

PRÉSENTATION DU NUMÉRO

Cinq ans déjà de publications pour le Courrier des statistiques nouvelle mouture, plus de 70 articles, et toujours une exigence de pédagogie, la volonté de s'ouvrir à de nouveaux sujets, de nouveaux services... avec désormais des numéros à deux chiffres, et une maquette modernisée.

Ce numéro N10 commence par une problématique jamais abordée dans le nouveau Courrier (mais à plusieurs reprises dans l'ancien), celle de la diffusion des statistiques. Il s'agit plus précisément de la visualisation des données, ou *datavisualisation* (*dataviz* pour les intimes). L'article de Christine Lagarenne, Frédéric Minodier et Odile Samson est riche et multiplie les exemples, y compris anciens (l'extraordinaire carte de Charles Joseph Minard sur la campagne napoléonienne de Russie) pour nous faire toucher du doigt toutes les facettes de cette activité. Cependant, il ne faudrait pas hâtivement la résumer à la réalisation de « jolis » graphiques. À la frontière entre diffusion et communication, la *dataviz* cherche à simplifier les messages pour une compréhension au premier coup d'œil mais doit aussi donner envie aux lecteurs de lire. Pour cela, elle fait appel à différentes techniques : infographie, défilement narratif, datavisualisation interactive... L'article soulève également des questions opérationnelles, techniques et d'organisation, sur un sujet désormais incontournable pour les statisticiennes et statisticiens.

Une préoccupation permanente du Courrier est de donner la parole à tous les services statistiques ministériels, afin de donner à voir la manière dont se décline la statistique dans différents secteurs d'activité. Jusqu'à présent, le Courrier a proposé des articles provenant des SSM de la justice (N1), de l'intérieur (N2, voire N7), de la santé (N4), du SDES pour la partie logement (N4), de l'éducation (N5, N6, N7), de l'agriculture (N7), et des collectivités locales (N8). Dans le présent numéro, deux nouveaux domaines sont abordés, la défense et le sport.

Dans l'article sur les statistiques de la défense, Pierre Greffet aborde les spécificités du domaine, non soumis à des règlements européens, ne se fondant pas facilement dans les nomenclatures générales (NAF), et entretenant des liens réguliers et très formalisés avec le monde de la recherche. Il met aussi en évidence un paradoxe : de telles données sont souvent très confidentielles, avec une sensibilité qui dépasse le secret statistique... et dans le même temps il existe une volonté d'ouverture, notamment aux chercheurs. On peut résoudre cette contradiction apparente en soulignant que si certains sujets peuvent être sensibles (économie de défense), d'autres ne le sont pas (fréquentation des lieux de mémoire) ; ainsi, un projet pour organiser une ouverture très maîtrisée est en cours de réflexion avec le principe de *data room*.

Avec les statistiques sur le sport, on retrouve un sujet non encadré par une réglementation internationale... et peu encadré de façon générale. Augustin Vicard présente les sources disponibles dans ce domaine et met en évidence les limites des données administratives, les enjeux de caractérisation de la notion de pratique sportive ainsi que les difficultés

liées à la multiplicité des sports, certains plus rares. Enquêter sur le sujet requiert donc des choix forts, par exemple en se concentrant sur la pratique sportive régulière. Au-delà du processus de production, la question même des enjeux de ces statistiques se pose : envisage-t-on le sport comme fait social ? Comme activité physique, avec en particulier une approche sanitaire, et des politiques publiques qui vont avec ? L'auteur aborde enfin la question des données issues d'applications connectées, de capteurs, intéressantes mais qui ne permettent pas un suivi harmonisé de la pratique.

Il s'ensuit une série de deux articles à la structure très proche, portant cette fois sur les domaines de l'éducation et de la santé, mais qui ne sont pas stricto sensu des articles de SSM : en effet, ils ne portent pas sur les statistiques mais sur les répertoires. De ce point de vue, ils font écho au dossier sur les répertoires du numéro N8, qui contenait des articles sur le RNIPP, le SNGI, SIRUS, et la Base permanente des équipements. Dans le présent numéro, deux répertoires d'établissements sont décrits : le répertoire FINESS des établissements sanitaires et sociaux et le répertoire Ramsese des établissements du système éducatif. On ne saurait trop recommander aux lectrices et lecteurs de lire les deux articles « en miroir » pour en percevoir les points communs et les différences.

FINESS est un répertoire connu dont la création date de plus de 40 ans (1979) et qui joue un rôle fondamental dans la régulation, l'évaluation, le pilotage, le financement et l'identification des structures qui le constituent. Il se caractérise par un cadre formel très exigeant, tout enregistrement de données requérant l'existence d'actes juridiques ou administratifs. Chaque établissement possède un numéro d'identification. FINESS joue un rôle majeur dans l'écosystème des systèmes d'information de santé. Articulé avec deux autres référentiels du domaine, le RPPS et le ROR¹, il est aussi apparié avec Sirene. Il fait l'objet de nombreuses utilisations par les administrations centrales, le grand public et les établissements eux-mêmes. Son rôle central, sa diffusion large, induisent de fortes exigences de qualité. Pour pallier certaines limites, il fait l'objet d'une refonte pilotée par l'Agence du numérique en santé.

Créé en 1977, à la même époque que FINESS, le répertoire académique et ministériel sur les établissements du système éducatif Ramsese s'appuie sur une démarche très différente, profitant de l'organisation territoriale du système éducatif. Ainsi, la gestion déconcentrée du répertoire est-elle assurée par les services statistiques académiques (SSA), le travail des gestionnaires au niveau local étant déterminant pour garantir la qualité des données du répertoire. Ses usages sont très variés et permettent de répondre à des besoins statistiques, de pilotage, de gestion ou d'interopérabilité, chaque établissement disposant d'un numéro d'identification unique. Ramsese joue un rôle central dans le système éducatif et veille à la cohérence des données des structures publiques avec Sirene. Sa diffusion fondée sur des API facilite le partage des données dans le cadre des projets d'urbanisation des applicatifs. La visibilité de Ramsese s'étend, avec la mise à disposition en *Open data* d'une partie de ses données.

¹ Répertoire partagé des professionnels de santé et répertoire opérationnel des ressources

Ce numéro s'achève avec un thème *a priori* plus classique dans l'univers statistique, celui des sondages. Enfin, classique... jusqu'à un certain point, car Pascal Ardilly y aborde un sujet peu évoqué en statistique publique : les sondages empiriques. N'étant pas fondés sur une sélection aléatoire *a priori* d'un échantillon, ils constituent, en revanche, la norme dans les instituts de sondage. De façon pédagogique, l'auteur décline les différences entre sondages aléatoire et empirique, avec force schémas. Les sondages empiriques, efficaces pour maîtriser les coûts, le sont moins pour maîtriser les erreurs, avec en particulier un problème spécifique de biais. Mais ils possèdent eux aussi une véritable justification théorique. Pour finir, à travers deux exemples, l'article éclaire d'un jour nouveau l'utilisation du big data, en nous délivrant un message : attention, la quantité n'est pas gage de qualité...

Pascal Rivière
Directeur de la collection, Insee