

Autorité de la statistique publique

Séance du 23 mars 2016

Remarque

Les comptes rendus détaillés des débats de l'Autorité de la statistique publique ne sont pas publics.

Le texte ci-dessous expose toutefois les sujets abordés et les décisions ou avis qui en sont éventuellement issus.

**La séance est ouverte à 14h30
sous la présidence de M. Dominique Bureau**

Le Président de l'ASP accueille deux nouveaux membres du collègue. Patricia Blancard est désignée par le président du Cese en remplacement de Philippe Le Clézio. Elle est secrétaire générale adjointe de la Cfdt Cadre, membre du Cese et elle participe aux travaux du Cnis.

Eric Dubois remplace François Ecalte, Il est conseiller membre à la première chambre de la Cour des comptes et a occupé le poste de directeur des études et synthèses économiques à l'Insee.

Anne-Marie Brocas a été nommée par la chef de l'Igas en remplacement de Stéphane Paul. Elle a rejoint l'Igas , après avoir occupé différents postes de responsabilité dans les domaines social et de la santé.

I) Points divers

1) Nomination du vice-président de l'ASP

Le Président de l'ASP propose que Denis Badré soit nommé vice-président de l'ASP.

Les membres de l'ASP nomment Denis Badré vice-président de l'ASP.

2) Suivi des « petits SSM »

Le Président de l'ASP rappelle qu'il avait été jugé nécessaire de revoir les « petits » SSM Défense et Culture dans le cadre des clauses de revoyure. Les deux services ont proposé des projets de services qui sont actuellement instruits par la Direction de la méthodologie et de la coordination statistique (DMCSI) de l'Insee.

Les deux SSM seront de nouveau auditionnés par l'ASP dès lors que les projets de service auront été approuvés par la DMCSI.

Par ailleurs, les deux SSM de la DGFIP et de la Justice vont publier pour la première fois en 2016

un calendrier prévisionnel de leurs publications.

Enfin, au terme de la revue par les pairs, la recommandation de respecter des règles d'embargo plus strictes a pour date limite le deuxième semestre de 2017. Les SSM se mettent actuellement aux normes pour anticiper cette date.

3) Approbation du rapport annuel 2015 de l'ASP

Un projet de rapport a été transmis aux membres du collège.

Suite aux remarques des membres, une annexe listant toutes les recommandations de la revue par les pairs a été ajoutée et il sera noté dans l'avant-propos que les enjeux liés au numérique nourriront le programme de travail de l'Autorité en 2016.

Un rapport critique sur la statistique anglaise- rapport BEAN- a été publié récemment, qui affirme que la statistique anglaise est de bonne qualité, mais qu'elle rend très mal compte de l'évolution de la structure de la production et des activités de service dans un monde où le numérique joue un rôle croissant.

Les enjeux liés au numérique étant importants, il a été décidé d'en faire un axe stratégique du programme de travail de l'ASP.

4) Suivi par l'ASP de la nomination de la nouvelle directrice de la DARES

La revue par les pairs avait relevé que les personnes nommées chefs de SSM devraient l'être uniquement pour leurs compétences et leur indépendance, alors que les directeurs de l'administration centrale étaient nommés de manière discrétionnaire et révocable. Le remplacement à la tête de la Dares a précipité la nécessité de voir comment l'ASP traite de ces situations.

En l'espèce, le ministère du Travail avait sollicité le directeur général de l'INSEE pour qu'il lui propose des personnes susceptibles de diriger la DARES. L'ASP a suivi cette nomination, dans le cadre réglementaire existant. Le retour d'expérience de ce cas conduit à pointer les éléments suivants : l'ASP doit fondamentalement vérifier la conformité du profil de la personne proposée avec les exigences requises, sans élargir à d'autres candidatures éventuelles, son rôle n'étant pas d'opportunité. Le moment de son intervention le plus propice se situe avant le choix du Ministre et la présentation au Président de la République. Il serait donc utile de confier plus spécifiquement à l'ASP la mission de donner un avis au ministre qui émettra la proposition.

Le président de l'ASP suggère que les membres lui donnent mandat pour qu'il en parle avec le SGG.

En tout état de cause, il propose d'adapter en conséquence le règlement intérieur de l'ASP. En effet, si le collège a de fait pour mission de s'assurer que les nominations sont conformes au règlement de la statistique, il est nécessaire de modifier le règlement intérieur pour permettre au Président d'informer les membres dans des délais appropriés. Cette démarche permettrait de vérifier l'existence d'un consensus au sein du collège avant de donner un avis au Ministre.

II) Suivi des recommandations de l'ASP relatives à la labellisation des statistiques mensuelles des demandeurs d'emploi (DEFM)

Le président de l'ASP rappelle le contexte de la labellisation des DEFM. Le 26 mars 2014, l'Autorité a émis un avis favorable à la labellisation des statistiques mensuelles de demandeurs d'emploi inscrits à Pôle Emploi, publiées conjointement par la Dares et Pôle Emploi.

La mise en place de cette labellisation s'est appuyée sur des travaux des trois inspections générales IGF, IGAS et IG de l'Insee. La labellisation a été accordée pour une période de cinq ans, avec une clause de revoyure après deux ans. La labellisation a été assortie de six recommandations :

- la publication de séries avec recul ;
- la publication au cas par cas de séries rétropolées tenant compte des incidents et des changements de procédure ;
- la nécessité d'ajouter à la publication mensuelle des DEFM un commentaire privilégiant la tendance des derniers mois ;
- la nécessité d'attirer l'attention des lecteurs sur la faible signification de la variation d'un mois sur l'autre en dessous d'un certain seuil ;
- la publication régulière des séries d'indicateurs visant à évaluer la variabilité statistique des DEFM en stocks et en flux ;
- la conduite d'une analyse préventive des risques des processus de production et de publication et la prévision d'une procédure de gestion de crise.

En raison de l'importance prise dans le débat public par les DEFM, Pôle Emploi et la Dares ont souhaité associer le Cnis. Un groupe de consultation du Cnis a été mis en place pour étudier les modalités pratiques de mise en œuvre des recommandations de l'ASP. Au-delà des évolutions demandées par l'ASP, Pôle Emploi et la Dares ont souhaité mettre à profit la réflexion ouverte par les recommandations de l'ASP pour faire évoluer le format de la publication mensuelle et en améliorer la lisibilité. Tous ces travaux ont été présentés au public en conférence de presse en janvier 2016 lorsque le format de la publication mensuelle a été modifié.

- 1) **Audition de Monsieur Stéphane DUCATEZ, directeur des statistiques de Pôle Emploi, accompagné de Monsieur Cyril NOUVEAU, son adjoint**
- 2) **Audition de Madame Selma MAHFOUZ, Directrice de l'animation, de la recherche, des études et des statistiques (DARES), accompagnée de Monsieur Pierre BISCOURP, sous-directeur Emploi et marché du travail**

Les deux auditions se déroulent conjointement.

Le directeur des statistiques de Pôle Emploi, Stéphane DUCATEZ débute avec le rappel des recommandations 1,2 et 6, et la réponse qui leur a été apportée.

Première recommandation : « Publier sur une base annuelle des séries révisées en droits constatés (séries "avec recul") issues du fichier historique statistique des demandeurs d'emploi »

La STMT est une statistique mensuelle sur le marché de travail produite rapidement, sans recul, sur ce qui s'est produit au cours du mois précédent. Dans ces statistiques administratives, lorsque les données sont publiées, certaines informations ne sont pas nécessairement attestées. Il peut donc exister un écart entre la statistique publiée et la statistique obtenue après recul et enregistrement par Pôle Emploi des pièces administratives. L'indicateur le plus pertinent retenu consiste à publier les données avec six mois de recul, ce qui semble être le meilleur compromis entre la complétude des données et la date de disponibilité de l'information.

Le fichier historique statistique des parcours des demandeurs d'emploi compile toutes les informations mensuelles et permet de reconstituer les parcours des demandeurs d'emploi sur une période de dix ans.

S'agissant de la mise en œuvre de la recommandation, une note méthodologique a été publiée sur les sites internet de la DARES et de Pôle Emploi. Elle explique les deux concepts et présente la série statistique avec un recul de six mois, ainsi que les écarts entre la série avec six mois de recul et la série publiée dans la statistique. Cette série sera mise à jour chaque année en septembre.

Deuxième recommandation : « Publier, au cas par cas, les séries rétropolées tenant compte des incidents et des changements de procédure »

Un document recensant l'ensemble des incidents depuis 2011 sera enrichi au fil du temps avec les incidents et changements de procédures qui auront été relevés, pour la bonne interprétation des statistiques des demandeurs d'emploi. Une estimation de l'impact et des séries contrefactuelles sont ajoutés lorsque cela est possible. Les principaux incidents relatés sont l'incident sur les relances d'août 2013 (« bug SFR »), la réforme des règles de gestion des radiations administratives de janvier 2013, les particularités de la période d'actualisation (mai 2015) et les opérations de fiabilisation de la liste des demandeurs d'emploi, lorsque Pôle Emploi a pu disposer de données plus complètes et rapides sur les demandeurs d'emploi en formation ou en contrat aidé. Toute personne consultant la publication est informée de l'existence de ce document et peut y accéder rapidement en ligne.

Sixième recommandation : Conduire une analyse préventive des risques des processus de production et de publication et prévoir une procédure de gestion de crise

Le document rédigé conjointement par la DARES et Pôle Emploi permet d'enrichir la gestion de l'élaboration de cette statistique par la DARES et Pôle Emploi. Il permet également de formaliser la procédure d'échanges entre Pôle Emploi et la DARES en cas d'événement inhabituel.

Le document présente une cartographie des risques susceptibles d'avoir des conséquences en matière de retard des publications, soit par manque d'informations, soit en cas d'événement qui aurait un impact sur un indicateur et conduirait à une mauvaise interprétation de la situation ou de l'évolution mensuelle publiée, ou tout autre incident qui conduirait à publier des données erronées ou à ne pas publier.

Des solutions sont mises en place pour prévenir les risques de production et d'autres sont envisagées. Elles sont regroupées en cinq catégories :

- la remontée des données opérationnelles ;
- les risques relatifs à la constitution ou à la transmission de fichiers ;
- les risques humains dans les traitements ;
- les risques statistiques dans l'exploitation des fichiers (indicateurs erronés, évolution atypique) ;
- la protection et la confidentialité des données pour les protéger d'un risque informatique, et la réaction à observer en cas de fuite ou de rupture d'embargo.

Le processus consiste à mettre en place des instances de suivi chez Pôle Emploi et entre Pôle Emploi et la DARES, puis à créer une cellule de crise comprenant la direction de Pôle Emploi et de la DARES pour échanger régulièrement pendant la production et intervenir à J-1 en cas d'événement grave conduisant à annuler la publication ou à effectuer une publication partielle.

La directrice de l'animation, de la recherche, des études et des statistiques (DARES), Madame Selma MAHFOUZ, poursuit avec la présentation des réponses aux recommandations 3, 4 et 5 et de la refonte des publications.

Troisième recommandation : « Ajouter à la publication mensuelle des DEFM un commentaire privilégiant la tendance des derniers mois »

La directrice de l'animation, de la recherche, des études et des statistiques (DARES), Madame Selma MAHFOUZ, indique que la démarche suivie pour répondre à cette recommandation a consisté à calculer un indicateur statistique de tendance, désormais mentionné dans le commentaire des évolutions. Les critères pour choisir cet indicateur sont les suivants :

- l'indicateur doit pouvoir être calculé chaque mois ;
- il doit être relativement facile à expliquer et interpréter ;
- il doit permettre un certain lissage, une orientation persistante ;
- il doit être capable de détecter rapidement les retournements.

Un équilibre reste donc à trouver entre la volonté de lisser et de capter les retournements. Des travaux ont été menés pour juger des deux derniers critères, en comparant la tendance sous-jacente des séries, calculées à posteriori par moyennes mobiles (procédure dite « X12 »), avec des calculs d'indicateurs de tendance pouvant être calculés en temps réel. Pour les séries d'effectifs, la variation sur trois mois a été retenue, et, pour les séries de flux, une moyenne trimestrielle glissante. Dans la nouvelle publication, cet indicateur de tendance complète l'information mensuelle pour les séries d'effectifs et se substitue à l'information mensuelle pour les séries de flux.

Quatrième et cinquième recommandations : « Attirer l'attention des lecteurs sur la faible signification de la variation d'un mois sur l'autre en dessous d'un certain seuil » et « Publier régulièrement des séries d'indicateurs visant à évaluer la variabilité statistique des DEFM en stocks et en flux »

Les données administratives présentent une certaine volatilité mensuelle en raison des chocs sur le marché du travail, de la variabilité due aux procédures de gestion et d'imprécisions dans l'estimation de la saisonnalité. Cette volatilité perturbe l'interprétation conjoncturelle, ce qui conduit à privilégier les analyses en tendance. Des seuils ont été déterminés, au-delà desquels la variation des effectifs de demandeurs d'emploi est de même signe que la tendance sous-jacente. Au-delà de ces seuils, sur la base des évolutions passées, on estime qu'il existe 95 % de chances que la variation de la tendance soit de même sens que la variation observée dans les données. Ces seuils seront ré-estimés une fois par an, lors de la publication de la DEFM de janvier. Ils permettent d'évaluer la volatilité de la série et indiquent ainsi dans quelle mesure les variations observées renseignent sur la tendance sous-jacente.

Les seuils sont communiqués dans la documentation. Par exemple, pour les demandeurs d'emploi en catégorie A, la variation sur un mois n'est significative que si elle est supérieure à 27 000, alors que le seuil de significativité est de 35 000 pour les variations sur trois mois.

L'information sur la tendance à trois mois est plus souvent informative de façon fiable que celle sur la tendance à un mois. Dans de nombreux cas, l'évolution sur un mois n'est pas significative, contrairement à celle sur trois mois.¹

La refonte de la publication

La refonte de la publication tient compte au mieux des recommandations de l'ASP. Un groupe du CNIS présidé par Jacques Freyssinet a été créé pour discuter des modalités d'application des recommandations de l'ASP et, plus largement, de la rénovation du format de la publication. Un nouveau format a été instauré en janvier 2016.

La publication a été recentrée sur les principaux indicateurs. Les données sur le RSA et les offres d'emploi sont désormais diffusées sur d'autres supports. Les publications des régions ont également été allégées. La publication a été rendue plus lisible. Les anciens avertissements ont été allégés et des graphiques ont été introduits, avec un historique plus long. Des aides à l'interprétation ont également été ajoutées : des indicateurs de tendance et des compléments sont intégrés dans la publication ou ses annexes, ainsi qu'un renvoi vers des publications ou données complémentaires.

¹ Pour plus de détails voir les actes du séminaire de méthodologie statistique du 3 novembre 2015 : <http://www.insee.fr/fr/insee-statistique-publique/default.asp?page=connaître/colloques/sms/sms.htm>

Après discussion, l'Autorité délibère de la manière suivante :

1 - Décision relative à l'adaptation du règlement intérieur, notamment pour permettre à l'Autorité de jouer pleinement son rôle concernant l'indicateur 1-8 du CBP.

Au titre de sa mission (art. I.1. du décret 2009-250), l'ASP se doit de vérifier la mise en œuvre du Code de bonnes pratiques de la statistique européenne (CBP), donc, en particulier, que les nominations au sein du SSP mettent au premier plan les critères professionnels (indicateur 1-8 du CBP) ; et que les conditions d'affectation éventuelle de l'un de ses responsables à un autre poste ne compromettent pas l'indépendance professionnelle. À cet égard, la situation la plus sensible concerne la nomination des responsables ayant le statut de directeur d'administration centrale.

Cette situation étant susceptible d'intervenir à tout moment, il convient de consolider les conditions dans lesquelles l'ASP en traite, notamment pour permettre : que l'ASP puisse exercer son rôle dans les délais contraints qui sont associés aux procédures de nomination ; que la nature de l'avis de l'ASP, les conditions dans lesquelles celui-ci est élaboré et rendu public, soient transparentes.

À cette fin, le règlement intérieur de l'Autorité doit expliciter la procédure dans ce cas spécifique. La proposition correspondante sera examinée à sa prochaine séance, en juin 2016. À cette occasion, l'adéquation du règlement intérieur et, notamment, l'opportunité de pouvoir recourir plus largement aux consultations électroniques de ses membres, seront systématiquement passés en revue.

2 - Proposition de délibéré pour la séance de l'ASP du 23 mars 2016 : statistiques mensuelles des demandeurs d'emploi inscrits à Pôle emploi.

La labellisation des statistiques mensuelles de demandeurs d'emploi inscrits à Pôle emploi (DEFM) a été accordée par l'Autorité le 26 mars 2014, pour une durée de cinq ans, associée à des recommandations dont la mise en œuvre devait être examinée à l'horizon de deux ans. Suite aux auditions de M. Stéphane Ducatez, directeur des statistiques de Pôle Emploi et de Mme Selma Mahfouz, directrice de la Dares, réalisées à ce titre, il apparaît que la Dares et Pôle Emploi ont pris en compte l'ensemble des recommandations de l'Autorité qui étaient associées à cette labellisation. Chacune des recommandations a fait l'objet de travaux méthodologiques approfondis, de nature à améliorer la qualité de cette statistique. La publication des DEFM a par ailleurs été refondue, en intégrant les conclusions du groupe de consultation réuni dans le cadre du Cnis.

L'Autorité salue la qualité des travaux réalisés par Pôle Emploi et la Dares. Ainsi, des séries « avec recul » seront publiées annuellement, et la documentation mise à la disposition du public sur les méthodologies utilisées, sur la significativité des résultats, sur l'impact des incidents et changements de procédure, apparaît conforme au Code de bonnes pratiques de la statistique européenne, notamment son principe 6 en matière d'impartialité et d'objectivité.

Compte tenu de l'importance des statistiques en ce domaine, l'Autorité recommande à Pôle Emploi et à la Dares de poursuivre les travaux engagés relatifs à :

- la présentation des résultats. Il convient de privilégier les évolutions en tendance plutôt qu'au mois le mois, trop volatiles. Ceci est documenté, y compris pour les « stocks », de manière très précise dans la documentation portée à la connaissance de l'Autorité. Dès lors, ceci ne devrait plus relever seulement d'un « avertissement », mais être intégré dans la conception de la présentation de la publication elle-même,

- l'analyse des écarts entre l'évolution des données d'enquête de l'Insee (« chômage au sens du BIT issu de l'enquête Emploi ») et des données d'origine administrative. Pour que la diversité des sources ne soit pas un facteur d'incertitude pour le public, il importe donc de développer tous les travaux de nature à permettre la compréhension de leurs écarts, et au public de les mobiliser de manière complémentaire.

- la gouvernance, notamment, en termes de partage des rôles et des responsabilités, de procédure de gestion de crise, de traçabilité des évolutions des règles de gestion ou processus de production, de démarche-qualité dans la collecte, d'anticipation des impacts des changements de procédure.

Compte tenu de l'acuité des enjeux associés à la statistique des DEFM dans le débat public, et à la poursuite des progrès dans les directions indiquées ci-dessus, l'Autorité propose en 2017 une revoyure des deux organismes. La question du champ de la labellisation (séries nationales, séries régionales etc...), ainsi que l'examen des possibilités de compléter l'information statistique sur le devenir des chômeurs seront aussi examinés à cette occasion.

III) **Présentation de la loi numérique/accès aux données par Madame Françoise DUPONT, co-rapporteur du groupe sur l'utilisation des données privées présidé par Monsieur Michel BON, Madame Pauline GIVORD, cheffe de la division des méthodes appliquées de l'économétrie et de l'évaluation à l'INSEE, Monsieur Michel ISNARD, chef de l'unité des affaires juridiques et contentieuses à l'Insee**

a) Utilisations du big data pour la statistique publique

La cheffe de la division des méthodes appliquées de l'économétrie et de l'évaluation à l'INSEE, **Pauline GIVORD** indique qu'elle assure depuis deux ans un rôle d'animation sur les nouvelles expérimentations relatives à l'utilisation du *big data* dans la statistique publique.

Ces expérimentations s'inscrivent dans le cadre d'un intérêt croissant pour l'utilisation de données privées à des fins statistiques. Ces données (données issues d'internet, données de prix par exemple) sont caractérisées par de très grands volumes. Elles peuvent mélanger différents formats et elles comportent des flux importants de données, ce qui représente des difficultés pour leur traitement.

Les projets les plus aboutis en matière de statistique sont les projets de données de caisse (facturation de la grande distribution qui peut être utilisée pour constituer les indices de prix) et les projets relatifs aux données de téléphonie mobile.

De son point de vue, la statistique publique pourrait être confrontée à la concurrence de sociétés privées relativement à la production d'indicateurs statistiques (par exemple le MIT avec le *Billion prices project*, ADP, ou Orange, qui développe des indicateurs de fréquentation de sites à partir de ses données de téléphonie mobile). Dans leur communication, ces sociétés se présentent comme des concurrents sérieux capables de produire des indicateurs avec des délais de publications restreints.

L'accès à des données privées permet de donner plus rapidement des informations, ce qui est un enjeu fort pour la statistique publique. Grâce à l'importance des volumes de données, il est possible d'obtenir une granularité plus fine. En outre, ces données sont plus objectives sur certaines grandeurs et pourraient permettre de réduire la charge liée au questionnement. Par exemple, une enquête sur le temps de transport des usagers est coûteuse, alors que la téléphonie mobile permet d'obtenir des données précises sur les déplacements des individus.

Ces nouvelles données permettent aussi de compléter le dispositif des indicateurs de développement durable et de constituer des indicateurs relatifs à l'économie numérique, aujourd'hui inexistantes.

L'accès aux données, qui sont en général privées, représente cependant une difficulté. Un cadre juridique doit être défini, car certaines données peuvent être très personnelles, et il est impératif de garantir la confidentialité des données utilisées, dans la mesure où il existe un risque important de ré-identification.

Des questions se posent également sur la qualité, la représentativité et la pérennité de ces données et des indicateurs statistiques qui en découlent, sachant que le mode de production et le contenu de données issues par exemple d'opérateurs privés ne sont pas toujours connus avec précision pour les utilisateurs extérieurs.

S'agissant des ressources internes à l'INSEE, des compétences spécifiques sont nécessaires pour utiliser les outils permettant de traiter ces données. Or, cette expertise n'existe pas de façon standard dans les services de la statistique.

L'Insee a commencé à investir dans le domaine du *big data*. Des groupes de travail ont été mis en place au niveau européen. S'agissant du système statistique européen, le mémorandum de Scheveningen a permis de définir une stratégie pour l'utilisation du *big data* dans la statistique publique permettant une mutualisation entre les différents INS traitant le sujet et identifier les sources intéressantes. Un ESSNET a été créé pour conduire des projets pilotes sur l'utilisation de données privées pour la statistique publique. Ces projets concernent notamment les sites d'offres

d'emploi, les données de téléphonie mobile et les compteurs intelligents.

En termes de coopération internationale, un projet UNECE portait sur la création d'une Sandbox (plateforme informatique possédant les logiciels spécifiques au *big data*), offrant une infrastructure pour permettre des premières expérimentations.

A l'Insee, le projet le plus abouti est le projet « données de caisse » lancé en 2015, avec un objectif d'intégration dans la production à compter de 2019.

Depuis deux ans, un statisticien (qui sera appuyé à partir de septembre 2016 d'un second statisticien) se consacre à la veille active et à la recherche et développement sur ces projets, en participant en particulier aux groupes de travail européens et à des projets de recherche. Les expérimentations ont pour objectif de définir les données qui sont importantes et les niveaux d'agrégation de ces données, ainsi que d'acquérir une expertise sur les méthodologies nécessaires. Une maille plus agrégée permettrait de gérer plus facilement les questions de confidentialité.

Il est également possible d'utiliser les requêtes internet pour améliorer les indicateurs pour la prévision économique (*nowcasting*).

b) Précisions sur les projets d'utilisation des données privées

Le corapporteur du groupe sur l'utilisation des données privées présidé par Monsieur Michel Bon, **Françoise DUPONT**, rappelle, au sujet du groupe présidé par Michel Bon, que les instituts cherchent à améliorer la production et à réduire les coûts. Un parallèle peut être établi avec l'utilisation des fichiers administratifs. Il s'agit en effet de trouver les modalités permettant de pérenniser les nouvelles sources de données et de comprendre leur potentiel.

Tous les pays européens sont dans cette démarche. Les données de caisse sont étudiées depuis les années 2000, alors que les données de la téléphonie mobiles ne sont exploitées que depuis 2013, tout comme les nouveaux compteurs électriques. L'expérimentation relative à toutes ces sources exige du temps pour comprendre ce qu'elles peuvent apporter.

Les données de la téléphonie mobile sont des données dont il est délicat de comprendre la nature de la matière première et les paramètres jouant sur sa pérennité. Dix ans de stabilité technique seraient nécessaires pour prendre une décision d'investissement.

La matière première est souvent très précise, localisée dans le temps et l'espace. Il est plus délicat de trouver de bonnes modalités de procédures, tout en respectant la vie privée. L'expérimentation permet de cerner des aspects de méthode (agrégation, redressement des personnes surreprésentées) pour extraire des statistiques extrapolant correctement sur la France entière.

Pour les données de caisse, la France a progressé en analysant le contenu de ces données et en vérifiant qu'elles apportent un supplément de précision. Il est important de montrer aux fournisseurs la matière première à récupérer et les précautions à prendre pour protéger les données et le secret des affaires. La confiance s'installe en effet progressivement, et une expérimentation est nécessaire avant d'envisager des développements plus larges. En revanche, le coût de ces données est bas.

Même si les professionnels centralisent ces données pour leur bénéfice, un tiers de confiance effectue les analyses, qu'il leur restitue ensuite. Pour les données de téléphonie mobile, les opérateurs ont travaillé en premier lieu avec les professionnels du tourisme et de la route. Convaincre prendra cependant du temps, car les retours sur investissement ne sont pas encore réalisés. L'expérimentation précède nécessairement la production.

Seule l'Estonie a commencé à rechercher et exploiter les données de téléphonie mobile, après un long cheminement. Tous les pays s'intéressent à ces données, qui présentent un intérêt pour les secteurs du tourisme et des transports, ainsi que pour les équipements locaux. Toutefois, de nombreux acteurs essaient d'obtenir des données pour expérimenter, ce qui rend la démarche difficile. Elle est plus facile dans les pays qui n'ont pas encore installé un marché privé.

La concertation s'est déroulée en 2015, avec deux réunions sur les données de caisse, dont la

première a fait apparaître un consensus sur l'intérêt général, avec la nécessité de le matérialiser par un cadrage juridique. Quelques enseignes n'ont pas souhaité faire partie de la concertation.

Le caractère général de la loi numérique a suscité des interrogations. Une rencontre a été organisée pour expliquer que les bonnes pratiques permettant de rassurer seront décrites dans des textes ultérieurs. La loi donne la possibilité de faire une enquête, sachant que chaque sujet fera l'objet d'une analyse de faisabilité et d'une présentation à la CNIL.

Lors de réunions sur la téléphonie, un projet de calcul de la population sur un territoire pour calibrer les équipements d'urgence a été présenté.

Les cartes bancaires sont utilisées par la Banque de France dans le cadre d'un observatoire sur la fraude. Ce projet consiste à établir un partenariat avec la Banque de France pour étudier la consommation.

c) Projet de loi pour une république numérique

Le chef de l'unité des affaires juridiques et contentieuses à l'INSEE, **Michel ISNARD**, rappelle que le projet de loi pour une république numérique est le premier à avoir été soumis à une concertation publique par internet. Il a été présenté le 9 décembre 2015 à l'Assemblée nationale. Le vote a eu lieu le 26 janvier 2016, et une seule opposition a été exprimée. Les discussions étaient constructives et le projet semble engendrer peu de clivage. Le projet a été transmis au Sénat. La séance de la commission des lois aura lieu le 5 avril, suivie de séances du 25 au 27 avril, et le vote du Sénat aura lieu le 3 mai. Quelques amendements sont déjà apparus.

Ce projet de loi concerne tous les sujets de la république numérique.

L'article 1 mentionne la gratuité des échanges entre les administrations de l'État et les établissements publics administratifs de l'État. La CADA devient compétente dans le cadre des relations entre les administrations. L'article 7bis stipule qu'aucune redevance ne sera demandée pour la réutilisation des données produites par le service de statistique publique (SSP). En revanche, d'autres SSM ont des données qui nécessiteront la mise en place d'une redevance. La gratuité ne commencera que le 1^{er} janvier 2017. Une compensation budgétaire de 10 millions d'euros est annoncée pour SIRENE, et probablement pour l'ensemble des données du SSP.

Les données de SIRENE sont revendues à des rediffuseurs et des licences d'usage final sont attribuées. Le gouvernement souhaite ouvrir les données à d'autres entreprises dans l'attente de retrouver des bénéfices en matière d'impôt et de valeur ajoutée. Les assurances de compensation budgétaire sont donc réelles.

L'article 18 est relatif au Nir haché. Des traitements statistiques ou de recherche reposent sur des appariements de fichiers, qui sont effectués en utilisant le numéro de Sécurité sociale et le Nir. Or, utiliser le Nir exige un décret en conseil d'État, ce que les chercheurs publics ont des difficultés à obtenir. Le Nir haché permet de résoudre ce problème : l'ensemble du SSP pourra simplement déclarer le traitement à la CNIL et les chercheurs demanderont l'autorisation à la CNIL pour les traitements de recherche, sans passer par le Conseil d'État. La CNIL a donné son accord. Une seule clé de hachage existera pour les traitements du SSP. En revanche, il existerait une clé de hachage par projet de recherche. Le décret en Conseil d'État est prévu début 2017, et le Nir en clair et le Nir haché coexisteront pendant une période de transition.

L'article 12 est relatif aux données de caisses : il oblige les entreprises à transmettre de manière électronique les réponses à une enquête. Le ministre chargé de l'Économie prendra sa décision à la suite d'une étude de faisabilité et d'opportunité et d'un avis du CNIS.

Actuellement, les données des enquêtes peuvent être communiquées aux chercheurs, mais la loi interdit au SSP de communiquer des données permettant l'identification des entreprises. Les données transmises aux chercheurs ne pourront donc concerner que des agrégations de produits. En outre, seules les données agrégées pourront être archivées.

Il est possible d'infliger des amendes en cas de refus de réponse à une enquête obligatoire, de

25 000 à 50 000 euros, contre 150 à 2 500 euros actuellement. L'article de loi peut s'appliquer à toutes sortes de données ; les décisions de transmissions apporteront des précisions sur les modalités et les destinataires des transferts.

La séance est levée à 18 heures 30.

Autorité de la statistique publique
Séance du 23 mars 2016

ANNEXE

ETAIENT PRESENTS

Monsieur Dominique BUREAU
Président

Monsieur Denis BADRE
Ancien Sénateur

Monsieur Abdeldjellil BOUZIDI
Economiste, Université Sorbonne Nouvelle Paris 3

Monsieur Bruno DURIEUX
Ancien ministre
Président du comité national des conseillers du commerce extérieur de la France

Monsieur Éric DUBOIS
Conseiller Maître à la Cour des comptes

Monsieur Jean GAEREMYNCK
Président du comité du secret statistique
Membre du Conseil d'État

Madame Véronique HESPEL
Inspectrice générale des finances

Madame Anne-Marie BROCAS
Inspectrice générale des affaires sociales

Madame Patricia BLANCARD
Membre du Conseil économique social et environnemental

ASSISTAIENT EGALEMENT À LA SEANCE

Monsieur Philippe CUNEO
Chef de l'Inspection générale de l'Insee

Madame Claudine GASNIER
Rapporteur de l'Autorité de la statistique publique

Madame Béatrice GIMARD
Secrétaire de l'Autorité de la statistique publique