOES, puis le CNS, avaient admis la possibilité de libre circulation des données statistiques individuelles collectées par un service statistique ministériel en faveur d'un autre service statistique ministériel intéressé par ces données (Remarquons que si le vocabulaire général de la loi de 1951 et du décret de 1984 utilise "service enquêteur", ou "service producteur", la loi du 26 décembre 1986, qui insère dans la loi de 1951 un article 7bis, y fait apparaître le concept de "service statistique ministériel", qui par contre n'apparaît toujours pas dans le décret de 1984 modifié).

Au cours de sa première réunion (29 mars 1985), le comité du secret a approuvé une cinquantaine d'avis confirmant des décisions antérieures de ce type qui, s'appliquant à des services statistiques ministériels, ont un caractère permanent (c'est-à-dire ne nécessitant pas de confirmation à l'occasion de chaque nouvelle collecte disponible concernant une enquête habilitée) tant que dure la reconnaissance par le comité du caractère de service statistique ministériel du demandeur. Au fil de ses réunions, le comité a pris d'autres décisions de ce type et la liste des organismes susceptibles de bénéficier d'une habilitation permanente a été élargie, notamment (mais pas exclusivement) en faveur de nouveaux services statistiques ministériels. Il y a actuellement 86 décisions de ce type en vigueur

Dans la mesure où un texte précis (loi ou décret) charge un organisme d'une mission de statistique publique qui nécessite évidemment l'accès à certaines données statistiques que cet organisme ne produit pas, le comité du secret a admis l'élargissement à un tel organisme (Banque de France), pour l'accomplissement de cette mission, de la possibilité d'une habilitation permanente.

3.4.2. Les autres habilitations d'accès à des données individuelles font l'objet d'une décision concertée du ministre dont relève l'INSEE et du ministre dont relève le service enquêteur, prise après avis (favorable ; il n'y a jamais eu de divergence à ce jour) du comité du secret statistique concernant les entreprises.

Ces habilitations sont prises en faveur d'un organisme (personne morale, ou quasi personne morale) ayant présenté au comité un dossier de demande, et au sein de cet organisme d'une liste nominative (mais révisable) de personnes physiques, chacune pénalement responsable en cas de divulgation des données confidentielles transmises.

Les fichiers de données ne sont transmis par les services enquêteurs concernés qu'après la signature de la décision concertée, qui fait elle-même suite à la réception par le secrétariat du comité du secret d'une lettre d'engagement au secret signée par chacune des personnes habilitées. Ces habilitations sont toujours temporaires, s'appliquant à des collectes statistiques précisément définies (y compris les années) qui doivent avoir été effectuées (et dont les fichiers doivent avoir été constitués par le service enquêteur) au jour de la réunion du comité du secret statuant sur cette demande. Les fichiers transmis ont, en principe, une durée d'utilisation et sont réputés avoir été détruits par l'organisme habilité au terme de cette durée. Les fichiers transmis sont, en principe, identifiés (SIREN, SIRET, raison sociale, adresse) sans aucune restriction, sauf disposition explicite contraire dans l'avis et dans la décision correspondante.

4 – EVOLUTION DE L'ENVIRONNEMENT JURIDIQUE

4.1. Droit communautaire et européen

4.1.1. Le règlement n° 1588/90 du conseil du 11 juin 1990 relatif à la transmission à l'Office Statistique des Communautés Européennes (OSCE, devenu EUROSTAT) d'informations statistiques couvertes par le secret a créé (article 7) un "comité [européen] du secret statistique". Ce comité, composé d'un représentant par Etat membre de la Communauté et présidé par un représentant de la Commission, est chargé d'examiner les mesures (proposées par la Commission) relatives :

- aux modalités de transmission à l'OSCE des données statistiques confidentielles (article 4, paragraphe 3) ;
- aux modalités d'accès à ces données confidentielles dans les locaux de l'OSCE (article 5, paragraphe 3) ;
- et, plus généralement, à toute question soulevée par un membre du comité et portant sur l'application du règlement du 11 juin 1990 (article 8).

Le Comité, qui est ainsi compétent sur les questions de confidentialité statistique touchant aussi bien les personnes physiques que les personnes morales, se réunit une fois par an (en application de son règlement intérieur ; article 7).

Sont alors considérées comme confidentielles (article 2-1) les données que la loi ou la pratique de chaque Etat membre considère comme confidentielles.

Signalons, en outre, que l'article 6 de ce règlement contient l'une des très rares apparitions du droit pénal dans les actes communautaires dérivés du Traité de Rome.

4.1.2. La directive n° 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données prévoit que chaque Etat membre doit, dans un délai de 3 ans, transposer dans son droit interne les dispositions permettant à cet Etat d'élever la protection des personnes physiques au niveau minimal requis par la directive.

La directive pose le principe de finalité, selon lequel chaque collecte est faite en vue d'une finalité précisément définie, et qu'est interdite toute utilisation desdites données en vue d'une autre finalité (détournement de finalité). Toutefois la directive autorise les Etats membres à déroger au principe de finalité, à l'occasion de la transposition, en prévoyant qu'est considéré comme compatible avec la finalité initiale tout "traitement ultérieur à des fins historiques, statistiques ou scientifiques, ..., pour autant que les Etats membres prévoient des garanties appropriées".

Ainsi, pour autant que la loi française de transposition¹ le prévoit (ce qui est le cas dans le projet), la directive permet de consolider l'article 7 bis de la loi de 1951 et d'ouvrir une voie analogue en faveur de la recherche historique et scientifique.

Les dispositions de la directive s'appliquent aux données individuelles relatives aux personnes physiques, mais les Etats membres peuvent décider d'élargir les garanties aux données individuelles concernant les personnes morales.

4.1.3. Le règlement n° 322/97 du Conseil du 17 février 1997 relatif à la statistique communautaire (complété par la décision de la Commission du 21 avril 1997 concernant le rôle d'Eurostat en matière de production statistique communautaire d'une part, et par l'article 285 introduit dans le traité de Rome par le traité d'Amsterdam d'autre part) constitue la "loi fondamentale de la statistique publique communautaire".

¹ - Dont un projet a été arrêté début 2001, mais non définitivement approuvé, ce qui met la France en infraction [et la Commission l'a déférée devant la Cour de Justice des Communautés européennes].

L'article 10 du règlement définit, parmi les six principes de la statistique communautaire (qui sont eux-mêmes repris de la résolution sur les principes fondamentaux de la statistique officielle adoptée à l'unanimité le 14 avril 1994 par le Conseil Economique et Social des Nations-Unies, sur la proposition de sa Commission de statistique), celui du secret statistique qui :

"signifie que les données relatives à des unités statistiques individuelles qui sont obtenues directement à des fins statistiques ou indirectement à partir des sources administratives ou autres sont protégées contre toute violation du droit à la confidentialité. Cela implique que toute utilisation non statistique des données obtenues et toute divulgation illicite soient empêchées".

Le chapitre V (Secret statistique : articles 13 à 18) développe les mesures à prendre pour garantir la confidentialité, cependant que l'article 20 du règlement étend la compétence du comité (européen) du secret statistique créé par le règlement du 11 juin 1990 à la mise en œuvre de toutes les mesures visées au chapitre V, notamment celles relatives à l'accès aux données confidentielles à des fins scientifiques (article 17). Implicitement, le comité du secret européen est également chargé de vérifier que la diffusion des données statistiques communautaires respecte les règles de confidentialité. D'ailleurs l'article 21 du règlement de 1997 modifie le texte du 2.1) du règlement de 1990 qui dit désormais :

"1) Données statistiques confidentielles : les données définies à l'article 13 du règlement du Conseil du 17 février 1997, relatif à la statistique communautaire", tandis que l'article 13 du règlement de 1997 dit :

"1 – Les données utilisées par les autorités nationales et l'autorité communautaire pour la production de statistiques communautaires sont considérées comme confidentielles lorsqu'elles permettent l'identification, directe ou indirecte, d'unités statistiques, ce qui aurait pour effet de divulguer des informations confidentielles.

2 – Par dérogation au paragraphe 1, les données qui sont tirées de sources accessibles au public et qui restent accessibles à celui-ci auprès des autorités nationales conformément à la législation nationale, ne sont pas considérées comme confidentielles".

Ainsi l'article 13 du règlement de 1997 produit une définition formelle de la notion de donnée statistique confidentielle au niveau communautaire, dont l'interprétation (confiée au comité européen du secret statistique) n'est pas évidente (notamment le paragraphe 2).

- 4.1.4. Le droit européen en vigueur comprend également la recommandation n° R(97)18 adoptée le 30 septembre 1997 par le Comité des ministres du Conseil de l'Europe concernant la protection des données à caractère personnel collectées et traitées à des fins statistiques, prise en application de la convention (dite : "n° 108") pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel signée à Strasbourg le 28 janvier 1981, qui s'inscrit elle-même dans le cadre de la Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales, faite à Rome le 4 novembre 1950.
- 4.1.5. Le règlement n° 45/2001 du Parlement européen et du Conseil du 18 décembre 2000 a porté application de la directive n° 95/46 du 24 octobre 1995 sur la protection des données personnelles pour ce qui concerne l'ensemble des institutions et organes communautaires.
- 4.1.6. En application de l'article 18 du règlement de 1997, EUROSTAT (au nom de la Commission) a présenté au comité européen du secret statistique plusieurs versions successives d'un projet de règlement relatif à l'accès par les chercheurs aux données confidentielles transmises à EUROSTAT par les Etats membres. La version de ce projet présentée lors de la réunion du comité du secret européen tenue le 7 février 2001 à Luxembourg, qui tenait compte de quelques remarques antérieures de la France, convenait grosso modo aux Etats membres, et la procédure se poursuit. A l'occasion de la réunion, l'Allemagne annonce son intention de modifier sa législation pour la rendre compatible avec le principe de l'accès des chercheurs européens aux données statistiques confidentielles.

4.2. Droit français

4.2.1. Le projet de loi modifiant la loi "Informatique et libertés" et transposant la directive communautaire du 24 octobre 1995 a fait l'objet de nombreux travaux (notamment le rapport BRAIBANT "Données personnelles et société de l'information" et le rapport du CNIS n° 55 de janvier 2000 sur les attentes spécifiques des statisticiens dans le cadre de cette transposition).

Le Conseil des ministres a approuvé le 18 juillet 2001 le projet de loi de transposition "relatif à la protection des personnes physiques à l'égard du traitement de données à caractère personnel et modifiant la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés", transmis le même jour à l'Assemblée nationale. Mais le calendrier d'examen de ce texte par le Parlement demeure inconnu, vu l'encombrement actuel, ce qui occasionne une grande frustration du côté de la recherche.

4.2.2. La loi (dite "DCRA") n° 2000-321 du 12 avril 2000 relative aux droits des citoyens dans leurs relations avec les administrations a notamment, dans le chapitre 2 (articles 4 à 9 ; "Dispositions relatives à la transparence administrative") de son titre I^{er}, organisé la conciliation entre les dispositions des lois CNIL, Archives et CADA dont le rapport du Conseil d'Etat de 1995 avait souligné la nécessité.

Notamment, l'article 5 modifie la loi CNIL en prévoyant que :

"Au-delà de la durée nécessaire à la réalisation des finalités pour lesquelles elle ont été collectées ou traitées, les informations ne peuvent être conservées sous une forme nominative qu'en vue de leur traitement à des fins historiques, statistiques ou scientifiques. Le choix des informations qui seront ainsi conservées est opéré dans les conditions prévues à l'article 4-1 de la loi du 3 janvier 1979 sur les archives".

Tandis que l'article 6 modifie le code pénal, l'article 7 modifie le titre I^{er} de la loi du 17 juillet 1978 et étend la compétence de la CADA au respect de la liberté d'accès aux archives publiques, dans les conditions prévues par le titre II de la loi n° 79-18 du 3 janvier 1979, c'est-à-dire notamment aux dérogations permises par l'article 8 de la loi sur les archives.

L'article 9 modifie la loi du 3 janvier 1979 sur les archives publiques et y introduit un article nouveau 4-1:

"Lorsque les documents [d'archives publiques] comportent des informations nominatives collectées dans le cadre des traitements automatisés régis par la loi du 6 janvier 1978, ces informations font l'objet, à l'expiration de la durée prévue à l'article 28 de ladite loi, d'un tri pour déterminer les informations destinées à être conservées et celles, dépourvues d'intérêt scientifique, statistique ou historique destinées à être détruites.

Les catégories d'informations destinées à la destruction ainsi que les conditions de leur destruction sont fixées par accord entre l'autorité qui les a produites ou reçues et l'administration des archives".

Ainsi la loi DCRA permet de prendre les mesures utiles à la conservation des données en vue de leur réutilisation à des fins scientifique, statistique ou historique, mais n'autorise pas pour autant cette réutilisation elle-même.

4.2.3. Le décret n° 2001-139 du 12 février 2001 porte création du comité de concertation pour les données en sciences humaines et sociales. Ce comité, présidé par le ministre chargé de la recherche, avec l'appui des ministres chargés de l'économie, de l'emploi et de l'éducation nationale compte parmi ses missions :

- de s'informer des activités de production et d'utilisation des données ;
- de déterminer les ensembles de données dont l'intérêt scientifique justifie le recueil, la préservation et la diffusion pour la recherche ;
- de faire des propositions afin de contribuer à la promotion de la formation à l'utilisation des données, l'accès des chercheurs et des enseignants-chercheurs aux données utiles aux sciences humaines et sociales ainsi qu'au développement des recherches utilisant les grands fichiers de durées ;
- de formuler des propositions en vue de la production de grandes enquêtes utiles à la recherche.

Le secrétaire général du comité interministériel et les membres du comité scientifique qui en dépend ont été nommés par arrêté du 5 novembre 2001.

4.2.4. Un projet bouleversant le régime de la loi de 1979 sur les archives a été présenté par le ministère de la culture en avril 1999. Dans la version initiale de ce projet, il n'y avait plus de protection de base, plus de mention du secret statistique, et la conséquence pour les données statistiques publiques était que :

(i) il n'y avait plus de protection du tout pour les données concernant les entreprises, qui devenaient immédiatement accessibles (le ministère de la culture pensait que le secret en matière industrielle et commerciale pouvait protéger ces données, mais cela est inexact, notamment vis-à-vis du droit de communication fiscal et douanier [point 1.3.3.] ;

(ii) la durée de protection des données statistiques relatives aux personnes physiques était ramenée de 100 à 50 (voire 25) ans.

Ce projet a donné lieu à des réunions interministérielles houleuses, et pour ce qui concerne la statistique publique à un avis pris le 30 novembre 1999 par le Comité du secret statistique concernant les entreprises suivie d'une lettre de son président au Directeur des Archives de France dont le contenu a été repris, dans l'avis général sur les programmes statistiques 2000 adopté par le CNIS le 9 décembre 1999, puis d'une audition du Directeur des Archives de France par le bureau du CNIS, le 16 mars 2000. Une négociation avait finalement ouvert la voie à une solution satisfaisante pour la statistique publique, consistant à abaisser la durée de protection de 30 à 25 ans pour les données d'entreprises et à conserver la durée de 100 ans pour les données sur les personnes physiques, tout en abandonnant le privilège (exorbitant !) de l'impossibilité de dérogation à ce délai à condition de subordonner ces dérogations à notre avis favorable. Mais ce projet n'a pas abouti.

Toutefois le ministère de la culture a saisi l'occasion du projet de loi sur la société de l'information (LSI) pour y inscrire un projet beaucoup plus court, mais revenant à sa position initiale pour ce qui concerne la statistique publique. Ce projet vient d'être transmis à la CNIL, et au Conseil d'Etat. Dans son état actuel, le chapitre III ("L'accès aux archives publiques", articles 7 et 8) du projet de la loi LSI reprend les conséquences énoncées au (i) et (ii) ci-dessus, et modifie en conséquence le premier alinéa de l'article 6 de la loi de 1951. L'adoption d'un tel projet serait de nature à ruiner la confiance des répondants aux enquêtes statistiques publiques et à entraîner une baisse spectaculaire du taux de réponse à ces enquêtes. En vérité, en cas d'adoption d'une telle disposition, une réflexion sur le secret statistique concernant les entreprises devient futile.

5 – AMELIORATIONS ENVISAGEES AU NIVEAU DES REGLES DE FONCTIONNEMENT DU COMITE

5.1. La loi du 7 juin 1951

On l'a déjà dit : c'est l'article 6 de la loi qui fonde le secret statistique, et c'est le 2^e alinéa de cet article qui justifie une jurisprudence s'agissant du secret statistique concernant les renseignements individuels d'ordre économique ou financier, que l'on convient d'appeler le "secret statistique concernant les entreprises".

La seule petite difficulté (voir le point 2.0 et le point 3.2.2.) concerne le membre de phrase "figurant sur les questionnaires revêtus du visa prévu à l'article 2" et concerne tant le premier que le deuxième alinéa de l'article 6, mais cette difficulté est aujourd'hui purement formelle et il convient de ne surtout pas toucher à la loi du 7 juin 1951, qui fonctionne bien.

5.2. Le décret n° 84-628 du 24 juillet 1984

5.2.1. En ce qui concerne la composition du comité du secret, un membre du groupe a suggéré d'y ajouter un représentant de la recherche. En considération des liens à établir, sans doute, avec le comité de concertation sur les données en sciences humaines et sociales, cette proposition paraît raisonnable. Elle aurait formellement l'avantage de porter à 15 (impair) le nombre des membres du comité. Signalons que, sauf pendant les trois premières années d'existence du comité, il a été impossible d'obtenir du ministre concerné la désignation d'un représentant "des utilisateurs régionaux et locaux de l'information statistique". On pourrait donc aussi supprimer ce représentant, ou le remplacer par celui de la recherche, ce qui entraînerait aussi une modification du décret. Une autre possibilité à envisager serait d'intégrer un représentant des archives de France dans la composition du comité.

5.2.2. En considération des dispositions à prendre en cas d'adoption du projet de règlement sur l'accès des chercheurs aux données confidentielles transmises par les Etats membres à Eurostat (point 4.1.6.), il a été suggéré que, pour réduire le délai de réponse pour l'examen des dossiers de ce type transmis par Eurostat, le comité du secret statistique concernant les entreprises pourrait décider de déléguer ce type de décision à un bureau permanent nommé en son sein.

La mise en place d'un tel dispositif nécessiterait une modification du décret. Sur le fonds, la légitimité de restreindre le débat sur ce type de dossier à certain des membres du comité seulement pose un véritable problème.

5.2.3. Une autre voie possible pour accélérer les décisions du comité sur certains types de dossiers serait d'envisager la possibilité d'une procédure écrite. Rien dans la loi de 1951, ni dans le décret de 1984, ne semble s'opposer explicitement à la mise en place par le comité du secret d'une procédure écrite pour régler certaines questions, sauf peut-être l'apparition de l'expression "tient séance" dans le 3^e alinéa de l'article 21 du décret.

D'ailleurs une lecture possible de cet alinéa laisse penser qu'il ne s'applique qu'en cas d'absence ou d'empêchement du président, et qu'il est donc tout à fait loisible à celui-ci d'organiser une procédure écrite ; mais il convient vraisemblablement de faire mieux.

5.3. La pratique

5.3.1. Traditionnellement, le comité se réunit trois fois par an. Les demandeurs jugent cette cadence un peu trop lâche, et le comité a admis qu'il pourrait se réunir quatre fois par an. Mais, en fait, cela n'a pas été mis en œuvre et serait assez délicat en raison des disponibilités des membres du comité. Un tel passage de trois à quatre séances annuelles ne serait, de toutes façons, pas suffisant pour résoudre la question des liaisons avec Eurostat et le comité européen du secret statistique qui entraîne les considérations évoquées aux points 5.2.2. et 5.2.3.

5.3.2. Comme il a été dit précédemment, le comité distingue les avis (qui concernent les collectes statistiques obligatoires) et les recommandations (qui concernent soit les collectes statistiques non obligatoires, soit des collectes non statistiques).

Cela ne paraît pas optimal. On pourrait :

- soit requalifier en avis le résultat de l'examen des collectes statistiques non obligatoires ;
- soit qualifier "d'opinion du comité" le résultat de l'examen des collectes non statistiques.

5.3.3. La pratique habituelle actuelle consiste à convoquer l'intégralité des demandeurs à l'heure du début de la réunion, et à laisser tout le monde assister à l'ensemble de la réunion, sauf dans le cas exceptionnel où le secrétariat, en accord avec le président, envoie deux convocations distinctes avec une participation restreinte pour celle correspondant au début de la réunion. Cette pratique semble satisfaisante, mais peut être questionnée.

5.3.4. Seuls, d'une part, le compte rendu de la réunion rédigé par le secrétariat et approuvé par le comité lors de sa réunion suivante et, d'autre part, les décisions ministérielles communiquées aux demandeurs et archivées par le secrétariat du comité (accompagnées des engagements au secret des personnes habilitées pour chaque décision) gardent la trace des réunions du comité. Naturellement la loi sur l'accès aux documents administratifs s'applique à ces documents, et chacun peut en principe y accéder. Mais aucune publicité n'est faite des avis du comité (sauf pour ceux qui assistent aux réunions) et des décisions ministérielles qui les suivent, en tout cas pour ce qui concerne les décisions individuelles.

Il n'est pas proposé de modifier cette pratique, d'autant qu'il ne serait pas raisonnable de passer au Journal officiel et que le bulletin officiel du ministère, qui serait le vecteur naturel d'une telle publicité, a disparu. La proposition d'utiliser à cet effet le site Internet du CNIS paraît bien adaptée.

6 – AMELIORATIONS ENVISAGEES AU NIVEAU DE LA JURISPRUDENCE DU COMITE

6.1. Déroptions générales

- 6.1.1. La mise en œuvre du paragraphe 2 de l'article 13 du règlement de 1997, concernant la dérogation du secret statistique accordée aux "données qui sont tirées de sources accessibles au public et qui restent accessibles à celui-ci auprès des autorités nationales conformément à la législation nationale" ne paraît pas simple.

Ainsi, en vue d'une application uniforme de cette disposition dans les Etats membres, la France avait demandé, en comité du secret statistique européen, à la Commission de procéder à un recensement des actes du droit communautaire imposant de rendre accessible au public des données concernant les entreprises, ainsi que la liste des données concernées. Aucune réponse n'a été obtenue.

La signification de "restent accessibles.....législation nationale" est également loin d'être claire. Elle pourrait être relative à des données pour lesquelles la durée d'accès est très courte (comme l'affichage des résultats à des examens !).

En ce qui concerne la notion de "source accessible au public", il paraît clair qu'il ne peut s'agir que du public général, et non pas d'une catégorie particulière, comme les habitants d'une commune ou les contribuables d'un centre des impôts. Il paraît clair aussi que par accès on entend non seulement "avoir le droit de voir les données", mais aussi avoir le droit de les reproduire, non seulement pour soi-même, mais aussi pour les diffuser (que cela soit payant ou gratuit, cela n'entre pas en ligne de compte). C'est-à-dire que (dans le langage mis en place par le rapport MANDELKERN) la source doit être non seulement "querable" mais aussi "légalement portable".

Ainsi le cas des données d'entreprises soumises à publicité légale entre clairement dans le cadre du dispositif envisagé (même si ces obligations ne sont pas strictement homogènes au sein des 15 Etats membres de la Communauté). Mais qu'en est-il si une entreprise ne dépose pas au greffe les données soumises à publicité légale, et décide de payer les amendes correspondantes ? Est-ce ce cas qui correspond à des données "accessibles", mais qui ne restent pas "accessibles auprès des autorités nationales conformément à la législation nationale". Autrement dit : si l'entreprise paie une amende en application de la législation, s'est-elle acquittée de son obligation légale et ne peut-on plus exiger d'elle l'accès aux données soumises à publicité légale ?

- 6.1.2. Une proposition consiste à utiliser le dispositif du paragraphe 2 de l'article 13 du règlement de 1997 non pas pour l'accès à des données individuelles, mais de manière indirecte pour la construction de données agrégées. Cette proposition est élégante et prudente, elle mérite certainement d'être retenue.

- 6.1.3. Une question connexe est celle du statut, au regard du secret statistique concernant les entreprises, d'une donnée collectée par une enquête statistique et pour laquelle un responsable habilité d'une entreprise a explicitement donné son accord pour la diffusion. La diffusion de cette donnée individuelle serait déontologiquement évidente, mais peut-être bien juridiquement contestable. Une solution possible serait de demander à l'entreprise de confirmer son accord sur la diffusion de cette donnée, en la transmettant à l'identique, mais en dehors de la collecte statistique.

Une meilleure solution serait d'admettre, par la voie jurisprudentielle, que le secret statistique ne s'applique plus dans le cas.

- 6.1.4. Un membre du groupe a suggéré de supprimer le droit d'opposition, c'est-à-dire la possibilité pour une entreprise concernée d'adresser un courrier au secrétariat du comité du secret statistique pour déclarer qu'elle s'oppose à ce que les données individuelles statistiques la concernant soient rendues accessibles, lorsqu'une décision relative à l'accès public d'une telle catégorie de données a été prise. La majorité du groupe de travail est opposée à la suppression du droit d'opposition. D'une part la conservation d'un tel principe paraît nécessaire pour préserver la confiance de certains répondants ; d'autre part le nombre des entreprises ayant demandé, à ce jour, à profiter du droit d'opposition n'est que de l'ordre de la centaine (dont une seule depuis trois ans) notamment du fait de la publicité très modérée faite à l'existence du droit d'opposition.

6.1.5. En ce qui concerne les avis du comité du secret qui, tels celui du 3 juillet 1986, ont pour objet de déclarer diffusables (sauf objection de la part de l'entreprise considérée), des informations d'intérêt général la concernant collectées par des enquêtes statistiques, on peut considérer que, jusqu'à présent, ces avis concernent l'ensemble des secteurs économiques, et l'ensemble des entreprises au-dessus d'une certaine taille.

Il a été proposé, au sein du groupe de travail, d'étendre la possibilité de rendre diffusables (soit auprès du public général, soit auprès de publics restreints) des informations individuelles collectées par des enquêtes statistiques à des variables plus spécifiques issues d'enquêtes moins généralistes. Il a été suggéré, en outre, de donner au comité du label la capacité de proposer, à l'occasion de l'examen de telles collectes, les données susceptibles d'une telle diffusion, à charge ensuite au comité du secret, puis à l'assemblée plénière du CNIS, de prendre les avis correspondants.

Il s'agit de généraliser, avec une procédure appropriée, les exemples que représentent dans cette voie, d'une part, la possibilité d'accès à certaines variables collectées par l'enquête sur la commercialisation des logements neufs à la suite d'un accord entre le ministère de l'équipement et la Fédération Nationale des promoteurs-constructeurs (FNPC) approuvé par le comité du secret (point 3.3.5.) et, d'autre part, l'accès restreint aux données de l'enquête sur les liaisons financières (LIFI ; point 6.1.12).

6.1.6. Bien que le texte de l'avis adopté par l'assemblée plénière du CNIS du 3 juillet 1986 ne le mentionne pas explicitement, il est clair que cet avis fait implicitement référence à la source enquête annuelle d'entreprise (EAE), pour les quatre variables dont la diffusion individuelle est autorisée.

Il devient progressivement clair, aussi, que le vecteur naturel de diffusion de ces données individuelles est SIRENE. Peut-être pas le répertoire SIRENE strictement défini par le décret n° 73-314 du 14 mars 1973 modifié, qui n'autorise explicitement que la diffusion de la "catégorie d'importance de l'effectif salarié de l'entreprise et de chacun de ses établissements" (sans précision sur les tranches à retenir), d'une part, et les "numéros de la nomenclature des activités économiques caractérisant l'activité de l'entreprise et de chacun de ses établissements (cela vise-t-il uniquement les activités principales, ou cela comprend-il également les activités secondaires ?), d'autre part. Mais une extension de ce répertoire serait justifiée certes par la jurisprudence du comité du secret, mais aussi par le règlement n° 2186/93 du Conseil du 22 juillet 1993 relatif à la coordination communautaire du développement des répertoires d'entreprises utilisés à des fins statistiques. En effet, ce règlement donne une base juridique à l'introduction dans SIRENE (qui fait l'application de ce règlement pour la France) :

- des activités secondaires (pour celles qui atteignent 10 % du total des activités de l'entreprise ou représentent plus de 5 % du total de l'activité nationale de ce type) ;
- de la dimension, mesuré par le nombre de personnes occupées, de l'unité légale et de chacune de ses unités locales ;
- du montant net du chiffre d'affaires de l'unité légale, éventuellement mesuré en tranches libellées en euros.

6.1.7. Toutefois, pour ce qui concerne les effectifs, l'intérêt du point 1 de l'avis du 3 juillet 1986 est devenu moins important depuis le développement d'un nouveau traitement exhaustif des DADS (source de type CERFA).

En effet l'arrêté initial du 24 juin 1998 "portant création d'un traitement automatisé de l'ensemble des informations issues des DADS" a été ensuite modifié par un arrêté du 10 décembre 1998, pris après avis favorable de l'ensemble des organismes associés à la collecte DADS, d'une part, et de la CNIL, d'autre part, qui précise :

"La diffusion des effectifs salariés totaux, au 31 décembre de chaque année, des entreprises et de leurs établissements est autorisée, sauf pour les entreprises qui notifieraient leur opposition au directeur général de l'INSEE".

Il se trouve que la source annuelle DADS est supérieure à la source EAE :

- par l'extension de son champ, tant au niveau des activités qu'au niveau de la taille des entreprises ;
- par l'uniformité réglementaire du concept d'effectif à déclarer dans les DADS .

Cela est loin de rendre, pour ce qui concerne les effectifs, l'avis du 3 juillet 1986 inutile, du fait :

- que l'on peut utiliser la source EAE en substitution de la source DADS si l'entreprise s'est opposée à la diffusion de cette dernière, mais pas de la première ;
- que l'on peut utiliser la source EAE pour les établissements, dans le cas où la déclaration DADS regroupe certains des établissements de l'entreprise (ce qui n'est pas rare), soit en substitution, soit en clef de répartition ;
- que, surtout, en élargissant à d'autres sources statistiques que l'EAE l'interprétation de l'avis de 1986 (ce qui est parfaitement licite) on peut mobiliser certaines données statistiques relatives aux effectifs là où la source DADS soit n'existe pas, soit n'est pas correcte. C'est notamment le cas pour les collectivités locales, pour lesquelles existe une enquête statistique sur les effectifs.

Le groupe de travail s'est également penché sur la question de la diffusion de données sur l'emploi à un niveau géographique détaillé. L'avis du 3 juillet 1986 emporte comme conséquence que toute agrégation de données individuelles tenant compte des remarques précédentes est diffusable et qu'il est donc possible, par exemple, de diffuser des données sur l'emploi salarié total par commune. La demande est toutefois forte pour que des données détaillées soient également diffusées à un tel niveau géographique fin, par exemple des ventilations des effectifs par CSP au niveau communal. En l'état, cette demande ne peut être satisfaite mais, sur la base d'une étude préalable et en combinant les règles proposées par le groupe de travail tant en matière de données sur les effectifs qu'en termes de procédures de protection des données statistiques confidentielles (cf. partie 7), une procédure de génération d'estimations de cette nature respectant l'ensemble des contraintes paraît envisageable.

6.1.8. Pour ce qui concerne les codes NAF des activités principales et secondaires des entreprises et des établissements, on a déjà vu que l'utilisation de l'EAE pour leur mise à jour dans SIRENE a été rendue licite, et que leur diffusion est permise.

Une décision du directeur général de l'INSEE du 13 septembre 1989, prise après l'avis adopté le 21 juin 1989 par le comité du secret statistique concernant les entreprises a d'ailleurs indiqué :

"Il est possible d'utiliser les enquêtes annuelles d'entreprises pour effectuer la mise à jour du code APE (Activité Principale Exercée) de SIRENE, sous réserve que cette utilisation soit signalée sur le questionnaire des enquêtes et que l'avis SIRENE porte la mention du service enquêteur."

Un problème connexe, en principe réglé par l'article 4 du décret n° 92-1129 du 2 octobre 1992 portant approbation des nomenclatures d'activités et de produits (et faisant suite à des jugements répétés de la Cour de cassation), est celui de l'usage abusif, constituant un défaut de procédure, du code APE déterminé par l'INSEE pour des raisons uniquement statistiques que font (parfois systématiquement) certaines administrations dans l'application de leur propre réglementation. Il est légitime que le code APE puisse faire présomption, mais il n'est pas normal pour une administration de s'en servir systématiquement, sans autre considération.

Un membre du groupe de travail a proposé que d'autres sources statistiques publiques, notamment les enquêtes de branche, puissent être utilisées pour établir la liste des activités et le code d'activité principale servant à nourrir SIRENE. Cette proposition n'a pas rencontré d'opposition au sein du groupe de travail. Dans cette voie, l'utilisation du recensement général de l'agriculture serait également bien adaptée.

6.1.9. Pour ce qui concerne le chiffre d'affaires, on a déjà vu que l'utilisation de l'EAE pour introduire dans SIRENE une tranche d'importance était licite. Il convient simplement de mettre à jour l'avis de 1986 sur les seuils des tranches, en raison du passage à l'euro au 1^{er} janvier 2002 et en tenant compte des seuils figurant dans le règlement de 1993. La publication des données individuelles d'entreprises sur la tranche d'importance du chiffre d'affaires (implicitement, à partir de la source EAE) est également autorisée par le point 2 de l'avis du CNIS de 1986 et SIRENE a manifestement vocation pour devenir le support d'une telle diffusion.

6.1.10. Le point 3 de l'avis du 3 juillet 1986 du CNIS est relatif à la possibilité de diffuser (toujours implicitement à partir de l'EAE) une information individuelle sur l'entreprise relative à la part du chiffre d'affaires réalisée à l'exportation (selon une nomenclature prenant 4 modalités). Le règlement européen du 22 juillet 1993 ne cite pas cette variable et ne fournit donc aucune base juridique pour l'introduction de cette variable dans SIRENE, qui ne pourrait donc résulter que du droit français seulement. La question de l'éventuelle diffusion de cette donnée est également posée.

6.1.11. Le point 4 de l'avis du 3 juillet 1986 du CNIS est relatif à la possibilité de diffuser (toujours implicitement à partir de l'EAE) une information individuelle sur l'entreprise relative à l'indication d'une activité de recherche au sein de l'entreprise (variable 0,1). Le règlement européen du 22 juillet 1983 ne cite pas cette variable et ne fournit donc aucune base juridique pour l'introduction de cette variable dans SIRENE, qui ne pourrait donc résulter que du droit français seulement. La question de l'éventuelle diffusion de cette variable est également posée.

6.1.12. L'avis du comité du secret statistique du 10 mai 1998 (point 3.3.4.) relatif à la diffusion restreinte de données collectées dans l'enquête "Liaisons financières (LIFI)" doit-il être mis en œuvre ? Dans le cas d'une réponse positive, il convient de poursuivre la procédure au sein du CNIS en assemblée plénière, puis d'organiser la publicité de cette décision.

6.1.13. L'un des membres du comité a proposé d'aller au-delà en rendant publique (soit auprès du public général, soit auprès des chercheurs) la variable d'indication de l'appartenance d'une entreprise à un groupe résultant de l'enquête LIFI. Dans le cas d'une réponse positive concernant le public en général, la question de la présence de cette variable dans le fichier SIRENE et de son éventuelle diffusion se poserait dans des conditions analogues à celles exposées aux points 6.1.11 et 6.1.12. En fait, de telles perspectives n'emportent pas l'assentiment de la majorité du groupe de travail, pour de multiples raisons :

- grande sensibilité du monde des entreprises sur ces informations (même si d'autres sources, privées, existent sur ce sujet, et bien plus précises) ;
- comparaison avec la situation sur cette question dans les autres Etats membres de l'Union européenne ;
- absence d'une définition faisant consensus de la définition du groupe ;
- approche partielle de la notion de groupe, résultant de la nature même de la source LIFI.

6.2. Habilitations permanentes

6.2.1. Cas des services statistiques ministériels

Peut-être serait-il utile de procéder à une révision générale de l'ensemble des habilitations permanentes examinées par le comité du secret, en faveur des services statistiques ministériels, dans les conditions décrites en 3.4.1., notamment pour tenir compte des évolutions de la jurisprudence du comité intervenues depuis 1985.

6.2.2. Autres cas

Cette révision devrait, sans doute, englober l'ensemble des habilitations permanentes.

6.3. Habilitations temporaires

En ce qui concerne les habilitations temporaires, examinées dans les conditions décrites en 3.4.2., plusieurs questions méritent examen.

6.3.1. Extension de finalité d'une habilitation

Actuellement, le comité examine une demande d'habilitation au vu d'une description de l'objet de la demande figurant dans le dossier, qui vise une (ou des) finalité(s) déterminée(s). Un membre du comité a expliqué que, notamment dans le cas de la recherche, la finalité pouvait se trouver modifiée ou étendue par l'obtention de résultats intermédiaires, et qu'il conviendrait d'accepter une telle modification de finalité sans avoir à repasser devant le comité. Une forte majorité du groupe de travail, bien que consciente du fait qu'une recherche procède en général par étapes et que le résultat de chaque étape peut modifier la suite de cette recherche, est opposée à cette proposition, qui constituerait un détournement substantiel de la procédure actuelle

6.3.2. Fichiers non disponibles

Le fait que les habilitations temporaires ne puissent s'appliquer qu'à des fichiers réputés constitués au jour de l'examen de la demande par le comité est souvent jugé contraignant par les demandeurs, obligé parfois de repasser devant le comité. Le groupe de travail n'estime pas possible de modifier cette règle ; par contre il est envisageable d'assouplir la procédure en chargeant le service enquêteur de présenter lui-même la demande si la délibération initiale du comité a prévu ce cas.

6.3.3. Délai, notamment en vue du projet de règlement d'Eurostat

Comme il a été dit en 5.2.2. et 5.2.3., la réduction du délai entre le dépôt d'un dossier de demande, la délibération du comité et la décision ministérielle doit être recherchée, notamment dans le cadre du projet de règlement sur l'accès des chercheurs aux données confidentielles transmises par les Etats membres à Eurostat. L'examen de la possibilité d'une procédure écrite semble indispensable pour aller dans cette voie.

6.4. Enseignement et recherche

6.4.0. Information des chercheurs

Le groupe de travail a constaté que de nombreux chercheurs directement intéressés par les fichiers statistiques concernant les entreprises ne connaissent même pas l'existence du comité du secret statistique concernant les entreprises et de la procédure leur permettant de demander un accès privilégié à ces données individuelles. Le groupe de travail propose que les organismes producteurs de données sur les entreprises fassent un effort d'information (site web spécifique, par exemple) en direction du monde de la recherche sur les données disponibles et la procédure du comité du secret statistique.

6.4.1. Représentant de la recherche au Comité du secret statistique concernant les entreprises

Le groupe de travail recommande que la sélection du chercheur qui siègerait au comité du secret statistique concernant les entreprises soit approuvée par le comité de concertation du « Centre Quételet ».

6.4.2. Enseignement (cas général)

La statistique d'entreprise intéresse de plus en plus les enseignants en sciences sociales, à la fois du point de vue de ses résultats et du point de vue de la formation des étudiants à l'analyse de fichiers individuels. Cette demande sociale est appelée à croître. Elle fait partie des demandes que les organismes comme l'INSEE ont pour mission de satisfaire. Le groupe de travail recommande que les organismes producteurs d'enquêtes statistiques sur les entreprises créent des fichiers didactiques individuels, aussi « anonymisés » que possible (par exemple en utilisant des procédés de micro-agrégation), en vue d'application pour l'enseignement.

6.4.3. Enseignement et ancienneté des données dont le format est particulièrement original

Lorsque le format des données est original, comme l'a été par exemple le fichier des données de l'Enquête "Liaisons industrielles" du SESSI, il est possible que ce fichier n'ait pas été inclus dans les bases prévues pour les applications pour l'enseignement. Il n'existe donc pas de données statistiques anciennes permettant aux étudiants de s'exercer à l'analyse de ce type de données. Le groupe de travail recommande donc que le comité du secret statistique concernant les entreprises tienne compte de l'ancienneté des données individuelles (de type original) auxquelles il est ainsi donné accès aux enseignants. Un seuil de dix ans (au lieu de trente ans) d'ancienneté pour l'accès à ces fichiers pédagogiques non transformés par micro-agrégation pourrait être fixé.

6.4.4. "Arbitre" pour les publications scientifiques

Les données d'enquête accessibles par la procédure du Comité du secret statistique concernant les entreprises sont parfois appariées avec des données administratives. Certains chercheurs ont accès à ces bases mixtes. Le groupe de travail rappelle le principe d'égalité des chances d'accès de tous les chercheurs aux données sur la base desquelles ils publient. La publication d'articles basés sur des

fichiers de données auxquels la communauté scientifique n'a pas accès soulève un problème d'éthique pour les comités de rédaction des revues scientifiques, et un problème déontologique pour les chercheurs qui tentent de publier sur la base de ces données. Le groupe de travail recommande que le comité du secret statistique concernant les entreprises exige des chercheurs qui ont accès aux données mixtes qu'ils le signalent clairement dans les manuscrits qu'ils soumettent aux revues scientifiques. Ainsi, un certain nombre d'items obligés, standardisés, devraient être définis pour identifier clairement la nature des données utilisées. Par ailleurs, le comité du secret pourrait permettre la vérification des analyses effectuées sur des données mixtes par un arbitre ("referee", ou "relecteur") extérieur aux organismes producteurs des données et/ou auxquels sont affiliés les auteurs. Cela pourrait se faire en autorisant l'accès par cet arbitre aux données individuelles mixtes et aux traitements effectués sur les données individuelles avant la publication dans une revue scientifique.

6.4.5. Comité d'éthique

Le groupe de travail recommande que le comité du secret statistique concernant les entreprises envisage l'intervention de comités d'éthique comme celui qui est prévu dans le décret portant création d'un comité de concertation pour les données en sciences humaines et sociales, dit « Centre Quételet ».

6.4.6. Remplacement d'enquêtes par des sources administratives

Le remplacement d'enquêtes par des sources administratives peut engendrer des restrictions des données auxquelles les chercheurs auront accès. Ces données ne pourront plus être cédées par l'INSEE dans le cadre habituel du comité du secret statistique concernant les entreprises. Même si les administrations cèdent des données à des chercheurs au coup par coup, cette cession peut n'être que partielle, ou dissymétrique (dans le cas de fichier à apparier contrôlés par des administrations différentes). Le groupe de travail recommande que la réflexion se poursuive sur les moyens de rétablir l'accès des chercheurs à ces données dans le cadre habituel du comité du secret statistique concernant les entreprises.

6.5. Information statistique et régulation de la concurrence

Le lien entre l'information statistique et la régulation de la concurrence constitue un enjeu d'importance pour les entreprises qui a déjà été abordé au 2.3.5.. Cette question fait l'objet de la note "Du bon usage de la statistique dans la régulation de la concurrence" que l'on trouvera en annexe du présent rapport.

7 – METHODES POUR PROTEGER LES DONNEES STATISTIQUES CONFIDENTIELLES

Le traitement des données statistiques pour préserver la confidentialité est aujourd'hui un domaine à part entière de la technique statistique, et la recherche théorique y est très active depuis une dizaine d'années. L'objectif de ces méthodes est de réaliser un compromis entre la mise à disposition d'informations dans l'intérêt de la collectivité et la protection du secret que peut demander toute personne morale ou physique en ce qui concerne les données qui lui sont relatives². Dit en termes simples, le problème est le suivant : en utilisant des informations rendues publiques par une organisation statistique, peut-on découvrir des informations portant sur un individu particulier (une entreprise dans le cas traité par ce rapport) appartenant à la population concernée, informations qui pourraient nuire à ses intérêts ?

Le problème se pose différemment selon les deux types de la mise à disposition de l'information :

- sous forme agrégée (totaux par exemple)
- sous forme individuelle (les valeurs de certaines caractéristiques individuelles).

Le statisticien ne peut donner l'accès à des informations qui permettent l'identification directe des entreprises, sauf celles qui font l'objet d'une dérogation explicite (voir le paragraphe sur la publication de l'activité principale, l'effectif annuel et d'une tranche de chiffre d'affaires de chaque entreprise).

Avant tout approfondissement, il est bon de réaliser que presque toute diffusion d'information donne de l'information sur les unités statistiques de la population. Mais en général, cette information est tellement peu précise qu'elle ne peut porter préjudice aux dites unités. Par exemple, lorsque l'on publie qu'en 1999 l'effectif total du secteur des entreprises de « *production et distribution d'électricité, de gaz et de chaleur* » est de 164 000 personnes, on dit ainsi qu'aucune des entreprises de ce secteur n'emploie à elle-seule en 1999 plus de 164 000 personnes. C'est une information qui concerne chacune des entreprises du secteur. Mais on imagine mal en quoi cette information pourrait nuire à l'une quelconque d'entre elles.

7.1. Protection de la confidentialité statistique lors de la diffusion de données agrégées

Les grands types de méthodes utilisées sont la suppression pure et simple de certaines informations, la réduction de la finesse des ventilations, les arrondis, les perturbations aléatoires.

7.1.1. La suppression pure et simple

Elle consiste à ne pas diffuser d'information sur une sous-population quand elle ne comprend qu'un trop petit nombre d'unités ou qu'au sein de cette sous-population l'une d'elles est tellement prépondérante que les informations sur l'ensemble de la sous-population permettent d'estimer avec une assez bonne précision les informations relatives à l'unité prépondérante. Le nombre minimum d'unités garantit qu'une unité statistique au sein de ce sous-groupe n'a pas pour son propre usage une information trop précise sur chacune des autres unités de cette sous-population, information qu'elle pourrait utiliser à des fins concurrentielles.

A noter que si cette suppression est une méthode simple, elle conduit à une perte totale d'information.

Pour limiter cette perte, on cherche néanmoins à publier l'information pour un niveau de population plus large, au sein duquel la prépondérance d'une unité n'existe plus. Mais alors, il faut s'assurer que l'on ne peut pas par recalcul estimer la cellule supprimée. Il faut dans ce cas éliminer d'autres cellules, qui sont dites relever de la « confidentialité secondaire », les cellules posant directement le problème signalé ci-dessus étant dites relever de la « confidentialité primaire ». Le choix des cellules à éliminer pour cause de confidentialité secondaire peut être optimisé, de façon à limiter la perte d'information. De plus, dans les ventilations à plusieurs dimensions (selon l'activité, la taille, la localisation géographique, etc.), il faut s'assurer que la suppression d'une cellule primaire et des cellules secondaires est complète, c'est à dire qu'il n'y a aucun moyen de recalculer par combinaison les statistiques non publiées. Ce type de contrainte conduit à des algorithmes compliqués, qui nécessitent le recours à un ordinateur. Parfois, les calculs sont tellement complexes que seules des solutions approchées ou sous-optimales sont fournies.

² - Pour une présentation pédagogique et détaillée des différentes méthodes, on pourra se reporter à «Manual on Disclosure Control Methods », Eurostat, 1996, ISBN 92-827-6045-6. Pour une revue des principaux thèmes de recherche actuelle, on consultera « Statistical data protection - Proceedings of the conference, Lisbon, 25 to 27 March 1998 - 1999 edition », Eurostat, numéro de catalogue : CA-24-99-615-EN-C.

Dans toute sa généralité, la protection contre la divulgation de données confidentielles par gestion des confidentialités primaire et secondaire est probablement impossible. Ce qui est décrit ci-dessus peut être mis en œuvre au niveau d'une publication donnée. Peut-il l'être pour l'ensemble des publications et bases de données produites à partir d'une source au sein d'institutions qui peuvent publier des tableaux à partir de demandes adressées à différents services (nationaux et régionaux) ? Cela supposerait au minimum une centralisation systématique de toutes les demandes reçues sur une même source de données. Ensuite, il faudrait choisir entre restreindre a priori les tabulations sur demande autorisées et se doter d'outils puissants d'identification des cas de cellules à confidentialiser. Il n'est pas certain que la complexité de ce problème, au sens de la technique algorithmique, permette de traiter ce défi sans restreindre a priori l'information diffusable. Les logiciels disponibles actuellement ne semblent pas en mesure de traiter ce problème. Des investissements supplémentaires en recherche seraient nécessaires pour permettre une telle approche. Se protéger contre quelqu'un qui voudrait avoir accès à des données confidentielles en multipliant les demandes de tableaux croisant divers critères est extrêmement difficile. On retrouve le constat général que la protection des données consiste en pratique à un arbitrage entre le niveau de garantie du secret recherché et les coûts induits par cet objectif.

C'est la méthode la plus classiquement utilisée en statistique d'entreprises, à l'INSEE et dans les SSM, ainsi qu'au niveau européen.

7.1.2. La réduction de finesse des ventilations

Les tableaux obtenus par la méthode précédente possèdent des cases non renseignées, et ne sont plus « rectangulaires ». Pour éviter cet inconvénient, on procède parfois à une modification des critères de ventilation, de façon à s'assurer qu'on ne rencontrera plus aucun cas de confidentialité primaire. La difficulté consiste à mettre au point un algorithme de modification des critères qui limite la perte d'information.

Par exemple, s'il apparaît un problème avec la ventilation d'une population selon l'activité principale au niveau des classes de la NAF (3 chiffres et une lettre), on pourra se limiter à une ventilation uniquement selon les groupes d'activité (3 chiffres). Mais si on perd ainsi trop d'information, on pourra se limiter à cette opération pour certains groupes seulement, et conserver les classes là où elles ne posent pas de problème.

7.1.3. Les arrondis

Cette méthode modifie les statistiques publiées, de façon à ne pas fournir d'information trop précise sur les individus de la population.

Elle semble naturellement adaptée aux comptages. Elle consiste par exemple à arrondir par différentes méthodes les comptages qui figurent dans des tableaux à des chiffres multiples de 5 ou de 10 (ou plus généralement de tout nombre choisi convenablement).

7.1.4. Les perturbations aléatoires

Cette méthode modifie aussi les statistiques publiées, de façon à ne pas fournir d'information trop précise sur les individus de la population.

Elle consiste à ajouter aux cellules un montant aléatoire, de moyenne nulle et de variance choisie par le statisticien. Il est également possible de perturber non pas les statistiques dans les tableaux, mais plutôt directement les données individuelles dans le fichier initial et de tabuler ensuite les données perturbées.

7.2. Protection de la confidentialité statistique lors de la diffusion de données individuelles

Tout d'abord, il est des cas où il s'avère indispensable de diffuser des données individuelles. Certains utilisateurs ont besoin de données individuelles pour estimer des statistiques qui ne sont pas publiées, et qu'en général ils ne peuvent pas définir a priori. Ils ont besoin de faire certaines analyses préalables sur les fichiers de données individuelles pour effectuer des choix de traitement. Ils ne peuvent donc pas demander aux unités de diffusion des services statistiques de produire ces statistiques à la demande, selon des spécifications initiales. Ces utilisateurs sont des chargés d'études, des chercheurs. Ou encore des étudiants, dans le cadre de leur formation professionnelle.

L'objectif du statisticien est d'empêcher, malgré la connaissance d'informations individuelles, l'identification de l'unité statistique correspondante.

On considère que l'on peut classer les variables pour lesquelles on a de l'information en trois catégories : variables d'identification, variables de quasi-identification et variables sensibles. Un point de vocabulaire spécialisé : il est d'usage de distinguer l'**identification directe** d'une unité statistique (son nom et son adresse, son identifiant dans un répertoire officiel comme le répertoire SIRENE pour les entreprises et les établissements), de l'**identification indirecte** (qui consiste à reconnaître un individu que l'on sait unique dans l'ensemble de la population). Par exemple, des données individuelles relatives à une entreprise française produisant de l'électricité et réalisant un chiffre d'affaires de plus de dix milliards d'€ conduisent très facilement à l'identification de cette entreprise. Il est très difficile d'assurer scientifiquement que des variables ne permettent pas l'identification indirecte d'informations individuelles, car pour cela il est nécessaire de connaître toutes les informations individuelles publiées en dehors de la sphère statistique. Un concept assez souvent utilisé par les méthodologues de la confidentialité est le « coût » de l'identification des données individuelles, et le rapport en le coût de l'identification et l'avantage espéré de l'information individuelle ainsi obtenue.

Les méthodes envisageables sont l'anonymisation, la réduction de la finesse de l'information, la suppression, la perturbation, la permutation, la microagrégation.

7.2.1. L'anonymisation

Elle consiste à retirer les variables d'identification (nom, adresse) de l'unité statistique et celles de quasi-identification pour ne conserver que les variables pour lesquelles l'analyse doit être conduite. Cette méthode s'applique beaucoup mieux aux unités personnes physiques dans le cadre de la statistique sociale qu'aux entreprises. En effet, beaucoup des variables d'analyse économique nécessaires à toute enquête sur la sphère productive sont des variables de quasi-identification. En connaissant le secteur d'activité et le nombre d'employés (ou le chiffre d'affaires), les plus grandes entreprises françaises peuvent être identifiées. On pourrait toutefois envisager d'utiliser cette méthode sur des sous-populations comme les petites entreprises, l'artisanat, etc....

La grande difficulté de cette méthode consiste à maîtriser le risque de divulgation d'information. En effet, comment s'assurer que l'on connaît toutes les variables de quasi-identification ? En toute rigueur, il faudrait être en possession d'un recensement de la population, et effectuer tous les comptages correspondant à toutes les possibilités de croisement de modalités de toutes les variables susceptibles de servir de variables de quasi-identification, afin de s'assurer qu'il y a toujours plus d'un individu pour chaque combinaison de modalités. Or ce type d'opération donne un nombre tellement important de possibilités dès que la taille de la population initiale et le nombre de variables deviennent réalistes, que les calculs numériques ne sont plus possibles.

Il faut également noter que pour certains travaux, les chercheurs doivent pouvoir apparier des données anonymisées, par exemple pour constituer des panels par adjonction de vagues périodiques de fichiers anonymisés. On parle alors d'« appariement sécurisé ».

7.2.2. La réduction de finesse

L'idée est de réduire le nombre de modalités et de transformer les variables quantitatives de quasi-identification afin d'aboutir à un nombre de combinaison de ces variables suffisamment peu important pour que l'on soit assuré qu'il n'y a plus d'unité statistique concernée par la confidentialité primaire par combinaison de variable. Clairement, le prix à payer est une perte d'information.

7.2.3. La suppression

Dans certains cas extrêmes, on peut décider de ne pas diffuser les enregistrements relatifs à des unités atypiques de la population. Les informations ainsi obtenues du fichier seront biaisées, mais on estime préférable cette perte d'information à celle qui serait liée au remplacement des variables quantitatives pour lesquelles des extrema sont atteints par des variables qualitatives.

7.2.4. La perturbation

En ce qui concerne les variables quantitatives, un moyen d'empêcher l'identification est d'ajouter aux variables qui permettraient l'identification un terme perturbateur positif ou négatif, qui suit une variable aléatoire de moyenne nulle et de variance choisie de façon appropriée. On voit très clairement dans cette méthode que plus le terme perturbateur ajouté est important (variance élevée), plus le risque d'identification est faible.

7.2.5. La permutation

Pour des variables qualitatives (codes), la méthode spontanée est d'échanger les valeurs prises par certains codes entre différents enregistrements individuels, et cela de façon contrôlée. Les contrôles sont les statistiques sur chaque variable prise indépendamment qui sont choisis égales à celles du fichier avant permutation. Parmi ces statistiques, on peut choisir non seulement les statistiques calculées séparément sur chacune des variables, mais aussi des statistiques relatives à la liaison entre les variables (coefficients de corrélation, etc.). Ces méthodes font appel à des algorithmes assez complexes.

7.2.6. La micro-agrégation

La méthode consiste à remplacer des unités individuelles par des unités représentatives, en plus petit nombre, mais permettant de conserver l'essentiel de l'information initiale sur la population considérée. Chacune des unités représentatives est en quelque sorte un résumé local au sein de la population, fondé sur des « moyennes » (respectivement « médianes ») pour les variables quantitatives (respectivement qualitatives). Il existe un lien conceptuel avec les méthodes d'analyse des données. Cette méthode a été développée et promue à Eurostat³, afin de proposer des pistes alternatives aux règles très strictes imposées au niveau européen sur l'accès aux données individuelles.

D'une façon générale, toutes les méthodes fondées sur la perturbation ou la réduction d'informations individuelles modifient les propriétés statistiques par rapport à la population initiale. Elles permettent néanmoins d'approximer une analyse des populations observées. Il est donc nécessaire de valider sur les données initiales les conclusions statistiques tirées des données transformées. Des travaux de simulation sur la robustesse des différentes méthodes ont été réalisés.

7.3. Logiciels de protection des données confidentielles

Comme les méthodes statistiques en général, les méthodes de protection de la confidentialité ne sont utilisables qu'à l'aide des programmes informatiques correspondants. Les principaux produits sur le marché ont été développés par des Offices Statistiques Publics ou en sont dérivés. On peut citer USBSCUP, du US Bureau of the Census, ACSSuprs, dérivé du logiciel CONFID de Statistics Canada, ARGUS, développé par un consortium européen fédéré par Statistics Netherlands et financé par le 5^{ème} programme cadre de recherche européen et GHQUAR, développé par l'office statistique régional allemand de Nordrhein-Westfalen. Ces logiciels traitent essentiellement de la confidentialité par suppression, sauf l'option Argus-μ qui traite de l'anonymisation de données individuelles § 7.2.1.

Pour une étude comparative de ces produits, on pourra se référer à l'article « *Transferable software for automated secondary cell suppression* », Sarah Giessing, (Statistisches Bundesamt), Research in Official Statistics, 1/2000, pp 119-132.⁴ Comme pour tous les logiciels, la comparaison de ces produits implique de prendre en compte les méthodes qu'ils utilisent, leur portabilité, l'accessibilité de leur documentation (en particulier la/ les langue(s) disponibles), leur rapidité de calcul.

Remarque : on notera qu'on n'a pas traité ici des problèmes de sécurité (physique et logique) dans l'accès aux données. En particulier, les enjeux liés au choix des identifiants, aux méthodes de chiffrement des données avant transport ou transmission entre sites protégés n'ont pas été abordés ici, même s'ils sont une des conditions premières de la protection effective des données individuelles. Dans le cas des « appariements sécurisés » mentionné ci-dessus, il est important de conserver une « formule » de passage entre les identifiants réels des unités statistiques observées, qui sont des identifiants officiels dans un répertoire, et les identifiants ad hoc utilisés pour les fichiers anonymisés. De telles formules de passage revêtent évidemment une grande importance et doivent être totalement inaccessibles aux tentatives de ré-identification des unités individuelles. Des méthodes de chiffrement sophistiquées peuvent alors être utilisées (méthodes dites de « hachage »). De telles méthodes sont occasionnellement par l'INSEE en statistiques d'entreprises (TPE, entreprises individuelles, artisanat).

³ - voir Defays, D and Nanopoulos, P. "Panels of enterprises and confidentiality : the small aggregates method", Proceedings of Statistics Canada Symposium 1992, Design and Analysis of Longitudinal Surveys, November 1992, pp195-204.

⁴ - consultable sur <http://europa.eu.int/comm/eurostat/>

8 - CONCLUSION

Ecrit notamment dans la perspective de la célébration par le CNIS du 50^e anniversaire de l'adoption de la loi du 7 juin 1951, aujourd'hui encore loi fondamentale de la statistique publique française, ce rapport s'efforce de faire le point, dans un certain détail, de la situation actuelle du secret statistique concernant les entreprises (au moment où, à l'entrée du troisième millénaire de notre civilisation, explose la société mondiale de l'information) et de ses perspectives d'évolution (alors que nos concitoyens s'inquiètent de plus en plus des dangers potentiels résultant de l'accroissement incessant des capacités de stockage, de transmission et de traitement des données individuelles que permettent les nouvelles technologies).

C'est pourquoi, s'il convient d'explorer les possibilités d'amélioration de l'accès aux données individuelles en faveur de la recherche scientifique, voire dans certains cas en faveur de l'ensemble de la société, il convient aussi de ne modifier que très prudemment l'équilibre, très sage, aujourd'hui établi entre la transparence et le secret, en considérant attentivement les avantages et les inconvénients potentiels liés à chaque modification, et en prenant en compte les pratiques existantes tant dans chacun des autres Etats de l'Union européenne, qu'au niveau de l'Union elle-même, et en tenant compte également des autres pays de l'OCDE.

Il convient également de rendre plus visibles les avancées dans la connaissance rendues possibles par les habilitations d'accès résultant des avis favorables rendus par le comité du secret statistique concernant les entreprises en archivant systématiquement les retours émanant des organismes habilités dans un endroit accessible à tous les intéressés.

