

16 mai 2007

## Mesure de la diversité et protection des données personnelles

### Les dix recommandations de la CNIL

#### 1. La lutte contre les discriminations est une priorité pour tous.

La Commission Nationale de l'Informatique et des Libertés est consciente, comme toute autre institution publique, de la nécessité de lutter contre les discriminations et contribue, dans le champ de compétences qui est le sien, aux actions menées en ce sens.

Or, pour lutter contre les discriminations, encore faut-il pouvoir les identifier les mesurer. Pour cela, il paraît naturel de procéder à l'observation statistique des différences, de la diversité. Mais alors quelle diversité mesurer ? La diversité sociale, ethnique, religieuse, culturelle... ? Quels critères utiliser pour analyser cette, ces diversité(s) ? Quelles méthodes employer ? Qui peut le faire ?

Cette problématique est complexe et délicate.

Complexe, car comme notre Commission a pu le constater, cette « mesure de la diversité » suscite une véritable « effervescence méthodologique », les chercheurs et statisticiens faisant preuve d'une grande imagination en ce domaine et n'étant pas forcément tous d'accord sur les outils à employer.

Délicate, car elle touche à l'essence même de ce qui fait notre identité, à notre conception de la République, à la façon dont on se perçoit et dont on est perçu par les autres. Délicate également car elle ne saurait en aucune manière remettre en cause le fait que la notion de race n'a aucune valeur scientifique.

Compte tenu des multiples enjeux soulevés par cette question, le Président de la CNIL, Alex TÜRK approuvé par l'ensemble de la Commission, a souhaité constituer, en 2005 un groupe de travail, présidé par Madame Anne Debet, professeur des universités et membre de la CNIL.

A la suite des travaux et auditions menés par ce groupe, notre Commission a émis, en juillet 2005, ses recommandations en matière de mesure de la diversité et de lutte contre les discriminations dans le domaine de l'emploi. Or, ses positions ont été parfois mal comprises, certains reprochant même à la Commission de s'opposer au développement des études et enquêtes portant sur ces sujets ! En outre, il est apparu que ce sujet suscitait de plus en plus de débats et controverses.

## **2. Le groupe de travail**

La CNIL a donc souhaité approfondir et élargir sa réflexion en recueillant le point de vue de l'ensemble des acteurs concernés : chercheurs, statisticiens, organisations syndicales, représentants des grandes religions, mouvements associatifs, personnalités qualifiées, chefs d'entreprises...

Le groupe de travail a ainsi réalisé plus de soixante auditions entre novembre 2006 et février 2007. Parmi ces auditions, celles des 18 et 25 janvier 2007 ont été ouvertes à la presse et diffusées en direct sur Publicsénat. C'était une première.

La CNIL a également souhaité aussi recueillir le point de vue de simples citoyens et a mis en ligne sur son site un questionnaire sur le thème de la mesure de la diversité. Plus de 1000 internautes ont ainsi pu exprimer leurs opinions sur le sujet. Bien entendu, ce « sondage » n'a aucune valeur représentative et n'a d'autre objet que de permettre à chacun d'émettre un avis sur cette question délicate.

A l'issue de ces travaux, notre Commission estime aujourd'hui nécessaire d'exprimer publiquement sa position et de proposer des pistes de réflexion constructives qui permettront, elle l'espère, de faire progresser la connaissance de notre société et, par là même, de mieux lutter contre les discriminations.

Ces recommandations, au nombre de dix, s'adressent donc à l'ensemble des acteurs concernés : aux décideurs politiques, chefs d'entreprises, syndicats, administrations afin de les guider dans la lutte contre les discriminations, mais aussi aux chercheurs et statisticiens dont la CNIL reconnaît pleinement la légitimité de la démarche.

Surtout, ces dix recommandations sont rendues publiques afin que chacun, quelle que soit son origine, comprenne bien les enjeux, connaisse mieux ses droits comme ses devoirs et concoure ainsi à une meilleure protection des données personnelles.

### **Recommandation n°1 : ouvrir plus largement aux chercheurs l'accès aux bases de données statistiques et aux fichiers de gestion**

#### **☞ L'accès aux bases statistiques publiques**

L'INSEE, les services statistiques ministériels et les organismes de recherche publique (CNRS, INED...) disposent de gisements de données considérables qui permettent déjà de constituer une « photographie » fidèle et détaillée de la société française, de ses flux migratoires, de ses pratiques d'intégration ainsi que de ses diversités culturelles, sociales, professionnelles...

**La CNIL estime que les bases des données statistiques publiques doivent être ouvertes plus largement au monde de la recherche dès lors que cet accès s'opère dans des conditions garantissant la protection des données et l'anonymat des personnes. A cet égard, la création de centres d'accès sécurisés, permettant aux chercheurs habilités de disposer des données dans des conditions garantissant la confidentialité de celles-ci, constitue une piste qui doit être approfondie, vu l'intérêt général évident que revêt leur création.**

Il convient aussi de procéder à un réexamen d'ensemble des conditions générales de diffusion des données issues du recensement de la population. A cet égard, des propositions

de modification, allant dans le sens d'une plus grande accessibilité, font actuellement l'objet de réflexions entre la CNIL et l'INSEE.

#### ☞ **L'accès aux fichiers de gestion**

Les fichiers de gestion, qu'il s'agisse des fichiers de personnel, d'élèves, d'étudiants... comportent déjà un certain nombre de données (ex : nom, prénom, nationalité, lieu de naissance, adresse) susceptibles d'être utilisées à des fins statistiques pour mesurer la diversité, suivre les trajectoires et évaluer les politiques de lutte contre les discriminations.

**La CNIL estime que ces fichiers, qui constituent des sources importantes d'information pour les chercheurs, peuvent leur être ouverts plus largement dès lors que ces exploitations s'opèrent dans des conditions et selon une méthodologie rigoureuse garantissant la protection des données et l'anonymat des personnes.**

#### **Recommandation n°2 : utiliser les données « objectives » relatives à l'ascendance des personnes (nationalité et/ou lieu de naissance des parents) dans les enquêtes pour mesurer la diversité**

Dans ses recommandations de 2005, la CNIL a estimé que les données relatives à la nationalité ainsi qu'au lieu de naissance des parents, dès lors qu'elle seraient recueillies directement auprès des personnes concernées, pouvaient être utilisées pour mesurer la diversité au sein des entreprises et, plus largement, dans la société française.

La Commission a examiné dans quelle mesure ces données sur l'ascendance pourraient être utilisées dans les fichiers suivants.

#### ☞ **Dans les fichiers d'enquêtes adossées au recensement et éventuellement dans le recensement**

**La CNIL estime que l'intégration de questions sur la nationalité et le lieu de naissance des parents peut être admise dans le cadre d'enquêtes spécifiques adossées au recensement. De telles questions pourraient aussi être posées dans le cadre du recensement de la population dès lors que toutes précautions méthodologiques auraient été prises pour garantir la protection des données et que l'acceptabilité publique de ces questions aurait été préalablement testée.**

#### ☞ **Dans les fichiers d'enquêtes réalisées au sein des entreprises et des administrations**

**La Commission considère que l'intégration de questions sur la nationalité et le lieu de naissance des parents peut aussi être admise dans le cadre d'enquêtes par questionnaires menées dans les entreprises et les administrations, dès lors que celles-ci s'inscrivent dans le cadre d'un programme national de lutte contre les discriminations. Les modalités de ce programme seraient validées par la HALDE et par une instance d'expertise**

statistique (en lien avec le Conseil National de l'Information Statistique), afin que des précautions méthodologiques nécessaires soient prises pour garantir la protection des données

**Recommandation n°3 : ne pas intégrer dans les fichiers des entreprises et des administrations (personnel et usagers) de données sur l'ascendance des personnes**

La Commission estime que l'intégration, dans les fichiers de gestion des administrations et des entreprises, de données sur l'ascendance des personnes, ne peut être envisagée dans la mesure où elle impliquerait l'enrichissement, obligatoire, des fichiers par des données supplémentaires, sensibles, qui ne serviraient donc pas strictement à la gestion. De surcroît, elle serait susceptible de comporter, au moins aux yeux des personnes concernées, le risque d'une utilisation détournée de ces informations à d'autres fins.

La Commission considère en conséquence qu'il est préférable de privilégier la voie des enquêtes par questionnaires auprès des personnes. Ces enquêtes pourraient, pour certaines, reposer sur la constitution de panels permettant ainsi de suivre les trajectoires des personnes.

**Recommandation n°4 : développer des études sur le « ressenti » des discriminations, incluant le recueil de données sur l'apparence physique des personnes.**

Pour la plupart des personnes auditionnées par la CNIL, il est particulièrement important aujourd'hui de pouvoir mesurer la réalité de la discrimination vécue, ce qui suppose d'abord de demander aux personnes si elles s'estiment victime d'une discrimination, comment elles la subissent au quotidien...

Toutes s'accordent aussi sur la nécessité, en ce domaine, d'une démarche prudente et progressive, fondée sur le volontariat et l'auto déclaration.

La CNIL estime que des enquêtes sur le « ressenti » des discriminations, incluant le recueil d'informations sur les caractéristiques physiques des personnes, pourraient être réalisées dès lors que :

- 1) Celles-ci seraient conduites dans le cadre de la statistique publique ;
- 2) les questionnaires auraient fait l'objet d'une validation scientifique (par exemple par le CNIS ou une autre instance scientifique indépendante : cf. ci-après) ;
- 3) des tests préalables auraient été menés de façon à mesurer le degré d'acceptabilité des questions qui devraient être formulées sous forme de questions ouvertes ;
- 4) le caractère facultatif de ces questions serait expressément rappelé ;
- 5) des précautions méthodologiques rigoureuses auraient été prises (cf. ci-après).

## **Recommandation n°5 : admettre, sous certaines conditions, l'analyse des prénoms et des patronymes pour détecter d'éventuelles pratiques discriminatoires**

Ainsi que la CNIL l'a rappelé en 2005 et comme cela a été souligné à plusieurs reprises lors des auditions, on ne peut établir de lien fiable entre les noms et prénoms et l'appartenance à une catégorie raciale ou ethnique déterminée, qu'il faudrait d'ailleurs au préalable pouvoir définir. Le port de tel ou tel prénom reflète avant tout le choix des parents qui peut certes faire référence indirectement à une appartenance sociale, culturelle ou religieuse, mais qui correspond le plus souvent à un choix personnel (parfois influencé par des phénomènes de mode).

Le fait de porter tel ou tel nom de famille ne peut non plus être considéré comme révélateur de l'appartenance à une catégorie « ethno-raciale » donnée. La francisation des noms, le choix des femmes de porter le nom de leur époux, la transmission du nom du père... sont autant de raisons qui conduisent à rejeter l'utilisation de la méthode patronymique comme critère de mesure de la diversité.

Au delà, une telle méthode, si elle était utilisée aux fins de classement ethnique, comporterait un risque grave de stigmatisation et d'atteinte à l'identité humaine. Reposant sur le postulat, contestable, que tel nom = telle origine ethnique, les annuaires publics, fichiers de gestion... seraient ainsi susceptibles d'être traités, sur ces critères, à des fins notamment discriminatoires.

**La CNIL estime donc que l'analyse des noms et prénoms, aux fins de classement dans des catégories « ethno-raciales » n'est pas pertinente en raison du manque de fiabilité de cette méthode et du risque de stigmatisation qui pourrait néanmoins en découler.**

**En revanche, elle considère que le recours au prénom ainsi que, le cas échéant au nom de famille pour détecter d'éventuelles pratiques discriminatoires dans le parcours scolaire, universitaire ou professionnel, à l'exclusion de tout classement dans des catégories « ethno-raciales », peut constituer un indicateur intéressant sur le plan statistique dès lors que les conditions suivantes sont réunies :**

- **l'analyse de ce seul critère ne saurait être suffisante pour apprécier la discrimination ;** il convient de procéder à une analyse réellement objective (**multi-critères**) des parcours ou des trajectoires de vie **prenant en compte l'ensemble des autres facteurs discriminants** (ex : sexe, âge, lieu de résidence, CSP...) **ou susceptibles d'expliquer la différence de parcours** (ex : diplômes, compétences, ancienneté et expérience professionnelle ... ) ;
- **le recours au prénom et au nom de famille peut constituer un repère pour étudier les trajectoires de populations issues de l'immigration, quand il n'existe pas d'autre source disponible ;**
- **de telles études doivent être conduites selon une méthodologie rigoureuse :**
  - **l'analyse des noms et des prénoms doit seulement permettre un classement de ces données en catégories « potentiellement discriminant » / « non-discriminant » ;**
  - **les fichiers de gestion (fichiers de personnel, fichiers d'élèves, d'étudiants, ...), les annuaires professionnels et d'anciens élèves ne**

peuvent être utilisés qu'à des fins statistiques de suivi des trajectoires et d'évaluation des politiques de lutte contre les discriminations ;

- des mesures doivent être prises pour assurer la confidentialité des données traitées, garantir l'anonymat des résultats et informer les personnes concernées de la finalité de l'étude, des conditions de sa réalisation et de leur droit de s'opposer au traitement de leurs données.

### **Recommandation n°6 : modifier la loi informatique et libertés pour assurer une meilleure protection des données sensibles en garantissant le caractère scientifique des recherches et en harmonisant les procédures de contrôle des fichiers de recherche**

L'article 8 de la loi informatique et libertés interdit le traitement des données sensibles et notamment "les données qui font apparaître directement ou indirectement les origines raciales ou ethniques (...), sauf exceptions prévues par la loi. Ainsi, de tels traitements de données ne sont possibles, sous certaines conditions, qu'avec l'autorisation de la CNIL ou qu'avec le consentement exprès des personnes - auquel cas une simple déclaration suffit auprès de la CNIL qui ne peut alors exercer réellement un contrôle de pertinence de l'enquête.

Cette situation n'est pas satisfaisante. On ne peut en effet considérer que le consentement de la personne constitue réellement une garantie pour elle, en particulier lorsqu'il s'agit d'enquêtes réalisées dans les entreprises, dans la mesure où la relation de travail par principe « hiérarchisée » est de nature à « biaiser » le consentement. En outre, dans ce cas, comment garantir le sérieux des études menées en ce domaine ?

Ne conviendrait-il pas, dans le domaine de la recherche et des statistiques, dès lors qu'il s'agit de recueillir des données sensibles telles que celles sur les origines raciales ou ethniques, de prévoir un régime d'autorisation identique à celui qui existe déjà pour les fichiers de recherche médicale ? Ceux-ci sont soumis à avis d'un comité scientifique puis à autorisation de la CNIL, les personnes concernées disposant du droit de s'opposer au traitement de leurs données .

**Dans un souci d'harmonisation, il pourrait être envisagé de s'inspirer de cette procédure pour l'élargir aux traitements des données sensibles ayant pour fin la recherche et les statistiques (et notamment la mesure de la diversité).** Un comité consultatif, composé de statisticiens, sociologues, démographes, ... pourrait être créé à cet effet. Il apprécierait la pertinence du traitement, de la méthode utilisée, des données collectées, etc. Son avis serait ensuite adressé à la CNIL qui disposerait ainsi de l'expertise nécessaire pour délivrer ou refuser son autorisation.

**La CNIL estime en conséquence qu'il convient d'envisager une modification des dispositions de la loi informatique et libertés et en particulier de ses articles 8 et 25 afin d'assurer une meilleure protection des données personnelles sensibles recueillies et traitées dans le cadre des études sur la mesure de la diversité et des discriminations, garantir le sérieux scientifique de ces études et harmoniser les régimes de formalités préalables applicables aux traitements de données sensibles mis en oeuvre à des fins de recherche.**

## **Recommandation n°7: refuser en l'état la création d'un référentiel national « ethno-racial »**

En 2005, la CNIL, avait émis de fortes réserves sur la création d'une nomenclature nationale de catégories « ethno- raciales », estimant que le principe de la création d'un tel référentiel relevait du Législateur.

**Même si une certaine évolution des esprits se fait jour sur cette question délicate, comme en témoignent la multiplication des débats et des tribunes, les prises de positions de certaines personnalités politiques, les résultats de sondages et enquêtes, force est de constater que sur ce sujet, les auditions n'ont pas permis de dégager un réel consensus en faveur de l'élaboration d'un tel référentiel. Au contraire, les personnes auditionnées sont, dans leur grande majorité, hostiles à une telle nomenclature ou, à tout le moins, pour certaines, partisans d'une démarche prudente et progressive. Dès lors, l'idée d'une classification nationale « ethno-raciale » qui s'imposerait dans toute la statistique publique n'est pas aujourd'hui acquise.**

L'intervention d'un cadre législatif en la matière apparaît, aux yeux des personnes auditionnées, prématurée.

Risques de renforcement des stéréotypes, de stigmatisation, de communautarisme, classification incertaine, non scientifique, réductrice, approximative... autant de raisons, bonnes ou mauvaises, qui expliquent les réticences actuelles et qui justifient une attitude mesurée sur ce sujet.

Toutefois, certains chercheurs et experts, ainsi que les associations de défense des minorités visibles, y sont plutôt favorables. Selon eux, de telles statistiques seraient de nature à favoriser l'intégration nationale et la diminution des discriminations. Si la réalisation de ces études est encouragée, ces associations préconisent que les réponses soient facultatives, favorisant l'auto déclaration et recueillies selon une procédure garantissant l'anonymat. Les associations en question se prononcent d'ailleurs en faveur de l'intégration de questions, facultatives, sur les origines ethniques ou raciales au sein du recensement.

Mais la plupart des chercheurs interrogés sont prudents et même ceux qui seraient le plus enclins à recourir à une classification « ethno-raciale » reconnaissent la nécessité de démarches exploratoires pour tester l'acceptabilité de ce type de questions.

**La CNIL émet de fortes réserves sur la création d'une nomenclature nationale de catégories « ethno- raciales » et estime que la décision de principe de créer une telle nomenclature, si elle devait être utilisée, de façon obligatoire, en particulier pour les statistiques publiques et pour le recensement, appartiendrait au Législateur sous le contrôle du Conseil Constitutionnel.**

## **Recommandation n°8 : développer le recours à des experts, tiers de confiance pour mener les études de mesure de la diversité**

Afin de garantir à la fois le sérieux scientifique des études et la protection de la vie privée des personnes, la Commission recommande aux entreprises, établissements scolaires, universités

et de façon générale à tout organisme qui souhaiterait réaliser en son sein une mesure de la diversité de faire appel à une expertise extérieure indépendante (ex : organisme de recherche) qui soit également un « tiers de confiance ».

**Ce tiers de confiance devrait ainsi s'assurer que :**

- 1) les traitements seront réalisés dans un cadre confidentiel, c'est-à-dire par un nombre limité de personnes spécialisées et dans un environnement informatique sécurisé ;**
- 2) les résultats de l'enquête seront produits sous une forme statistique agrégée, de façon à garantir l'anonymat des personnes concernées ;**
- 3) les fichiers de données individuelles constitués pour la réalisation de l'étude (échantillons, réponses) seront effacés à l'issue de l'enquête ou archivés dans des conditions préservant leur confidentialité.**

### **Recommandation n°9 : garantir la confidentialité et l'anonymat par le recours aux techniques d'anonymisation**

Il importe que toutes les précautions soient prises pour préserver la confidentialité des données et garantir l'anonymat des personnes.

La CNIL a toujours favorisé la mise en place et le développement de mesures de sécurité spécifiques, tout particulièrement pour les fichiers de recherche portant sur des données sensibles telles que la santé : chiffrement (cryptage), séparation des données d'identité et des renseignements sensibles, recours aux techniques d'anonymisation à la source des données d'identité.

Elle estime que de telles mesures doivent aussi être développées dans le domaine de la recherche en sciences humaines, tout particulièrement lorsque des données sensibles sur l'origine des personnes sont recueillies.

**En conséquence la CNIL considère que les pouvoirs publics doivent encourager, notamment dans le domaine de la statistique publique, un recours beaucoup plus systématique aux techniques de chiffrement et d'anonymisation.**

### **Recommandation n°10 : garantir l'effectivité des droits « Informatique et Libertés » en assurant la transparence**

Dès lors que des enquêtes sont réalisées par voie de questionnaires auprès des personnes ou par exploitation des fichiers de gestion et que des données directement ou indirectement nominatives sont recueillies, les personnes concernées doivent être parfaitement informées des objectifs poursuivis, des conditions de réalisation de l'enquête, (en particulier des destinataires des données), du caractère facultatif de celle-ci, ainsi que de leurs droits d'opposition, d'accès et de rectification.

**La Commission constate que cette information, pourtant essentielle, est trop souvent négligée. Or, dans un domaine aussi sensible que la mesure de la diversité et la lutte contre les discriminations, elle constitue un facteur clé pour assurer l'adhésion, garantir la confiance et la participation pleine et entière de chacun.**

**De même la Commission préconise une consultation des instances représentatives du personnel en cas d'enquêtes menée dans le domaine de l'emploi. De façon plus générale, il est souhaitable que le lancement des enquêtes nationales soit annoncée publiquement de façon à sensibiliser la population sur ces questions.**