Les principaux investissements méthodologiques moyen terme du service statistique public

JMS 31 mars 2022

S. Lagarde

Insee

Mesurer pour comprendre

Beaucoup de résonance avec le cadre européen 1/2

- **Data revolution and trusted smart statistics** : axe 7 du MAP (multi annual action plan) du programme statistique européen 2021-2027

  - Mobiliser de nouvelles sources de données (plateformes, web, téléphonie mobile, données satellites, compteurs intelligents, logiciels de gestion mobilisés par des professionnels, transactions bancaires, données de caisses…)
  - Explorer de nouvelles méthodes de traitement (NLP, ML et IA) pour : exploiter ces sources nouvelles, répondre à des problèmes plus traditionnels (codification automatique, traitement de la non réponse, redressements...)
  - Intégrer de nouveaux modes de recueil de l'information dans la collecte des enquêtes ménages (application smartphone pour la gestion de carnets, traces GPS, fichiers sons...)
  - Progresser en matière de gestion de la confidentialité et du traitement des séries temporelles (Centre d'excellence européen)

Beaucoup de résonance avec le cadre européen 2/2

- Développer des exploitations rapides ou avancées (timeliness) : transversal à l'ensemble des axes du MAP
  - Nowcasting : expériences pendant la crise sanitaire en mobilisant de nouvelles données imparfaites mais disponibles rapidement (pour le suivi conjoncturel)
  - Méthodes de microsimulation (taux de pauvreté)
  - Exploitations avancées de sources exhaustives (DSN…)
- Développer les combinaisons de sources (données administratives, enquêtes, autres sources), les techniques d'appariements (axe 3 du MAP), l'intégration de la géographie et de la statistique (axe 5 du MAP)
- Nécessite un cadre méthodologique permettant de rendre compte de la qualité des sources mobilisées mais aussi des résultats produits (biais, variance)

Des collaborations précieuses au sein du SSE, avec l'Unece

- Du fait de ces fortes résonances entre préoccupations nationales et européennes/internationales :
  - Poursuivre les travaux méthodologiques communs (TF, ESSnets), projets HLG-MOS (modernisation of official statistics de l'Unece par exemple sur ML)
  - Agir sur les freins à ces innovations (accès aux données privées, partenariats/conventions avec des propriétaires de données privées multinationaux, directive e-privacy…)
  - Investir sur des outils communs de webscraping, des plateformes collaboratives de traitements innovants de données, des applications permettant le recueil de carnets dans les enquêtes (emploi du temps, budget des familles)

Des spécificités françaises au sein du cadre européen

- Des domaines sur lesquels investir davantage ou progresser :
  - Confidentialité : traitement encore artisanal ; explorer de nouvelles méthodes (méthodes perturbatives, differential privacy...)
  - Données administratives et appariements : domaine nouveau pour la méthodologie transversale, jusqu'à présent au niveau de chaque direction métier/SSM. S'appuyer sur les travaux de la Vision 2020 d'Eurostat, profiter du lancement du programme Résil (répertoire statistique d'individus et de logements), du CSNS (code statistique non signifiant)

Le multimode : un enjeu national...

- Poursuivre l'appui méthodologique au développement du multimode dans les enquêtes ménages

  - Agrégation des données d'enquêtes multimode : principes opérationnels, protocoles permettant d'identifier de potentiels effets de mode (biais de sélection et erreur de mesure modale)
  - Conception de e-questionnaire. Quels tests (objectifs et mise en oeuvre) pour la conception d'enquêtes multimodes ? Apport des tests cognitifs dans le cadre d'enquêtes internet
  - Quel impact sur l'échantillonnage des enquêtes ménages : quid de l' échantillon maître ?

…mais aussi pour l'ensemble du Système statistique européen (SSE)

- La crise sanitaire a forcé tous les pays à expérimenter des enquêtes ménages en multimode
  - Position paper (fin 2021) d'un GT de 6 pays européens sous la houlette de l'Insee (P. Sillard), mixant experts des enquêtes ménages et méthodologues
  - Recommandations pour poursuivre ces travaux méthodologiques au niveau européen et aboutir à des préconisations opérationnelles de mise en œuvre du multimode généralisé (suites de MIMOD)
  - Un axe de la stratégie innovation 2022-2027 du Système statistique européen ?

Des forces sur lesquelles s'appuyer au niveau national, des partenariats à développer

- Des atouts :
  - Formation initiale de nos cadres (ENSAE, ENSAI…)
  - Des unités SSP Lab et DIIT chargées d'animer/appuyer l'innovation en data science au sein du SSP
  - Pratique d'études, d'analyse conjoncturelle et de travaux exploratoires et pas seulement de production statistique

- Des axes de développement
  - Liens avec la recherche : chaires (nowcasting, finance digitale,...), accueil expérimental de doctorants, participation à des projets ANR
  - Partenariats avec les propriétaires de données administratives et privées

De l'expérimentation à la mise en production : un enjeu pour le SSP et pour le SSE

- Beaucoup d'expérimentations mobilisant de nouvelles sources/nouvelles méthodes de datasciences au sein du SSP et du SSE. Peu de passages en production encore (à l'exception notable des données de caisses)
- Méthodes de ML pour la codification automatique : du potentiel à transformer pour l'intégration en production de la codification de la PCS 2020 dans le recensement
- Comment embarquer les directions métiers et l'informatique suffisamment tôt dans ces expérimentations pour favoriser et accélérer le passage en production, sans lancer un projet traditionnel ?
- Un cas d'usage intéressant au niveau européen : le prochain changement de NAF (NACE au niveau européen). Développer des échantillons pré-entrainés au niveau européen pour la NACE ?

En guise de conclusion

- Un intérêt fort à échanger/travailler ensemble au sein du Système statistique européen sur les innovations méthodologiques moyen terme
- Une mobilisation commune des méthodologues, experts métiers, informaticiens indispensable, au niveau national comme européen pour réussir le passage en production
- Des partenariats à nouer, notamment avec le monde académique et les propriétaires de nouvelles sources de données, au niveau national comme européen
- Quels nouveaux vecteurs européens pour y répondre ? ESSnets et grants aux formats/contours nouveaux ?
- Des JMS à venir aux contours élargis à certaines innovations informatiques ?

Merci de votre attention !