



## Commission SERVICES PUBLICS et SERVICES AUX PUBLICS

---

Réunion du 24 novembre 2010

PROJET D'AVIS.....	2
COMPTE RENDU DE LA RÉUNION .....	3
LISTE DES DOCUMENTS PRÉPARATOIRES .....	31
AVIS D'OPPORTUNITÉ et AVIS AU TITRE DE L'ART 7bis.....	32

---

PROJET D'AVIS

---

proposé par la commission **SERVICES PUBLICS ET SERVICES AUX PUBLICS** du 24 novembre 2010

1. Le Conseil redit l'importance qu'il accorde à la **constitution d'un système d'information permettant de suivre les parcours scolaires des élèves et étudiants**. Il encourage la poursuite d'enquêtes par panel pour collecter des informations sur le jeune et sa famille qui sont absentes des bases de données administratives alors qu'elles peuvent jouer un rôle déterminant pour le déroulement de la scolarité. Il prend note que seule la mobilisation à des fins statistiques de données administratives permettra de suivre régulièrement des « cohortes » d'élèves ou d'étudiants pour étudier des populations particulières ou produire des indicateurs de parcours à un niveau géographique infranational.
2. Le Conseil est conscient de la **nécessité de disposer d'un identifiant national élève pour réaliser des études longitudinales** et soutient en conséquence les actions menées en ce sens sous la coordination de la DEPP. Il demande que toutes les mesures soient prises pour **garantir la confidentialité des données individuelles, notamment par le cryptage des identifiants**. Ce cryptage protégera les individus tout en permettant des recherches ciblées. Il demande que cet identifiant ne soit pas conservé plus de cinq ans après la sortie du jeune du système scolaire. Il rappelle que des formalités doivent être réalisées auprès de la Cnil avant tout appariement de fichiers administratifs.  
Compte tenu de l'importance de ce projet, le Cnis souhaite être informé chaque année de son avancement.
3. Le Conseil confirme l'importance de disposer d'information sur la **mesure des résultats du système éducatif**. Il soutient la consolidation d'un dispositif d'enquêtes visant à appréhender les transitions entre les systèmes d'enseignement, scolaire ou supérieur, et la vie active. A ce titre, il suivra tout particulièrement les résultats de l'enquête « génération 2007 » et demande que le dispositif de mesure de l'insertion professionnelle des diplômés de l'université remplisse tous les critères de qualité statistique, notamment dans le respect du secret statistique, le protocole d'enquête et la diffusion des résultats.  
Il rappelle que la diffusion statistique des résultats doit être distinguée de la communication politique conformément au code de bonnes pratiques de la statistique européenne.
4. Le Conseil a pris connaissance de **l'état d'avancement du projet d'étude longitudinale française depuis l'enfance (Elfe)** qui doit permettre de mieux comprendre comment les parcours familiaux, sociaux, scolaires, sanitaires, le contexte environnemental affectent le devenir des enfants. Il a noté que les enquêtes nationales en maternité et à 6-8 semaines auront lieu sur l'année 2011.

Il prend bonne note des modalités novatrices adoptées pour garantir la confidentialité des données. Il rappelle le besoin d'être informé régulièrement sur les travaux des différentes instances et sur les conditions de fonctionnement de l'instance *ad hoc* chargée de statuer sur les demandes d'accès aux données à des fins d'études et de recherches.

COMPTE RENDU DE LA RÉUNION DE LA  
COMMISSION Services publics et services aux publics

- 24 novembre 2010 -

---

Président : Pierre-Yves GEOFFARD, CNRS et École d'Économie de Paris (EEP)

Rapporteurs : Baudouin SEYS, Drees (Direction de la recherche, des études, de l'évaluation et des statistiques), ministère du travail, des relations sociales et de la solidarité, ministère de la santé, de la jeunesse et des sports, ministère du budget, des comptes publics et de la fonction publique.

Cédric AFSA, Depp (Direction de l'Évaluation, de la prospective et de la performance), ministère de l'éducation nationale

Responsable de la Commission : Nadine LEGENDRE (01 41 17 52 65)

**Adoption du compte rendu**

*En l'absence d'observations sur ce compte rendu, celui-ci est adopté dans le mois qui suit sa date de diffusion.*

**RAPPEL DE L'ORDRE DU JOUR**

I. Diversité des trajectoires des élèves et étudiants.....	7
II. Mesures des résultats du système éducatif.....	19
Présentation pour opportunité d'enquêtes .....	23
III. Point sur le projet ELFE .....	27
IV. Propositions d'avis.....	2
V. Examen de demandes d'accès à des fichiers au titre de l'article 7bis de la loi de 1951 présentées par la Direction des études et des synthèses économiques de l'Insee .....	30

## LISTE DES PARTICIPANTS :

Cédric	AFSA	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Martine	AUBERT-BIAS	Ministère de la santé
Florence	AUDIER	Université Paris 1 Panthéon Sorbonne
Jean-Paul	CAILLE	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Nathalie	CARON	Ministère de l'enseignement supérieur et de la recherche
Caroline	CAVADINI	Secrétariat d'état et d'outre mer
Marie-Aline	CHARLES	INSERM
Marie Luce	COURTOUX	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Laurette	CRETIN	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Joël	CREUSAT	Ministère de la justice et des libertés
Laurent	DAVEZIES	Insee
Florence	DEFRESNE	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Françoise	DUMONT	Ligue des droits de l'homme (LDH)
Françoise	DUSSERT	Cnis
Raphaëlle	FLEUREUX	CNRS Centre M Halbwachs
Dominique	GARREAU	Délégation inter ministérielle à l'orientation
Alain	GELY	CGT
Pierre-Yves	GEOFFARD	PSE Paris Jourdan sciences économiques
Julien	GRENET	CNRS
Ghislaine	GRIMLER	Insee
Danièle	GUILLEMOT	Insee
Salah	IDMACHICHE	Ministère du budget, des comptes publics, de la fonction publique
Steve	JAKOUBOVITCH	Ministère de la santé
Stéphane	JUGNOT	CEREQ
Annick	KIEFFER	CNRS Centre M Halbwachs
Jean-Pierre	LE GLEAU	Insee
Olivier	LEFEBVRE	Ministère de l'enseignement supérieur et de la recherche
Nadine	LEGENDRE	Cnis
Henri	LERIDON	INED
Carine	LHOSTE	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Bruno	LUTINIER	Ministère de l'enseignement supérieur et de la recherche
Isabelle	MAETZ	Ministère de l'enseignement supérieur et de la recherche
Catherine	NAVIAUX	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Laura	ORTUSI	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Sylvain	PAPON	Ministère de l'enseignement supérieur et de la recherche
Hélène	PARIS	CNAF
Hélène	PEYTAVI	Ministère de l'éducation nationale, de la jeunesse, et de la vie associative
Vincent	POUBELLE	CNAVTS
Stéphanie	POUGET	Collectif national de résistance à la Base élèves
Claude	POULAIN	Ligue des droits de l'homme
Jean-François	ROYER	Société Française de Statistique
Karine	STEBLER	Ministère de la santé
Bernard	SUJOBERT	CGT
Laura	SULLI	Conseil général de Seine Saint Denis
Xavier	THIERRY	INED
Christine	THIESSET	Insee
Dalila	VIENNE	Autorité de la statistique publique
Jean-Claude	VITRAN	Ligue des droits de l'homme (LDH)

*EXCUSÉS :*

Jacques	BICHOT	Professeur des universités
Michel	BOISSONNAT	FAFIH
Bertrand	BOIVIN-CHAMPEAUX	Centre techniques des institutions de prévoyance
Michel	CHASSANG	Confédération des syndicats médicaux français
Florence	DEFRESNE	Ministère de l'éducation nationale
Diane	DORE	ANIA
Catherine	FILPA	OREFQ Lorraine
Magali	GRENOUILLEAU	Familles de France
Anne	KERKHOVE	PEEP
Pierre	LEVY	CSMF
Alain	MOREAU	SGAR Aquitaine
Frédéric	VAN ROEKEGHEM	CNAMTS
Jean-Philippe	VINQUANT	HCAAM

*La séance est ouverte à 14 heures 35, sous la présidence de Pierre-Yves GEOFFARD.*

## **INTRODUCTION**

### **Pierre-Yves GEOFFARD**

Merci à tous d'être présents. De fait, les questions et les enjeux de méthode qui nous occupent aujourd'hui sont fortement mobilisateurs.

En préambule, je rappellerai que la commission « services publics et services aux publics » s'intéresse à l'ensemble des services publics, c'est-à-dire à l'État et aux autres opérateurs producteurs de services publics. L'éducation demeure un des domaines d'intervention privilégiés de l'État. Cette situation soulève des interrogations nombreuses, notamment en termes d'efficacité du système éducatif. Quel impact du système éducatif sur la formation des élèves et des étudiants ? Quelle transmission des savoirs ? Dans quelle mesure le système éducatif remplit-il sa mission d'éducation ? Parvient-il à assurer un avenir professionnel et citoyen à ses usagers ? Qu'en est-il de l'insertion professionnelle ?

Un autre enjeu est la lutte contre les inégalités. Le système éducatif tel qu'il existe aujourd'hui contribue-t-il ou non à réduire les inégalités de situations socioéconomiques des élèves et des étudiants ?

Au-delà des constats, ces enjeux posent la question des études à mener pour évaluer l'impact des politiques publiques et des réformes ?

La problématique sur laquelle nous nous penchons présente des déterminants de long terme. En effet, l'impact du système éducatif sur l'avenir des élèves et des étudiants s'inscrit souvent dans la durée. A cela s'ajoute une extrême diversité des situations de départ, des parcours au sein du système éducatif et des devenir à la sortie de celui-ci. De la même façon que pour les questions ayant trait aux retraites ou la santé, ces constats montrent l'impossibilité de raisonner à partir de moyennes. Une double nécessité s'impose dès lors : appréhender la diversité des parcours et adopter une perspective de long terme.

Ces éclairages plaident en faveur de la construction d'outils de suivi statistique des trajectoires dans et à la sortie du système éducatif. Cet objectif requiert de mobiliser des populations suffisamment nombreuses pour prendre en compte toute la diversité des situations et des parcours, dans le respect de la confidentialité des données individuelles. Différentes approches méthodologiques existent pour cela, à travers des dispositifs d'enquêtes répétées et des appariements avec des données d'origine administrative.

Notre ordre du jour sera entièrement consacré au système éducatif par un ensemble de questions portant sur l'appréhension de la diversité des trajectoires des élèves et des étudiants, ainsi que sur la mesure des résultats obtenus par le système.

### **Bernard SUJOBERT, CGT**

Je regrette au nom de la CGT que l'avant projet de programme de travail ne soit pas davantage débattu dans cette commission ou dans les autres d'ailleurs. Je souhaiterai pour ma part revenir sur le programme de travail de la Commission. Les thèmes de la santé et de l'éducation auront été abordés au cours des deux premières réunions. Cela étant, d'autres services statistiques ont présenté un programme de travail pour 2011 portant sur d'autres thématiques.

La commission « services publics et services aux publics » recouvre les champs de plusieurs anciennes formations du Cnis. De surcroît, elle présente la particularité de ne compter aucun département de l'Insee parmi les services producteurs qui lui sont rattachés. Sept services statistiques ministériels et sept organismes producteurs extérieurs au service statistique public y sont en revanche représentés. Dans un tel contexte, un dialogue transversal mériterait d'autant plus d'être instauré que

des réflexions globales ont été engagées dernièrement en matière d'étalonnage de sources privées, de labellisation de données administratives extérieures au service statistique public, etc. Par ailleurs, des innovations devraient être soumises à la discussion, parmi lesquelles la création de l'Institut national des hautes études sur la sécurité et la justice, la refondation du service statistique du ministère de la justice, etc. Pour assurer la cohérence d'ensemble des travaux du Cnis, il serait nécessaire d'aborder tous ces points dans le cadre de cette commission.

### **Pierre-Yves GEOFFARD**

Je partage la frustration exprimée. Effectivement, le champ de la Commission s'avère extrêmement large. Dans ces conditions, l'option a été prise d'aborder séquentiellement un certain nombre de thèmes fondamentaux. A l'issue de ce premier tour d'horizon, les domaines de la santé, des retraites, de l'éducation, de la sécurité et de la justice auront été couverts. Un nouveau fonctionnement devra alors être trouvé. Une réunion du Cnis pourrait ainsi être consacrée aux questions transversales.

### **Bernard SUJOBERT**

Tous les SSM représentés au sein de la Commission sont fortement impliqués dans la remontée des informations au sein de leurs administrations. Cette problématique revêt un caractère transversal.

Par ailleurs, les SSM disposent généralement d'instances de concertation avec leurs administrations de rattachement. Si le débat au sein du Cnis ignore les concertations au sein de ces instances, des divergences d'orientations pourraient apparaître dans la durée et le Cnis manquerait à sa fonction de mise en cohérence de l'ensemble de la statistique publique.

### **Pierre-Yves GEOFFARD**

Ces questions sont effectivement importantes, d'autant que la configuration actuelle de la Commission ne permet pas de confronter régulièrement chaque SSM avec la demande des utilisateurs de statistiques publiques. A ce stade, je suis preneur de toute suggestion.

## **I. Diversité des trajectoires des élèves et étudiants**

### **1. Suivi et déterminants des trajectoires – perspectives d'évolution des dispositifs d'observation – point sur le répertoire national des identifiants élèves**

#### **Nathalie CARON, SIES – MESR**

Mon propos sera centré sur l'enseignement supérieur. Cédric AFSA abordera quant à lui les questions ayant trait au système scolaire.

#### *a. Les principaux objectifs des travaux sur les parcours*

#### **Nathalie CARON**

Les travaux sur les parcours ont pour premier objectif de produire un panorama d'ensemble. Il s'agit d'éclairer le public (en publiant des notes d'information sur les sites des ministères de l'éducation nationale et de l'enseignement supérieur ou en répondant à des questions parlementaires), d'informer les apprenants (élèves et étudiants) ainsi que leur famille, de connaître les situations récentes et actuelles pour comprendre les trajectoires et leurs déterminants, de participer à la collecte de données européennes (dans la perspective de répondre à un règlement européen à compter de 2012) et d'informer les décideurs sur les modalités de scolarisation, de poursuite d'études et d'orientation des élèves, étudiants et apprentis.

Les travaux sur les parcours ont également un objectif de pilotage. L'enjeu est alors d'évaluer l'efficacité des politiques publiques et des dispositifs de formation et d'orientation tels que le Plan Licence (conçu pour attirer et accompagner les étudiants en licence) et la réforme de la voie professionnelle (à travers la transformation du bac professionnel). Il s'agit d'apporter des réponses homogènes entre les établissements (avec des indicateurs comparables entre eux), avec le niveau académique et le niveau national. De nombreux indicateurs LOLF sont également calculés à partir de parcours, en particulier dans l'enseignement supérieur.

Le troisième objectif des travaux sur les parcours se trouve être la rationalisation et l'urbanisation des systèmes d'information, en vue notamment d'alléger le questionnement des établissements. Des données concernant le baccalauréat (série, année d'obtention, etc.) pourront ainsi être récupérées directement dans les fichiers du baccalauréat sans avoir à les redemander systématiquement dans les enquêtes auprès des établissements du supérieur.

#### *b. Les données disponibles actuellement*

### **Nathalie CARON**

Dans l'enseignement supérieur, il existe deux instruments principaux pour observer les parcours : les enquêtes par panel et la construction de cohortes à partir des remontées individuelles administratives. La construction de cohortes nécessite de « fusionner » les fichiers individuels avec une clé d'appariement : l'identifiant national des étudiants (INE). Ce dernier est attribué aux étudiants au moment de leur passage à l'examen du baccalauréat.

En pratique, les enquêtes par panel et les cohortes s'avèrent complémentaires. Les panels d'étudiants sont des enquêtes par sondage reposant sur des échantillons de 10 000 à 12 000 bacheliers. Ces étudiants sont ensuite interrogés individuellement chaque année, jusqu'à ce qu'ils quittent l'enseignement supérieur. Les données sont collectées via un questionnaire papier, avec possibilité pour les étudiants de répondre par Internet. Des relances téléphoniques permettent généralement d'obtenir des taux de réponse proches de 90 %. Le dispositif repose sur un tronc commun de questions, visant notamment à préciser le lieu d'inscription au 31 octobre et au 1er mars. Des modules spécifiques peuvent être implémentés d'une année sur l'autre, en vue de caractériser les étapes clés telles que l'entrée dans l'enseignement supérieur (choix de formation, canaux d'information, satisfaction à l'égard de l'orientation, etc.), la deuxième année de cursus (choix de réorientation, mobilité, etc.), la troisième année de cursus (stages effectués, etc.), etc.

Les panels servent à donner des cadrages généraux. Combien de bacheliers s'orientent dès la première année vers l'université ? Combien d'étudiants se réorientent à l'issue d'une première année à l'université et où se réorientent-ils ? Des parcours relativement fins peuvent être ainsi observés, en fonction de la taille des échantillons. Les panels offrent de surcroît la possibilité de recueillir des données qualitatives permettant de rendre compte du vécu et des difficultés éventuelles des étudiants (difficultés financières, difficultés liées à la formation choisie, etc.). *A contrario* des sources administratives, les panels permettent également de repérer les inscrits et présents dans la formation au 31 octobre – l'inscription administrative n'impliquant pas nécessairement une présence effective. Des bilans généraux sur le déroulement des études peuvent alors être établis, à travers des profils. Le champ couvert par les panels est très vaste, avec des informations sur les formations suivies en France, à l'étranger, par correspondance, par apprentissage, etc.

Cela étant, les panels d'étudiants ne permettent aucune analyse université par université ou par catégorie d'étudiants (dès lors que ceux-ci sont trop peu nombreux au sein de l'échantillon). De plus les panels ne sont constitués que de bacheliers. Or l'enseignement supérieur n'accueille pas que des titulaires du baccalauréat. D'autres équivalences, en particulier celles liées à des diplômes étrangers, ne sont pas prises en compte dans ces panels. Une autre limite de l'outil réside dans le fait que les échantillons reposent sur des millésimes précis. Le panel constitué à partir des bacheliers de 2008 ne pourra ainsi fournir aucune indication sur l'impact d'une réforme de l'orientation opérée en 2009. Les difficultés pour contacter les étudiants s'intensifient au cours les années, malgré un suivi minutieux des adresses.

Dans l'enseignement supérieur, un premier panel a été mobilisé en 1996, à partir d'un panel constitué en 1989 dans l'enseignement secondaire. Un second panel a ensuite été mobilisé en 2002, à partir d'un panel constitué en 1995 dans l'enseignement secondaire. A l'heure actuelle, des étudiants de ce



panel font encore l'objet d'un suivi. En mars 2005, ce panel a été repris par l'Insee dans le cadre d'une enquête sur l'entrée dans la vie active. Un nouveau panel a finalement été construit en 2008 : il est constitué de 12 000 bacheliers de 2008 qui ont été interrogés pour la première fois en mars 2009.

Parmi les sources relatives à l'enseignement supérieur figure également le recensement annuel des inscriptions administratives. En pratique, 15 opérations statistiques sont nécessaires pour appréhender toute la diversité des formations de l'enseignement supérieur français. Ces opérations statistiques correspondent à des remontées administratives (les universités notamment) ou à des enquêtes statistiques (auprès des écoles d'ingénieurs privées, des écoles de management privées, etc.). Deux types de remontées d'information existent : des remontées individuelles (une ligne = un étudiant) et des remontées agrégées (tableaux sur la composition des effectifs). L'observation des parcours nécessite quant à elle de mobiliser des données individuelles. A ce jour, 80 % des remontées portent sur des données individuelles.

L'apport principal des remontées d'informations individuelles réside dans l'exhaustivité des données produites (sur champ). Des analyses peuvent alors être menées sur des sous-populations fines, ainsi qu'au niveau académique ou par université. Des limites apparaissent néanmoins dans le taux de couverture des formations, et dans l'appréhension de l'apprentissage et de la formation continue. L'exploitation des données requiert de plus des manipulations relativement lourdes. Enfin, les remontées individuelles s'avèrent exclusivement quantitatives et ne portent que sur des inscriptions administratives.

Au niveau de l'enseignement supérieur, la première perspective d'enrichissement consiste à maximiser les remontées d'informations individuelles. L'objectif est d'atteindre un taux de couverture du champ de l'ordre de 90 %. Des progrès seront notamment possibles à travers la remontée des informations concernant les formations paramédicales et sociales (gérées par la Drees) qui devraient être intégrées à l'INE.

Plusieurs études ont été réalisées à partir des outils existants. A partir des cohortes, des taux de réussite en licence (en 3 ans) ont été calculés par université, ainsi que des taux de réussite en IUT (en 2 ans) par IUT. Des notes d'information ont été publiées à ce sujet (ces notes sont disponibles sur le site du ministère de l'enseignement supérieur et de la recherche). A partir des panels, des trajectoires de bacheliers de « première génération » – c'est-à-dire de bacheliers dont les parents n'étaient pas titulaires du baccalauréat – ont pu être observées, avec en point de mire une interrogation sur l'impact d'ambitions moindres sur les parcours dans l'enseignement supérieur. Un article a été publié sur le sujet dans l'édition 2009 de *France portrait social*. Une étude devrait être menée prochainement sur les bacheliers « S », à partir du panel constitué en 2008.

### **Cédric AFSA, DEPP**

Au niveau de l'enseignement scolaire (primaire et secondaire), les panels d'élèves constituent la seule source d'observation des trajectoires. Un panel d'élèves entrant en sixième et suivis durant leur scolarité secondaire a été mis en place dès 1973. Le panel d'entrants en 6<sup>ème</sup> en 1989 a étendu le suivi à l'enseignement supérieur. La Depp a donc une longue expérience de l'utilisation de ce type d'outils. Des panels d'entrants au premier niveau de l'enseignement élémentaire ou au premier niveau de l'enseignement secondaire ont été construits, avec des effectifs et des périodicités variables. Ils s'appuient sur des données administratives (permettant d'actualiser la situation scolaire des élèves), ainsi que sur des données d'enquêtes auprès des responsables d'établissements, des familles et des élèves eux-mêmes (en vue d'obtenir des informations n'existant pas dans les bases administratives, portant notamment sur la situation socioprofessionnelle des parents, le rapport à l'école des familles, les souhaits d'études et les projets professionnels des enfants, les environnements extrascolaires, les difficultés perçues, etc.).

Les panels permettent un suivi longitudinal. Ils reposent généralement sur des échantillons importants. Ils combinent la précision des données provenant des sources administratives avec la pertinence et la richesse des données d'enquêtes. Pour ces raisons, les panels sont irremplaçables. Leur remplacement à plus ou moins long terme par des données exclusivement administratives ne saurait être envisagé.

Cela étant, les échantillons constitués dans le cadre des panels ne permettent pas d'opérer des zooms particuliers. Ils ne permettent pas non plus une actualisation rapide des données, le cas échéant pour calculer des indicateurs de performance.

Le suivi de cohortes à partir de données administratives présente quant à lui l'avantage de reposer sur des informations existantes. En les utilisant la Depp met en œuvre le Code de bonnes pratiques de la statistique européenne, en particulier son article 9 consacré à l'allègement de la charge des répondants. Ce type d'outils permet le calcul actualisé d'indicateurs longitudinaux de performance.

Le suivi régulier de cohortes à partir de données administratives revêt également des enjeux en termes d'évaluation des politiques publiques. A titre d'exemple, il est aujourd'hui impossible de suivre en temps réel la mise en place de la réforme de la voie professionnelle. Les outils actuels ne couvrent pas non plus le champ des formations ne relevant pas du ministère de l'éducation nationale, dont les formations par l'apprentissage.

### *c. Les préalables à la construction de parcours sur cohortes*

#### **Cédric AFSA**

Pour construire des cohortes susceptibles de mettre en évidence des parcours, il faut tout d'abord mettre en place des nomenclatures et des répertoires partagés afin de disposer d'un système d'information cohérent. L'objectif est également de définir des règles communes de traitement des incohérences telles que les inversions de genre. Il est donc nécessaire de mener une réflexion qualitative sur les appariements.

L'enjeu est par ailleurs de mettre en œuvre des protections juridiques suffisantes pour appairer les fichiers dans le respect du droit (en lien avec la Cnil) et de l'anonymat des personnes. Sur ce point, un double cryptage interviendra au niveau de chaque académie puis au moment de la remontée des informations vers la Depp. Par ce biais, toute identification des individus devient impossible. Des règles de diffusion doivent également d'être définies, en fonction des interlocuteurs.

Un autre préalable à la mise en œuvre de cohortes est de disposer d'un identifiant commun – c'est-à-dire d'une clé d'appariement unique et fiable. L'idée est de conserver pour une durée limitée (i.e. tant que le jeune n'est pas sorti définitivement du système éducatif) des INE ne variant pas dans le temps ni dans l'espace. A ce jour, un « INE » existe dans l'enseignement secondaire, mais il n'est pas national. Il est modifié lors de chaque changement d'académie. Une stabilité dans l'espace doit être instaurée. Pour être efficace et pertinent, le dispositif devra également être généralisé à l'ensemble des sources, y compris celles ne relevant pas des ministères de l'éducation nationale et de l'enseignement supérieur.

C'est dans une telle optique que la construction d'un répertoire national des identifiants d'élèves a été initiée. Celui-ci doit reposer sur une distinction fondamentale entre la fonction d'immatriculation et la fonction d'usage des identifiants. Le service d'immatriculation serait assuré par la Depp qui aurait en charge d'attribuer des identifiants uniques. Les identifiants seraient ensuite utilisés sous la responsabilité de leurs utilisateurs, dans le strict respect des obligations légales notamment en matière de confidentialité.

A terme, l'ambition est de remplacer l'actuelle base nationale des identifiants élèves (BNIE) ne s'appliquant qu'au niveau primaire. Le répertoire national des identifiants d'élèves (RNIE) couvrirait quant à lui l'ensemble du système éducatif. Il s'agirait ainsi de disposer d'une base de « pure » immatriculation, sans historique et à durée de vie limitée. Cinq ans après la sortie du système éducatif, les identifiants seraient supprimés.

## 2. Introduction à la discussion

### Julien GRENET, CNRS

En poursuivant des travaux de recherche dans le domaine de l'économie de l'éducation, j'ai pu prendre conscience de la richesse des données permettant d'étudier ces questions en France, ainsi que des difficultés rencontrées par les chercheurs pour y accéder.

L'économie de l'éducation s'intéresse aux effets économiques de la dépense éducative, en termes d'efficacité et d'effets redistributifs (dans une perspective de lutte contre les inégalités). L'ambition des chercheurs en économie de l'éducation est d'alimenter le débat public et de formuler des recommandations en matière de politique publique. Ceci implique de pouvoir suivre les trajectoires des élèves et étudiants dans le temps. Parmi les questionnements nécessitant des analyses longitudinales figurent notamment : la mesure des effets du redoublement et de l'orientation scolaire sur la réussite scolaire et l'insertion professionnelle, l'évaluation du bénéfice de l'assouplissement de la carte scolaire pour les élèves issus de milieux défavorisés, la mesure de l'impact des aides publiques directes aux étudiants (bourses, aides au logement, etc.) sur l'accès aux études supérieures et la réussite aux diplômes, ainsi que l'appréhension des moyens d'améliorer l'insertion professionnelle des étudiants.

Le suivi des trajectoires représente un triple défi pour la recherche. Tout d'abord, il s'avère nécessaire de pouvoir suivre les individus au cours du temps. Il convient ensuite de caractériser le contexte dans lequel s'inscrivent les trajectoires. Enfin, il importe de respecter les règles de confidentialité s'appliquant aux données individuelles.

En pratique, le suivi longitudinal impose de disposer d'un identifiant individuel fiable et stable (dans le temps et l'espace), de pouvoir croiser les informations provenant de sources statistiques diverses et de définir des règles précises concernant la nature des données utilisables par les chercheurs et leurs conditions d'accès.

Traditionnellement, des données d'enquêtes sont utilisées par les chercheurs. Celles-ci sont d'une grande diversité et d'une grande richesse. Les panels de l'éducation nationale permettent ainsi de suivre des trajectoires dans l'enseignement (primaire, secondaire et supérieur). Le suivi de l'insertion professionnelle repose sur des enquêtes générales telles que l'enquête emploi de l'Insee ou des enquêtes plus spécifiques telles que l'enquête « Génération » du Cereq. Des enquêtes sont également réalisées par les chercheurs eux-mêmes, en particulier sur les dispositifs expérimentaux tels que les écoles de la deuxième chance, les internats d'excellence, etc. Les données ainsi produites à partir d'échantillons représentatifs présentent une grande qualité statistique et fournissent des variables riches et des informations subjectives. De surcroît, elles demeurent relativement faciles d'accès, notamment grâce au centre Quételet. Toutefois, la taille des échantillons limite souvent les possibilités et n'autorise pas l'analyse des situations locales. L'anonymisation des données en vue de respecter le secret statistique induit par ailleurs une perte d'information, notamment au niveau des variables de localisation géographique. Les informations diffusées aux chercheurs conservent ainsi un niveau de détail relativement grossier – ce qui pose problème pour analyser certains impacts locaux, en lien par exemple avec l'assouplissement de la carte scolaire.

Dans ce contexte, les chercheurs se tournent de plus en plus vers des sources administratives décentralisées (BEA, AFFELNET, APOGEE, HEBERG) ou nationales (OCEAN, AGLAE, SCOLARITE, SISE). Ces sources présentent l'avantage d'être quasi-exhaustives, ce qui permet l'étude de parcours fins et la mesure d'impacts parfois limités dans le temps. Elles autorisent également des appariements, permettant notamment de comparer les parcours des élèves boursiers et non-boursiers. Collectées essentiellement à des fins de gestion, les données issues de sources administratives ne contiennent toutefois pas toute la richesse des données d'enquêtes. En outre, elles n'affichent pas toujours un taux de couverture de 100 % et ne bénéficient pas d'un archivage systématique.

A ce jour, les sources administratives demeurent largement sous-exploitées par les chercheurs, du fait notamment des règles juridiques qui encadrent leur utilisation. Jusqu'à une période très récente, les règles de confidentialité conditionnaient l'accès à ce type de données à une anonymisation lourde par les services statistiques producteurs. La multiplicité des interlocuteurs génère également des contraintes institutionnelles. Dans le domaine de l'éducation, des rencontres s'avèrent nécessaires

avec la Depp, les rectorats académiques, le centre national des œuvres universitaires et scolaires (CNOUS), le service inter académique des examens et concours (SIEC), etc. Dans ce paysage, les responsabilités vis-à-vis des sources ne sont pas toujours clairement identifiées. De plus, les demandes des chercheurs entrent parfois en conflit avec les missions de gestion des administrations. La lourdeur des traitements entraîne alors des refus ou génère des délais d'obtention parfois supérieurs à un an.

Afin de faciliter l'étude des trajectoires tout en garantissant le respect des règles de confidentialité, la mise en place d'un identifiant commun pourrait être envisagée, le cas échéant à travers la généralisation du RNIE. Des appariements entre sources seraient alors possibles, notamment en vue de mieux caractériser la réussite éducative à partir des trajectoires scolaires et des résultats aux examens. Un enrichissement du contenu des données devrait par ailleurs être entrepris tout d'abord en augmentant le taux de couverture, en particulier dans l'enseignement supérieur. Il faudrait également améliorer l'information sur les résultats scolaires (au-delà des variables binaires de réussite ou d'échec) et le suivi de l'insertion professionnelle (dans la transition entre les études supérieures et le marché du travail). L'enquête sur l'insertion professionnelle des étudiants devrait ainsi apporter des éléments de réponse supplémentaires. En dernier lieu, l'enjeu est d'assurer l'archivage et la conservation des sources. Dans cette optique, la durée de conservation de certaines données de gestion serait être allongée à travers, par exemple, des extractions de fichiers anonymisés pour des études sur des réformes ponctuelles.

En ce qui concerne l'accès aux données, la loi sur les archives du 15 juillet 2008 constitue un cadre nouveau. Celle-ci a étendu les possibilités d'accès aux données personnelles indirectement nominatives, à des fins statistiques ou de recherche scientifique ou historique. L'avis du Comité du secret demeure toutefois requis, de même que l'accord du service producteur et le dépôt d'une déclaration normale de traitement statistique auprès de la Cnil. Le Comité du secret a toutefois été réorganisé par le décret du 20 mars 2009, pour faciliter le traitement des demandes d'accès des chercheurs.

Pour que ce nouveau cadre puisse élargir l'accès aux données administratives, il conviendrait cependant qu'un dispositif d'accès suffisamment sécurisé soit mis en œuvre. Le centre d'accès sécurisé à distance de l'Insee offre aujourd'hui une solution technique en la matière, reposant sur un système de contrôle d'accès biométrique et de blocage des possibilités de copie et de rapatriement des données. Au jour d'aujourd'hui, ce dispositif est utilisé pour accéder à des données confidentielles de l'Insee (fichiers individuels du recensement, DADS, etc.). A terme, certaines données administratives scolaires pourraient être rendues accessibles par ce biais.

### **Pierre-Yves GEOFFARD**

La multiplication des interlocuteurs génère des inégalités dans l'accès aux sources, ce qui n'est guère acceptable. Un cadre plus précis et plus transparent mériterait d'être défini. Certains chercheurs français travaillent aujourd'hui sur le système éducatif en utilisant des données d'origine étrangère, faute de pouvoir disposer de données sur le système français. A l'avenir, les centres d'accès sécurisés pourraient permettre la mise à disposition de données sensibles, en vue de procéder à des analyses statistiques, sans risque de diffusion inappropriée.

### **Françoise DUMONT, vice-présidente de la LDH**

Mon domaine d'intervention privilégié au sein de la LDH se trouve être le droit des enfants et des jeunes. C'est à ce titre que je souhaiterais aujourd'hui m'exprimer. L'invitation adressée à la LDH témoigne d'une volonté de dialogue qui, au cours des derniers mois, a fait défaut. Nombre d'organisations dont la LDH ont regretté ce manque de transparence et de débat citoyen autour des questions liées au fichier « Base élèves premier degré » et à la base nationale des identifiants élèves (BNIE). A ce jour, le ministère de l'Éducation nationale (MEN) n'a toujours pas publié de texte concernant cette dernière. L'INE, quant à lui, n'a été évoqué qu'au travers d'un arrêté datant du mois octobre 2008. Le MEN donne ainsi l'impression d'agir comme si la mise en place de tels outils était anodine – ce qui ne saurait être le cas.

La BNIE<sup>1</sup>, le RNIE et l'INE sont aujourd'hui examinés sous l'angle statistique. Cependant, la mise en place de tels outils technologiques interpelle sur la constitution de méga-bases de données. Cette problématique est au cœur des réflexions de la LDH, en lien avec ses engagements en matière de défense des libertés et de protection de la vie privée. Au-delà des préoccupations associées au contexte sécuritaire (chasse aux sans-papiers, chasse aux fraudeurs, etc.), les méga-bases de données recouvrent des enjeux économiques considérables, sur des marchés en développement.

Les outils dont nous discutons aujourd'hui méritent une attention particulière, du fait de leur utilisation en ce qui concerne des enfants et des jeunes. Le Conseil d'État et le Comité des droits de l'enfant de l'ONU ont rappelé, directement ou indirectement, la nécessité de prendre des précautions à cet égard. Le Conseil d'État a sanctionné l'omission, dans la déclaration du fichier Base élèves à la Cnil, de la mention des rapprochements possibles avec des données provenant d'autres fichiers. Le Conseil d'État a également jugé injustifiée la durée de conservation de 35 ans des informations contenues dans la BNIE. Le Comité des droits de l'enfant de l'ONU a insisté, en juin 2009, sur la nécessité de veiller au respect du caractère anonyme des données enregistrées.

Les outils évoqués ce jour sont mis en œuvre à travers l'école. La scolarité étant obligatoire, ces outils concerneront, à terme, l'ensemble de la population française. Dans sa déclaration à la Cnil, le MEN a évalué à 13 millions le nombre de personnes susceptibles, à terme, de recevoir un INE. Par son ampleur, le dispositif rappelle le projet SAFARI du ministère de l'Intérieur, que la LDH avait combattu, obtenant la restriction du champ d'application du NIR à une partie de la sphère sociale.

Les enjeux évoqués nous imposent également de veiller à la conformité des outils mis en place avec les dispositions de la loi informatique et libertés. La question se pose vis-à-vis de la finalité et de la proportionnalité des informations recueillies à travers le RNIE. La durée de conservation de ces informations suscite également des interrogations. Qu'en sera-t-il du droit à l'oubli ? Qu'en sera-t-il, enfin, du droit d'opposition des parents des mineurs concernés ?

L'utilisation d'outils de gestion informatiques ne saurait être dénoncée en tant que telle. La société de surveillance nécessite en revanche d'être combattue, à travers une vigilance exercée concernant les interconnexions qui la sous-tendent. Depuis 1998, l'utilisation du NIR n'est plus protégée – l'administration fiscale y ayant accès. Cette offensive se poursuit aujourd'hui avec le répertoire commun des prestations sociales.

L'éducation nationale a besoin d'outils lui permettant d'assurer une gestion pertinente et il est juste qu'elle se dote des moyens d'assurer des suivis de cohortes et de mettre en place des études longitudinales. Les fiches de la Depp fournissent d'ores et déjà des renseignements précieux sur le fonctionnement et l'évolution du système éducatif. Reste à évaluer le caractère indispensable pour la recherche du RNIE et de l'INE. Compte tenu des risques d'interconnexions liés à ces outils, ne serait-il pas envisageable d'obtenir des variables pertinentes par d'autres voies ?

### **Jean-Claude VITRAN, LDH**

En tant que membre du bureau national de la LDH, je m'occupe principalement des questions ayant trait aux libertés et aux nouvelles technologies. A ce jour, aucune réponse claire n'a été apportée concernant la finalité de l'INE (et, par extension, du RNIE). S'agit-il d'un outil de gestion ou d'un moyen simple de fichier, à terme, l'ensemble de la population ? La question a déjà été posée lors de la réunion du 12 octobre 2009, dans les termes suivants : « *Le problème essentiel est lié à l'absence de précision concernant la finalité de l'INE et celle de l'enregistrement d'informations personnelles dans les bases de données de MEN. (...) On peut légitimement demander pourquoi il n'est pas fait appel aux techniques bien connues d'échantillonnage qui permettent d'obtenir des réponses sous forme de pourcentages – ce qui évite l'enregistrement de données nominatives pour la totalité des élèves ? Et*

---

<sup>1</sup> La Base Nationale des Identifiants Élèves (BNIE) qui ne couvre que le premier degré sera remplacée à terme par le Répertoire National des Identifiants Élèves et étudiants (RNIE). Ce répertoire, sous la responsabilité du Service Statistique Ministériel de l'Éducation Nationale, a comme finalité d'attribuer un numéro (INE=Identifiant national élève-étudiant) unique à tout élève scolarisé et retrouver ce numéro tout au long de sa scolarité. Il couvrira donc les premier et second degrés, l'enseignement supérieur, l'apprentissage et l'enseignement délivré dans des établissements sous tutelle d'autres ministères que celui de l'éducation nationale.

*pourquoi n'a-t-on pas mis en place un cryptage des informations qui aurait empêché l'exploitation ultérieure de données personnelles susceptibles de stigmatiser les individus ? »*

La question du cryptage est fondamentale. Elle a d'ailleurs été évoquée par un de vos pairs dans un écrit datant de 2005, dans les termes suivants : « *La Cnil considère que la constitution de fichiers centraux avec un identifiant tel que l'INE présente un risque. Elle estime en effet qu'il est actuellement facile de se procurer l'INE d'un élève particulier et que la multiplication de fichiers d'études nationaux comportant l'INE pourrait permettre trop aisément à des personnes malintentionnées d'accéder à des informations individuelles nominatives. (...) Afin d'effectuer un suivi longitudinal de leurs étudiants, les universités ont, à plusieurs reprises, sollicité la Depp pour obtenir un fichier national avec l'INE. Ce genre d'information est sensible dans la mesure où il permettrait, par exemple, la fabrication de CV électroniques des étudiants. Les contacts informels avec la Cnil ont confirmé cette idée. Elle considérerait que la présence dans un fichier de l'INE suffisait à le rendre nominatif et amenait à interdire sa diffusion large. Elle a donc suggéré que l'on tire parti des travaux déjà effectués, notamment par le CHU de Dijon, en matière de cryptage sans retour (hachage) des identifiants. »*

On peut s'interroger également sur la nécessité de conserver ces données pendant une durée de 35 ans, alors même que la Cnil avait évoqué une durée de 10 ans dans l'un de ses avis. Pourquoi est-on revenu sur des décisions qui semblaient acquises en 2005 ?

Des associations d'enseignants et de parents s'inquiètent aujourd'hui des dérives liberticides possibles dans l'utilisation de tels outils et demandent l'abrogation pure et simple du projet INE. Ces inquiétudes, par ailleurs liées à la perception « d'arrière-pensées » de l'État, nous semblent particulièrement légitimes. Le Comité des droits de l'homme de l'ONU a exprimé des préoccupations similaires à travers l'un de ses avis : « *Le Comité prend note avec préoccupation de la multiplication des bases de données servant à la collecte, à la conservation et à l'utilisation à long terme de données personnelles sur les enfants, qui pourrait aller à l'encontre du droit de l'enfant et de sa famille à la protection de leur vie privée. (...) Le Comité engage instamment l'État à prendre toutes les mesures voulues pour garantir que la collecte, le stockage et l'utilisation de données personnelles sensibles sont compatibles avec les obligations qui lui incombent en vertu de l'article 16 de la Convention des droits de l'enfant. »*

A ce stade, il convient d'évoquer également l'amendement Warsmann qui voudrait légaliser l'interconnexion de l'ensemble des fichiers – ce qui reviendrait à mettre en place un nouveau projet SAFARI.

La LDH estime aujourd'hui que l'ensemble des données collectées doivent être anonymisées avant toute exploitation en dehors du cadre scolaire local, en vue de respecter les droits fondamentaux des personnes. Les données utilisées dans le cadre d'études statistiques devront pour cela être cryptées, sans possibilité de retour.

### 3. Débat

#### **Stéphanie POUGET, Collectif national de résistance à la Base élèves (CNRBE)**

Des besoins en termes d'informations ont été exprimés ce jour. La différence entre les données obtenues au travers des panels et celles obtenues au travers des systèmes d'informations individuelles a également été clairement mise en évidence. Cependant, le souci de respecter les libertés individuelles et la vie privée n'a jamais été évoqué parmi les limites à l'utilisation des outils de suivi des trajectoires. Or cette question se pose dès lors que des bases exhaustives sont constituées. Le risque n'est pas le même lorsque les travaux portent sur des panels. Dans ce contexte, le besoin d'attribuer un identifiant national introduit un risque supplémentaire.

On assiste aujourd'hui, dans l'éducation nationale, à une multiplication des traitements de données nominatives, pour la gestion des élèves, le suivi de la validation des compétences du socle commun, le suivi des élèves décrocheurs, etc. L'identifiant académique actuel figure dans tous les fichiers ainsi constitués. Son remplacement par un identifiant national pourrait autoriser quantité de recoupements, sans possibilité de contrôle de la part des familles. Sans aller jusqu'à parler « d'arrière-pensées », ces potentialités inquiètent aujourd'hui, à plus forte raison dans un contexte marqué par la mise en œuvre

de lois telles que celle relative à la prévention de la délinquance. Certaines barrières semblent ne pas devoir être franchies.

### **Annick KIEFFER, CMH – CNRS**

Des tensions apparaissent entre les associations citoyennes telles que la LDH et les chercheurs. Les unes recherchent la protection des données individuelles, tandis que les autres souhaitent pouvoir y accéder plus largement. Au moment d'envisager un arbitrage, la réflexion mériterait aussi d'être menée sur les conditions d'accès des chercheurs aux données susceptibles de permettre des appariements entre les enquêtes et les sources administratives. Des réponses ont été apportées dans ce domaine grâce au centre d'accès sécurisé de l'Insee. Cependant, les ressources pourraient être insuffisantes pour généraliser ce modèle.

L'utilisation de données d'origine étrangère a par ailleurs été évoquée. En réalité, les pays qui ont développé des accès sécurisés demeurent peu nombreux. Il en va de même pour les pays mettant à disposition des données sur l'éducation. En France, un grand nombre de données sont aujourd'hui disponibles, notamment *via* le centre Quételet. Il serait opportun de défendre cette richesse, sans avoir une vision idyllique des dispositifs mis en œuvre à l'étranger.

Un travail avait du reste été engagé autour de l'accessibilité des données du recensement de la population. Cette réflexion pourrait être étendue aux données sur l'éducation, en considérant la finalité des travaux de recherche et d'amélioration de la connaissance plutôt que celle des travaux de gestion ou d'évaluation des établissements. A cet endroit, il importe d'insister sur les risques de stigmatisation des établissements, au-delà de la stigmatisation des élèves et des pratiques familiales. Une réflexion nécessiterait d'être engagée sur la finalité des enquêtes.

Des critiques ont également été émises à l'encontre des panels. Il faut néanmoins rappeler que des données rétrospectives existent aujourd'hui. La source FQP (enquête Formation et qualification professionnelle de l'Insee) est ainsi largement utilisée en sociologie de l'éducation et en économie de l'éducation. Il est clair que les panels sont, par nature, des instruments lourds et qu'ils ne sauraient fournir rapidement des données permettant de répondre à des besoins précis. De toute évidence, certaines interrogations, concernant notamment l'impact de la réforme du baccalauréat professionnel, nécessiteront des enquêtes particulières. Les enquêtes d'insertion du Cereq, quant à elles, ne peuvent répondre aux besoins d'évaluation établissement par établissement.

### **Stéphane JUGNOT, CEREP**

Julien Grenet a évoqué de manière générale un manque d'informations sur l'insertion professionnelle des étudiants. Or des centaines de milliers d'euros sont investis depuis plusieurs années dans les enquêtes Génération du Cereq, avec pour objectif d'étudier l'insertion professionnelle des diplômés de tous niveaux. Un certain nombre d'études ont été menées sur le sujet. Le constat mériterait donc d'être nuancé, afin d'identifier plus précisément les besoins réels.

Vis-à-vis du secret statistique (longtemps perçu comme une entrave à l'activité des chercheurs), un changement de législation est tout de même intervenu récemment, permettant aux chercheurs de contourner un certain nombre de difficultés.

Quoi qu'il en soit, les appariements avec des sources administratives semblent aujourd'hui incontournables pour la recherche, en dépit des préoccupations légitimes exprimées concernant la confidentialité des données. Les organismes administratifs eux-mêmes utilisent de plus en plus cette possibilité. L'enjeu est toutefois de faire en sorte que les exploitations soient proportionnées et répondent à des besoins réels et significatifs.

Un autre axe de progrès se situe dans la documentation des données – ce qui renvoie à la question des moyens. En pratique, la charge de travail des organismes producteurs s'avère déjà importante. Ce point devrait d'être pris en compte dans le cadre du plan recherche lancé par le ministère de l'enseignement supérieur et de la recherche (MESR) pour que des moyens soient consacrés au dimensionnement des centres d'accès sécurisés.

### **Claude POULAIN, LDH – CGT**

Les parcours scolaires sont des éléments importants de la vie privée. Leur divulgation constitue une atteinte au respect des droits fondamentaux et peut avoir, dans certain cas, des conséquences graves. En début d'année 2010, un étudiant tunisien a ainsi reçu une obligation de quitter le territoire français, au motif que son parcours scolaire était « erratique ». Cet étudiant en mathématiques appliquées à la gestion s'était inscrit en lettres modernes afin de répondre aux attentes de son nouvel employeur. Orwell rencontre ici Kafka !

Les enjeux associés au double cryptage des identifiants ont par ailleurs été explicités. Celui-ci permet de protéger les individus, tout en laissant aux chercheurs la possibilité de poursuivre leurs travaux. Je suggère que cette attente figure dans l'avis émis ce jour par la Commission, dans l'espoir que la Depp bénéficie de moyens adaptés.

Des propos rassurant ont été tenus concernant la confidentialité des données individuelles dans le cadre des travaux statistiques et de recherche. Le problème se situe toutefois en amont. En pratique, les fichiers de gestion de la direction générale de l'enseignement scolaire (DGESCO) (Base élève 1<sup>er</sup> degré, Base élèves 2<sup>nd</sup> degré, etc.) contiendront les identifiants et les données personnelles associées. Toutes les interconnexions seront donc possibles.

Le public perçoit la mise en place d'un nouveau dispositif de type SAFARI, sans nécessairement opérer de distinction entre la Depp et la DGESCO. Le débat risque d'enfler sur ce thème et il n'est pas certain que la statistique publique en sorte indemne.

### **Florence AUDIER, CES – Université Paris I**

*In fine*, les chercheurs et les services statistiques devraient pouvoir s'accorder. Les acteurs politiques, quant à eux, se saisissent de plus en plus des données statistiques pour asseoir leurs préconisations. Des expérimentations pragmatiques ont ainsi été lancées dernièrement à partir de statistiques, notamment à l'initiative du Haut commissaire aux solidarités actives. Cette évolution confère aux services producteurs une responsabilité en matière de contrôle éthique et scientifique de la diffusion des données.

### **Bernard SUJOBERT**

Le thème évoqué ce jour soulève des préoccupations depuis des années, sans que la question ait été véritablement posée. Il est satisfaisant qu'un débat s'instaure aujourd'hui. Au sein même de la profession statistique, le débat est vif. Certains déplorent, à l'instar des chercheurs, la perte d'informations suite notamment au retrait de certaines variables de Base élèves ; d'autres revendiquent une certaine prudence, en cohérence avec les positions exprimées par la LDH.

La question plus générale de l'articulation des fichiers administratifs et de leur utilisation à des fins statistiques sera également abordée lors d'un colloque intersyndical organisé par plusieurs syndicats de l'Insee en mars 2011. Les collègues de la Depp pourront aussi s'exprimer dans ce cadre.

A ce stade, la direction de la Depp semble vouloir prendre en compte les objections importantes formulées à l'encontre du projet. Dans ce contexte, la CGT appuie et relaye sans réserve les préoccupations exprimées par la LDH et d'autres personnes ou groupes tels que le CNRBE qui se sont positionnés sur la Base-élèves, tout en regrettant les pertes d'informations résultant de décisions ministérielles se voulant apaisantes. La CGT soutient la production de connaissances utiles aux sociologues et économistes de l'éducation, ainsi qu'aux citoyens, mais assortie de toutes les garanties qui s'imposent. Elle suggère à la Depp de s'appuyer sur le Cnis pour faire valoir le point de vue de la connaissance statistique, notamment dans son dialogue au sein du MEN, en bénéficiant d'un véritable conseil en matière d'information statistique.

La CGT demande qu'une information précise sur les projets en cours soit dispensée dans le cadre du Cnis, rejoignant ainsi la position de l'Autorité de la statistique publique telle qu'exprimée dans l'avis suivant, datant du mois de juin 2010 : « *Les membres de l'Autorité ont relevé en premier lieu l'importance de poursuivre les échanges avec la Depp sur le thème de la construction de la trajectoire*



*scolaire. Ils soulignent, en second lieu, que cet échange très riche, demande à être approfondi pour une meilleure compréhension du sujet. Enfin, ils observent un manque d'éléments de comparaison avec des systèmes d'information analogues délivrant de l'information statistique dans un environnement sécurisé. »*

### **Pierre-Yves GEOFFARD**

S'agissant de l'usage des données individuelles à des fins de gestion, de fichage ou d'évaluation, le cas échéant par des personnes « malintentionnées », l'erreur serait de vouloir exercer un contrôle sur des intentions. En pratique, les meilleures intentions ne sauraient garantir les meilleurs usages. Une vigilance particulière doit être exercée sur les accès aux données individuelles.

Le cryptage, quant à lui, permettra d'améliorer la diffusion des données individuelles, tout en renforçant la protection.

La question des moyens attribués aux SSM et aux centres d'accès a également été abordée. De fait, les opérations de vérification et de protection des bases de données requièrent des moyens techniques et humains, de même que la préparation des bases de données à d'autres usages.

### **Cédric AFSA**

L'accès aux sources locales soulève en effet plusieurs difficultés. Il importe au préalable d'opérer une distinction entre les bases de gestion et les bases statistiques. Les bases statistiques sont construites à partir des bases de gestion. Or certaines opérations de transformation de ces bases de gestion sont extrêmement lourdes, avec des contrôles de qualité et une mise en cohérence des données. De plus, les bases de gestion ne s'appuient pas nécessairement sur des nomenclatures harmonisées. Cette transformation demande donc souvent des moyens substantiels, qui ne sont pas toujours disponibles au sein des Services Statistiques Académiques. Sur ce point, il convient par ailleurs de rappeler que ces services n'ont pas le statut de SSM et demeurent placés sous l'autorité des recteurs.

L'absence de transparence et de communication au cours des dernières années a également été pointée, à juste titre. Nous héritons aujourd'hui d'une situation difficile. De surcroît, cette opacité a abouti à des décisions largement contreproductives. La base élèves du 1<sup>er</sup> degré a ainsi été vidée de son contenu, ce dont nous ne saurions nous satisfaire collectivement. A ce jour, nous ne disposons plus d'aucune information sur la catégorie sociale des élèves, ce qui interdit toute analyse spatiale, en vue notamment d'appréhender les conséquences de l'assouplissement de la carte scolaire. Il nous faut aujourd'hui reprendre le dialogue pour construire un système d'information qui réponde véritablement aux attentes. La question des bases de gestion, qui suscitent à la fois des attentes et des inquiétudes, devra notamment être abordée.

La réflexion doit s'appuyer sur le Cnis. Compte tenu des enjeux en termes de développement de la connaissance et d'évaluation des politiques publiques, il devient aujourd'hui nécessaire d'avancer sur ce sujet. De plus, le problème se situe surtout en amont. Les statisticiens observent scrupuleusement les règles du secret statistique. Des outils de double-cryptage sont notamment mis en œuvre. D'autres acteurs interviennent cependant en amont, parmi lesquels les établissements à l'origine des données, les inspecteurs d'académies, les rectorats, etc. Aucun contrôle ne saurait être exercé à tous les niveaux.

S'agissant du RNIE, la fonction d'immatriculation sera clairement isolée. Le répertoire lui-même contiendra donc très peu d'informations. De surcroît, ces informations seront détruites une fois l'élève sorti définitivement du système éducatif. Les choses auraient été différentes si l'option de collecter le NIR n'avait pas été écartée d'emblée.

L'utilisation des outils de suivi des trajectoires doit donc être encadrée. Cette préoccupation est intégrée dans les réflexions de la Depp. Nous progressons aujourd'hui pas à pas. Le dossier à présenter à la Cnil pour le RNIE est encore en cours de finalisation. Dans ce cadre, nous nous efforçons d'anticiper les conséquences que pourraient avoir la mise en place de l'outil. Un toilettage d'ensemble des fichiers utilisant l'INE est envisageable, avec un regard sur les destinataires des

informations et les interconnexions potentielles. Ce chantier est aujourd'hui ouvert, avec des ressources dédiées.

**Stéphanie POUGET**

Le processus d'attribution de l'INE en lui-même présente un risque. Un processus centralisé permettra de repérer systématiquement tous les enfants nouvellement entrés sur le territoire. On imagine facilement quelle utilisation pourrait être faite d'un tel outil.

**Cédric AFSA**

Seuls les gestionnaires du répertoire auront accès à cette information.

**Stéphanie POUGET**

Une liste des enfants immatriculés à un âge anormalement élevé par rapport aux standards du système éducatif français pourrait néanmoins être éditée.

**Cédric AFSA**

Ce critère n'est pas suffisant. L'enfant peut avoir effectué le début de sa scolarité à l'étranger.

**Stéphanie POUGET**

Des recoupements peuvent être opérés avec le patronyme. Des dérives sont ainsi toujours possibles. En tout état de cause, la preuve n'a pas encore été apportée que les panels ne suffisent pas à répondre à l'ensemble des questions soulevées. Compte tenu des risques, l'opportunité d'utiliser des bases de données exhaustives et un identifiant national mérite d'être appréciée au regard des besoins réels.

En ce qui concerne Base élèves, le problème se trouvait être la remontée de données nominatives par les établissements scolaires. Une enquête anonyme aurait permis de réunir les données nécessaires.

**Hélène PEYTAVI, DGESCO – MEN**

En ce qui concerne l'usage des statistiques en matière de pilotage des politiques éducatives, des enjeux associés au court terme ont été évoqués. Les difficultés rencontrées au niveau des territoires et vis-à-vis de la cohérence nationale de l'outil statistique doivent cependant être explicitées. En matière de pilotage, quelles sont les limites actuelles de l'outil statistique de la Depp ?

**Julien GRENET**

La « voracité » des chercheurs ne saurait remettre en question le respect des règles du secret statistique. Les risques de dérives se situent donc davantage en amont. En pratique, les chercheurs n'ont jamais accès aux données nominatives et n'accèdent que très rarement aux données indirectement nominatives. La publication de statistiques directement ou indirectement nominatives ne saurait non plus être envisagée.

De la même manière, la demande des chercheurs ne porte pas sur l'archivage de données de gestion mais bien sur l'archivage de données anonymisées.

S'agissant de l'utilisation des panels, plusieurs limites peuvent être mises en évidence. Au moment d'étudier l'impact de l'assouplissement de la carte scolaire, il importe de pouvoir analyser les

causalités en mobilisant des cohortes précédant et suivant la réforme. Or les panels ne permettent le suivi que d'une seule et même cohorte. Il faut également pouvoir descendre à un niveau suffisamment fin pour prendre en compte le contexte propre à chaque académie, ce que les panels ne permettent pas non plus.

Au moment d'étudier le développement des inégalités scolaires par les effets de pairs, il est important de mesurer les influences réciproques au sein d'une même classe. Or les panels reposent sur des échantillons ne recouvrant pas les classes de manière exhaustive.

L'enquête génération, quant à elle, ne permet pas d'analyser certains phénomènes tels que l'impact de la création des Parcours actifs d'insertion professionnelle (PAIP) dans les universités. De manière générale, les outils actuels manquent de puissance statistique pour appréhender l'impact des réformes.

### **Cédric AFSA**

Pour construire des indicateurs précis à un niveau infranational, le cas échéant dans une optique de pilotage, il est nécessaire de mobiliser des cohortes plus importantes, donc d'exploiter mieux les données issues de sources administratives.

## **II. Mesures des résultats du système éducatif**

### **1. Les différentes mesures des résultats des élèves**

#### **Cédric AFSA**

Trois éléments permettent de mesurer les compétences acquises par les élèves : les résultats aux examens, les évaluations des acquis des élèves et l'insertion professionnelle.

Les résultats aux examens traduisent le niveau atteint par les élèves. Cette mesure parle immédiatement et demeure largement utilisée en dehors de nos frontières, notamment dans le cadre de la construction d'indicateurs européens. La part des jeunes de 20 à 24 ans possédant un diplôme de l'enseignement secondaire constitue ainsi un indicateur de la stratégie de Lisbonne, avec un objectif de 85 % à atteindre. Des informations plus fines peuvent être également obtenues à travers la collecte des notes attribuées.

Bien que simple à mettre en œuvre et à analyser, la mesure des résultats aux examens s'avèrent souvent imparfaite et incomplète. Par exemple, les résultats au brevet des collèges, reposant sur une part d'évaluation par contrôle continu, ne permettent pas de rendre en compte des effets liés à l'enseignant et empêche les comparaisons entre élèves de classes différentes. L'hétérogénéité des diplômes gêne également les comparaisons entre les niveaux atteints dans l'enseignement secondaire.

Au fil du temps, des dispositifs d'évaluation des acquis des élèves ont été mis en place par la Depp. Ceux-ci permettent d'avoir une bonne mesure du niveau des élèves et fournissent une aide aux enseignants pour situer leurs élèves (dans une logique de pilotage pédagogique). Pour les décideurs politiques, les évaluations des acquis servent également à calculer des indicateurs de pilotage, dans le cadre de la LOLF notamment.

Les évaluations reposent sur un principe de révélation et non d'observation du niveau de l'élève. Le niveau de l'élève constitue ainsi une variable « latente », dont la distribution est modélisée en tenant compte de plusieurs paramètres tels que la difficulté du questionnement. Cette approche assure une comparabilité temporelle et territoriale des résultats entre eux. Elle implique toutefois des protocoles très lourds, avec des tirages d'échantillons et des passations de tests extrêmement contrôlés.

Les évaluations s'articulent autour de différents critères : le moment de l'évaluation (CM2, 6<sup>ième</sup> ou 3<sup>ième</sup>), l'âge de l'individu (9 ans avec l'enquête internationale sur les langues, 15 ans avec l'enquête PISA et 17 ans avec les journées de la citoyenneté), les disciplines évaluées (avec des évaluations mono ou pluridisciplinaires) et les objectifs de l'évaluation (pilotage du système éducatif, comparaisons internationales, comparaisons territoriales ou temporelles).

La mesure de l'insertion professionnelle, enfin, se révèle plus complexe. On touche ici à l'articulation entre les compétences acquises au sein du système scolaire et celles demandées par le marché du travail. La mesure intervient alors à la sortie du système de formation initiale.

## 2. L'insertion professionnelle

### **Olivier LEFEBVRE, SIES - MINISTÈRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE**

Le dispositif d'ensemble pour la collecte nationale de données sur l'insertion professionnelle des diplômés de l'Université est en construction, avec pour objectif de produire des analyses au niveau national et au niveau de chaque université. Des indicateurs bruts seront produits (taux d'insertion, ventilation par discipline, caractéristiques des emplois occupés, etc.), ainsi que des indicateurs enrichis (prenant en compte les effets de structure de l'offre de formation). Les mesures devront être homogènes et comparables d'une université à l'autre, d'où l'importance de construire un dispositif coordonné.

Le dispositif suscite aujourd'hui des attentes fortes au niveau national et au niveau local. En lien avec les missions du système d'enseignement supérieur telles que définies par la loi LRU (loi relative aux libertés et responsabilités des universités), chaque établissement d'enseignement supérieur est tenu de favoriser l'insertion professionnelle de ses diplômés, mais également de produire, de mobiliser et de diffuser des indicateurs sur le sujet. Cette volonté de mesure de l'insertion professionnelle s'exprime aussi au niveau national, à travers la LOLF et le dialogue avec les parlementaires. Des attentes ont été exprimées en termes d'information (des étudiants, des enseignants, des universités, etc.), de pilotage (à l'échelle nationale et au niveau des universités) et de mesure de la performance d'ensemble du système éducatif (dans le cadre de la LOLF et du dialogue entre le MESR et les universités).

La construction du dispositif s'est appuyée sur les enquêtes déjà menées par un certain nombre d'universités auprès de leurs diplômés de masters. Ces enquêtes n'étaient toutefois pas homogènes, avec des divergences dans les dates d'interrogation, les questionnements, etc. D'autres enquêtes nationales sur les titulaires de DUT et de licences professionnelles (pilotees par le MESR et l'association des directeurs d'IUT) ont été mobilisées.

Entre décembre 2009 et avril 2010, une première enquête coordonnée a été menée sur les étudiants diplômés en 2007. Il s'agissait de produire des informations dans toutes les universités sur la situation des diplômés 30 mois après l'obtention du diplôme, afin d'avoir une vision homogène et stabilisée de l'insertion professionnelle. Un tronc commun de 27 questions a été administré. Chaque université était tenue d'interroger les mêmes catégories de diplômés. Le taux de réponse moyen est de 64 %. Des premiers résultats sur le champ des masters ont été publiés le 15 octobre 2010. L'enquête a par ailleurs été coordonnée avec les enquêtes nationales sur les titulaires de DUT et de licences professionnelles. La coordination a porté sur la date d'interrogation, le questionnement et le champ des enquêtes.

Il est envisagé de reconduire le dispositif à l'identique pour interroger les diplômés de 2008. Une consolidation sera opérée à partir des acquis d'expérience de l'enquête précédente. L'objectif est ensuite d'élargir progressivement le champ du dispositif aux licences générales et aux doctorants, en concertation avec les universités. Les aspects de poursuite d'études devront être traités.

Les résultats de la première enquête ont été diffusés sur le site du MESR, pour les 70 universités dont les résultats étaient exploitables (avec un taux de réponse et des effectifs suffisants). Les résultats par université sont présentés par ordre alphabétique, sans logique de classement. Les effectifs et les taux de réponse sont indiqués pour apporter un éclairage sur la qualité des informations proposées et la

pertinence des comparaisons. Des taux d'insertion globaux et par discipline sont fournis pour chaque université, de même que des indicateurs de caractérisation des emplois occupés (catégorie professionnelle, type de contrat, secteur d'activité, répartition public/privé/associatif, etc.). Des indicateurs prenant en compte les effets de structures de l'offre de formation (voire de la population étudiante) seront produits ultérieurement. Des résultats nationaux ont également été présentés.

L'enquête sur les diplômés de 2007 a porté sur les diplômés français ayant obtenu leur diplôme dans le cadre de la formation initiale et n'ayant pas poursuivi ni repris d'études après leur diplôme, soit 48 000 diplômés de master, 16 000 titulaires d'un DUT et 26 000 titulaires d'une licence professionnelle. Le champ ainsi constitué a été interrogé de manière exhaustive, 30 mois après l'obtention des diplômes. Un tronc commun de 27 questions a été utilisé pour identifier les reprises d'études, connaître les situations professionnelles à la date d'observation et caractériser les emplois occupés. L'enquête sur les diplômés de 2008 intégrera une question sur la localisation des emplois, en vue de permettre une analyse géographique des débouchés.

Les universités sont responsables de la collecte des données et de leur qualité mais doivent en rendre compte (un bilan de collecte permettant d'apprécier la qualité des données transmises leur était demandé). Le choix des modes de collecte était laissé à leur appréciation (questionnaire papier, réponse par internet, enquête téléphonique). Nombre d'universités ont privilégié des approches multi-modes. Les données individuelles collectées étaient ensuite transmises au MESR après anonymisation.

Une charte a été signée par les universités pour organiser la collecte. Les universités se sont ainsi engagées à garantir l'homogénéité, la comparabilité et la confidentialité des données. En matière de traitement et de protection des données collectées, le dispositif a été approuvé par la Cnil. Les universités ne disposant pas par ailleurs d'un dispositif centralisé s'appuyant sur un correspondant Cnil étaient tenues d'adresser une déclaration à la Cnil. Des informations sur l'utilisation purement statistique des données étaient portées sur les questionnaires.

Le MESR assurait un soutien financier et technique dans le cadre de la collecte, en exigeant en retour un bilan qualité sur la collecte et les données.

Les enquêtes d'insertion (annuelles et exhaustives sur champ) s'inscrivent dans une logique de complémentarité avec l'enquête génération (triennale et portant sur un échantillon représentatif de l'ensemble des sortants du système éducatif). Les enquêtes d'insertion autorisent des exploitations par université et par domaine d'études alors que l'enquête génération permet des comparaisons entre diplômés et non diplômés, ainsi que des comparaisons entre formations.

## **Stéphane JUGNOT**

Les enquêtes « Génération » réalisées par le Cereq ont pour objet d'étudier l'insertion professionnelle des jeunes sortis de formation initiale, en fonction de leur niveau de formation et de leur spécialité. Elles permettent l'étude des disparités existantes selon différentes caractéristiques (l'origine sociale, le parcours scolaire, l'origine nationale, etc.), ainsi que des analyses en termes d'évolution.

Contrairement aux panels constitués à partir d'une date d'entrée à un niveau précis du système éducatif, les enquêtes « Génération » sont bâties à partir d'une date d'entrée sur le marché du travail. Les populations interrogées connaissent donc une même situation sur le marché du travail à une date donnée. Les trajectoires peuvent ensuite être comparées en fonction des niveaux de diplômes, sans nécessité d'isoler les effets conjoncturels.

Il n'existe pas de définition unique de l'insertion professionnelle. Une approche multicritères est donc privilégiée. Le questionnement permet ainsi d'étudier plusieurs dimensions telles que le fait d'être en emploi ou non, la nature du contrat de travail, le secteur d'activité, le niveau de rémunération, l'adéquation entre la formation suivie et l'emploi occupé, etc. L'accent est également mis sur les trajectoires, avec un suivi du parcours mensuel des individus durant les trois premières années suivant la sortie de formation initiale. Le délai d'accès à un emploi à durée indéterminée peut ainsi être étudié. Grâce au calendrier mensuel, différents types de parcours peuvent aussi être cernés : les situations d'accès rapide à l'emploi, les situations de décrochage, les situations d'accès différé à l'emploi, les situations de retour à la formation, etc. De telles analyses peuvent ensuite être déclinées selon le type de formation (niveau de diplôme, filière générale ou professionnelle, apprentissage, etc.).

L'insertion est donc étudiée avec du recul et dans la durée. Une enquête à 6 mois aurait tendance à accentuer les différences d'efficacité dans l'accès au marché du travail en fonction des formations (grandes écoles, masters, etc.).

Les enquêtes « Génération » s'intéressent aux jeunes sortis de formation (qu'ils soient diplômés ou non) et résidant en France. La formation initiale est définie par l'absence d'interruption de plus de 12 mois au cours de la scolarité.

*In fine*, le dispositif permet d'étudier les disparités à tous les niveaux de formation, avec un traitement homogène en termes de questionnaire, de date d'enquête et de traitement de la non-réponse. Le dispositif est ainsi complémentaire avec les enquêtes ciblées telles que les enquêtes sur les sortants de masters ou les enquêtes d'insertion sur les établissements d'enseignement agricole. Cependant, le dispositif nécessite de mobiliser de gros échantillons, de 10 000 à 30 000 répondants. Des extensions thématiques ou géographiques sont également proposées.

Les enquêtes « Génération » sont des enquêtes de la statistique publique, présentées au Cnis pour avis d'opportunité et soumises à un examen par le Comité du label.

En termes de calendrier, une première interrogation intervient trois ans après la sortie de la formation initiale. Tous les trois ans une nouvelle cohorte est interrogée. Certaines font l'objet de réinterrogations. L'échantillon de l'enquête « Génération 2007 » a ainsi été interrogé en 2010 et ne sera pas réinterrogé (la publication des résultats aura lieu en 2011). La cohorte suivante, la « Génération 2010 », devrait être interrogée en 2013 puis réinterrogée en 2015 et 2017.

En pratique, les enquêtes « Génération » nécessitent de constituer des bases de sondage *ad hoc*, avec l'appui de la SIES et de la Depp. Des fichiers d'inscrits sont récupérés auprès des établissements pour permettre le tirage des échantillons. La récupération des numéros de téléphone des individus est assurée par un prestataire externe. Cette opération est complexe et coûteuse, du fait de la multiplication des opérateurs et des difficultés rencontrées pour retrouver certains individus. La collecte est ensuite réalisée par CATI.

Pour l'enquête génération 2004, un premier échantillon de 372 000 individus a été constitué. Parmi ces individus, 84 000 n'ont pu être associés à des coordonnées téléphoniques, 172 000 n'ont pu être contactés ou ont refusé de répondre, 50 000 se sont révélés être hors champ (car ayant poursuivi leurs études). En définitive, 55 000 individus ont été interrogés. Des variables complémentaires (au-delà de la formation et de l'établissement) auraient permis de mieux traiter les non-réponses.

Des contraintes budgétaires fortes conduisent aujourd'hui à envisager de redessiner le dispositif. Des sources administratives telles que le fichier historique des demandeurs d'emploi inscrits à Pôle Emploi et celui des DADS pourraient notamment être mobilisées. L'inscription à Pôle Emploi et la caractérisation des situations d'emploi pourraient faciliter le traitement des non-réponses. Des informations complémentaires à celles récoltées à travers le calendrier mensuel pourraient également être obtenues. Les sous-déclarations d'emplois de courte durée devraient ainsi être évitées.

Le dispositif pourrait par ailleurs évoluer vers un questionnement par Internet, ce qui permettrait d'intégrer les DOM et les jeunes partis à l'étranger dans le champ des enquêtes. Enfin, le questionnaire pourrait être adapté à des problématiques d'intérêt.

En conclusion, je rappellerai qu'un certain nombre de résultats et d'études sont aujourd'hui disponibles sur le site du Cereq. Un document méthodologique centré sur l'enquête génération 2004 a également été édité en 2010.

### 3. Présentation pour opportunité d'enquêtes

#### a. Enquête sur l'insertion des sortants de Master – SIES

##### **Pierre-Yves GEOFFARD**

Cette enquête vient d'être largement évoquée. Je propose donc d'ouvrir le débat sur son opportunité, ainsi que sur l'ensemble des points qui viennent d'être abordés.

##### **Jean-Pierre LE GLEAU, INSEE**

J'interviendrai au nom du secrétaire général du Cnis, retenu à l'étranger. L'enquête sur l'insertion professionnelle des sortants de masters présente la particularité d'être réalisée par quelque 80 universités et non pas par un seul et même service producteur. Cette approche soulève des interrogations quant à l'homogénéité et à la qualité des données.

Le sujet traité par cette enquête mérite certainement un approfondissement de la connaissance. L'enjeu est donc de manifester un soutien au projet présenté, tout en formulant des exigences en matière de respect des critères de qualité statistique. Un avis d'opportunité favorable devrait être ainsi pondéré.

Le projet d'enquête précise bien que le responsable de l'enquête est le SSM du MESR. La collecte, quant à elle, étant déléguée aux universités et IUT.

En matière de protection du secret statistique, il s'agit d'affirmer clairement que les universités et IUT auront en charge de réaliser une enquête statistique. Les informations individuelles recueillies ne pourront donc être utilisées qu'à des fins statistiques et non pas à des fins de gestion administrative, de mise à jour de fichiers, de mailing, etc. L'accès aux données devra pour cela être contrôlé à l'intérieur même des universités et réservé à des personnes connues du service producteur (à savoir le SSM du MESR).

Le protocole de l'enquête, reposant sur un tronc commun, ne pourra pas faire l'objet d'adaptations (rajout/suppression de questions, modifications de présentation, etc.). Le mode d'administration du questionnaire peut varier, mais devra se conformer aux prescriptions du service producteur. Les biais susceptibles d'être introduits par les modes de collecte devront être écartés. Les éventuelles questions complémentaires, ne relevant pas du tronc commun, seront clairement distinguées. Leur caractère non obligatoire sera très clairement indiqué aux enquêtés.

Le service producteur assume la responsabilité de diffuser les résultats nationaux – c'est-à-dire des résultats d'ensemble de l'enquête. Chaque établissement pourra quant à lui diffuser les résultats le concernant. Conformément aux principes du Code de bonnes pratiques de la statistique européenne, la diffusion des résultats statistiques sera bien distincte de la communication politique,

Si ces exigences étaient inscrites en réserve dans un avis d'opportunité, c'est au Comité du label qu'il appartiendrait d'en vérifier la prise en compte.

##### **Pierre-Yves GEOFFARD**

Il est effectivement prévu que le Code de bonnes pratiques de la statistique européenne s'applique à l'ensemble de la statistique publique et des pouvoirs publics, pour encadrer la collecte, l'exploitation et la diffusion des données statistiques.

##### **Jean-François ROYER, Société française de statistique**

Une alternative méthodologique aux enquêtes d'insertion professionnelle des élèves et étudiants devra d'être étudiée à moyen terme. La Société française de statistique (SFdS) promeut depuis plusieurs années l'utilisation d'appariements sécurisés de fichiers administratifs exhaustifs. Le champ de

l'insertion professionnelle pourrait être ainsi couvert. Des identifiants hachés et cryptés éviteraient les risques d'utilisations inappropriées détaillées lors du précédent débat. De surcroît, les fichiers seraient constitués, uniquement pour la production de statistiques et non de manière permanente.

De telles procédures se répandent aujourd'hui, en particulier dans le champ de la santé. Dans le champ examiné aujourd'hui, des informations provenant des établissements d'enseignement pourraient être rapprochées d'informations provenant du monde de l'emploi (demandeurs d'emploi, DADS, etc.). Des fichiers exhaustifs seraient ainsi constitués (les taux de réponse constituant une des limites des enquêtes présentées), avec toutefois des possibilités moindres d'adaptation aux questionnements (dans leur dimension qualitative notamment).

En tout état de cause, les difficultés rencontrées dans l'introduction de ces procédures ne sauraient conduire à écarter d'emblée cette approche, dorénavant bien acceptée, voire encouragée par la Cnil. Des dispositions législatives ont même été adoptées, dans certains cas, pour en faciliter la mise en œuvre.

### **Florence AUDIER**

L'enquête sur l'insertion des sortants de masters a fait l'objet d'une première publication dans la grande presse, et non pas dans les supports statistiques habituels. L'objectif affiché était de mesurer l'insertion professionnelle des diplômés de l'université. En réalité, il s'est agi de produire une photographie ponctuelle à 30 mois, avec des éléments très sommaires concernant les emplois et les statuts. Le questionnaire utilisé était « lacunaire et pauvre », ne permettant pas de caractériser réellement l'insertion professionnelle.

Le porte-parole du réseau des observatoires universitaires a fait part de sa « consternation » et de son « indignation » à l'annonce du premier palmarès des universités issu de cette enquête, les résultats ayant été présentés sous forme de « classements » à la performance. Pierre Dubois, créateur du site educpro.fr, a quant à lui dénoncé « une pure instrumentation de la statistique par le politique ». En tant qu'ancien directeur d'observatoire universitaire, il a fait part de sa « honte », en titrant son commentaire de la façon suivante : « le classement des universités est juste bon à jeter à la poubelle ».

Pour ma part, je rappellerai que, si des résultats ont bien été publiés sans commentaire sur le site du MESR, ils ont également fait l'objet d'une publication exclusive dans le Figaro Magazine, avec des commentaires de la ministre de l'enseignement supérieur et de la recherche.

Sur le fonds, l'enquête souffre d'importantes imprécisions méthodologiques. La définition du champ manque de clarté, en particulier dans la prise en compte ou non des doctorants. La base de sondage, soit disant exhaustive, ne correspond pas au nombre des diplômés mentionnés dans d'autres publications du MESR ou sur les sites des universités.

Au final, six classements dont un général ont été publiés dans le Figaro Magazine. Les résultats concernant 68 universités (sur un total de 89) ont été diffusés, avec des taux de réponses exploitables extrêmement variables d'une université à l'autre (de 37 à 93 %). Aucun redressement n'a été opéré. De surcroît, aucune distinction n'était faite entre les masters professionnels et les masters recherche ce qui est indispensable vu la thématique de l'enquête. En outre, le fait de procéder par classement sous-entend que les universités ont une maîtrise du marché du travail. Le classement général, enfin, s'appuie sur des écarts extrêmement faibles entre universités. Pour ce qui est de l'indicateur général, par exemple, 28 universités se situent dans une fourchette de 2 %, très en deçà du taux d'erreur de la statistique.

En ce qui concerne l'attribution des caractéristiques d'excellence, l'exemple lyonnais, parmi d'autres, s'avère révélateur. L'université de Lyon II, qui est une université de sciences humaines et sociales (SHS), se trouve positionnée à la 40<sup>ième</sup> place au classement général en sciences humaines et sociales ainsi qu'en droit, économie et gestion et en 18<sup>ième</sup> et dernière place en langues, arts et lettres. En revanche, elle se retrouve à la 1<sup>ère</sup> place en sciences. Cette aberration est due à une concentration, en sciences, de diplômés en informatique qui bénéficient d'un taux d'emploi de 97,8 %. En réalité, ces formations ne recouvrent que 3 des 68 masters de l'université Lyon II. L'université Lyon I, à dominante scientifique a des performances excellentes, quant à elle ... en droit, économie et gestion.



Dans ce contexte, la responsabilité endossée par la ministre de l'enseignement supérieur et de la recherche est lourde. En réalité, l'enquête est indissociable de son utilisation (mesures nationales et locales de l'insertion pour éclairer les étudiants, indicateurs de pilotage, mesure des effets de structure, indicateurs pour le Parlement, indicateurs pour la LOLF, outil du dialogue de performance, etc.). La ministre a ajouté que les dotations financières aux universités seraient régulées en fonction des performances en termes d'insertion professionnelle.

Un avis d'opportunité favorable accordé à cette enquête irait à l'encontre des efforts fournis légitimement par les services statistiques des universités. Les risques d'utilisations inappropriées – car trop éloignées des critères de sérieux scientifiques – seraient, de surcroît, trop importants.

### **Bernard SUJOBERT**

Le projet d'enquête est peu explicite. La marge de manœuvre très large laissée aux universités suscite des inquiétudes quant à l'homogénéité et à la qualité des données. De plus, les conditions de publication des résultats sous la responsabilité du SIES demeurent imprécises et paraissent minimales sur un sujet associé à des enjeux politiques aussi importants.

La ministre de l'enseignement supérieur et de la recherche, en promouvant spectaculairement un « classement des universités » fin septembre 2010, a, pour sa part, mis en évidence les risques de mauvais usage de ce type d'enquête, en affirmant qu'il s'agissait d'un « palmarès » et d'un outil de pilotage de l'enseignement supérieur. L'opportunité d'une enquête de la statistique publique ne saurait être appréciée sans prendre en compte cette dimension.

De fait, les compléments apportés en séance par Florence Audier confirment bien les interrogations que nous avons à la lecture de la fiche de présentation de l'enquête. La seule invocation du Code de bonnes pratiques de la statistique européenne semble représenter une garantie insuffisante sur un terrain aussi sensible. En pratique, les universités ne possèdent ni l'expérience ni la culture des enquêtes statistiques telles que mises en œuvre par la statistique publique. L'encadrement par le SIES dans cette opération, quant à lui, n'est guère défini.

Une option serait d'accorder un avis d'opportunité favorable, avec un nombre important de réserves ; une autre option serait de solliciter la production par le SIES d'un projet plus détaillé, notamment pour les aspects ayant trait à la méthodologie de collecte et à la diffusion des informations statistiques. L'enjeu est de garantir la production et la publication de statistiques complètes, indépendantes et fiables, sous la responsabilité du SIES. La CGT, pour ce qui la concerne, est favorable à la seconde option.

### **Joël CREUSAT, SSM Justice**

Dispose-t-on d'une mesure du biais de sélection entre les répondants et les non répondants à l'enquête ? Le ministère de la justice fait face à des problématiques similaires dans le cadre des enquêtes auprès des victimes.

### **Stéphane JUGNOT**

Le cryptage, bien qu'utile pour procéder à des appariements de fichiers, ne semble pas constituer une « solution miracle » pour anonymiser les données. Dans certains cas, des données relatives à l'âge, au sexe ou à la localisation géographique peuvent s'avérer indirectement nominatives.

### **Pierre-Yves GEOFFARD**

Le Cnis jouerait son rôle en rappelant la nécessité pour la statistique publique et les pouvoirs publics d'appliquer le Code de bonnes pratiques de la statistique européenne. Pour ce qui est du projet d'enquête tel que présenté ce jour, un avis d'opportunité favorable ne pourrait être accordé qu'avec énormément de réserves. Si la question semble pertinente, il convient de faire en sorte qu'elle soit traitée à travers une enquête conforme aux critères de la statistique publique. Sans doute serait-il

préférable de réexaminer à nouveau la demande d'avis d'opportunité, sur la base d'un projet prenant en compte les remarques formulées.

### **Olivier LEFEBVRE**

La base d'enquête comprend les diplômés français de masters, ayant obtenu leur diplôme dans le cadre de la formation initiale et n'ayant pas repris ou poursuivi d'études. Cette définition du champ d'observation exclut les doctorants et explique l'écart avec le nombre total des diplômés.

Des redressements de la non-réponse ont bien été réalisés, à partir d'une analyse des différents facteurs de non-réponse (l'université d'appartenance, le secteur disciplinaire, etc.). Du reste, des taux de réponse relativement proches ont été obtenus par grande discipline (65 % pour les sciences et les SHS, 62 % pour le droit l'économie et la gestion, 60 % pour les arts et lettres).

L'objectif était de produire des informations homogènes et comparables d'une université à l'autre, d'où l'adoption d'un questionnaire commun et d'une charte de l'interrogation. Pour la première enquête, un questionnaire relativement simple a été retenu afin de maximiser les chances d'obtenir des taux de réponse pertinents. Ce choix a été fait en concertation avec la Conférence des Présidents d'Universités (CPU), le Cereq et différents utilisateurs des données.

En ce qui concerne l'allocation des moyens aux universités, le modèle actuel repose à 80 % sur des indicateurs d'activité (étudiants présents aux examens) et à 20 % sur des indicateurs de performance (publications et réussite aux examens). Pour l'heure, l'insertion professionnelle ne figure pas dans les indicateurs de performance, du fait même que le dispositif est en cours de construction. Dans un premier temps, l'objectif est que les données produites servent dans le cadre du dialogue contractuel entre l'État et les établissements, autour de la définition des politiques d'insertion professionnelle.

Les données publiées sur le site du MESR sont accompagnées d'éléments sur les effectifs de la base d'enquête et les taux de réponse par université, afin que les utilisateurs puissent avoir des indications sur la significativité des comparaisons.

Effectivement, la publication de données par université facilite la production de classements. Cela étant, telle ou telle utilisation qui serait jugée non-conforme ne saurait remettre en cause le dispositif.

En ce qui concerne l'encadrement par le SIES, la fiche descriptive d'enquête a probablement été présentée de manière trop synthétique pour le faire ressortir ; en réalité, la sous-direction a piloté l'ensemble de la concertation autour de l'enquête et a participé à la construction du questionnaire, à la mise en place de la charte de l'interrogation, etc. Un système sécurisé de remontée des données a également été développé. Un travail a été mené sur la qualité des données (correction des incohérences, redressement de la non-réponse). Une analyse des bilans de la collecte a été ébauchée. Le SIES continuera également de travailler à la production d'indicateurs enrichis.

### **Pierre-Yves GEOFFARD**

Je propose que le SIES soumette, à la prochaine commission, un projet plus détaillé et tenant compte des suggestions méthodologiques formulées, pour permettre de traiter cette question pertinente à travers une enquête statistique adaptée.

*b. Panel d'élèves entrant au CP en 2011 – Depp*

### **Jean-Paul CAILLE, DEPP**

La Depp souhaiterait mettre en place un 8<sup>ème</sup> panel d'élèves, le 3<sup>ème</sup> consacré aux élèves du 1<sup>er</sup> degré. Le déroulement de la scolarité élémentaire joue un rôle capital dans la poursuite des parcours scolaires. Or les données dont nous disposons actuellement sont issues du panel de 1997. De surcroît, le fait que la base élèves du 1<sup>er</sup> degré ne contienne plus d'informations sur les familles nous prive de statistiques descriptives élémentaires telles que les taux de redoublement en fonction du milieu social.

Le nouveau panel a pour objectif de décrire et d'expliquer les carrières et performances scolaires des élèves depuis le cours préparatoire jusqu'à l'entrée en sixième, de mesurer l'égalité des chances dans l'enseignement élémentaire (par rapport à l'origine sociale, l'immigration, les niveaux de revenus, etc.), d'analyser les facteurs de progression d'acquis du cours préparatoire à la fin du CM2, d'évaluer l'effet des politiques mises en place dans l'enseignement élémentaire (stages de remise à niveau, programmes personnalisés de réussite éducative, etc.) et d'observer l'évolution des pratiques culturelles (en collaboration avec le SSM du ministère de la culture).

Ce nouveau panel présente deux nouveautés par rapport à celui de 1997. D'une part, il comportera une évaluation des acquis en fin de CM2, et pas seulement au début du CP ; d'autre part, les familles seront aussi réinterrogées en fin d'enquête pour prendre en compte les évolutions survenues dans la situation familiale du jeune depuis son entrée à l'école élémentaire.

### **Annick KIEFFER**

Tous les travaux sur les parcours dans le secondaire ont démontré l'importance d'approfondir également la connaissance des parcours dans le primaire. Jusqu'à présent, les données de panels concernant le primaire ont été peu communiquées au centre Quételet. Une transmission plus rapide serait souhaitable.

### **Jean-Paul CAILLE**

Les données du panel de 1997 ont été communiquées à un certain nombre de laboratoires. En ce qui concerne la mise à disposition des données *via* le centre Quételet, une documentation devrait être réalisée en 2011.

*Le projet de panel d'élèves entrant au CP en 2011 recueille un avis d'opportunité favorable.*

## **III. Point sur le projet ELFE**

### **Henri LERIDON, INED – ELFE**

Notre intervention sera focalisée sur les écarts entre le projet tel qu'il est aujourd'hui et le projet initial. En préambule, je rappellerai que le projet ELFE a reçu un avis d'opportunité favorable du Cnis en octobre 2006. La première étape du projet, à savoir l'enquête en maternité à 2 mois, a ensuite recueilli un avis favorable du Comité du label en mai 2008. Compte tenu de certains changements intervenus dans la collecte, ce dernier avis devra être renouvelé.

En matière de gouvernance, le projet est passé d'un pilotage à travers un groupement d'intérêt scientifique à un pilotage par une unité mixte INED-Inserm, dirigée par Marie-Aline Charles. La gestion du projet incombe toujours à l'INED, avec une responsabilité attribuée à l'Inserm vis-à-vis des prélèvements biologiques. J'assume pour ma part un rôle de conseiller scientifique auprès du projet. Le comité de pilotage regroupe toujours les partenaires historiques du projet. Seule la Depp du ministère de l'Éducation nationale s'en est retirée formellement. La direction générale de la prévention des risques (DGPR) du ministère de l'écologie, du développement durable, des transports et du logement s'y est en revanche associée. Le conseil scientifique continue d'examiner l'ensemble des projets de recherche. Un groupe de représentants des familles participant à l'étude pilote a été constitué à côté du Groupe Éthique.

Du point de vue des partenariats institutionnels, l'Insee s'est désengagé de la réalisation de la collecte, pour des raisons budgétaires. L'enquête en face-à-face à 6-8 semaines a alors dû être remplacée par une enquête téléphonique. L'enquête à domicile à 3 ans, en revanche, a été maintenue. Des tests sont d'ailleurs en cours de réalisation avec un institut privé.

Des contacts ont par ailleurs été maintenus avec la Depp et la DGESCO du ministère de l'éducation nationale (MEN). Le groupe thématique « école » a notamment été relancé, afin que des données

précises puissent être mobilisées sur les parcours scolaires et qu'en retour, le MEN puisse bénéficier d'une meilleure contextualisation de l'environnement de l'enfant.

Des contacts ont également été établis avec la CNAM pour obtenir l'accès aux données du SNIIRAM. Un partenariat national a été défini avec l'établissement français du sang (EFS) pour la collecte biologique en maternités.

En matière de financement, les phases pilotes ont été soutenues par les partenaires historiques et par le biais de contrats avec l'agence nationale de la recherche (ANR) et le centre commun de documentation en sciences humaines et sociales (CCDSHS). A la mi-2009, un arbitrage interministériel a été obtenu pour garantir le financement des premières opérations de collecte (de 0 à 1 an) par les ministères de la recherche, de la santé et de l'environnement, en partie *via* l'appel d'offres TGIR-Cohortes<sup>2</sup>. Un autre dossier, en cours d'instruction dans le cadre du grand emprunt, pourrait déboucher sur un appui financier sur 10 ans. La recherche des financements a entraîné un décalage du démarrage du projet. Celui-ci est aujourd'hui prévu pour le mois de mars 2011.

### **Marie-Aline CHARLES, Inserm – ELFE**

Des modifications méthodologiques ont été décidées suite à la réalisation du pilote. Elles visent surtout à simplifier la logistique de collecte. Alors qu'il était prévu de travailler sur les 542 maternités de France pour recruter les 20 000 enfants de la cohorte, un tirage au sort des maternités a finalement été décidé. De ce fait, le nombre de jours d'enquête a dû être augmenté (de 16 à 24), avec pour conséquence un découplage partiel (à hauteur de 50 %) avec l'échantillon démographique permanent (EDP). La vague d'enquête du 2 janvier a été supprimée, au vu de la complexité induite par la collecte et de la justesse du calendrier.

Des contraintes additionnelles, techniques et financières, ont par ailleurs été imposées dans le cadre de la collecte biologique. L'objectif est d'obtenir des prélèvements sur 10 000 enfants et non plus sur 20 000 enfants.

L'échantillonnage des maternités a été calibré pour aboutir à une surreprésentation des établissements de grande taille, susceptibles de permettre des recrutements plus importants. Cet élément devra être pris en compte lors de l'exploitation statistique des résultats. 342 maternités ont ainsi été sélectionnées. Avec un taux d'abandon de la part des maternités estimé à 10 % et un taux d'acceptation de l'enquête par les mères estimé à 55 %, le recrutement de 20 000 enfants devrait être assuré.

L'enquête sera lancée le 28 mars 2011 et s'étalera sur 4 plages de 6 jours : du 28 mars au 4 avril, du 27 juin au 4 juillet, du 27 septembre au 4 octobre et du 12 décembre au 20 décembre. La dernière vague étant complètement découplée de l'EDP.

Du fait du report de l'opération, la collecte d'ELFE dans les maternités aura lieu en même temps que celle de l'enquête Epipage 2<sup>3</sup>. Certains moyens logistiques pourront ainsi être mutualisés (recrutement des coordinateurs, mise en place du Système d'information, etc.). L'enquête Epipage continuera de suivre les grands prématurés tandis que la cohorte d'enfants témoins sera suivie dans le cadre du projet ELFE.

Pour le lancement de l'enquête nationale, l'accord du CCTIRS (comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé du ministère de l'enseignement supérieur et de la recherche) a été obtenu. Les démarches auprès de la Cnil devraient être engagées en décembre 2010. Une demande d'autorisation de collection biologique sera déposée auprès de l'Inserm, ainsi qu'une demande d'avis d'opportunité auprès du comité de protection des personnes compétent.

---

<sup>2</sup> Le projet Très grandes infrastructures de recherches-Cohortes mis en place par le ministère de l'enseignement supérieur et de la recherche et celui de la santé vise à favoriser le développement des grandes cohortes qui sont indispensables à la recherche médicale. Dans ce cadre, l'IRESF lance chaque année un appel d'offre.

<sup>3</sup> L'enquête Epipage est menée par l'Inserm et porte sur la prématurité.

## Henri LERIDON

En ce qui concerne le stockage de l'information, l'ambition de départ était de développer un système qui dissocie les données proprement dites et les données nominatives, avec également un stockage différencié des bases correspondant aux différentes enquêtes. L'idée est dorénavant d'atomiser et d'autonomiser la totalité des données. Chaque variable sera dotée d'un identifiant spécifique. Une table de correspondance permettra de faire le lien entre les identifiants. Toute demande de traitement sera instruite et donnera lieu à la production d'une base *ad hoc*. Le dispositif permettra une traçabilité complète des opérations, y compris celles réalisées par l'équipe de gestion du projet. Des systèmes de protection et de certification classiques seront mobilisés par ailleurs. Le Cnis pourra être informé plus en détail de la mise en place de ce système novateur qui fera l'objet d'une demande de certification ISO.

Une structure spécifique, le Comité d'accès aux données Elfe (CADE), examinera toutes les demandes d'accès aux données, en veillant notamment à minimiser les risques d'identification indirecte. Après son accord, le traitement sera entièrement automatisé et intégralement « tracé ». La composition du CADE sera définie par le Conseil scientifique du projet, en accord avec les règles définies par la structure de gestion des cohortes à l'IRESP.

## Marie-Aline CHARLES

L'enquête en maternités reposera sur une collecte informatisée, de façon à limiter les erreurs de saisie. La télétransmission des données se fera *via* un serveur sécurisé – ce qui permettra un suivi en temps réel de la collecte. Le recrutement du personnel de collecte est en cours.

En parallèle, la cohorte pilote continue de faire l'objet d'un suivi, avec le démarrage de l'enquête à domicile à 3 ans.

## Bernard SUJOBERT

Les modifications apportées ne répondent pas totalement aux attentes, notamment elles n'assurent pas une meilleure visibilité du rôle des différentes instances pilotant le projet. Depuis 2007, l'évolution du projet demeure relativement opaque.

En ce qui concerne la protection des données, nous prenons acte d'une volonté d'ouverture à une expertise extérieure.

Des incertitudes persistent en revanche sur la gestion du consentement des personnes. Le fait que l'étude porte sur des enfants et corresponde à un engagement sur une durée de 20 ans soulève de très nombreuses interrogations.

D'un point de vue philosophique, le sens d'une accumulation de traces numériques sur les individus mériterait également de faire l'objet d'une réflexion approfondie. En pratique, le destin d'un individu ne saurait être résumé à ses traces numériques. Or, il n'existe guère de lieu pour aborder ces questions, alors même que le coût du projet est très élevé.

## Henri LERIDON

L'enjeu est de privilégier une approche pluridisciplinaire et multifactorielle. Il s'agit, par exemple, de faire en sorte que l'étude des trajectoires scolaires ne soit pas limitée à un certain nombre de facteurs explicatifs mais puisse prendre en compte des facteurs économiques, d'environnement social, d'environnement familial, etc. L'ambition est de mettre en évidence les interactions entre ces différents facteurs, d'où la nécessité de recueillir un grand nombre d'informations dans le cadre d'un même suivi.

Pour ce qui est de l'information du Cnis, un point a été fait chaque année sur l'avancement du projet. Un courrier lui a également été adressé en 2010. Dans l'intervalle, des contacts ont été pris avec la Cnil pour aborder notamment les questions relatives au consentement, au droit des individus devenant plus âgés, etc. Les chartes correspondantes sont en cours d'élaboration.

**Pierre-Yves GEOFFARD**

Le projet permettra notamment de travailler sur le lien présumé entre inégalités de santé et inégalités socio-économiques. Des impacts à très long terme des conditions de vie au plus jeune âge pourraient ainsi être étudiés.

#### **IV. Propositions d'avis**

*Les projets d'avis ont été adoptés et figurent en annexe.*

#### **V. Examen de demandes d'accès à des fichiers au titre de l'article 7bis de la loi de 1951 présentées par la Direction des études et des synthèses économiques de l'Insee**

*La demande d'accès aux données des bases RELAIS de 2003/2004 à 2010/2011, adressée par la DESE de l'Insee à la Depp, est acceptée.*

### **Conclusion**

**Pierre-Yves GEOFFARD**

Merci à tous.

*La séance est levée à 18 heures 45.*

---

DOCUMENTS PRÉPARATOIRES

---

- Fiches descriptives des enquêtes présentées en opportunité
- Fiche de demande d'accès au titre de l'article 7bis de la loi de 1951 à des données administratives
- Avant-projet de programmes statistiques 2011 de la commission Services publics et services aux publics (consultable sur le site : [http://www.cnis.fr/agenda/PDE/PDE\\_0152.pdf](http://www.cnis.fr/agenda/PDE/PDE_0152.pdf))
- Document ELFE

---

AVIS FAVORABLES

---

**Avis d'opportunité pour les enquêtes présentées**

- Panel d'élèves du premier degré recrutés en 2011 ..... 33

**Avis favorables aux demandes d'accès à des données au titre de l'article 7bis de la loi de 1951**

- Demande d'accès par l'Insee à des données détenues par la direction de l'évaluation, de la prospective et de la performance (DEPP) ..... 35



Paris, le 22 décembre 2010 n°291/D130

## AVIS D'OPPORTUNITÉ

### Panel d'élèves du premier degré recrutés en 2011

---

*Type d'opportunité* : enquête nouvelle

*Périodicité* : enquête annuelle

*Demandeur* : Ministère de l'éducation nationale - Direction de l'évaluation, de la prospective et de la performance (DEPP).

Au cours de sa réunion du 24 novembre 2010, la commission Services publics et services aux publics a examiné le projet de panel d'élèves du premier degré recrutés en 2011.

Depuis le début des années soixante-dix, la Depp analyse les trajectoires des élèves et l'évolution de leur niveau d'acquis en suivant des cohortes dans le temps. Le panel d'élèves recrutés en 2011 se situe ainsi dans la droite ligne du panel 1978 (22 000 écoliers suivis pendant huit ans) et de celui de 1997 (9 000 écoliers suivis depuis leur entrée au cours préparatoire). Le nouveau panel d'écoliers sera la seule source permettant de mesurer l'égalité des chances et d'identifier les facteurs familiaux d'échec ou de réussite dans l'enseignement du premier degré.

Le nouveau panel, comme les précédents, a pour objectif de décrire et d'expliquer les carrières et performances scolaires des élèves depuis l'entrée à l'école élémentaire jusqu'à l'entrée en sixième. Il permet de suivre de manière précise et détaillée les cheminements des élèves de l'échantillon au sein de l'école élémentaire, de mesurer l'équité de l'enseignement élémentaire, d'évaluer l'effet de mesures de politiques éducatives mises en place dans l'enseignement primaire, de mesurer l'évolution des acquis cognitifs des élèves du cours préparatoire à la fin du CM2 en identifiant les facteurs de progression, et également de mesurer l'évolution des acquis des élèves à l'entrée au cours préparatoire entre 1997 et 2011. Le panel 2011 est également l'occasion, en association avec le service statistique ministériel de la culture, de mesurer l'évolution des pratiques culturelles des jeunes enfants au cours de la scolarité primaire. Pour atteindre ces objectifs, le panel recueille des informations sur les caractéristiques sociodémographiques de l'élève, son milieu familial (composition, situation par rapport à l'emploi, PCS, niveau de diplôme, revenu, origine immigrée ou non, conditions de logement), le degré d'implication des parents dans le suivi de la scolarité de l'enfant, l'opinion des parents sur l'école et le fonctionnement du système éducatif, le déroulement de la scolarité à l'école maternelle, le niveau d'acquis cognitifs des élèves au début et en fin de scolarité élémentaire, le parcours scolaire des élèves dans l'enseignement primaire et les dispositifs pédagogiques dont a bénéficié l'élève.

Il s'agit d'une enquête d'initiative nationale qui concerne les écoliers entrés pour la première fois au cours préparatoire dans une école publique ou privée de France métropolitaine à la rentrée scolaire 2011. Le panel est observé pendant toute la scolarité élémentaire des élèves, soit une durée de cinq à sept ans selon les cas. Les modalités de collecte varient selon le type d'informations demandées :

- la situation scolaire des élèves est actualisée chaque année, en début d'année scolaire, soit par appariement avec les bases d'élèves du premier degré, si c'est possible, soit par enquête par Internet ou par questionnaire papier ;

- les acquis des élèves sont mesurés au début et à la fin de la scolarité élémentaire par des questionnaires papier qui seront administrés, en classe par les enseignants dans les écoles retenues, selon un protocole qui garantit la confidentialité des réponses. La première prise d'information a lieu en septembre 2011 ; la seconde est prévue en avril - mai 2016, à la fin de la cinquième année d'études.

Cette prise d'information pourrait être complétée par une nouvelle mesure des acquis des écoliers ayant atteint le CM2 avec un an de retard, au début du deuxième trimestre 2017. Selon cette même périodicité, des enquêtes à destination des directeurs d'école et des enseignants pourraient être mises en œuvre pour recueillir des informations sur les différents dispositifs pédagogiques mis en place dans l'école et ceux dont l'élève a bénéficié.

- pour recueillir les informations sur le milieu familial de l'élève et la manière dont les parents vivent la scolarité de leur enfant, une enquête postale, avec relance par téléphone, est menée auprès des familles, à deux moments : la première année d'enquête, en avril – mai 2012, puis en fin de cinquième année de scolarité élémentaire (avril-mai 2016).

Les collectes annuelles sont réalisées par la Depp, tandis que les collectes sur les acquis des élèves et auprès des familles le sont par un organisme de sondage privé sélectionné dans le cadre d'un appel d'offre. Dans tous les cas, la construction des instruments de mesure et l'exploitation des résultats ont lieu exclusivement à la Depp.

Pour éviter une trop grande dispersion des élèves dans un nombre important d'écoles, l'échantillon est issu d'un sondage à deux degrés. L'échantillon de départ est constitué de 1 500 à 2 000 écoles dans lesquelles 10 000 à 15 000 élèves sont tirés aléatoirement. En absence de base de sondage exhaustive des élèves du premier degré, le tirage des élèves de l'échantillon est effectué par les directeurs d'école selon une méthode aléatoire fixée et fournie par la Depp. Les caractéristiques sociodémographiques et les informations concernant la situation scolaire à la rentrée 2011 des élèves du panel sont recueillies par un questionnaire renseigné par les directeurs d'école au moment du recrutement de l'échantillon.

Le comité opérationnel chargé de mettre au point l'architecture du dispositif réunit des membres des bureaux de la Depp, maîtres d'ouvrage et d'oeuvre du panel, et des représentants de la Direction générale de l'enseignement scolaire (DGESCO), des Inspections générales du Ministère de l'éducation nationale (Inspection générale de l'éducation nationale et Inspection générale de l'administration de l'éducation nationale et de la recherche) et du Département des études, de la prospective et des statistiques (DEPS) du ministère de la culture et de la communication. Le comité de pilotage comprend, outre les membres du Comité opérationnel, un recteur, un inspecteur d'académie, un directeur d'école, ainsi que des représentants des directions du ministère de l'Éducation nationale non représentées au comité opérationnel (Direction de l'encadrement, Direction générale des ressources humaines, Service des technologies et des systèmes d'information). Un comité de concertation comprenant des représentants des syndicats d'enseignants, de grandes fédérations de parents d'élèves et de l'Association des maires de France sera réuni avant le lancement de l'enquête.

Les données seront exploitées par la Depp. Comme lors des opérations précédentes, les résultats de l'enquête donneront lieu à des études diffusées dans les supports éditoriaux de la DEPP (*Éducation et Formations, Notes d'information, Repères et références statistiques*) ou extérieurs (publications de l'Insee, revues scientifiques dédiées aux sciences de l'éducation). A terme, le fichier des données anonymisées sera mis à disposition des chercheurs via le réseau Quételet.

Le Président, après avoir entendu la commission, émet un **avis d'opportunité favorable** à cette opération. La commission rappelle que les résultats de l'enquête doivent être largement diffusés et ses données mises à disposition du plus large public sous les conditions légales et de confidentialité en vigueur.

L'opportunité est accordée pour sept années scolaires à partir de la date d'entrée en vigueur de la conformité délivrée par le comité du label.

Paris, le 10 décembre 2010 n° 280/D030

**AVIS DU CNIS SUR UNE DEMANDE D'ACCÈS À DES DONNÉES  
AU TITRE DE L'ARTICLE 7BIS DE LA LOI DU 7 JUIN 1951 MODIFIÉE  
Données détenues par la direction de l'évaluation, de la prospective et de la  
performance (DEPP) du ministère de l'Éducation nationale**

---

*Demandeur* : Insee, Direction des études et des synthèses économiques (DESE).

*Détenteur des données* : Ministère de l'Éducation nationale, Secrétariat général, direction de l'évaluation, de la prospective et de la performance (Depp).

Au cours de sa réunion du 24 novembre 2010, la commission Services publics et services aux publics a examiné la demande d'accès à des données détenues par la direction de l'évaluation, de la prospective et de la performance (Depp).

La division Redistribution et politiques sociales souhaite développer, en collaboration avec la Depp, des études microéconomiques et des évaluations de politiques éducatives. Pour cela, elle doit connaître la structure enseignante par âge, par ancienneté, par corps et grade dans chaque collège.

Les données souhaitées sont celles des bases RELAIS.

Le premier projet mené en 2010 visait à étudier l'effet du classement en réseau ambition réussite (RAR) de certains collèges, sur la distribution des notes en mathématiques et en français au diplôme national du brevet des collèges. Le classement en RAR des établissements s'est en effet fait selon certains critères précis (par exemple la part de professions défavorisées devait être supérieure à 67 %). De ce fait des méthodes économétriques avancées (*regression discontinuity design*) ont été appliquées afin d'analyser l'effet de la politique des RAR sur la réussite scolaire des élèves, ainsi que sur une possible stigmatisation des collèges classés en RAR.

Le second projet a pour objectif d'analyser les éventuelles modifications de structures du corps enseignant suite au classement en RAR des collèges. Ce travail complètera et approfondira une première étude descriptive menée par la Depp (Note d'information n°09.09, mai 2009, *Les réseaux « ambition réussite » : état des lieux en 2006-2007*), ainsi qu'un travail joint de l'Insee et du CREST sur l'évaluation des réseaux « ambition réussite » présenté en atelier Depp le 7 octobre 2010, et en séminaire D3E-INSEE le 11 octobre 2010.

La transmission des données s'effectue chaque année à partir de la base RELAIS relative à l'année scolaire 2003/2004. Les modalités de transmission seront définies par convention entre l'Insee et la Depp.

L'étude fera l'objet d'un document de travail de l'Insee en milieu ou fin d'année 2011, qui sera présentée aux journées de la Depp au 1<sup>er</sup> semestre 2011. Elle sera éventuellement publiée dans une revue scientifique.

La commission émet un **avis favorable** à la demande d'accès aux données détenues par la Depp.