

Qu'ont en commun l'astronome de la Renaissance Nicolas Copernic, le pianiste et compositeur du ^{xix}^e siècle Frédéric Chopin et la physicienne du début du ^{xx}^e siècle, double prix Nobel de physique et chimie, Marie Curie ? Réponse : ils étaient tous les trois polonais ou d'origine polonaise.

La Pologne ne se résume pas seulement à ses grands hommes et à ses grandes femmes : ce pays est aussi, depuis 2004, l'un des membres de l'Union européenne (UE), comparable à l'Italie par sa superficie et à l'Espagne par sa population.

Le *Courrier des statistiques* se devait donc de faire connaître à ses lecteurs le système statistique du pays du *zloty* (la monnaie polonaise) et ses relations – nombreuses – avec l'Insee.

Pascal Rivière présente tout d'abord le système statistique polonais, au centre duquel se trouve son Office statistique, le GUS, équivalent de notre Institut.

Jan Robert Suesser raconte ensuite la transition statistique de la Pologne au cours des années 1990, tandis que le pays entrait progressivement dans une économie de marché. C'est dans le cadre de ses actions de coopération internationale que l'Insee a alors aidé la Pologne à réaliser cette transition.

Quelques années plus tard, au début des années 2000, la Pologne a dû, pour pouvoir adhérer à l'UE, harmoniser ses statistiques avec les exigences européennes et développer ses statistiques locales. **Victor-Pierre Morales** explique comment cette stratégie de préadhésion s'est concrétisée, sous la forme d'un jumelage mis en place entre l'Insee et le GUS.

Jean-Louis Brillet expose ensuite la problématique et le fonctionnement de modèles macroéconométriques qu'il a réalisés récemment pour le

ministère des Finances polonais, modèles destinés à la prévision et à l'analyse économiques.

Enfin, pour clore ce dossier, **Daniel Verger** et **Thibaut de Saint Pol** décrivent la coopération statistique entre la France et la Pologne en matière d'enquêtes sur la pauvreté et l'exclusion, coopération devenue aujourd'hui une véritable collaboration.

L'une des caractéristiques bien connues du système statistique public français est qu'il regroupe, autour de l'Insee, une vingtaine de services statistiques ministériels (SSM), de tailles diverses. Ces derniers travaillent bien sûr étroitement avec l'Institut, mais sont aussi de plus en plus impliqués dans la statistique européenne. Cette implication se fait *via* l'Insee, ou bien directement avec Eurostat. **Bernard Guibert** analyse ces relations et en propose une typologie.

Un de ces SSM est celui des « collectivités locales » au sein du ministère de l'Intérieur. Il y a trente ans, dans le *Courrier des statistiques*, paraissait un article signé du responsable, **Jean Bouinot**, décrivant les missions et les travaux de son service.

Cet article est à nouveau publié aujourd'hui, accompagné d'une interview de **Jean-Luc Heller**, chef actuel de ce SSM. Il explique la façon dont les statistiques produites par son service se sont transformées depuis trente ans, afin de répondre aux demandes croissantes des collectivités locales, elles-mêmes en évolution.

Il y a quelques mois, Raymond Lévy-Bruhl est décédé. **Alain Desrosières** et **Béatrice Touchelay** retracent le parcours professionnel exemplaire de ce statisticien, qui a terminé sa carrière en tant que Secrétaire général de l'Insee, poste qu'il a quitté il y a vingt ans, en 1988.

La rubrique de techniques rédactionnelles « Savoir compter, savoir conter » continue d'explorer les dimensions du *bien écrire*, cette fois sur des ques-

tions d'ordre typographique plus que stylistique. L'auteur de ces lignes rappelle ainsi quelles sont les règles à respecter pour écrire les nombres et les unités de mesure dans les publications statistiques.

Marcel Goldberg, Catherine Quantin, Alice Guéguen et **Marie Zins** présentent ensuite les problèmes statistiques qui se posent aujourd'hui aux épidémiologistes : il s'agit notamment de traiter les biais et la non-réponse dans les données issues de cohortes décrivant les phénomènes de santé de la population. Des solutions existent, qui nécessitent d'avoir recours à des bases de données nationales. Les auteurs évoquent ces solutions d'un point de vue méthodologique, sans oublier celui de la confidentialité des données.

Le métier d'astronome n'a pas toujours été de tout repos : Copernic, déjà cité ci-dessus, n'a pu publier sa théorie, l'héliocentrisme, de son vivant, car elle remettait en cause de manière trop radicale la vision du monde de l'époque. Aujourd'hui, ce métier peut fort heureusement s'exercer en toute sérénité. Il s'agit par exemple

– tâche astronomique, c'est le cas de le dire – de recenser les objets, visibles ou non, présents dans l'univers ; de les identifier, souvent de manière indirecte et au moyen de tests statistiques ; enfin, de les répertorier dans de gigantesques bases de données.

C'est donc un véritable métier de nomenclaturiste et de statisticien qu'exercent certains astronomes ; deux d'entre eux, **Françoise Genova** et **François Mignard**, ont accepté de répondre à une interview par le *Courrier des statistiques* pour présenter leur travail, permettant aux lecteurs de la revue, l'espace d'un instant, d'avoir la tête dans les étoiles...

Ce numéro se termine par les résumés en anglais des articles publiés ; quant aux annonces de séminaires et colloques, elles figurent désormais sur le portail de la statistique publique (www.statistique-publique.fr), dans la rubrique « événements ». Les internautes auront ainsi accès à une information mise à jour plus régulièrement. ■

 Serge Darriné

Le système statistique polonais

 Pascal Rivière*

L'avènement d'une société démocratique et le passage à une économie de marché sont les défis que la Pologne a dû relever au cours des années 1990 et au début des années 2000. L'adhésion de la Pologne à l'Union européenne le 1^{er} mai 2004 illustre les progrès réalisés par ce pays au cours de cette période. À l'échelle de la production de statistiques, les défis étaient également lourds pour les statisticiens officiels. La loi statistique de 1995 établit les nouvelles règles du système statistique public polonais. Elle confirme le rôle central du GUS, son Office statistique, et crée le Conseil statistique. Celui-ci propose au Gouvernement un programme d'enquêtes statistiques et fait des recommandations sur les travaux à mener à un horizon pouvant aller jusqu'à dix ans. La « revue par les pairs » menée en avril 2007 confirme que les efforts soutenus du GUS, notamment pendant les années 1990, lui ont permis de se mettre en ligne avec les principes du code de bonnes pratiques de la statistique européenne.

Après des décennies d'un régime politique imposé, la Pologne a retrouvé, au début des années 1990, le chemin de la démocratie. Le système statistique national a pleinement participé au processus de reconversion de la société polonaise et à son prolongement, l'adhésion à l'Union européenne.

La Pologne, un grand partenaire européen

L'importance politique de la Pologne

Comme pour neuf autres pays d'Europe centrale et orientale (PECO), l'entrée de la Pologne dans l'Union européenne a été rapide : quinze années seulement séparent la chute du mur de Berlin (1989) de l'adhésion de la Pologne à l'Union européenne (2004). Pour les gouvernements et les habitants de tous ces pays, ce rapprochement était vivement souhaité du point de vue politique, bien sûr, mais aussi économique ; sur ce dernier point, il a été matérialisé par l'abandon des systèmes d'échanges qui prévalaient du temps du Conseil d'aide économique mutuelle¹ et leur remplacement par les mécanismes de l'Union européenne.

Pour la France, l'ancrage à l'ouest d'un partenariat aussi important (cf. encadré 1 page suivante) représentait également un enjeu politique majeur. Dès 1991, le processus dit du Triangle de Weimar a été lancé. Ce dialogue tri-latéral franco-germano-polonais, des-



Vue de Varsovie en 1919

tiné à l'origine à faciliter les relations germano-polonaises, est devenu par la suite l'un des instruments de l'intégration euro-atlantique de la Pologne. En 2004, le principe d'un Sommet annuel entre la France et la Pologne a été arrêté. En outre, la tenue de rencontres gouvernementales annuelles, restreintes et opérationnelles, dont le principe a été retenu par les deux Premiers ministres d'alors, offre désormais un cadre de dialogue régulier au plus haut niveau.

Des liens privilégiés entre l'Insee et son homologue polonais, le GUS

La coopération avec le GUS (Office central de statistique) polonais a com-

mencé en 1991 avec la signature d'un accord entre le GUS et l'Insee. Au cours des années 1990, la Pologne a été l'un des principaux partenaires de l'Insee, avec 30 à 35 « actions » par an (visites, missions, etc.), indicateur significatif à la fois de l'importance et de la continuité de cette coopération. Parmi les actions engagées, on mentionnera les nomenclatures, les

* Pascal Rivière est chef du pôle Méthodes comparées, au sein de l'unité Europe et relations multilatérales de la direction de la coordination statistique et des relations internationales de l'Insee.

1. Comecon en anglais. Dans le contexte d'un refus du Plan Marshall américain, il est fondé en 1949 sur l'initiative de l'URSS. Il se donne pour objectif l'intégration économique de l'ensemble de ses adhérents par le biais d'une « division socialiste du travail ».

enquêtes structurelles d'entreprises, les comptes nationaux, les indicateurs de pauvreté et les prévisions.

La coopération sur les statistiques régionales a marqué une étape importante dans les relations entre les deux instituts. Elle a tout d'abord été l'occasion d'établir un lien sur plusieurs années entre les instances politiques et statistiques de la région de Katowice⁴ et de la région Nord – Pas-de-Calais. Elle a ensuite donné lieu à un contrat de jumelage⁵. Cette nouvelle forme de coopération technique pour l'Insee a été relativement lourde. Étendue de mi-2001 à fin 2003, elle s'est traduite par différentes formes d'actions : des missions d'experts français, mais aussi allemands et suédois, dans les établissements du GUS ; des visites de cadres du GUS en France, notamment dans les directions régionales de l'Insee ; des sessions de formation en Pologne assurées par des organismes polonais.

Par la suite, le programme Phare⁶ a continué, avec notamment une formation dispensée sur la comitologie (et plus généralement les processus de décision en matière statistique au niveau européen).

Vers la fin 2004, soit peu après l'adhésion de la Pologne à l'Union européenne, les aides ont basculé sur les

2. Après la France (547 000 km²), l'Espagne (505 000), la Suède (449 000), l'Allemagne (357 000) et la Finlande (337 000).

3. Après l'Allemagne (82 millions d'habitants), la France (64), le Royaume-Uni (60), l'Italie (58) et l'Espagne (40).

4. L'office statistique régional de Katowice est responsable des statistiques régionales au niveau national, pour l'ensemble de la Pologne.

5. Créés en 1998 par la Commission européenne, les jumelages consistent en un contrat passé entre les administrations des pays candidats à l'adhésion à l'Union européenne et celles d'États membres. Ces contrats, financés par les « programmes Phare » nationaux (définis plus précisément en note de bas de page n° 7), doivent permettre la mise en place d'une organisation efficace afin que les pays candidats puissent remplir leurs obligations vis à vis de l'Union européenne. Pour en savoir plus sur ces jumelages, voir l'article de Victor-Pierre Morales dans ce même dossier.

6. Aide attribuée par l'Union européenne au pays candidat, ici la Pologne, pour financer des actions de coopération où le pays aidant est *leader* (en ce sens que la Pologne demande une assistance sur un sujet spécifique et c'est le pays aidant qui en détermine la forme).

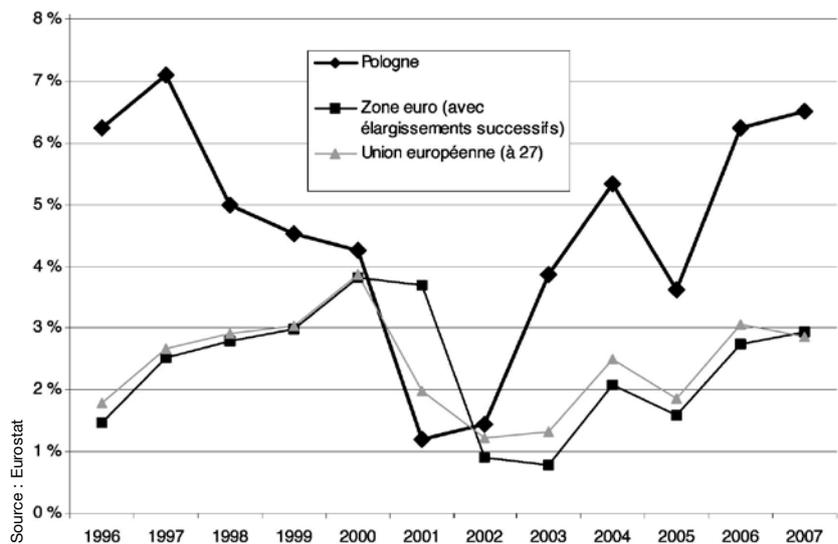
Encadré 1 – Quelques éléments géographiques et économiques sur la Pologne

S'étendant sur 313 000 km², la Pologne est le sixième pays de l'Union européenne par sa superficie², devant l'Italie et le Royaume-Uni. Sa situation entre l'Allemagne, la Scandinavie, les pays baltes, le sud-est de l'Europe et les pays de la Communauté des États indépendants (CEI) en fait un carrefour obligé pour les échanges économiques et politiques entre ces zones.

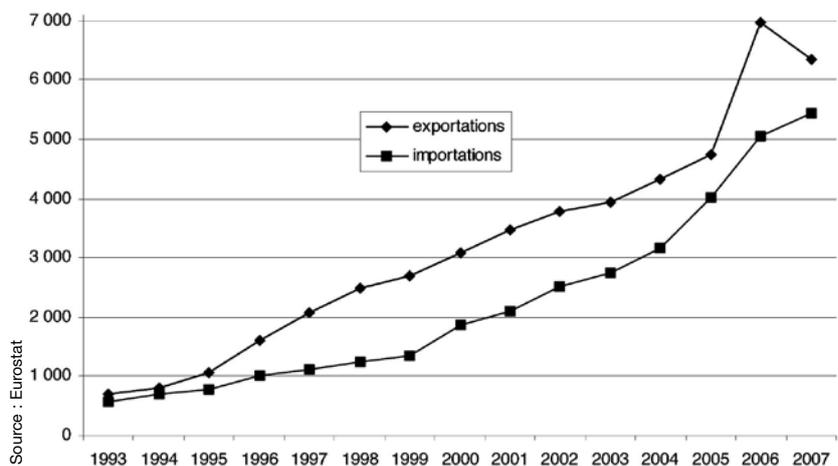
Également sixième pays de l'Union européenne par sa population³, la Pologne, avec 38 millions d'habitants en 2006, compte pour 8 % de la population de l'Union européenne. Pesant 2,5 % du PIB de l'Union en 2007, ce pays se classe en neuvième position suivant ce critère. Le retard économique relatif de ce grand pays signale le rattrapage potentiel dont il est capable. La croissance économique polonaise est ainsi plus vive que dans la moyenne du reste de l'Union européenne (cf. graphique 1).

La France profite de ce dynamisme : nos échanges commerciaux ont ainsi été multipliés par dix en quinze ans (cf. graphique 2). Les données de 2006 placent la France au 5^e rang des fournisseurs de la Pologne et au 3^e rang de ses clients. La Pologne est son 13^e débouché dans le monde. En 2006, la France demeurait le 1^{er} investisseur étranger en Pologne, devant les États-Unis, l'Allemagne et les Pays-Bas.

Graphique 1 : croissance du PIB en volume (évolution annuelle en %)



Graphique 2 : commerce extérieur de la France avec la Pologne (en millions d'euros)



*transition facilities*⁷ ; des séminaires (pour les pays de l'Europe centrale et orientale) ont été organisés (statistiques d'entreprises, indicateurs de la pauvreté, nomenclatures, groupes d'entreprises, emploi et chômage), ainsi que des formations spécifiques (méthodologie d'enquêtes, comptes nationaux, corrections des variations saisonnières...). Désormais, les relations entre l'Insee et le système statistique polonais se font sur la base d'échanges entre experts, financés le cas échéant par les fonds propres des organismes concernés ; on mentionnera ici les travaux sur l'indice de la production industrielle, les statistiques de transports et les travaux sur l'enquête exclusion-pauvreté, dont l'enquête pilote a lieu cette année⁸.

Un système statistique public organisé autour du GUS et du Conseil statistique

Le GUS fête ses 90 ans d'existence en 2008

Le Conseil de régence du Royaume de Pologne a créé le GUS le 13 juillet 1918. C'était la première unité d'administration centrale de l'État polonais indépendant. Tout au long de son existence, le GUS a conservé son nom, son programme et les principes qui le guident en vue de la fourniture d'information à la société et à l'État.

Au cours des quatre-vingt-dix années écoulées, le GUS a conduit des recherches et des études, fait des collectes et publié des résultats. Il est resté en activité entre les deux guerres, durant la Seconde Guerre mondiale et bien sûr dans la période postérieure jusqu'à nos jours.

Le 90^e anniversaire du GUS est marqué en 2008 par toute une série de séminaires consacrés à différents domaines de la statistique. Après les statistiques sociales (en janvier dernier) et les « statistiques publiques comme fondement de l'administration locale dans un État démocratique » (en avril), une conférence (en septembre) a été consacrée aux « statis-

tiques publiques dans le processus d'intégration européenne ». La série s'achèvera en novembre 2008 par un séminaire intitulé « bonnes statistiques pour un bon gouvernement ».

Le GUS coordonne fortement les activités statistiques

La loi statistique nationale du 29 juin 1995 confère un rôle primordial au GUS : celui-ci est au cœur du système statistique polonais. Au-delà de son rôle naturel de direction du GUS et de ses seize offices régionaux⁹, le président du GUS se voit en effet conférer le droit de proposer au Premier ministre de constituer, et de dissoudre, des services ou agences de la statistique officielle.

En vertu de l'article 20 de la loi statistique nationale, des services dans les ministères ont reçu la délégation de produire des informations statistiques officielles sectorielles : finances, agriculture et développement rural, environnement, économie, travail et politique sociale, éducation nationale et sport, santé, affaires internationales et administration, justice. En l'occurrence, ces services effectuent environ 20 % des enquêtes prévues dans le programme de travail.

Le Conseil statistique soumet pour adoption le projet officiel de programme d'enquêtes

Comme le note le rapport de la « revue par les pairs »¹⁰, menée au GUS en avril 2007, le rôle de coordination de l'Office central polonais de statistique sur les autres services producteurs est réel. Le processus de coordination s'appuie sur différents canaux :

- de nombreux contacts informels permettent aux experts de coordonner leurs actions ;
- la Commission de programme et de méthodologie permet d'échanger sur le programme d'enquêtes statistiques ;
- le Conseil statistique scientifique se réunit trois fois par an et statue sur les méthodologies statistiques employées.

Ces différents mécanismes doivent notamment permettre au président du GUS d'élaborer un projet de programme d'enquêtes statistiques concernant l'ensemble des services.

Ce programme est transmis au Conseil statistique, lequel est seul habilité à le présenter au Conseil des ministres pour adoption. Les membres du Conseil statistique sont nommés pour cinq ans par le Premier ministre. Ils sont issus des administrations publiques, des administrations locales, de la Banque centrale et des fédérations professionnelles et syndicales. Le président du GUS participe aux travaux menés par le Conseil. Outre son rôle de proposition au Gouvernement du programme d'enquêtes statistiques, le Conseil fait des recommandations sur les travaux à mener à un horizon pouvant aller jusqu'à dix ans, il évalue la mise en place effective du programme statistique et étudie les questions stratégiques concernant les statistiques officielles.

Parmi les autres instances gravitant autour des statistiques officielles, il convient de noter l'organe de gouvernance du GUS (« CSO Governing Body »). Composé des hauts cadres du GUS, celui-ci se réunit au moins deux fois par an en vue d'approuver le projet d'enquêtes statistiques et

7. Aide attribuée par l'Union européenne, ici à la Pologne, pour financer des actions de coopération où le pays bénéficiaire est *leader* (en ce sens que la Pologne demande une assistance sur un sujet spécifique et lance un appel d'offres dont elle détermine les termes pour ce faire).

8. Voir à ce sujet l'article de Daniel Verger dans ce même dossier.

9. De même que pour l'Insee et certaines de ses directions régionales, le GUS délègue des travaux nationaux à certains de ses offices de « Voivod » : c'est par exemple le cas des méthodologies d'enquêtes sur l'environnement, des statistiques d'investissements et de stocks de capital, des statistiques maritimes (transports, tourisme...), des enquêtes sur le secteur informel/caché, ou encore des statistiques de santé.

10. Adopté en juin 2005 par le Conseil Ecofin, le code de bonnes pratiques de la statistique européenne institue quinze principes devant régir le cadre institutionnel et les conditions de la production statistique par les acteurs du système statistique européen. Menées de fin 2006 à fin 2007, les « revues par les pairs » (trois experts européens, différents pour chaque audit) ont pour but d'auditer les instituts nationaux de statistique (INS) européens pour analyser leur degré de conformité avec les exigences du code. Pour l'instant, elles sont centrées sur l'environnement institutionnel de l'INS et la qualité de la diffusion (principes 1 à 6 et principe 15 du code).

de budget. Par ailleurs, il existe un organisme relativement original dans le paysage statistique européen : le Conseil gouvernemental de la population¹¹ chargé d'établir un bilan annuel de la population en Pologne et de faire des propositions de loi dans le domaine de la « politique de la population ». Il est aussi plus classiquement chargé d'initier des travaux de recherche en matière de démographie.

Le SSP polonais est bien à la hauteur de l'acquis communautaire

Le GUS a pris sa part pour relever le défi menant la société polonaise jusqu'à son adhésion à l'Union européenne en 2004

La chute du mur de Berlin a été, pour la Pologne comme pour les autres pays d'Europe centrale et orientale, le début d'une période de transition, à la fois politique, vers une démocratie parlementaire, et économique, vers une économie de marché. L'effort considérable consenti entre 1989 et la seconde moitié des années 1990 a concerné tous les acteurs de la société polonaise. Cet effort a également été fourni par le GUS, lequel s'est attaché à ce que les statistiques officielles soient objectives, indépendantes et conformes aux principes déontologiques, notamment celui de la confidentialité.

La période préparatoire à l'adhésion à l'Union européenne (effective le 1^{er} mai 2004) a été éprouvante pour le GUS. M. Toczynski, son président jusqu'en octobre 2006, reconnaît que la fin des

11. Établi en 1997 pour prendre la suite de la Commission gouvernementale sur la population, ses membres sont désignés en Conseil des ministres. Ce sont des universitaires et des représentants des administrations polonaises.

12. Jean-Étienne Chapron était l'un de ces « pairs » (cf. encadré 2).

13. Cette « revue » était dirigée par le statisticien norvégien Olaf Ljones. Elle comprenait également Peter Bekx d'Eurostat. Jean-Étienne Chapron a dirigé quant à lui les « revues par les pairs » de l'ISTAT italien (octobre 2006) et de l'ONS britannique (septembre 2007). Par ailleurs, dans le cadre du programme Phare, Jean-Étienne Chapron, alors chef du département des normes statistiques et comptables puis du département des comptes nationaux de l'Insee, avait animé le partenariat avec le GUS sur les comptes nationaux de 1991 à 1996.

Encadré 2 – Entretien de l'auteur avec Jean-Étienne Chapron

Jean-Étienne Chapron, inspecteur général de l'Insee, a participé à la « revue par les pairs » du GUS polonais en avril 2007¹³.

Comment pourrait-on caractériser le GUS en une seule phrase ?

« De la transition à l'adhésion ». Cela traduit la période de la transition politique et économique des années 1990 et les préparatifs de l'adhésion à l'Union européenne, effective en mai 2004. Le personnel du GUS a alors fait montre d'une motivation que ses cadres d'aujourd'hui continuent à afficher.

Comment pourrait-on caractériser le système statistique polonais en une seule phrase ?

« Un système assez centralisé », afin de rendre compte du rôle prédominant du GUS, qui produit et publie environ 80 % des statistiques publiques. Il faut compléter ce constat en mentionnant la grande importance du Conseil statistique, notamment dans l'élaboration du programme de travail national. Il convient également de souligner les différents et efficaces canaux de concertation, aussi bien entre les producteurs de statistiques qu'entre les producteurs et les utilisateurs. On mentionnera enfin le groupe de travail sur les statistiques du compte des administrations publiques ; créé en 2004, il réunit des représentants du GUS, du ministère des Finances et de la Banque nationale de Pologne ; il établit un rapport annuel sur l'établissement des comptes des administrations publiques, tant annuels que trimestriels.

Quelques mots sur le personnel du GUS ?

Dans la veine de ce que j'ai dit auparavant, il s'agit d'un personnel motivé, soucieux de bien faire et généralement bien formé. Certains des jeunes cadres ont fait des études à l'étranger. Cette solide formation est un atout, elle est également un risque : les salaires dans la fonction publique polonaise, notamment au GUS, ne sont objectivement pas attractifs pour de tels jeunes diplômés qui sont tentés de partir vers d'autres emplois.

Quel bilan la « revue par les pairs » du GUS a-t-elle établi ?

Il convient ici de rappeler que le champ de cet « audit » est limité. Il est centré sur l'environnement institutionnel du GUS : indépendance professionnelle, mandat légal de collecte des données, ressources, suivi de la qualité, confidentialité, etc. Sur ce champ, le bilan est globalement encourageant et montre que le GUS a su se mettre au niveau des standards européens. Deux aspects mériteraient toutefois d'être améliorés : le suivi interne de la qualité selon les standards européens et la mise à disposition des métadonnées (surtout en anglais).

années 1990 a été rude [1] : « Durant les cinq premières années de coopération, il ne nous a pas été possible de faire autre chose qu'écouter ».

Les premières difficultés portaient sur des points pouvant paraître triviaux, comme le vocabulaire statistique (s'ajoutant au problème classique de traduction), sensiblement différent entre les pays de l'Est et de l'Ouest. Les méthodes ont dû être radicalement revues. Ainsi, avec environ 80 000 entreprises, l'économie planifiée permettait au GUS de mener des enquêtes exhaustives ; les plus de 3 millions d'unités statistiques de l'économie polonaise d'aujourd'hui obligent *de facto* à recourir à des sondages.

Néanmoins, les deux premiers domaines statistiques réformés au sein du GUS ont été : 1) les prix, pour mesurer au mieux la forte inflation qui a caractérisé les débuts de la phase d'adhésion polonaise ; 2) les comptes nationaux, piliers incontournables de la statistique et de la conduite de la politique économique.

Le GUS est bien arrivé au niveau des standards statistiques européens

Le rapport de la « revue par les pairs »¹², menée en avril 2007, montre que le GUS est bien parvenu au niveau des standards définis par le code de bonnes pratiques de la statistique européenne. Ce rapport

souligne en particulier le souci de l'indépendance professionnelle des producteurs, le respect de l'égalité d'accès aux informations, le respect de la confidentialité, et la transparence du processus d'adoption du programme de travail. À l'inverse, il pointe des manques quant au suivi de la qualité, la publication des métadonnées (notamment en anglais) et la diffusion des informations pour répondre à des demandes spécifiques. Le rapport conclut en mettant en avant les bonnes pratiques du GUS, en particulier la bonne coordination des travaux entre administrations pour les enquêtes agricoles, ainsi que celles sur l'énergie et l'environnement. La banque de données régionales est également mentionnée, comme un exemple d'outil de suivi et d'analyse de long terme des phénomènes socio-économiques locaux. ■



Source : Wikipédia

Centre de Varsovie

Références bibliographiques

- [1] « Statistics Poland, ready for take-off », Sigma 1/2002, Eurostat.
- Courrier des pays de l'Est, *Documentation française* n° 1063, septembre 2007 : « Élargissement et flux migratoires. L'expérience de la Pologne en perspective ».
- La lettre de la coopération technique internationale n° 2, décembre 2003, Insee.
- Panorama 2000 des relations internationales et des actions de coopération, Insee.
- Peer review on the implementation of the European Statistics Code of Practice », 23 juillet 2007, disponible sur le site internet d'Eurostat.

La Pologne à la recherche des chemins de sa transition statistique (1989 – 1994)

 Jan Robert Suesser*

Cet article est dédié à Hanna Zaluska-Tucholska, l'infatigable soutien de l'Insee au sein de l'unité de coopération du GUS.

Les premières années de la transition statistique en Europe centrale (1989-1994) se font sans que la perspective de l'adhésion à l'Union européenne soit ouverte. Il s'agit alors uniquement de construire de nouveaux instruments statistiques dans les domaines économiques et sociaux, adaptés à une société de marché. La reprise d'un acquis communautaire statistique ne viendra que plus tard compléter les objectifs poursuivis. Si quelques domaines de connaissance sont vite couverts, il faut rapidement constater qu'une part essentielle du cœur du système statistique s'avère difficile à développer pour nos collègues des pays dits en transition. Persistance d'éléments du passé, émergence chaotique du nouveau système économique, profondeur des effets sociaux de la transition, turn-over des responsables statistiques, inadaptation de certains conseils prodigués... L'article revient sur tous ces éléments en indiquant comment la coopération de l'Insee y a répondu.

Lorsque le *Courrier des statistiques* m'a demandé d'écrire un article sur la période initiale de la transition statistique en Pologne, beaucoup d'anecdotes me sont revenues en mémoire sans que cela fasse une véritable analyse historique construite. Une telle analyse nécessiterait en effet un travail à partir des archives de la coopération de l'Insee et d'Eurostat¹ et, si elles existent, des pays concernés.

Mobilisant donc ma mémoire et quelques documents écrits à l'époque, le présent article se borne à présenter les enjeux de la transition statistique telle que nous la comprenions à l'époque initiale de la coopération (1989-1994), période qui précède l'ouverture de la préparation de la Pologne à son adhésion à l'Union européenne ; il évoque aussi la réflexion qu'a eue l'Insee sur la façon de coopérer avec nos collègues polonais et ceux des autres pays

d'Europe centrale et orientale². Enfin, les succès de ces premières années de coopération suivant la chute du mur de Berlin (et leurs limites) sont analysés.

L'Insee a été le premier institut de statistique européen à s'engager avec Eurostat dans une coopération avec la Pologne. Début 1989, dans le contexte du processus politique de la table ronde visant à associer l'opposition pourtant illégale à la vie politique du pays, le président du GUS³ polonais s'était adressé au directeur d'Eurostat pour lui demander une assistance « pour un rapprochement des standards et méthodes utilisés par le GUS et ceux utilisés en Europe de l'Ouest ».

En septembre de la même année, juste après la session bi-annuelle de l'Institut International de Statistique tenue à Paris, M. Kubiczek s'adressait aussi au directeur général de l'Insee, Jean-Claude Milleron, pour lui demander « de bénéficier de l'expertise impressionnante » de l'Institut.

En réponse à ces sollicitations, une mission Eurostat et Insee se rendait à Varsovie en octobre 1989⁴.

Dès mars 1990, l'Insee reçoit à Paris le président de la statistique polonaise pour discuter d'une coopération bila-

* Jan Robert Suesser, administrateur de l'Insee, a travaillé au sein de l'Unité de la coopération technique internationale de l'Insee entre 1990 et 1996 où il était responsable de la coopération avec l'Europe Centrale et Orientale et l'ex-URSS. Aujourd'hui, il est directeur de l'Adetef, l'agence de coopération technique internationale des ministères français chargés de l'économie et du budget.

1. L'histoire de la transition statistique qu'ont réalisée les pays passant d'une économie administrée à une économie de marché présente un réel intérêt tant elle est un reflet fidèle des difficultés de ce processus de transformation économique et sociale qui se poursuit depuis près de vingt ans. Elle mérite sûrement un travail historique.

2. Dans les années 1990-1996, la coopération de l'Insee se concentre alors sur huit pays : Albanie, Moldavie, Pologne, Roumanie, Russie, Tchécoslovaquie (puis Slovaquie et République tchèque), Ukraine.

3. GUS – Główny Urząd Statystyczny – Office Central de Statistique.

4. Cette ouverture précoce de la coopération de l'Insee à l'Europe centrale et orientale correspond au volontarisme de Jean-Claude Milleron, directeur général, Jean-Louis Bodin, chef du département des relations internationales et de la coopération, et Alain Azouvi, chef de la division des relations bilatérales et des programmes de coopération, qui en comprennent immédiatement les enjeux pour l'avenir.



Hanna Zaluska-Tucholska dans son bureau du GUS avec l'affiche du recensement français

térale articulée avec la coopération européenne. En septembre 1990, est créé un poste de responsable de la coopération avec l'Europe centrale et orientale et l'URSS, auquel je suis affecté, et une normalienne tout juste agrégée en sciences sociales, Cécile Lefèvre, rejoint la division de la coopération pour un an afin d'analyser l'existant statistique en Pologne et en Tchécoslovaquie⁵. Mieux comprendre pour mieux assister est l'obsession de l'équipe de coopération de l'Insee.

Avec l'assistance au GUS, l'Insee introduit deux aspects novateurs dans sa pratique de coopération internationale : une coopération avec un pays à l'économie développée et une coopération menée dans un cadre multilatéral. Coopérer avec les pays en transition puis les pays dits « émergents » signifie que les nouveaux outils statistiques mis en place doivent s'inscrire dans un système administratif complexe, et donc, peuvent moins que jamais être de simples décalques des outils existants dans le pays fournisseur d'assistance. Quant à l'introduction d'une coopération multilatérale, elle reflète le fait que l'Union européenne s'impose alors comme une référence jugée plus riche que la seule relation bilatérale. C'est donc dans cette période que ces deux éléments deviennent une réalité maintenant naturelle.

Pour en revenir aux débuts de la coopération avec le GUS, alors que les responsables de la statistique polonaise conçoivent l'assistance demandée comme une aide pour un simple aménagement technique (nomenclatures, nouveaux agrégats), les responsables de la coopération de l'Insee, de leur côté, considèrent que c'est une assistance pour une transformation en profondeur de l'ensemble du système statistique polonais.

En mars 1991, l'Insee et le GUS signent un accord de coopération de durée indéterminée, incluant huit

thèmes de travail (quatre sur financement européen et quatre sur financement franco-polonais). De premières missions permettent de défricher les objectifs des coopérations à engager (comptabilité nationale, statistique d'entreprises, formation, système de base de données, répertoire d'entreprises). Couvrant la période initiale, trois rapports de programmation de la coopération entre le GUS et l'Insee vont analyser en temps réel les réussites, les difficultés et les potentialités de cette collaboration (décembre 1991, juillet 1992 et novembre 1992).

Dès décembre 1991, on peut constater que de nombreux outils statistiques ont été introduits ou testés : enquête Emploi pilote, enquêtes de conjoncture pour certains secteurs, déclaration administrative unique (DAU) pour le commerce extérieur, indice des prix à la consommation. La NACE rev1 a été traduite et des agrégats ont été publiés selon le Système européen des comptes.

En fait, la statistique polonaise, comme l'ensemble des autres offices d'Europe centrale, avance assez rapidement là où il s'agit de construire des instruments nouveaux qui n'avaient pas de correspondance dans des instruments de la période passée et là où la production d'agrégats nouveaux peut se faire en utilisant des sources existantes, même partielles ou inadaptées au prix de conventions pas toujours très étayées.

Dans les autres domaines comme la statistique structurelle d'entreprises ou les enquêtes de conditions de vie des ménages, les progrès se font attendre. Pour la statistique d'entreprises, cela traduit sans doute le manque de base de sondages (les entreprises se créent en Pologne par centaines de milliers en ces premiers temps de la transition) et le manque d'expérience des enquêtes par sondages (voire une réticence de responsables polonais vis-à-vis de la statistique économique).

De plus, certains pans de la réalité économique changent moins rapidement que les discours et les entreprises issues de l'économie administrée

gardent une place majeure dans la production. Les autorités continuent à solliciter les statisticiens pour qu'ils fournissent l'information de gestion extensive concernant les entreprises non privatisées, à l'identique de la période précédente. Quant au secteur privé en extension massive, son caractère largement informel n'aide pas à sa mesure statistique. Il n'est objectivement pas simple de construire une information statistique au sens que cela a dans une économie de marché.

Pour les enquêtes sur les conditions de vie, les statisticiens ont une évidente réticence à prendre la responsabilité de décrire les effets sociaux négatifs de la transition, d'autant qu'ils ne peuvent appuyer cela sur une méthodologie d'enquête validée internationalement.

Le rapport de décembre 1991 souligne la difficulté pour le GUS d'avoir une stratégie de transition statistique dans un contexte administratif et économique instable. Les questions clefs sont notamment les suivantes : sur quelles sources administratives s'appuyer alors que tout bouge rapidement ? Faut-il mesurer la transition ou construire un système pour une période future plus stabilisée ? Comment répondre au *turnover* du personnel qui trouve de nombreuses opportunités dans un marché du travail en mutation rapide et alors que les moyens humains sont globalement insuffisants ? Comment construire une crédibilité pour l'institution statistique, dont l'image dans le public est liée à son rôle dans l'ancien système et qui continue à collecter une information aux fins de gestion qu'elle doit transmettre aux autorités ?

Autre problème délicat : l'abondance d'offres de coopération a aussi de forts effets pervers. La multiplication des visites d'études dans les instituts statistiques européens ou nord-américains, si elle donne une information variée aux statisticiens polonais, n'aide pas à construire une stratégie cohérente et suivie dans le temps. Chaque responsable de département ou de division du GUS privilégie les contacts internationaux

5. Il en résultera un ouvrage de la collection Études de l'Insee, publié en français et en anglais en mai 1992 : « Pologne – Tchécoslovaquie : transition économique et sociale, statistique en mouvement ».

qu'il a noués, sans qu'une logique d'ensemble ne soit donnée aux propositions de changement qui résultent de ces échanges. Enfin, il se trouve que le découpage des thèmes de la coopération coordonnée par Eurostat, qui n'a résulté d'aucune approche réfléchie, ne favorise pas partout une articulation logique des outils à construire.

Dans ce contexte, plusieurs responsables d'assistance technique thématique de l'Insee s'interrogent sur l'utilité de leur investissement. L'Insee propose au GUS de resserrer la coopération autour d'un bloc de projets prioritaires structurant pour un système statistique cible que le GUS devrait plus précisément définir à partir des connaissances accumulées⁶.

Après l'introduction des nouveaux instruments statistiques indiqués ci-dessus, quatre éléments prioritaires sont proposés : le répertoire d'entreprises, les statistiques d'entreprises de base (surtout les enquêtes structurelles annuelles), la comptabilité nationale et un système de bases de données. La nécessité d'arbitrer entre moyens et besoins est soulignée car la persistance d'une demande d'informations fines par les autorités nationales et locales constitue alors d'évidence un obstacle à la construction d'un système statistique analogue à ce qui peut exister dans une économie de marché⁷.

En juillet 1992, le GUS exprime à l'Insee son souhait qu'il soit dorénavant « *project manager* » de la coopération pour les comptes nationaux, les statistiques d'entreprises, le répertoire d'entreprises et la conjoncture, marquant ainsi sa volonté que l'Institut coordonne la coopération pour les travaux dans ces domaines. Il en informera Eurostat pour que cela soit pris en compte dans l'organisation de l'assistance européenne. Le GUS annonce également la modification envisagée de son organigramme, avec la mise en place d'une unité fonctionnelle chargée du nouveau système statistique d'entreprises articulé sur les besoins de la comptabilité nationale et l'analyse conjoncturelle. Dépasser son organisation historique



Source : JRS

Discussion de bilan entre les experts français et polonais. De gauche à droite : M. Suesser, responsable de la coopération avec la Pologne à l'Insee ; M. Olenski, président du GUS ; M. Bruneau, expert nomenclatures Insee ; Mme Zaluska, responsable de la coopération avec l'Insee au GUS ; M. Camus, expert statistique d'entreprises Insee ; Mme Zagodzinska, chef du département entreprises et répertoire au GUS ; M. Sawinski, vice-président en charge des statistiques économiques et de la comptabilité nationale au GUS.

« branche/secteur » est en fait une rupture significative avec son passé statistique de fournisseur d'une information de gestion administrative et de contrôle sur les entreprises.

Quant au répertoire d'entreprises, le GUS constate que son projet d'imiter le « modèle » français construit avec Sirene est mis à mal avec le développement par l'administration fiscale de son propre outil. Alors que le GUS dispose d'un répertoire dont la qualité ne correspond ni à celle d'un répertoire statistique, ni à celle d'un répertoire inter-administratif, il souhaite dorénavant envisager, avec l'appui de l'Insee, toute une série de variantes pour avancer dans la construction d'un outil permettant un usage statistique (démographie des entreprises, tirage d'échantillons...).

Quelques mois plus tard, en novembre 1992, le GUS n'a toujours pas avancé sur une conception du répertoire d'entreprises et la priorité donnée à la mise en place de l'enquête structurelle ne s'est pas encore traduite par un programme de travail permettant de construire un programme de coopération. Les demandes de coopération restent un patchwork sans priorités ni vision d'ensemble. Les affectations de personnels ne correspondent pas aux changements sta-

tistiques que les demandes de coopérations devraient appuyer, rendant fragiles les ambitions affichées. Les moyens financiers permettant de tester les nouvelles enquêtes annoncées ne sont pas toujours disponibles.

Dans cette période, ces décalages entre les souhaits de changements et les moyens qui le permettent se retrouvent à des degrés divers dans tous les pays d'Europe centrale. Ils vont perdurer pendant plusieurs années encore. L'exception roumaine, où la statistique progresse plus rapidement qu'ailleurs en ces premières années de la transition, interroge la coopération de l'Insee.

6. Il faut rappeler qu'à cette période la perspective de l'adhésion à l'UE n'est pas ouverte institutionnellement et que la transition statistique ne vise pas à reprendre l'acquis communautaire. Elle vise à construire une information statistique adaptée à décrire la situation économique et sociale d'une société démocratique avec une économie de marché.

7. Les collègues polonais ont des réticences à discuter avec les statisticiens français de ces demandes dont ils savent qu'elles ne correspondent pas aux standards de secret statistique pratiqués dans l'UE. Lorsqu'ils évoquent cependant ces transmissions de données individuelles d'entreprises aux autorités, ils mettent en avant le fait qu'il faut que les statisticiens soient utiles aux autorités en répondant aux demandes justifiées, comme par exemple lorsqu'il s'agit des entreprises qui sont la propriété de l'État. Il faudra encore plusieurs années pour que cette question de la non-transmission de données individuelles d'entreprises ne se pose plus.

Pourquoi des pays, ayant *a priori* davantage de moyens budgétaires que la Roumanie et affirmant plus fortement leur adhésion au changement de système économique, n'introduisent-ils pas les changements au même rythme ? L'étude de ce phénomène me semble encore à faire. Cela dit, si les experts se posent alors des questions sur l'efficacité immédiate de leur engagement, leur enthousiasme s'alimente à la conviction de contribuer à un changement historique.

À lire avec le recul les rapports faits sur la coopération en Europe centrale dans ces années initiales de l'après 1989, je suis frappé par la franchise des échanges quant aux difficultés rencontrées, difficultés objectives dues à la situation économique, sociale, politique, financière, mais aussi difficultés subjectives dues à la capacité d'imaginer les voies des changements.

Au-delà des caractéristiques personnelles des acteurs, et donc de leur façon de communiquer, cela me paraît refléter la confiance qui s'est installée entre l'Insee et les offices statistiques de la région. Car si la critique est ouverte, l'engage-

ment de l'assistance n'est jamais remis en question. Et la critique va de pair avec l'écoute des contraintes rencontrées par les statisticiens des pays d'Europe centrale, dès lors qu'elles sont comprises par les experts de l'Insee.

Car pour bien comprendre les difficultés du dialogue, on doit constater les limites de l'expertise délivrée dans ces premiers temps de la coopération en Europe centrale. Il n'était pas naturel pour les experts de l'Insee de prendre en compte la légitimité de la persistance d'une demande d'information des autorités sur les entreprises contrôlées par l'État en l'absence de sources alternatives disponibles.

Plus généralement, leur propositions pouvaient sous-estimer les conséquences du fait qu'elles n'intégraient pas suffisamment l'héritage statistique. Il faudra attendre 1993-1994 pour que les statisticiens d'Europe centrale accordent leur juste place – centrale – aux enquêtes structurelles d'entreprises (plus tôt en Roumanie), après l'élaboration d'argumentaires *ad hoc* aux réalités locales, faite par l'Insee à partir de l'expérience accumulée.

Pour passer à la seconde étape de la transition statistique, deux éléments essentiels vont intervenir : un début de stabilisation des processus productifs et sociaux, et surtout, la perspective d'adhésion à l'Union européenne (1994). Le premier élément permet de construire l'information avec un horizon temporel correspondant au rythme de développement des outils statistiques. Le second élément sera une contrainte « impérieuse » pour réussir les changements, au moins dans les domaines concernés par l'acquis communautaire.

En revenant brièvement sur ces années 1989-1994, je voudrais insister sur les aspects humains autant qu'organisationnels qu'impliquent une coopération statistique. Si la technique statistique est l'outil indispensable pour la réussite de tous transferts et adaptations des savoir-faire, elle ne suffit pas, en effet, à assurer le succès de la coopération. Depuis ces années-là, le concept de gouvernance statistique s'est construit et imposé comme central pour toute transition statistique. À relire ce qui s'écrivait à l'époque, la coopération de l'Insee avec l'Europe centrale portait sûrement ce concept déjà en germe. ■

Le jumelage entre l'Insee et son équivalent polonais, le GUS, sur les statistiques régionales

 Victor-Pierre Morales*

Dans le cadre de la stratégie de préadhésion, l'Union Européenne a financé un jumelage entre l'Insee et son équivalent polonais, le GUS, sur les statistiques régionales. Trois objectifs y étaient assignés : l'harmonisation des statistiques polonaises avec les exigences européennes, le renforcement du rôle des offices statistiques régionaux et le développement du système de statistiques régionales. Les différentes actions ont pris la forme de missions d'experts essentiellement français dans les établissements du GUS, de visites d'études de Polonais en France, notamment dans des directions régionales, ou encore de sessions de formation en Pologne.

Le concept de jumelage s'est adressé aux pays candidats à l'adhésion à l'Union européenne (UE) qui, à ce titre, avaient entamé l'adaptation de leurs institutions administratives et démocratiques pour se conformer aux critères d'adhésion. En juin 2001, la Pologne a signé avec la France une convention de jumelage relative aux statistiques régionales. Pendant 28 mois, c'est l'Insee qui a accompagné l'office statistique polonais, le GUS¹, dans l'application de « l'acquis communautaire » en matière de statistiques régionales.

La stratégie de préadhésion

Au mois de juillet 1997, la Commission européenne a publié un document intitulé « Agenda 2000 ». Elle y formulait des propositions en vue de l'élargissement, des réformes politiques et du financement futur de l'UE. Elle proposait également de rassembler les différentes formes d'assistance fournies par l'Union dans le cadre de la préadhésion au sein d'une structure commune, appelée « les partenariats pour l'adhésion ».

Ces partenariats devaient aider les pays candidats d'Europe centrale à remplir pleinement les trois critères

* Victor-Pierre Morales est statisticien de l'Insee en poste à la division des programmes de coopération, au sein de l'unité de la coopération technique internationale (direction de la coordination statistique et des relations internationales).

1. Główny Urząd Statystyczny.



d'adhésion énoncés lors du Conseil européen de Copenhague :

– l'existence d'institutions stables garantissant la démocratie, la primauté du droit, les droits de l'homme, le respect des minorités et leur protection ;

– l'existence d'une économie de marché viable, ainsi que la capacité de faire face à la pression concurrentielle et aux forces du marché à l'intérieur de l'Union ;

– la capacité d'assumer les obligations, et notamment de souscrire aux

objectifs de l'union politique, économique et monétaire.

Cette stratégie de préadhésion passait notamment par la préparation des pays candidats d'Europe centrale et orientale à leur entrée dans l'Union européenne. Ceux-ci devaient se doter d'administrations et d'institutions modernes et efficaces qui soient capables d'appliquer l'acquis communautaire de la même manière que celles des États membres de l'UE.

Un partenariat entre un État membre et un pays candidat

C'est en 1999 qu'ont réellement démarré les premiers programmes de jumelage. Ces programmes ont été lancés à la suite de la première vague de projets « PHARE-Institution Building »², développés par la Commission européenne au bénéfice des États candidats, pays d'Europe centrale et orientale (PECO).

Les grands principes des jumelages

Le jumelage est un contrat financé par l'Union européenne entre un État membre et un pays candidat. Ce contrat doit viser la mise en place d'un élément de l'acquis communautaire.

2. Le processus d'élargissement aux pays d'Europe centrale et orientale a nécessité l'élaboration d'instruments de soutien et d'aide financière à destination de ces pays. Parmi ces instruments, le projet PHARE est l'instrument principal de la stratégie de pré-adhésion. Son action se concentre sur deux objectifs essentiels liés à la reprise de l'acquis communautaire : l'« institution building » (ou renforcement des capacités administratives et institutionnelles des pays candidats) et l'aide au financement des investissements des pays (à l'exception des financements correspondants à d'autres instruments que PHARE). Sur la période 2000-2006, le budget total disponible du programme PHARE s'élève à un montant de 1,5 milliards d'euros par an [règlement du Conseil n° 3906/89].

3. Le benchmark est une mesure étalon ou un standard qui sert de point de référence pour évaluer les performances ou l'état d'un processus.

4. Opération de contrôle consistant à suivre méticuleusement le fonctionnement d'un système, d'un processus, en temps réel.

Les jumelages pour l'« Institution Building » sont fondés sur cinq principes simples :

– les pays candidats choisissent librement les États membres avec lesquels ils souhaitent coopérer ;

– tout jumelage doit aboutir à un résultat concret, opérationnel, permettant le respect de l'acquis communautaire par le pays candidat dans un domaine donné ;

– en signant une convention de jumelage, les deux jumeaux ne prennent pas un engagement de moyens, mais de résultat : à la fin du jumelage, un nouveau système doit fonctionner effectivement dans le pays candidat, sous la seule responsabilité et avec les seuls moyens du pays candidat ;

– un jumelage ne se résume pas à une assistance technique à sens unique d'un État membre vers un pays candidat ; c'est un travail commun au profit duquel les deux partenaires prennent chacun des engagements précis, le pays candidat vis-à-vis des réformes qu'il va réaliser et des moyens qu'il va leur affecter, l'État membre vis-à-vis de l'aide qu'il va y apporter pendant la durée du jumelage ;

– pour fonder la crédibilité d'une telle démarche, les deux jumeaux établissent, avant de commencer la mise en œuvre, un plan de travail détaillé, jalonné par des *benchmarks*³, qui pourra certes être adapté en cours de route, mais qui permettra par un contrôle ou *monitoring*⁴ sans faiblesse d'évaluer le progrès des travaux jusqu'au résultat final attendu.

Ces principes n'ont d'autre finalité que de favoriser l'efficacité du travail et, en conséquence, l'admission aussi rapide que possible des pays candidats dans l'Union.

Les spécificités d'un jumelage

La mise à disposition de conseillers préadhésion (CPA) des États membres auprès des administrations des pays candidats constitue une première spécificité. Ces conseillers, en fonction de contextes institutionnels particuliers et de projets de nature et

d'envergure variables, jouent un rôle à la fois de coordonnateurs de projet, mais aussi d'experts dans les domaines concernés. Les CPA, pendant tout ou partie de la durée de vie de la convention de jumelage, sont installés au sein de l'administration partenaire.

La seconde spécificité des jumelages réside dans la détermination initiale d'objectifs qualitatifs, dont la réalisation appartient en coresponsabilité aux administrations partenaires (pays candidat et État membre). Les jumelages comprennent ainsi une « obligation de résultats », évaluée *in fine* par les services de la Commission européenne.

Enfin, les jumelages se distinguent, par rapport aux projets PHARE traditionnels (coopération bilatérale sur des projets d'importance moindre), par l'importance des moyens consacrés et, dans nombre de cas, par une multilatéralité accrue des échanges, par le biais de l'association de plusieurs pays membres dans la réalisation d'un projet.

Le jumelage entre le GUS et l'Insee

Dans le cadre de la préadhésion de la Pologne à l'Union européenne, le Gouvernement polonais a souhaité bénéficier de la procédure de jumelage mise en place par la Commission européenne, financée par le programme PHARE 2000 (tranche de crédits mis à disposition en 2000 et à une nouvelle vague de jumelages) et portant sur les statistiques régionales.

Sollicité par le GUS, l'Insee, représentant la France, a accepté d'être partenaire de ce jumelage financé par l'Union européenne. La convention de jumelage, préparée en commun par le GUS et l'Insee et validée par la Commission, a été signée le 19 juin 2001.

Trois objectifs essentiels, chacun en deux phases

L'objectif de ce jumelage était triple : harmoniser les statistiques polonai-

ses avec les exigences européennes ; renforcer le rôle des offices statistiques régionaux, sur le plan du management, de l'organisation et des matériels ; enfin, développer le système d'information des statistiques régionales (système d'information géographique ou SIG, bases de données). L'ensemble s'accompagnait d'un dispositif de formation pour chacune des composantes décrites ci-dessus.

Le planning de travail s'est déroulé en deux phases, pour chacun des objectifs retenus (y compris pour l'accompagnement en termes de formations). Une première phase, désignée sous le nom de « définition du plan stratégique » (annexe 1), s'est déroulée pendant deux trimestres et avait pour but de réaliser l'évaluation de la situation et des besoins. La deuxième phase est consacrée à la « mise en place du plan stratégique » ; il s'agissait d'évaluer des résultats du jumelage, notamment à travers un séminaire-bilan, et de veiller à l'exécution du budget. Tous les objectifs définis dans la convention devaient être atteints avant le 31 octobre 2003.

Le cadre institutionnel

Pour mener à bien ce projet, l'Insee a mis à disposition un conseiller de préadhésion⁵. En poste à Varsovie de juillet 2001 à juillet 2002, son rôle a consisté notamment à organiser les différentes actions programmées, à veiller à leur bon déroulement, à préparer les réunions du comité de pilotage et à rédiger les rapports demandés par la Commission européenne. Puis la gestion du jumelage jusqu'à son terme en février 2004 a été effectuée depuis l'Insee.

Le chef du projet à l'Insee⁶ était garant, avec son homologue du GUS, le vice-président Janusz Witkowski, du bon déroulement du projet.

Un comité de pilotage s'est réuni chaque trimestre jusqu'à la fin du jumelage. Les rapports et propositions examinés dans ce cadre étaient adressés à la Commission européenne.



Source : droits réservés

Comité de jumelage : de gauche à droite : M. Tranap ; M. Witkowski, responsable polonais du jumelage ; M. Hautcœur responsable français du jumelage ; M. Morales, conseiller de pré-adhésion du jumelage ; Mme Masluk, assistante du conseiller.

Les actions de jumelage

Différents types d'actions (annexe 2) ont été réalisés : des missions d'expertise en Pologne ; des visites d'études (dénommées *study visit*) de responsables polonais en France ; des conférences internationales, séminaires et ateliers ; des préparations d'appel d'offre ; des formations ; et un séminaire final. Chacun de ces points est passé en revue ci-après.

Les missions d'expertise en Pologne

Ces missions ont été déclinées selon trois composantes.

- **Composante 1** : *harmonisation des statistiques polonaises régionales avec les exigences européennes, dans le cadre de la nomenclature NUTS (nomenclature des unités territoriales statistiques).*

Chaque domaine statistique principal de l'acquis communautaire a été couvert par un ou deux experts : emploi et chômage, agriculture, transport, tourisme, comptes régionaux, environnement, statistiques urbaines.

Au cours de la première phase de définition du plan stratégique, huit experts ont examiné et évalué la situation méthodologique et opérationnelle, par rapport à la nomenclature nationale NUTS, établie en 2000, aux conditions de l'acquis communautaire et en fonction de l'adaptabilité aux nouveaux besoins régionaux.

Le programme préparatoire spécial pour les fonds structurels en Pologne, présentant la description des statistiques régionales polonaises existantes, a servi de référence pour l'évaluation de la situation et la définition du plan stratégique.

Les rapports d'évaluation ont détaillé les imperfections ou les insuffisances possibles des statistiques existantes, les besoins en formations de base ou avancées, et proposé des corrections ou des améliorations pour la deuxième phase.

Dans cette dernière – exécution du plan stratégique –, les 13 missions d'experts ont aidé les unités respon-

5. Il s'agissait de Victor-Pierre Morales, l'auteur de cet article, et ce, pour toute la durée du jumelage.

6. C'était Jean-Claude Hautcœur, à l'époque directeur régional du Nord – Pas-de-Calais.

sables du GUS à mettre en application les adaptations nécessaires dans les statistiques régionales, ainsi qu'à en préparer l'amélioration de la diffusion.

• *Composante 2 : renforcement des Offices statistiques régionaux.*

Cette composante du projet a été plus particulièrement pilotée par le chef de projet français, aidé par ses collègues directeurs régionaux, et le conseiller de préadhésion. Elle a porté d'une part sur la gestion et la coordination des établissements régionaux polonais (WUS, Wojewodzki Urząd Statystyczny), d'autre part sur une aide relative au matériel informatique et aux communications télématiques.

Deux directeurs régionaux français, en mission d'une semaine, ont échangé avec leurs homologues polonais sur leur rôle, leurs missions, et notamment sur la coordination des missions nationales et régionales dans les bureaux régionaux polonais, la gestion des compétences nationales et régionales et la manière de répondre aux besoins locaux de statistiques.

Une mission d'une semaine d'un expert européen en réseaux informatiques a fourni un diagnostic de la situation et des besoins de l'informatique statistique et de la communication de données dans les offices statistiques régionaux.

Dans la seconde phase d'exécution – mise en place du plan stratégique régional –, quatre missions de directeurs régionaux de l'Insee dans les offices statistiques régionaux polonais ont permis d'évaluer leurs plans régionaux stratégiques, toujours dans l'optique d'une production de statistiques régionales conforme aux demandes européennes.

Sur le plan informatique, l'expert en réseaux a aidé le GUS à décrire les caractéristiques techniques pour l'appel d'offre d'acquisition de matériel, la restructuration et la modernisation des réseaux.

• *Composante 3 : évolution et développement du système d'information des statistiques régionales.*

Cette composante visait à préparer les investissements dédiés à trois outils informatiques qui seront utilisés dans des statistiques régionales, pour les besoins internes du GUS mais aussi pour ceux des utilisateurs externes : un système d'information géographique (SIG), le registre national géographique TERYT et la base de données régionales.

L'élaboration du plan stratégique s'appuie sur le principe que les investissements couvriront la conception, le développement de logiciels ainsi que l'acquisition de serveurs dédiés, et la formation du personnel correspondant, pour remplir trois objectifs :

– faire évoluer le registre territorial national existant (TERYT) en une base publique de données, y compris dans la nomenclature NUTS, et qui soit accessible par internet ;

– améliorer la banque de données existante sur les statistiques régionales, l'installer sur le serveur web du GUS, pour un accès public, en conformité avec la base de données REGIO d'Eurostat ;

– développer le système d'information géographique, avec acquisition d'un serveur opérationnel de données, de ses logiciels et de ses cartes numérisées, utilisable par les départements du GUS et les offices statistiques régionaux.

Des experts de l'institut statistique de Suède (SCB Statistiska CentralByrån) ont réalisé les missions relatives au SIG et à TERYT.

L'évolution prévue de la banque de données régionales existante et de ses bases de données d'agrégats, en conformité avec les exigences d'Eurostat, a été évaluée par un expert français.

Dans la deuxième phase, les experts suédois ont assisté le GUS dans la définition des caractéristiques techniques pour l'appel d'offres sur le nouveau SIG et celui sur l'évolution du registre territorial. De son côté, l'expert français a aidé le GUS à définir les caractéristiques techniques pour l'appel d'offres sur l'évolution de la banque de données régionales.

Les visites d'études de responsables polonais en France

Afin de connaître ou d'approfondir les thèmes développés dans le jumelage, des visites d'études de spécialistes polonais ont été organisées. Chaque délégation était composée de cinq spécialistes et un interprète pour une visite de trois à cinq jours. Ici aussi, le travail s'est décliné selon trois composantes.

Dans la phase d'exécution du plan stratégique relatif à l'harmonisation des statistiques polonaises, quinze visites d'études en France de statisticiens polonais ont été organisées afin de connaître les solutions pratiques et les expériences utilisées en France.

Sur le thème du renforcement des WUS, les échanges entre Français et directeurs régionaux polonais se sont poursuivis afin de mettre en application le plan régional de soutien à l'amélioration de la gestion des offices statistiques régionaux.

Au total, huit visites d'études dans différentes directions régionales de l'Insee ont été effectuées, essentiellement sur les thèmes suivants : organisation et gestion des offices régionaux ; coordination entre les niveaux nationaux et régionaux ; partenariats avec les utilisateurs locaux (monde politique, administrations, médias, chercheurs...).

Enfin, une visite d'études en France de cinq spécialistes polonais, consacrée à l'organisation et au fonctionnement de la banque de données régionale française (SEDDL) a montré l'importance du développement d'un tel système dans l'amélioration de la diffusion des statistiques polonaises.

Conférences internationales, séminaires et ateliers

Une bonne occasion de peaufiner la connaissance liée aux activités du jumelage a été la participation à des conférences internationales, séminaires et ateliers de travail relatifs aux statistiques et à l'informatique orientée vers les applications statistiques, orga-



Source : droits réservés

Dernier bâtiment du ghetto de Varsovie, présentant une exposition de photographies de juifs polonais y ayant vécu

nisées dans les États membres de l'UE et dans quelques pays candidats.

En ce qui concerne la composante 1, quatre statisticiens du GUS ont participé à la conférence sur le projet Eurorégion de base de données frontalières, qui a eu lieu à Kamenz (Allemagne).

En ce qui concerne la composante 2, un spécialiste du GUS a assisté à une conférence à Manchester (Royaume-Uni) sur le thème « Intranets, extranets dans le secteur public », et cinq spécialistes ont participé à une conférence sur le thème « Gouvernance et informatique » à Londres (R-U). Enfin, trois spécialistes polonais ont participé à la sixième conférence de sécurité de l'information à Bristol (R-U).

Enfin, en ce qui concerne la composante 3, deux statisticiens du GUS ont participé à un atelier sur des applications des GIS dans les statistiques à Tallin (Estonie), et deux autres à une conférence sur les GIS à Ostrava (République tchèque). Toujours dans le même domaine, deux polonais ont

suivi un cours de formation sur les métadonnées en statistique organisé par l'institut TES, à Oslo (Norvège).

Préparation d'appels d'offre

Plusieurs appels d'offre ont été préparés dans le cadre du jumelage, sur le plan de la formation comme de l'équipement informatique. Ils se sont révélés partiellement fructueux.

Formation

Les formations relatives aux statistiques régionales, en cohérence avec les exigences européennes et les besoins relatifs à la mise en place proche des fonds structurels, ont été dispensées en polonais dans les structures de formation du GUS. 514 participants ont aussi bénéficié de 436 heures de formations non individuelles.

Les cours d'informatique dispensés en polonais et dans les installations du GUS ont été délivrés à 100 personnes au cours de 350 heures.

Séminaire final

En septembre 2003, la conférence finale d'évaluation a accueilli les partenaires du projet de jumelage, mais également des représentants des collectivités locales et des statisticiens des offices statistiques de pays candidats.

Trois objectifs étaient assignés à cette conférence : présenter une évaluation finale du jumelage ; profiter des autres expériences présentées par les États membres ou les pays candidats ; présenter l'évolution de la statistique régionale aux représentants locaux et en profiter pour prendre en compte leurs besoins.

Les statistiques régionales comme mesure de l'utilisation des fonds structurels, la coopération entre les offices régionaux de statistique et les collectivités locales et le développement des statistiques régionales dans les pays représentés à cette conférence, autant de sujets qui ont été largement abordés et commentés au cours de ce séminaire qui a réuni une cinquantaine de participants. ■

Annexe 1 – Le plan stratégique

Ce plan stratégique⁷, validé par le comité de pilotage du jumelage, a été construit à partir des conclusions et des suggestions des missions d'experts réalisées au cours du quatrième trimestre 2001.

Adaptation des statistiques régionales aux exigences communautaires

Trois recommandations ont pu être retenues à l'issue des états des lieux sectoriels :

– Améliorer le système d'enquêtes de manière à produire et utiliser les résultats aux échelons géographiques requis.

– Améliorer, dans le moyen ou le long terme, l'utilisation des sources administratives. La génération actuelle des fonds structurels s'appuie aussi bien au niveau de l'éligibilité qu'au niveau du suivi et de l'évaluation sur des zonages courants (NUTS3, NUTS4). Mais il semble que la génération suivante sera définie davantage en termes de projets faisant moins référence à une logique spatiale pré-déterminée. De plus, les besoins des acteurs locaux ne pourront être totalement satisfaits par les résultats d'enquête.

– Faciliter l'alimentation de la banque de données régionales pour une meilleure utilisation des informations produites.

Renforcement des offices statistiques régionaux

Cet axe stratégique comporte deux volets : la définition de spécifica-

tions techniques pour l'équipement informatique et la mise en place d'un réseau entre les différentes unités de l'office statistique ; le renforcement, sur le plan de l'organisation et du management, des offices statistiques régionaux.

Banques de données régionales, GIS et TERYT

Pour ce qui concerne la banque de données régionales et TERYT, deux types d'actions ont été conduits : l'un par des experts suédois, pour définir les termes de référence pour un appel d'offres sur la mise en place de GIS ; l'autre par un expert français, sur l'administration et l'utilisation des banques de données régionales (SEDDL). Il convient de trouver une articulation entre ces deux champs d'expertise. La proposition consisterait à définir les produits et les utilisateurs de cette banque de données régionales de manière à ce que les termes de référence de l'appel d'offre, définis par les experts suédois, tiennent compte des spécifications nécessaires.

Formation

Au cours des missions réalisées par les experts, des besoins en formation ont été mis en évidence. Le GUS a entrepris un recensement des besoins en formation pour les personnels, besoins que les objectifs décrits dans ce plan stratégique devraient permettre de préciser. Sur la base de ces matériaux, un plan de formation a été élaboré.

Annexe 2 – Le nombre d'actions réalisées

La première action s'est déroulée en septembre 2001 et la dernière a eu lieu en novembre 2003. La plus grande partie des actions a concerné évidemment l'harmonisation des statistiques régionales polonaises avec les statistiques européennes. Le tableau ci-dessous ne prend pas en compte les actions de formation.

7. Rédigé par V-P Morales et J-C Hautcœur.

Type d'action	Missions d'expertise	Visites d'études	Ateliers, conférences, séminaires
I. Adaptation des statistiques régionales aux exigences européennes	25	16	2
I.1. Définition du plan stratégique	8		
I.2. Mise en œuvre du plan stratégique	17	16	2
II. Renforcement des offices statistiques régionaux	8	10	3
II.1. Définition du plan stratégique et d'équipement des régions	2		
II.2. Mise en œuvre du plan stratégique et d'équipement des régions	3	8	3
II.3. Suivi et évaluation du plan de formation	3	2	
III. Mise à niveau et développement du système d'information statistique polonais	7	1	3
I.1. Définition du plan stratégique	2		
I.2. Mise en œuvre du plan stratégique	5	1	3
TOTAL	40	27	8

Modèles macroéconométriques pour le ministère des Finances polonais

 Jean-Louis Brillet*

Parmi ses missions, l'Unité de la coopération technique internationale de l'Insee assiste notamment ses partenaires étrangers à la construction de modèles macroéconométriques, destinés à la prévision et à l'analyse. Cet article présente l'expérience menée au cours des dernières années avec le ministère des Finances polonais. Après une description de la problématique, il aborde les deux phases du projet. La première a consisté à diriger la production d'un modèle avec collaboration polonaise. Même si cet outil s'est révélé opérationnel, l'évolution de la situation a réclamé que l'on en construise un second. Les compétences polonaises ayant évolué, le rôle de la France a pu se limiter à une assistance dans les domaines stratégique, économique et technique. Le modèle est maintenant opérationnel, et permet d'appuyer des études de politique économique.

Au cours des dernières années, l'Insee a coopéré avec le ministère des Finances polonais à la construction de modèles macroéconométriques, destinés à éclairer les évolutions possibles de l'économie nationale, ainsi que les conséquences à attendre d'événements extérieurs ou de décisions de politique économique.

Cette coopération a pris deux formes successives, avec l'évolution des compétences des partenaires polonais. Dans un premier temps, une équipe française (dont l'auteur de cet article, Jean Louis Brillet, était le seul membre permanent) a dirigé la production d'un modèle, avec une assistance locale. Puis, lorsqu'il s'est agi d'élaborer un nouvel outil, pour des raisons qui seront expliquées ci-après, une deuxième phase a pu commencer, où les partenaires polonais sont devenus maîtres d'œuvre, la participation française prenant essentiellement la forme de conseils économiques et informatiques, sur des plans organisationnels et techniques.

Afin de rendre cette présentation accessible à des non-spécialistes, voici d'abord des éléments méthodologiques élémentaires. On abordera ensuite successivement les deux phases.

Qu'est-ce qu'un modèle macroéconométrique ?

L'objectif du projet était de construire un modèle quantifié macroéconomé-



Vue de Varsovie : Palais de la culture et de la science

trique structurel. Derrière ce terme technique se cache un objet relativement simple dans sa conception, même si sa réalisation se révèle souvent complexe.

Il s'agit ici :

- d'un modèle : une représentation simplifiée de la réalité, reposant sur des concepts explicites. Par exemple : l'entreprise investit d'autant plus que la demande est forte et que les profits attendus sont élevés ;
- quantifié : le raisonnement peut être appliqué à des données chiffrées. Par exemple, au raisonnement sur l'équi-

libre offre-demande, peut être associée la comptabilité nationale ;

- macroéconomique : on traite comme un tout indivisible les grands agents économiques : les entreprises, les ménages, les banques, l'État, le reste du monde. On peut cependant séparer les entreprises selon leur type de production. Le modèle ne le fait pas ;

* Jean-Louis Brillet travaille au sein de la direction de la coordination statistique et des relations internationales de l'Insee, en tant que chargé de mission auprès du chef de l'Unité de la coopération technique internationale.

- structurel : on décrit les concepts économiques par des variables, qui seront soit produites par le modèle (variables endogènes), soit utilisées comme hypothèse (variables exogènes). Les éléments endogènes peuvent venir de comportements des agents (par exemple : l'investissement, ou les importations qui dépendront de la compétitivité étrangère, de la demande locale et des difficultés locales à la satisfaire) ;
- économétrique : en plus du respect de la théorie économique, les formules choisies doivent reproduire assez fidèlement la réalité passée, et chaque élément explicatif doit apporter une contribution significative à l'explication.

Cet outil a essentiellement deux objectifs. D'une part, donner une image cohérente de l'évolution future de l'équilibre économique. Cette image peut représenter le futur le plus probable, être associée à des hypothèses spécifiques (une croissance mondiale déprimée) ou respecter certains objectifs (le retour de la France à l'équilibre budgétaire en 2012). D'autre part, cet outil doit présenter les évolutions de l'équilibre à la suite d'une modification des hypothèses, soit extérieures (un pétrole plus cher), soit de politique économique (une baisse de la TVA).

L'objectif de cette coopération

Les deux phases de la collaboration avaient en commun la plupart de leurs caractéristiques, les différences concernant certaines options théoriques, et surtout, l'organisation du travail et la nationalité du maître d'œuvre.

L'objectif recherché était, dans les deux cas, de produire un modèle qui soit :

- à un seul produit marchand, agrégeant donc les biens primaires (l'agriculture), secondaires (les biens manufacturés) et tertiaires (les services) ;

1. Côté français, ce projet a mis à contribution, outre Jean-Louis Brillet : Stéphane Capet, économiste au bureau de la prévision et des études économiques de la direction de la prévision, et Julien Deroyon, en poste à la division croissance et projections macroéconomiques de l'Insee.

- trimestriel. En effet, l'office statistique polonais (le GUS) produisait dès le début du projet une comptabilité nationale trimestrielle. Cette périodicité était requise par l'application de méthodes d'estimation, ainsi que par l'établissement d'un diagnostic conjoncturel ;

- privilégiant le secteur réel (biens et services non financiers). Ce choix était dicté par les données, la stabilité des comportements et la difficulté d'établir une formalisation cohérente. Mais le maximum a été fait pour introduire des éléments financiers ;

- keynésien à court terme, avec des équilibres de type néo-classique. Il s'agit de l'option la plus courante. En d'autres termes, la demande est satisfaite à court terme (dans la mesure des capacités disponibles), mais l'offre de long terme est conditionnée par la rentabilité (niveau des capacités et choix des facteurs de production), et le partage des fruits de la croissance par les pouvoirs de négociation des agents.

Côté polonais, les deux phases ont mis à contribution le même partenaire : le département pour la politique financière et l'analyse, du ministère des Finances. Malgré les bouleversements administratifs caractéristiques des pays en transition, il a été possible de conserver les mêmes partenaires tout au long du projet. Il s'agissait des personnes suivantes : M. Ludwig Kotecki, chef du département ; M. Slawomir Dudek, chef du bureau des simulations et des prévisions ; Mme Iwona Fudala-Poradzinska, Mme Monika Janiek et M. Dawid Pachucki, économistes du département en question.

Côté français, l'auteur de cet article a piloté l'équipe, mais ses partenaires ont changé au cours du temps.

Le premier projet : avril 2000 - juin 2003

Le déroulement du projet

Le projet¹ a commencé en avril 2000. Le modèle lui-même a été transmis en juillet 2002 et les travaux se sont terminés en juin 2003.

Pendant cette période, le travail a été mené simultanément à Paris (surtout) et à Varsovie. Les missions trimestrielles (trois ou quatre jours) à Varsovie ont été essentiellement consacrées à des séances de travail (données et formulations) et au suivi du programme. Les visites biennuelles à Paris ont permis aux partenaires polonais de découvrir l'Insee, la direction de la prévision et d'autres instituts, d'y rencontrer les spécialistes de la modélisation, et de présenter le modèle lors d'un séminaire à Bercy, en décembre 2001.

Puis, les phases suivantes ont été réalisées :

- identification des données disponibles, venant essentiellement du GUS mais traitées par le ministère des Finances ;
- choix du cadre comptable et des options de comportement ;
- transfert des données originales et constitution de la base de données spécifique au modèle ;
- estimation des équations de comportement ;
- contrôle de la cohérence des informations ;
- simulations sur le long terme et tests des propriétés ;
- rédaction d'un document descriptif et d'un mode d'emploi.

Le document complet, transmis en juillet 2002, comprend une description complète du modèle et de ses propriétés, ainsi qu'un mode d'emploi sous le logiciel utilisé. Il a été présenté lors d'une conférence se déroulant à Cedzyna (au sud de Varsovie), du 5 au 7 décembre 2002.

Le modèle obtenu : principales caractéristiques

Il s'agit d'un modèle macroéconométrique de petite taille aux caractéristiques assez traditionnelles :

- un seul produit marchand ;
- des comportements adoptant systématiquement une forme à correction d'erreurs, assurant une stabilité de la croissance structurelle sous-jacente



Source : droits réservés

Place du château à Varsovie

et une séparation interprétable entre dynamique de court terme et équilibres de long terme ;

- une fonction de production de type Cobb-Douglas², Putty-Clay³, prenant en compte explicitement l'influence des coûts relatifs des facteurs ;
- une boucle prix-salaires influencée par l'offre, de type WS-PS⁴ pour le taux de salaire ;
- un commerce extérieur défini par la demande, la compétitivité et les capacités disponibles ;
- un domaine financier simple mais cohérent ;
- un budget de l'État détaillé, aux instruments clairement identifiés.

Bilan et conclusion ; un succès provisoire

Une fois le modèle transmis, les partenaires ont pu se rencontrer, en

décembre 2002 et 2003, à l'occasion d'un congrès annuel auquel l'Insee participe régulièrement, et en juillet 2002 lors d'un séminaire à Varsovie.

Cela a été l'occasion, en particulier lors d'un entretien avec Ludwig Kotecki, d'avoir la confirmation que le modèle a été bien intégré au processus d'analyse, permettant de réaliser de façon autonome des études de politique économique, qui ont joué dans le processus de décision, mais que la poursuite de cet usage posait problème.

Au cours des deux années précédentes, le ministère des Finances avait établi une base de données très complète de l'économie polonaise, bien documentée et organisée, sur laquelle il appuyait ses études courantes. L'adaptation du modèle à ces nouvelles informations nécessitait trois choses : tout d'abord, une

2. La fonction de Cobb-Douglas est largement utilisée en économie pour représenter le lien qui existe entre intrant et extrant. Cette fonction a été proposée et testée économétriquement par les américains Paul Douglas, économiste, et Richard Cobb, mathématicien, en 1928. Supposons par exemple qu'une entreprise désire installer de nouvelles capacités de production et doit choisir entre plusieurs technologies. Le rapport des niveaux du capital et du travail installés évoluera comme le rapport de leurs coûts, lien représenté par la fonction de Cobb-Douglas.

3. Dans une entreprise, lorsqu'il y a une substitution constante entre le travail et le capital, aussi bien lors de l'installation de machines que pendant toute leur durée d'utilisation, on parle de modèle putty-putty (le mot anglais « putty » signifie « pâte à modeler »). Lorsqu'il n'y a aucune substitution des facteurs, on parle de modèle clay-clay (« clay » signifie « argile », l'argile n'étant pas malléable comme la pâte à modeler). Enfin, lorsque cette substitution n'a lieu que pour des machines nouvelles (ce qui signifie qu'il y a des possibilités de substitution ex ante mais non ex post), on parle de modèle putty-clay.

4. Le modèle WS-PS (*wage-setting, price-setting*), introduit par Layard-Nickell-Jackman (1991) et approfondi par Cahuc-Zylberberg (1999), constitue une référence pour fonder un taux de chômage d'équilibre. Il suppose que la part des salaires dans la valeur ajoutée résulte d'un processus de négociation s'appuyant sur le niveau du taux de chômage.

refonte totale des programmes d'alimentation, l'organisation technique des données étant complètement différente ; ensuite, une remise à niveau des concepts, les nomenclatures et certaines définitions ayant changé ; enfin, une réestimation des équations, à la suite de l'extension des périodes disponibles et du changement de base. Ceci ne concerne pas seulement les dernières périodes : leur ajout devait être mis à profit pour abandonner les premières années de la transition (1994 et 1995) aux mécanismes économiquement douteux et aux données basées sur des méthodes obsolètes.

En conséquence, même si le premier projet avait atteint son but, la continuité de cette coopération réclamait, de fait, non pas une assistance technique de routine, mais l'établissement d'un projet nouveau. Et si une volonté de coopération existait des deux côtés, elle restait à formaliser.

Le second projet : mars 2005 - février 2006

Début 2005, une demande formelle de poursuivre cette collaboration a été adressée à l'Insee. Cette demande faisait suite à une visite de Jean-Michel Charpin, alors directeur général de l'Insee, à Varsovie début novembre 2004 et à une réunion, début décembre à Varsovie, avec Ludwig Kotecki (chef du département) et Slawomir Dudek (chef du bureau).

5. Direction générale du trésor et de la politique économique, du ministère de l'Économie, des Finances et de l'Emploi.

La forme de cette collaboration s'est cependant révélée assez différente de la précédente. Il s'agissait cette fois plutôt d'une expertise et d'un suivi, intégrant les éléments suivants :

- assistance technique conceptuelle sur les aspects offre, prix et salaires du modèle ;
- appui à l'amélioration d'autres blocs du modèle ;
- évaluation des résultats et de la cohérence de l'outil ;
- discussion et consultation sur les propriétés du bloc de finances publiques ;
- expertise sur l'introduction des transferts de l'Union européenne et ses effets d'offre ;
- appui à l'utilisation opérationnelle du modèle (réestimations, prévisions, simulations, politique économique) ;
- appui technique sur la programmation du modèle sous EvIEWS.

Cette action a pris la forme de missions courtes, en 2005 et 2006, à peu près tous les trimestres.

La première mission (réalisée par une seule personne) a eu lieu du 23 au 25 mars 2005. Elle a permis de réaliser les transferts de données, d'établir le cadre théorique général et de discuter point par point des spécifications, et de définir la méthodologie du projet de modélisation.

La seconde mission, qui a eu lieu du 22 au 24 juin 2005 (avec Salvatore

Serravalle, de la division croissance et projections macroéconomiques de l'Insee), a permis de commencer à tester un modèle complet, en insistant sur la fonction de production qui conditionne la suite des travaux.

L'été 2005 a été consacré à des tests de simulation, dont les enseignements ont conduit à revenir fréquemment sur les spécifications.

La troisième mission a eu lieu du 20 au 22 septembre 2005, avec Thomas Le Barbanchon (même division que précédemment). Elle a essentiellement porté sur des tests du modèle. Les implications de la formulation sophistiquée des capacités de production étaient difficiles à mesurer. Pour y voir plus clair, on est revenu à une fonction plus simple de type Cobb-Douglas, démarche qui a permis des progrès sensibles.

La dernière mission (du 15 au 17 février 2006) a permis des progrès certains et la visite des experts polonais début juin à Paris a donné lieu à une présentation, qui a montré un modèle aux propriétés acceptables. Cette présentation a soulevé un grand intérêt de la partie française, en particulier des membres de la DGTPÉ⁵, qui ont établi des contacts formels avec les partenaires polonais.

Un dernier contact direct entre partenaires a eu lieu lors du congrès « Macromodels » organisé à Varsovie en décembre 2006. On a pu constater que le modèle était devenu opérationnel, même si des interrogations subsistent. ■

De la coopération à la collaboration : petite histoire de la construction d'une enquête sur la pauvreté et l'exclusion sociale en Pologne

 Daniel Verger et Thibaut de Saint Pol*

Entamée il y a plus de dix ans, la coopération avec la Pologne dans le domaine des enquêtes statistiques sur la pauvreté et l'exclusion sociale se poursuit encore aujourd'hui. Il s'agissait initialement d'une situation de coopération dans laquelle des experts français transmettaient à leurs collègues polonais leurs savoirs et leurs habitudes. Aujourd'hui, c'est une véritable collaboration, dans laquelle les statisticiens des deux pays travaillent ensemble pour élaborer une enquête novatrice, qui réponde au mieux au défi que représente l'étude des phénomènes de la pauvreté et de l'exclusion sociale dans un pays européen.

La coopération de l'Insee avec la Pologne, consacrée à la mise en place et à l'exploitation d'enquêtes statistiques sur le thème de l'exclusion sociale, s'inscrit dans un ensemble plus vaste de coopération avec plusieurs pays d'Europe de l'Est dans le domaine des inégalités et de la pauvreté.

Dès 1992, dans le cadre d'un projet de modernisation de la statistique publique en Tchécoslovaquie, république alors indépendante depuis deux ans, il avait été décidé d'étendre une action de l'Insee, déjà engagée dans le domaine de l'emploi et des salaires, au monde des enquêtes auprès des ménages. Il s'agissait alors de fournir aux statisticiens tchécoslovaques une analyse critique de leur système d'enquêtes, menée à la lumière des expériences françaises. Une mutation technologique s'imposait pour faire face aux exigences d'une future intégration européenne. En 1993, la partition du pays en deux, avec la création de la République tchèque et de la Slovaquie, a conduit à une révision des objectifs de cette coopération, les liens de la France se resserrant surtout avec la Slovaquie.

La première action continue menée avec la Slovaquie a consisté à monter une enquête sur le thème de la pauvreté, thème qui correspondait à une réelle urgence politique. Le processus de conception de l'enquête a permis de mettre en place un partenariat intellectuel avec les sociologues et les politiques du secteur social de



Vue de Varsovie : quartier populaire de Praga

ce pays, afin de bien recenser leurs besoins. Le modèle d'inspiration a été le panel européen, que la France mettait alors en œuvre sur son territoire, puisque l'idée était de mettre en place une opération allant dans le sens d'un rapprochement avec l'Europe.

Les résultats français et slovaques ont donc été élaborés presque simultanément et ont abouti à la rédaction d'un article dans le numéro spécial d'*Économie et Statistique* sur la pauvreté, publication qui, bien que datant de 1997, est encore classiquement citée comme référence.

Le projet polonais se rattache au projet slovaque

C'est à ce moment-là que la coopération avec la Pologne, initialisée de façon indépendante autour des responsables de l'enquête française

* Daniel Verger est inspecteur général de l'Insee, chef de l'unité des méthodes statistiques de la direction des statistiques démographiques et sociales de l'Institut ; Thibaut de Saint Pol est administrateur de l'Insee, chargé d'études à la division des conditions de vie des ménages, au sein du département des prix à la consommation, des ressources et des conditions de vie des ménages de la même direction de l'Insee.

Situations défavorisées, a rejoint le projet slovaque. Le questionnaire de l'enquête française a de fortes ressemblances avec celui du panel européen des ménages (ECHP), mais les exploitations traditionnelles qui en étaient faites n'avaient jamais été spécifiquement tournées vers le thème de la pauvreté. L'idée de faire converger les deux actions, de façon à obtenir une étude tripolaire avec une comparaison France-Pologne-Slovaquie, est alors apparue et, au gré des mutations diverses, l'équipe qui coopérait avec la Slovaquie a pris également en charge les actions avec la Pologne. La démarche était en effet identique et a donné peut-être encore plus de résultats dans le deuxième cas.

Au sein de l'office statistique polonais (GUS)¹, le soutien de la hiérarchie, surtout au niveau élevé du Vice-Président, a été essentiel. L'aide de chercheurs très motivés l'a été tout autant, en particulier Anna Bienkunska, spécialiste reconnue du domaine et désormais chargée de représenter la Pologne auprès d'Eurostat. Malgré des différences de questionnaires, il a été possible de réaliser une comparaison France-Pologne tout à fait interprétable et qui a fait l'objet d'un ouvrage en français et en polonais, présenté lors d'une conférence de presse à Varsovie.

Changement d'ampleur : le colloque de Bratislava en juin 2000

Les comparaisons France-Slovaquie et France-Pologne ont eu toutes les deux un certain retentissement, grâce à quelques présentations dans des colloques internationaux.

L'Insee, avec l'accord et le soutien financier d'Eurostat, y a vu l'occasion de donner une autre ampleur à ce projet et d'étendre ce genre de comparaisons à tous les pays dotés de statistiques de qualité équivalente, c'est-à-dire tous les pays participant au panel européen ainsi que ceux,

en Europe centrale, qui disposaient d'enquêtes plus ou moins proches. Avec certains pays, il a été possible de mettre en œuvre une méthodologie commune de manière complète, tandis qu'avec d'autres, on a procédé par un système de contributions associées, avec une méthodologie légèrement adaptée.

Ainsi est né le projet de colloque de Bratislava, qui s'est tenu dans la capitale slovaque en juin 2000, en hommage au pays qui était à l'origine de la démarche. Plusieurs déplacements et un important travail éditorial ont été nécessaires, afin de repérer les équipes nationales (parfois dans les offices statistiques, parfois parmi les universitaires, souvent en combinaison), les motiver pour le projet, présenter la méthodologie de recherche, l'adapter et accompagner la production des résultats en veillant à leur comparabilité.

Sous l'égide d'Eurostat, des participants de 22 pays ont assisté à la présentation de communications portant sur 13 pays : outre la France, la Pologne et la Slovaquie, les contributions ont porté sur l'Espagne, le Portugal, la Roumanie, la Hongrie, la République tchèque, l'Italie, la Belgique, l'Albanie, et même Madagascar et le Brésil.

Un nouveau défi : la rédaction d'un numéro spécial d'*Économie et Statistique*

Le succès de ce séminaire, qui devait être initialement l'aboutissement du travail, a convaincu les auteurs d'aller au-delà de la publication des actes et d'aboutir à un vrai volume de comparaison internationale, d'où le projet d'un numéro spécial d'*Économie et Statistique*. Pour ce faire, il fallait un éventail de pays plus divers, avec de nouvelles études, ainsi qu'une certaine actualisation des études originelles.

Le travail a été de longue haleine, souffrant de la mobilité de plusieurs responsables du projet, tant en France qu'à l'étranger, mais le numéro a fini par paraître en décembre 2005. Les

contributions ont bien été actualisées, la comparabilité des méthodes et des présentations renforcée.

Des pays qui n'avaient pas participé au colloque ont fourni des contributions inédites (Royaume-Uni, Russie...), de nouvelles synthèses ont été spécialement réécrites (Dell-Herpin). Quelques éclairages complémentaires, impossibles à réaliser dans tous les pays, ont été introduits afin d'illustrer la manière de faire évoluer l'approche principalement suivie dans le dossier : exploitant explicitement les évolutions, S. Lollivier et D. Verger ont tracé la voie permettant de dépasser le cadre statique restreint habituel, tandis que S. Paugam illustre la richesse potentielle d'une ouverture vers des dimensions jusqu'alors insuffisamment prises en compte, que ce soit l'expérience subjective de la pauvreté ou l'exclusion sociale.

Une certaine imbrication des actions

La vie du projet a ainsi continué, toujours plus vaste, plus complexe, plus diversifiée. Il est devenu de plus en plus difficile de distinguer ce projet d'autres actions de coopération. En effet, quand on traite d'enquêtes auprès des ménages, on est vite amené à traiter de toutes les opérations qui, enchaînées, forment le processus de production de l'enquête : l'échantillonnage en premier lieu, mais aussi les techniques d'imputation pour données manquantes et la pondération.

Le récent élargissement de l'Union européenne en direction de l'Est, au premier rang desquels la Pologne, a aussi suscité de nouvelles occasions de coopération, les offices statistiques nationaux étant amenés à mettre en œuvre sur leur territoire les enquêtes européennes.

En 2004, Eurostat lance le successeur du panel européen (SILC), enquête dont le principal sujet est la mesure des revenus, et un des objectifs majeurs la mesure de la pauvreté ; un séminaire d'une semaine a d'ores et déjà été organisé avec succès, et des cours

1. Pour une description du GUS, voir l'article de Pascal Rivière dans ce même dossier.



Vue de Varsovie : quartier populaire de Praga

dont l'ambition était de dépasser la mesure de la pauvreté pour la replacer dans le cadre plus général de l'exclusion sociale.

Dès l'été 2007, un test a été réalisé à l'office statistique régional de Bydgoszcz afin de vérifier la pertinence et l'efficacité des nouvelles questions. En septembre 2007, à Varsovie, une réunion de travail a permis de tirer parti des résultats de ce test pour élaborer le questionnaire d'une enquête pilote qui doit donner des résultats publiables concernant la région de Bydgoszcz, et dont la collecte a débuté en mars 2008. Le projet est d'étendre cette opération à l'ensemble du territoire en 2009, sous réserve que des moyens financiers adéquats soient trouvés.

Le caractère multidimensionnel de la pauvreté nécessite en effet que les savoirs et les pratiques de recherche sur ce thème soient harmonisés en Europe. Des travaux plus novateurs doivent donc être réalisés, de façon à améliorer l'appréhension de ce phénomène complexe.

C'est d'ailleurs dans ce cadre que se situe le travail actuel, car il s'agit de construire une nouvelle opération destinée à intégrer les aspects relationnels à l'analyse désormais plus classique de la pauvreté monétaire et de la pauvreté en termes de conditions de vie. Le projet d'étude de la pauvreté et de l'exclusion sociale sera en outre mené en collaboration avec des administrations et des instituts de recherche extérieurs au GUS (en l'occurrence le ministère polonais chargé des affaires sociales, ainsi que l'Académie des sciences polonaise), sur le modèle des collaborations que l'on met sur pied lors de la préparation des nouvelles enquêtes à l'Insee. Les autorités locales polonaises se sont aussi révélées intéressées par le projet, auquel certaines d'entre elles envisagent même d'apporter un financement.

La maîtrise d'œuvre polonaise de l'opération s'est imposé les mêmes standards de qualité que ceux de l'Insee (réalisation de plusieurs tests, avec formation par les concepteurs

sur ce thème mis sur pied au centre de formation de l'Insee à Libourne : une sorte de retour aux sources, puisque c'est ce panel européen qui, dès le début, a inspiré notre action.

Vers de véritables partenariats d'étude

La Pologne ayant désormais intégré l'espace européen, nous avons quitté le registre de la coopération avec ce pays. Mais les travaux franco-polonais se sont poursuivis, avec une approche inédite en matière d'exclusion, le questionnement polonais allant au-delà du questionnaire français.

Les résultats obtenus, publiés dans le numéro spécial d'*Économie et statistique* dont il est question précédemment, constituent donc une première qu'il était nécessaire de prolonger, quoique déjà stimulante. La coopération et l'aide apportées à la Pologne ont ainsi abouti à un partenariat dans lequel la France et la Pologne explorent ensemble de nouvelles voies. On n'en est plus aujourd'hui à faire connaître et à diffuser nos pratiques, mais à découvrir

des voies d'amélioration et d'approfondissement.

Une expérience similaire a été tentée avec la Russie au cours de la même période, certains chercheurs de l'Académie des Sciences de ce pays ayant étudié nos travaux, les ayant adaptés et, sur certains points, développés.

C'est dans ce nouveau contexte, qu'à l'automne 2006, ont été lancés conjointement deux nouveaux projets visant à élaborer une enquête sur l'exclusion sociale, l'un en Russie, l'autre en Pologne, avec une méthodologie d'approche commune, alors même qu'il n'existe pas de véritable enquête sur l'exclusion sociale dans le dispositif des enquêtes auprès des ménages de l'Insee. C'est donc ensemble qu'il a fallu construire de toute pièce ce projet.

L'élaboration d'une nouvelle enquête sur l'exclusion sociale en Pologne

C'est en janvier 2007 que les bases d'un programme de travail en commun avec la Pologne ont été posées, comprenant les grandes lignes que devait respecter un questionnement

et bilans détaillés en particulier) ; les experts français ont ainsi pu, au cours d'une mission à Varsovie, rencontrer deux représentants de l'office régional où le test a été réalisé. Ils ont pu établir que, dans ses grandes lignes, l'opération avait fonctionné, et ils se sont aussi rendu compte de certaines difficultés rencontrées.

Pour qui est accoutumé à de telles séances de bilan en France, le dépaysement n'est pas très fort, du fait d'un grand nombre de similitudes : il a ainsi été noté, comme c'est fréquemment le cas en France, que la longueur de l'enquête avait paru excessive à certains enquêtés et que certaines questions étaient répétitives, trop détaillées ou trop intimes.

Mais quelques éléments différencient tout de même les deux pays : les questions relatives à la perception de la pauvreté ont soulevé un malaise chez un grand nombre de ménages polonais, alors que les questions similaires avaient été plutôt bien acceptées en France. La difficulté spécifique à la Pologne vient sans doute du fait que la part de ménages victimes de ces privations matérielles y est beaucoup plus forte que dans des pays comme la France. L'enquête renvoie donc à un plus grand nombre de personnes une image dévalorisée d'eux-mêmes.

La formulation doit donc faire l'objet d'une attention particulière, voire de nouveaux tests. Les questions ouvertes ont également posé problème et ont été souvent mal remplies, sans doute parce que les enquêteurs ne sont pas habitués à ce type de questions.

L'analyse du test a conduit à rappeler l'importance de la formation des enquêteurs. L'accompagnement de ces derniers par des concepteurs et des responsables de l'office régional, fonctionnement encore inhabituel en Pologne, est une piste d'amélioration à ne pas négliger. Quant à la réalisation de l'opération de collecte avec un ordinateur portable, elle n'est pas apparue possible, pour des raisons purement matérielles, et l'enquête sera effectuée au moyen de ques-

tionnaires papier. Nul doute que la convergence des pratiques va continuer à l'avenir, quand le contexte matériel aura évolué.

Comme pour le test, l'enquête pilote est limitée à la zone de Bydgoszcz sur la base de 1 500 ménages représentatifs de la région. Un cofinancement par l'administration régionale est prévu avec, à la clef, une ou plusieurs publications locales. L'enquête pilote a eu lieu de mars à mai 2008 ; ses résultats seront exploités durant le second semestre 2008.

L'objectif est d'introduire l'enquête sur l'exclusion sociale en elle-même (qui fera suite à l'enquête pilote) dans le programme des enquêtes du GUS, et ce, dès le dernier trimestre 2009. Une nouvelle mission à Varsovie a eu lieu en septembre 2008, afin de définir les grandes lignes de l'exploitation. Une dernière mission, à Bydgoszcz, pourrait avoir lieu en fin d'année afin de participer à la conférence de presse au cours de laquelle les résultats seront présentés.

Quelques considérations sur les facteurs de succès ou d'échec

Au travers du récit précédent, on voit clairement se dessiner les facteurs d'une coopération efficace, mais aussi les difficultés et les obstacles qui peuvent encore faire échouer certaines initiatives. Tentons de lister quelques éléments décisifs, d'après cette expérience.

- **Des problèmes de communication** : il ne faut pas sous-estimer les difficultés de compréhension dues aux différences de langue. Le domaine décrit nécessite en effet d'avoir des notions de statistique, de sociologie et d'économie, de maîtriser des concepts théoriques parfois complexes et d'être en mesure de s'interroger sur la compréhension fine d'une question par l'enquêté, afin de savoir si, d'un pays à l'autre, une même question est bien interprétée de la même façon.

On dépasse donc de très loin les exigences habituelles demandées à un interprète. On ressent de façon impérieuse la nécessité d'avoir un interprète compétent, spécialisé sur le thème étudié et qui accepte de reconnaître quand il n'est pas sûr d'avoir exactement compris le message. Quand il a fallu, pour des raisons conjoncturelles, s'en passer, on a vite mesuré la perte d'efficacité. Et ce n'est pas un problème spécifique au français. Même en anglais, la communication se révèle appauvrie, réduite, l'anglais étant rarement maîtrisé de part et d'autre à un niveau suffisamment élevé pour communiquer sur les nuances d'une question ou pour comprendre les langages scientifiques spécifiques. De ce point de vue, la mise au point du texte franco-polonais a été une expérience « inoubliable » : on a parfois passé plus d'une demi-heure sur une seule phrase pour trouver l'équivalence exacte dans les deux langues. On était loin de se contenter d'une traduction mot à mot.

Les problèmes de communication ne sont pas seulement affaire de langue, ils se déclinent aussi en termes informatiques. Les différences de logiciels utilisés compliquent la transmission du savoir-faire : en Pologne, l'absence du logiciel SAS, très coûteux pour le pays, a parfois été un frein.

- **La nécessité d'avoir du temps** : toutes ces démarches sont à l'évidence de long terme ; déjà en France, les délais de production d'une enquête sont longs. Dans un pays où l'on innove, il faut en plus que les savoirs soient assimilés, les techniques apprivoisées, les habitudes anciennes modifiées. Une condition impérative est que, des deux côtés, on puisse jouir d'une certaine stabilité des équipes.

- **La nécessité de disposer du fichier de données** : il est vite apparu que le travail sur les données nécessitait que ces dernières soient accessibles aux deux équipes polonaise et française. C'est la seule façon d'avoir la souplesse nécessaire pour tester des variantes



Source : droits réservés

Vue de Varsovie

de modèle : envoyer un message au statisticien local, qui le transmet pour exécution à son informaticien, serait en effet une procédure à la fois trop lourde et trop lente dans la phase de réglage de la fin de la méthode.

Nos partenaires ont en général réussi à obtenir de leur hiérarchie l'autorisation de sortir du territoire les données (anonymisées), même si cela a parfois été difficile et a demandé du temps. Mais cette possibilité, encore une fois, s'avère essentielle pour l'efficacité de la coopération.

Finalement, un bénéfice mutuel

On le voit, les défis associés à un tel projet sont de taille. Pour nos partenaires, la réussite s'est traduite par un transfert de la France vers

la Pologne de technologie statistique, notamment dans les domaines suivants : promotion de nouvelles études comparatives, diffusion de méthodes d'exploitation performantes (économétrie des données qualitatives grâce au modèle *Logit* en particulier), sensibilisation aux problèmes d'échantillonnage et au respect du caractère aléatoire des échantillons, perfectionnement dans les méthodes de calage et de détection-correction des valeurs extrêmes aberrantes, et passage de données en classe à des données en continu (« résidus simulés »).

De notre point de vue, on peut parler de réussite globale et d'enrichissement mutuel ; le retour sur investissement est en effet indéniable. Le projet a été l'occasion d'élargir notre horizon et de nous inciter à une réflexion critique sur les insuffi-

sances de nos propres approches : par exemple, le fait de pouvoir étudier des sociétés autres, d'une complexité différente, car en évolution rapide, avec une part plus importante du secteur non monétaire, de l'informel, et parfois plus hétérogènes géographiquement, permet de se réinterroger sur le bien-fondé de certains des choix adoptés dans nos enquêtes.

L'expérience acquise nous a aussi permis d'être plus convaincants sur le plan pédagogique : on a par exemple découvert que si l'on appliquait sans précautions les définitions habituelles, on observait une pauvreté plus grande en France qu'en Slovaquie, alors que le niveau de vie de la première est trois à quatre fois supérieur à celui de la seconde ! Cela a permis à tous, y compris aux décideurs français, de mieux comprendre

les avantages et les inconvénients d'une mesure de la pauvreté relative qui, jusqu'alors, étaient connus d'un point de vue théorique, mais n'avaient pas été clairement identifiés comme susceptibles de créer de graves contresens.

La légitimité internationale confortant la crédibilité nationale, l'expérience plus large acquise grâce à la coopération nous a ainsi permis de mieux communiquer, d'avoir davantage

d'arguments concrets pour défendre nos choix de méthode mais aussi pour relativiser nos résultats, ce qui a été fort utile dans un contexte où le dialogue scientifique serein n'est pas toujours facile à établir.

De nombreux choix implicites à nos méthodes, acceptés jusqu'ici comme allant de soi, montrent leurs limites dès qu'il s'agit de rendre compte de sociétés ou d'économies moins stables, moins homogènes, moins

monétarisées que les nôtres : et cela nous incite à plus de modestie, nos approches de l'inégalité et de la pauvreté apparaissant en effet loin d'être abouties, car trop simplificatrices, ignorant trop de phénomènes qui, pour minoritaires qu'ils soient chez nous, n'en sont pas pour autant négligeables ailleurs, quand il s'agit d'étudier les démunis, les pauvres et les exclus. La coopération nous indique en quelque sorte les voies de progrès pour notre propre démarche. ■

L'implication croissante des services statistiques ministériels dans la construction de la statistique européenne

 Bernard Guibert*

La plupart des services statistiques ministériels (SSM) sont fortement impliqués dans la statistique européenne. Cette implication s'est accélérée au cours du « plan statistique quinquennal européen » 2003 – 2008, mais de manière inégale selon les départements ministériels et l'importance, pour ces derniers, des politiques européennes. Cette accélération a été induite par celle qu'a connue l'ensemble de la statistique européenne, à la suite de trois événements : la création de l'euro, celle de la Banque centrale européenne et la mise en place d'une « gouvernance économique européenne » (avec le conseil Ecofin) après l'adoption du traité de Maastricht. Les relations des SSM avec la statistique européenne empruntent deux voies, l'une passant par l'Insee et Eurostat, l'autre par le ministère du SSM en question et la Commission européenne à Bruxelles. C'est par cette seconde voie que les ministres français travaillent avec les directions générales de la Commission pour concevoir et mettre en œuvre des outils d'évaluation des politiques publiques. Ces travaux d'élaboration des indicateurs européens ne manquent pas d'avoir des effets en retour, d'une part sur les indicateurs nationaux, d'autre part sur les demandes de statistiques européennes.

Le bilan du programme 2004-2008 du Conseil national de l'information statistique (Cnis), réalisé pour préparer le programme à moyen terme 2009 – 2013, montre, entre autres, que la construction de la statistique européenne détermine de plus en plus les orientations et les programmes de travail du système statistique public (SSP) français. Or, celui-ci a une organisation relativement originale¹ en Europe : certes, comme dans tous les autres pays, il existe un institut statistique national (l'Insee), mais celui-ci coordonne en outre un ensemble complémentaire de services statistiques ministériels (SSM), au nombre de 18 au 1^{er} janvier 2008 (encadré page suivante).

Le présent article montre que la croissance récente et spectaculaire de la statistique européenne s'est répercutée au moins aussi fortement auprès de ces SSM qu'auprès de l'institut statistique national lui-même (autrement dit l'Insee), mais de manière très inégale, selon la dimension et l'éventail des missions assurées de chaque SSM et selon l'intensité et la nature de l'implication de leur ministère dans les politiques européennes.



Symbole de l'euro devant la Banque centrale européenne (BCE)

Pour cela, l'article rappelle brièvement l'hétérogénéité des SSM et la participation inégale à la construction politique européenne des départements ministériels où ils se trouvent, afin d'éclairer les intensités et les dynamiques différenciées de leur contribution à la statistique européenne. Il esquisse en conclusion les

formes que prendra vraisemblablement l'augmentation de l'importance de cette mission, qui restera très variable d'un ministère à l'autre mais se généralisera.

* Bernard Guibert, administrateur de l'Insee, est chargé de mission auprès du directeur de la coordination statistique et des relations internationales (DCSRI) de l'Institut.

1. La mission d'analyse comparative internationale des instituts statistiques nationaux, supervisée par l'Inspecteur général des finances Philippe Dane en décembre 2004 (*Mission d'analyse comparative internationale de l'Insee*, rapport de l'Inspection générale des finances, décembre 2004, point 2, page 4), insiste sur l'originalité de l'organisation du système statistique national français. Ce rapport caractérise le système français, d'une part, par une forte décentralisation fonctionnelle et, d'autre part, par une forte coordination statistique assurée par l'institut statistique national (l'Insee). La décentralisation fonctionnelle se traduit par l'existence de services statistiques de ministères, les SSM, aux effectifs non négligeables et remplissant des missions statistiques complémentaires de celles de l'Insee. En particulier, les annexes par pays (Allemagne, Pays-Bas, Royaume-Uni, Canada, Danemark, États-Unis d'Amérique, Italie, Suède) ne décrivent pas de manière exhaustive les systèmes statistiques publics européens et d'Amérique du Nord. Mais quand ces derniers possèdent l'équivalent des SSM français, le rapport pondère leur importance par leurs effectifs. Dans ces conditions (page 9), sur huit pays examinés, le rapport en dénombre quatre « fonctionnellement très fortement centralisés », deux à 100 % (Pays-Bas, Canada) et deux à plus de 80 % (Suède, Danemark).

L'influence croissante de la statistique européenne sur les statistiques nationales

Les SSM participent fortement à la contribution de la statistique française à la statistique européenne

« Le développement de la statistique européenne a eu et aura de plus en plus d'influence sur l'évolution de la statistique dans les États membres ». C'est la première phrase d'un article que Michel Glaude a consacré aux « Enjeux et défis de la statistique européenne »². Le phénomène touche tous les États membres de l'Union européenne. Ainsi, pour l'Office fédéral statistique allemand³, les 4/5 des statistiques de base et les 2/3 des exploitations secondaires seraient déterminées dans ce pays par le « programme statistique communautaire ». Le nombre de règlements statistiques examinés par le Comité des programmes statistiques (CPS) d'Eurostat à Luxembourg est croissant. Le nombre de transmissions de données du système statistique français à Eurostat⁴ a été multiplié par près de quatre en huit ans : la France a en effet transmis à Eurostat 164 flux en 1999 et 621 en 2007.

Et c'est bien la statistique européenne qui est à l'origine de cette mutation. Certes, le système statistique public (SSP) français n'adresse pas des flux internationaux de données statistiques seulement à Eurostat, mais également à d'autres organisations internationales comme le Conseil de

2. GLAUDE, Michel, Enjeux et défis de la statistique européenne, *Courrier des statistiques* n° 121-122, mai-décembre 2007, p. 19.

3. RADERMACHER, Wilhelm, Knowledge Economy : Challenges of Official Statistics, Eurostat Conference, Luxembourg, 8 et 9 décembre 2005.

4. Pour plus de détails, voir MICHEL DE PRÉVIA, Bertrand, Transmission internationale de données statistiques françaises, *Courrier des statistiques* n° 121-122, mai-décembre 2007, p. 43.

5. Organisation de coopération et de développement économiques.

6. Organisation des Nations unies.

7. Fonds monétaire international.

8. Bureau international du travail.

9. Op. cit. MICHEL DE PRÉVIA, graphique n° 3, p. 44.

10. Op. cit. MICHEL DE PRÉVIA, graphique n° 2, p. 44.

Encadré : les SSM et leur évolution dans le cadre de la RGPP

Au 1^{er} janvier 2008, la liste des SSM était la suivante :

Département ministériel	Service statistique
Agriculture	Service central des enquêtes et études statistiques (Scees)
Collectivités locales	Département des études et des statistiques locales (DESL)
Culture	Département des études, de la prospective et des statistiques (DEPS)
Défense	Observatoire économique de la défense (OED)
Douanes	Département des statistiques et des études économiques (DSEE)
Éducation, recherche	SSM éducation recherche, sous-ensemble de la direction de l'évaluation, de la prospective et de la performance (DEPP)
Énergie	Observatoire de l'économie de l'énergie et des matières premières (OEEMP)
Environnement	Institut français de l'environnement (Ifen)
Fonction publique	Bureau des statistiques, des études et de l'évaluation (BSEE)
Jeunesse, sports	Mission des études, de l'observation et des statistiques (MEOS)
Industrie	Service des études et des statistiques industrielles (Sessi)
Justice	Sous-direction de la statistique, des études et de la documentation (S/D SED)
Pêche maritime, aquaculture	Bureau central des statistiques (BCS)
PME, commerce, artisanat, services, prof. libérales	Observatoire des activités commerciales, artisanales, de services et des professions libérales (OACASPL)
Santé, solidarité	Direction de la recherche, des études, de l'évaluation et des statistiques (Drees)
Transports, construction	Service économie, statistiques et prospective (SESP)
Travail, emploi	Direction de l'animation de la recherche, des études et des statistiques (Dares)
Tourisme	Département de la stratégie, de la prospective, de l'évaluation et des statistiques (DSPES)

La Révision générale des politiques publiques (RGPP) se traduit d'ores et déjà par une réorganisation en profondeur des architectures de certains ministères, ce qui a des conséquences substantielles sur les périmètres, les places dans les organigrammes et les dénominations d'un certain nombre de SSM.

Ainsi, au cours de l'été 2008, la plus grande partie du Service des études et des statistiques industrielles (Sessi) est intégrée à l'Insee, et les services statistiques de l'environnement (Ifen : Institut français de l'environnement), de l'énergie (OEEMP : Observatoire de l'économie de l'énergie et des matières premières) et des transports et de la construction (SESP : Service économie, statistiques et prospective) sont regroupés au sein du Service de l'observation statistique du Commissariat au développement durable du Ministère de l'écologie, de l'énergie, du développement durable et de l'aménagement du territoire (MEEDDAT).



Le siège du FMI à Washington

l'Europe, l'OCDE⁵, l'ONU⁶, le FMI⁷, le BIT⁸, etc. Mais si on se limite à l'Insee, 59 % de ces émissions sont réservées à Eurostat⁹. C'est dire la place centrale qu'occupe ce dernier.

Quant aux SSM, ils ne sont pas en reste, puisqu'au 1^{er} janvier 2007, alors que leurs effectifs représentent 30 % de ceux de l'ensemble de la statistique publique française, ils sont à l'origine de près de 50 % des flux de données statistiques transmis de la part de ce même ensemble vers Eurostat¹⁰.

Le fait qu'à lui seul, le service statistique de l'agriculture transmette plus du quart (160) des flux laisse deviner que l'implication des différents segments du SSP français est très inégale. On peut supposer que plus la politique est ancienne, plus la part dans le budget communautaire est élevée et plus les flux statistiques

émis sont importants. C'est le cas de la politique agricole commune (PAC).

Le caractère récent de l'implication forte des SSM

L'implication des SSM dans la statistique européenne est donc quantitativement importante en 2007. Mais, à l'exception de l'agriculture et des SSM directement impliqués dans la constitution du marché unique (douanes, industrie, transport), cette implication est relativement récente. En 2000, elle paraissait faible aux responsables des SSM. Dans le cadre d'une mission confiée en 2000 à Michel Blanc et à Alain Desrosières, pour décrire le réseau des SSM « du point de vue de leur spécificité et de leurs attentes par rapport aux différentes formes de coordination existantes, à développer et à créer »¹¹, les auteurs constataient :

« Le système statistique public (SSP) français est pratiquement le seul qui, en Europe, soit organisé sur le double principe de décentralisation et de coordination. Dans la plupart des autres pays européens, l'activité statistique est en principe centralisée dans un institut national statistique, même si d'autres organismes que cet institut national statistique peuvent avoir des attributions statistiques immergées dans leur activité principale. C'est également le cas pour la Commission européenne, dont la statistique est en principe concentrée à Eurostat, mais où, de fait, nombre de questions qui, chez nous, relèvent de la responsabilité de SSM, sont traitées par diverses directions générales (DG) autres qu'Eurostat. Ceci explique que, interrogés sur leurs relations avec les autres pays européens et avec les institutions de l'Union, les responsables des SSM fournissent des réponses variées, où souvent Eurostat ne joue pas un rôle important, sinon aucun rôle. »¹²

Nous étions en 2001, soit deux ans avant le début du programme à moyen terme du Cnis 2003 – 2008 et un an avant le début du programme statistique communautaire quinquennal 2002 – 2007.



Siège de l'ONU à New York

La création de l'euro et l'élargissement de l'Union ont accéléré la construction de la statistique européenne

Ce caractère récent de l'implication des SSM correspond à l'accélération de la construction de la statistique européenne, à la suite de la création de l'euro (1^{er} janvier 1999), de la Banque centrale européenne (juin 1999) et de l'adoption concomitante du pacte de stabilité. Le Conseil ECOFIN (ministres des finances de l'Union économique et monétaire européenne) de juin 2000 a en effet adopté un « EMU Action Plan »¹³ qui visait à satisfaire, à partir de décembre 2002, les besoins statistiques urgents définis par la Banque centrale européenne (BCE) pour conduire la politique monétaire commune et vérifier le respect par les États membres des critères de Maastricht.

Pour préciser les données requises par la BCE, une mission de *benchmarking*¹⁴ a été conduite en 2000 et 2001 pour comparer la statistique européenne à celle des États-Unis. Cette évaluation a induit une demande d'indicateurs de court terme (PEEI)¹⁵ et de raccourcissement des délais de transmission des données statistiques à Eurostat et à l'institut monétaire européen. Cette demande pressante et forte de l'ECOFIN¹⁶ concernait plus particulièrement cinq « lignes de

produits statistiques » : les comptes nationaux trimestriels, les comptes trimestriels des administrations publiques, les statistiques du marché du travail, les chiffres de la conjoncture économique et les statistiques douanières.

L'Insee a été le plus directement concerné par cette accélération de la construction de la statistique européenne (indices des prix, comptes, conjoncture, statistiques de court terme, emploi, etc.). Mais un certain nombre de règlements statistiques adoptés pendant la décennie précédente, les règlements STS (*Short Term Statistics*), SBS (*Structural Business Statistics*), PROSTAT (statistiques de productions industrielles) et

11. BLANC Michel et DESROSIÈRES Alain, Entre décentralisation et coordination. Une analyse des spécificités des SSM, *Courrier des statistiques* n° 104, décembre 2002.

12. Les mots soulignés au moyen de l'italique le sont par l'auteur de cet article.

13. EMU : European Monetary Union.

14. Le *benchmarking* (en français : étalonnage, analyse comparative ou parangonnage) est une technique de marketing ou de gestion de la qualité qui consiste à étudier et analyser les techniques de gestion et les modes d'organisation des autres entreprises, afin de s'en inspirer, pour améliorer la performance des processus dans une organisation [NDLR].

15. Principal European Economic Indicators, ou principaux indicateurs économiques européens.

16. Voir le communiqué de l'ECOFIN du 6 novembre 2001 : <http://europa.eu/rapid/pressReleasesAction.do?reference=PRES/01/401&format=HTML&aged=0&lg=fr&guiLanguage=en>, page 6.

INTRASTAT (statistiques de commerce extérieur intra-communautaire), ont été réformés. Les services statistiques de l'agriculture, de l'industrie, des transports et des douanes ont supporté les réformes de ces règlements.

Le passage, au début des années 2000, d'une Europe à 15 membres à une Europe à 25 puis à 27 alourdit et allonge les procédures de concertation et rend incontournable son aspect multilatéral. Cet élargissement a eu un effet qualitatif analogue à celui de la création de l'euro en incitant la statistique européenne à s'impliquer dans les procédures d'évaluation des politiques publiques européennes. En effet, les services statistiques nationaux des nouveaux États membres ont souvent été désorganisés et affaiblis pendant la phase de transition à l'économie de marché.

La remise à niveau – appelée « reprise des acquis communautaires » –, y compris dans le domaine de la technique statistique et des financements des enquêtes, prend dès lors des aspects politiques qui renforcent le poids des DG de Bruxelles dans la détermination du programme statistique communautaire quinquennal. Ce dernier est un acte juridique fort puisqu'il est adopté par le Conseil et le Parlement européen sur initiative de la Commission. Les priorités statistiques qu'il affiche reprennent les différents titres du traité de Maastricht. Il planifie en quelque sorte les « règlements statistiques cadres », dont la Commission prendra l'initiative, et détermine indirectement, par l'intermédiaire du Comité des programmes statistiques (CPS) de Luxembourg, les programmes des services producteurs de statistiques français des ministères les plus impliqués par les règlements statistiques en question.

L'équilibre de la gouvernance de la statistique européenne, fragilisée par la crise d'Eurostat au début des années 2000, s'est donc déplacé en faveur des autorités politiques de l'Union, que ce soit celles de l'ECOFIN et de

Source (de gauche à droite et de haut en bas) : Drees, Insee, Ifen, ministère de l'Éducation nationale, Statistique publique, ministère de la Culture, Dares, Cnis et Sessi



Quelques logos de la statistique publique

la Banque centrale européenne ou des DG à Bruxelles. À cause de ce déplacement, la demande statistique européenne est de plus en plus déterminée par le souci de l'évaluation des politiques, souci partagé par les DG et par les départements ministériels français auxquels appartiennent les SSM.

Les délais de préparation des « règlements statistiques cadres » sont approximativement d'un an et demi pour formuler la proposition de la Commission qui en a l'initiative, et d'un an pour l'adoption en codécision (Conseil et Parlement), soit, en tout, de l'ordre de deux ans et demi. Les délais d'adoption des « règlements statistiques d'application » sont plus courts, de l'ordre de l'année et demie. Ainsi, le processus complet dure quatre à cinq ans. Dans ces conditions, les changements du début 2000 ont pleinement manifesté leurs effets au début du programme à moyen terme du Cnis 2004 – 2008, ainsi qu'au début du programme statistique communautaire quinquennal 2003 – 2007.

La complexité des relations entre la statistique publique française et la statistique européenne

L'organisation « fonctionnellement décentralisée » de la statistique publique française

Les effets de cette brusque et importante augmentation de la demande européenne, pour les besoins de l'évaluation des politiques de l'Union à la suite de la création de l'euro et de l'élargissement de l'Union, ont été proportionnés à l'intensité des liaisons existant auparavant entre les SSM et la statistique européenne. Or, les SSM sont extrêmement hétérogènes en taille et en nature.

L'originalité de la statistique publique française a été caractérisée dans le rapport de Philippe Dane¹⁷, à la fois par :

- une *décentralisation* fonctionnelle : existence de 18 SSM dans des minis-

17. DANE, Philippe, *Mission d'analyse comparative internationale de l'Insee*, Rapport de l'Inspection générale des finances, décembre 2004, point 2, page 4.

tères administrativement distincts de l'Insee *stricto sensu* ;

- une *puissante coordination technique* par l'Insee, via la direction de la coordination statistique et des relations internationales (DCSRI) et les directions verticales spécialisées, et une concertation avec les utilisateurs par l'intermédiaire du Cnis.

Du fait même de cette organisation « fonctionnelle décentralisée », la coordination des programmes statistiques des SSM et de l'Insee, et leur intégration en un seul programme national cohérent à moyen terme, est une tâche complexe. Chaque SSM doit en effet répondre à la fois aux besoins de la statistique publique en général et à ceux, spécifiques, de son propre département ministériel. La « puissante coordination technique » par l'Insee – pour reprendre les termes du rapport Dane – fait en sorte que ces besoins ne soient pas fondamentalement différents. Mais les priorités et les urgences des SSM au sein de leur ministère se déclinent selon un agenda politique qui n'obéit pas à la même temporalité que la programmation des investissements lourds de la statistique publique.

À cause de cette organisation « fonctionnelle décentralisée », originale en Europe, les relations entre le SSP français et Eurostat sont plus *complexes* en France que dans d'autres pays, notamment dans ceux où le SSP, complètement *centralisé*, coïncide avec le seul institut national statistique (INS)¹⁸. En effet, chacun des 18 SSM français est, d'une part, un des services du département ministériel spécialisé où il se trouve (agriculture, équipement, industrie, etc.). Il est donc sous l'autorité du ministre. D'autre part, en tant que service statistique, ses activités sont coordonnées dans le domaine technique (méthodes, nomenclatures, qualité statistique, etc.) par l'Insee¹⁹, direction générale du ministère de l'Économie, de l'Industrie et de l'Emploi (MEIE).

S'il existe par ailleurs, au niveau européen, une « politique commune » ou une « compétence partagée » dans un



Conseil de l'Europe à Strasbourg

domaine donné (agriculture par exemple), les attributions du ministre français correspondant sont telles qu'il a pour interlocuteurs politiques privilégiés, à la Commission de Bruxelles, à la fois un Commissaire et une DG. Ce ministre privilégie alors naturellement ses relations politiques, par rapport à celles, techniques, que son SSM peut entretenir avec Eurostat, d'autant plus que ces relations techniques se font par l'intermédiaire de l'Insee.

La demande statistique s'élargit aux indicateurs permettant d'évaluer les politiques

En outre, chacun des SSM a une histoire plus ou moins longue et a acquis une position plus ou moins élevée dans la hiérarchie du ministère. Il a ainsi pu compléter son « cœur de métier statistique » avec une ou plusieurs des quatre fonctions qu'Edmond Malinvaud jugeait complémentaires dans son rapport de 1997²⁰ : études, recherche, prospective et évaluation des politiques publiques. Chaque SSM est ainsi plus ou moins mis à contribution, à condition qu'il soit suffisamment étoffé et outillé pour le faire, par son ministre, pour contribuer à l'évaluation de la politique publique dont il a à répondre au sein du gouvernement.

Lors de leur mission déjà évoquée, Michel Blanc et Alain Desrosières²¹ avaient donné une description simplifiée du système complexe des SSM grâce à une typologie de ces derniers²². Elle combinait la taille du SSM et l'adjonction éventuelle, à leur mission statistique de base, de fonctions complémentaires d'études, de recherche et d'évaluation, conformément aux analyses et aux recommandations du rapport Malinvaud. Cette typologie ayant en commun avec celle présentée plus bas le critère de la taille, elle n'en diffère pas radicalement. Si le critère de la taille a peu varié, en revanche le deuxième critère, l'intensité de la détermination par la statistique européenne, qui était alors assez faible²³ est devenu inégalement importante selon le SSM.

18. C'est par exemple le cas aux Pays-Bas et au Danemark. En Suède, il existe une trentaine de SSM, mais leurs effectifs ne représentent à eux tous que 10 % du SSP.

19. Le décret de 1946 qui fixe les attributions et l'organisation de l'Insee lui donne la responsabilité de la coordination technique de l'ensemble du SSP.

20. MALINVAUD Edmond, *La fonction « statistique et étude économique » dans les services de l'État* (rapport rendu au Premier ministre en 1997).

21. BLANC Michel et DESROSIÈRES Alain, *Entre décentralisation et coordination. Une analyse des spécificités des SSM*, *Courrier des statistiques* n° 104, décembre 2002.

22. Cette typologie est présentée en pages 15 et 16 de l'article cité dans la note précédente.

23. Voir supra (p. 35) la citation extraite de l'article.



Source : Wikipédia

Le Berlaymont à Bruxelles, siège de la Commission européenne

Les besoins d'évaluation des politiques publiques en Europe

Au niveau européen également, des besoins d'évaluation des politiques communes, à commencer par celle de la Banque centrale, ont des conséquences importantes sur les demandes de statistiques. Il peut s'agir de *statistiques* proprement dites, demandées par Eurostat dans le cadre des « règlements statistiques » de plus en plus nombreux, statistiques recueillies par voie d'enquête ou, de plus en plus, par exploitation de fichiers administratifs.

Mais il peut s'agir également, et cela, de manière croissante, d'*indicateurs* permettant de faire des comparaisons entre États membres (*benchmarking*) pour les politiques qui ne sont pas communes mais qui sont coordon-

nées sur une base volontaire dans le cadre de la Méthode ouverte de coordination (MOC). C'est le cas, par exemple, des « indicateurs structurels » liés à la stratégie de Lisbonne lancée en 2000 et relancée en 2005, après une évaluation à mi-parcours jugée plutôt décevante. C'est également le cas des stratégies pour l'emploi (stratégie de Barcelone), de la lutte contre la pauvreté (indicateurs de Laeken) ou des « indicateurs de développement durable » liés à la stratégie européenne de développement durable (stratégie de Göteborg de 2006).

Une typologie des SSM dans leurs relations avec la statistique européenne

Pour comprendre l'implication de chaque SSM dans la construction de la statistique européenne, on a choisi de caractériser la nature et l'intensité de ses relations avec quatre pôles : deux en France, le ministre qui a autorité sur le SSM, et l'Insee, qui coordonne le système statistique public français ; et deux pôles au niveau européen, avec, du côté politique, la DG et le Commissaire de la Commission qui sont les interlocuteurs du ministre français à Bruxelles, et du côté statis-

tique, Eurostat, qui est l'office statistique de l'Union européenne.

À partir de ces quatre relations, on a dressé²⁴ une typologie des SSM dans leurs rapports avec la statistique européenne au début du programme à moyen terme 2004 – 2008 du Cnis. Le croisement de l'opposition entre « gros » et « petits » SSM, d'une part, et de l'opposition entre les SSM dont les relations avec les institutions européennes sont intenses ou faibles (voire inexistantes), d'autre part, donne en théorie une typologie à quatre classes, mais en pratique seulement trois, car la classe des « gros services statistiques ministériels avec des relations internationales faibles » est vide (tableau 1).

La plupart des ministères qui sont les instruments des interventions économiques et sociales de l'État, et qui ont été créés essentiellement au moment de l'essor des politiques de l'État-providence (économie, industrie, affaires sociales, emploi, santé, équipement, etc.), ont besoin de statistiques pour conduire leur politique d'intervention. En conséquence, ils ont développé des SSM importants. *A contrario*, ce n'est pas le cas de la plupart des ministères dits régaliens, comme ceux des affaires étrangères, de l'intérieur et de la justice. Ceux-ci agissent en effet plus au moyen de normes législatives et réglementaires que par des interventions économiques et sociales. Leurs SSM sont donc relativement modestes par rapport à leur importance politique ou même à l'effectif global des fonctionnaires qu'ils emploient.

Ces derniers cas mis à part, le rang élevé d'un SSM dans l'organigramme de son ministère et un effectif relativement important de ses agents déterminent bien souvent le degré d'influence auquel peut prétendre le responsable du SSM dans d'éventuels *groupes de travail* ou *task force* de la statistique européenne²⁵ à Luxembourg (Eurostat) ou dans des commissions *ad hoc* à Bruxelles. Ces dernières sont créées à l'initiative de la DG homologue du département ministériel français, notamment pour convenir de méthodes d'évaluation des politiques européennes et éven-

24. Le matériau à partir duquel a été construite cette typologie est une série d'entretiens conduits au dernier trimestre 2005 avec tous les responsables de SSM sur la base d'un guide les invitant à caractériser le positionnement de leur service dans le champ institutionnel constitué par le quadrilatère « Bruxelles » (DG), « Luxembourg » (Eurostat), « Paris Ministère » et « Paris Insee ». Il y avait alors 19 SSM, celui des « médias » ayant été supprimé en 2008.

25. Plus de 170 « groupes de travail » (où tous les États membres sont représentés) et « task force » (constituées sur une base volontaire) ont été recensés en 2008.

tuellement les politiques nationales coordonnées par la Méthode ouverte de coordination (MOC).

La commission, avec éventuellement un groupe de travail statistique *ad hoc*, est amenée à définir les *indicateurs* quantitatifs nécessaires à ces évaluations et à ces comparaisons (*benchmarking*). Ces travaux de définition, de normalisation et d'homogénéisation ne manquent pas d'avoir des effets en retour, d'une part sur les indicateurs que chaque ministre demande à son SSM de concevoir et de chiffrer pour éclairer sa politique nationale, d'autre part sur les demandes statistiques adressées par la DG à Eurostat et, par l'intermédiaire de ce dernier, à la statistique européenne.

Des SSM plus ou moins impliqués

Quelques SSM ont des relations intenses avec la statistique européenne, soit parce que leurs départements ministériels sont associés à des politiques communes fondamentales, ce qui rend leurs statistiques particulièrement stratégiques (statistiques conjoncturelles pour le Sessi, statistiques des produits agricoles pour le ministère de l'agriculture), soit parce qu'ils sont impliqués dans des stratégies globales (politiques nationales coordonnées sur une base volontaire par la MOC) politiquement prioritaires, comme celles de l'emploi (Dares), de la lutte contre la pauvreté (Drees) ou de la stratégie de Lisbonne (éducation, recherche, emploi, industrie).

Le cas le plus extrême est sans doute celui de l'agriculture (Scees). Du fait d'une des politiques communes les plus substantielles du point de vue budgétaire et les plus anciennes historiquement – la « politique agricole commune » (PAC) –, ce service statistique possède son propre « comité des programmes statistiques », ancêtre sectoriel, en quelque sorte, du Comité des programmes statistiques (CPS) d'Eurostat, qui lui, est plus récent, mais à compétence universelle (en dehors de l'agriculture).

Le service statistique de la pêche (pêche et aquaculture plus précisément) est également étroitement

Tableau 1 : typologie des services statistiques de ministère selon leurs relations avec Eurostat

SSM « importants » avec des relations intenses	SSM « petits ou moyens » avec des relations moyennement intenses	SSM « petits » avec des relations faibles ou nulles
Agriculture (Scees) Douanes (DSED) Éducation (SDES) Emploi et travail (Dares) Équipement (SESP) Industrie (Sessi) Santé et social (Drees)	Énergie (OEEMP) Environnement (IFEN) Justice (SDSED) Pêche (BCS) Tourisme (DSPES)	Coll. Loc. (DESL) Culture (DEPS) Défense (OED) Fonction Publique (DGAFP) PME (DCASPL) Sports (MEOS)

Remarque : les sigles et acronymes sont ceux qui avaient cours en septembre 2007.

impliqué dans la politique des quotas, politique analogue à celle de la PAC. Du fait de sa taille, plus réduite, et du caractère impératif des délais très courts à respecter, il intervient moins que le service statistique de l'agriculture à Bruxelles, et encore moins en amont, de telle sorte qu'il possède très peu de degrés de liberté dans la définition des statistiques à communiquer à Bruxelles.

Bien qu'il soit important par la taille, le service statistique des douanes est intégré à un dispositif qui lui laisse peu de souplesse en termes de contraintes temporelles : il a à fournir les données conjoncturelles sur le commerce extérieur dans le champ qui est le sien, circonscrit à la zone euro, à l'autre grand prescripteur statistique de l'Union européenne, la Banque centrale européenne (BCE).

La *production* statistique de ces trois services statistiques (agriculture, douanes, pêche) est donc contrainte par les impératifs de rapidité (statistiques conjoncturelles) et par leur intégration dans des politiques communes fortes de l'Union. Cela atténue le caractère déterminant du CPS d'Eurostat, dans ces champs particuliers où les DG de Bruxelles sont très influentes et où leurs crédits d'études importants permettent de financer éventuellement des enquêtes statistiques, soit directement, soit par l'intermédiaire d'Eurostat.

Le classement du SSM de l'éducation nationale parmi ceux qui sont « impor-

tants et ayant des relations intenses avec la statistique européenne » peut surprendre. En effet, conformément au principe de subsidiarité, les politiques d'éducation pour le primaire et le secondaire sont nationales. En revanche, les bureaux de ce service qui s'occupent des enquêtes sur la recherche, l'innovation, l'enseignement supérieur et la formation professionnelle²⁶, en collaboration, dans ce dernier cas, avec la Dares, sont en relation avec les commissions spécialisées à Bruxelles et les groupes de travail à Luxembourg.

Les liens entre la formation initiale, la formation continue et l'emploi conduisent de plus en plus le SSM « Éducation nationale », en collaboration avec d'autres SSM et en particulier la Dares, à s'impliquer dans la définition des indicateurs statistiques nécessaires aux stratégies pour l'emploi et à la stratégie de Lisbonne dans un cadre international où l'Unesco et l'OCDE jouent depuis longtemps un rôle de standardisation des nomenclatures, des concepts et des méthodes (statistiques dites « UEO », comme Unesco, Europe, OCDE). Mais en termes d'effectifs, ces services ne représentent qu'une petite partie du SSM de l'éducation nationale dont, quantitativement, l'essentiel des

26. Ce travail se fait dans le cadre de la politique à « compétences partagées » de la recherche et de la stratégie de Lisbonne, lancée en 2000 et relancée en 2005, qui combine des ambitions en termes de recherche, d'enseignement supérieur, d'emploi et de politique industrielle.

activités statistiques est tourné vers le primaire et le secondaire.

Pour le SSM de l'équipement, il faut distinguer les statistiques des transports qui sont fortement demandées par l'Europe (marché unique), et celles de l'habitat, de l'urbanisme et de la construction qui, pour le moment, ne font pas partie de la statistique européenne.

Les statistiques *conjoncturelles* entrent dans la confection des indicateurs demandés par la BCE, et les contraintes de définition, de rapidité et de qualité sont très fortes. Sinon, l'influence de la statistique européenne, notamment quand il s'agit de définir de nouvelles données structurelles, est relativement faible.

À l'opposé des grands SSM impliqués, on trouve les six SSM de petite taille qui ne sont pratiquement pas impliqués dans la statistique européenne, soit parce qu'il n'y a pas eu de transferts de souveraineté dans le domaine politique du département ministériel qui les abrite, comme les domaines de la défense (OED), de l'intérieur (collectivités territoriales) et de la fonction publique (DGAFP), soit parce que les politiques correspondantes ne font pas partie, pour le moment, des priorités des politiques de coordination par la MOC, et donc des priorités du programme statistique communautaire quinquennal (c'est le cas de la culture, des petites et moyennes entreprises, et des sports).

Enfin, la classe intermédiaire, très hétérogène suivant le critère de la taille, est constituée de « services statistiques petits et moyens qui ont des relations internationales relativement intenses » avec l'Union européenne (DG et Eurostat), mais aussi, compte tenu de leur champ relativement étroit de compétence, avec des organisations internationales spécialisées dans les études relatives à un secteur déterminé : l'Organisation mondiale du tourisme (OMT), l'Agence internationale de l'énergie (AIE), l'Agence européenne de l'environnement (AEE), la Politique européenne de sécurité commune (PESC), l'Organisation du traité de l'Atlantique Nord (OTAN), etc.

Les interactions entre la statistique européenne et la « mondialisation » de la statistique

Le formalisme du « quadrilatère » (Bruxelles, Luxembourg, Paris-Ministère, Paris-Insee) s'avère en réalité rapidement trop simple pour décrire le champ des relations internationales entre les SSM, Eurostat et les autres organisations internationales qui interviennent dans le champ de la statistique publique.

Les relations entre les SSM et l'Union européenne sont dans certain cas régulées par des instances spécifiques. Du côté de la BCE et du Conseil ECOFIN, les décisions du sous-groupe « statistique » ont moins de conséquences directes pour les SSM que pour l'Insee. La BCE a un rôle prescripteur de fait important pour des services statistiques peu nombreux (douanes, agriculture, industrie, travail et emploi) mais importants par leur taille. La création de la zone euro fait jouer un rôle central à un certain nombre de statistiques conjoncturelles. La BCE a profondément modifié la nature du processus de production de la statistique européenne en donnant la priorité aux délais et à la qualité (pertinence et précision). Mais en contrepartie, l'impératif de crédibilité des statistiques auprès des marchés dont elle est la garantie est un puissant moyen de protection de la rigueur et de la transparence statistique, et donc de légitimation du code de bonnes pratiques de la statistique européenne auprès des États membres.

Dans le domaine des statistiques de l'agriculture, le « Comité des programmes statistiques agricoles » (CPSA) de Bruxelles est un « comité des programmes statistiques » spécialisé. D'autres instances ne relèvent pas de la statistique européenne institutionnelle mais sont placées dans les DG correspondantes à Bruxelles des départements ministériels auxquels appartiennent les SSM français. C'est d'abord le « Comité des douanes », dont les décisions ont des retombées statistiques importantes et immédiates pour le SSM des douanes. Les travaux du Comité de

l'emploi à Bruxelles ont une grande influence méthodologique pour la définition des indicateurs utilisés par la Dares. Dans le domaine de la santé (Drees), l'animation du sous-groupe de travail consacré aux indicateurs du Comité de la protection sociale joue également un rôle important en amont de la statistique sanitaire et sociale européenne (système européen des statistiques de protection sociale : SESPROS).

Différentes organisations internationales qui n'appartiennent pas au champ institutionnel de la statistique européenne y jouent néanmoins un rôle très important, notamment pour les méthodes, les nomenclatures et les définitions conceptuelles. Certaines sont spécialisées dans des domaines sectoriels spécifiques : l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), la Politique européenne de sécurité commune (PESC) et l'OTAN pour la défense, l'Organisation mondiale du tourisme (OMT) pour le tourisme, l'Unesco pour l'éducation et la culture, l'Organisation mondiale de la santé (OMS) pour la santé, le Bureau international du travail (BIT) pour le travail les relations sociales, l'Agence internationale de l'énergie (AIE) pour l'énergie et l'Agence Européenne de l'environnement de Copenhague (AEE) pour l'environnement. Chacune de ces institutions a des contacts avec le SSM responsable du domaine correspondant à ses activités.

Il faut considérer à part l'Organisation de coopération et de développement économiques (OCDE). Certes, il s'agit d'une institution généraliste, mais elle a à la fois multiplié des expériences d'évaluation de politiques économiques, sociales et environnementales, et élaboré toute une panoplie de méthodes et d'instruments statistiques généraux « clés en main », à la disposition de l'Union européenne ou des États membres, pour évaluer des politiques, des pays ou des secteurs déterminés. Elle n'intervient évidemment pas directement dans la définition des politiques publiques européennes, mais elle détermine indirectement beaucoup de « standards » de la statistique européenne. Dans la quasi-totalité des secteurs, et plus particulièrement dans ceux



Source : Wikipédia

Bâtiment d'Eurostat à Luxembourg

de l'emploi et de l'environnement, l'OCDE, et la Commission économique pour l'Europe des Nations unies (CEE-NU) à un moindre degré, jouent un rôle pionnier pour la méthodologie de la statistique et de l'évaluation des politiques publiques.

Quelle dynamique en œuvre pour la suite ?

Cette typologie, qui peut paraître statique, permet d'éclairer la dynamique qui sera sans doute à l'œuvre dans les prochaines années. En effet, la demande d'évaluation des politiques, tant au niveau national qu'europpéen, devrait vraisemblablement croître.

Au niveau européen, les besoins statistiques pour l'évaluation des politiques communautaires conduiront à de nouveaux *règlements statistiques* ou des réformes des anciens. D'autre part, la *méthode ouverte de coordination* s'étendra à de nouveaux domaines, comme on peut le

voir dans le programme statistique communautaire quinquennal 2008 – 2012 : innovations, migrations, sécurité, justice, etc.

Mais ce n'est pas pour autant que de nouvelles politiques communes devraient apparaître à l'agenda de la construction européenne, politiques qui auraient un effet aussi structurant pour la statistique européenne, aussi énergique et aussi rapide, que la création de l'euro au début des années 2000.

L'intensification des relations avec les statistiques européennes devrait affecter les SSM « importants » par leur taille, déjà impliqués fortement dans la statistique européenne.

Il semble que l'augmentation de l'influence de la statistique européenne sur les SSM français doive plutôt se faire par extension de la MOC aux domaines qui correspondent aux SSM caractérisés aujourd'hui comme « petits et moyens » et impliqués de manière encore relativement modeste dans la statistique européenne.

Les relations avec cette dernière devraient s'intensifier, notamment dans les domaines de la justice, des migrations internationales et du développement durable (SSM de l'environnement, de l'énergie, des transports, de l'habitat, du tourisme et de la santé). L'implication de tous les SSM (à l'exception de ceux pour lesquels il n'y a pas de transfert de souveraineté) devrait également

augmenter par suite du renforcement des MOC déjà en place pour les stratégies appelées à couvrir plusieurs exercices quinquennaux de la programmation statistique communautaire (stratégie de Lisbonne, stratégie pour l'emploi, lutte contre la pauvreté).

Les effets sur les procédures et sur la nature des chiffres (*statistiques* versus *indicateurs*) seront de nature différente. D'un côté, le pouvoir d'initiative de la Commission, via Eurostat, aura tendance à diminuer les degrés de liberté des systèmes statistiques publics nationaux. D'un autre, pour toutes les politiques qui relèvent de la MOC, les SSM pourront, *via* les départements ministériels où ils se trouvent, participer à la conception des indicateurs que les directions générales de Bruxelles demanderont à la statistique européenne de lui fournir pour faire du *benchmarking*. Si l'originalité de l'organisation « fonctionnelle décentralisée » de la statistique publique française rend plus complexe sa coordination, elle pourrait constituer un avantage comparatif dans la gouvernance de la statistique européenne.

Enfin, il conviendra de prendre en compte cette implication croissante des SSM dans la construction de la statistique européenne, que ce soit dans la fourniture des *statistiques* ou dans la définition des *indicateurs* utilisés pour la MOC, afin de définir le futur programme à moyen terme du Cnis 2009 – 2013. ■



Source : Wikipédia

Les drapeaux des pays de l'Union européenne

La statistique à la direction générale des collectivités locales – ministère de l'Intérieur

Article initialement paru dans le *Courrier des statistiques* n° 5 de janvier 1978, écrit par Jean Bouinot, à l'époque chef de la mission de statistiques et d'analyses financières à la Direction générale des collectivités locales, ministère de l'Intérieur.

Les statistiques sur les collectivités locales, et les travaux d'études qui les accompagnent nécessairement, sont effectués au ministère de l'Intérieur par la Mission de statistiques et d'analyses financières de la Direction générale des collectivités locales¹.

Cette mission, [structure] horizontale rattachée au directeur, a pris la suite, le 28 mai 1974, du Service de statistiques et d'analyses financières, structure verticale, dont la création résultait d'un arrêté du 13 janvier 1967.

Les statistiques ainsi produites présentent deux caractéristiques essentielles : elles sont d'une part à dominante financière – mais ceci n'est qu'un trait provisoire, les développements attendus dans l'avenir devant déboucher sur une prise en compte croissante des aspects socio-économiques des institutions locales ; elles sont d'autre part orientées vers l'aide à la gestion, d'où le rattachement des activités de RCB² municipales à la mission de statistique.

Pour des activités très variées, un système de gestion par projet

- Les différents travaux qui incombent à la Mission peuvent être regroupés autour de trois axes. Il s'agit respectivement :

en pourcentage colonne

secteurs d'activité	travaux en régie (1)			sous-traitance (2)		
	1975	1976	1977 (3)	1975	1976	1977 (3)
statistiques	51	36	35	36	20	20
« management »	6	23	22	23	34	46
études générales	9	4	8	41	46	34
activités diverses et non ventilables	34	37	35	–	–	–
total	100	100	100	100	100	100

(1) Évaluation en temps de travail
 (2) Évaluation en pourcentage des crédits alloués
 (3) Chiffres prévisionnels (résultats définitifs janvier 1978)

- des travaux de statistiques, *stricto sensu*, de la production à la diffusion ;

- des études aboutissant à la création d'instruments modernes de gestion (secteur « management ») ;

- des études générales dont les analyses financières.

À cet ensemble, s'ajoutent des travaux divers d'organisation générale (documentation, gestion de crédits, etc.). Le tableau ci-après indique la part respective des différents types d'activités et donne une idée de leur évolution au cours des dernières années.

- Les ressources humaines de la Mission s'établissent actuellement à seize personnes, dont quatorze cadres A. Parmi ces derniers, figurent

cinq attachés de l'Insee. Rappelons, à titre de comparaison, que l'effectif, à la création de la Mission en mai 1974, était seulement de six personnes.

L'importance du nombre de cadres A peut surprendre pour un service ayant en charge des productions statistiques. L'explication est simple : les tâches habituellement imparties aux autres catégories [d'agents] sont pour l'essentiel sous-traitées.

Plus généralement, les crédits d'études dont on a donné la répartition par secteur d'activité constituent un puissant coefficient multiplicateur

1. Les collectivités locales assurent la maîtrise d'ouvrage de plus de 50 % des équipements publics. Ce seul chiffre suffit à situer l'importance des institutions locales dans la vie nationale.
 2. Rationalisation des choix budgétaires [NDLR].

d'activité permettant, à peu près, de quadrupler la capacité interne de production.

- La recherche de la souplesse dans le fonctionnement interne de la Mission a conduit à mettre en place un système de gestion par projet : le projet étant l'unique unité de responsabilité, substituée à la formule habituelle de l'organigramme hiérarchique.

En bref, à la fin de l'exercice N, le personnel se répartit entre les différents projets à réaliser pendant l'exercice N+1 et détermine pour chacun d'entre eux une date-objectif d'achèvement. Le point est effectué à la fin de chaque trimestre.

L'ensemble du dispositif planification-contrôle de gestion repose sur un système de comptabilité-temps individuelle consistant à répertorier, chaque fin de semaine, les temps passés par projet.

La rénovation de l'appareil statistique

- La revitalisation du dispositif statistique sur les collectivités locales a été recherchée, dès la création de la Mission, comme la priorité des priorités, en retenant comme objectif central celui de l'utilité maximale des documents produits. En termes opérationnels, l'objectif d'utilité a été traduit en trois sous-objectifs :

- l'adaptation du contenu des principales statistiques aux besoins à partir d'enquêtes préalables auprès des usagers potentiels ;

- la diversification des champs d'observation ;

- la réduction des délais de mise à disposition des résultats statistiques.

- Ainsi, la reconstruction des statistiques sur les communes de plus de 10 000 habitants a démarré en

mai 1975 par une opération préliminaire de consultation des principales catégories d'usagers autour d'un dossier de dialogue. Présentement, les résultats sortent 17 mois après la clôture de l'exercice sous la forme de guides des ratios et sont disponibles en collection à la Documentation française.

- La rénovation des statistiques sur les finances départementales a suivi un schéma identique de restructuration, le délai de mise à disposition dans la collection précitée étant de 11 mois après la clôture de l'exercice.

- Dans les communes de moins de 10 000 habitants, le processus antérieur d'élaboration a été provisoirement conservé ; le guide des ratios donne, comme par le passé, des indicateurs par strates de taille démographique croisées avec les régions. L'ambition pour le futur serait de présenter les résultats par groupes socio-économiques homogènes de communes : un premier pas dans cette direction sera franchi au début de 1978 avec l'édition d'une statistique financière spécifique au groupe des communes touristiques.

- Dans ce survol, la statistique sur les voiries locales mérite aussi un bref signalement individuel car elle déborde nettement la seule collecte de données financières pour les rapprocher de données physiques (longueurs classées, état de viabilité, trafic, etc.).

- Les autres travaux statistiques (fichier sur la fiscalité directe de l'ensemble des collectivités locales, emprunts auprès d'organismes privés, etc.) sont, dans l'ensemble, plus légers à réaliser que ceux qui viennent d'être cités. À cette règle générale, une exception : l'enquête sur le personnel communal (situation au 1^{er} janvier 1977), dont les premiers résultats seront connus au début de l'année 1978.

- Enfin, mention doit être faite d'un travail hybride entre le secteur statistique et les autres secteurs : il s'agit du fichier « coûts des équipements

collectifs » réalisé en collaboration avec les services du ministre de l'Équipement et les différentes institutions intéressées. Les résultats obtenus sont diffusés notamment sous la forme de dossiers spécialisés d'équipement ; à ce jour, dix d'entre eux sont parus à la Documentation française dans la collection « planification urbaine » (en fascicules séparés : maternelles, crèches, etc.).

Le développement des études générales et des analyses financières

Ces études sont effectuées fréquemment en liaison avec d'autres administrations, parmi lesquelles on peut citer :

- le Commissariat général du Plan, notamment pour les travaux du comité de financement ;

- la DATAR³, pour les études de finances locales nécessitées par les grandes opérations d'aménagement ;

- le Conseil de l'Europe, du fait d'une participation constante aux travaux du Comité de coopération pour les affaires municipales et régionales.

- À la rubrique des études générales, appartiennent :

- les diagnostics financiers de villes et de départements, tous travaux qui justifient la « raison sociale » de la Mission ;

- les études en cours, d'une part en matière de compte prospectif, d'autre part en matière d'analyse rétrospective des finances de grandes villes sur la période 1965-1974 (analyses en composantes principales, analyses de correspondances, etc.).

- Quelle que soit leur nature, toutes ces études prennent très directement appui sur l'appareil statistique en place et leurs développements respectifs, sur le passé, se sont avérés effectivement corrélés ; dire que la réponse aux demandes a crû en liaison étroite avec l'enrichissement des données quantitatives disponibles n'est en définitive qu'évidence.

3. Délégation à l'aménagement du territoire et à l'action régionale, devenue depuis la Diact (Délégation interministérielle à l'aménagement et à la compétitivité des territoires) [NDLR].

L'ouverture d'un secteur d'aide à la gestion des collectivités locales

▪ La finalité des efforts entrepris depuis 1974 se résume à un objectif clef : aider les collectivités locales à réaliser des gains de productivité et de qualité dans la gestion de leurs services.

La notion de productivité est prise ici dans son sens habituel : pour sa part, le terme de qualité d'un service public désigne son adaptation aux attentes des usagers. Il s'agit, en quelque sorte, d'une rencontre réussie entre une offre et une demande de service public.

▪ Les efforts entrepris, et dont le bilan a été présenté dans une plaquette spéciale établie en août 1977⁴, s'ordonnent autour de trois champs d'investigation :

– le domaine de la politique générale d'une municipalité (politique urbaine, politique du personnel, politique financière) au service de laquelle ont été conçus et testés différents outils de gestion : modèle financier prévisionnel, consolidation des comptes, planification de la trésorerie, etc ;

– la gestion des services communs (ateliers, magasins...) pour lesquels les premières réalisations sont en voie d'achèvement ;

– la gestion des services directs (voirie, transports collectifs, équipements de superstructures de tous ordres...) qui a déjà donné lieu à plusieurs opérations pilotes importantes.

▪ Tous les travaux précités sont conduits en liaison étroite et constante avec les responsables communaux ; leurs résultats sont diffusés en faisant appel aux supports documentaires classiques, mais la faveur se porte désormais sur les techniques vivantes d'information et de formation : participation à des colloques, organisation de journées d'études et, depuis le début de 1977, montage et animation de sessions de 3 à 5 jours de formation permanente.

Conclusion : trois axes de développement pour le futur

À court terme, tout d'abord, l'accent se portera sur l'essor des statistiques financières conjoncturelles. Si leur production atteint un rythme de croisière pour les départements, elle se réduit en revanche à un essai expérimental sur les communes et ce

dernier appelle un effort sérieux de consolidation.

En second lieu, se dessinent sur l'avenir des perspectives de diversification et d'affinement de l'appareil statistique en place par le jeu d'une articulation progressive avec certains outils de gestion implantés sur le terrain. Ainsi, la multiplication actuelle de bases de données patrimoniales⁵ autorisera, à terme rapproché, l'établissement consécutif de statistiques détaillées sur les coûts des équipements collectifs.

Dans cette dialectique d'enrichissement réciproque « du centre et de la périphérie », les services statistiques de villes – dont on souhaite la création rapide – formeront le pivot essentiel ; ils joueront de plus le rôle d'observatoire du changement socio-économique de la vie locale, ouvrant ainsi la voie à un troisième axe de développement. ■

4. Cette plaquette est disponible à la mission de statistiques et d'analyses financières – Direction générale des collectivités locales – 4 à 12, rue d'Aguesseau – 75008 Paris.

5. Une dizaine de communes mettent en place des bases de données patrimoniales, c'est-à-dire des fichiers décrivant les caractéristiques des équipements collectifs qu'elles gèrent, ainsi que leurs coûts de fonctionnement.

Le SSM collectivités locales : interview de Jean-Luc Heller

Courrier des statistiques : Jean-Luc Heller, vous êtes le chef du service statistique ministériel (SSM) « collectivités locales », le DESL¹. En quoi consistent les missions de ce service et quels en sont les travaux phares ?

Jean-Luc Heller : Les missions du service sont très classiquement celles d'un SSM : recueillir, contrôler, traiter, analyser et diffuser l'information statistique dans le domaine de son département ministériel, ici celui chargé des collectivités locales.

À cet effet, le SSM bénéficie pleinement de son insertion dans la Direction générale des collectivités locales (DGCL), à parité avec les sous-directions qui ont en charge les finances, les compétences ou les statuts des personnels des collectivités. Cela permet au SSM d'être associé aux problématiques en cours et d'utiliser les sources administratives disponibles, au sein du ministère ou par le réseau des préfectures. Cette proximité de la décision, qui fait l'intérêt du travail en SSM, est aussi source de tension, car le temps du statisticien et des investissements dans la production statistique n'est pas spontanément celui de l'action d'une administration centrale comme celle du ministère de l'Intérieur.

Ce qui est moins classique, c'est à la fois la grande hétérogénéité et le développement rapide des collectivités locales à la suite des vagues de décentralisation intervenues depuis le début des années 1980. À l'intervalle

de trente ans, que l'article reproduit ci-avant nous invite à parcourir, il n'est pas exagéré de parler de bouleversement.



Jean-Luc Heller

Source : Courrier des statistiques

Les régions sont apparues comme collectivités locales de plein exercice au côté des départements. Les regroupements intercommunaux – communautés d'agglomération, communautés urbaines, communautés de communes – couvrent la quasi-totalité du territoire. Construction et entretien des lycées et des collèges, formation professionnelle, routes, transports ferroviaires de voyageurs, action sociale, versement du revenu minimum d'insertion (RMI) et de l'allocation personnalisée d'autonomie, autant de domaines qui sont passés de la compétence de l'État à celle des collectivités territoriales, si bien que, désormais, plus de 72 % de

l'investissement public est le fait des administrations publiques locales.

De nombreuses réformes fiscales et financières ont accompagné ce mouvement. La fonction publique territoriale représente maintenant plus de 1,6 millions d'agents, soit près d'un tiers de l'ensemble des trois fonctions publiques (d'État, territoriale et hospitalière).

Le service statistique mène son activité pour l'essentiel à partir de la mobilisation de sources administratives, donc de données issues de la gestion ou de l'activité des collectivités : tenue des comptes par le comptable public, déclarations de fiscalité, transmission des budgets et des comptes pour le contrôle de légalité, déclarations de données sociales, volet statistique des rapports aux comités techniques paritaires, suivi de l'évolution des structures pour les besoins de répartition des dotations de l'État.

Cette situation a des forces bien connues (exhaustivité, coût de collecte marginal) mais aussi des limites (délais et pertinence des données pour l'exploitation statistique). Il est d'ailleurs proposé de compléter cet ensemble par un dispositif d'enquête par internet sur un échantillon de collectivités locales. Cette méthode serait en effet mieux adaptée à une collecte

1. L'intitulé exact de ce service statistique ministériel est le Département des études et des statistiques locales (DESL), au sein de la Direction générale des collectivités locales (DGCL) du ministère de l'Intérieur, de l'Outre-Mer et des Collectivités territoriales.

rapide d'informations, d'autant plus si l'on souhaite que ces informations soient à la fois plus ciblées et plus évolutives que celles des sources administratives traditionnelles.

Les travaux du SSM s'organisent autour de trois missions : observation des finances locales, de la fonction publique territoriale et des structures territoriales (et parmi ces dernières, les structures intercommunales en particulier).

Les travaux sur les finances sont marqués par le rythme annuel du vote par les collectivités locales, des budgets primitifs et des comptes administratifs, une fois la gestion terminée. Pour chaque type de collectivité (régions, départements, communes et intercommunalités), il s'agit de recueillir les principales données financières et d'en tirer une agrégation et une analyse nationale.

Le nouveau couple commune – intercommunalité pose des problèmes méthodologiques et pratiques pour parvenir à construire une vision du secteur communal consolidée, c'est-à-dire sans double compte. En plus de ces résultats agrégés par type de collectivité, sont aussi diffusées des données financières pour chacune des plus grandes collectivités.

La fiscalité locale fait l'objet d'un traitement particulier. Dès le vote des taux de fiscalité, les données sont recueillies, directement ou par l'intermédiaire des services fiscaux. Ceci permet de disposer de résultats sur les taux et produits de la fiscalité locale directe, au cours même de l'année étudiée. C'est naturellement un domaine particulièrement sensible.

2. Bureau des statistiques, des études et de l'évaluation, Direction générale de l'administration et de la fonction publique, ministère du Budget, des Comptes publics et de la Fonction publique.

3. Le SSM du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche est un sous-ensemble de la Direction de l'évaluation, de la prospective et de la performance (DEPP).

4. Direction de la recherche, des études, de l'évaluation et des statistiques, actuellement rattachée à trois ministères : Travail, Relations sociales et Solidarité ; Santé, Jeunesse et Sports ; Budget, Comptes publics et Fonction publique.

En matière de fonction publique territoriale, notre deuxième mission d'observation, le SSM conduit l'exploitation statistique issue des rapports (bilans sociaux) que la loi impose aux employeurs locaux de fournir aux comités techniques paritaires. Cette exploitation a lieu tous les deux ans sur la situation au 31 décembre des années impaires. C'est un dispositif lourd (puisqu'on vise l'exhaustivité), mais en progrès constant dans sa qualité et ses délais. Et surtout, c'est la seule source disponible sur les emplois par filière détaillée (jusqu'au niveau du grade), sur la formation, les droits sociaux, le temps de travail ou encore les conditions de travail.

Enfin, les structures territoriales font l'objet d'un suivi en continu, puisque le périmètre des groupements de communes à fiscalité propre – ils sont désormais plus de 2 500 – est publié par le DESL courant janvier, pour la situation au 1^{er} janvier de la même année.

Le site internet ouvert récemment par notre service statistique (BASE NATIONALE sur l'InterCommunalité, Banatic : www.banatic.interieur.gouv.fr) constitue un apport décisif dans la mise à disposition de données exhaustives s'étendant à l'ensemble des formes de regroupements intercommunaux (groupements à fiscalité propre et syndicats), données mises à jour trimestriellement. De plus, Banatic met à disposition des informations sur tous les groupements de collectivités territoriales, notamment leur périmètre, leur mode d'organisation et de financement ainsi que leurs compétences.

Courrier des statistiques : Parmi ces travaux, lesquels sont réalisés en liaison avec d'autres administrations (Insee, autres SSM, etc.) ?

Jean-Luc Heller : Les premiers partenaires sont nos collègues de Bercy qui s'occupent de la gestion des comptes des collectivités (la Direction générale de la comptabilité publique, DGCP) et du suivi de leur fiscalité (la Direction générale des impôts, DGI), appuyés par leurs réseaux de comptables et de fiscalistes sur le territoire.

Par leur activité administrative, ils disposent de sources riches et exhaustives sur les comptes des collectivités qui nous sont transmises pour exploitation à des fins statistiques.

Alors que la restitution individuelle en ligne est le fait du site Bercy Collectivités locales (www.colloc.bercy.gouv.fr), le SSM de la DGCL cherche à développer les synthèses et les analyses, le calcul d'agrégats et de ratios selon des strates significatives pour la comparaison des situations des collectivités.

Des liens étroits sont également tissés avec le SSM « fonction publique »², car lorsqu'il s'agit de rassembler des données sur l'emploi, les rémunérations ou les conditions de travail des agents des trois fonctions publiques, la coordination est de mise.

Les collectivités locales prennent un part croissante dans les politiques publiques, en matière d'éducation et de formation ou d'action sociale par exemple. Nous sommes en contact avec le SSM « éducation »³ qui établit le compte de l'Éducation pour estimer la contribution des collectivités locales.

Les statistiques sur les politiques sociales sont traitées quant à elles par le SSM « santé, solidarité », la Drees⁴, parfois par des enquêtes directes auprès des départements et nous devons alors souvent « croiser » nos informations. Par exemple, durant les premières années de transfert des dépenses du RMI aux départements, nous avons dû travailler à partir de données des comptables du Trésor, des enquêtes de la Drees et des comptes administratifs transmis aux préfetures après leur vote par les assemblées départementales pour établir les montants en cause pouvant donner lieu à compensation.

Il faudrait citer aussi les transports, le tourisme, la culture ou le sport, autant de domaines où les collectivités locales sont actives. En pratique, le SSM de la DGCL traite plutôt les questions de la dépense publique locale d'un point de vue d'ensemble et de celui des grands équilibres

et des grandes masses, réservant les investigations plus approfondies aux SSM qui disposent de plus de moyens et d'informations pour maîtriser les problématiques propres à leur domaine. L'important est de veiller à la cohérence des approches et à ne pas alourdir la charge de réponse des collectivités.

L'Insee est naturellement un de nos partenaires essentiels, et ce, dans plusieurs domaines. Tout d'abord, les collectivités locales sont immatriculées au répertoire Sirene. L'opération, menée en commun, de mise en cohérence de l'immatriculation des nombreux syndicats intercommunaux, entre les bases des préfectures et celles de l'Insee, a constitué une véritable enquête de balayage pour cette partie de Sirene dite « Secteur public ». Le suivi des modifications de la géographie des communes (ou d'autres circonscriptions, cantons, arrondissements), tel que l'Insee l'établit par le Code officiel géographique, suppose donc des échanges constants entre le DESL et l'Insee à propos des textes qui sont à la source des changements de superficie et de population, des fusions, des éclatements et des changements de nom des communes.

Dans le domaine de l'emploi public ou des rémunérations des fonctionnaires territoriaux, l'Insee nous procure les sources de référence. Des conventions sont passées pour permettre au SSM d'utiliser l'enquête Insee dite *Colter* (pour COLlectivités TERritoriales), dont le questionnaire fait l'objet d'une mise à jour annuelle concertée. Le système d'information sur les agents du secteur public est un grand projet qui doit très prochainement déboucher sur une mise en œuvre opérationnelle pour tous les partenaires, à partir de l'exploitation des fichiers administratifs. Le bon usage de la nomenclature des emplois territoriaux (NET) dont est responsable la DGCL dans les déclarations annuelles de données sociales (DADS), à la base de ce système d'information pour la fonction publique territoriale, était un pré requis ; la sensibilisation des acteurs a été conduite en commun. Dès mainte-

nant, les études menées à l'Insee à partir de la source DADS sur la fonction publique territoriale font partie des références de base.

Il n'est pas possible d'oublier les questions liées au recensement de population. À la suite de la rénovation du recensement de la population par l'Insee, la prise en compte, à compter de 2009, de l'actualisation des chiffres de population de toutes les circonscriptions administratives, à un rythme annuel, va intervenir dans les données statistiques comme dans les nombreuses dispositions faisant référence à la population. Les communes qui préparent et réalisent les opérations de recensement en sont les premières utilisatrices. Le service statistique est placé à l'interface des nombreuses questions que les communes posent à ce sujet.

Courrier des statistiques : Dans son article de 1978, Jean Bouinot expliquait que son service produisait principalement des « statistiques financières à finalité de gestion », mais qu'il était prévu qu'à l'avenir, les statistiques socioéconomiques se développent fortement... Trente ans après, diriez-vous que les choses ont évolué ainsi ?

Jean-Luc Heller : La place des statistiques financières reste centrale pour le DESL, mais leur finalité va au-delà de l'aide à la bonne gestion de chaque collectivité, apportée par des guides de ratios financiers ou de références moyens. L'appréciation d'ensemble de la situation financière et des grands équilibres comptables des collectivités est la première demande qui nous est adressée : évolution de la dépense de fonctionnement, dynamique de l'investissement, importance du recours à l'emprunt et charge de la dette.

L'attention et les demandes issues de débats parfois vifs entre les acteurs se portent particulièrement sur les domaines dont la compétence a été transférée de l'État aux collectivités locales. La compensation est-elle équitable devant l'évolution des charges ? Quelles sont les conséquences de la décentralisation dans la dyna-

mique des dépenses ? Une gestion plus proche est-elle génératrice de meilleur emploi des fonds publics ou, à l'inverse, plus sensible à la proximité de la demande ?

L'impact des réformes institutionnelles est aussi questionné. Avec l'apparition des régions dans l'intervention économique, l'association fréquente de plusieurs niveaux de collectivités dans des projets communs d'investissements ou de développement est-elle un gage de mutualisation efficace de moyens ou à l'opposé génératrice de coûts de transactions importants ? L'évolution des dépenses de l'intercommunalité comparée à celle des seules communes permet-elle de conclure à des économies d'échelle ou, au contraire, traduit-elle les effets d'un doublonnage dans les structures ? Autant de questions aux enjeux très sensibles, posées par le débat public, auxquelles le DESL tente de répondre en mobilisant les statistiques financières.

L'autre approche, où statistiques financières et préoccupations socioéconomiques se rejoignent, est celle des domaines de l'action publique locale. Combien les collectivités locales dépensent-elles pour l'action sociale, l'enseignement, la formation, les transports, le sport, la culture, le tourisme ? Les nomenclatures fonctionnelles et la manière dont les utilisent les collectivités ne permettent pas toujours de répondre très précisément à ces problématiques. Encore moins quand il s'agit d'en évaluer l'efficacité ou l'efficience à l'aide d'indicateurs de résultats ou de 'produit'.

Mais ce faisant, il faut faire appel à de nombreuses données, disponibles à l'échelon territorial pertinent. Ce besoin de statistiques détaillées à des échelons territoriaux fins, allant parfois jusqu'aux quartiers, est bien présent et c'est sans doute ce que Jean Bouinot anticipait. Le recours aux sources administratives contenant des données localisables est l'outil principal – en dehors du recensement – pour disposer de données statistiques locales. Mais la production largement répandue de résultats statistiques aux niveaux infra-nationaux n'est pas le

fait du service statistique des collectivités locales : elle a heureusement été prise en charge par l'ensemble du système statistique public.

L'autre élément d'évolution majeur sur ces trente ans est bien sûr celui de l'informatique, qui apporte de nouvelles possibilités, tout particulièrement s'agissant de données et d'utilisateurs très dispersés. La dématérialisation des échanges de données, bien qu'à parfaire dans le domaine des finances locales, permet les traitements et les croisements de masses d'informations. Pour les données locales, la diffusion par internet, avec la souplesse de l'interrogation et les progrès de la cartographie dynamique, sont à la base de développements qu'il était difficile d'imaginer.

Courrier des statistiques : *À la fin de son article d'il y a trente ans, l'auteur exprimait le vœu que soient rapidement créés des services statistiques dans les [grandes] villes, services qui joueraient un rôle « d'observatoire du*

changement socioéconomique de la vie locale ». *Ce vœu a-t-il été exaucé, sous cette forme ou sous une autre, et si oui, quelles conséquences cela a-t-il sur les missions et les travaux de votre service ?*

Jean-Luc Heller : Plus que les seules grandes villes, ce sont maintenant toutes les grandes collectivités territoriales, les régions, les départements, les communautés urbaines ou les grandes communautés d'agglomération qui structurent le territoire.

Dans leur mise en œuvre des politiques publiques locales, elles sont à la fois productrices de données, qui peuvent être des sources pour les statistiques, et demandeuses d'indicateurs statistiques et de données comparables, utiles au pilotage et à l'évaluation. L'enjeu est le positionnement de la statistique publique dans le contexte issu de la décentralisation. Avec l'éclatement des systèmes d'information et de gestion, il faut préserver la comparabilité des statis-

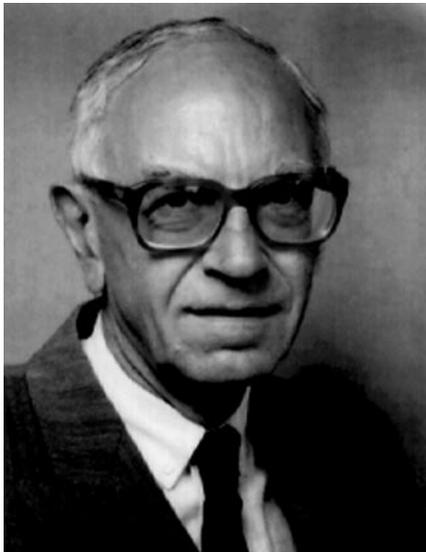
tiques, ce qui passe par un partage des données mais aussi par des normes, des concepts, des méthodes et nomenclatures, au moins harmonisées, au mieux communes.

C'est là que les statisticiens doivent intervenir. La statistique publique est déjà présente sur le territoire grâce au réseau de l'action régionale des directions régionales de l'Insee et à celui des statisticiens en poste dans les services déconcentrés de l'État. Elle doit se positionner vis-à-vis de ces autres collectivités publiques, actives dans de très nombreux domaines, que sont maintenant les collectivités locales. Le Conseil national de l'information statistique a débattu de ces questions à plusieurs reprises et certains verraient bien une mise à disposition de statisticiens de l'Insee dans des services statistiques au sein des collectivités locales. Des SSCL en complément des SSM ? Rendez-vous dans trente ans pour un prochain bilan... ■

Raymond Lévy-Bruhl (1922-2008) : un statisticien innovateur et un serviteur de l'État

 Alain Desrosières et Béatrice Touchelay*

Raymond Lévy-Bruhl, statisticien et ancien Secrétaire général de l'Insee, est mort le 19 février 2008. Né en 1922, il a subi les conséquences des mesures antisémites de Vichy. Il a participé à la Résistance et aux combats de la Libération. Entré à l'Insee en 1946, il est envoyé aux États-Unis où il étudie les nouvelles méthodes d'enquêtes par sondage. Il les développe ensuite en France avec Pierre Thionet, en concevant et lançant en 1950 la première enquête sur l'emploi. Responsable du service statistique du ministère du Travail de 1951 à 1961, il organise les enquêtes sur l'activité et les conditions d'emploi de la main d'œuvre (Acemo). Revenu à l'Insee en 1961, il est successivement directeur de la coordination générale, directeur de cabinet du Directeur général et enfin, Secrétaire général de 1972 à sa retraite, en 1988. Il participe ensuite, avec Jean-Pierre Azéma et Béatrice Touchelay, aux recherches sur le rôle du Service national de statistique (SNS) à l'époque de l'Occupation, entre 1941 et 1944.



Source : droits réservés

Raymond Lévy-Bruhl

Raymond Lévy-Bruhl, Secrétaire général de l'Insee de 1972 à 1988, est mort le 19 février 2008. Avant d'exercer ces fonctions administratives importantes, il avait joué un rôle clé, d'une part, en important en France, dans les années 1940, les méthodes d'enquêtes par sondage aléatoire, puis, dans les années 1950, en développant le service statistique du ministère du Travail. Cet article vise à rappeler son rôle dans cette période fondatrice du système statistique public français.

Il est issu d'une famille de grands scientifiques. Son grand-père, Lucien Lévy-Bruhl (1857-1939) a été, avec Émile Durkheim, un des fondateurs

de la sociologie française. Son oncle, Henri Lévy-Bruhl (1884-1964) juriste et sociologue du droit, avait milité pour le développement des statistiques criminelles.

Né en 1922, Raymond Lévy-Bruhl fait ses études au lycée Pasteur à Neuilly, et il fréquente régulièrement son grand-père. Intéressé par l'histoire, il a envie d'entrer en hypokhâgne (première année de classes préparatoires littéraires), mais la guerre puis la débâcle de 1940 bouleversent ses projets¹. Il souhaite alors entrer dans une hypotaube (première année de classes préparatoires scientifiques), mais le statut des juifs d'octobre 1940 le contraint à renoncer aux grandes écoles et aux carrières administratives. Sa famille se réfugie à Arreau, dans les Hautes-Pyrénées, jusqu'en 1942. Il parvient cependant à suivre des cours de droit, de mathématiques et de chimie à Toulouse et obtient les licences de ces disciplines. Après l'invasion de la zone Sud en novembre 1942, il se procure des faux papiers qui lui permettent, d'une part, d'échapper au STO², d'autre part, de changer de patronyme. Il obtiendra, après la guerre, le statut de réfractaire au STO en produisant des procès-verbaux des recherches menées pour le retrouver par la gendarmerie d'Arreau. Il est engagé en 1943 auprès de l'Ingénieur en chef des Ponts et Chaussées du Gers. Ses fonctions lui permettent de transmettre aux Alliés, par l'intermédiaire des organismes de

résistance, des informations sur les ponts du département et de délivrer, à l'insu de son administration, un nombre important de permis de circulation à la Résistance.

Le jour du débarquement (6 juin 1944), il rejoint le maquis au côté de l'armée secrète et s'engage jusqu'en septembre 1944 dans les FFI³. Il rentre à Paris où il réussit à se faire incorporer dans la 1^{re} Division française libre sur le front d'Alsace, puis participe à la campagne des Alpes, dans la région de Sospel. Il est démobilisé le 5 novembre 1945. Il obtiendra la croix du combattant volontaire.

Il reprend alors ses études et obtient, en 1946, trois diplômes d'études supérieures, en droit, en économie politique et en sciences économiques. Il s'intéresse alors à la statistique et, sur les conseils d'André Mayer, un professeur du collège de France, il entre en contact avec Eugène Morice, directeur de l'école de l'Insee (la future Ensae). Celui-ci lui indique qu'il

* Alain Desrosières est administrateur de l'Insee et membre du Centre Alexandre Koyré d'histoire des sciences. Il a publié « La politique des grands nombres. Histoire de la raison statistique », La Découverte / Poche, 2000. Béatrice Touchelay est maître de conférences en histoire contemporaine à l'université de Paris XII-Créteil.

1. Nous remercions vivement la famille de Raymond Lévy-Bruhl de nous avoir fourni nombre d'informations sur sa jeunesse et sur son activité pendant la guerre et l'Occupation.

2. Service du travail obligatoire.

3. Forces françaises de l'intérieur.

peut se présenter à un concours lui permettant d'entrer sur titre à l'Insee, comme administrateur. L'Institut vient tout juste d'être créé, en avril 1946, à partir du Service national de la statistique (SNS) de René Carmille (créé par ce dernier en 1941). Obtenant une bourse d'études Rockefeller pour aller aux États-Unis, Raymond Lévy-Bruhl est autorisé par Francis-Louis Closon, le tout nouveau directeur de l'Insee, à ne pas suivre les cours de l'école et à partir en Amérique. Ce séjour de quelques mois, en 1946-1947, sera très profitable pour l'Institut.

Lévy-Bruhl passe tout d'abord six mois à l'Université de Columbia, à New York, à étudier les statistiques mathématiques. Puis, à la demande de Closon, il va visiter les services statistiques américains. Il passe un mois au *Bureau of Labor Statistics*, puis deux mois au *Census Bureau*, le service où les méthodes de sondages aléatoires ont été mises au point dans les années 1930, notamment pour établir des statistiques sur l'emploi. Il y fait la connaissance des pionniers de ces méthodes, Hansen, Hurwitz et Deming, et conservera des liens avec eux après son retour.

Revenu en France avec une malle de documents, il souhaite en tirer une thèse sur la méthode des sondages, mais ses activités ultérieures ne lui en laisseront pas le temps⁴. Il s'associe avec Pierre Thionet (1916-2002)⁵, le spécialiste français des sondages aléatoires⁶, responsable à l'Insee du service des sondages. Michel Armatte

(2003) a décrit en détail l'introduction des enquêtes par sondages aléatoires en France.

Il présente ainsi le « tandem Thionet - Lévy-Bruhl » : « *Lorsque Lévy-Bruhl rentre des États-Unis en septembre 1947, il est affecté aux sondages et, même s'il consacre une première année à suivre les cours de l'école d'application, il forme avec Thionet un tandem qui va se trouver maître de la conception des enquêtes par sondages, tandis que la partie opérationnelle (liaison avec les directions régionales et les enquêteurs) allait vite leur échapper au profit de la direction de l'exploitation (Chevry, assisté de Brichler et Desabie après 1949). Tous deux vont pouvoir passer du transfert de technologie à un programme d'enquêtes par sondage.* » (Armatte, 2003).



Raymond Lévy-Bruhl autour des années 1945

Source : droits réservés

Les premières enquêtes par sondage effectuées par le « tandem » portent sur les effets de la suppression partielle du rationnement des textiles (1948) et sur la consommation de produits alimentaires non contingents (1949). Ces enquêtes ont été les prototypes d'un modèle de sondage aléatoire que Thionet et Lévy-Bruhl amélioreront et généraliseront ensuite. L'étape suivante est la réalisation de la première enquête emploi, avec la méthode du sondage aréolaire que Lévy-Bruhl a étudiée au *Census Bureau* américain. La première version date de 1950 (Lévy-Bruhl, 1951a et 1951b). Thionet construit le plan

de sondage et fait les calculs de précision, tandis que Lévy-Bruhl met au point le questionnaire et les instructions aux enquêteurs. De ce point de vue, leur choix essentiel est d'avoir privilégié la méthode des sondages aléatoires à celle des quotas, cette dernière étant utilisée par les instituts de sondages, depuis Gallup, notamment par Jean Stoetzel. Lévy-Bruhl publie même un article (1951c) pour railler l'échec de Gallup au moment de l'élection présidentielle américaine de 1948, où Dewey, annoncé vainqueur par les sondeurs, a en réalité été battu par Truman.

À l'occasion de la publication des résultats de l'enquête Emploi de 1950, il mentionne les insuffisances des sources antérieures sur l'emploi, issues du ministère du Travail. Il sera ensuite affecté à ce ministère en 1951, pour y effectuer l'enquête sur l'Activité et les conditions d'emploi de la main d'œuvre (Acemo). Il contribue aussi à la diffusion et à l'enseignement des méthodes de sondage. Il cosigne avec Thionet le cours publié en 1947-1948 de théorie et pratique des sondages, où il rédige deux leçons sur « l'enquête verbale » et sur « les sondages aux États-Unis ».

Il enseigne en 1949 au *Centre européen d'application statistique agricole et démographique*⁷ à l'École nationale d'administration (ENA) de 1955 à 1957, puis à l'Institut des sciences sociales du travail (ISST) de 1958 à 1968.

Il est affecté à la division de la statistique sociale de l'Insee en 1950. En septembre 1951, André Aboughanem, chef de la division statistique du ministère du Travail, lui demande de le remplacer lorsqu'il est envoyé au Bureau international du travail (BIT) à Genève. Il arrive ainsi « un peu par hasard »⁸ dans une division créée en 1946 et dont il représente le seul technicien. À 29 ans, il se trouve à la tête d'une soixantaine d'agents du ministère du Travail, dans l'une des rares structures déconcentrées de l'Insee. Le cadre traditionnel du ministère l'éloigne de « l'esprit de camaraderie » qui régnait à l'Institut parmi les jeunes polytechniciens⁹.

4. Quarante ans plus tard, un jeune historien des sciences, Emmanuel Didier, fera lui aussi un périple aux États-Unis pour préparer une thèse sur l'histoire des sondages. Raymond Lévy-Bruhl lui donnera généreusement ces précieux documents, qui dormaient dans la fameuse malle.

5. Toute cette partie doit beaucoup à un article très complet sur Thionet et Lévy-Bruhl, publié par Michel Armatte (2003) dans le *Journal de la Société française de statistique*.

6. Jean Stoetzel, créateur de l'IFOP, avait introduit dès 1938 les sondages d'opinion, sur le modèle des enquêtes américaines de Gallup.

7. Centre créé par G. Darmais avec le soutien de l'Organisation des Nations unies (ONU), l'Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO) et l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO).

8. Entretien de Raymond Lévy-Bruhl le 23 juin 2005 avec É. Penissat et B. Touchelay.

9. *Id.*

La division a toutefois le grand avantage d'être rattachée aux deux directions du ministère (la direction de la main d'œuvre et la direction du travail), ce qui est un gage d'indépendance. Elle dispose en outre, depuis sa création en 1946, d'un atelier mécanographique qui est bien équipé. Grâce à cet atelier, le chef de la division maîtrise le calendrier des travaux.

Lévy-Bruhl considère que le fait que les statistiques du travail soient établies par le ministère, et par conséquent en dehors de l'Insee, présente « un gros avantage » car il « permet un rapport direct » entre les producteurs de ces statistiques et « ceux qui font la réglementation ». Cette organisation facilite aussi la compréhension des données par ses utilisateurs. De surcroît, la position d'administrateur détaché de l'Insee garantit au chef de cette division une indépendance par rapport au cabinet du ministre du Travail. Enfin, la division est également autonome vis-à-vis de l'Institut qui ne centralise pas les données. Cette autonomie lui permet de répondre à des « préoccupations statistiques de moyen terme » et de laisser au second plan les « préoccupations politiques du moment » lorsqu'il établit ses questionnaires¹⁰.

Selon son témoignage, en mai 2007¹¹ : « J'ai été pendant les années 1950 le chef de la division statistique du ministère du Travail et j'aimerais répondre à quelques-unes des questions... La première concerne les rapports avec les différentes parties prenantes de l'administration. Je crois que le ministre du Travail, dans les années 1950, avait une position un peu intermédiaire, peut-être meilleure qu'au début du siècle mais moins bonne que dans les années plus récentes. En revanche, le ministre du Travail, pendant les années 1950, a été un patron agréable et solide pour la statistique. Aussi bien Paul Bacon qu'Albert Gazier étaient des hommes qui pensaient que l'information devait être à la fois fiable et désagréable. Elle était ce qu'elle était. J'ai trouvé dans ces deux ministres une aide importante pour le développement du service statistique dont j'avais la responsabilité. [...] En revanche, l'admini-

stration du ministère du Travail était restée un peu sclérosée sur le plan de la statistique ; sans être hostile, elle avait une position très neutre et ne voyait pas l'intérêt des travaux que nous faisons ».

Pourtant, en 1951, les collaborateurs de Lévy-Bruhl au ministère sont convaincus qu'il va rapidement les quitter pour aller occuper d'autres fonctions plus prestigieuses. Il y reste dix ans pendant lesquels il contribue très largement à faire reconnaître la nécessité des statistiques par ses collaborateurs et conforte l'image du technicien. La réputation de la division statistique, et surtout celle de son atelier mécanographique, est celle d'un « bague » où le travail est strictement évalué mais le *turnover* du personnel particulièrement faible.

Sa tâche principale consiste à organiser et à exploiter les enquêtes trimestrielles Acemo dont la qualité et la fiabilité sont améliorées. Il décharge les inspecteurs du travail de toute intervention dans ces enquêtes à l'occasion d'une « grève opportune » en 1951 et sépare ainsi la fonction statistique de celle du contrôle. À partir de cette date, les questionnaires sont transmis directement aux entreprises par l'Insee, qui tire l'échantillon de son fichier, ce qui réduit considérablement les délais. Il parvient aussi à simplifier le questionnaire pour limiter les non-réponses. Il aboutit à un questionnaire qui tient en deux pages recto verso et qui comporte des questions extrêmement simples. De ce fait, l'exploitation des questionnaires et la diffusion des résultats de l'enquête prennent d'abord deux mois, puis sept semaines après 1951, ce qui est un record pour un échantillon qui se stabilise autour de 30 000 entreprises. Ainsi, pour une enquête organisée le 1^{er} avril, les résultats sont disponibles le 20 mai. Les principaux tableaux sont publiés dès juin par l'Insee.

L'enquête Acemo est la seule disponible pour éclairer la politique économique. Elle informe en particulier sur l'impact des variations du Smig¹² sur le niveau des prix, information qui retient toute l'attention des autorités politiques en période de forte infla-

tion. L'enquête sert aussi à établir des indices d'évolution des effectifs salariés par classe d'effectifs des établissements et par secteur d'activité. Les indices sont pondérés à l'aide de coefficients proportionnels aux effectifs de l'année de base arrêtée en 1946. L'enquête fournit également un indice de la durée hebdomadaire moyenne du travail (base 100 en 1938). Elle permet en outre de suivre l'évolution des écarts entre les salaires masculins et féminins et selon les qualifications.

Lévy-Bruhl indique que cette enquête « est faite pour le mieux » malgré « le flou » qui entoure encore les notions de salaires et de chômage comme celle de durée du travail. Ses résultats, pour évaluer les effectifs de salariés par exemple, sont devenus plus fiables à mesure de la stabilisation de l'échantillon, même si la connaissance du temps partiel reste aléatoire. Enfin, la régularité de cette enquête en fait tout l'intérêt.

Une autre facette de l'activité de Lévy-Bruhl au ministère du Travail a été sa participation aux études pour établir des statistiques sur les charges sociales et salariales des pays membres, au moment de la création de l'Union de l'Europe occidentale (UEO), puis de la Communauté européenne du charbon et de l'acier (CECA)¹³. Il noue alors des contacts avec des statisticiens étrangers, ainsi qu'avec des représentants du patronat français, comme le montrent ses propos :

« [...] À cette époque, les travaux internationaux ont commencé, dans le domaine statistique, à se développer beaucoup plus. Il y avait effectivement le BIT mais, du fait de la création de la CECA et de l'obligation pour elle d'obtenir des données de bonne qualité, une activité interna-

10. *Id.*

11. Témoignage de R. Lévy-Bruhl au séminaire sur le centenaire du ministère du Travail, DARES-CHATEFP, 23 mai 2007.

12. Salaire minimum interprofessionnel garanti, devenu le Smic (salaire minimum interprofessionnel de croissance) en 1970.

13. Témoignage de R. Lévy-Bruhl au séminaire sur le centenaire du ministère du Travail, DARES-CHATEFP, 23 mai 2007.

tionale importante s'est développée à Luxembourg. Cette activité s'est poursuivie ensuite dans le cadre de la Communauté économique européenne, mais elle a débuté avant. Elle a été un aiguillon pour faire comprendre au patronat d'une part, aux syndicats ouvriers d'autre part, l'importance du développement de la statistique et de la concertation entre ces États. Je signalerai enfin qu'il y avait une organisation dont on parle très peu, l'Union de l'Europe occidentale qui, sur le plan statistique, avait un comité qui se réunissait souvent, soucieux de mieux connaître ce qui se faisait dans les autres pays. »



Source : droits réservés

Raymond Lévy-Bruhl à la fin des années 1960

En 1961, Lévy-Bruhl revient à l'Insee, dont Claude Gruson vient d'être nommé directeur général. Celui-ci lui confie la direction de la coordination générale, qu'il conserve jusqu'en 1966. Puis il est directeur de cabinet du Directeur général de 1966 à 1971. À cette époque, il a à gérer, sous la direction de Jean Ripert, les vastes « remues-ménages » issus des débats de mai 1968, puis la « réforme MacKinsey » de 1970¹⁴. Il est enfin Secrétaire général de l'Insee de 1972 à 1988, c'est à dire pendant toute la période (1974-1987) où Edmond Malinvaud est directeur général. Il a eu notamment à préparer et organiser, de 1973 à 1975, la construction du nouveau bâtiment de l'Insee à la porte de Vanves, où se trouve l'Institut encore aujourd'hui, ainsi que le déménagement depuis les anciens locaux du Quai Branly (espace aujourd'hui occupé par le Musée des arts premiers).

Après son départ à la retraite en juillet 1988, Lévy-Bruhl continue à avoir une activité importante. En 1992, Jean-Claude Milleron, alors Directeur général, lui demande d'entreprendre une

recherche sur la question du rôle du SNS pendant la période de l'Occupation allemande. Cette question tient particulièrement à cœur à Lévy-Bruhl. En effet, c'est à ce moment que des

interrogations ont surgi sur le fichage policier des juifs et l'Insee souhaite que la question du rôle éventuel du SNS soit étudiée de façon approfondie, avec l'aide d'historiens spécialisés et par l'étude des archives.

Ce sera fait avec l'historien Jean-Pierre Azéma, puis avec Béatrice Touchelay. Il en résultera, en 1998, un rapport signé de ces trois personnes, qui sera le point de départ d'autres recherches sur le sujet, menées notamment par Robert Carmille (fils de René Carmille) et Xavier Jacquey (fils de Pierre Jacquey, un ancien inspecteur général de l'Insee et ancien du SNS).

Ces divers travaux montrent que les registres du SNS n'ont pas servi aux rafles de juifs, ni aux envois de jeunes hommes au Service du travail obligatoire (STO) en Allemagne, les deux questions qui inquiétaient Lévy-Bruhl. ■

Bibliographie

- ARMATTE M., 2003 : « Pierre Thionet et l'introduction en France des méthodes de sondage aléatoire », *Journal de la Société française de statistique*, Vol. 14, n° 1-2, p. 227-255.
- AZEMA J.P., LEVY-BRUHL R., TOUCHELAY B., 1998 : Rapport de la Mission d'analyse historique sur le système statistique français de 1940 à 1945, Insee.
- Insee 1996 : *Cinquante ans d'Insee... ou la conquête du chiffre*, (ouvrage publié à l'occasion du cinquantenaire de l'Insee).
- LÉVY-BRUHL R., 1951a. « Enquêtes récentes par sondage sur l'emploi ; avec discussion », *Journal de la Société statistique de Paris*, 92, p. 108-125.
- LÉVY-BRUHL R., 1951b : « Une enquête par sondage sur l'emploi », *Supplément au BMS*, janv.-mars.
- LÉVY-BRUHL R., 1951c : « Les sondages d'opinion et l'échec des prévisions électorales de Gallup en 1948 », *Journal de la Société statistique de Paris*, 92, p. 63-70.
- LÉVY-BRUHL R., « Souvenirs de la période 1945-1960 », in Insee, *Pour une histoire de la statistique*, tome 1, contributions, Insee-Economica, p. 559-564.

14. Sur ces épisodes, on peut se reporter au volume du Cinquantenaire de l'Insee (1996), p. 101 à 105, qui évoque le rôle alors joué par Lévy-Bruhl.

Épisode n° 9 - Écrire les expressions numériques : les pièges à éviter

 Serge Darriné*

Remerciements : l'auteur remercie vivement Jean-Jacques Ribier, correcteur de l'Imprimerie nationale mis à disposition de l'Insee au sein du secrétariat technique aux éditions de la division production éditoriale (département de l'offre éditoriale, direction de la diffusion et de l'action régionale), pour ses conseils avisés et sa relecture attentive.

Les statisticiens manient quotidiennement des nombres et, plus généralement, des expressions numériques. Celles-ci sont ensuite reprises dans des publications, destinées à un lectorat plus ou moins spécialisé. Comment les écrire ? Dans quels cas faut-il le faire en toutes lettres, dans quels cas en chiffres ?

Telles sont les questions auxquelles ce neuvième épisode de la rubrique « Savoir compter, savoir conter » répond, en trois parties. La première porte sur les nombres, également appelés « adjectifs numéraux cardinaux » (un, deux, trois, etc.). La deuxième partie traite des « adjectifs numéraux ordinaux », autrement dit ceux qui décrivent, non pas une quantité, mais un rang (premier, deuxième, troisième, etc.). Enfin, la troisième partie de cet article s'intéresse à la manière d'écrire les unités de mesure et les signes mathématiques.

Des chiffres et des lettres... ou comment écrire les nombres

Les nombres, ou adjectifs numéraux cardinaux, peuvent s'écrire de plusieurs manières : en toutes lettres (un, cinq, dix) ou bien en chiffres, ces derniers pouvant être arabes (1, 5, 10) ou romains (I, V, X). Dans quel cas faut-il écrire les nombres en toutes lettres ou bien en chiffres ?

Dans une publication statistique, surtout si elle contient beaucoup de

nombres, mieux vaut les écrire au moyen de chiffres. L'écriture en chiffres est en effet à la fois plus courte et plus lisible.

Pour la même raison, les nombres écrits au moyen de chiffres se retrouvent dans les ouvrages d'économie, finance, sciences, etc.

Les nombres s'écrivent également ainsi quand ils visent à comparaison, font partie d'énumérations, opérations, tableaux, ou sont placés entre parenthèses. Exemples : La ville comptait 782 320 habitants au dernier recensement. L'importance du domaine (400 ha) exigeait de gros investissements. En 2001 : 17 200 inscrits, 15 309 votants.

S'écrivent aussi avec des chiffres : les dates (sauf les mois, bien sûr, si l'on choisit la forme substantivée telle que « janvier »), les heures, minutes et secondes, ainsi que les âges. Exemples : L'armistice a été signé le 11 novembre 1918. L'avion décolle à 18 h 21. L'exposition ouvrira le 10 du mois prochain. Il est âgé de 45 ans.

Quelques exceptions cependant : si le texte ou une partie de celui-ci ne contient que des nombres entiers inférieurs à 10, autrement dit des nombres qui s'écrivent avec un seul chiffre, mieux vaut les écrire en toutes lettres. Exemple : Au cours des cinq années observées, le chômage a augmenté dans quatre départements de la région Île-de-France.

En revanche, si, dans une énumération, les nombres utilisés sont tantôt supérieurs ou égaux à 10, tantôt inférieurs à ce nombre, il est préférable d'unifier la manière de les écrire. On utilise dans ce cas les chiffres plutôt que les lettres. Exemple : L'agglomération dispose de 3 supermarchés et de 12 commerces de proximité.

Mais les nombres s'écrivent en toutes lettres lorsqu'ils correspondent à des mesures approchées, arrondies, telles que les durées, ou encore dans le cas de fractions marquant des quantités approximatives. Exemples : Le train avait deux heures de retard. Les trois quarts de la population de cette commune sont des retraités.

Il existe par ailleurs quelques règles d'utilisation des chiffres ou des lettres pour écrire les nombres en dehors des publications statistiques (encadré 1 page suivante).

Écrire les nombres en toutes lettres

Rien de plus simple, en apparence, que d'écrire les nombres en toutes lettres : un, deux, trois, quatre, tout le monde connaît cela. Il faut néanmoins

* Au moment de la rédaction de cet article, Serge Darriné était rédacteur en chef du *Courrier des statistiques*, au sein de la division des relations avec les services statistiques des ministères, département de la coordination statistique, direction de la coordination statistique et des relations internationales de l'Insee.

se rappeler un certain nombre de règles, apprises à l'école, mais peut-être oubliées depuis...

Les nombres, ou adjectifs numéraux cardinaux, peuvent être simples – c'est-à-dire composés d'un seul mot, comme « sept » et « huit » –, ou bien complexes – composés d'au moins deux mots, comme « trente-quatre ». Les nombres complexes s'écrivent avec des tirets entre les mots exprimant des dizaines et des unités, sauf si ces mots sont séparés par la conjonction « et ». C'est ainsi qu'on écrit « vingt et un » (sans tiret), « vingt-deux » (avec tiret), « quatre-vingt-dix-neuf » (trois tirets) et « cent trente-cinq » (sans tiret entre « cent » et « trente », car « cent » n'est pas l'expression d'une unité ni d'une dizaine, mais d'une centaine).

Par ailleurs, les nombres sont invariables, autrement dit ils ne s'accordent ni en genre (masculin / féminin) ni en nombre (singulier / pluriel) avec le nom qu'ils qualifient. Cette règle a cependant trois exceptions notables : un, vingt et cent.

La première de ces trois exceptions, le nombre « un », s'accorde en *genre* : on écrit par exemple « une élève » lorsqu'on veut parler d'un élève de sexe féminin.

Les deux autres exceptions, « vingt » et « cent », s'accordent quant à elles en *nombre*, prenant un « s » final lorsqu'il y a plusieurs dizaines ou centaines ; mais attention, cela n'est vrai que si le nombre en question se termine par les mots « vingt » ou « cent » ou bien, pour le dire autrement, si ce nombre exprime des dizaines ou des centaines entières. C'est ainsi qu'on écrit « quatre-vingts » (avec s, puisque « vingt » est le dernier mot du nombre), mais « quatre-vingt-un » (sans s à « vingt », puisque 81 n'est pas un multiple de 20). De la même façon, on écrit « deux cents » (avec s) mais « deux cent soixante » (sans s à « cent »).

Écrire les nombres en chiffres

Règles d'écriture : les nombres en chiffres exprimant une quantité s'écrivent par tranches de trois chiffres,

Encadré 1 : règles les plus fréquentes d'écriture des nombres en dehors des publications statistiques

- Dans les adresses postales, les numéros d'immeubles, arrondissements, codes postaux, etc., s'écrivent avec des chiffres.
Exemple : 18 boulevard Adolphe Pinard, 75675 PARIS CEDEX 14 ;
- dans les ouvrages imprimés, les numéros de notes et les appels de notes, les numéros de colonnes, paragraphes, articles, pages, lignes, etc., s'écrivent avec des chiffres ;
- dans les actes judiciaires et notariés, les nombres sont écrits en toutes lettres pour des raisons de sécurité ou pour éviter toute ambiguïté (par exemple la date et le montant de la vente d'un bien immobilier). Idem sur les chèques, c'est le montant écrit en toutes lettres qui fait foi ;
- à l'inverse, lorsqu'il s'agit de citer un texte ayant une valeur juridique (loi et article de loi, décret, circulaire, etc.), le numéro de ce dernier est écrit en chiffres : « Vu l'article 5 de la loi n° 59-473... » ;
- dans les œuvres littéraires, les nombres sont généralement écrits en toutes lettres (sauf dans le cas, assez rare, où cela exige trop de caractères) ;
- c'est aussi le cas des nombres entrant dans la composition des noms de rue (à l'exception des dates historiques, en chiffres arabes, et des numéros de souverains, en chiffres romains).
Exemples : La rue des Trois-Bornes. La rue du 4-Septembre. L'avenue Louis XIV ;
- c'est encore le cas pour les nombres substantivés, comme les classes de l'enseignement secondaire.
Exemple : Redoubler la cinquième ;
- enfin, les nombres s'écrivent parfois en chiffres romains : outre le cas cité ci-dessus (numéros de souverains), c'est aussi le cas pour les siècles (XVIII^e, et non 18^e ni dix-huitième), les divisions d'un ouvrage, indications de tomes, livres, volumes, titres, parties, actes, chants, psaumes, planches, appendices.
Exemple : Bulletin mensuel, tome XXVI.

séparées par une espace insécable (encadré 2).

Cela revient concrètement à placer ces espaces insécables entre les centaines et les milliers.

Exemple : 35 700.

Le cas des dates historiques et des années est particulier : il est en effet admis d'écrire certaines dates célèbres en chiffres et en ne gardant, parmi ceux-ci, que les deux derniers. C'est ainsi qu'on lira parfois « les principes de 89 » (pour « 1789, année de la Révolution française »), « les événements de 68 » (pour « mai 1968 »), ou encore « la guerre de 14-18 » (pour « 1914-1918 »).

Mais dans une publication statistique amenée à citer des dates, même célèbres, mieux vaut éviter toute ambiguïté et donner les années entières, au moyen des quatre chiffres qui les composent. De la même façon, pour parler des années 1930 à 1939, on écrira « les années 1930 » (ou, à la rigueur, « les années trente »), mais pas « les années 30 », qui, hors

contexte, renvoient autant à l'empereur romain Tibère qu'au Front populaire.

Deux erreurs orthographiques et une erreur de formulation fréquentes autour des nombres

Le nom de l'objet décrit au moyen de nombres se met au pluriel à partir de deux, autrement dit, l'accord se fait selon la partie entière du nombre : on écrira donc « 1,8 enfant par femme » (sans s à « enfant » car la partie entière de 1,8 est 1) mais « 2,1 enfants par femme » (avec un s à « enfants » car la partie entière de 2,1 est 2).

Les fractions s'écrivent quant à elles sans trait d'union : deux tiers, trois quarts, quatre cinquièmes, etc.

Enfin, sous prétexte de gagner du temps lorsqu'elles donnent un ordre de grandeur, certaines personnes écrivent « 5 à 6 000 » pour signifier « 5 000 à 6 000 ». Si ce raccourci, sorte de « mise en facteur des milliers », est toléré à l'oral, il est à proscrire à l'écrit car il prête à confusion : l'ordre de

grandeur du phénomène étudié semble être en effet d'une amplitude, non pas d'un millier (différence entre 6 000 et 5 000), mais de presque 6 000 (différence entre 6 000 et 5)...

Un peu d'ordre, s'il vous plaît ! ou comment écrire les adjectifs et adverbes ordinaux

Les adjectifs ordinaux, tout comme les adverbes correspondants, indiquent l'ordre ou le rang. Exemples : « Elle a été reçue *première* à l'école Polytechnique » ; « Il voyage en *seconde* classe » ; « *Troisièmement*, l'inflation n'a pas joué en faveur de la reprise économique ».

Comme les nombres, les adjectifs ordinaux peuvent s'écrire de plusieurs manières, en lettres ou en chiffres : « premier » ou « 1^{er} », « deuxième » ou « 2^e », « dix-huitième » ou « 18^e », voire « XVIII^e », etc. Le choix de l'une ou l'autre des solutions dépend du contexte, mais il est globalement le même que dans le cas des nombres vu précédemment.

C'est ainsi que la forme abrégée sera choisie, dès lors que le nombre correspondant est composé de plusieurs chiffres, ou bien dans le cas d'un texte contenant un grand nombre d'adjectifs ordinaux, dont l'écriture (et la lecture) en toutes lettres serait fastidieuse.

Enfin, les adjectifs ordinaux ne s'abrègent pas s'ils désignent la classe de voyage en train, avion... On écrira donc « première classe » mais pas « 1^{re} classe ».

Les adverbes ordinaux peuvent eux aussi s'écrire en toutes lettres (« premièrement » ou « primo », « deuxièmement » ou « secundo ») ou en abrégé (« 1^o », « 2^o »). C'est la présentation abrégée qui sera privilégiée pour numéroter des titres ou intertitres, tandis que l'écriture de l'adverbe en toutes lettres est préférable quand celui-ci se trouve dans le corps même de l'article (énumération de différents points au sein d'un raisonnement, par exemple).

Encadré 2 : « un » ou « une » espace ?

Le mot « espace » est du genre féminin lorsqu'il désigne, en imprimerie, le blanc servant à séparer les mots. Pour insérer une espace insécable dans un texte, il existe au moins trois méthodes, valables pour le logiciel Word : taper sur son clavier « Alt » et simultanément « 0160 » ; ou bien, simultanément « Ctrl », « Maj » et « barre d'espace » ; ou encore, taper un signe de ponctuation automatiquement précédé d'une espace insécable (deux points, point virgule, point d'exclamation...) et effacer le signe de ponctuation afin de ne laisser que l'espace insécable qui le précède.

Écrire les ordinaux en toutes lettres

C'est bien connu, les adjectifs ordinaux s'accordent en genre et en nombre avec le nom qu'ils qualifient : on écrit par exemple « les *premiers* résultats » (accord au pluriel) et « la *seconde* baisse au cours du trimestre » (accord au féminin).

Parfois, les adjectifs cardinaux, autrement dit les nombres, sont employés dans un sens ordinal, c'est-à-dire pour désigner un rang. Par exemple, la référence à un article publié en page 1158 d'une revue, elle-même parue en l'an 2005, signifie que l'article débute en page n° 1158 (autrement dit en 1158^e page de la revue), et que cette dernière a été publiée au cours de la 2005^e année de notre système de mesure du temps. Il ne s'agit donc pas d'une mesure, comme pour le nombre des chômeurs par exemple, mais d'un rang : les pages d'un livre et les années qui s'écoulent sont numérotées les unes après les autres, de manière incrémentée.

C'est pourquoi ces nombres s'écrivent sans espace séparant les centaines des milliers (« 1158 » et « 2005 »), contrairement au cas où ils seraient employés dans un sens cardinal (« La commune comprend 1 158 habitants au dernier recensement »).

De plus, lorsque l'adjectif cardinal est employé dans un sens ordinal, il est invariable. Exemples : « les années quatre-vingt » (sans s à vingt car, bien que « quatre-vingts » prenne normalement un s final en tant que cardinal, ici il s'agit d'une date exprimée en années, et donc d'ordinal...) ; « page

deux cent » (sans s à cent, pour la même raison, la notion de numéro de page étant ordinale) ; « page un » (et non « page une », toujours pour la même raison).

Écrire les ordinaux en abrégé

Comment s'abrègent les adjectifs ordinaux ? « Premier » s'abrège en « 1^{er} » (avec les lettres « er » en exposant), « première » en « 1^{re} » (et non en « 1^{ère} ») et « second » s'abrège en « 2nd ». Dans tous les autres cas, l'abréviation consiste à faire suivre le nombre écrit en chiffres par la lettre « e » en exposant, et rien d'autre. Autrement dit, « deuxième » s'abrège en « 2^e », et non en « 2^{ème} » et encore moins « 2^{ième} » comme on le voit trop souvent.

Les adverbes ordinaux (premièrement, deuxièmement...) s'abrègent quant à eux sous la forme d'un « o » en exposant placé immédiatement après le nombre. Exemples : « 1^o », « 2^o », etc.

De la mesure en toute chose, ou comment écrire les unités de mesure et les signes mathématiques

Le lecteur a besoin de connaître les unités de mesure des grandeurs décrites dans les publications statistiques, aussi bien quand il lit le texte que quand il consulte les figures (tableaux, graphiques et cartes). Cependant, « toujours » ne signifie pas « partout » : il suffit en effet dans certains cas (figures) que l'unité soit précisée une seule fois, à un endroit judicieux permettant au lecteur de la retrouver à tout moment.

Quelles sont les principales unités de mesure ? Comment les écrit-on ?

Les unités de mesure se placent toujours *après* le nombre correspondant à la mesure. C'est vrai pour les intitulés en toutes lettres comme pour les abréviations. Ces dernières sont toujours séparées du nombre qui les précède par une espace insécable. Une erreur courante consiste à oublier cette espace insécable avant l'abréviation « % » : elle doit pourtant systématiquement s'y trouver.

Exemples : « 2 000 € » (et non « € 2 000 », ni « 2 000€ ») ; « 30 % » (et non « 30% »).

Toujours à propos des pourcentages, une autre erreur courante consiste à écrire un intervalle sous la forme « 3 à 5 % », alors qu'il faut répéter le signe « % », ce qui donne : « 3 % à 5 % ».

Les unités de mesure fréquemment utilisées dans les publications statistiques sont peu nombreuses (tableau 1). Elles ont des abréviations officielles qu'il est impératif de respecter.

À partir de ces unités, il est possible de former des multiples (kilomètre au lieu de mètre) et des sous-multiples (centimètre), soit en faisant précéder l'intitulé de l'unité de mesure, s'il est en toutes lettres (mètre), d'un préfixe lui aussi en toutes lettres (kilo) ; soit en faisant précéder l'abréviation de l'unité de mesure (m) d'un symbole, abréviation du préfixe (k) (tableau 2).

Les préfixes et symboles ci-dessus doivent toujours être collés à l'unité qui les suit, qu'il s'agisse d'unités en toutes lettres ou abrégées. Attention à les écrire en lettres minuscules, sauf exceptions signalées dans le tableau 2, et à ne pas ajouter de « s » final, inutile. Exemples : « 20 k€ » (et non « 20 k € », ni « 20 K€ »). « 10 km » (et non « 10 kms », qui évoque des kilomètres multipliés par des secondes).

Concernant les plus grands multiples (10³ et au-delà), il existe une autre solution que de parler de « giga », « méga » et « kilo » à son lecteur : on peut tout simplement utiliser les expressions « milliards de » (Md en

Tableau 1 : intitulés en toutes lettres et abréviations officielles des principales unités de mesure

<i>Unités monétaires</i>		<i>Unité de longueur</i>	<i>Unités de masse</i>
euro	€	mètre	gramme
dollar	\$		tonne
<i>Unités de temps</i>		<i>Unité de surface</i>	<i>Unité de mesure d'angle</i>
jour	j	mètre carré	degré
heure	h		<i>Unité de température</i>
minute	min	<i>Unités de volume</i>	degré Celsius
seconde	s	mètre cube	
		litre	

Tableau 2 : préfixes en toutes lettres et abréviations officielles des multiples et sous-multiples décimaux les plus courants

Facteur par lequel l'unité est multipliée	Préfixe (en toutes lettres) à mettre devant le nom (en toutes lettres) de l'unité de mesure	Symbole à mettre devant l'abréviation de l'unité de mesure
10 ⁹	giga (ou "milliards de")	G (et non g, qui signifie gramme) (ou Md pour "milliards de")
10 ⁶	méga (ou "millions de")	M (et non m, qui signifie mètre)
10 ³	kilo (ou "milliers de")	k
10 ²	hecto	h
10	déca	da
10 ⁻¹	déci	d
10 ⁻²	centi	c
10 ⁻³	milli	m
10 ⁻⁶	micro	μ
10 ⁻⁹	nano	n

abrégé), « millions de » (M en abrégé, comme pour méga, ce qui tombe bien puisque les deux sont synonymes) et « milliers de » (pas d'abréviation spécifique dans ce cas, on utilisera donc le symbole k, comme pour kilo).

Si l'on utilise « Md » pour « milliards de », il faut penser à le signaler aux lecteurs car tous ne connaissent pas la signification de cette abréviation (certains croient qu'elle signifie « millions de ») ; attention également à ne pas ajouter un « s » après « Md », même s'il y a plusieurs milliards (l'abréviation « Mds » ne veut rien dire).

Dans les tableaux et les graphiques, il est fortement conseillé d'utiliser astucieusement les multiples des unités, afin d'éviter au lecteur des nombres trop longs : c'est ainsi que l'unité figurant sur l'axe d'un graphique ou dans la colonne d'un tableau sera exprimée par exemple en « milliers de salariés » ou en « millions d'euros ».

Toujours dans l'idée de faciliter la lecture, cette unité sera, dans le cas d'un tableau, précisée une fois et une seule dans la tête, c'est-à-dire

en tête de colonne, et non à côté de chaque chiffre de cette dernière. Cette règle s'applique aussi aux pourcentages : dans une colonne donnant des pourcentages, le signe « % » ne doit figurer qu'une fois, au sommet de celle-ci.

Les unités de mesure et les pourcentages dans un texte : en toutes lettres ou en abrégé ?

Dans le corps du texte, les unités de mesure doivent être écrites en toutes lettres (et non à l'aide d'abréviations), car c'est plus facile à lire : le lecteur peut en effet se concentrer sur le sens du texte, sans avoir à réfléchir *en même temps* à la signification des abréviations. Exemples : « En 2004, un salarié à temps complet de l'industrie percevait 23 000 euros nets en moyenne. » (et non « 23 000 € », ni « 23 k€ ») ; « Les 78 500 entreprises de l'immobilier ont réalisé cette année un chiffre d'affaires de 63 milliards d'euros » (et non « 63 Md€ »).

Il existe cependant une exception à cette règle, celle des unités de

mesure dont l'intitulé en toutes lettres est particulièrement long. Dans ce cas, la règle s'inverse : l'abréviation est préférable à l'intitulé, car elle simplifie la lecture. Exemple : « En 2004, l'Île-de-France avait en moyenne 940 habitants par km² » (et non « par kilomètre carré »).

Concernant les pourcentages : « pour cent » en toutes lettres n'est jamais employé ; il est systématiquement remplacé par son abréviation, « % » : « En 2004, en France métropolitaine, 57 % des ménages étaient propriétaires de leur résidence principale » (et non « 57 pour cent ») ; seule excep-

tion, les pourcentages employés adverbialement : « Un produit cent pour cent pure laine » (et non « 100 % pure laine »).

Pour finir... Les signes mathématiques

Les signes mathématiques les plus courants sont : « + » (plus), « - » (moins), « x » (multiplié par) et « / » (divisé par). Ils sont écrits systématiquement sous leur forme abrégée (+ - x /). Dans ce cas, ils sont tous séparés du texte auquel ils se réfèrent (y compris les chiffres qui les précèdent

et qui les suivent) à l'aide d'espaces insécables.

Le signe moins ne s'écrit pas comme un tiret « - » (symbole se trouvant sous le 6 sur le clavier, ou bien dans le coin supérieur droit du pavé numérique), mais ainsi : « - ». On l'obtient dans Word en tapant simultanément « Alt » et « 0150 » ; on le trouve aussi dans les caractères spéciaux. Il faut en effet que le signe moins soit de même largeur que la barre verticale du signe plus, ce qui n'est pas le cas avec le tiret. De plus, cela évite un éventuel passage à la ligne juste après le tiret, qui séparerait ce dernier du chiffre qui le suit. ■

Bases de données médico-administratives et épidémiologie : intérêts et limites

✍ Marcel Goldberg¹, Catherine Quantin², Alice Guéguen³ et Marie Zins⁴

Les épidémiologistes cherchent à décrire les phénomènes de santé des populations, à comprendre l'histoire naturelle des maladies et à juger, en termes de causalité, du rôle de facteurs de risque sur la santé. Dans tous ces domaines, les difficultés méthodologiques sont difficiles à contrôler. Elles proviennent essentiellement des biais potentiellement induits par divers phénomènes de sélection qui interviennent lors de l'inclusion des sujets dans l'étude épidémiologique et durant le suivi de celle-ci lorsqu'elle est longitudinale. Certaines solutions efficaces pour prendre en compte la non-participation et pour le calcul des pondérations impliquent le recours à des bases de données nationales réputées exhaustives (bases de données de santé et socioprofessionnelles essentiellement). Les problèmes rencontrés, aussi bien méthodologiques que ceux liés à des questions d'éthique et de confidentialité, sont illustrés par l'exemple de la cohorte Constances en cours de mise en place ; la plate-forme scientifique et technique « Plastico », actuellement en phase de préfiguration, peut apporter une aide considérable à la mise en œuvre des solutions proposées.

L'épidémiologie est la science qui étudie la fréquence et la distribution, dans le temps et dans l'espace, des problèmes de santé des populations, ainsi que le rôle des facteurs qui les déterminent. Son domaine d'intervention est tout ce qui concerne la santé de l'homme dans ses relations avec son environnement naturel, socio-économique ou culturel.

Dans un contexte épidémiologique, la prise en compte du temps joue un rôle capital. On peut considérer le temps de diverses façons, selon les questions que l'on se pose. Dans certains cas, on veut étudier l'évolution d'un phénomène de population dans le temps : la fréquence du tabagisme diminue-t-elle ? L'incidence des cancers augmente-t-elle ? Les approches pertinentes sont alors l'enquête transversale répétée (Enquête décennale santé par exemple) ou l'enregistrement systématique permanent (mortalité, registres du cancer).

Mais c'est aussi souvent à l'évolution de processus, à l'échelle individuelle, que les épidémiologistes s'intéressent. Deux questions, notamment, sont au cœur de la problématique épidémiologique. La première est « l'histoire naturelle » des processus physiopathologiques et des maladies (quelles sont les étapes qui vont de la primo-infection par le VIH⁵ à la maladie sida constituée et au décès ?) ; cet intérêt s'est particulièrement déve-



The Sick Girl (La jeune fille malade), tableau de Mikael Ancher, 1882

loppé dans le contexte des maladies chroniques, qui évoluent sur des décennies et qui dominent le paysage sanitaire des pays développés.

Outre le temps, une autre question qui intéresse particulièrement les épidémiologistes est celle de la causalité : l'exposition à un « facteur de risque » modifie-t-elle le risque de survenue d'une maladie ? Cette recherche de la causalité n'est pas l'apanage de l'épidémiologie traditionnelle « d'organe »

1. Marcel Goldberg est professeur à la Faculté de médecine de Paris Île-de-France – Ouest. Il a dirigé une unité de recherche d'épidémiologie de l'Institut national de la santé et des études médicales (Inserm) et est rattaché actuellement à l'Unité 687 (marcel.goldberg@inserm.fr).

2. Catherine Quantin est professeur au Centre hospitalo-universitaire (CHU) de Dijon, où elle dirige le service de biostatistique et d'informatique médicale (catherine.quantin@chu-dijon.fr).

3. Alice Guéguen est ingénieur de recherche à l'Inserm, Unité 687 (alice.gueguen@inserm.fr).

4. Marie Zins est chercheur et dirige, au sein de l'Inserm Unité 687, l'équipe Risques postprofessionnels-cohortes (RPPC) du Centre technique d'appui et de formation des centres d'examens de santé (Cetaf) (marie.zins@inserm.fr).

5. Virus de l'immunodéficience humaine.

(épidémiologie du cancer, des maladies cardiovasculaires...). Elle concerne aussi « l'épidémiologie sociale », qui s'intéresse aux déterminants sociaux de la santé : il ne s'agit pas seulement de décrire les inégalités de santé selon des variables sociales diverses, mais de tenter de comprendre les mécanismes par lesquels la situation sociale se traduit par la survenue de problèmes de santé, en tenant compte de l'exposition aux facteurs de risques « classiques » comme le tabac, l'alcool ou l'alimentation par exemple [Goldberg *et al.*, 2003].

La notion de cause en épidémiologie est probabiliste : une cause est un facteur qui modifie l'incidence d'une pathologie dans la population, en tenant compte des autres facteurs qui peuvent interagir dans la relation « exposition-effet ». Établir une relation de causalité entre une exposition à un facteur de risque et une maladie est le plus souvent difficile et long, car l'épidémiologie est avant tout une science d'observation où, pour des raisons notamment éthiques, il n'est usuellement pas possible d'expérimenter (c'est-à-dire de tirer au sort des sujets exposés et non exposés, et observer le cours naturel des choses).

Les épidémiologistes ont donc défini des « critères de causalité » pour juger de l'existence de relations causales : force de l'association entre exposition à un facteur de risque et maladie, mesurée par divers indices ; stabilité de l'association entre études épidémiologiques dans des populations diverses et avec des méthodes différentes ; existence d'une relation dose-réponse (plus l'exposition est forte, plus le risque est élevé) ; temporalité de l'association (la cause doit précéder l'effet) ; plausibilité biologique et cohérence de l'association avec les données connues. Réunir tous ces éléments peut être très long : par exemple, la première observation suggérant nettement que le nickel est un cancérigène date de la fin des années 1920, mais la confirmation « officielle » de son caractère cancérigène par le Centre international de recherche sur le cancer date du début des années 1990, soit après plus

de 60 ans et de nombreuses études dans différents pays !

Deux grands types d'études sont utilisés par l'épidémiologie analytique (ou étiologique⁶) : les études de cohorte (terme épidémiologique synonyme de panel) et les études cas-témoins ; les secondes diffèrent des premières notamment par la façon de prendre en compte le temps. Le principe dans la cohorte épidémiologique est le suivi individuel longitudinal d'un groupe de sujets. Il peut s'agir de personnes souffrant d'une maladie dont on veut étudier l'évolution, mais lorsqu'on se place dans une optique étiologique, les sujets sont *a priori* tous indemnes de la maladie étudiée, alors que l'étude cas-témoins est rétrospective : on part de sujets déjà malades (et de témoins indemnes de la pathologie étudiée) et on remonte dans le passé pour chercher l'exposition aux facteurs de risque soupçonnés qu'on compare entre groupes malades et témoins.

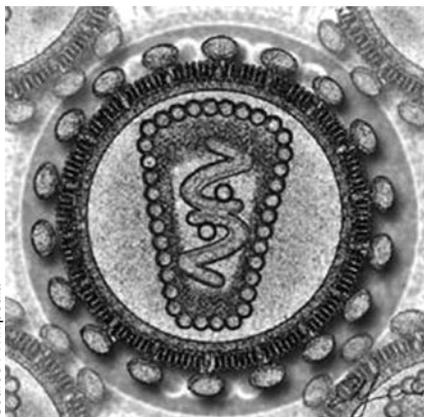


Schéma de la section d'un VIH

Dans ce qui suit, on envisagera essentiellement les études de cohorte qui, parce qu'elles assurent un suivi individuel des mêmes sujets, offrent la possibilité d'analyses épidémiologiques permettant de tenir compte au mieux de phénomènes liés au temps : séquence temporelle exposition – effet, effet de génération, effet de période. Il est ainsi possible de modéliser l'enchaînement et les effets conjoints des différents facteurs relatifs aux conditions de vie (alimentation, habitat, accès aux soins, réseau social...), à l'environnement (conditions de travail, expositions profes-

sionnelles et environnementales...) et à l'état de santé (chronologie des phénomènes pathologiques).

Globalement, les études de cohorte sont celles qui permettent d'étudier dans les meilleures conditions l'histoire naturelle des processus physiopathologiques et des maladies, et de juger, en termes de causalité, du rôle sur la santé de facteurs de risque ou d'interventions préventives en prenant en compte les évolutions temporelles et les effets conjoints de différents facteurs.

Le problème méthodologique majeur : les biais dus aux effets de sélection

Un biais est une erreur qui entraîne une différence systématique entre le paramètre que l'on cherche à estimer (par exemple l'incidence d'une maladie, une mesure d'association entre une maladie et un facteur de risque, ou encore entre un traitement et un état de santé) et le paramètre qui est estimé par l'étude.

Une des sources majeures de biais dans les enquêtes épidémiologiques provient des effets de sélection, qui surviennent lorsque la population observée diffère de la population cible en raison de phénomènes liés au recrutement ou au suivi des sujets. Les biais susceptibles de se produire peuvent concerner : (i) l'estimation de la fréquence de la maladie ou de l'exposition à un facteur de risque ; (ii) l'estimation de l'association entre exposition et maladie. L'estimation de la fréquence de la maladie ou de l'exposition, ou de l'association exposition – maladie, est biaisée si la probabilité d'être malade (ou exposé) n'est pas indépendante de la probabilité d'être inclus dans l'étude, ou si la relation exposition – maladie est différente entre les sujets inclus et ceux qui ne le sont pas.

Or, les cohortes épidémiologiques incluent souvent les sujets sur la base du volontariat, nécessaire dès qu'il s'agit de recueillir des données auprès des sujets eux-mêmes : des biais de sélection peuvent donc intervenir lors de la constitution de la

6. L'étiologie est l'étude des causes des maladies [NDLR].

cohorte et tout au long du suivi de celle-ci, notamment lorsque certains sujets faisant initialement partie de la cohorte cessent de participer à l'enquête (phénomène d'attrition) [Goldberg & Luce, 2001].

Les cohortes épidémiologiques sont conçues essentiellement pour des besoins de recherche épidémiologique, visant à étudier l'histoire naturelle des processus physiopathologiques et des maladies, et à comprendre les relations entre exposition à des facteurs de risque de nature diverse et survenue de maladies (objectif « analytique »). Elles peuvent également avoir des objectifs « descriptifs » d'information en santé publique (estimation de paramètres variés concernant l'ensemble de la population cible). Or, le problème des biais potentiels liés aux effets de sélection se présente de façon très différente selon que les objectifs sont analytiques ou descriptifs.

Étude analytique des relations entre expositions et maladies

La première phase d'une étude épidémiologique est l'inclusion des sujets qui y participent. Différentes procédures peuvent être mises en œuvre à cette fin : appel au volontariat par courrier ou appel téléphonique, sélection aléatoire ou non dans des fichiers divers, etc. Les procédures reposant sur le volontariat sont toujours à l'origine de biais de sélection qui sont parfois d'ampleur considérable. Au sein d'une cohorte dont les procédures d'inclusion ont été les mêmes pour tous les sujets, la relation exposition – maladie n'est cependant pas *a priori* différente entre les sujets volontaires et ceux qui ne le sont pas [Greenland, 1977 ; Criqui, 1979 ; Austin *et al.*, 1981]. Une des raisons en est qu'au moment de l'inclusion, tous les sujets sont indemnes des maladies qui seront analysées, seuls les cas incidents pendant la période de suivi étant pris en compte dans les études de cohorte : des conditions très particulières seraient en effet nécessaires pour entraîner un biais dans la mise en évidence ou la quantification d'une relation entre une exposition et une maladie.



The Doctor (Le docteur), tableau de Samuel Luke Fildes, 1891

Ainsi, pour analyser les effets du tabac sur le risque de cancer, il n'est pas nécessaire d'observer un échantillon représentatif de la population, mais de disposer d'effectifs suffisants de non-fumeurs et de fumeurs parmi lesquels le niveau d'exposition est contrasté : en effet, sur la base des connaissances actuelles, il est très vraisemblable que les mécanismes physiopathologiques et biologiques de la cancérogenèse liée au tabac soient identiques dans un échantillon de volontaires et dans l'ensemble de la population. Les effets de sélection dus au volontariat de la participation ne génèrent donc *a priori* pas de biais, ou seulement des biais minimes, lorsqu'il s'agit de comprendre comment les expositions à des facteurs de risque, les caractéristiques professionnelles et sociales, etc., influencent l'état de santé et peuvent être à l'origine de pathologies ou, au contraire, protectrices.

La probabilité de participation initiale est différente chez les sujets exposés et les non exposés au facteur d'intérêt, ce qui n'entraîne habituellement pas de biais. En revanche, le problème de l'attrition au cours du suivi peut, lui, être à l'origine de biais importants, si la probabilité de ne plus être suivi diffère chez les exposés et les non-exposés, ou bien si elle

diffère chez ceux qui sont devenus malades et ceux qui ne le sont pas, ce qui est souvent le cas.

Les épidémiologistes sont familiers de ces difficultés. On rappellera que toutes les cohortes épidémiologiques longitudinales reposant sur le volontariat (c'est-à-dire toutes celles qui impliquent la participation active des sujets pour fournir des données non disponibles par d'autres sources) présentent le même phénomène de sélection à l'inclusion et au cours du suivi. Certaines, célèbres dans l'histoire de l'épidémiologie, ont contribué de façon décisive à l'avancement des connaissances scientifiques : ainsi, c'est à partir d'enquêtes sur des volontaires qu'on a pu mettre en évidence le caractère pathogène du tabac, ou identifier les principaux facteurs de risque cardiovasculaires.

Étude descriptive de la fréquence des problèmes de santé et des expositions

Dans ce type d'étude, les paramètres d'intérêt doivent être estimés sur un échantillon représentatif de la population cible. Or, les cohortes épidémiologiques prospectives étant habituellement composées uniquement de volontaires, il existe inévitablement des effets de sélection,

même lorsqu'on procède à un tirage au sort aléatoire d'un échantillon dans une base de sondage appropriée.

Lors de l'inclusion initiale, les non-participants (personnes non retrouvées, refus, etc.), constituent une source potentielle de biais. Pour y remédier, on s'efforce de recueillir un minimum de données sur les non-participants – essentiellement sur leur âge, sexe, profession et catégorie socioprofessionnelle (PCS) –, afin de procéder ultérieurement à des redressements pour estimer les paramètres d'intérêt. Cette approche, utilisée dans de nombreuses enquêtes de santé (notamment l'Enquête décennale santé et l'Enquête santé et protection sociale), connaît cependant certaines limites. Ainsi, il n'est pas toujours possible de recueillir les données de redressement (âge, sexe, PCS) pour l'ensemble des sujets non participants.

De plus, il n'est pas toujours facile de savoir si ces données sont suffisantes pour contrôler les biais potentiels, car on sait par exemple qu'au sein de la même catégorie socio-économique, existent de larges différences à bien des égards, notamment en termes de santé, de comportements, de modes de vie, de réseaux sociaux, etc. Ceci a été bien montré par exemple dans la cohorte Gazel constituée de 20 000 agents de l'entreprise EDF-GDF volontaires suivis depuis 1989 [Goldberg *et al.*, 2007]. La comparaison de ceux-ci à leurs collègues non participants et de même catégorie socioprofessionnelle, sexe et âge montre que le statut marital, l'existence de maladies psychiatriques et de pathologies liées à des consommations élevées d'alcool et de tabac, sont fortement associés à la participation initiale à la cohorte [Goldberg *et al.*, 2001].

Finalement, on est rarement en situation de contrôler complètement les biais de sélection potentiels, faute de disposer de données pertinentes à la fois sur les participants et

l'ensemble des non-participants. Le même problème se pose tout au long du suivi, les non-répondants et les perdus de vue différant toujours des participants selon divers facteurs, en particulier les comportements de vie et les problèmes de santé qui jouent un rôle majeur, même à catégorie socioprofessionnelle égale. Là aussi, on a pu l'observer dans la cohorte Gazel : à sexe, âge et catégorie socioprofessionnelle équivalents, les risques d'attrition diffèrent en fonction des consommations d'alcool et de tabac, de l'état de santé perçue, de l'absentéisme médical⁷, de la survenue de problèmes de santé mentale et de cancers notamment [Goldberg *et al.*, 2006]. Or, c'est justement ce type de facteurs qui est étudié dans les cohortes épidémiologiques.

Des solutions pour l'analyse de la non-participation et le calcul des pondérations

Il existe des solutions pour obtenir un échantillon représentatif de la population cible et minimiser les biais liés à la sélection initiale et à l'attrition au cours du suivi. On s'appuiera ici sur celles qui ont été élaborées pour le projet Constances. Celui-ci vise à constituer une importante cohorte épidémiologique prospective, d'environ 200 000 individus⁸. Ces solutions s'inspirent d'une méthode déjà utilisée pour la cohorte Gazel, à partir de diverses bases de données portant sur les agents d'EDF-GDF. Cette méthode a permis d'identifier et de quantifier les effets de sélection liés au volontariat, à l'inclusion comme tout au long du suivi, de façon beaucoup plus fine qu'habituellement [Goldberg *et al.*, 2001 ; Goldberg *et al.*, 2006]. Afin de pouvoir procéder à des redressements permettant d'estimer les variables d'intérêt dans la population cible tout en minimisant les biais, la démarche générale retenue est la suivante.

Les sujets éligibles sont tirés au sort par sondage stratifié avec probabilités inégales, en surreprésentant les individus ayant une probabilité de non-volontariat plus forte. Les probabilités d'inclusion sont empirique-

ment définies à partir d'enquêtes déjà réalisées, où il est possible d'estimer les probabilités de participation en fonction des variables usuelles (sexe, âge, PCS).

Lors de l'inclusion, et chaque année pendant le suivi de la cohorte au cours du temps, un poids sera attribué à chaque sujet permettant de produire des estimations de certaines variables relatives à la population cible. Le principe de calcul de ces pondérations consiste à appliquer au poids de tirage initial, attribué à chaque individu, un coefficient de correction de « non-réponse » déterminé à partir de l'analyse des variables liées à la non-participation à la cohorte. La non-réponse observée une année donnée est le résultat de la non-participation initiale à la cohorte combinée à la non-réponse ponctuelle ou définitive. Avec le poids avant calage ainsi calculé, on peut alors déterminer une pondération finale par calage sur les marges décrivant de façon pertinente la population à laquelle on souhaite généraliser les résultats.

Cette démarche, relativement simple dans ses principes, présente des difficultés pour être mise en œuvre concrètement, notamment sur deux points : (i) dans la comparaison des participants et des non-participants pour le calcul des coefficients de redressement ; (ii) dans la définition des bases sur lesquelles il faut calculer les marges permettant le calage.

La comparaison des participants et des non-participants doit prendre en compte la non-participation et l'attrition au cours du suivi. L'analyse des facteurs liés à l'attrition ne pose pas de problème particulier, car on disposera de nombreuses informations sur tous les sujets qui abandonnent la cohorte puisque, par construction, tous auront été inclus. Il convient cependant de définir explicitement les critères pris en compte pour décider qu'une personne abandonne définitivement sa participation à la cohorte, la principale difficulté résidant dans cette définition, car la non-participation une année donnée ne préjuge pas forcément d'un abandon

7. L'absentéisme médical correspond aux arrêts de travail pour raison de santé.

8. www.constances.fr/index.html

(cf. l'expérience du suivi dans Gazel [Goldberg *et al.*, 2006]).

Le problème de la non-participation est à la fois plus classique et plus complexe à résoudre, puisque la non-participation équivaut le plus souvent, dans le cadre d'enquêtes reposant sur le volontariat, à une absence d'information, tant en ce qui concerne les caractéristiques des personnes que les motifs de leur non-participation. Une solution qui permet de contourner ces difficultés consiste à constituer et à suivre une « cohorte de non-participants » formée d'un échantillon aléatoire issu de la population des non-participants à partir de bases de données nationales existantes, l'ensemble (participants et non-participants) devant être représentatif de la population générale.

L'établissement des marges permettant le calage repose aussi sur l'utilisation des mêmes bases de données nationales afin de calculer les marges pertinentes qui, outre les habituelles caractéristiques (sexe, âge, PCS), doivent intégrer des variables relatives à la santé et caractéristiques du recours aux soins, toutes informations également disponibles pour les participants à la cohorte. La qualité du calage sera donc sensiblement améliorée par le calcul de marges ayant un rapport spécifique avec la finalité de la cohorte, c'est-à-dire des questions de santé.

En pratique, la cohorte de non-participants fera l'objet d'un suivi anonyme en ce qui concerne certaines

données également disponibles pour les participants de la cohorte, et qui sont habituellement associées à des phénomènes de sélection dans les études épidémiologiques [Goldberg & Luce, 2001], comme les hospitalisations, les accidents, l'absentéisme, la consommation de soins, les diagnostics d'affections de longue durée, le lieu d'habitation, la trajectoire socio-professionnelle.

Un tel dispositif permet le calcul d'un coefficient de traitement de non-réponse pour chaque individu. Pour cela, une fois l'analyse de la non-participation et de l'attrition réalisée, seules les variables explicatives de l'une et l'autre seront retenues pour la constitution de groupes homogènes quant à la probabilité de réponse. Dans un premier temps, le coefficient de traitement de la non-réponse « non-participation » est calculé comme l'inverse du taux de participation pondéré du groupe relatif à l'individu considéré. De la même façon, et dans un second temps, le coefficient de traitement de la non-réponse « attrition » est calculé comme l'inverse du taux de « non-attrition » pondéré du groupe relatif à l'individu considéré. Le coefficient de traitement de non-réponse final est alors calculé comme le produit de ces deux coefficients, ce qui permet d'obtenir des estimations non biaisées [Hernan *et al.*, 2004 ; Little & Vartivarian, 2003 ; Jones *et al.*, 2006].

Une telle approche nécessite donc de pouvoir utiliser des bases de données nationales exhaustives contenant des

données individuelles socioprofessionnelles et de santé pertinentes pour l'analyse épidémiologique. On verra dans la section suivante que les bases de données existantes ne sont pas indemnes d'imperfections de nature diverse, susceptibles d'induire des « erreurs de classement » (personnes non malades classées à tort malades, par exemple). Mais dans l'optique de la comparaison des participants et des non-participants, il s'agit d'erreurs « non différentielles », puisque la probabilité d'erreur est la même dans les deux groupes, la source d'information étant identique : ce type d'erreur diminue la puissance des tests, mais n'induit pas de biais [Bouyer *et al.*, 1993].

Bases de données nationales pertinentes pour l'épidémiologie

Il existe différentes bases de données nationales exhaustives⁹ réunissant des données individuelles diverses pouvant être utilisées dans des protocoles épidémiologiques, qu'il s'agisse, comme on vient de le voir, d'analyser la non-participation et de calculer les pondérations, ou de répondre à d'autres besoins (sélection de sujets présentant une pathologie, validation d'un diagnostic, suivi de personnes incluses dans des études, etc.).

Événements socioprofessionnels

Les bases de données de la Caisse nationale d'assurance vieillesse (Cnav) sont essentielles, à la fois pour permettre l'accès aux données socio-professionnelles et pour suivre les sujets dans les cohortes. Le rôle de cet organisme est notamment d'assurer le droit au paiement de la retraite pour toute personne ayant appartenu au moins une fois au régime général de sécurité sociale (RGSS) durant sa vie. Pour cela, la Cnav a mis en place un système permettant de collecter

9. On peut trouver une description plus complète des bases de données citées dans deux documents téléchargeables (www.rppc.fr/publications.html) : Ribet C, Genet J, Zins M. Les bases de données socioprofessionnelles. Rapport. Unité 684 Inserm – CNAMTS, 2006 et Coeuret-Pélissier M, Zins M. Les bases de données de l'assurance maladie. Rapport. Unité 687 Inserm – CNAMTS, 2006.



Source : site Gazel

Extrait du journal d'information de la cohorte Gazel

et traiter les données sociales issues de différents organismes et régimes gestionnaires des prestations sociales aux niveaux national, régional et local. Pour remplir son rôle, la Cnav a créé plusieurs bases de données qu'elle gère, dont les principales sont présentées succinctement ci-après.

Le Système national de gestion des identités (SNGI) contient l'ensemble des données (numéro d'inscription au répertoire ou NIR, état civil, statut vital) pour toute personne née en France métropolitaine ou dans les départements d'outre-mer (DOM) (premier cas). Il contient aussi les données d'identification des personnes nées à l'étranger dont l'inscription est demandée par un organisme autorisé sur la base d'un justificatif d'état civil, ou encore de celles nées dans les territoires d'outre-mer (TOM) et résidant sur le territoire français (second cas).

Le SNGI a pour finalité de certifier l'identité d'une personne. L'Insee a en charge l'immatriculation de toute personne se trouvant dans le premier cas ; ces informations sont contenues dans le Répertoire national d'identification des personnes physiques (RNIPP). Depuis 1981, il incombe à la Cnav, dans le cadre de sa mission déléguée par l'Insee, de procéder à l'immatriculation des personnes se trouvant dans le second cas. Le SNGI contient l'ensemble des éléments d'identification des personnes (NIR, nom patronymique, prénom(s), sexe, date et lieu de naissance, date et lieu de décès, numéros d'acte de naissance et d'actes de décès), soit reçus de l'Insee, soit intégrés par la Cnav elle-même.

Le Système national de gestion des carrières (SNGC) permet de retracer, pour chaque individu dès l'âge de 16 ans et jusqu'à la liquidation de ses droits à la retraite, ses différentes périodes d'activité : périodes d'activité professionnelle ou assimilées (chômage, maladie, maternité ou congés parentaux...). Le SNGC contient donc l'ensemble des données inhérentes à la carrière des assurés du régime général, y compris les données concernant d'éventuelles périodes effectuées dans d'autres

régimes de base (MSA, Cancava, Organic), ainsi que dans certains régimes particuliers ou spéciaux (SNCF, EDF-GDF, CNRACL, Mines).

De plus, pour le compte et sous le contrôle des organismes d'assurance maladie, la Cnav met en œuvre le Répertoire national inter-régimes des bénéficiaires de l'assurance maladie (RNIAM), qui est constitué, pour chaque bénéficiaire, en plus de son NIR et de son état civil, des informations de rattachement à l'organisme lui servant les prestations d'assurance maladie.

Pour la constitution et l'enrichissement de ces bases de données, la Cnav reçoit régulièrement des données en provenance de différentes sources. Les Déclarations annuelles de données sociales (DADS) sont transmises chaque année par les employeurs ayant un numéro Siret.

Les Données nominatives trimestrielles (DNT) sont transmises par les employeurs de personnel de maison. Les informations de périodes d'activité / non-activité des individus relevant de l'Unedic (chômage), de la CNAMTS (maladie), de la Caisse nationale d'allocations familiales (Cnaf, notamment les données sur la maternité), des régimes particuliers ou spéciaux (SNCF, EDF, RATP...), sont également transmises à la Cnav.

Cet ensemble de données est recueilli de façon prospective depuis 1995. Cependant, les données des autres régimes ne sont actuellement collectées que lorsque les sujets atteignent l'âge de 55 ans : c'est en effet à partir de cet âge que la vérification et le remplissage d'éventuelles absences d'informations débutent.

Les données issues de la Cnav sont pour des raisons évidentes (elles servent de base au calcul des retraites) complètes et particulièrement bien validées, notamment pour les périodes les plus récentes, et leur qualité (complétude et exactitude) s'améliore régulièrement au fil des années avec l'informatisation du recueil à la source.

Les bases de données de la Cnav peuvent grandement faciliter des opérations particulièrement lourdes et complexes, courantes dans de nombreuses enquêtes épidémiologiques, et dont l'efficacité est souvent médiocre. Elles permettent notamment le suivi des sujets : tous les épisodes socioprofessionnels de la quasi-totalité des personnes vivant en France étant enregistrés de façon prospective et détaillée, il est théoriquement possible de suivre les personnes incluses dans un protocole longitudinal tout au long de leur vie et de minimiser ainsi le taux de perdus de vue.

Les bases de données de la Cnav peuvent également être utilisées pour l'accès aux données socioprofessionnelles : certains domaines de l'épidémiologie, notamment l'épidémiologie des risques professionnels et des déterminants sociaux de la santé, s'intéressent particulièrement au statut social et professionnel et à son évolution dans le temps. Les données enregistrées dans les bases de la Cnav sont particulièrement riches de ce point de vue et susceptibles d'intéresser différentes équipes d'épidémiologistes, aussi bien pour sélectionner, sur des critères socioprofessionnels, des sujets à inclure dans des études de méthodologie variée (cas-témoins, cohorte, etc.), que pour avoir accès aux données socioprofessionnelles les concernant tout au long de suivis de longue durée.

Événements de santé

En ce qui concerne les données de mortalité : le statut vital et les causes de décès peuvent actuellement être obtenus selon la procédure décrite dans le décret n° 98-37 autorisant l'accès au RNIPP et à la base de données du Centre d'épidémiologie des causes de décès de l'Inserm (CépiDc), dans des conditions garantissant une confidentialité absolue des données à caractère personnel.

Concernant les données d'hospitalisation : le Programme de médicalisation du système d'information (PMSI) a pour objectif de produire des informations à contenu médical sur les fonc-

tions hospitalières et de permettre une allocation de ressources dépendant de l'activité hospitalière. Il consiste en un recueil exhaustif systématique et un traitement automatisé d'informations administratives et médicales soumises au secret médical. Les établissements transmettent les fichiers anonymisés à l'Agence régionale d'hospitalisation (ARH) et celles-ci les transmettent à l'Agence technique de l'information sur l'hospitalisation (ATIH), en vue de la constitution des bases de données nationales. Cette transmission se fait sous la forme de résumés de sortie anonymisés (RSA), qui contiennent essentiellement des données permettant l'identification du séjour, l'identification du patient par un numéro d'anonymat et des données médicales (diagnostic principal, ensemble des diagnostics associés et des actes pratiqués).

Enfin, parmi les données enregistrées par les systèmes d'information de l'assurance maladie, on distingue deux types de données : celles dites « de production », portant principalement sur les consommations de soins, et dont l'objectif premier est la liquidation des prestations d'assurance maladie ; et celles « de référentiels », qui concernent les informations sur les assurés, les établissements de santé et les professionnels de santé.

Par ailleurs, les services médicaux disposent de leurs propres fichiers comportant des informations médicales sur les affections de longue durée (ALD), les accidents du travail et les maladies professionnelles, et dont l'objectif initial est le contrôle, par les médecins conseil, des pathologies ouvrant droit à une prestation.

Toutes ces données sont désormais rassemblées au sein du Système national d'information inter-régimes de l'assurance maladie (SNIIR-AM). Les données du SNIIR-AM incluent tous les régimes de l'assurance maladie : CNAMTS, MSA, RSI et les 16 autres régimes spéciaux, et concernent aussi bien la médecine de ville que les hospitalisations. Le SNIIR-AM a pour objet la connaissance des dépenses de l'ensemble des régimes de l'assurance maladie, le retour de ces

informations auprès des professionnels de santé (informations pertinentes relatives à leur activité, leurs recettes, et s'il y a lieu, à leurs prescriptions), la définition, le suivi et l'évaluation des politiques de santé publique (loi de santé publique du 13 août 2004).

Le SNIIR-AM peut constituer une solution particulièrement efficace pour l'accès à des données individuelles permettant de réaliser des enquêtes épidémiologiques. Il s'agit en effet d'une base de données individuelles et anonymes conservées pendant deux ans au-delà de l'année en cours. L'anonymisation des variables identifiantes est réalisée par le module FOIN (fonction d'occultation des informations nominatives) [Quantin *et al.*, 2005 ; Lenormand, 2005]. Cette fonction repose sur le NIR de l'ouvrant-droit, la date de naissance et le sexe du bénéficiaire. Les données sont anonymisées en deux étapes : au niveau local, créant un numéro d'anonymat réversible (algorithme FOIN-1) ; puis au niveau national, créant un identifiant anonyme irréversible (algorithme FOIN-2) : à partir de cet identifiant, on ne peut pas retrouver les données nominatives qui ont servi au calcul.



Logo du Programme de médicalisation du système d'information - PMSI

Principaux problèmes méthodologiques

Bien que l'ensemble des bases de données citées n'ait pas fait l'objet d'analyses systématiques de validité, quelques études plus ou moins ponctuelles ont porté sur les données issues des différents fichiers.

L'utilisation du PMSI comme source d'information sur les pathologies s'avère délicate et ne peut reposer uniquement sur le diagnostic principal [Couris *et al.*, 2002, Couris *et al.*, 2007]. Il est nécessaire de développer des algorithmes plus complexes alliant les codes des diagnostics aux codes des actes spécifiques de la pathologie étudiée. Par ailleurs, lorsque la base du PMSI est utilisée pour estimer l'incidence d'une pathologie, il faut exclure les cas prévalents¹⁰ par la recherche de la pathologie dans les bases PMSI les années antérieures [Quantin *et al.*, 2000] ; cependant, ce problème ne se pose pas lorsque le PMSI est utilisé pour sélectionner des sujets à inclure en fonction d'une pathologie [Geoffroy-Perez *et al.*, 2006] ou pour connaître l'occurrence des pathologies dans le cadre du suivi longitudinal d'une cohorte [Borella *et al.*, 2000 ; Laroche *et al.*, 2002].

L'intérêt potentiel des bases de données de l'assurance maladie dans une optique épidémiologique apparaît clairement dans la mesure où elles fournissent des données individuelles médicalisées, structurées et codées de manière standardisée [Fender & Weill, 2004]. Leur utilisation dans une optique épidémiologique nécessite cependant un important travail de réflexion méthodologique [Quantin *et al.*, 1999], de contrôle et de validation de données. Ainsi, la base des ALD codées par des médecins reste une base de données à vocation médico-sociale [CNAMTS, 2004] et ses limites sont connues : imprécision des diagnostics, absence d'exhaustivité des cas déclarés, risque de double déclaration [Chinaud *et al.*, 2004]. La prévalence des affections classées en ALD est systématiquement inférieure à la prévalence réelle, car le patient peut être atteint de l'une de ces maladies, mais ne pas répondre aux critères de sévérité ou d'évolutivité exigés ; il peut déjà être exonéré du ticket modérateur à un autre titre (autre ALD, invalidité), ou il peut ne

10. L'incidence concerne les nouveaux cas de maladie apparus pendant une période donnée, alors que la prévalence concerne l'ensemble des cas pendant cette période (cas préexistants et nouveaux cas).

pas demander à être exonéré pour des raisons personnelles (assurance complémentaire satisfaisante, souci de confidentialité...).

Par ailleurs, la qualité du codage n'a jamais été évaluée, à notre connaissance. La base de données de remboursements de l'assurance maladie est adaptée aux objectifs d'analyse des pratiques de prescription [Deprez *et al.*, 2004], d'évaluation de l'impact de campagne d'information [Lecadet *et al.*, 2004]. En revanche, elle ne comporte pas d'information sur la nature des maladies traitées et exclut, par définition, l'automédication et les prestations non présentées au remboursement.

L'identification des sujets dans les bases de données recouvre deux aspects distincts : identifier, dans une base de données, des individus connus dans une autre base (appariement de bases de données, notamment pour les suivis de cohorte) ; identifier des individus dans une base de données pour les études d'incidence, de prévalence, la constitution d'études cas-témoins ou de cohorte.

Dans la première situation, il s'agit d'apparier au moins deux bases de données, puis d'interpréter les informations retrouvées. Dans la seconde, la principale stratégie est de définir un algorithme d'identification fiable des patients à identifier, qui peut être constitué à partir de diagnostics, d'actes ou de consommation de médicaments selon la base de données utilisée. La qualité d'un appariement quel qu'il soit se mesure par le nombre de collisions (même identifiant pour des personnes différentes) et de doublons (plusieurs identifiants pour une même personne) [Quantin *et al.*, 2005].

Les doublons et collisions peuvent exister dans une même base de données, ou peuvent être générés lors de l'appariement de deux ou plusieurs bases. Les collisions ou les doublons sont de nature à provoquer des erreurs de classement et donc, au mieux, une perte de puissance, au pire, un biais de classement si les erreurs sont liées à la maladie ou à une autre variable d'intérêt.

À l'heure actuelle, le seul identifiant stable et qui présente les meilleures qualités est le NIR. Pour garantir la confidentialité des données individuelles, l'identifiant encrypté¹¹ par la fonction FOIN est actuellement utilisé dans les bases de données médico-administratives. Si, pour chaque individu, les éléments permettant de reconstituer son identifiant FOIN [Quantin *et al.*, 2005 ; Lenormand, 2005] sont disponibles, il est alors possible de retrouver, pour une personne donnée, les enregistrements de données le concernant, y compris dans le SNIIR-AM. Des erreurs peuvent néanmoins se produire à divers niveaux limitant l'identification des sujets par l'identifiant FOIN, et il faut alors envisager des solutions d'appariement probabiliste [Jaro, 1995 ; Quantin *et al.*, 2005] à partir des autres variables disponibles dans ces bases de données.

Cependant, en l'état actuel des textes, l'utilisation directe du NIR est interdite aux épidémiologistes, seuls certains organismes ayant l'autorisation d'en avoir connaissance (notamment les organismes de protection sociale). Or, les données individuelles contenues dans les bases médicales du SNIIR-AM ne sont accessibles que *via* le numéro FOIN pour le calcul duquel le NIR est nécessaire. Lorsqu'il est possible d'inclure dans une enquête des sujets à partir d'un organisme détenteur du NIR (comme dans l'exemple de Constances présenté plus loin), cette difficulté peut être résolue et il est possible, comme on le verra, d'élaborer des procédures où le NIR n'est jamais connu des épidémiologistes. Mais cette situation n'est pas courante et ne peut être utilisée lorsque les participants d'une enquête épidémiologique sont recru-

tés à partir d'autres sources comme c'est le plus souvent la règle : il est alors impossible de recueillir le NIR, même avec l'accord des sujets, et donc de suivre ces derniers dans les bases du SNIIR-AM. Ce problème est bien connu des épidémiologistes et de la Commission nationale de l'informatique et des libertés (Cnil). Des solutions sont actuellement recherchées, qui pourraient reposer sur la mise en place d'un « identifiant national de santé » [Gensbittel *et al.*, 2007].

Éthique & confidentialité

L'utilisation des bases de données citées précédemment pose des problèmes relevant de l'éthique de la recherche. Cette utilisation nécessite en cela des dispositifs destinés à garantir aux données à caractère personnel une confidentialité conforme aux textes et suffisante pour être acceptable par la société.

On peut tout d'abord rappeler que, du fait du caractère particulièrement sensible des données qu'ils utilisent, les épidémiologistes voient leur activité particulièrement encadrée. Ainsi, dans le cadre de la loi Informatique et libertés, il existe un dispositif spécifique réglementant le traitement de l'information en matière de recherches dans le domaine de la santé, et tout projet d'enquête épidémiologique doit obtenir, préalablement à son examen par la Cnil, l'aval du Comité consultatif sur le traitement de l'information en matière de recherches dans le domaine de la santé (CCTIRS)¹². Certaines enquêtes épidémiologiques, notamment celles qui s'accompagnent des gestes invasifs¹³ (examen radiologique, par exemple), nécessitent l'accord préalable d'un comité de protection des personnes. Enfin, dans certaines situations complexes sur le plan de l'éthique, il est également possible de saisir le Comité consultatif national d'éthique pour les sciences de la vie et de la santé (CCNE). Il faut aussi souligner que la communauté des épidémiologistes s'est dotée, dès 1998, de recommandations de déontologie et bonnes pratiques en épidémiologie qui ont fait récemment l'objet d'une révision¹⁴.

11. L'encryptage est une procédure de chiffrement permettant de protéger un message ou des données qu'on souhaite garder confidentiels.

12. Loi 78-17 modifiée et décret n° 95-682 pris pour application du chapitre V bis de cette loi et modifiant lui-même le décret n° 78-774 du 17 juillet 1978.

13. Se dit d'une méthode d'exploration ou de soin nécessitant une lésion de l'organisme ou susceptible d'entraîner un risque pour la santé [NDLR].

14. Téléchargeable sur : <http://adelf.isped.u-bordeaux2.fr> (rubrique Actualités).

Sur le plan technique, l'utilisation des bases de données médico-administratives présente des difficultés particulières concernant la confidentialité des données à caractère personnel. À cet égard, le problème essentiel est celui des identifiants utilisables. Ce point a été traité en détail dans un numéro récent du *Courrier des statistiques* et nous n'y reviendrons pas ici [Gensbittel *et al.*, 2007]. D'autres techniques plus récentes viennent d'être proposées [Quantin *et al.*, 2008]. À titre d'illustration des solutions qui peuvent être mises en œuvre pour assurer une stricte confidentialité, tout en permettant un large accès aux bases de données médico-administratives, on résumera (en le simplifiant) le dispositif prévu pour la cohorte Constances déjà citée¹⁵. Les procédures comportent notamment un tirage au sort des participants potentiels, une inclusion dans un centre d'examen de santé (CES) de la sécurité sociale permettant le recueil de données médicalisées, et un suivi des volontaires, par questionnaire postal et dans les bases de données de la Cnav et du SNIIR-AM ; une cohorte de non-répondants doit également faire l'objet d'un suivi anonyme dans les bases de la Cnav et du SNIIR-AM.

Constitution de la cohorte

La Cnav procède à un tirage au sort à probabilités inégales selon des variables usuelles (sexe, âge, PCS) des sujets éligibles¹⁶. À cette fin, elle extrait du RNIAM les NIR des affiliés éligibles, qui sont ensuite appariés aux informations incluses dans d'autres fichiers sur la PCS actuelle pour les actifs, et sur la dernière PCS occupée pour les autres (retraités, chômeurs, inactifs). Chaque sujet tiré au sort se verra attribuer un « numéro Constances » ($N_{\text{Constances}}$). La correspondance $\text{NIR}_{\text{individuel}} - N_{\text{Constances}}$ sera conservée par la Cnav pour les échanges de données ultérieurs. Pour chaque sujet tiré au sort, la Cnav adresse cette table de correspondance à la CNAMTS (structure gérant le SNIIR-AM), ainsi que des données d'état civil (nom patronymique, prénom, date de naissance, sexe, lieu de naissance) nécessaires pour



Entrée et porche du bâtiment de la Cnav, 8 rue Vivienne, Paris 2^e



l'algorithme de calcul de l'identifiant crypté FOIN. La Cnav transmet également les $N_{\text{Constances}}$ et les données concernant la date de naissance et le sexe à un « Tiers de confiance » agréé par la Cnav.

La CNAMTS conserve les informations nécessaires à la génération des numéros FOIN associés aux $N_{\text{Constances}}$. Afin de pouvoir adresser les courriers d'invitation à participer aux sujets tirés au sort, elle adresse également la correspondance $\text{NIR}_{\text{individuel}} - N_{\text{Constances}}$, ainsi que les données d'état civil, aux Centres de traitement informatique (CTI) interrégionaux de la CNAMTS, qui disposent de l'adresse postale des sujets. Les CTI transmettent alors les fichiers d'adresses aux CES, pour envoi des invitations.

Informations préalables et inclusion des participants

Les participants de la cohorte seront invités par courrier : les personnes sollicitées recevront une plaquette de présentation de Constances et un coupon-réponse leur permettant de donner leur accord pour participer à la cohorte. Comme il est prévu de mettre en place un échantillon de non-participants suivi de façon totalement anonyme dans les bases de données nationales, les personnes

tirées au sort seront informées, lors de l'invitation, de la possibilité de faire partie de cet échantillon ; elles pourront explicitement refuser ce suivi en remplissant le coupon-réponse accompagnant l'invitation et seront alors exclues de la population des non-participants susceptibles d'être tirés au sort.

Lors de leur venue dans le CES, les personnes volontaires rempliront un consentement éclairé permettant d'exprimer l'acceptation ou le refus de transmission de tout ou partie des données les concernant, notamment dans le cadre du suivi : données du SNIIR-AM (consommations de médicaments, actes de laboratoires, hospitalisations, consultations de généralistes, consultations de spécialistes), données socioprofessionnelles de la Cnav, adresses postales de la Poste, données des causes de décès du CépiDC de l'Inserm. Une étiquette avec un code à barres correspondant au $N_{\text{Constances}}$ qui leur aura été attribué, sera apposée sur chaque document (consentement, questionnaires, données de l'examen de santé).

15. Au moment de la rédaction de cet article, le dossier Constances est en cours d'examen à la Cnav.

16. La population source de Constances est constituée des affiliés du Régime général de sécurité sociale âgés de 18 à 69 ans, résidant dans certains départements français.

Circuit de transmission des données

Deux types de données seront recueillis : des données nominatives et des données anonymisées, pour lesquelles des flux indépendants seront mis en place.

Dans le cas des données nominatives : après l'inclusion, les données d'état civil (nom patronymique, prénom, date de naissance, sexe, lieu de naissance) et l'adresse postale, associées aux $N_{\text{Constances}}$, seront transmises au Tiers de confiance et conservées sur un ordinateur indépendant de tout réseau (internet, intranet...). Le Tiers de confiance sera en mesure d'identifier par différentiel, après un délai de 3 mois, les non-participants, parmi lesquels un échantillon de non-participants sera tiré au sort pour faire l'objet d'un suivi dans les bases de données nationales.

Dans le cas des données anonymisées : les données recueillies dans les CES, identifiées par le $N_{\text{Constances}}$, seront adressées à l'équipe Constances. Ces données seront alors associées à un « numéro de cohortiste permanent » (NCP) et intégrées dans les bases de données de la cohorte, et accessibles uniquement aux personnes habilitées. La correspondance NCP- $N_{\text{Constances}}$ sera gérée par l'équipe Constances. De plus, le Tiers de confiance adressera à l'équipe Constances les informations suivantes : $N_{\text{Constances}}$, date de naissance, sexe, PCS et l'information quant à la participation éventuelle de chaque sujet.

Après les vagues d'inclusion, la Cnav et la CNAMTS détruiront les informations ne concernant ni la cohorte de participants ni la cohorte de non-participants. Le Tiers de confiance détruira toutes les informations ne concernant pas les participants (y compris les informations de sujets appartenant à la cohorte de non-participants, dont le suivi doit être strictement anonyme) ; ainsi, après l'inclusion, le Tiers de confiance disposera uniquement d'informations concernant les participants.

Le projet de plate-forme Plastico

Actuellement, la France dispose, comme on l'a vu, de bases de données médico-administratives couvrant la quasi-totalité de la population nationale (et ce, contrairement à la plupart des autres pays) ; de surcroît, ces bases de données sont centralisées et structurées. Mais notre pays se caractérise aussi par une sous-utilisation de cette ressource particulièrement importante.

Certes, comme l'exemple du projet Constances l'illustre, les procédures permettant l'accès aux bases de données médico-administratives nationales sont complexes et nécessitent des moyens importants, mobilisables sur une très longue durée. Il existe en outre divers problèmes légaux, mais ils pourraient être résolus.

L'absence de dispositifs destinés à surmonter les importantes difficultés scientifiques, techniques et logistiques inhérentes à ce type d'activités (qui nécessitent des moyens dont aucune équipe d'épidémiologie ne dispose en France) est donc, à l'évidence, la raison principale de la faible utilisation actuelle des dispositifs informationnels disponibles.

Dans ce contexte, le projet de création d'une plate-forme scientifique et technique pour l'aide à la gestion de cohortes et de grandes enquêtes (projet « Plastico ») correspond aux besoins de nombreuses équipes menant de grandes enquêtes relevant de divers organismes. La plate-forme Plastico, ouverte à la communauté scientifique spécialisée, repose sur la mutualisation de moyens sous la forme d'une plate-forme scientifique et technique : ressources méthodologiques et outils de recueil de données, compétences de natures diverses, développement d'activités partagées et de synergies scientifiques. Répondant à la nécessité de structures pérennes pour des opérations de très longue durée, elle devrait offrir des prestations diverses, dans des conditions de fonctionnement apportant des solutions à certaines des difficultés évoquées, tout en induisant d'importantes économies d'échelle.

Le projet Plastico a pour objectif de constituer une plate-forme limitée à l'aide à la réalisation d'enquêtes concernant des individus et impliquant l'utilisation de données à caractère personnel. Le rôle de la plate-forme Plastico concernant l'accès aux données individuelles des bases de données nationales est de servir d'interface entre les enquêtes épidémiologiques et les bases de données : mise au point avec les responsables concernés des modalités opérationnelles d'accès et de transfert de données ; réception des résultats des opérations informatiques de recherche des sujets et d'extraction des données individuelles ; contrôles de premier niveau (complétude, absence de doublons...) ; transmission sécurisée. La plate-forme pourra prendre en charge cette activité dans des conditions de qualité et de sécurité difficiles à réunir au sein de chaque équipe concernée, en raison des ressources nécessaires, de la compétence et de l'expérience du personnel spécialisé dans ces opérations.

Une autre fonction de Plastico est, pour les enquêtes qui font appel à plusieurs sources informatisées de données, d'assurer l'appariement de données individuelles issues de bases de données différentes. En effet, du fait des contraintes très fortes de sécurité et de confidentialité attachées au traitement des données à caractère personnel, et des restrictions pour l'utilisation du NIR, le croisement de données individuelles provenant de plusieurs bases de données est une opération complexe et particulièrement lourde. Plastico, structure inter-organismes, en relation permanente avec les organismes gérant les bases de données nationales, devrait disposer de moyens techniques importants et de personnel spécialisé de haut niveau, jouera le rôle d'un « Centre d'appariement sécurisé » [Quantin et al., 2008]. Les aspects légaux et réglementaires et les modalités de fonctionnement d'une telle fonction sont en cours de définition en relation avec la Cnil.

L'utilisation des données de morbidité extraites de bases de données nationales, comme le PMSI ou les

ALD, ne permet pas d'obtenir des diagnostics suffisamment fiables et précis par référence aux contraintes épidémiologiques.

Dans de nombreuses situations, il est donc nécessaire de mettre en place des procédures de validation des diagnostics extraits des bases de données. Dans le cadre de Plastico, un réseau de recherche, associant des épidémiologistes et des professionnels de l'information médicale des départements d'information médicale des hôpitaux qui recueillent les données de base du PMSI, a été constitué. Les objectifs principaux de ce réseau sont de développer des méthodes d'aide à la validation pour des diagnostics de pathologie issus des bases de données médicales nationales. Il n'existe cependant pas de méthode « générale » pour un tel objectif.

Une voie prometteuse est le développement d'algorithmes alliant un diagnostic à des actes médico-techniques, à des consommations de médicaments plus ou moins spécifiques de la pathologie concernée, etc. L'accès aux bases de données médicalisées citées ici rend possible de telles approches, dans la mesure où elles contiennent des données à la fois sur les diagnostics et sur les actes médico-techniques et les médicaments prescrits. Bien entendu, de tels algorithmes sont spécifiques des pathologies ; ils doivent de plus être constamment mis à jour en fonction de l'évolution des techniques médicales et de l'introduction de nouveaux

médicaments. À l'heure actuelle, on ne dispose de pratiquement aucun algorithme validé qui pourrait être utilisé en routine.

Enfin, Plastico assurera également une fonction de saisie automatisée de questionnaires au bénéfice d'enquêtes épidémiologiques par des techniques de lecture automatisée de documents (LAD) grâce à un atelier de LAD.

L'importance des besoins nécessaires implique l'association de plusieurs organismes pour la création et la gestion de la plate-forme. Actuellement, la mise en place d'une structure de préfiguration de la plate-forme Plastico a été confiée par l'Institut de recherche en santé publique (IReSP) à l'Unité mixte Inserm-CNAMTS 687 pour une durée de deux ans (2008-2009). Les objectifs de la structure de préfiguration sont d'accumuler une expérience suffisante, à la fois sur le plan scientifique, technique et des ressources à réunir, et sur celui des modalités institutionnelles et de gestion.

Pour une meilleure utilisation des bases de données médico-administratives

L'épidémiologie se trouve aujourd'hui face à la nécessité de développer des études de taille autrefois inimaginable. Qu'il s'agisse de mettre en évidence des risques de faible ampleur associés à l'exposition à des agents potentiellement pathogènes, d'évaluer l'efficacité d'interventions dont on attend des

bénéfices d'ampleur modeste, ou de décrire la distribution et l'évolution d'événements peu fréquents, ce sont aujourd'hui des effectifs très importants qu'il faut réunir dans les études épidémiologiques. Des cohortes de centaines de milliers de sujets, suivis de façon prospective pendant des périodes qui s'étendent sur des décennies, sont indispensables pour répondre à des questions multiples.

Dans ce paysage, la France ne se distingue pas particulièrement par ses grandes réalisations, les cohortes prospectives françaises ayant une taille relativement faible comparative-ment à ce qui existe, parfois depuis des décennies, dans d'autres pays. La modestie des études françaises s'explique en partie par une sous-utilisation des possibilités offertes par les bases de données nationales alimentées par les organismes de protection sociale et médicale, qui offrent pourtant un intérêt potentiel majeur pour la réalisation d'études épidémiologiques [Goldberg, 2006].

Or, si l'on veut que la France se dote d'outils épidémiologiques d'envergure comparable à ce qui existe dans plusieurs pays, de très grandes enquêtes et de nouvelles cohortes sont indispensables, dont l'effectif ne se comptera plus en dizaines, mais en centaines de milliers de sujets. Ceci nécessite le développement de structures du type de celle proposée par le projet Plastico, indispensable pour assurer une interface entre les équipes d'épidémiologie et les grandes bases de données nationales. ■

Références bibliographiques

- Austin MA, Criqui MH, Barrett et al. The effect of response bias on the odds-ratio. *Am J Epidemiol*, 1981;114,137-143.
- Borella L et al. Un essai d'exploitation de la base PMSI nationale pour évaluer le volume et les modes de prise en charge du cancer en secteur hospitalier non lucratif. *Rev Epidemiol Sante Publique*, 2000, 48 : 53-70.
- Bouyer J, Hémon D, Cordier S et al. *Épidémiologie, principes et méthodes quantitatives*. Paris, Éditions Inserm, 1993.
- Chinaud F, Weill A, Ricordeau P, Fender P, Allemand H. Incidence du cancer du poumon en France métropolitaine de 1997 à 2002 : les données du régime général de l'assurance maladie. *Revue Médicale de l'Assurance Maladie* Avril-juin 2004, 35: 69-79.
- CNAMTS-DSM-Mission des Soins de ville-Mission Statistique. Incidence médico-sociale des ALD30 en 1999. Avril 2004. www.ameli.fr/245/doc/1391/article_pdf.html.
- Couris CM et al. Method of correction to assess the number of hospitalized incident breast cancer cases based on claims databases. *J Clin Epidemiol*, 2002, 55 : 386-391.

- Couris CM, Forêt Dodelin C, Rabilloud M, Colin C, Bobin JY, Dargent D, Raudran D, Schott AM Sensibilité et spécificité de deux méthodes d'identification des cancers du sein incidents dans les services spécialisés à partir des données médico-administratives. *Rev Epidemiol Sante Publique* 2004, 52, 151-60.
- Criqui MH. Response bias and risk ratios in epidemiologic studies. *Am J Epidemiol*, 1979;109:394-399.
- Deprez Ph-H, Chinaud F, Clech S, Germanaud J, Weill A, Cornille JL, Fender P. La population traitée par médicaments de la classe des antihistaminiques en France Métropolitaine : données du régime général de l'assurance maladie, 2000. *Revue Médicale de l'Assurance Maladie*, Avril-Juin 2004, 35: 3-11.
- Fender P, Weill A. Épidémiologie, santé publique et bases de données médico-tarifaire. (Éditorial) *Rev Epidemiol Santé Publique*, 2004, 52,113-117.
- Gensbittel MH, Riandey B, Quantin C. Appariements sécurisés : statisticiens, ayez de l'audace ! *Courrier des statistiques*, 2007, 121-122: 49-58.
- Geoffroy-Perez B, Imbernon E, Gilg Soit Ilg A, Goldberg M. Comparison of the French DRG based information system (PMSI) with the National Mesothelioma Surveillance Program database. [Article in French]. *Rev Epidemiol Sante Publique*. 2006 54:475-83.
- Goldberg M, Luce D. Les effets de sélection dans les cohortes épidémiologiques. Nature, causes et conséquences. *Rev Epidemiol Santé Publique* 2001;49:477-92.
- Goldberg M, Chastang JF, Leclerc A, Zins M, Bonenfant S, Bugel I, Kaniewski N, Schmaus A, Niedhammer I, Piciotti M, Chevalier A, Godard C, Imbernon E. Socioeconomic, demographic, occupational and health factors associated with participation in a long-term epidemiologic survey. A prospective study of the French Gazel cohort and its target population. *Am J Epidemiol* 2001;154:373-84.
- Goldberg M, Melchior M, Leclerc A, Lert F. Épidémiologie et déterminants des inégalités sociales de santé. Apports récents et problèmes actuels. *Rev Epidemiol Santé Publique* 2003 ;51:381-401.
- Goldberg M. Les bases de données d'origine administrative peuvent-elles être utiles pour l'épidémiologie ? *Rev Epidemiol Sante Publique*, 2006, 54: 297-303.
- Goldberg M, Chastang JF, Zins M, Niedhammer I, Leclerc A. Attrition during follow-up: health problems are the strongest predictors. A Study of the Gazel Cohort. *J Clin Epid*. 2006, 59;1213-1221.
- Goldberg M, Leclerc A, Bonenfant S, Chastang JF, Schmaus A, Kaniewski N, Zins M. Cohort profile: the GAZEL Cohort Study. *Int J Epid* 2007, 36:32-39.
- Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol*, 1977;106,184-187.
- Hernan AH, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*, 2004b, 15:615-625.
- Jaro M. Probabilistic linkage of large public health data files. *Stat Med* 1995;14:491-8.
- Jones AM, Koolman X, Rice N. Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighted estimators in non-linear models. *J Royal Stat Society, Series A*. 2006;169:1-27.
- Laroche ML et al. Qualité des données PMSI au CHU de Limoges : application de la méthode L.Q.A.S. *Rev Epidemiol Sante Publique*, 2002, 50 : 433-439.
- Lecadet J, Vialaret K, Vidal P, Baris B, Fender P. Mesure à l'échelle d'une région des effets d'un programme national d'information sur le bon usage des antibiotiques. *Revue Médicale de l'Assurance Maladie* Avril-Juin 2004, 35: 81-91.
- Lenormand F. Le système d'information de l'assurance maladie. *Courrier des Statistiques* 2005 ;113-114 :33-51.
- Little RJ and Vartivarian S. On weighting the rates in non réponse weights. *Statistics in Médecine*, 2003, 22:1589-1599.
- Quantin C., Sauleau E., Bolard P., Mousson C., Kerkri M., Brunet-Lecomte P., Moreau T., Dusserre L. Modeling of high cost patient distribution within renal failure diagnosis related group. *Journal Clin Epidemiology*. 1999;52:251-258.
- Quantin C., Allaert F.A., Bouzelat H., Rodrigues J.M., Trombert-Paviot B., Brunet-Lecomte P., Gremy F., Dusserre L. La sécurité des réseaux d'informations médicales : application aux études épidémiologiques. *Rev Epidémiol et Santé Publ*. 2000 ; 48:89-99.
- Quantin C, Gouyon B, Allaert FA, Cohen O. Méthodologie pour le chaînage de données sensibles tout en respectant l'anonymat : application au suivi des informations médicales. *Courrier des Statistiques* 2005 ; 113-114 :15-26.
- Quantin C, Biquet C, Allaert FA, Cornet B, Pattisina R, Le Teuff G, Ferdynus C, Gouyon JB. Decision analysis for the assessment of a record linkage procedure: application to a perinatal network. *Meth Inf Med*, 2005;44:72-79.

Le métier de statisticien en dehors du système statistique public

Le Centre de données astronomiques de Strasbourg – Interview de Françoise Genova

Courrier des statistiques : Françoise Genova, quel est votre parcours professionnel ?

Françoise Genova : J'ai commencé à étudier l'astronomie lorsque j'étais à l'École normale supérieure de Fontenay – une évolution à partir de mon domaine d'origine, les mathématiques, vers une discipline plus liée à la physique. J'ai alors travaillé dans un laboratoire de l'Observatoire de Paris pendant plus de dix ans : j'y ai obtenu une thèse de troisième cycle, un poste au CNRS¹ (en 1981), puis une thèse d'État. J'ai ensuite passé un peu plus de trois ans au Centre national d'études spatiales (CNES), comme « ingénieur thématique » en charge des expériences françaises spatiales d'astronomie. J'ai rejoint l'Observatoire astronomique de Strasbourg en 1993, et j'ai pris la direction du Centre de données astronomiques (CDS) en 1995.



Source : Observatoire de Strasbourg

Françoise Genova

CdS : En quoi consiste votre travail actuel ?

FG : Le centre que je dirige produit et diffuse des services en ligne largement utilisés par la communauté astronomique mondiale. Le CDS a été créé en 1972 par le prédécesseur de l'Institut national des sciences de l'univers (INSU), qui s'appelait l'Institut national d'astronomie et de géophysique (INAG).

La mission qui lui a été donnée depuis l'origine était très en avance pour l'époque et mérite sans doute d'être décrite en détail :

- collecter l'information « utile » sur les objets astronomiques (encadré), sous forme électronique ;
- l'améliorer par une évaluation critique et en combinant les informations ;
- distribuer les résultats à la communauté internationale ;
- produire des recherches utilisant ces données.

Les mots clés de l'activité du CDS, qui continuent à guider quotidiennement nos activités, se retrouvent dès cette charte initiale : construire des outils pour la recherche ; prendre en charge l'information sous forme électronique ; construire une connaissance experte sur les données ; jouer un rôle international.

Nos services regroupent donc en particulier des informations sur les objets astronomiques. Je donnerai quelques détails par la suite sur trois d'entre eux : Simbad, la base de données de référence pour la nomenclature et la bibliographie des objets astronomiques (hors système solaire), le Dictionnaire de nomenclature des objets astronomiques, et VizieR, service en ligne donnant accès aux catalogues astronomiques et aux tables de données publiées dans les journaux académiques d'astronomie.

Encadré : qu'est-ce qu'un objet astronomique ?

Un « objet astronomique » est, en jargon d'astronome, un objet qu'on observe dans le ciel (directement, ou indirectement par ses effets sur son environnement), ou un ensemble d'objets qui forment un système que l'on peut décrire et étudier en tant que tel.

Certains objets se trouvent dans le système solaire : celui-ci, qui est l'ensemble des objets liés au Soleil, est un objet astronomique, le Soleil lui-même (qui est une étoile plutôt banale) en est un aussi, de même que les planètes – dont la Terre –, les satellites des planètes – y compris la Lune –, les comètes et les astéroïdes.

Les étoiles, les nuages de gaz et de poussière qui appartiennent à notre galaxie ou à d'autres galaxies sont également des objets astronomiques, ainsi que les galaxies, les amas de galaxies, et même les trous noirs. En effet, les astronomes étudient comment se forment les trous noirs qui sont au centre de certaines galaxies et quel est leur rôle dans le comportement et l'évolution de ces dernières. L'anneau d'une planète, par exemple l'anneau de Saturne, est aussi un objet astronomique. Je ne suis pas sûre, en revanche, que chacune des poussières qui composent l'anneau puisse être considérée comme un « objet astronomique » !

1. Centre national de la recherche scientifique.

L'équipe du CDS regroupe des astronomes, des documentalistes et des ingénieurs informaticiens qui assurent les différents aspects du travail : d'une part, la construction du contenu des bases de données à partir de diverses sources (articles dans les journaux académiques, listes des observations conservées dans des archives, images de référence...); d'autre part, la programmation informatique nécessaire à la mise en place des bases de données, des pages web pour l'interrogation des bases, etc. Il faut aussi vérifier que les systèmes informatiques sont aussi fiables que possible pour assurer la continuité des opérations en cas de problème, une obligation que connaissent bien tous les responsables de centres informatiques : par exemple, il faut continuer à alimenter les machines en courant électrique en cas de panne de secteur.

On pourrait penser qu'il s'agit pour l'essentiel d'un travail de routine, mais en fait, l'un des défis auxquels nous sommes constamment confrontés – en plus du flux toujours croissant de l'information à traiter – est que nous travaillons dans un contexte en perpétuelle évolution.

C'est le cas, bien sûr, de notre champ scientifique : il faut se tenir au courant au jour le jour des avancées scientifiques en astronomie pour assurer la pertinence du contenu des bases de données et des fonctions offertes par les services : par exemple, lorsque la première planète extrasolaire² a été découverte, il a fallu créer un nouveau type d'objets dans Simbad, appelé en anglais 'Extra-solar Planet Candidate', et discuter de la manière dont il fallait faire apparaître ces objets dans les bases de données : *via* le nom de l'étoile autour de laquelle ces planètes orbitent, en tant qu'objets séparés, ou bien une combinaison des deux ?

C'est également le cas de la technique : elle aussi évolue très rapidement et il nous faut identifier les nouveautés

prometteuses et évaluer leur intérêt pour les services en ligne que nous proposons. Nous avons donc une activité significative de veille technologique, et de recherche et développement. Qu'on se souvienne par exemple de l'irruption d'internet au début des années 1990 : à l'époque, dès 1993, nous avons mis en ligne l'un de nos services, puis progressivement tous les autres. Il devenait beaucoup plus facile, en particulier, de construire des liens entre des services différents via les 'ancres' html. Nous en avons immédiatement tiré parti en collaborant avec notamment les journaux académiques, pour mettre en réseau les informations bibliographiques, les journaux électroniques, des bases de données, la base bibliographique de référence de la discipline qui est développée par la NASA, etc.

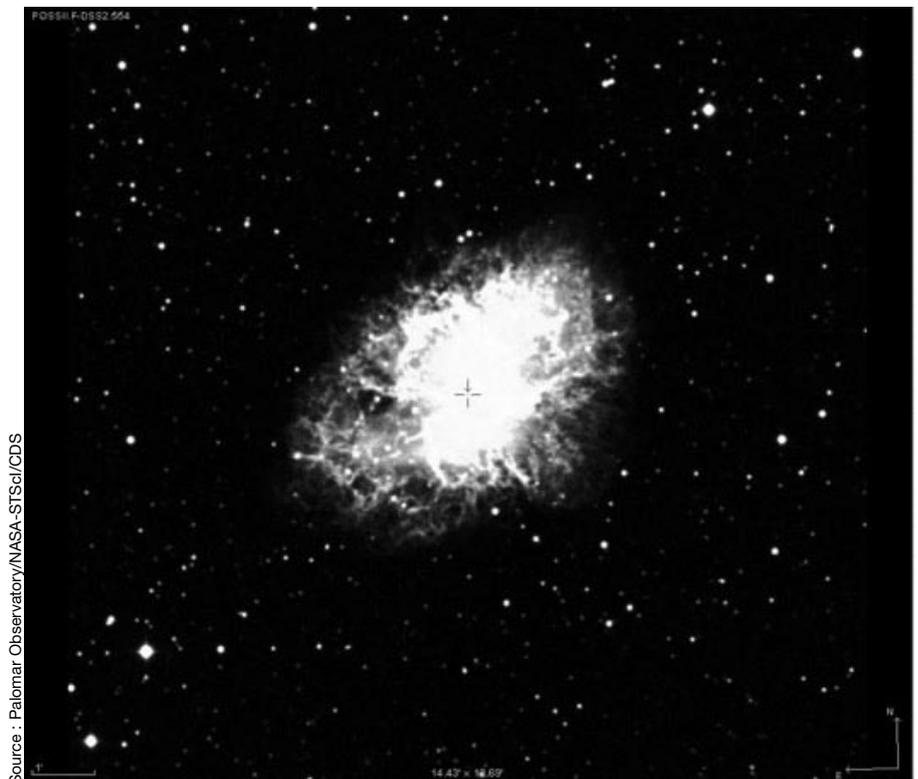
CdS : Pouvez-vous décrire les outils (nomenclatures, répertoires) que vous utilisez ?

FG : Nous développons un certain nombre d'outils pour la commu-

nauté astronomique, par exemple le Dictionnaire de nomenclature des objets astronomiques (hors système solaire), qui collecte les désignations d'objets astronomiques. En effet, un objet astronomique n'a pas nécessairement une désignation unique (sauf ceux du système solaire, dont le nom est attribué sous la responsabilité de l'Union astronomique internationale).

Très souvent, un nouveau catalogue d'objets ou un article assignera un nom qui lui est propre à chacun des objets qu'il répertorie. Par exemple, un objet très connu comme la Nébuleuse du Crabe, qui est ce qui reste d'une étoile ayant explosé il y a près d'un millénaire, en 1054, a 73 noms différents ! Tous les noms d'un objet sont répertoriés dans une de nos bases de données, Simbad, référence en matière de nomenclature et de bibliographie de nombreux objets astronomiques, en particulier ceux qui sont cités dans les articles scientifiques.

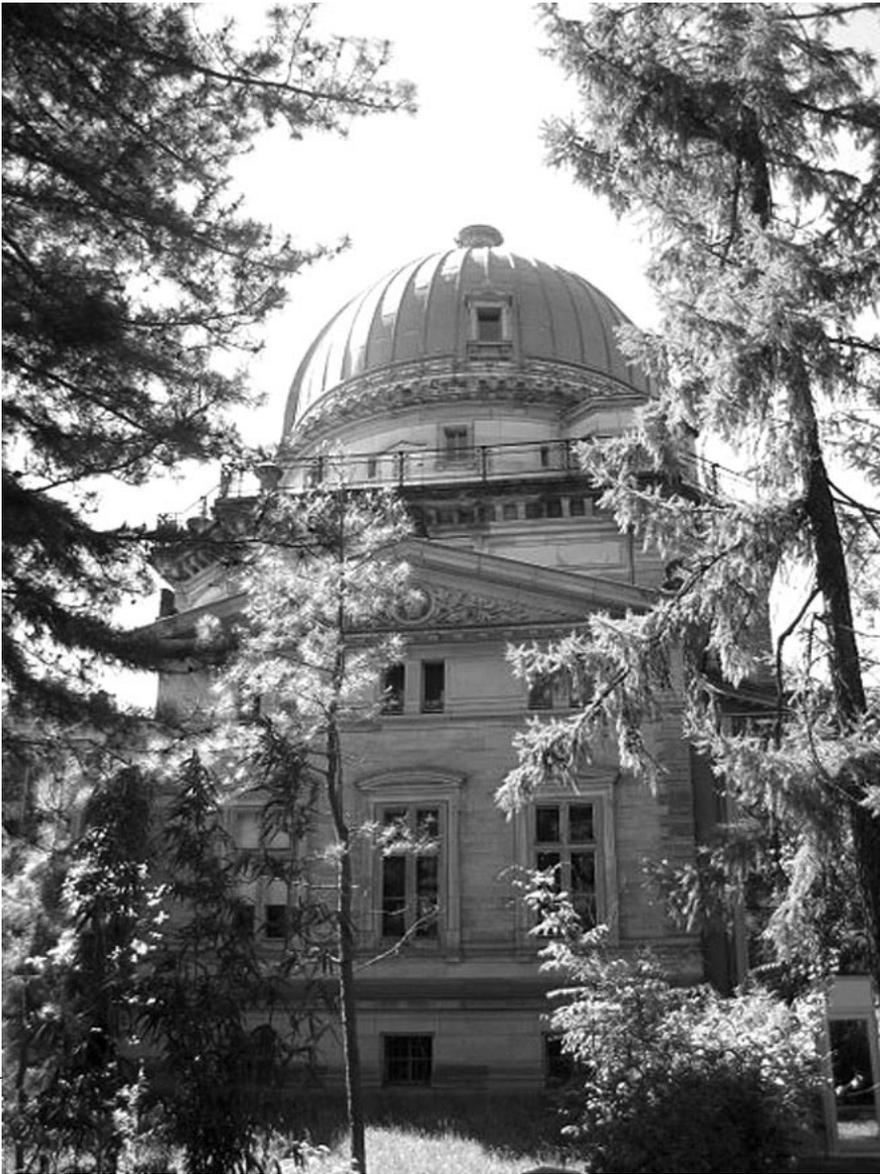
Malheureusement, les auteurs sont souvent peu soigneux lorsqu'il s'agit de la nomenclature des objets, et



Source : Palomar Observatory/NASA-STScI/CDS

Messier 1 (la Nébuleuse du Crabe). L'image provient d'une plaque photo de l'Observatoire du Mont Palomar, digitalisée par le Space Telescope Institute de la NASA pour le Digitized Sky Survey. Visualisation en fausses couleurs à l'aide du programme Aladin du CDS (CDS, CNRS/ULP)

2. C'est-à-dire gravitant autour d'une autre étoile que le Soleil [NDLR].



Source : Wikipédia

Observatoire d'astronomie de Strasbourg

ne se soucient guère de possibles ambiguïtés. Par exemple, toujours à propos de la Nébuleuse du Crabe, c'est le premier objet référencé dans le « Catalogue des nébuleuses et des amas d'étoiles » de Messier, publié en 1784, si bien que l'un de ses noms courants est « Messier 1 » ou « M 1 ». Mais on peut trouver, dans la littérature, des objets appelés « M » suivi d'un nombre, qui proviennent, non pas du catalogue de Messier, mais d'autres listes d'objets, elles-mêmes publiées dans des articles dont le nom du premier auteur commence par la lettre M !

Les astronomes lisant la littérature en question connaissent le contexte

scientifique et s'y retrouvent en général, mais on ne peut pas se permettre ce genre d'ambiguïté dans une base de données de référence. Dans Simbad, nous utilisons donc une désignation non ambiguë pour chaque catalogue ou liste d'objets publiée ; par exemple, le premier objet de la liste publiée en 1979 par Minniti, Alonso, Goudefrooij *et al.* s'appelle « [MAG95] 1 » dans Simbad, alors que cette liste est parfois appelée 'M'.

À la mi-avril 2008, la base de données Simbad donne des informations sur près de 4 millions d'objets. Ces derniers sont dits « cross-identifiés », au sens où, lorsque nous trouvons un objet cité dans un article et que

nous voulons l'ajouter à Simbad, nous cherchons d'abord s'il n'y figure pas déjà, sous le même nom ou sous d'autres. Si c'est le cas, on ne crée évidemment pas un nouvel objet dans la base, en revanche on ajoute la référence de l'article aux informations concernant cet objet, et éventuellement un nouveau nom ou de nouvelles données.

Beaucoup d'objets répertoriés dans des catalogues ne sont pas dans Simbad – l'une des raisons majeures étant que le travail de cross-identification prend beaucoup de temps et que nous ne sommes pas assez nombreux pour traiter tous les objets qui figurent dans les publications. En complément de Simbad, le CDS tient aussi à jour une collection aussi complète que possible de catalogues et de tables publiées dans les journaux académiques. Cette collection est la référence, au niveau international, pour ce type d'information.

Chaque catalogue de la collection comporte une ou plusieurs tables, ainsi que la description de ces dernières : organisation physique des tables, avec le nombre et le format de chacune des colonnes, et information sur leur contenu – par exemple, telle colonne contient une position dans tel système de coordonnées, telle autre renseigne sur l'abondance des éléments ou bien sur l'erreur portant sur une des mesures.

Les catalogues, sauf ceux qui sont très volumineux, sont disponibles en ligne. Les utilisateurs peuvent en télécharger une copie, ou utiliser le service Vizier, qui permet de sélectionner les informations contenues dans les catalogues en posant des contraintes sur n'importe lequel des champs des tables : par exemple, il est possible de retrouver tous les objets astronomiques dont, d'une part, on a mesuré s'ils contenaient du fer (et pour lesquels cette mesure est par exemple comprise entre deux valeurs données) et qui, d'autre part, sont situés dans une région donnée du ciel. Les plus grands catalogues de cette collection comptent plus d'un milliard d'objets. Pour ces derniers, les requêtes se font seulement 'par position' : l'utili-

sateur donne une position dans le ciel et récupère la liste de tous les objets proches de celle-ci.

Une différence importante entre Simbad et les catalogues est que ces derniers ne sont pas « cross-identifiés », c'est-à-dire que Vizier donne accès à tous les objets présents dans les catalogues dans une région du ciel, mais ne dit pas si un objet provenant d'un catalogue est le même qu'un autre provenant d'un autre catalogue – l'astronome qui utilise le service devra en décider lui-même, si c'est possible, d'après les informations disponibles.

Les services du CDS sont très utilisés par la communauté scientifique : en 2007, nous avons compté en moyenne 160 000 requêtes par jour ! Une partie de ces requêtes correspond à des questions posées par les utilisateurs *via* nos pages web, mais la plupart d'entre elles sont des interrogations d'autres services qui réutilisent nos données pour les besoins de leurs propres utilisateurs. Par exemple, les archives en ligne des observatoires astronomiques peuvent être interrogées en donnant sur leur page web le nom de l'objet au sujet duquel l'utilisateur souhaite obtenir des données. Le site de l'archive soumet une requête à Simbad, qui renvoie alors la position de l'objet ; cette dernière est ensuite le paramètre utilisé pour accéder à la base de données d'archive.

CdS : Comment les nomenclatures évoluent-elles ?

FG : J'ai expliqué ci-dessus que la nomenclature d'un objet hors système solaire est plutôt un phénomène cumulatif, au sens où les nouveaux catalogues ou les nouvelles listes d'objets peuvent introduire de nouveaux noms pour les objets. Les désignations données dans les catalogues permettent de « lire » certaines propriétés des objets, par exemple lorsqu'il s'agit de noms provenant de relevés du ciel dans une gamme de longueur d'onde donnée (ondes radio, infrarouge, rayons X, etc.).

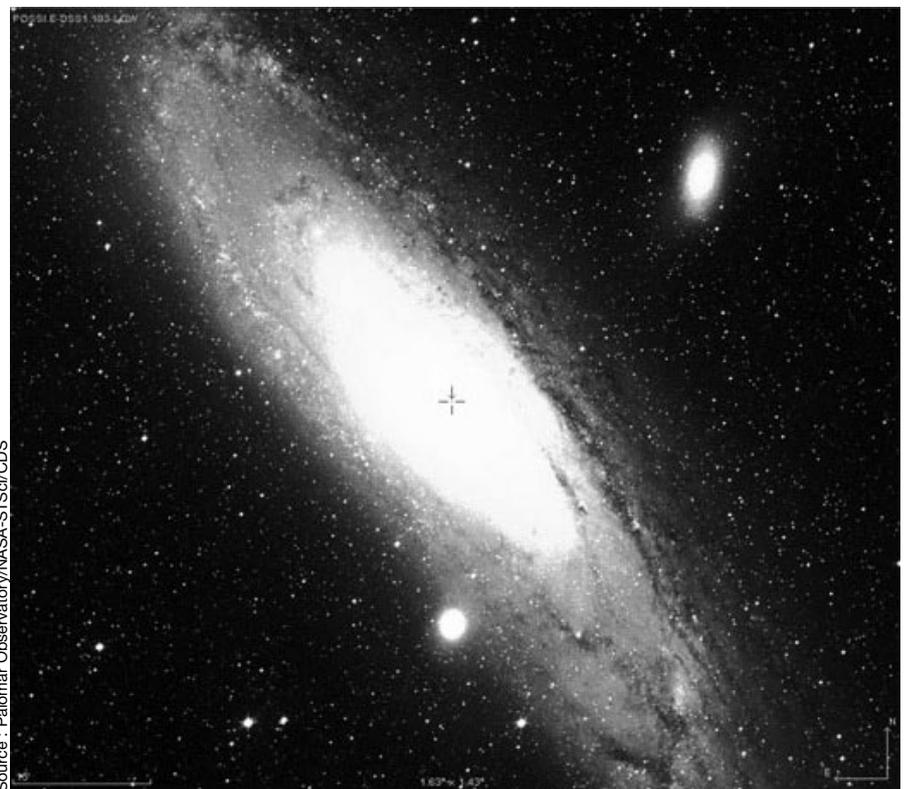
Pour les listes d'objets provenant des articles, on utilise le nom donné à la liste par l'auteur si celui-ci en a attribué un et que ce nom n'est pas ambigu, sinon on construit un nom à partir des initiales des trois premiers auteurs de l'article et de la date de publication, par exemple le nom [MAG95] cité précédemment. Le nom de catalogue ou de liste est suivi de caractères (chiffres dans une liste numérotée, lettre si c'est le format employé dans l'article, coordonnées du centre de l'objet, formats composites plus complexes). Par exemple, la liste des étoiles observées dans le champ d'un objet repéré par sa position J0558-6707 peut s'écrire [XYZ2001] J0558-6707 NNN, où NNN est le numéro d'ordre de l'étoile dans la liste concernant J0558-6707 qui figure dans l'article de X, Y et Z publié en 2001. Le ou les formats utilisés sont répertoriés dans le Dictionnaire de nomenclature.

Les auteurs peuvent d'ailleurs soumettre une nouvelle entrée pour le Dictionnaire de nomenclature avant

la publication de leurs articles (service mis en place par l'Union astronomique internationale, qui publie aussi des recommandations sur la meilleure manière de désigner des objets), mais seule une infime minorité d'entre eux prend la peine de le faire. Pour les autres articles, nous construisons un nom au moment de l'entrée dans Simbad comme je l'ai déjà expliqué.

CdS : Pour finir, quelques mots sur le projet d'« accès transparent à des données hétérogènes et distribuées » et l'Observatoire virtuel ?

FG : Les astronomes ont depuis longtemps l'habitude de réutiliser des données obtenues par d'autres. Au fil du temps, de nombreux centres de données et services d'archives d'observatoires astronomiques ont en effet mis à la disposition de la communauté scientifique des données et des services. L'utilisation d'informations en ligne est ainsi devenue une pratique de travail quotidienne de la communauté scientifique dans



Source : Palomar Observatory/NASA-STScI/CDS

Messier 31 (la galaxie d'Andromède). L'image provient d'une plaque photo de l'Observatoire du Mont Palomar, digitalisée par le Space Telescope Institute de la NASA pour le Digitized Sky Survey. Visualisation en fausses couleurs à l'aide du programme Aladin du CDS (CDS, CNRS/ULP)



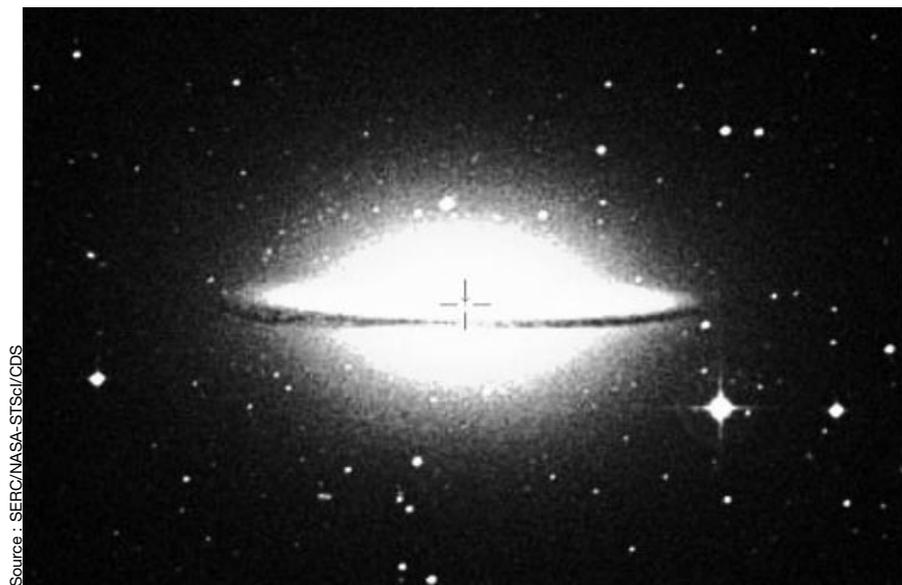
Source : Palomar Observatory/NASA-STScI/CDS

Messier 51 (les galaxies en interaction des Chiens de Chasse). L'image provient d'une plaque photo de l'Observatoire du Mont Palomar, digitalisée par le Space Telescope Institute de la NASA pour le Digitized Sky Survey. Visualisation en fausses couleurs à l'aide du programme Aladin du CDS (CDS, CNRS/ULP)

notre domaine. Par ailleurs, très vite après la généralisation d'internet, ces informations ont été mises en réseau : comme je l'ai expliqué précédemment, des liens entre archives d'observation, journaux électroniques et bases de données ont été construits, grâce à une collaboration étroite entre tous les acteurs, et on peut facilement naviguer d'un service à l'autre en cliquant sur des pages web.

Vers 2000, on a ensuite vu apparaître le concept d'« Observatoire virtuel astronomique » : l'objectif était d'aller plus loin, en donnant à tous les astronomes un accès à toutes les données de façon transparente – en particulier sans avoir besoin d'apprendre à utiliser l'interface propre à chaque service. Des portails permettent par exemple de visualiser toutes les données disponibles sur un objet ou dans une région du ciel, de les comparer, de les combiner.

C'est ainsi que l'*International Virtual Observatory Alliance* (OV pour Observatoire virtuel en français) regroupe tous les projets nationaux qui participent au développement de l'OV (17 partenaires, dont la France et plusieurs autres pays européens, ainsi que le Canada, la Chine, les États-Unis, l'Inde et le Japon), afin de définir les standards communs indispensables pour assurer la transparence de l'accès aux données. Les centres de données souhaitant inclure leurs données dans l'OV doivent simplement enregistrer leurs services dans un de ses répertoires et mettre en œuvre une « couche d'interopérabilité » qui permet d'explorer les données disponibles, d'y accéder dans le langage de requête commun, etc. ■



Source : SERC/NASA-STScI/CDS

Messier 104 (la galaxie du Sombrero, qui est vue par la tranche). L'image provient d'une plaque photo du SERC, digitalisée par le Space Telescope Institute de la NASA pour le Digitized Sky Survey. Visualisation en fausses couleurs à l'aide du programme Aladin du CDS (CDS, CNRS/ULP)

Le métier de statisticien en dehors du système statistique public

Les techniques statistiques utilisées en astronomie – Interview de François Mignard

Courrier des statistiques : François Mignard, quel est votre parcours professionnel ?

François Mignard : Ancien élève de l'École normale supérieure de la rue d'Ulm, je suis chercheur au CNRS depuis 1975 et actuellement directeur de recherche à l'observatoire de la Côte d'Azur, au sein du laboratoire Cassiopée. Mes domaines d'activités sont l'astronomie fondamentale, la mécanique céleste, la physique du système solaire et les tests de physique fondamentale avec les techniques astronomiques.

CdS : En quoi consiste votre travail actuel ?

FM : Je suis responsable du consortium européen en charge du traitement des données de la mission d'astronomie spatiale Gaia, en développement à l'Agence spatiale européenne (ESA en anglais). Cette mission doit effectuer un relevé complet du ciel avec des mesures de position, de distance, de vitesse radiale et de luminosité d'environ un milliard d'objets astronomiques, essentiellement des étoiles. L'objectif principal de la mission concerne les étoiles de la Voie lactée, mais les observations vont également inclure près d'un million d'objets en dehors de notre galaxie, ainsi que 300 000 petits corps du système solaire. Ces derniers, plus communément appelés astéroïdes, sont des objets de quelques dizaines de kilomètres de diamètre en orbite autour du Soleil, principalement entre Mars et Jupiter.

En tant que scientifique, mes domaines de prédilection pour cette mission demeurent le système solaire

d'un côté, et les tests de la théorie de la relativité générale appliqués aux étoiles et aux quasars, de l'autre. Les quasars sont les objets astronomiques les plus éloignés que nous connaissons et qui se présentent comme des points lumineux, bien qu'il s'agisse de galaxies dont le centre est occupé par un trou noir. Ces trous noirs sont la source de processus très énergétiques à l'origine du rayonnement des quasars. Ces derniers n'ont aucun mouvement dans l'espace et constituent donc des repères fondamentaux pour le repérage céleste.



François Mignard

CdS : Quels sont les méthodes et traitements statistiques que vous utilisez en astronomie et que vous enseignez ?

FM : Je suis aujourd'hui chercheur à plein temps, sans charge permanente d'enseignement. Mais jusqu'à ces toutes dernières années, j'ai enseigné régulièrement en troisième cycle universitaire l'astronomie fondamentale,

la théorie des probabilités et la statistique, ainsi que l'analyse numérique.

Les outils statistiques que j'utilise dans mon travail sont nombreux.

Il y a tout d'abord ceux de statistique descriptive : il s'agit par exemple de définir les caractéristiques principales de grands échantillons (pouvant aller jusqu'à un milliard d'objets astronomiques, avec une vingtaine de variables décrivant ces objets). On se sert pour cela des indicateurs classiques de tendance centrale (moyenne, médiane, centrage robuste) et de dispersion, ainsi que des méthodes de corrélation dans les distributions spatiales. On cherche aussi les relations liant des paramètres de position et de vitesse des objets astronomiques à des quantités physiques relatives à ces mêmes objets, comme la couleur des étoiles ou leur composition chimique. Ces analyses se généralisent à plus de deux paramètres et, dans ce cas, les outils d'analyse en composantes principales peuvent être utilisés.

J'utilise beaucoup les méthodes inverses : l'outil statistique de base est ici la théorie de l'estimation ponctuelle, utilisée pour retrouver des paramètres inconnus à partir de mesures indirectes et un modèle mathématique liant ces paramètres aux quantités mesurées. La formulation mathématique est en général assez complexe, et l'inversion combine les méthodes statistiques et l'analyse numérique des systèmes d'équations, linéaires ou non.

Par exemple, dans le cadre du projet Gaia, nous effectuons une inversion de ce type, faisant intervenir 10 milliards de mesures affectées d'une

erreur plus ou moins aléatoire, pour retrouver environ 500 millions de données inconnues relatives aux étoiles, telles que leurs position, distance et vitesse. L'estimateur est construit à partir d'une minimisation en norme quadratique du vecteur des résidus (c'est à dire les différences, pour chaque mesure, entre la valeur observée et la valeur prédite par le modèle), pondérés par les poids statistiques de chaque observation. La théorie statistique sous-jacente (méthode du maximum de vraisemblance) permet également d'obtenir une estimation de la précision du résultat de type fourchette d'erreur ou intervalle de confiance.

J'effectue par ailleurs des tests statistiques. Ils servent par exemple à déterminer l'appartenance d'objets astronomiques à des classes d'objets (étoiles ou quasars, astéroïdes ou satellites, étoiles chaudes ou froides, géantes ou naines, étoile à éclat variable ou constant...), ou encore à valider des résultats de traitements par comparaison à des distributions statistiques d'erreurs. Ces tests peuvent être paramétriques ou non, selon les objectifs.

Je réalise enfin des simulations. Les astronomes en sont de gros utilisateurs pour mettre au point des modèles ou préparer des expériences. Ces simulations passent par la génération de distributions aléatoires particulières (uniformes, normales, log-normales, puissances, poissonniennes), nécessaires pour imiter la nature. La distribution de Poisson, par exemple, est à la base de nombreuses modélisations du comportement aléatoire de l'arrivée des photons sur les détecteurs de lumière, que l'on peut assimiler à un processus de file d'attente sans mémoire.

Ces simulations sont indispensables pour produire à l'avance des signaux semblables à ceux qui sont attendus lors du fonctionnement d'un instrument, afin de mettre au point les méthodes mathématiques et statistiques d'analyse des données. Mais c'est également au travers de simulations que l'on peut évaluer les performances de la chaîne constituée d'un

système de mesure (téléscope pour collecter la lumière, caméra ou détecteur numérique pour l'enregistrer) et de tous les traitements qui permettent de remonter aux paramètres astronomiques ou physiques des objets observés.

L'utilisation de la statistique mathématique, comme l'inégalité de Cramer-Rao, fournit des bornes sur les meilleures performances que l'on pourra obtenir pour les estimateurs convergents et sans biais. C'est un outil fondamental dans les analyses de conception d'un instrument au sol ou dans l'espace, qui est donc largement utilisé par les astronomes. Cette inégalité quantifie en pratique l'information disponible (information de Fisher) dans un ensemble fini de mesures affectées d'une erreur statistique.

CdS : Au moyen de quelles méthodes, de quels recoupements, s'assure-t-on qu'un objet astronomique que l'on vient de repérer est réellement nouveau, ou bien qu'il correspond en réalité à un objet que l'on avait déjà recensé ? Et dans le second cas, comment fait-on pour prouver qu'il s'agit d'un seul et même objet ?

FM : C'est une question très importante, qui se pose en astronomie lorsqu'un même objet est observé par différents instruments et dans des longueurs d'onde différentes : l'observation peut en effet se faire dans le visible, l'infra-rouge, le domaine millimétrique ou radio, les rayons X ou gammas. L'unique méthode employée est celle de l'identification de l'objet par sa position, ce qui suppose que tous ces instruments soient capables de fournir des coordonnées assez précises. Mais les aspects statistiques sont très secondaires dans cette activité : cela relève plus de la théorie de la décision, plusieurs choix étant possibles.

Il faut également s'entendre sur ce que l'on appelle un objet astronomique. Une galaxie, qui est un objet dans un catalogue de galaxies, comprend des centaines de millions d'étoiles, qui sont elles-mêmes des objets astrono-

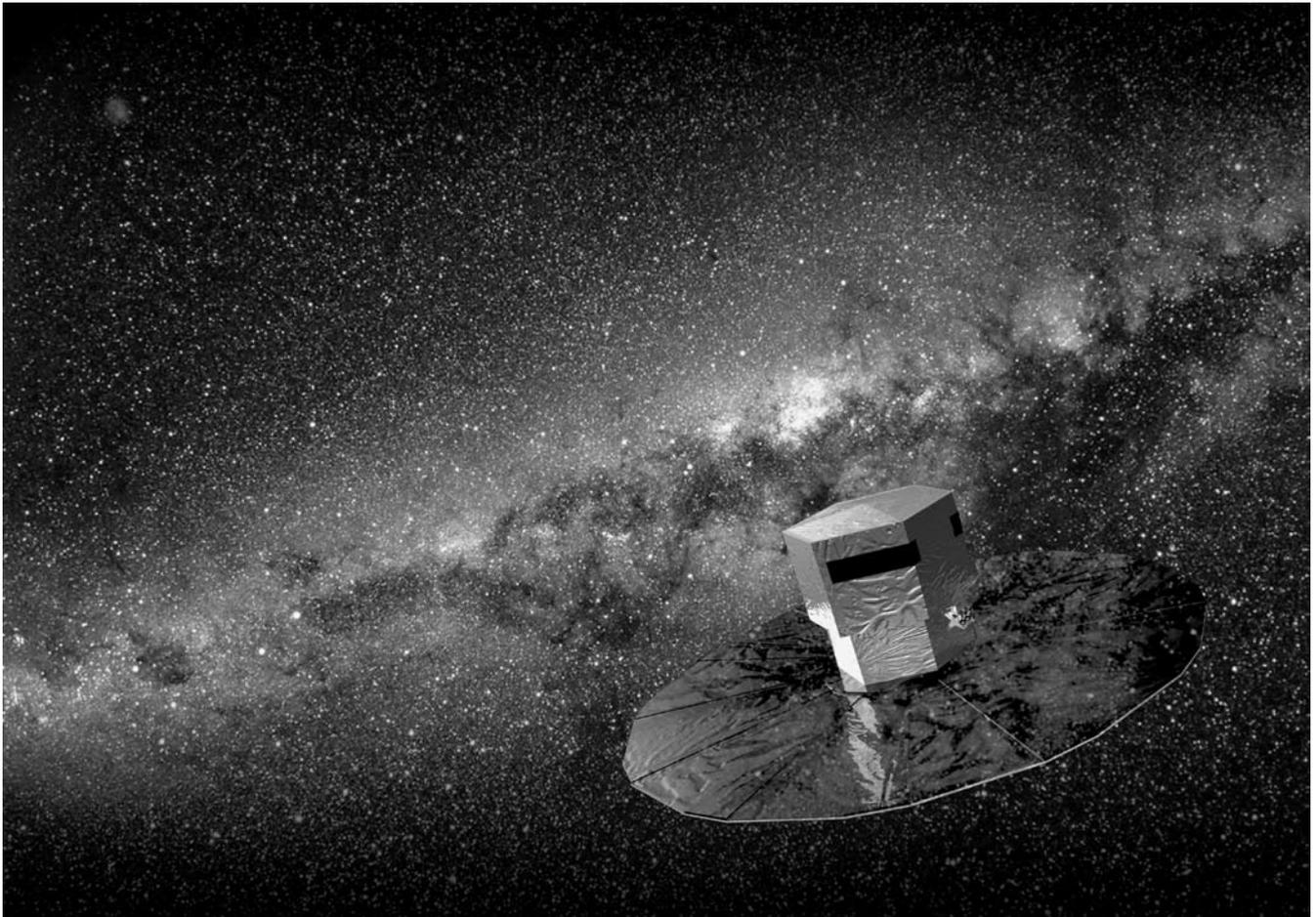
miques dans un catalogue d'étoiles. Lors de leur naissance, les étoiles peuvent se former par petit groupe (souvent par deux ou trois) et rester liées au cours de leur évolution par la force d'attraction gravitationnelle. Elle apparaissent alors dans le ciel comme des sources lumineuses très proches l'une de l'autre et constituent un système multiple. Ce système est en lui-même un objet astronomique, différent de chacune de ses composantes.

Bref, une entrée d'une base de données astronomiques peut être un système en lui-même ou ses composantes individuelles. D'ailleurs, tous les instruments ne permettent pas forcément de voir les composantes au sein d'un système et, dans ce cas, c'est tout naturellement le système qui est identifié (sans même savoir qu'il peut être réduit à des composantes plus simple). Avec un autre instrument, plus gros ou plus spécialisé, on sera en mesure de comprendre qu'il s'agit d'un système complexe ou même d'en détecter les composantes, qu'il faudra alors nommer et intégrer dans la base de données. On dit dans ce cas que le système est résolu par l'instrument. Ainsi, selon l'instrument utilisé, on aura accès au système ou à l'une de ses composantes, et les problèmes d'identification des objets en question ne seront pas résolus de la même façon. Plus que de la statistique, il s'agit là de structure et de description de données, autrement dit de nomenclatures¹.

CdS : Dans les années à venir, les progrès qu'il est possible de faire dans votre domaine de recherche seront-ils, selon vous, plutôt du côté de la statistique ou de l'astronomie ?

FM : D'un point de vue historique, les astronomes ont eu une place particulière dans la naissance de la statistique, tout simplement parce qu'ils étaient les seuls à effectuer des mesures quantitatives répétables qu'il fallait combiner. Les premières

1. Voir à ce sujet l'interview de Françoise Genova dans ce même numéro [NDLR].



Source : ESA

Montage photo du Satellite Gaia sur un fond d'étoiles et de nuages de gaz de la Voie Lactée

analyses de mortalité aux XVII^e et XVIII^e siècles sont le fait d'astronomes (Halley, les frères Huygens, Deparcieux, Wargentin), probablement parce qu'ils avaient l'habitude de manipuler des données numériques abondantes.

Quetelet, le fondateur des statistiques sociales, était initialement un astronome, fondateur de l'Observatoire de Bruxelles. Enfin, avec les géodésiens, les astronomes-mathématiciens du XIX^e siècle, ont donné naissance à la théorie des erreurs et à la méthode des moindres carrés, ce qui a abouti à la formalisation de la théorie de l'estimation. Ensuite, la statistique est devenue une discipline autonome, avec le rôle majeur des bureaux nationaux de statistiques des

populations pour les aspects descriptifs. On voit ainsi le lien étroit qui a uni l'astronomie et la statistique.

Aujourd'hui, le mouvement s'est inversé et les astronomes vont puiser dans la statistique théorique des outils fondamentaux.

Dans ma pratique, mélangeant collecte des observations, traitement du signal et analyse des données, les limites sont fixées en premier lieu par la performance des instruments d'observation et en second lieu par la puissance de calcul, bien avant les avancées en statistique. Les outils mathématiques et statistiques permettent de tirer le meilleur parti des données, mais n'apparaissent pas comme des facteurs limitants des

progrès attendus dans les décennies à venir. Aller dans l'espace, observer dans toutes les longueurs d'onde, traquer les rayons cosmiques ou les ondes gravitationnelles, construire au sol des grands télescopes optiques ou des radio télescopes, tout cela sera plus déterminant que les avancées en statistique théorique. C'est un peu moins vrai des méthodes du traitement du signal et de l'image, pour lesquelles l'exploitation des caractéristiques statistiques des bruits (puissance spectrale) et l'échantillonnage sont déterminants pour extraire l'information utile. Dans ce cas, les progrès théoriques sur ces questions devraient bénéficier aux astronomes, y compris pour reprendre l'analyse de données existantes. ■