

Images satellites, SPOT en ciel pour la statistique publique ?

À l'Insee, la gestion de l'information géographique a longtemps été étroitement associée à l'organisation du recensement de la population et à la diffusion de ses résultats. La demande croissante d'une information statistique finement localisée pour éclairer le débat public a conduit à décloisonner et étendre cette gestion à l'ensemble du processus de production, d'analyse et de diffusion statistique. En amont, il s'agit de puiser dans un gisement croissant d'information géographique afin de localiser, au moment opportun du processus de production, les unités d'intérêt de la statistique publique (établissements, logements, individus...). En aval, il s'agit, d'une part, de mobiliser ce positionnement pour enrichir les analyses avec des méthodes statistiques adaptées et, d'autre part, de doter les outils de diffusion de manière plus systématique d'une dimension cartographique, tout en veillant scrupuleusement au respect du secret statistique.

Une contextualisation accrue et une analyse plus fine des phénomènes économiques et sociaux sont par ailleurs rendues possibles par la mobilisation d'un nouveau type de données pour les instituts de statistique : les données satellites. Ces données, qui font l'objet de cette lettre, sont désormais plus facilement accessibles. Elles posent cependant des défis particuliers de traitement et d'analyse, d'architecture informatique, de collaboration interne ou externe avec des acteurs nouveaux.

Nul doute que l'Insee, qui a inscrit une gestion efficiente du fait géographique au titre de ses ambitions 2025 et qui vient de lancer une réflexion sur la place des études en économie géographique pour mieux connaître et comprendre les disparités territoriales, saura mobiliser son expérience, son goût pour l'innovation, son insertion dans un réseau international dynamique et ses compétences académiques pour répondre au défi spatial.

Vincent Loonis, chef de la Division méthodes et référentiels géographiques (DMCS)

Vers une statistique officielle vue du ciel ?

Dans le foisonnement actuel des données ouvertes, les informations géospatiales, en particulier issues de l'imagerie satellite, occupent une place particulière du fait de leur richesse, leur potentiel pour divers domaines, leur complexité et la part de rêve qu'elles nous rapportent de l'espace. Information accessible immédiatement a priori sur toute la surface du globe, sans coût de réponse supplémentaire ni besoin de recourir à l'infrastructure statistique en place dans le pays, leur potentiel pour la statistique publique a été repéré depuis quelques années par la Commission Statistique des Nations-Unies. Plus précisément son groupe de travail mondial sur les mégadonnées (Global Working Group on Big Data) créé en 2014 et qui oriente sa stratégie sur les données massives a constitué une task force pour recenser les cas d'usages, initiatives publiques et privées s'appuyant sur l'imagerie satellitaire et les données géospatiales et conduire quelques projets pilotes en la matière. Le bilan de ces travaux, une revue de la méthodologie adaptée et des bonnes pratiques concernant l'usage de ces données ont donné lieu à un guide à l'attention des Instituts Nationaux de Statistiques qui envisageraient d'utiliser ce type de données pour la statistique officielle. Ce *handbook* "Earth Observations for Official Statistics Satellite Imagery and Geospatial Data Task Team report" est disponible en ligne depuis fin 2017 sur la page de la task force (<https://unstats.un.org/bigdata/taskteams/satellite/>).

Parmi les projets pilote, on compte en particulier la couverture et l'utilisation du sol (la classification de récoltes) pour le suivi et les statistiques agricoles (Australie, Colombie, Canada), un projet avec Google pour l'évaluation des stocks de pétrole. Ces projets sont hébergés sur la UN Global platform for Official Statistics qui offre une infrastructure et des outils adaptés à l'accès et au traitement de ces données, encourage l'ouverture des codes et le partage d'expérience en fournissant des outils collaboratifs et une forte incitation à la reproductibilité des analyses par divers instituts statistiques. Elle fournit aussi des services de formation et d'accompagnement.

Eurostat s'intéresse également à cette source notamment via le groupe de travail (WPH) sur l'observation de la terre ("[Earth Observation](#)") de l'ESSnet Big Data 2, qui travaillera entre 2018

et 2020 sur ces questions. Ce groupe de travail réunit huit pays partenaires (Pologne, Belgique, Allemagne, France, Italie, Finlande, Pays-Bas, Portugal) dont plusieurs acteurs pour la France (Insee, SDES, SSP). Ce groupe de travail vise à explorer le potentiel des données satellitaires et de la photographie aérienne, au travers de cas d'usages concrets, et à élaborer un cadre méthodologique de référence pour traiter ces données. Les cas d'usages couvrent les statistiques agricoles (classification/monitoring des cultures, de la végétation hors saison), les zones de bâti et les objectifs de développement durable reliés (étalement urbain/espaces publics ouverts en ville), la couverture du sol, et l'enrichissement de la directive Inspire (infrastructure for spatial information in Europe).

Des données de plus en plus accessibles

Des initiatives nationales (Equipex GEOSUD, Pôle THEIA) mais également des programmes européens comme le programme Copernicus ont pour but de rendre les données satellitaires plus facilement accessibles aux collectivités locales, à la communauté scientifique, aux entreprises ou au grand public.

Citons trois exemples de données satellitaires sur la France accessibles pour la statistique publique :

- LandSat est le premier programme spatial d'observation de la Terre destiné à des fins civiles. Il est développé par l'agence spatiale américaine, la NASA. Huit satellites Landsat ont été lancés entre 1972 et 2013. Les images, distribuées gratuitement et sans licence d'utilisation par leur propriétaire ont donné lieu à de nombreuses applications dans les domaines du climat, de l'agriculture, de la géologie, de la sylviculture, etc. Une application est décrite plus en détails au paragraphe suivant.
- Les images sur France entière des satellites SPOT (les dernières versions disponibles sont SPOT 6/7) via l'Equipex Geosud : l'accès est gratuit pour la recherche et les acteurs publics. Le projet compte de multiples partenaires : l'IGN, l'Irstea, etc.
- Initialement connu sous le nom de GMES (Global Monitoring for Environment and Security), le programme européen dénommé Copernicus depuis juillet 2013 a pour objectif la surveillance de l'environnement et la sécurité.

Les données images satellites produites au sein du programme Copernicus sont mises à disposition avec un accès total et gratuit (les images Sentinel-1 et 2).

Néanmoins, l'accès aux images satellites à très haute résolution (plus fine qu'un mètre), qui sont produites par des entreprises privées, européennes ou américaines principalement, reste payant.

Une liste des principales données satellitaires est disponible dans le *handbook* des Nations Unies (2017) précédemment cité.

Les expériences au sein du Service Statistique Public

Le Service Statistique Public accumule aussi ses expériences d'usages des images satellites. Ces initiatives ont alimenté en partie une [session](#) aux Journées de méthodologie statistique en juin dernier.

Ainsi, une [contribution](#) a présenté comment les données satellitaires (données LandSat) permettent de mesurer des îlots de chaleur urbains. En effet, les ondes infrarouges captées par le satellite offrent la possibilité de reconstruire la température au-dessus de l'atmosphère, laquelle évolue comme la température au sol. L'étude a ainsi mis en évidence sur 10 métropoles françaises un écart de 2°C en moyenne entre les centres-villes et leur périphérie. Cet effort d'appropriation de la source satellitaire à l'Insee a soulevé les difficultés à la mobiliser, les données sont en effet particulièrement volumineuses et les limites liées à la couverture nuageuse peuvent s'avérer importantes.

Le ministère de l'agriculture conduit annuellement l'enquête Teruti-Lucas sur la nature et l'usage du sol. Des points sont tirés aléatoirement sur le territoire et certains visités par des enquêteurs. Afin d'optimiser les déplacements d'enquêteurs, il serait particulièrement intéressant d'estimer la probabilité de changement dans l'usage du sol au niveau d'un point. C'est pourquoi le SSM agriculture (SSP) a noué une collaboration avec le Centre d'études spatiales de la biosphère (CESBIO), expert dans le traitement des données satellitaires avec lequel il produit la couche [OSO](#) d'occupation des sols. Cette collaboration entre spécialistes de la télédétection et le SSP s'inscrit notamment dans le groupe de travail WPH "Earth observation" de l'ESSnet Big Data 2.

L'Insee participe également à ce groupe de travail avec pour objectif de mesurer l'accès aux espaces ouverts dans les villes, en vue d'implémenter l'indicateur 11.7.1 du cadre international d'indicateurs associés aux Objectifs de développement durable. Les images satellites permettent de construire des indicateurs comparables internationalement puisque reposant sur une même source de données avec une même méthodologie.

De la télédétection aux données cartographiques

Pour autant, les images satellites ne sont pas les seules sources d'information géospatiale mobilisables pour la statistique publique.

S'y ajoutent les informations cartographiques. L'Institut national de l'information géographique et forestière (IGN), avec lequel l'Insee collabore depuis de longues années, est l'opérateur de l'État en matière de production et de diffusion de l'information géographique de référence en France. Son champ de compétence couvre aussi bien la géographie physique (altimétrie, fleuves...) qu'humaine (découpages administratifs, bâtiments, routes...). L'IGN entreprend avec la géoplateforme, prévue à l'horizon 2020, une transformation de son offre de diffusion et plus largement de son positionnement institutionnel. Cette plateforme répond à une triple vocation : coordonner et fédérer les organismes qui produisent des données géographiques,

garantir l'accessibilité aux données géographiques souveraines et enfin permettre à l'IGN de certifier la production des données cartographiques.

Depuis le début du 21^e siècle, de nouveaux acteurs privés, mais aussi provenant de la sphère de l'open source proposent des services de cartographie en ligne. Par exemple, Google Maps et Bing Maps sont respectivement les services de Google et de Microsoft. Ils permettent entre autres choses de visualiser des cartes, de suivre un itinéraire et d'obtenir des informations sur les lieux accessibles aux publics. Bien que non payant pour un usage domestique, ces services ne sont pour autant pas gratuits. Généralement, leur tarif est fonction du nombre de consultations.

A contrario, OpenStreetMap (OSM), soutenu par l'organisation éponyme à but non lucratif enregistrée au Royaume-Uni, a pour ambition de créer une base de données géographiques mondiales libre. Les données cartographiques sont élaborées par une communauté de cartographes bénévoles et sont libres d'utilisation. Pour toutes publications à partir de ces données, la source « OpenStreetMap » doit néanmoins être citée. Plusieurs sites, tels que geofabrik, permettent de télécharger gratuitement ces données. Outre ces utilisations domestiques et commerciales, OSM est exploité par le monde de la recherche pour produire des analyses géographiques (variabilité des formes urbaines, organisation des espaces ruraux...).

Enfin, l'État français (via Etalab, service du Premier ministre en charge de l'ouverture des données publiques et du gouvernement ouvert, au sein de la Direction interministérielle du numérique et du système d'information et de communication de l'État) et les acteurs para-publics mettent à disposition progressivement auprès du grand public des données cartographiques provenant de leurs administrations. À titre d'exemple, Etalab diffuse les données cadastrales de la DGFIP.

Actualités / Brèves

Save the date ! Le prochain séminaire Big data se tiendra **mardi 14 mai 2019**, de 10h00 à 12h00 en Salle Closon Malinvaud. Il présentera les résultats de l'expérimentation « identification employeur » réalisée à la suite du hackathon « Les Champs de Sirene » qui a eu lieu en janvier 2018. Paola Tubaro du CNRS interviendra également pour porter un regard réflexif sur les transformations du travail liées à « l'intelligence artificielle ».

Le document de travail « Les méthodes de décomposition appliquées à l'analyse des inégalités » de Béatrice Boutchenik, Sophie Maillard et Élise Coudin est [sorti](#) !

L'Insee a accueilli la Task Force Big data d'Eurostat ces 4 et 5 avril derniers. Et pour en savoir plus sur l'ESSnet Big Data, [cliquez ici](#).

Consulter la page « [Parlons Innovation statistique](#) » pour en savoir plus sur la conférence NTTs de mars dernier ainsi que sur la visite de labs innovants à l'université de Berkeley, Californie.

Cinq ateliers autour du webscraping ont été organisés par le SSP Lab et la DIIT, plus d'information en rejoignant le réseau Yammer sur le [groupe dédié](#).

Présentation du potentiel des données de téléphonie mobile à la prochaine commission Territoires du Cnis « [Mesurer l'accès aux équipements et services](#) » le mercredi 15 mai 2019 à 14h30 au Ministère de l'Économie et des Finances.

Ont participé à ce numéro Élise Coudin, Vincent Loonis, Mathilde Poulhes, Benjamin Sakarovitch et François Sémécurbe.

Cette lettre est une occasion d'informer largement et d'échanger. N'hésitez pas à nous [transmettre](#) vos réactions et suggestions d'articles à mathilde.poulhes@insee.fr et benjamin.sakarovitch@insee.fr
Les archives de cette lettre sont disponibles sur : <https://www.agora.insee.fr/cms/sites/dmcsi/home/ssp-lab/BigData.html>
Demande d'inscription individuelle à la lettre : dg75-l101@insee.fr