



**Conseil Economique
et Social**

Distr.
GENERALE

CES/2003/11
20 mai 2003

FRANCAIS¹

COMMISSION DE STATISTIQUE et COMMISSION ECONOMIQUE POUR L'EUROPE
CONFERENCE DES STATISTICIENS EUROPEENS

Cinquante et unième réunion plénière
(Genève, 10-12 juin 2003)

**LEGISLATION RECENTE DE L'UE CONCERNANT L'ACCES AUX
DONNEES CONFIDENTIELLES A DES FINS DE RECHERCHE — MISE EN
OEUVRE ET REPERCUSSIONS**

Document soumis par Eurostat²

RESUME

Le présent document décrit le contexte dans lequel s'inscrit un nouveau règlement (831/2002 concernant l'accès aux données confidentielles à des fins scientifiques) de même que les principales dispositions de celui-ci. Les conséquences du règlement sur Eurostat, les États membres, la communauté scientifique et même les personnes concernées sont également étudiées. De plus, le document met en évidence les activités d'Eurostat dans la mise en œuvre du règlement et fournit des indications sur certaines questions en suspens. L'application du règlement a par ailleurs donné lieu à des interrogations sur des sujets aussi divers que la possibilité d'un accès à distance aux données confidentielles et sur la signification de l'expression «à des fins scientifiques».

¹ Le document original a été soumis en anglais; la traduction en français est fournie par Eurostat.

² Pedro Diaz Muñoz (email: pedro.diaz@cec.eu.int), Commission européenne, L-2920 Luxembourg. John King (email: johnrbking@waitrose.com).

I. LEGISLATION RECENTE DE L'UE CONCERNANT L'ACCES AUX DONNEES CONFIDENTIELLES A DES FINS DE RECHERCHE — MISE EN OEUVRE ET

1. Il y a environ un an, la Commission européenne a adopté le règlement 831/2002 concernant l'accès aux données confidentielles à des fins scientifiques, ce qui a permis de progresser sensiblement dans ce domaine. Le présent document apporte des éclaircissements sur le contexte dans lequel s'inscrit le règlement, expose les dispositions de celui-ci et met en évidence les mesures qu'Eurostat a prises pour l'appliquer. Il étudie en outre les répercussions de ces travaux et renvoie à certaines questions qui en découlent.

II. CONTEXTE DU REGLEMENT 831/2002

2. Les ensembles de micro-données deviennent de plus en plus importants, car l'accès à ceux-ci revêt un intérêt croissant pour les chercheurs. Cet intérêt s'explique par deux raisons dont la première est liée aux exigences de la vie moderne, à savoir une administration responsable et la transparence des pratiques. Cet aspect des choses se traduit par un intérêt croissant pour les politiques basées sur des éléments tangibles, pour l'analyse politique ainsi que pour le suivi des politiques et de leurs répercussions. Les activités de ce type exigent des informations actuelles détaillées et nécessitent fréquemment des analyses plus fines que celles qui sont aujourd'hui publiées par les instituts de statistique. Parfois, ces analyses sont considérées comme des activités extérieures au domaine de compétence des instituts nationaux de statistique (INS) ou même comme des activités susceptibles de porter préjudice à l'indépendance supposée des INS. En effet, ces analyses sont souvent réalisées par des instituts universitaires ou des organismes de recherche indépendants.

3. Des données plus détaillées sont requises pour ce type d'analyses. Les données collectées et stockées par les INS apparaissent souvent comme la source la plus évidente et la plus pertinente à consulter. Aussi les INS et d'autres instituts statistiques sont-ils soumis à une pression croissante en vue de la fourniture de données détaillées sur des thèmes très divers. Au niveau de l'Union européenne (UE) en particulier, les analyses «pan-UE» revêtent une importance croissante. Cette remarque s'applique également dans le cas de la zone euro. L'accès aux ensembles de données «pan-UE» est donc très recherché. Eurostat détient de nombreux ensembles de données de ce type, si bien qu'Eurostat est considéré - par analogie avec la situation au niveau national - comme la source potentielle la plus naturelle, la plus simple et la plus directe pour ces ensembles de données.

4. La seconde raison correspond à la nature changeante de la recherche elle-même. Souvent, la recherche moderne ne se satisfait pas de valeurs agrégées: des micro-données sont requises pour les analyses affinées et la modélisation. Ce phénomène s'est accompagné d'une évolution (ou peut-être plutôt d'une révolution) des capacités informatiques - tant en ce qui concerne le matériel que les logiciels - ainsi que du nombre de chercheurs et d'instituts de recherche. Ces facteurs ont fait augmenter considérablement le nombre de demandes d'accès à des fichiers de micro-données en vue de calculer des matrices de corrélation, d'estimer des modèles et d'effectuer d'autres analyses, suivant le contexte des travaux concernés.

5. Des exemples concernant les besoins de données de la communauté scientifique ont été cités dans

la presse, notamment par Westergaard-Nielsen et Blundell lors du 19e séminaire du CEIES (Comité consultatif européen sur l'information statistique dans les sphères économique et sociale) qui s'est tenu en septembre dernier à Lisbonne et qui a été consacré aux « Solutions innovatrices en matière d'accès aux micro-données ». D'autres exemples ont également été cités, notamment par Dilnot, Vickers et Blundell en décembre 2001, à l'occasion de la conférence inaugurale du centre de recherche *cemmap* de Londres (*Centre for microdata methods and practice*).

6. Dans le même temps, du côté de l'offre, les instituts de statistique - tant les INS que les instituts supranationaux et internationaux - jugent de plus en plus qu'une exploitation renforcée des données qu'ils détiennent représente une contribution à la vie de la société et fait partie intégrante de l'obligation de mieux valoriser leurs ressources (données). Il y a toutefois des limites à ce que peuvent faire les organisations statistiques, en particulier les INS, et à la façon dont elles peuvent le faire. C'est pourquoi les chercheurs et les instituts de recherche ont un rôle important et de plus en plus déterminant à jouer.

7. Du fait de sa mission, qui consiste à élaborer des statistiques pour les besoins de l'Union européenne, Eurostat collecte des données auprès des États membres (ÉM) sur de nombreux aspects de la vie économique et sociale. Ces ensembles de données sont globalement comparables entre ÉM et sont basés sur des définitions harmonisées. Les ensembles de données d'Eurostat représentent une source statistique riche et utile pour la Commission, les ÉM et - potentiellement - les chercheurs. Les données collectées et stockées par Eurostat sont soumises à des règlements qui constituent des accords entre la Commission et les ÉM quant aux besoins auxquels les données sont censées répondre et aux conditions dans lesquelles les données sont fournies (en substance, il s'agit de définir ce qu'il est permis de faire des données et ce qui est interdit). Les données sont maintenues en fonction des exigences et des conditions imposées par les ÉM, c'est ce qui est indiqué explicitement dans certains de ces règlements.

8. Le principe de secret statistique est le contrat qui met effectivement en relation le statisticien avec les personnes qui fournissent des données à caractère personnel, que ce soit volontairement - comme c'est souvent le cas - ou par obligation légale, en vue de l'élaboration de statistiques essentielles à la société dans son ensemble. Du point de vue purement légal, la plupart des pays européens ont arrêté des dispositions relatives à la confidentialité statistique il y a déjà longtemps. Au niveau européen, c'est l'article 285 du traité instituant la Communauté européenne qui traite de ce principe fondamental pour les statistiques communautaires: il dispose ainsi que l'établissement de statistiques se fait dans le respect de l'impartialité, de la fiabilité, de l'objectivité, de l'indépendance scientifique, de l'efficacité au regard du coût et de la confidentialité des informations statistiques. Le principe de confidentialité fait donc partie intégrante de la charte européenne de base et a ainsi acquis le statut le plus élevé qu'il y ait au plan légal.

9. Ce principe a été précisé encore davantage et les données reçues, stockées, exploitées et diffusées par Eurostat sont contrôlées en vertu d'un ensemble de textes élaborés depuis la mise au point du traité instituant la Communauté européenne. En 1990, le Conseil a adopté le règlement 1588/90 relatif à la transmission à l'Office statistique des Communautés européennes de données couvertes par le secret: ce règlement fixe des règles et des garanties de base pour le traitement des données confidentielles. En 1997, ces dispositions ont été étendues dans le cadre de la «loi statistique» (règlement 322/97 relatif à la statistique communautaire) qui contient notamment une définition du secret statistique à l'article 13:

« 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 individuelles.

Pour déterminer si une unité statistique est identifiable, il est tenu compte de tous les moyens qui pourraient raisonnablement être utilisés par un tiers pour identifier ladite unité statistique.

10. Cette définition a remplacé celle qui figurait dans le règlement 1588/90 selon laquelle les données confidentielles correspondaient aux «données déclarées confidentielles par les États membres en vertu des législations ou pratiques nationales en matière de secret statistique». La notion de données confidentielles est donc devenue une notion objective comportant une dimension communautaire bien établie.

11. L'article 13 dispose en outre:

«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»

12. La loi statistique stipule également que les données confidentielles doivent être exploitées uniquement à des fins statistiques, à moins que les répondants n'aient sans équivoque donné leur consentement à ce que les données soient utilisées à d'autres fins (article 15). La loi prévoit de plus l'accès aux données confidentielles à des fins scientifiques (article 17).

13. Moyennant l'accord de l'ensemble des États membres, cette dernière disposition a été utilisée afin d'accéder simplement aux données du panel communautaire des ménages (PCM). Un ensemble de micro-données anonymisées a été élaboré (par Eurostat, en coopération avec les États membres) et peut être consulté par les chercheurs sous certaines conditions.

14. En outre, l'article 17 a été mis à profit par plusieurs chercheurs en entreprises qui souhaitent exploiter des micro-données pan-UE pour les besoins de leurs travaux. Ces chercheurs ont dû contacter l'autorité statistique nationale de chaque État membre afin d'obtenir l'autorisation de consulter les données de ce pays. Eurostat est alors habilité à fournir un droit d'accès aux données de l'État membre ayant donné son accord. Cette approche a donné des résultats mitigés qui dépendent en fait du type d'enquête ou de données recherchées (parfois les États membres rejettent les demandes d'accès aux données).

III. FINALITE DU REGLEMENT 831/2002

15. Le règlement 831/2002 met en œuvre certaines dispositions de la loi statistique (règlement 322/97), et en particulier les articles 17, paragraphe 2, et 20, paragraphe 1. En substance, le règlement 831/2002 définit des procédures simplifiées régissant l'accès aux données confidentielles à des fins scientifiques. Le règlement cherche à lever certains des obstacles de la loi statistique, même si l'accès aux données confidentielles reste soumis aux observations des autorités statistiques nationales de chaque État membre et à différentes conditions. Le règlement renvoie à quatre sources importantes:

- le panel communautaire des ménages (PCM);
- l'enquête sur les forces de travail (EFT);
- l'enquête communautaire sur l'innovation (ECI);
- l'enquête communautaire sur la formation professionnelle (ECFP).

16. En résumé, les chercheurs doivent appartenir à des instituts ou des organismes de recherche des États membres (dans le cas contraire, la procédure est plus longue). Une proposition détaillée doit être établie, faisant apparaître l'objectif des travaux de recherche ainsi que des détails sur les données à exploiter. Des garde-fous concernant la sécurité du stockage des données sont nécessaires et des contrôles d'accès sont exigés. L'approbation des conditions et des garanties passe par l'établissement d'un contrat avec l'organisme de recherche. Le règlement ne prévoit aucun droit d'accès aux données confidentielles. En outre, Les États membres peuvent refuser l'accès à leurs données pour toute demande de recherche. La consultation des données confidentielles peut se faire dans les locaux d'Eurostat, moyennant des vérifications des résultats pour des raisons de confidentialité, ou à travers des distributions d'ensembles de micro-données anonymisées. L'accord des chercheurs sur les conditions et les garanties passe par un contrat avec l'organisme auquel ils appartiennent.

17. Accessoirement, le nouveau règlement 831/2002 fournit une définition légale des ensembles de micro-données anonymisées: par «microdonnées anonymisées», il faut désormais entendre «les fichiers statistiques individuels qui ont été modifiés afin de minimiser, selon les meilleures pratiques actuelles, le risque d'identification des unités statistiques auxquels ils se rapportent».

IV. MISE EN OEUVRE DU REGLEMENT 831/2002 AU SEIN D'EUROSTAT

18. Pour Eurostat, les répercussions du règlement et de sa mise en œuvre sont considérables. Il y a toutefois des précédents et des expériences susceptibles de servir de références. Par exemple, le panel communautaire des ménages (PCM) a déjà montré la direction à suivre - en fournissant initialement des accès contrôlés à des micro-données et, plus récemment, en créant et en permettant de consulter des micro-données anonymisées. Des approches similaires sont en cours d'élaboration et d'extension (aux autres enquêtes mentionnées dans le règlement et à un plus grand nombre de chercheurs).

19. De nouvelles procédures sont mises au point pour la réception des demandes d'accès, pour l'évaluation des chercheurs et de leurs demandes, pour la conclusion de contrats de même que pour la consultation des autorités statistiques des États membres telle qu'elle est exigée par le règlement. De nouveaux contrats types sont actuellement rédigés et des clauses de confidentialité ont été élaborées. Les parties aux contrats seront Eurostat, d'une part, et les instituts ou les organismes de recherche concernés, d'autre part. Aussi doit-il également y avoir une relation contractuelle entre le chercheur et l'institut/l'organisme auquel il appartient. Le règlement ne permet pas aux individus en tant que tels d'accéder à des données confidentielles.

20. Au bout du compte, les dispositifs prévus par le règlement 831/2002 doivent être conviviaux et fournir un service à la communauté scientifique. Pour Eurostat, la consultation des chercheurs quant à leurs

besoins en matière de données et de ressources revêt une grande importance. Parallèlement, Eurostat doit expliquer certaines contraintes à la communauté scientifique et doit s'efforcer de les faire comprendre et accepter. Le fait d'agir en coopération étroite avec les chercheurs, de comprendre leurs besoins et leurs intérêts et d'expliquer les contraintes à respecter, tout cela constitue une activité relativement nouvelle pour Eurostat. Toutefois, le dialogue s'est engagé, des contacts ayant récemment été pris avec le CEIES, l'ESF (Fondation européenne pour la science) et d'autres organismes de recherche internationaux.

21. Il ne s'agit cependant pas d'un territoire entièrement inconnu. Le présent document a déjà fait référence aux attentes et aux besoins des chercheurs. Des ressources sont ainsi été mises à la disposition des chercheurs dans plusieurs États membres et d'autres possibilités pourraient également être citées. L'étude sur les revenus réalisée au Luxembourg constitue un proche exemple de données comparables au niveau international pouvant être consultées à distance par certains chercheurs. Quelques États membres, au nombre desquels figure le Royaume-Uni, ont acquis une vaste expérience dans l'élaboration de micro-données anonymisées pour les besoins des travaux de recherche des universités ou des instituts spécialisés. Aux États-Unis, il convient de passer par les centres de recherche de données du «Census Bureau» pour consulter des données confidentielles. Toutefois, ce type d'accès n'est pas commun à tous les pays et l'on note des différences dans les pratiques, les attentes, les cultures et les cadres législatifs.

22. L'article 3 du règlement 831/2002 prévoit une procédure assez directe pour les demandes de données émanant de chercheurs appartenant à deux catégories d'organismes:

- «a) les universités et les autres établissements d'enseignement supérieur établis au titre du droit communautaire ou du droit d'un État membre;
- b) les organisations ou les institutions de recherche scientifique établies au titre du droit communautaire ou du droit d'un État membre».

23. En ce qui concerne les «autres établissements», l'article 3 du règlement 831/2002 dispose qu'ils doivent être préalablement approuvés par le comité «Secret statistique» avant de pouvoir soumettre des demandes de consultation de données confidentielles à des fins de recherche scientifique. Par «autres établissements» il faut entendre tous ceux qui sont indiqués à l'article 3, paragraphe 1, point c), du règlement. Il s'agit essentiellement des organismes qui ne relèvent pas des catégories visées au paragraphe 1, points a) et b), et qui n'ont pas été mandatés par les services de la Commission ou les administrations des États membres pour réaliser des travaux de recherche spécifiques.

24. En tant que tel, le règlement 831/2002 ne précise pas les critères que le comité doit prendre en considération pour formuler un avis. Certains facteurs pertinents sont toutefois indiqués dans ce règlement et dans le règlement 322/97, à savoir:

- la prévention de tout usage des données à des fins non statistiques (règlement 322/97, articles 10 et 18, et règlement 831/2002, article 8, paragraphe 1);
- l'accès aux données à des fins de recherche scientifique (règlement 322/97, article 17, et règlement 831/2002, article 1);
- la protection des données (règlement 831/2002, article 8, paragraphe 1).

25. Compte tenu des principes de transparence et d'équité, il convient en outre que les critères soient clairs et bien connus.

26. Le comité «Secret statistique» a décidé que les facteurs ci-dessous devaient être pris en considération lors de la formulation de ses avis:

- la finalité première de l'organisme;
- les modalités d'organisation de la recherche au sein de l'organisme;
- les garde-fous mis en place par l'organisme;
- les dispositions relatives à la diffusion des résultats des travaux de recherche.

27. Eurostat transpose actuellement ces conditions et ces facteurs en procédures opérationnelles. Par exemple, la question préalable de «l'habilitation» d'un organisme à formuler une demande (indépendamment des mérites de la demande elle-même) est réglée par une série de questions (questionnaire) permettant:

- d'identifier l'organisme et sa finalité première;
- de décrire succinctement le(s) projet(s) de recherche;
- de préciser les modalités d'organisation de la recherche au sein de l'organisme ainsi que les moyens financiers existants;
- d'indiquer les mesures de sécurité prises par l'organisme;
- de spécifier les dispositions relatives à la diffusion des résultats des travaux de recherche.

28. Ces informations seront transmises à l'autorité statistique nationale de chaque INS pour avis, ce qui se fera sans doute par une procédure écrite afin d'accélérer le processus.

29. Le règlement prévoit la consultation des données confidentielles par des chercheurs dans les locaux d'Eurostat. Il prévoit une consultation similaire dans les locaux des autorités statistiques nationales des États membres si le niveau des dispositifs de sécurité et de vérification est identique à celui d'Eurostat. Pour ce type de consultation, on parle souvent d'accès «contrôlé» ou d'accès par un «centre sécurisé».

V. REPERCUSSIONS POUR LES ETATS MEMBRES ET NES INSTITUTS NATIONAUX DE STATISTIQUE

30. Le règlement encourage les INS des États membres ainsi qu'Eurostat à travailler en coopération étroite à l'élaboration d'un système permettant d'accéder à des données confidentielles pour les besoins de recherches scientifiques. Le champ des activités concernées est vaste puisqu'il convient de s'accorder sur les méthodes de vérification et de protection des résultats de la recherche, de mettre au point des garde-fous et des moyens de contrôle ainsi que des méthodes permettant de créer des micro-données anonymisées et, enfin, d'arrêter des procédures pour le traitement des demandes de recherche et la

consultation réciproque. Tous ces éléments sont en cours d'élaboration et seront discutés avec les États membres. Les garde-fous, les contrôles et les vérifications seront basés sur des méthodes et des approches existantes et refléteront les pratiques nationales en vigueur moyennant quelques adaptations. Supposons par exemple qu'un État membre ait une procédure bien définie pour étudier les demandes de recherche plusieurs fois par an. Le règlement dispose toutefois que tout État membre doit répondre à une demande de recherche dans un délai de six semaines. Prenons un autre exemple: un État membre applique une certaine procédure pour approuver les demandes de recherche émanant de chercheurs et d'organismes de recherche de ce pays. Là encore, le règlement prévoit la possibilité d'une consultation non seulement pour les chercheurs des autres États membres, mais aussi ceux de pays tiers.

31. Même si le règlement dispose que tout État membre doit être informé des demandes de données qui le concernent, il suppose que les États membres donneront accès à leurs données, pour autant que les chercheurs respectent l'ensemble des conditions et des exigences prévues.

32. La façon dont les données sont collectées peut avoir des répercussions sur les INS. En particulier, si l'usage qui peut être fait des données doit être indiqué au répondant, le type de recherche envisagé par le règlement devra peut-être y figurer. Ce point est discuté ci-dessous. Chaque INS doit en outre élaborer des procédures visant à anonymiser les données et à protéger les résultats contre un accès direct aux données confidentielles et doit faire agréer ces procédures par Eurostat. En pratique, une approche commune de l'ensemble des INS permettrait d'obtenir une meilleure protection ainsi que des ensembles de données plus utiles. Certains domaines nécessiteront davantage de recherches et de réflexion, car ils ne sont encore ni très développés ni très bien compris. Au nombre de ceux-ci figurent des problèmes liés au potentiel de divulgation des résultats de modélisation. En ce qui concerne les données tabulaires, ces problèmes sont bien compris - une méthodologie existe d'ailleurs en la matière et elle continue à être développée - mais la situation est moins claire, surtout pour ce qui est des solutions, en ce qui concerne la modélisation. L'une des restrictions à mettre en œuvre consiste à supprimer les informations sur les résidus - même si elles revêtent un grand intérêt statistique pour les chercheurs - car elles apportent des informations sur des valeurs atypiques qui correspondent souvent aux unités rares. Il est également nécessaire de mieux comprendre le potentiel de divulgation des estimations de paramètres, en particulier lorsque des modèles similaires sont exploités et comparés.

VI. REPERCUSSIONS DU REGLEMENT 831/2002 SUR LA COMMUNAUTE SCIENTIFIQUE

33. Les répercussions du nouveau règlement pour la communauté scientifique illustrent la nature du partenariat entre les organisations statistiques et la communauté des chercheurs. D'une part, le règlement ouvre la voie à de nouvelles opportunités et, d'autre part, il impose des restrictions et des limitations strictes en échange de ces ouvertures.

34. La communauté scientifique doit tout d'abord accepter qu'elle n'a pas de droit d'accès. Elle doit également reconnaître qu'elle aura la responsabilité de maintenir et de protéger la confidentialité des données auxquelles elle a accès. Les restrictions et les garde-fous sont peut-être plus stricts que ceux qui existent dans les organismes de recherche ou pour d'autres séries de données, mais il convient de les respecter. La documentation des centres de recherche de données (RDC) du «Census Bureau» des

États-Unis est volumineuse, mais elle est aussi exhaustive. Les chapitres consacrés aux différences culturelles entre les RDC et les universités sont particulièrement intéressants. Ils rappellent en outre qu'il n'y a pas lieu de se baser sur l'hypothèse d'une culture ou d'une finalité commune. Les chercheurs devront également accepter qu'un autre organisme encore soit habilité à leur poser des questions précises, pas seulement sur les recherches menées et leur objectif, mais aussi - dans le cas des micro-données anonymisées - sur le mode de stockage des données et les contrôles d'accès. Il leur faudra enfin accepter que ces réponses soient transmises à l'INS de chaque État membre pour étude. En outre, en cas de consultation de données confidentielles, les résultats prévisibles doivent être soumis à une vérification avant publication ou diffusion.

35. En échange, les chercheurs bénéficieront d'un accès simplifié aux données couvrant l'ensemble des États membres. Jusqu'à présent, les dispositions de la loi statistique étaient telles que la consultation des données de chaque État membre était soumise à une procédure de longue durée visant à obtenir l'approbation de tous les États membres. Maintenant, les chercheurs auront la possibilité de faire des recherches et des analyses au niveau pan-européen. Le règlement couvre quatre ensembles de données importants et d'autres données devraient également devenir accessibles en temps utile.

VII. REPERCUSSIONS DU REGLEMENT 831/2002 POUR LES PERSONNES CONCERNEES

36. Même si le règlement 831/2002 vise à améliorer l'accès des chercheurs aux données, il a aussi des répercussions pour les personnes qui fournissent les données de départ. Ces informations sont transmises aux instituts statistiques du pays concerné, dans le cadre d'une enquête facultative ou obligatoire. Elles peuvent aussi être tirées de répertoires administratifs dans le cadre de travaux statistiques. L'institut statistique transmet les données à Eurostat après avoir supprimé toute donnée permettant d'identifier les personnes concernées. Ce processus a en outre d'autres conséquences pour les instituts de statistique. Le principe qui sous-tend la collecte de données est celui du *consentement informé*, selon lequel la personne concernée a le droit de savoir à quelles fins ses données seront utilisées et à qui elles seront transmises. Si l'on ajoute une nouvelle dimension – nouveaux utilisateurs, nouveaux besoins – il convient d'en informer les personnes concernées. Dans certains États membres, il sera peut-être nécessaire de modifier la législation qui régit la collecte des données afin de préciser à quelles fins les données peuvent désormais être collectées.

37. Dans le cadre d'une enquête statistique, les personnes concernées doivent être informées que les données fournies seront utilisées à des fins statistiques, y compris - le cas échéant - des travaux de recherche réalisés par des organismes externes en plus des tâches habituelles de l'institut de statistique. En vertu du règlement, les chercheurs peuvent maintenant appartenir à des institutions situées dans les États membres ou dans des pays tiers, et pas seulement dans le pays de la personne concernée. À l'heure actuelle, les pratiques varient d'un État membre à l'autre, de sorte qu'il n'est pas facile de déterminer ce qui devra changer. En fait, il s'agira peut-être de peu de choses, car les formes actuelles de consentement sont susceptibles de couvrir implicitement l'accès des chercheurs aux données d'un autre pays.

38. Le tout est une question d'équilibre, puisqu'il y a lieu de peser la valeur explicite-implicite de l'aspect «informé» du consentement. Des travaux sur le terrain seront peut-être nécessaires pour répondre à

cette question, y compris des recherches qualitatives portant sur les personnes concernées. Cette question joue un rôle important dans le contrat entre la personne concernée et l'institut de statistique et sera considéré par ce dernier comme un élément ayant un effet sur les taux de réponse aux enquêtes facultatives.

VIII. PRECISIONS SUPPLEMENTAIRES

39. *«Que faut-il entendre par «à des fins scientifiques» ?* Cette question s'est déjà posé. Pour certains, le qualificatif scientifique est synonyme d'universitaire. Mais dans ce cas aussi, que couvre-t-il exactement ? On peut supposer que les chercheurs reconnus de niveau post-doctoral réalisent des travaux scientifiques (même si une certaine partie peut être de nature «commerciale»). En dessous de ce niveau de recherche bien défini, les distinctions sont plus difficiles à réaliser. Et le débat tend à mettre l'accent sur les qualifications ou le statut du chercheur plutôt que sur la nature réellement scientifique des travaux proposés. Qu'en est-il dans ce cas des étudiants en doctorat ? Une grande partie des travaux réalisés par ces étudiants sont à la pointe des connaissances scientifiques, et donc a priori de nature scientifique. En outre, les recherches effectuées pour l'obtention de diplômes de type «maîtrise» sont souvent supervisées par un chercheur/scientifique et peuvent faire partie intégrante d'un projet plus vaste à finalité clairement scientifique. Les étudiants de premier cycle doivent-ils avoir accès aux données à des «fins scientifiques» et pour se familiariser avec les grands ensembles de données ? Après tout, la formation peut être considérée comme un besoin scientifique. L'argument de la formation s'applique également à des niveaux plus élevés et la frontière est difficile à déterminer dans ce domaine. Par souci de pragmatisme, on pourrait parfaitement déterminer la limite sur la base d'arguments juridiques et non scientifiques, à savoir: la personne désirant consulter les données est-elle liée par contrat à l'institut, de sorte que des dommages-intérêts pourraient être exigés en cas de non-respect des conditions d'accès ?

40. *Accès à distance.* Pour certains, le concept de «centre sécurisé» physique est déjà dépassé – en raison des nouvelles technologies. Cette option comporte plusieurs inconvénients: le coût, la difficulté d'accès (le Luxembourg n'est pas spécialement bon marché et il n'est pas particulièrement facile d'accès), le décalage entre le lancement du programme et l'obtention de résultats, le manque de spontanéité des analyses et l'absence de convivialité du système. Aussi de nombreux chercheurs semblent-ils disposés à se passer d'un accès à des données confidentielles dans le cadre d'un centre sécurisé au bénéfice d'ensembles de micro-données anonymisées, directement consultables depuis leur bureau.

41. Plusieurs systèmes d'accès à distance ont toutefois bien fonctionné. L'enquête luxembourgeoise sur les revenus (ELR) en est le meilleur exemple. En utilisant l'approche et les logiciels propres à l'ELR, un autre groupe de chercheurs – travaillant pour le projet « PiEP » (inégalités de paiement et productivité) - a développé des procédures d'accès à distance à l'enquête sur la structure des rémunérations réalisée par Eurostat. Dans les deux cas, des concessions doivent être faites pour obtenir l'accès aux données. L'ELR est basée sur un ensemble de données fortement réduit composé d'un assez petit nombre de variables clés dont certaines sont des variables catégorielles. Le projet PiEP accepte certaines réductions de l'ensemble de données et certaines restrictions de résultats – pas de tabulations, pas d'informations sur les résidus, etc. Ces restrictions visent, avec l'accord des INS des États membres concernés, à limiter à un niveau minimal le risque d'identification ou de divulgation.

42. Le défi consiste à fournir un accès à distance aux chercheurs. La question est de savoir si le

règlement peut être interprété de façon à permettre ce type d'accès. Dans le cadre d'un tel système, le traitement et l'analyse des données se ferait dans les locaux d'Eurostat. Les contrôles en vigueur – pour les individus, les autorisations d'accès et les résultats – pourraient être les mêmes que pour un centre sécurisé, mais l'accès aux données serait beaucoup plus aisé. Toutefois, cette question doit être étudiée de manière plus approfondie avant d'être mise à l'ordre du jour du comité «Secret statistique».

43. *Valeur des travaux de recherche pour le fournisseur de données.* Le «US Census Bureau» pose explicitement comme condition préalable que les travaux de recherche proposés doivent revêtir un intérêt pour le Bureau – en effet, les travaux de recherche doivent être utiles aux programmes du Bureau (c'est une exigence légale). Le projet de protocole discuté au Royaume Uni contient une formule similaire, mais les fondements juridiques sur lesquels elle sera basée ne sont pas clairs. Le règlement 831/2002 ne prévoit pas de telle exigence.

44. *Les données anonymisées sont-elles des données confidentielles ?* Si la procédure d'anonymisation permet de réduire au minimum les risques d'identification ou de divulgation, les données anonymisées sont-elles «confidentielles» ? Cette question semble faire intervenir l'approche légale, basée sur la classification, et l'approche statistique, de nature pragmatique. Si une identification directe n'est pas possible et que le risque d'identification indirecte est négligeable (minimisé conformément aux meilleures pratiques actuelles) – données anonymisées – alors les données ne sont pas de nature à divulguer quoi que ce soit et ne sont donc pas confidentielles. L'hypothèse inverse consisterait à penser que ces données sont toujours confidentielles indépendamment de l'étendue des modifications qui y ont été apportées.

IX. L'AVENIR

45. L'article 13, paragraphe 2, de la loi statistique porte sur un aspect de l'obligation légale faite aux INS et aux États membres qui mérite d'être étudié de plus près: «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 ». Ce paragraphe représente un problème très épineux du point de vue de la confidentialité statistique, puisque les données obtenues par un INS lors d'une enquête sont traitées comme des informations confidentielles même si elles sont accessibles au public et même si la personne concernée fournit l'information. Dans certains pays, il existe une échappatoire pour les INS – si les personnes concernées libèrent l'INS de l'exigence de confidentialité. Sans cette possibilité, des efforts considérables (essentiellement économiques) doivent être déployés pour protéger les données statistiques accessibles au public. Peut-être la réponse consiste-t-elle à réaliser les enquêtes statistiques en deux parties, portant l'une sur les données accessibles au public, souvent en vertu d'un règlement, et l'autre sur les données confidentielles à protéger.

46. Certaines des exigences et des obligations visées dans les textes législatifs ne sont pas définitives, mais évoluent dans le temps. Aussi les INS et Eurostat sont-ils tenus de revoir ponctuellement leurs pratiques et leurs méthodes. Selon le règlement 831/2002 par exemple, il faut entendre par «micro-données anonymisées » des données «qui ont été modifiés afin de minimiser, selon les meilleures pratiques actuelles, le risque d'identification des unités statistiques auxquels ils se rapportent». Comme la meilleure pratique évolue au fil du temps, les procédures utilisées doivent changer elles aussi. Une fois encore, la loi statistique exige «[qu'il soit tenu] compte de tous les moyens qui pourraient raisonnablement être utilisés par un tiers

pour identifier ladite unité statistique». Les moyens dont disposent les tierces parties évoluent également – accès plus étendu à d'autres bases et matériel ainsi que logiciels plus performants. Le règlement 1588/90 peut également être cité comme exemple. Eurostat doit assurer les mêmes garanties de sécurité que les INS pour le transfert de données des États membres vers Eurostat. L'évolution observée dans ce domaine au sein des États membres devra se refléter sur les procédures d'Eurostat.

X. CONCLUSIONS

47. Des efforts considérables ont été nécessaires pour élaborer le règlement et le faire approuver par les États membres à travers le comité «Secret statistique» et par la Commission. Toutefois, cette approbation n'est qu'un élément dans un processus plus vaste de mise en œuvre et de création de systèmes et de procédures, sans même parler des données elles-mêmes, visant à permettre aux chercheurs d'accéder aux données confidentielles. Ce processus de mise en œuvre a soulevé d'autres questions, tant statistiques que légales, qui doivent également être étudiées.