



COURRIER --- DES STATISTIQUES

N3 - Décembre 2019

Rédaction en chef

Odile Rascol

Contribution

Insee : Véronique Alexandre,
Jean-Marc Béguin, Chantal Cases,
Jean-Pierre Cling, Christel Colin,
Franck Cotton, Thomas Dubois
Olivier Haag, Anne Husseini-Skalitz,
Heïdi Koumarios, Marie Leclair,
Hervé Piffeteau,
Éric Sigaud,

CASD : Kamel Gadouche

Directeur de la publication

Jean-Luc Tavernier

Directeur de la collection

Pascal Rivière

Rédaction

Maryse Cadalanu, Pierre Glénat, Odile
Rascol, Pascal Rivière, Nelly Verrière

Composition

Agence **LATITUDE** Nantes

0620/19

02 51 25 06 06

www.agence-latitude.fr

Photo de couverture

Adobe Stock®

Éditeur

Institut national de la statistique
et des études économiques
88, avenue Verdier
92541 MONTROUGE CEDEX

www.insee.fr

© Insee 2019 « Reproduction partielle
autorisée sous réserve de la mention
de la source et de l'auteur ».

Courrier des statistiques N3

SOMMAIRE

Présentation du numéro <i>Odile Rascol</i>	4
Naissance d'une innovation en production statistique <i>Jean-Marc Béguin</i>	6
Pogues, un outil de conception de questionnaires <i>Franck Cotton, Thomas Dubois</i>	17
Eno, un générateur d'instruments de collecte <i>Heidi Koumarios, Éric Sigaud</i>	29
Collecte par internet des enquêtes auprès des entreprises : l'Insee entre dans l'ère Coltrane <i>Olivier Haag, Anne Husseini-Skalitz</i>	45
Utiliser les données de caisses pour le calcul de l'indice des prix à la consommation <i>Marie Leclair</i>	61
Le Centre d'accès sécurisé aux données (CASD), un service pour la <i>data science</i> et la recherche scientifique <i>Kamel Gadouche</i>	76
Un nouveau triptyque juridique pour les statistiques européennes <i>Hervé Piffeteau</i>	93
FRIBS : un nouveau cadre commun pour les statistiques d'entreprises européennes <i>Christel Colin</i>	110
IESS : l'Europe harmonise ses statistiques sociales pour mieux éclairer les politiques <i>Chantal Cases</i>	125
Le système statistique allemand, une évolution facilitée par la réglementation européenne <i>Véronique Alexandre, Jean-Pierre Cling</i>	140

PRÉSENTATION DU NUMÉRO

Ce troisième numéro de la nouvelle collection du Courrier des Statistiques est placé sous le signe de l'innovation : en termes de méthodes, d'outils informatiques, comme sur le plan juridique.

Mais comment naît une innovation ? Jean-Marc Béguin, ancien directeur des Statistiques d'entreprises, retrace le parcours de l'Insee pour concevoir et construire tout un écosystème autour de la plateforme de collecte par internet des enquêtes auprès des entreprises ; il constate que pour rénover significativement, il faut parfois sortir des sentiers battus et suivre des voies inattendues, en profitant des circonstances.

Quelles sont ces nouveautés ? Heïdi Koumarios et Éric Sigaud présentent Eno, outil d'une grande généralité permettant entre autres de générer automatiquement des supports de collecte, selon les différents modes (*web*, papier, face-à-face, etc.) ; le tout à partir du questionnaire formalisé selon des métadonnées respectant des standards internationaux. Écrire ces métadonnées requiert une technicité particulière, qu'il ne faut pas imposer au concepteur d'enquête... d'où l'intérêt d'une application ergonomique permettant au statisticien de les produire aisément : c'est le rôle de Pogues, dont Franck Cotton et Thomas Dubois explicitent la logique et les fonctionnalités.

Enfin, Anne Husseini-Skalitz et Olivier Haag décrivent la plateforme de la collecte par internet et ses services associés : Coltrane. L'ensemble est guidé par un choix structurant : pouvoir activer des métadonnées pour piloter les processus de production statistique (notamment les phases de Conception, de Construction des instruments de collecte). Coltrane, Eno et Pogues ont donc des liens forts et le lecteur constatera ainsi quelques redondances, assumées, entre les différents articles.

Autre innovation, relative à nouveau au processus de collecte : l'élaboration de l'indice des prix à la consommation (IPC) va s'appuyer pour partie, à compter de 2020 sur des « données massives », les données de caisses. Marie Leclair aborde ainsi l'ensemble des possibilités qui s'offrent à l'Insee avec ces nouvelles données, mais également les difficultés soulevées et les réponses apportées, pour l'IPC comme pour d'autres utilisations à venir de cette source d'un nouveau type.

Une production statistique ne présente d'intérêt que si des utilisateurs peuvent s'en emparer : Kamel Gadouche revient sur le dispositif de mise à disposition des données de la statistique publique aux chercheurs et aux *data scientists*. Avec près de dix ans d'expérience, le Centre d'accès sécurisé aux données (CASD) est une structure dont le rayonnement s'est élargi à l'international, et qui innove cette année avec la mise en place d'une certification de recherches fondées sur des données confidentielles.

Enfin pour ce numéro N3, l'Europe est plus que jamais présente, avec une réforme en profondeur du cadre juridique qui sous-tend les activités statistiques. De façon transversale, le socle juridique de la statistique européenne évolue : Hervé Piffeteau nous en décrit la genèse et les perspectives. Mais ce n'est pas tout : deux grands piliers de la statistique européenne, attendus depuis longtemps, se mettent en place. Christel Colin nous explique que le règlement-cadre sur les statistiques d'entreprise, FRIBS, permet d'unifier et de clarifier la mosaïque préexistante de règlements par domaine qui prévalait jusqu'alors. Chantal Cases nous conte une histoire similaire pour les statistiques sociales, avec le nouveau règlement-cadre, IESS. Dans chacun de ces deux papiers on se pose la question


des conséquences concrètes pour les utilisateurs, pour le rapport coût-efficacité, pour l'Insee. Jean-Pierre Cling et Véronique Alexandre concluent ce numéro en présentant le système statistique allemand, dans la lignée du numéro précédent : son organisation est facilitée par les évolutions réglementaires européennes, qui ont introduit une certaine souplesse dans le dispositif législatif allemand.

Odile Rascol
Rédactrice en chef, Insee

NAISSANCE D'UNE INNOVATION EN PRODUCTION STATISTIQUE

Jean-Marc Béguin*

Au départ, l'Insee visait à passer sur internet la collecte de l'enquête sectorielle annuelle (ESA) dans l'espoir d'en améliorer la qualité intrinsèque et le taux de réponse. Les arbitrages de moyens conduisirent à agréger le projet avec deux autres, pour construire une plateforme de collecte commune à toutes les enquêtes auprès des entreprises de la statistique publique. Une demande gouvernementale de simplification des enquêtes obligea l'Insee à précipiter la mise en ligne de l'ESA sur internet. Cette contrainte allait à nouveau modifier le projet et, du fait de la complexité du questionnaire conduire à une première innovation de rupture : le développement d'un générateur de questionnaires et la validation du concept de métadonnées actives. La gouvernance du projet faillit aussi être bouleversée par la difficulté d'adapter la future plateforme de collecte aux enquêtes auprès des ménages. Le premier « hackathon » de l'Insee, à l'initiative de la maîtrise d'œuvre, permit une deuxième innovation de rupture : le développement d'outils mutualisables, permettant de générer automatiquement, dans un standard international et de façon ergonomique, les métadonnées d'enquêtes, fournissant les briques d'un véritable atelier de conception. Voici ainsi un exemple rare et réussi de transformation d'un projet et d'adaptation à son environnement, puis de récupération par l'institution d'innovations qui ont bousculé les procédures classiques de genèse de projets.

 *Initially, INSEE aimed to collect the annual sectoral survey (ESA) on the Internet expecting improving its intrinsic quality and response rate. The decision on resources allocation led to merge the project with two others, and to build a common data collection platform for all businesses surveys of official statistics. A government request to simplify surveys forced INSEE to speed up the publication of ESA on the Internet. Both this requirement and the questionnaire complexity modified again the project and led to a first breakthrough innovation: the development of a questionnaire generator and the validation of the concept of active metadata. The project's governance was also affected by the difficulty to choose a web collection platform adapted to household surveys. INSEE's first «hackathon», at the initiative of the IT project management, resulted in a second breakthrough innovation: the development of common tools to automatically generate survey metadata according to an international standard and in an ergonomic manner. It provided building blocks for a real design workshop. This is a rare and successful example of how a project can be transformed and adapted to its environment. It also shows how an institution can reuse innovations even if they do not adopt the usual project management approach.*

* Ancien directeur des Statistiques d'entreprises à l'Insee de 2003 à 2013.

Dans une organisation humaine d'une certaine taille (disons plusieurs milliers de personnes), l'introduction d'une innovation résulte pratiquement toujours d'un processus complexe et souvent imprévisible. Il y a d'abord la formalisation du contenu potentiel d'une innovation à partir d'un besoin ressenti par des utilisateurs (par exemple *via* un service marketing chargé de « révéler » les besoins des clients), ou à partir d'idées nouvelles émises par des agents particulièrement créatifs. Il s'ensuit logiquement un processus de gestion des différentes idées ou projets pour les trier, quantifier les moyens dont ils bénéficieront et les ordonnancer dans le temps. Ce processus de gestion de l'innovation est étroitement lié au processus d'arbitrage des moyens, ce qui

« Transformer les chocs tant internes qu'externes en opportunités : profiter de ces chocs plutôt que les subir. »

n'est pas sans impact sur la vie des projets, leur déroulement, leurs acteurs. Il arrive cependant qu'on puisse transformer les chocs tant internes qu'externes en opportunités : profiter de ces chocs plutôt que les subir. Ceci peut donner naissance à des innovations importantes et totalement différentes du besoin exprimé au départ. C'est ce qu'illustre l'histoire des projets Coltrane, Eno et Pogues¹.

🎯 LE BESOIN INITIAL : UN PROJET CIRCONSCRIT ET CLASSIQUE —

À l'été 2009, une grande préoccupation de la direction des Statistiques d'entreprises (DSE) de l'Insee était le retard et la moindre qualité des résultats sectoriels fins² d'Ésane, le nouveau système d'information de la statistique structurelle d'entreprises, dont c'était la première année de mise en production.

L'élaboration des résultats à ce niveau fin de nomenclature repose sur la ventilation du chiffre d'affaires des entreprises entre leur activité principale et d'éventuelles multiples activités secondaires. Cette ventilation est la principale information que l'on cherche à collecter dans les deux enquêtes (ESA et EAP³) qui forment l'un des deux piliers⁴ du système Ésane. Pour de nombreuses entreprises, dont les répondants ne sont pas des spécialistes des nomenclatures, l'explicitation des diverses activités entre lesquelles ils doivent ventiler leur chiffre d'affaires se fait sous forme de libellés en français qu'il faut ensuite coder pour pouvoir exploiter l'enquête (« vente de ceci... », « fabrication de cela... », « fourniture de tel service... », etc.). C'est cette activité de codage, dont le volume avait été mal anticipé, qui était en assez grande partie responsable du retard de la production et de la moindre qualité des données d'Ésane.

Bref, à la fin de l'année 2009, l'Insee souhaitait améliorer la qualité de la réponse à cette question, couplé si possible à une amélioration du taux de réponse. Il y avait pour cela

-
1. Les trois articles suivants reviennent en détail sur Pogues (Franck Cotton et Thomas Dubois), Eno (Heidi Koumarios et Éric Sigaud) et Coltrane (Olivier Haag et Anne Hussein-Skalitz).
 2. La Nace, nomenclature d'activité européenne, comprend 615 classes et la nomenclature française (NAF) 732 sous-classes : un secteur fin est constitué de l'ensemble des entreprises dont l'activité principale correspond à une classe ou une sous-classe.
 3. L'EAP (enquête annuelle de production) a la même fonction que l'ESA (enquête sectorielle annuelle), mais elle ne concerne que l'industrie, la ventilation du chiffre d'affaires se basant sur les 3 000 postes de la Prodcom (nomenclature européenne spécifique).
 4. L'autre pilier est l'exploitation des données administratives que constituent les déclarations fiscales des entreprises.

une idée naturelle : *passer du mode de collecte papier, en vigueur à l'époque, à la collecte par internet*. Cette dernière présentait potentiellement de nombreux avantages :

I - elle répondait à une demande de modernisation et de simplification⁵ de la part des entreprises ;

II - elle correspondait aux orientations du programme à moyen terme de l'Insee et à celles de son ministère de rattachement (Économie), dont une structure était chargée, pour le compte de tous les autres ministères, de l'adaptation aux nouvelles technologies ;

III - elle permettait de réduire significativement les frais d'impression, d'envoi, et de saisie, d'autant plus que l'ESA avait un échantillon très important (plus de 100 000 entreprises) et un dispositif d'impression complexe (plus de 200 modèles de questionnaires selon les secteurs !) ;

IV - on imaginait enfin, et peut-être surtout, à la fois améliorer le taux de réponse à l'enquête et à la question de ventilation du chiffre d'affaires en particulier, en la rendant par exemple obligatoire et bloquante, et accroître la qualité de la réponse en proposant des questions annexes, dans le cadre interactif du *web*, lesquelles auraient permis aux entreprises de mieux signaler leurs activités secondaires au regard de leur activité principale. On imaginait

« Un projet innovant, rentable, cohérent avec le programme à moyen terme, présentant de fortes externalités positives et somme toute relativement classique et circonscrit. »

aussi améliorer la qualité générale de l'enquête en introduisant des contrôles interactifs, éventuellement bloquants. Tout cela nécessitait néanmoins des tests de faisabilité et de performance.

C'est ainsi qu'en avril 2010 la DSE effectuait une première

demande de moyens⁶ pour lancer un projet de passage de la collecte de l'ESA sur internet : il s'agissait d'un projet innovant (point IV), rentable (point III), cohérent avec le programme à moyen terme (point II), présentant de fortes externalités positives (point I) et somme toute relativement classique et circonscrit.

● D'AUTRES DEMANDES DE MOYENS CONCOMITANTES

Par ailleurs, en 2008, à la suite d'une décision politique de simplification des structures ministérielles, la production statistique du Sessi⁷ venait d'être rattachée à l'Insee. L'institut avait en particulier récupéré la responsabilité de l'ensemble des enquêtes d'entreprises dites « thématiques⁸ », quel que soit leur secteur d'interrogation. Celles qui étaient gérées auparavant par le Sessi étaient déjà accessibles par internet mais *via* une plateforme que l'Insee souhaitait abandonner pour des raisons de sécurité et de maintenabilité. L'institut disposait bien d'une infrastructure de collecte par internet des enquêtes auprès des entreprises, mise en place depuis 2004 : celle-ci était adaptée à des enquêtes infra-annuelles aux questions répétitives mais pas à des enquêtes thématiques dont les questionnaires pouvaient évoluer à chaque occurrence, comme ceux exigés par les règlements européens.

5. Une étude du Sessi avait montré en effet que, même si les questions posées étaient les mêmes par internet et sur papier, les entreprises voyaient l'interrogation par internet comme une simplification par rapport aux enquêtes papier et les plébiscitaient par conséquent.
6. À l'Insee, la plupart des projets sont développés en interne et il n'y a que très peu de sous-traitance externe. L'attribution de moyens spécifiques dits « d'investissement » est donc une condition souvent nécessaire pour faire aboutir un projet.
7. Le Service des études et des statistiques industrielles (Sessi) est l'ancien service statistique du ministère de l'Industrie. À la fin des années quatre-vingt-dix, ce ministère avait été intégré au ministère des Finances pour former le Minefi.
8. Il s'agit de cinq ou six enquêtes par an, plus ou moins périodiques (certaines tous les ans ou deux ans, certaines tous les cinq ou dix ans, voire totalement nouvelles), dont la plupart répondent à un règlement européen sur un thème donné.

Concomitamment, en avril 2010 donc, avec des objectifs similaires à ceux de l'ESA (simplification, modernisation, qualité, économie), la DSE effectua une seconde demande de moyens informatiques, pour assurer le passage sur internet des enquêtes thématiques. Entre autres raisons, il paraissait souhaitable, vu le sujet, d'être capable de gérer par internet l'enquête sur l'utilisation des technologies de l'information et de la communication par les entreprises (TIC), menée à la demande d'Eurostat tous les ans depuis 2006. L'échéance de disponibilité visée était janvier 2012.

Côté informatique, on s'inquiétait de voir se développer autant de solutions que de besoins ponctuels pour toutes les collectes par internet, alors que les fonctions assurées ou les services rendus étaient manifestement les mêmes, quelle que soit l'enquête envisagée : créer le questionnaire, authentifier les enquêtés, mettre à disposition le questionnaire, contrôler les données, rapatrier les données à l'Insee de façon sécurisée, etc. On mettait donc l'accent sur la nécessaire mutualisation de ces fonctions et demandait aussi des moyens pour développer une infrastructure moderne qui serait commune et partagée entre les diverses maîtrises d'ouvrage d'enquêtes, y compris celles des enquêtes auprès des ménages.

ARBITRAGES, NAISSANCE D'UN PROJET...

Face à la multiplicité des demandes de moyens, inhérente à toute organisation, l'Insee dispose de son propre dispositif de régulation, un comité *ad hoc* (dit comité des investissements) qui analyse toutes les demandes, les classe en fonction d'un certain nombre de critères et les soumet à l'arbitrage du comité de direction. Il proposa assez logiquement de regrouper les trois demandes précédentes qui concernaient chacune la collecte par internet en un seul et même projet, et de lui attribuer deux chefs de projet spécifiques⁹, un avec une compétence informatique et un avec une compétence métier (donc à la DSE). Sur la base de cet arbitrage, un nouveau projet était né, même si les moyens en question furent loin d'être immédiatement disponibles.

À ce stade, le projet devenait évidemment beaucoup plus large que la simple collecte par internet de l'ESA : il s'agissait de construire une plateforme de collecte offrant tous les services mentionnés plus haut, la demande spécifique pour l'ESA devenant ici un ensemble de services particuliers à rendre et donc à intégrer dans le projet. La plateforme devrait ensuite être utilisée aussi bien par les enquêtes auprès des entreprises déjà disponibles sur le *web* que par les enquêtes thématiques auprès des entreprises, voire par les enquêtes auprès des ménages. Le projet impliquait surtout une organisation dans laquelle la DSE, en tant que premier utilisateur de la plateforme, était maître d'ouvrage. Côté enquêtes auprès des ménages en effet, la collecte est le plus souvent administrée en face-à-face *via* un réseau d'enquêteurs. *Via* internet, elle devenait auto-administrée, ce qui posait des problèmes de biais qu'il était important de pouvoir mesurer avant même de changer de mode de collecte. Côté enquêtes auprès des entreprises, les questionnaires étaient déjà auto-administrés, la plupart des questions quantitatives, et la réponse des entreprises ainsi bien moins influencée par le mode de collecte (papier ou électronique).

Les deux chefs de projet baptisèrent le projet Coltrane¹⁰, dès leur arrivée en septembre 2011, et il était déjà sensiblement différent de l'idée initiale.

9. En juillet 2010, seuls deux autres projets sur vingt-six demandes bénéficièrent ainsi de moyens supplémentaires.

10. COLlecte TRANsversale d'Enquêtes renvoie à John Coltrane : l'équipe avait volontairement choisi d'identifier son projet au musicien, considéré comme « l'un des saxophonistes les plus influents et novateurs de l'histoire du jazz ».

1 ... ET DÉJÀ UNE VOLONTÉ DE GÉNÉRALISATION

La situation au lancement du projet était la suivante :

- ① côté maîtrise d'ouvrage, l'expérience des collectes par internet existantes (ex-Sessi, enquêtes infra-annuelles) et l'étude d'une demi-douzaine de réalisations étrangères permettaient de bien cerner l'ensemble des services à offrir à la fois aux entreprises enquêtées et aux gestionnaires d'enquêtes en interne. En juillet 2011, pour plus de généralité encore, une proposition fut faite aux services statistiques ministériels (SSM) impliqués dans des enquêtes auprès des entreprises, lesquels ne collectaient pratiquement pas par internet : mener une étude de faisabilité d'un portail de collecte unique pour les entreprises (les répondants se connectent à un même site), voire partager au sein du Système statistique public la future plateforme Coltrane (ils bénéficient de tous les services offerts par la plateforme) ;
- ① côté informatique, l'idée d'une construction modulaire préfigurant les développements agiles et s'appuyant sur le langage XML était présente dès l'origine. Elle incluait plusieurs chantiers indépendants : authentification des répondants, gestion des contacts, portail des services proposés aux enquêtés, et plateforme de collecte proprement dite.

« L'informatique apportait une idée tout à fait nouvelle : optimiser la mise en ligne d'une nouvelle enquête par le recours à un générateur de questionnaires à partir des métadonnées décrivant l'enquête. »

L'informatique apportait une idée tout à fait nouvelle : optimiser la mise en ligne d'une nouvelle enquête par le recours à un générateur de questionnaires à partir des métadonnées décrivant l'enquête. Cela permettait d'aller encore plus loin dans la généralité, sans se limiter à la collecte par internet : on envisageait une collecte multimodale, fondée sur une description unique de l'enquête, s'appliquant à tous les cas de figure connus (papier, échange de formulaires informatisés, dépôt/retrait de fichiers, etc.), potentiellement adaptable aux enquêtes auprès des ménages. Le générateur de questionnaires à partir des métadonnées devait logiquement être

fourni à Coltrane par le référentiel de métadonnées statistiques de l'Insee RMÉS, sous réserve que ce dernier soit prêt à temps (Bonnans, 2019).

Ainsi, au démarrage effectif du projet Coltrane, les idées de la maîtrise d'ouvrage et des informaticiens conduisaient toutes deux, de façon complémentaire, à un projet bien plus ambitieux et général qu'il ne l'était au départ. C'est alors qu'un événement imprévu vint modifier le cours normal du projet.

2 COMMENT UTILISER EFFICACEMENT UN CHOC EXOGÈNE

En avril 2011, quelques mois avant le démarrage proprement dit du projet, les assises nationales de la simplification firent remonter des entreprises, principalement de la part des PME, une forte demande d'allègement de la charge de réponse des entreprises. Le cabinet du secrétariat d'État qui les avaient organisées, demanda des propositions en ce sens à court terme. Comme la quasi-totalité des enquêtes de la Statistique publique auprès des entreprises répondaient à un règlement européen, et qu'une réduction importante de la charge venait d'avoir lieu avec la mise en place d'Ésane, l'allègement ne pouvait pas provenir du contenu des enquêtes. On proposa une autre forme de simplification : accélérer la mise en ligne sur internet des enquêtes auprès des entreprises ; avec *in fine* deux dates cibles : permettre une réponse dématérialisée à l'ESA dès janvier 2012, et à la totalité des enquêtes pour 2013.

Or les seuls moyens réellement mobilisables à cette fin étaient ceux du projet Coltrane. La seule solution viable était alors de les utiliser pour un développement rapide, satisfaisant la demande politique à court terme, puis de revenir au projet ensuite. Il n'était évidemment plus question de tester comment améliorer la qualité de la réponse à la ventilation du chiffre d'affaires ; on n'en avait plus le temps. Ainsi le projet initial disparaissait-il avant même d'avoir commencé. Il s'avéra rapidement que la date de janvier 2012 était inatteignable (avec plus de 200 déclinaisons sectorielles, le questionnaire de l'ESA était de loin le plus complexe de toute la statistique d'entreprises) mais, après mûre réflexion, l'équipe de projet fit deux propositions originales qui facilitaient bien les choses :

- ❶ intégrer la dématérialisation de l'ESA dans le cadre du projet Coltrane : on ne perdait pas de temps ainsi sur le projet et on pourrait tenir quand même la date de janvier 2013 pour l'ESA suivante ;
- ❷ se servir de la description de l'enquête avec ses métadonnées pour générer les 200 modèles de questionnaires différents ; en d'autres termes « tester » le concept de « métadonnées actives » : l'outil de génération de supports de collecte « Eno » allait naître¹¹.

L'équipe de projet travailla en parallèle à la mise au point de l'expression des besoins¹² de Coltrane (achevée en avril 2012) et à la préparation de la dématérialisation de l'ESA pour janvier 2013. Elle se chargea de la description de l'ESA selon les standards internationaux en matière de métadonnées (DDI¹³). En revanche l'urgence des délais trancha le débat de savoir qui se chargerait du générateur de questionnaire : ce ne pouvait être que le projet Coltrane.

« Il n'est pas certain que l'Insee ait perçu à l'époque la puissance du concept de métadonnées actives et de la filière qu'inaugurerait le générateur de supports de collecte Eno. »

Comme c'est souvent le cas en matière d'innovation, il n'est pas certain que l'Insee ait perçu à l'époque la puissance du concept de métadonnées actives et de la filière qu'inaugurerait le générateur de supports de collecte Eno. Grâce à ces nouveaux outils, fussent-ils intermédiaires, le délai de janvier 2013 fut tenu *in extremis* pour l'ESA. Surtout, le succès technique engrangé¹⁴ apporta

à toute l'équipe de projet la preuve de l'opérationnalité du concept. Il est peu probable que sans cette contrainte exogène sur les délais et sur l'ESA le projet se soit déroulé de la même façon et en eût apporté la preuve aussi tôt.

11. Pour plus de détail, voir l'article de Heïdi Koumarios et Éric Sigaud sur Eno dans ce même numéro.

12. Dans la méthodologie des projets Insee, cette première étape importante permet de s'approprier l'existant, précise les objectifs, l'organisation, le planning, le budget et les risques encourus. Son analyse par le comité des investissements conditionne l'affectation des moyens.

13. *Data Documentation Initiative* (DDI) : pour plus d'information voir le site <https://ddialliance.org>.

14. À ce stade, il s'agissait seulement de proposer aux entreprises de répondre à un questionnaire LibreOffice remplissable, image du questionnaire papier habituel. La « véritable » collecte par internet ne serait atteinte que plus tard. Techniquement tout marchait, mais le taux de réponse restait faible.

● NOUVEAU REBONDISSEMENT, CETTE FOIS ENDOGÈNE

Début 2012, sachant qu'Eurostat devait lancer une réflexion stratégique sur la modernisation des enquêtes principalement sociales, l'Insee fit de même, avec un rapport prospectif formulant des recommandations stratégiques pour le développement de la collecte par internet et plus généralement des enquêtes multimodales, à la fois pour les ménages et les entreprises. Il fut rendu quasiment en même temps que l'expression des besoins du projet Coltrane, en avril 2012.

Les recommandations stratégiques concernaient principalement les enquêtes auprès des ménages, pour lesquelles les enjeux du passage à une collecte multimode apparaissaient extrêmement forts. Pour la partie concernant les entreprises, elles reprenaient les orientations de Coltrane qui se voyait ainsi conforté, et soulignaient entre autres l'importance des métadonnées actives. Mais elles engageaient aussi à réfléchir, sur un plan technique, à l'unicité, ou pas, d'une plateforme de collecte multimode commune aux ménages et aux entreprises, ou de la seule mise en commun de certains modules. De même, pour les enquêtes auprès des entreprises, le rapport questionnait le partage d'une plateforme entre les SSM et l'Insee. À court terme, la plupart des recommandations portaient sur les seules enquêtes auprès des ménages et pour ce qui concernait aussi les entreprises, sur la mise en place d'une coordination sur le thème général du multimode dans l'ensemble du Service statistique public.

En parallèle, l'analyse de l'expression des besoins de Coltrane par le comité des investissements, tout en ne remettant pas le projet en cause, le replaçait dans le contexte du système d'information existant pour les enquêtes auprès des ménages : or ce dernier accordait une place centrale à l'outil Blaise, largement distribué dans le monde, plus particulièrement adapté aux enquêtes en face-à-face, mais peu compatible avec l'architecture qui se dessinait pour Coltrane. Ces questions étaient cependant spécifiques à l'univers des enquêtes auprès des ménages. Les enjeux étaient totalement différents dans la sphère des enquêtes auprès des entreprises où la question de l'unicité d'une plateforme entre l'Insee et les SSM était prépondérante. Les recommandations du comité des investissements suite à l'expression des besoins, en insistant sur la nécessité d'une convergence stratégique entre les deux sphères « ménages » et « entreprises », abordaient une généralisation différente, qui n'était finalement pas la question centrale. De fait, une grande partie de cet avis consultatif n'a pas été mise en œuvre, et c'est heureux : en introduisant ce souhait de convergence non réellement pertinent à ce stade, elles auraient finalement été dommageables au déroulement du projet ; car elles lui auraient fait porter une complexité supplémentaire, en particulier dans la gouvernance liée aux enquêtes auprès des ménages, ce qui l'aurait certainement retardé.

● LA SPÉCIFICITÉ D'UNE PLATEFORME PARTAGÉE AVEC UN AUTRE MINISTÈRE

Une recommandation du comité des investissements restait toutefois à l'ordre du jour : la nécessité d'approfondir l'impact organisationnel d'une gestion commune des contacts dans les entreprises, entre les SSM et l'Insee : il apparaissait de plus en plus probable, en juin 2012, que la Dares¹⁵ allait adopter la plateforme Coltrane pour faire basculer l'enquête Acemo¹⁶ trimestrielle sur internet.

15. Direction de l'Animation de la recherche, des études et des statistiques au sein du ministère du Travail (service statistique de ce ministère).

16. Activité et Conditions d'Emploi de la Main d'Œuvre : l'enquête Acemo réalisée par la Dares est utilisée principalement pour la mesure des salaires de base, mais également pour les estimations provisoires d'emploi salarié.

De fait cela allait constituer un sous-projet de Coltrane. C'était tout à fait cohérent avec l'objectif de disposer d'une plateforme de collecte partagée entre l'Insee et les SSM, et constituait une vraie simplification pour les entreprises jusque-là interrogées séparément par les divers services statistiques. En revanche, si la plateforme était commune, cela signifiait par exemple que la *hot line* serait commune : les personnels de l'Insee chargés de cette *hot line* allaient répondre à des entreprises interrogées par la Dares, celle-ci devant conserver le contact avec les entreprises pour les questions « métier » ; on se demandait comment les entreprises allaient réagir et il fallait clarifier l'organisation entre tous les acteurs.

En réalité, la mise en place d'une *hot line* (en l'occurrence dans les « pôles Insee Contact » qui dépendent de la direction de la diffusion (et de l'action régionale), donc pas de la DSE) engendra beaucoup plus de questions à résoudre en interne à l'Insee, notamment d'organisation (horaires, compétences nouvelles à acquérir, champ de responsabilité par rapport aux

gestionnaires des enquêtes, représentation dans la gouvernance du projet), que le partage de la *hot line* par deux administrations : lorsqu'une entreprise a perdu son identifiant, elle se moque de savoir qui lui répond pourvu qu'on lui réponde.

« *Lorsqu'une entreprise a perdu son identifiant, elle se moque de savoir qui lui répond pourvu qu'on lui réponde.* »

L'ensemble des solutions apportées aux problèmes posés par le partage d'une même plateforme de collecte fut consignée dans une convention qui fixe les droits et les devoirs de chaque administration :

l'Insee mettait à disposition la plateforme de collecte, la *hot line*, et dû développer quelques fonctions spécifiques pour la Dares ; celle-ci fournissait le questionnaire qu'il fallut décrire avec le standard DDI, ainsi bien sûr que l'échantillon d'enquêtés ; en retour l'Insee renvoyait les réponses des enquêtés à la Dares ; les flux de données, les engagements de services de l'Insee et bien sûr une compensation financière y sont décrits en détail.

📍 **NAISSANCE DU PETIT DERNIER, L'OUTIL POGUES**

Coltrane poursuivait le cours normal d'un projet informatique d'ampleur, sans fait notable qui le distingue des autres... Il devait déboucher logiquement, à partir de la fin 2015, sur le passage progressif de l'ensemble des enquêtes sur la nouvelle plateforme, ce qu'on appelait la « phase deux » du projet, dont l'histoire nécessiterait un article en soi. Avançons donc dans le temps pour narrer une dernière péripétie. Nous sommes au printemps 2015. Les deux chefs de projet demandent l'autorisation d'interrompre leur travail courant pour se réunir une semaine avec d'autres informaticiens pour un « *hackathon* », un séminaire en résidentiel consacré à de la réflexion et de la programmation à haute dose ; ils ne veulent pas trop en dire, car ils ne sont pas sûrs de réussir. Ils se proposent de construire un outil qui générerait les métadonnées au standard DDI à partir de spécifications que pourrait donner un concepteur d'enquêtes au travers d'une interface ergonomique très simple.

La preuve complète de la faisabilité de l'outil baptisé Pogues ne fut donnée que quelques semaines plus tard à l'issue d'un nouveau séminaire de trois jours¹⁷. Il s'agissait essentiellement de montrer que la génération des métadonnées en DDI pouvait se faire sans investissement de la part des concepteurs d'enquêtes, et que le processus de mise en ligne

17. Voir l'article de Franck Cotton et Thomas Dubois sur Pogues, dans ce même numéro.

d'une enquête *via* la nouvelle filière leur ferait finalement gagner du temps. La plupart des participants au *hackathon* avaient suivi un cours sur le standard DDI ensemble et c'est de là que naquit l'idée d'aller plus loin et de créer une véritable « filière » qui en un seul « clic » allait de la formulation des questions de l'enquête à la génération de toute l'infrastructure de collecte permettant d'en récupérer la réponse.

La démonstration de leur prototype fut d'autant plus convaincante que la génération des métadonnées en DDI était le talon d'Achille de Coltrane : pour la phase initiale de mise au point des différents modules du projet, il n'y avait en gros qu'une enquête concernée tous les ans ; la mobilisation des experts en DDI était donc assez légère. Mais pour la seconde phase de généralisation, il allait falloir le faire à la chaîne pour une trentaine d'enquêtes et le processus devenait extrêmement dépendant d'un petit nombre d'experts. L'introduction du nouvel outil renversait le problème : soit l'outil allait considérablement aider les experts, soit il serait même suffisamment avancé pour être mis dans les mains des différents responsables d'enquêtes. Dans tous les cas, la phase de généralisation devenait beaucoup moins risquée.

Au bout du compte le projet Coltrane n'avait pas produit ce pour quoi il avait été programmé au départ, mais il avait produit beaucoup plus : les briques pour un futur atelier complet de conception et d'exploitation d'enquêtes, comportant une plateforme de collecte par internet nommée Coltrane, un outil ergonomique permettant aux concepteurs de produire les métadonnées d'enquêtes selon le standard DDI, nommé Pogues, et un générateur des supports de collecte à partir de ces métadonnées, applicable à tout mode de collecte, nommé Eno.

CONCLUSION

La gestion de l'innovation dans une organisation humaine est un sujet d'étude en sociologie des organisations (Alter, 2000 ; Brion, 2012 ; Laborde, 2017 ; Loilier et Tellier, 2013) : qui doit être à l'initiative ? Dans quelle proportion ? Sans un dispositif institutionnel pour gérer les demandes d'innovation venant de ses maîtrises d'ouvrage, un organisme se scléroserait et risquerait un lent dépérissement. Mais si on laissait réciproquement fleurir toutes les idées que pourraient avoir ses agents, une institution exploserait aussi sous le poids de la charge et l'incohérence finale de son système d'information. D'où la nécessité des procédures d'analyse des demandes et d'arbitrage.

“ *L'histoire de Coltrane montre qu'il est rare, si le processus disruptif va à son terme, que le produit final corresponde exactement à la demande initiale.* ”

Ce processus de régulation est particulièrement difficile à faire fonctionner dans le cas des projets qui se caractérisent par une rupture importante avec le passé. L'histoire de Coltrane montre qu'il est rare, si le processus disruptif va à son terme, que le produit final corres-

ponde exactement à la demande initiale ; c'est inévitable et montre la capacité d'adaptation de l'institution. Dans le cas de Coltrane, elle a été bouleversée par des événements imprévisibles exogènes et endogènes, que le projet et l'institut ont su utiliser à leur profit. Si on analyse *ex post* la demande initiale, on se rend compte que sans plateforme de collecte, on n'aurait eu en réalité aucune chance de mener à bien les tests de faisabilité et de performance correspondant à une nouvelle forme d'interrogation de l'ESA ; au passage, en 2019, ce besoin initial demeure et pourrait constituer une évolution de Coltrane ; en tout cas, les outils sont maintenant réunis pour la mener à bien.

L'histoire d'Eno et de Pogues montre enfin que les idées ne viennent pas toutes des maîtrises d'ouvrage et que les maîtrises d'œuvre jouent aussi un rôle quelquefois majeur dans la création directe d'innovation, en dehors des procédures normales de lancement des projets. Lorsque cela se produit, il faut que l'institution sache laisser une liberté de manœuvre aux quelques individus plus créatifs que les autres mais dont la créativité s'accommode souvent mal du parcours du combattant des procédures internes. Il faut aussi qu'*ex post* elle sache récupérer ces innovations pour réellement les institutionnaliser. On s'engage alors dans un autre long parcours, semé d'embûches, qui fait appel à d'autres types de compétences, pour transformer un prototype en filière opérationnelle, maintenable et accessible à un ensemble varié d'utilisateurs. Mais ce serait là le sujet d'un autre article...

BIBLIOGRAPHIE

ALTER, Norbert, 2000. L'innovation ordinaire. 1^{er} septembre 2000. Presses Universitaires de France, Collection Sociologies. ISBN 978-2-13-050902-8

BONNANS, Dominique, 2019. RMÉS, le référentiel de métadonnées statistiques de l'Insee. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 46-55. [Consulté le 14 octobre 2019]. Disponible à l'adresse :

<https://www.insee.fr/fr/statistiques/fichier/4168396/courstat-2-6.pdf>

BRION, Sébastien, 2012. L'organisation comme levier de l'innovation. [en ligne]. 6 février 2012. Gestion et management, Université Grenoble-Alpes, dossier d'habilitation à diriger des recherches. [Consulté le 30 octobre 2019]. Disponible à l'adresse :

<https://halshs.archives-ouvertes.fr/tel-01616646>


LABORDE, Olivier, 2017. *Innover ou disparaître*. Mai 2017. Dunod, Collection Stratégie d'entreprise. ISBN 978-2-1007-5824-1

LOILIER, Thomas et TELLIER, Albéric, 2013. *Gestion de l'innovation*. 29 août 2013. Éditions EMS, Collection Les essentiels de la gestion, 2^e édition. ISBN 978-2-84769-508-3

POGUES, UN OUTIL DE CONCEPTION DE QUESTIONNAIRES

Franck Cotton* et Thomas Dubois**

Permettre à un concepteur d'enquête de construire ses supports de collecte sans avoir le souci de connaître la moindre technologie, voilà tout l'enjeu de Pogues, nouvel outil de conception de questionnaires. Un questionnaire est un objet complexe composé de questions, filtres, contrôles, etc. représentant un ensemble de métadonnées statistiques de questionnement. Donner la main au concepteur pour à la fois documenter son questionnaire et réaliser les supports de collecte représentait un réel défi pour l'Insee. Cet outil de conception a été construit selon une démarche innovante, notamment au plan des méthodes de développement et des technologies. Il s'agit du premier développement en open source à l'Insee, offrant ainsi des opportunités de collaboration avec d'autres instituts statistiques. Il s'agit enfin de la démonstration par l'exemple de la notion de métadonnées actives, laquelle restait encore floue et théorique pour beaucoup. Avec cet « atelier du concepteur », une nouvelle page s'écrit dans la démarche d'industrialisation de la filière de collecte.

 *Allowing a survey designer to build his data collection instrument, without having to worry about knowing the slightest technology, is what Pogues, the new questionnaire design tool, is all about. A questionnaire is a complex object composed of questions, filters, checks, etc. representing a set of statistical metadata for questioning. Giving the designer a hand to both document his questionnaire and produce the collection instrument was a real challenge for INSEE. This design tool has been built using an innovative approach, particularly in terms of development methods and technologies. This is the first Open Source development at INSEE offering opportunities for collaboration with other statistical offices. Finally, it is a demonstration by the example of the notion of active metadata, which was still vague and theoretical for many. With this "designer workshop", a new page is being written in the approach of industrialising the collection process.*

En 2015, l'Insee s'est lancé dans la construction d'une filière d'industrialisation des enquêtes : une nouvelle plateforme de collecte par internet (Coltrane¹), dotée d'un générateur de questionnaires (Eno²), lui-même basé sur une description formelle des questionnaires dans le standard international DDI³, largement reconnu.

* Conseiller scientifique de la direction du Système d'information, Insee, franck.cotton@insee.fr

** Responsable de l'équipe Métadonnées statistiques, Unité Qualité, Insee, thomas.dubois@insee.fr

1. Voir l'article d'Olivier Haag et Anne Husseini-Skalitz dans ce même numéro.

2. Voir l'article de Heidi Koumarianos et Éric Sigaud dans ce même numéro.

3. *Data Documentation Initiative* (DDI) est un langage de description de métadonnées, entre autres les questionnaires. Pour plus d'information voir le site <https://ddialliance.org>.

De nouvelles enquêtes ont été progressivement intégrées dans ce dispositif. Mais l'écriture préalable des questionnaires en DDI est vite devenue un goulot d'étranglement, car la richesse de ce langage de description de métadonnées se paie par une grande verbosité⁴, qui saturait de plus en plus les rares ressources d'expertise dont l'institut disposait : les allers-retours entre concepteurs d'enquête et programmeurs avaient été déplacés vers les deux ou trois experts DDI, mais pas supprimés.

Pour rendre les concepteurs autonomes et maîtres de bout en bout du processus de création des supports de collecte, il fallait leur permettre d'écrire du DDI sans le savoir. Il fallait de plus qu'ils puissent observer rapidement les résultats de leurs modifications, sans avoir à recourir à un expert. L'idée était née : un outil de conception de questionnaire (Iverson et Smith, 2013), au sein d'un « atelier du concepteur », permettant de créer les modèles de questionnaires et de les visualiser en un clic, tout en alimentant directement le référentiel de métadonnées statistiques RMÉS (Bonnans, 2019).

🎯 UNE INITIATIVE COLLECTIVE ET INNOVANTE POUR L'INSEE...

L'idée de créer l'outil de conception de questionnaire, dénommé « Pogues », visait dans un premier temps à résorber la charge de modélisation des questionnaires en DDI. Elle devait bien vite s'enrichir d'autres problématiques présentes alors dans la sphère de la méthodologie et de l'informatique : le besoin de démontrer par l'exemple cette notion

« Le besoin de démontrer par l'exemple cette notion de métadonnées actives. »

de métadonnées actives (Iverson, 2010), qui restait encore floue et théorique pour beaucoup, le souhait de tester de nouveaux outils techniques pour la création d'applications *web*, ou celui d'expérimenter des méthodes de développement collaboratives, agiles et légères. Cette conjonction d'intérêts

a permis de mobiliser rapidement les ressources et les compétences nécessaires pour organiser le lancement de l'opération. Il va sans dire que les procédures de planification, d'attribution et de gestion des moyens furent un peu bousculées au passage.

Le développement initial de Pogues a été réalisé pour l'essentiel en avril 2015, lors d'un « *hackathon* »⁵ d'une semaine au pôle EuraTechnologies⁶ de Lille. Le *hackathon* rassemblait huit informaticiens et deux personnes du métier (métadonnées et collecte). Il fut précédé de quelques sessions de spécification selon des méthodes agiles, et suivi d'un autre *hackathon* plus court et en plus petit comité.



De bonnes fées s'étaient apparemment penchées sur le berceau de Pogues, car le premier *hackathon* se révéla parfait à tous points de vue : opportunité en termes de calendrier, mélange de compétences, déroulement, localisation, et même météo... Les résultats dépassaient les espérances sur la plupart des objectifs :

4. Près de 15 000 lignes pour 11 pages papier du questionnaire sur les TIC en entreprises.

5. Voir <https://fr.wikipedia.org/wiki/Hackathon>.

6. EuraTechnologies est un pôle d'expertise et d'innovation (<https://www.euratechnologies.com>).

- ❶ une application *web* suffisamment fonctionnelle pour pouvoir être montrée à de futurs utilisateurs afin de recueillir leurs avis et leurs souhaits d'évolution ;
 - ❷ une architecture applicative d'ensemble bien spécifiée, notamment du point de vue des liens entre les différents composants (interface Pogues, services de *back-office*, générateur Eno, référentiel RMÉS...);
 - ❸ conséquemment, une répartition des rôles et un processus bien défini pour les développements informatiques (plus de formalisation et l'appui plus systématique sur des standards) ;
 - ❹ surtout, la réussite de l'objectif principal de génération du formulaire *web* en un clic, grâce à l'appel par Pogues du générateur Eno, adapté pour l'occasion sous forme de service *web*.
- Sur le plan de l'innovation, Pogues s'inscrivait dans la lignée d'Eno, avec trois nouveautés pour l'Insee :
- ❶ Pogues a inauguré ce qui allait devenir la filière de développement d'applications *web* clientes (filière JavaScript) de l'institut ;
 - ❷ le projet a été déposé en *open source* sur la plateforme en ligne de développement de logiciels et de gestion de versions GitHub : l'organisation InseeFr y a été créée à cet effet, et abrite aujourd'hui le code d'une vingtaine d'applications ou de modèles⁷ ;
 - ❸ son développement a été internationalisé dès l'origine, avec code, commentaires et documentation en anglais.

🌐 ...À RÉINTÉGRER AU SEIN D'UNE ORGANISATION

L'application a continué de s'enrichir durant les mois qui ont suivi le *hackathon*, sur la base de contributions volontaires d'abord, ensuite par le biais de prestations de services légères sur différents aspects : développements complémentaires, documentation, ergonomie. En parallèle, l'atterrissage « organisationnel » de Pogues se mettait en place : il fallait préciser

« L'option retenue fut de considérer que Pogues était essentiellement une application de gestion de métadonnées. »

à qui appartenait vraiment cet objet collectif surgi sans crier gare. L'option retenue fut de considérer que Pogues était essentiellement une application de gestion de métadonnées⁸, en l'occurrence des descriptions de questionnaires, et qu'il s'agissait donc d'un « satellite » du référentiel de métadonnées RMÉS.

Ce positionnement devait s'avérer d'une grande utilité pour donner un sens moins théorique au référentiel des métadonnées statistiques : Pogues, outil destiné aux concepteurs, devenait de fait le moyen de concrétiser l'apport de RMÉS à leur métier.

L'unité Qualité, au sein de la direction de la Méthodologie et de la coordination statistique et internationale de l'Insee (DMCSI), propriétaire de RMÉS, a donc pris en charge le devenir de l'outil Pogues. En particulier, elle a organisé en octobre 2016, un groupe de travail au sein duquel les concepteurs d'enquête ont pu tester Pogues d'une façon approfondie et exprimer leurs souhaits d'évolution sur la base de cas d'utilisation concrets. Cette rencontre a permis de formaliser une feuille de route pour les futurs développements de Pogues

7. Voir sur le site <https://github.com/InseeFr>.

8. Voir (Bakkomen, Orten et Prestage, 2014) et (Butt, Norland et Orten, 2018).

(et d'Eno). Elle a offert à ces outils une bonne visibilité parmi les responsables d'enquêtes auprès des entreprises comme auprès des ménages.

L'unité Qualité s'occupait également des contacts avec la DDI Alliance, l'organisation responsable de l'évolution du standard DDI⁹. Les besoins de Pogues et Eno, qui apparaissaient au fur et à mesure de la prise en compte de questionnaires de plus en plus complexes, nécessitaient en effet d'ajouter des précisions ou des compléments à DDI¹⁰. Il aurait été aisé d'intégrer ces modifications dans une déclinaison du standard spécifique à l'Insee. Cependant, conformément aux principes de RMÉS, et même si cela requerrait plus de

temps et d'énergie, il a été décidé de contribuer à l'enrichissement du standard DDI lui-même, afin que toute la communauté puisse bénéficier des nouvelles fonctionnalités.

« Contribuer à l'enrichissement du standard DDI lui-même, afin que toute la communauté puisse bénéficier des nouvelles fonctionnalités. »

Une prestation de services a démarré au printemps 2017, pour remettre à plat et développer la couverture fonctionnelle de Pogues : intégration de questions sous forme de tableaux, augmentation du nombre de cas de questions-filtres, etc. En interne, d'importants travaux de consolidation de l'architecte

ture ont conduit à la mise en production « officielle » d'une première version de l'application en avril 2018. Enfin, des supports de documentation et de communication (guides de l'utilisateur et du développeur, vidéo de démonstration, création de logos, présentations dans diverses réunions nationales ou internationales...) ont permis de donner à l'outil Pogues une bonne visibilité à l'Insee et au-delà.

❶ L'AMBITION DE LA COUVERTURE FONCTIONNELLE : SUPPORTER TOUS LES CAS DE FIGURE

Pogues permet au concepteur de structurer son questionnaire en séquences (modules et sous-modules) regroupant des questions et des éléments textuels, ainsi que de définir la logique de parcours du questionnaire par l'enquêté et les contraintes d'ensemble pour le remplissage (**encadré 1**).

Différentes sortes de questions sont supportées : simple à champ de saisie unique (texte, numérique, booléen, date), à choix unique ou multiple, tableau, les valeurs pouvant être contraintes par des bornes ou des listes de codes prédéfinies¹¹.

À chaque question correspondent une ou plusieurs variables dites « collectées » (c'est-à-dire que leur valeur sera fournie par le répondant), mais on peut également définir des variables calculées ou externes. Les premières s'obtiennent par des formules généralement simples à partir des variables collectées, et servent notamment à spécifier des contrôles de cohérence. Les variables externes ne sont pas collectées dans le questionnaire mais sont utiles pour sa personnalisation : il peut s'agir de valeurs obtenues antérieurement et rappelées pour faciliter le remplissage, contrôler les évolutions ou fixer certains éléments du questionnaire (vague de collecte, zonage géographique, dernier effectif connu, etc.).

9. Voir le site <https://ddialliance.org>

10. En particulier pour la modélisation de questions prenant la forme de tableaux complexes.

11. Pour plus de détail sur les fonctionnalités de l'outil Pogues voir le guide de l'utilisateur disponible sur la plateforme GitHub : <http://insee.fr.github.io/Pogues/fr/guide-utilisateur/guide-utilisateur.html>.

Outre les séquences et les questions, Pogues permet de spécifier différents éléments textuels : commentaires, consignes ou aide pour le remplissage, messages d'avertissement, etc. Selon le type de support de collecte qui sera produit, ces éléments apparaîtront sous des formes variées, par exemple des infobulles sur les formulaires *web*. Enfin, il est possible de décrire certains composants logiques, pour les modes interactifs (questionnaire *web* auto administré, ou enquête en face-à-face) :

- ❶ des expressions conditionnelles : typiquement, l'expression d'une question pourra dépendre des valeurs des réponses déjà obtenues ;
- ❷ des contrôles permettant de valider les données recueillies : ces contrôles peuvent porter sur une question ou sur plusieurs (par exemple : vérifier la cohérence entre le total du chiffre d'affaires déclaré et sa répartition par activité) ;
- ❸ des filtres : en fonction des données déjà collectées, le répondant peut être redirigé sur des questions différentes ou certaines parties du questionnaire peuvent être désactivées.

Ces expressions logiques sont écrites dans un langage *ad hoc* simple. Des recherches sont en cours pour étudier l'utilisation d'un langage plus standard, par exemple VTL¹².

« Pogues propose une interface graphique pratique et intuitive. »

Pogues propose une interface graphique pratique et intuitive pour décrire le questionnaire et ses composants (**encadré 1**). Les questions et séquences peuvent être déplacées à la souris ou dupliquées, et une représentation simplifiée de la structure du questionnaire permet de s'y déplacer

rapidement. À tout moment, l'utilisateur peut visualiser son questionnaire sous la forme de son choix (*web* ou pdf), l'exporter en DDI ou produire le document de spécification détaillée qui lui correspond.

❶ LA CONCEPTION D'UN QUESTIONNAIRE MULTIMODE

Le questionnement n'est pas identique selon le mode de collecte, que ce soit sur le fond ou sur la forme. Par exemple, sur le fond, une consigne peut différer selon qu'elle est adressée directement à un enquêté ou à un enquêteur (**encadré 2**). Sur la forme, les fonctionnalités dynamiques qu'offre le *web* (contrôles en ligne, filtres dynamiques, listes déroulantes, etc.) ne peuvent se retranscrire à l'identique sur le format papier. Ces deux aspects doivent ainsi être pris en compte dans la mise en place de la nouvelle filière, où le concepteur doit pouvoir être autonome dans la conception et la construction¹³ des questionnaires d'une enquête multimode.

L'outil Pogues doit offrir au concepteur la possibilité de distinguer des éléments relevant d'un seul mode, tout en mutualisant le plus possible ce qui peut l'être (les questions communes à tous les modes ne sont décrites qu'une seule fois).

12. *Validation and Transformation Language*, publié par l'initiative SDMX, voir https://sdmx.org/?page_id=5096.

13. Le GSBPM distingue les phases « Concevoir » et « Construire », mais dans le cas d'un processus statistique piloté par les métadonnées, la construction (ici du support de collecte) est en quelque sorte un produit fatal de la conception. Voir *infra*.

Sur la forme, il s'agit d'appliquer des principes déterministes, à savoir que le mode de collecte et le format du questionnaire suffisent à déterminer la présentation d'un composant du questionnaire. Ainsi, certaines différences sont interprétées directement par le générateur de questionnaires : disparition des comportements dynamiques, filtres remplacés par un simple texte explicitant la redirection ou présentation d'une liste déroulante du *web* en simple champ de saisie.

POGUES, ÉDITEUR DE MÉTADONNÉES STATISTIQUES

Un questionnaire et ses composants (consignes, formulations de question, formats de réponse, etc.) représentent un ensemble d'informations caractérisant la donnée statistique collectée. Le questionnaire est ainsi un ensemble de métadonnées statistiques parmi d'autres ayant vocation à se retrouver dans le référentiel dédié RMÉS. Pogues est à la fois un outil de conception de questionnaires et un outil de gestion des métadonnées décrivant les questionnaires.

Pogues n'a de sens en tant qu'éditeur de métadonnées que si les composants des questionnaires (séquences, questions, liste de codes) peuvent être facilement réutilisés. Il faut éviter de retomber dans la situation qui prévalait à l'Insee avant RMÉS : les métadonnées étaient réécrites d'une opération à l'autre, d'une année à l'autre, ou réutilisées mais adaptées au contexte de chaque enquête.

Encadré 2. Adaptation des consignes aux différents modes

Dans une enquête en face-à face, certaines consignes sont destinées à être lues par l'enquêteur pour l'enquêté, et d'autres sont à destination de l'enquêteur uniquement. Il peut s'agir par exemple de préciser des concepts compliqués. Pour passer à un questionnaire auto-administré, on doit reformuler, voire restructurer le questionnement. Ces différences de forme dans les modèles de questionnaires interviennent également dans la gestion de la non-réponse ou des réponses imprécises.

L'exemple ci-dessous illustre des consignes données à l'enquêteur pour la variable « date de naissance ». En auto-administré, on n'imagine pas les reproduire à l'identique, on préférera mettre des contrôles pour relancer en cas de non-réponse partielle, sur l'année par exemple.

**Quelle est la date de naissance exacte
de PRÉNOM ?
en clair JJ/MM/AAAA**

← question posée

Elle doit toujours être renseignée.

- Si le jour est inconnu, saisir 15.
- Si le mois est inconnu, saisir 06 (juin).
- Si l'année est inconnue, saisir une année vraisemblable.

← consignes données
à l'enquêteur

Pour cela, la communication entre Pogues et le référentiel de métadonnées statistiques doit permettre la recherche et la récupération de composants unitaires¹⁴ du questionnaire, ainsi que leur dépôt dans le référentiel lorsque la conception d'une version stable est achevée.

Jusqu'à ce jour, le développement de l'interface entre Pogues et RMÉS s'est heurté à des difficultés techniques et à des choix privilégiant d'autres priorités ; elle reste de ce fait fonctionnellement limitée. Il devient maintenant urgent d'achever la connexion de Pogues au référentiel RMÉS, ce devrait être entrepris dans les prochains mois.

❶ UN NOUVEAU PROCESSUS D'INTÉGRATION DU QUESTIONNAIRE DANS L'ÉCOSYSTÈME DES OUTILS DE COLLECTE

L'arrivée d'une première version de Pogues en production en avril 2018 a modifié le processus de spécification, de réalisation et de validation des questionnaires (d'enquêtes auprès des entreprises jusqu'à présent).

Désormais, il est nécessaire d'organiser une rencontre entre le concepteur d'enquête et les responsables du nouvel « écosystème de collecte » (*figure 1*) : l'administrateur de l'outil Eno qui est également le méthodologue de questionnement, l'administrateur de l'outil Pogues et l'administrateur de la plateforme de collecte Coltrane.

Au cours de cette réunion préparatoire, quatre points importants sont abordés :

- ❶ les bonnes pratiques de conception des questionnaires : l'intégration d'une enquête dans une nouvelle filière est l'occasion d'un échange entre concepteurs et méthodologues afin de réinterroger les choix en termes de contenu de questionnaire. L'objectif est d'aboutir à un questionnaire respectant les règles de bonnes pratiques de conception¹⁵, dont la plupart sont déjà embarquées dans Pogues et Eno. Cela peut conduire à une évolution non négligeable du questionnaire pré-existant ;
- ❶ la standardisation et ses conséquences sur le questionnaire : ici sont abordés les sujets de forme¹⁶ y compris ceux touchant à la structure du questionnaire, de distinction entre libellés de questions et consignes, numérotation, etc. La standardisation de la forme sert également les bonnes pratiques ;
- ❶ les besoins non couverts par l'ensemble des outils, et les délais de mise à disposition des évolutions éventuelles ;
- ❶ l'établissement d'un planning de réalisation du modèle de questionnaire.

« Avec Pogues, il s'agit de remettre le concepteur au centre du jeu. »

Cette rencontre permet de formaliser les améliorations en termes de conception de questionnaires, les engagements en termes de fonctionnalités offertes par l'écosystème de collecte mais aussi les limites fonctionnelles.

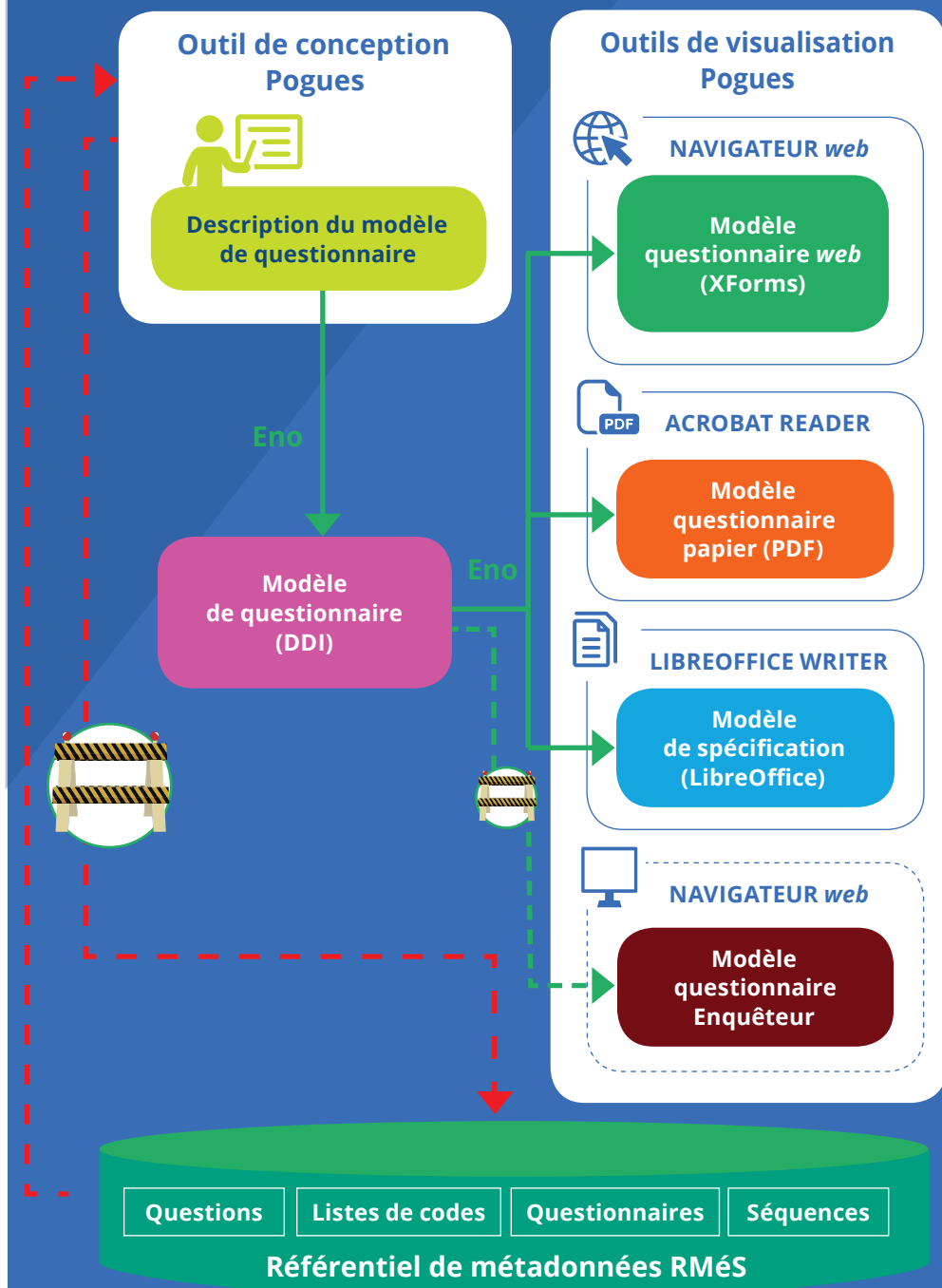
Avec Pogues, il s'agit de remettre le concepteur au centre du jeu. En effet, le concepteur saisit désormais son questionnaire directement dans Pogues

14. Voir (Greenough, Mechanda et Rizzolo, 2014).

15. Pour plus de précisions, voir les références de l'article de Heidi Koumarianos et Éric Sigaud dans ce même numéro.

16. Par exemple : pas de soulignement dans un questionnaire *web* si cela ne correspond pas à un hyperlien.

Figure 1. Sous le capot de Pogues



Note de lecture : le moteur Eno est utilisé à deux reprises, d'abord en sortie de Pogues pour générer le modèle DDI, et ensuite, à partir de celui-ci, pour générer les supports de collecte *web*, papier, « rendu spécification ». Le même procédé pourrait être utilisé dans le futur pour les modèles de questionnaires Enquêteur. Les liens avec RMÉS sont en cours de développement.

et, en un seul clic, visualise et valide le rendu *web*, le rendu papier voire le rendu « spécification » (au format bureautique). Les délais de mise au point d'un questionnaire sont fortement réduits du fait de la suppression des allers-retours entre le concepteur et le développeur de questionnaires ou entre le concepteur et l'expert DDI (quand le générateur de questionnaires Eno existait mais pas encore Pogues). Il est à noter qu'à ce stade, la recette est réalisée sans personnalisation du questionnaire. Cette étape sera réalisée au sein de la plateforme de collecte.

Au jour de parution de cet article, le périmètre fonctionnel des différents outils est encore incomplet. Cela implique encore certains ajustements manuels. Mais plus Pogues et Eno augmenteront leur périmètre fonctionnel, plus le concepteur deviendra autonome et plus les délais de mise au point d'un questionnaire se réduiront.

La question d'une intervention tierce pour la recette des questionnaires complexes reste cependant posée : les questionnaires des enquêtes auprès des ménages, qui n'ont pas encore basculé dans le nouvel écosystème, sont actuellement mis au point avec l'aide d'un intermédiaire¹⁷ entre le concepteur et l'informaticien ; la question du rôle qu'il pourrait jouer à l'avenir n'est pas encore tranchée.

AU-DELÀ DE LA COLLECTE ?

Pogues, on l'a indiqué, permet au responsable d'enquête de concevoir son questionnaire, y compris les contrôles qui s'exécuteront lors du remplissage. Cela correspond aux sous-processus « Conception des descriptions de variables » et « Conception de la collecte » dans la représentation du modèle générique de description d'un processus de production statistique (GSBPM¹⁸).

La structure formelle de questionnaire qui est ainsi spécifiée, avec les questions et les variables associées, peut être réutilisée plus en aval dans le déroulement du processus de production statistique, et en particulier lors de la phase de contrôle des données collectées. C'est bien ce qui a été noté lors des études préliminaires du projet de développement d'un poste générique de traitement des enquêtes¹⁹. Il a de ce fait été convenu que la structure des questionnaires serait transmise par RMÉS dans la phase de développement des contrôles, pour être réutilisée et ainsi générer automatiquement les écrans de reprise.

On peut aller plus loin et imaginer que les contrôles de reprise post-collecte soient spécifiés dans Pogues, où toutes les définitions de variables sont déjà disponibles pour ce faire. On étendrait alors la portée de l'outil Pogues au sous-processus « Conception du traitement et de l'analyse ». Et pourquoi s'arrêter là ? Pogues pourrait permettre également de spécifier le processus de collecte (« Conception de systèmes de production et du déroulement des travaux ») : modes, calendrier de collecte, etc. toutes métadonnées qui peuvent activer de futurs services automatisés (déclenchement des relances, etc). Pogues deviendrait alors une sorte d'atelier du concepteur d'enquête, qui sert non pas juste à concevoir le questionnaire, mais plus largement à concevoir le processus d'enquête.

17. Il s'agit du CPOS, ou chef de projet en organisation statistique

18. Le *Generic Statistical Business Process Model* est le modèle de référence dans la communauté mondiale de la statistique officielle pour la représentation et l'analyse de haut niveau des processus statistiques (<https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1>).

19. Prévu pour les enquêtes thématiques auprès des entreprises, le projet Générique vise à développer un poste de traitement de la collecte, dont les entrants sont les questionnaires et dont les contrôles constituent de nouvelles métadonnées.

UNE OUVERTURE AU-DELÀ DE L'INSEE

Pogues est aujourd'hui à l'Insee inclus dans l'offre de service du référentiel RMÉS. Du fait de la portée de ce dernier, Pogues sera ouvert à terme aux services statistiques ministériels (SSM). Dans le même temps, la plateforme de collecte Coltrane a pour objectif de collecter un maximum d'enquêtes du service statistique public (SSP) afin que les entreprises qui répondent à plusieurs enquêtes n'aient qu'une seule infrastructure à utiliser. Coltrane a ouvert fin 2019 ses applications (comme la gestion des contacts) aux producteurs du SSP qui ont adopté ses services. Il est donc naturel que cette offre soit complétée par une ouverture des outils Pogues et Eno aux SSM qui en ont besoin. Il en sera de même avec la nouvelle plateforme de collecte par internet des enquêtes auprès des ménages.

« Un système d'information complet et cohérent. »

En quelques années, le petit graphique originel Eno, issu du projet Coltrane, a donné naissance avec Pogues à un système d'information complet et cohérent, devenu le standard à l'Insee pour

les enquêtes auprès des entreprises. Cette transformation, à laquelle beaucoup ont contribué, s'est opérée grâce à une combinaison de constance dans l'adhésion à la vision stratégique et de souplesse dans l'adaptation tactique aux changements du contexte. Elle a été alimentée par une recherche permanente de l'innovation, aux plans technique, méthodologique et organisationnel. Il a parfois fallu prendre quelques risques, brusquer un peu l'organisation ou convaincre certaines résistances, parfois sans succès, et composer avec des moyens souvent très contraints, mais on peut considérer que le résultat est conforme aux ambitions initiales. Cette réussite a largement crédibilisé la démarche de pilotage des systèmes d'information par les métadonnées, qu'il paraît maintenant plus naturel d'étendre à d'autres domaines comme la diffusion.

Eno et Pogues ont aussi soulevé des problématiques importantes, pas totalement instruites à ce jour : comment organiser la maîtrise d'ouvrage des projets transverses et des outils qui en résultent, comment mieux gérer l'innovation technique en production, etc.

Mais l'histoire ne fait que commencer. Eno et Pogues s'engagent maintenant dans une nouvelle phase de développement, dans le cadre de la plateforme pour la collecte internet des enquêtes auprès des ménages. Des enrichissements fonctionnels sont déjà engagés, et des améliorations de l'architecture prévues. On espère aussi que les efforts de promotion auprès des Instituts nationaux de statistiques des autres pays finiront par payer et que les marques d'intérêt se convertiront en réutilisations ou en contributions effectives. Cela passera probablement par une professionnalisation de l'organisation en interne pour mieux documenter l'outil et améliorer le support aux utilisateurs.

Surtout, il faut souhaiter que le futur d'Eno et Pogues continuera de se construire sur cette volonté d'innovation que Brian Eno, d'une certaine façon, décrivait quand on l'interrogeait sur son processus créatif : « *The [...] thing is to set up a situation that presents you with something slightly beyond your reach* »²⁰.

20. « Le truc [...] c'est de mettre en place une situation qui vous présente quelque chose qui est légèrement hors de votre portée » (http://music.hyperreal.org/artists/brian_eno/interviews/musn79.html).

BIBLIOGRAPHIE

BAKKMOEN, Håvard Venge, ORTEN, Hilde et PRESTAGE, Yvette, 2014. The DASISH Questionnaire Design Documentation Tool: Keeping track of the Questionnaire Design Process. In : *EDDI14 – 6th Annual European DDI User Conference*. [en ligne]. 2-3 décembre 2014. Institute of Education, University of London. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi14/paper/download/164/98>

BONNANS, Dominique, 2019. RMÉS, le référentiel de métadonnées statistiques de l'Insee. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 46-55. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168396/courstat-2-6.pdf>

BUTT, Sarah, NORLAND, Stig et ORTEN, Hilde, 2018. The Questionnaire Design and Documentation Tool (QDDT) a DDI based tool for assisting questionnaire design teams in their work. In : *site de la plateforme Zenodo*. [en ligne]. 4-5 décembre 2018. EDDI1 8 – 10th Annual European DDI User Conference, Deutsche Institut für Wirtschaftsforschung, Berlin. [Consulté le 14 octobre 2019]. Disponible à l'adresse : https://zenodo.org/record/2530046/files/EDDI_2018_The%20Questionnaire%20Design%20and%20Documentation%20Tool%20%28QDDT%29%20v3_2.pptx?download=1

DANNEVANG, Flemming et NIELSEN, Mogens Grosen, 2017. Towards Common Metadata Using GSIM and DDI 3.2. In : *IASSIST Quarterly*. [en ligne]. 24 février 2017. Vol 40, N° 2 (2017), Summer 2016, pp. 6-17. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <https://www.iassistquarterly.com/index.php/iassist/article/view/782/774>

GREENOUGH, Carmen, MECHANDA, Kaveri et RIZZOLO, Flavio, 2014. Metadata in the modernization of statistical production at Statistics Canada. In : *European Conference on Quality in Official Statistics, Vienna*. [en ligne]. 2-5 juin 2014. [Consulté le 14 octobre 2019]. Disponible à l'adresse : http://www.q2014.net/fileadmin/user_upload/Metadata_in_the_modernization_of_statistical_production_15-05-2014.docx


IVERSON, Jeremy, 2010. Metadata-Driven Survey Design. In : *IASSIST Quarterly*. [en ligne]. 11 novembre 2010. Vol 33, N° 1-2 (2010), Spring/Summer 2009, pp. 7-9. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <https://www.iassistquarterly.com/index.php/iassist/article/view/644/636>

IVERSON, Jeremy et SMITH, Dan, 2013. Generating Blaise Surveys from the Data Documentation Initiative's Metadata Standard using Colectica. In : *EDDI13 – 5th Annual European DDI User Conference*. [en ligne]. 3-4 décembre 2013. Réseau Quetelet – French Data Archives for Social Sciences. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://www.eddi-conferences.eu/ocs/index.php/eddi/EDDI13/paper/view/98/67>

ENO, UN GÉNÉRATEUR D'INSTRUMENTS DE COLLECTE

Heidi Koumarianos* et Éric Sigaud**

Eno, c'est l'histoire d'une innovation qui est devenue un maillon important du processus de collecte d'information de l'Insee. Partant d'un concept novateur, l'activation de métadonnées pour industrialiser un processus statistique, Eno en a apporté une application concrète à l'Insee : la génération d'instruments de collecte à partir des métadonnées. La réalisation d'un prototype a permis de valider le concept, ainsi que le choix des standards et de la filière technique. L'objectif de généricité a conduit à mettre en place une architecture souple et évolutive, aujourd'hui utilisée pour produire la quasi-totalité des questionnaires d'enquêtes auprès des entreprises, sous format web et papier. En portant une exigence accrue de formalisation et de standardisation, l'outil a également contribué à une harmonisation des pratiques et, de ce fait, des questionnaires. Le caractère transversal d'Eno a par ailleurs nécessité l'adoption d'une gouvernance inédite pour l'Insee, afin notamment de faciliter son extension à la sphère sociale. Pour cette dernière, de nouvelles fonctionnalités et une évolution technologique seront nécessaires, afin d'outiller la collecte multimodale auprès des ménages. Enfin, les méthodes de développement employées ouvrent des perspectives pour l'utilisation d'Eno par d'autres organismes statistiques.

 *ENO is the story of an innovation that has become an important component of INSEE's data collection process. Starting from an innovative concept, the activation of metadata to industrialize a statistical process, ENO has provided INSEE with a concrete application: the generation of collection instruments based on metadata. The realization of a prototype has made it possible to validate the concept, as well as the choice of the standards and the technical field. The objective of genericity has led to the implementation of a flexible and scalable architecture, which is now used to produce almost all business survey questionnaires, in both web and paper formats. By increasing the need of formalisation and standardisation, the tool has also contributed to the harmonisation of questionnaires and practices. Its transversal nature also required the adoption of a new form of governance for INSEE, in order to facilitate its extension to the social domain in particular. This extension will require the development of new functionalities and technological evolution, to enable mixed-mode data collection for households. The development methods used open the use of ENO to other statistical agencies.*

* Méthodologue, direction de la Méthodologie et de la coordination statistique et internationale, Insee, heidi.koumarianos@insee.fr

** Directeur adjoint délégué au programme Métallica (MÉTadonnées Actives, Logiciels Libres et Infrastructure pour une Collecte Assistée), direction des Statistiques démographiques et sociales, Insee, eric.sigaud@insee.fr

Les métadonnées associées à une enquête ne se limitent pas à des informations facilitant la lisibilité : on trouve aussi des métadonnées « actives », permettant la normalisation et l'industrialisation des processus statistiques (Bonnans, 2019). La collecte des données constitue un domaine propice à la mise en œuvre de cette approche : il s'agit d'un des postes de dépenses les plus importants pour un institut statistique et les processus, souvent répétés, y sont en général bien identifiés et formalisés. À partir de 2012, l'Insee a totalement refondu son système de collecte des enquêtes auprès des entreprises, avec la mise en œuvre du projet Coltrane¹. Ce contexte a constitué une opportunité pour une première mise en œuvre de métadonnées actives pour la génération d'instruments de collecte, à travers l'outil Eno.

ENO, UNE IDÉE

Pour décrire les fonctionnalités d'Eno², on prend un raccourci en parlant de « générer automatiquement des questionnaires ». Cette idée est séduisante en soi : elle permet de réduire les coûts de développement, entraîne de facto une standardisation des pratiques et les gains en qualité afférents, une puissance et une capacité à même de supporter un rythme et une fréquence d'enquêtes plus importants.



Mais derrière Eno, il y a une idée plus générale. Lorsqu'on sait capturer des métadonnées, qu'elles soient relatives au questionnaire, à la gestion ou à la qualité d'un processus, aux produits de diffusion, cela offre potentiellement bien plus que des gains en qualité et en maîtrise du processus statistique : on peut également imaginer de générer un jour jusqu'aux outils nécessaires au déroulé dudit processus.

Pour commencer à concrétiser ce concept, il fallait démontrer sa viabilité sur un cas « simple » : la spécification des questionnaires. Le principe fondateur d'Eno est d'automatiser la production des « instruments de collecte » (formulaires papier ou *web*, modules de collecte des enquêteurs, etc.) à partir d'une description formelle et abstraite du questionnaire sous forme de métadonnées standardisées. On retrouve cette idée dès la première spécification d'Eno, qui prit la forme du schéma en **figure 1**, tracé sur un tableau blanc.

« Le principe fondateur d'Eno est d'automatiser la production des « instruments de collecte ». »

La vision d'ensemble était donc de produire automatiquement les supports associés aux différents modes de collecte à partir d'une description formelle³ et unique⁴ du modèle de questionnaire. Eno, figuré par la flèche verte dans la **figure 1**, serait au cœur de cette vision. Dès le début, il était clair également qu'une mise

La vision d'ensemble était donc de produire automatiquement les supports associés aux différents modes de collecte à partir d'une description formelle³ et unique⁴ du modèle de questionnaire. Eno, figuré par la flèche verte dans la **figure 1**, serait au cœur de cette vision. Dès le début, il était clair également qu'une mise

1. Voir l'article de Jean-Marc Béguin sur la naissance d'une innovation en production statistique dans ce même numéro.
2. En référence à Brian Eno, musicien et producteur anglais qui popularisa notamment le concept de musique générative (https://en.wikipedia.org/wiki/Generative_music). Les outils associés à la plate-forme Coltrane sont généralement baptisés avec des noms de musiciens.
3. S'appuyant sur DDI (voir *infra*).
4. On verra plus loin que l'extension aux enquêtes multimodales de la sphère sociale a conduit par la suite à questionner l'unicité du modèle, sans toutefois la remettre totalement en cause.

en œuvre efficace de celle-ci imposait des choix d'implémentation forts et novateurs, en particulier une architecture de transformation de documents qui n'existait alors pas vraiment à l'Insee.

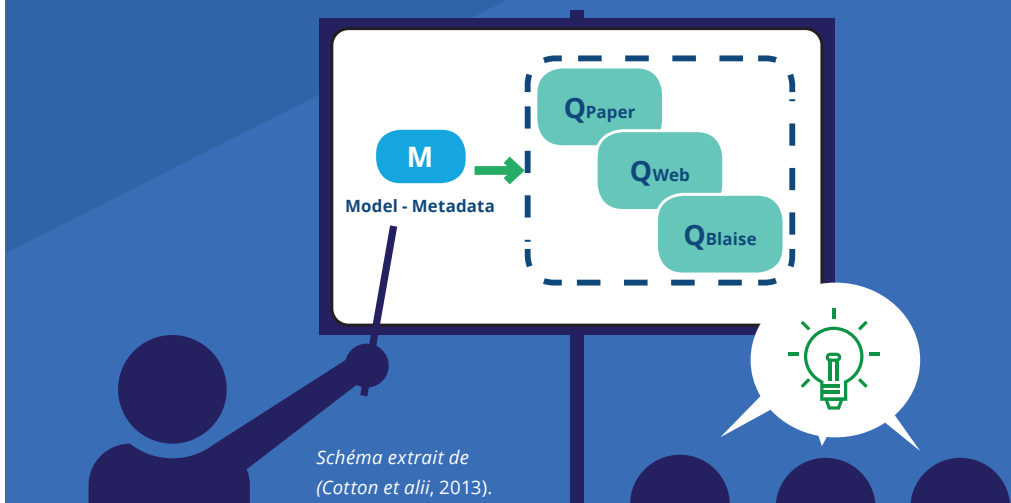
UN PROTOTYPE POUR VALIDER LE PRINCIPE DE GÉNÉRATION

Le projet Coltrane, plateforme pour la collecte internet auprès des entreprises⁵, offrait les circonstances favorables, les moyens humains et techniques, et un cas concret relativement urgent pour le développement d'un prototype de générateur : la dématérialisation de l'enquête structurelle annuelle auprès des entreprises (ESA). Celle-ci s'appuyait alors sur 200 modèles de questionnaires⁶ dont la réalisation était jusque-là confiée à un prestataire externe.

Le choix de cette enquête était doublement intéressant : d'une part sa complexité était à même de prouver la validité du principe, au-delà de son cas particulier ; et d'autre part, la maîtrise retrouvée du processus de réalisation des questionnaires était susceptible d'offrir un gain financier et en qualité.

Plus largement dans le contexte métier de l'époque, la production de questionnaires pour les enquêtes auprès des entreprises se faisait « à façon », sans capitalisation des développements d'une enquête à l'autre et avec peu de documentation. Il s'agissait d'un processus long (autour de six mois de délai), coûteux (de deux à trois développeurs mobilisés sur la période) et risqué (le caractère « artisanal » pouvait générer des bugs difficiles à identifier).

Figure 1. Eno, une idée : un modèle unique, plusieurs supports de collecte



5. Voir l'article de d'Olivier Haag et Anne Husseini-Skalitz dans ce même numéro.

6. Le nombre élevé de questionnaires était notamment hérité de l'empilement des spécificités sectorielles, lorsque l'Insee a fondu en un seul processus la quasi-totalité des enquêtes structurelles auprès des entreprises. Depuis, les ESA ne s'appuient plus que sur... 48 modèles.

Il y avait ainsi un intérêt certain à la mise en œuvre d'un générateur permettant d'industrialiser la production des questionnaires : gains en coût et en temps (automatisation), gain en qualité (standardisation, traçabilité), récupération de la maîtrise des spécifications (métadonnées comme « input » du générateur).

« La transversalité de l'outil était présente dès sa conception. »

Après le prototype pour l'enquête structurelle annuelle auprès des entreprises, une première version d'Eno a été développée afin de générer le questionnaire *web* d'une enquête du ministère du Travail⁷, puis enrichie progressivement pour répondre à l'ensemble des besoins des enquêtes intégrant la plateforme Coltrane. Ces enquêtes émanaient de différentes maîtrises d'ouvrage appartenant au Service statistique public⁸. La transversalité de l'outil était présente dès sa conception.

Le générateur a été développé par étapes, et le graphique simpliste des origines, s'il restait le cadre directeur, a évolué au cours des années pour devenir un système d'information riche, couvrant notamment la production de différents formats pour les supports de collecte.

❶ DÉCRIRE FORMELLEMENT LA COMPLEXITÉ SÉMANTIQUE DES QUESTIONNAIRES

Eno a été conçu dès l'origine dans le respect des principes qui fondaient également le développement concomitant du référentiel de métadonnées statistiques RMÉS (Bonnans, 2019) : utilisation des standards internationaux et activation des métadonnées statistiques.

Le questionnaire d'enquête, objet métier central du générateur Eno, est un objet complexe⁹. C'est pour le modéliser que le recours à un standard international a été étudié. En dehors de normes purement techniques, comme XForms¹⁰, seules deux options étaient véritablement à même de couvrir la complexité des questionnaires rencontrés dans la statistique publique : DDI (*Data Documentation Initiative*)¹¹ et SDMX (*Statistical Data and Metadata eXchange*)¹². Curieusement, ces deux standards répondaient au problème de façon complètement différente (**encadré 1**).

On avait donc schématiquement, avec DDI et SDMX, le choix entre d'un côté un standard complexe et sémantiquement précis et de l'autre un standard simple mais sémantiquement faible. Tous deux sont spécifiés par des schémas XML (*Extensible Markup Language*), et peuvent donc être utilisés directement avec les langages de programmation du monde XML, en particulier le langage de transformation de documents XSLT¹³. Tout modèle de

7. Il s'agit de l'enquête Activité et conditions d'emploi de la main-d'œuvre (Acemo) dont la maîtrise d'ouvrage est la direction de l'Animation de la recherche, des études et des statistiques (Dares).

8. Outre la Dares déjà citée, il s'agissait de trois directions de l'Insee : la direction des Statistiques d'entreprises, la direction des Études et synthèses économiques, et de la direction des Statistiques démographiques et sociales.

9. Voir notamment l'article sur l'outil de conception de questionnaires Pogues de Franck Cotton et Thomas Dubois, dans ce même numéro.

10. Recommandation du W3C, l'organisme de normalisation du *web*, pour la représentation et l'utilisation des formulaires *web*. XForms est en fait utilisée par Coltrane pour mettre à disposition les questionnaires des enquêtes par Internet.

11. Voir <https://www.ddialliance.org/>.

12. Voir <https://sdmx.org/>.

13. XSLT (<https://www.w3.org/TR/xslt-30/>) est un langage déclaratif (lui-même exprimé en XML) permettant de transformer des documents XML en d'autres documents XML. Eno est écrit en XSLT.

questionnaire exprimé en DDI ou SDMX est donc « activable » au sens où tout questionnaire conforme à ce modèle peut être passé sans modification en entrée d'une chaîne de traitement. On est donc bien dans les deux cas sur des exemples d'activation de métadonnées : le modèle de questionnaire, en l'occurrence.

Et de fait, en poussant la logique, on peut instruire très naturellement le choix entre DDI et SDMX. L'utilisation active des métadonnées est d'autant plus féconde que ces dernières sont sémantiquement riches, ce qui a conduit clairement à privilégier DDI.

Il est à noter qu'à l'époque, des travaux exploratoires similaires étaient conduits du côté de l'institut statistique australien (ABS) visant à développer un prototype de générateur

Encadré 1. DDI vs SDMX : deux standards, deux approches



DDI, issu essentiellement du monde de la recherche en sciences sociales et des bibliothèques, s'est développé depuis 1995 sous l'impulsion principale de l'Université du Michigan. Initialement axé sur la documentation des études scientifiques et de leurs résultats, DDI s'est étendu à partir de 2007 à la modélisation du cycle de

vie complet des données, de leur collecte à leur publication et leur archivage. Cette version, dite « *DDI Lifecycle* », couvre notamment d'une façon très précise la représentation des questionnaires statistiques, mais au prix d'une verbosité qui peut parfois en rendre l'utilisation difficile. En effet, ce standard, s'appuyant sur le profil d'archiviste ou de chercheur de ses concepteurs et qui se veut un standard de documentation, privilégie l'exhaustivité, la précision et les possibilités de réutilisation et de mutualisation des objets décrits, à la praticité d'utilisation et à la simplicité de représentation.



SDMX, de son côté, est une norme d'origine plus statistique et internationale, dont on peut faire remonter la naissance à une réunion organisée par le Fonds Monétaire International à Washington en septembre 2001. SDMX prenait la relève de travaux précédents sur le message EDIFACT-GESMES*, dont il constituait essentiellement une traduction en

XML (voir (Eurostat, Unece, 2002) pour plus d'information sur l'historique de SDMX). SDMX bénéficie de l'appui de sponsors influents (Fonds monétaire international, Banque centrale européenne, Banque mondiale, Banque des règlements internationaux, Eurostat, etc.) ; en particulier, Eurostat œuvre activement pour en imposer l'utilisation au sein du système statistique européen. SDMX est plus réputé pour l'échange et la diffusion de données tabulaires et des métadonnées structurelles associées. Il comporte également un modèle moins connu dédié aux métadonnées de référence. Ce modèle spécifie pour l'essentiel une structure d'annotations hiérarchiques et reste très générique. Il peut toutefois, au prix d'extensions dédiées (*plug-in* concernant les types, codes, etc.), représenter un questionnaire d'enquête, ou tout autre objet, si on l'adapte différemment. Comparativement à DDI, ce standard, dont les concepteurs avaient un profil plus opérationnel et technique et qui se veut un standard pour de l'échange d'informations, privilégie quant à lui le caractère pratique et générique d'utilisation à la précision et complexité des objets représentés (Arofan et Heus, 2007 ; Heus, Thomas et Vardigan, 2008).

* EDIFACT est un cadre développé par l'Unece et standardisé par l'ISO pour l'échange de données informatisé. GESMES (*Generic statistical message*) est un message EDIFACT pour l'échange de données et métadonnées statistiques.

de questionnaires *web* (XForms) à partir d'une description en DDI¹⁴. Les opportunités de collaboration avec les collègues australiens ont d'ailleurs encouragé l'équipe projet dans le choix du standard DDI et de la technologie XForms.

HARMONISATION INDUITE DES SUPPORTS DE COLLECTE

Le choix d'un standard pour la modélisation des questionnaires, en l'occurrence DDI, était un pré-requis à l'automatisation. La standardisation permet à terme des gains en coût de réalisation et en délai. Mais elle a également des impacts sur le fond et la forme des questionnaires conçus dans ce cadre.

L'utilisation d'un standard formel et sémantiquement précis comme DDI a provoqué des réflexions plus poussées sur la représentation de nos questionnaires : la formulation des questions, la distinction entre questions et instructions, etc.

La phase de conception était auparavant réalisée à l'aide d'outils bureautiques, et autorisait une construction linéaire du questionnement. Certaines parties comportaient parfois plusieurs questions, des éléments d'instruction, des champs de réponse, sans les distinguer formellement. La conceptualisation à l'œuvre dans le standard dévoile ces confusions.

Si le cadre formel n'est pas exempt de contraintes, l'exercice de conceptualisation qui en découle est souvent une opportunité de clarification. Pour exemple, la standardisation a conduit à identifier formellement chaque question, et à numéroter l'ensemble des questions de façon homogène, créant parfois une rupture avec la numérotation « historique » du questionnaire. Or celle-ci est souvent réutilisée dans les applications de gestion d'enquête, et sa modification a nécessité des évolutions applicatives, voire des changements dans les habitudes des gestionnaires d'enquêtes.

“*La conceptualisation plus poussée et la standardisation de la forme concourent toutes deux à une meilleure qualité des questionnaires.*”

La standardisation permet également lors de la phase de conception de normaliser les questionnaires générés : les choix de mise en forme (police, couleurs, espacements, disposition) sont génériques. Il en découle une présentation homogène des questionnaires des différentes enquêtes et une identité visuelle plus reconnaissable pour l'Insee et pour chaque SSM, que renforce par ailleurs la mise en place d'un portail de collecte unique (*figure 2*).

La conceptualisation plus poussée et la standardisation de la forme concourent toutes deux à une meilleure qualité des questionnaires, car elles permettent d'intégrer dans le processus de génération le respect de bonnes pratiques de conception de questionnaires¹⁵. La standardisation impose un changement de pratiques, et une rupture avec le passé, en particulier lorsque les choix de conception antérieurs visaient à optimiser la mise en page pour diminuer le coût de la collecte papier (par l'utilisation de tableaux notamment).

14. Voir le générateur Ramona dont les derniers développements datent de 2013 : <https://github.com/LegoStormtroopr/ramona-ddi-instrument-creator/>.

15. Voir à ce sujet (Haraldsen, Jones, Snijkers et Willimack, 2013), (Christian, Dillman et Morrison, 2010) et (Christian, Dillman et Smyth, 2014).

Elle peut aussi parfois susciter des difficultés d'acceptation, mais *in fine* elle décharge les concepteurs des choix de mise en forme, et leur libère ainsi du temps potentiel pour des tâches de conception à plus forte valeur ajoutée.

Figure 2. Une proximité forte dans l'identité visuelle des questionnaires

Deux enquêtes Insee au visuel identique

Un visuel pour la Dares qui ne diffère que par la couleur et le logo

Capacité à Innover et Stratégie (CIS) 2018

Notice en ligne
Exemple de questionnaire
Informations de diffusion
Contacter l'assistance
Déconnexion

Avancement 33 %

II - INNOVATION

Innovation de produits

Une innovation de produit est l'introduction sur le marché d'un bien ou d'un service nouveau ou significativement amélioré par rapport aux produits précédemment élaborés par l'entreprise.

Inclure :

- les changements significatifs de design
- les biens ou services numériques

Exclure :

- la vente en filat de nouveaux biens ou services
- les changements de nature esthétique

→ 1. Au cours des trois années 2016 à 2018, l'entreprise a-t-elle introduit un bien nouveau ou amélioré ?

→ 2. Ces biens ou services nouveaux ou améliorés sont-ils différents ou très similaires aux produits déjà offerts ?

Sectorielle Annuelle 2018

Notice en ligne
Exemple de questionnaire
Informations de diffusion
Contacter l'assistance
Déconnexion

Avancement 27 %

III - EMPLOI

Effectif non salarié

Les travailleurs non salariés (ou indépendants) se caractérisent par une affiliation à la sécurité sociale des indépendants (ex RSI) et non au régime général. Ils se distinguent des salariés par l'absence de lien de subordination et de contrat de travail avec l'entreprise pour laquelle ils exercent une mission. Ils incluent les exploitants individuels (artisans, commerçants, professions libérales, hors agriculteurs) ou les personnes liées familialement à ces derniers (conjoints collaborateurs, aides familiaux, les gérants de sociétés...)

Attention : les personnels prêtés par d'autres sociétés (notamment dans le cadre d'un groupe) ne doivent pas être comptabilisés comme non salariés.

→ 1. Le dirigeant de l'entreprise est-il non salarié ?

→ 2. Y a-t-il dans l'entreprise d'autres non salariés (associés, conjoints collaborateurs, aides familiaux par exemple) ?

Acemo sur le dialogue social en entreprise 2018

Notice en ligne
Exemple de questionnaire
Informations de diffusion
Contacter l'assistance
Déconnexion

Avancement 67 %

VI : LES ARRÊTS COLLECTIFS DE TRAVAIL EN 2018 DANS VOTRE ENTREPRISE OU L'UN DE SES ÉTABLISSEMENTS

Les informations ci-dessous concernent votre entreprise identifiée par son numéro SIREN (et non NUES ou le groupe).

Grève et débrayage

→ 1. Dans votre entreprise ou l'un de ses établissements, au moins une grève et/ou un débrayage se sont-ils produits au cours de l'année 2018 ?

→ 2. Pour quelles raisons ces grève(s) et/ou débrayage(s) se sont-ils produits ?

Plusieurs réponses possibles

- Des motifs spécifiques à l'entreprise ou l'un de ses établissements
- Des revendications sectorielles
- Des revendications interprofessionnelles, nationales

→ 3. Nombre total de journées individuelles non travaillées pour fait de grève et/ou de débrayage dans l'ensemble des établissements de votre entreprise en 2018

Journées individuelles non travaillées = nombre de jours de grève x nombre de salariés concernés.

- Lorsqu'il y a plus d'un arrêt collectif de travail dans l'année, on additionne les journées individuelles non travaillées de chaque arrêt. Toutes les journées (ou heures) non travaillées par les salariés participant ou débrayage ou à la grève doivent être incluses dans le calcul, y compris celles qui ont été rémunérées en totalité ou partiellement.

① UNE ARCHITECTURE CONÇUE POUR UN OUTIL MODULAIRE, ÉVOLUTIF ET EXTENSIBLE

Si l'objectif principal et initial d'Eno était la production automatique de questionnaires *web*, rappelons que la mise en œuvre de métadonnées actives en était la ligne directrice depuis le départ. En particulier, le développement s'est toujours orienté sur des objectifs de généralité, de ré-utilisabilité et d'extensibilité maximales. Ainsi :

- ① Eno permet de produire différents modèles de questionnaire (pour les collectes *web* ou papier) basés sur le standard de modélisation du cycle de vie des données DDI, donc indépendant de la sphère métier « ménage » ou « entreprise » (généricité).
- ① Eno permet d'ajouter des nouveaux formats de sortie et d'entrée, non identifiés au départ, comme un format Open Document de spécification des questionnaires, un format pour la collecte par enquêteur ou pour des besoins post-collecte de reprise des données (ré-utilisabilité et extensibilité).

Cette architecture (**encadré 2**) offre une gestion « douce » des opérations de migration technique. En effet, en cas d'obsolescence d'une technologie d'un format d'entrée ou de sortie, il s'agit de prendre en compte ce nouveau format et d'organiser ensuite la migra-

tion de l'ancien vers le nouveau format. Ainsi par exemple, la montée de version du standard DDI a été réalisée de façon transparente pour les interfaces de sortie. Il est également envisageable de remplacer le format XForms de la filière *web* par des technologies plus récentes comme JavaScript et ce, sans affecter la spécification DDI en entrée.

« Cette architecture offre une gestion « douce » des opérations de migration technique. »

① UN OUTIL UTILISÉ POUR LES ENQUÊTES AUPRÈS DES ENTREPRISES...

Si Eno a vu le jour à la direction des Statistiques d'entreprises (DSE), ce n'est pas un hasard. La statistique d'entreprises offrait en effet un terrain favorable à la naissance, puis à la maturation d'un générateur de supports de collecte. La longue expérience d'enquêtes auto-administrées, et le caractère central du répertoire d'entreprises dans le système d'information ont conduit, au fil des années, à élaborer des processus de production statistique plus structurés et des questionnaires moins complexes que dans la sphère sociale.

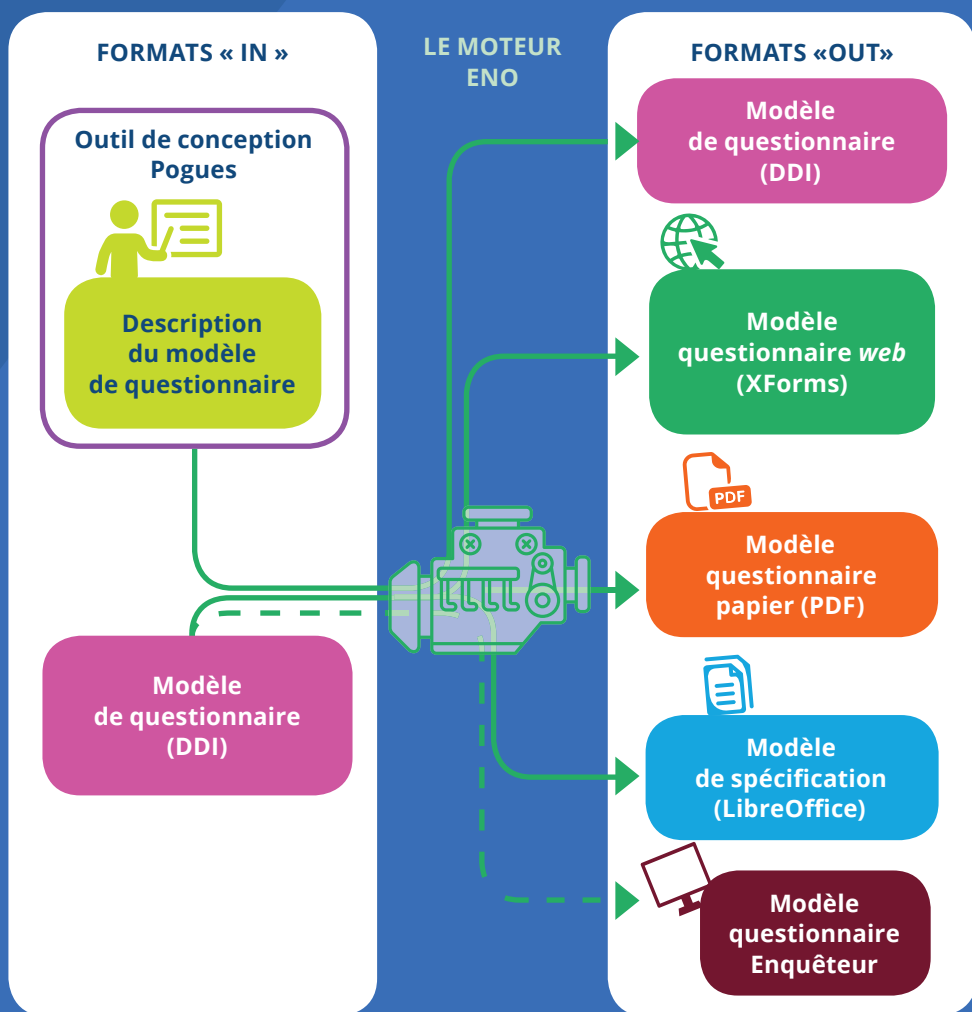
C'est dans ce contexte que le générateur s'est développé. Il est formellement aujourd'hui dans sa version 1. Celle-ci a déjà connu un certain nombre d'évolutions fonctionnelles qui lui permettent notamment de générer la grande majorité des questionnaires *web* des enquêtes entreprises de la filière Coltrane, ainsi que leur équivalent PDF.

Eno permet également de produire des spécifications dans un format bureautique de questionnaires décrits en DDI (**encadré 2**). Si ces spécifications ne concernent pas encore l'intégralité des éléments composant les questionnaires, la couverture actuelle permet déjà de disposer de premières spécifications pour les étapes de conception (relecture de la spécification) et de développement (pour des filières techniques tierces, telle que la filière Blaise).

Encadré 2. L'architecture du moteur Eno facilite la génération de supports de collecte sous plusieurs formats

Plus qu'un simple générateur de questionnaires *web*, Eno est un outil permettant de produire un ou plusieurs formats de sortie, génériquement désigné par « Out », à partir d'un format en entrée dit « In ». Cela est réalisé en s'appuyant sur une configuration déclarative, avec une technique dite de découplage fort (appelée « *separation of concerns* ») entre les formats d'entrée et de sortie, qui consiste à lier les éléments issus du format d'entrée avec ceux du format de sortie dans un fichier tableur, plutôt que sous la forme de code ou de fichiers de configuration technique réservés aux seuls spécialistes. C'est un choix de conception, intégré depuis le début dans le cœur même du moteur Eno, qui garantit une meilleure lisibilité et maintenabilité des fonctionnalités les plus complexes de l'outil.

En poussant cette logique jusqu'au bout, cela permet d'avoir des développeurs experts du format « In » et des développeurs experts du format « Out » (Une évolution du générateur est prévue, afin d'avoir un découplage total des formats d'entrée et de sortie).



📍 ... AU SEIN D'UN ÉCOSYSTÈME BASÉ SUR LES MÉTADONNÉES

La plateforme Coltrane offre les services nécessaires à la collecte *web* et papier, dont certains sont étroitement intégrés avec Eno (*figure 3*) : la plateforme de collecte permet au concepteur de visualiser les supports de collecte *web*, lesquels ont été produits par l'entremise d'Eno sous forme de descriptions techniques ; Coltrane permet ensuite de personnaliser les différents modèles de questionnaires pour les enquêtés.

Il en va de même pour les questionnaires papier : ceux-ci sont produits par un module de génération des courriers (opéré par un groupe dédié aux activités éditoriales). À l'instar de la plateforme de collecte *web*, ce module crée l'ensemble des supports de collecte papier à partir du modèle de questionnaires issu d'Eno et du fichier de personnalisation.

📍 DES ÉVOLUTIONS FONCTIONNELLES ENGAGÉES POUR LES MÉNAGES ET AU-DELÀ

Le développement du multimode dans les enquêtes auprès des ménages impose de mutualiser au maximum les développements nécessaires à la conception des supports de collecte, dans la mesure où le nombre de modes mobilisés est plus important que dans le cas des entreprises. Il implique des évolutions des protocoles de collecte et des questionnaires, mais également des outils dédiés à la collecte¹⁶.

L'utilisation d'Eno pour les questionnaires *web* des enquêtes ménages, va conduire à étendre ses fonctionnalités, notamment pour parvenir à générer les boucles de questions fréquemment utilisées dans ce type de questionnaires (par exemple l'articulation ménage/individus).

Ces évolutions sont toutefois mineures au regard de celles requises par l'intégration des enquêtes par enquêteur (face-à-face ou téléphone), qui exige un changement d'architecture et donc de format de sortie. (*figure 3*). En effet, il s'agit notamment de produire des questionnaires embarqués sur des postes enquêteurs, devant pouvoir fonctionner

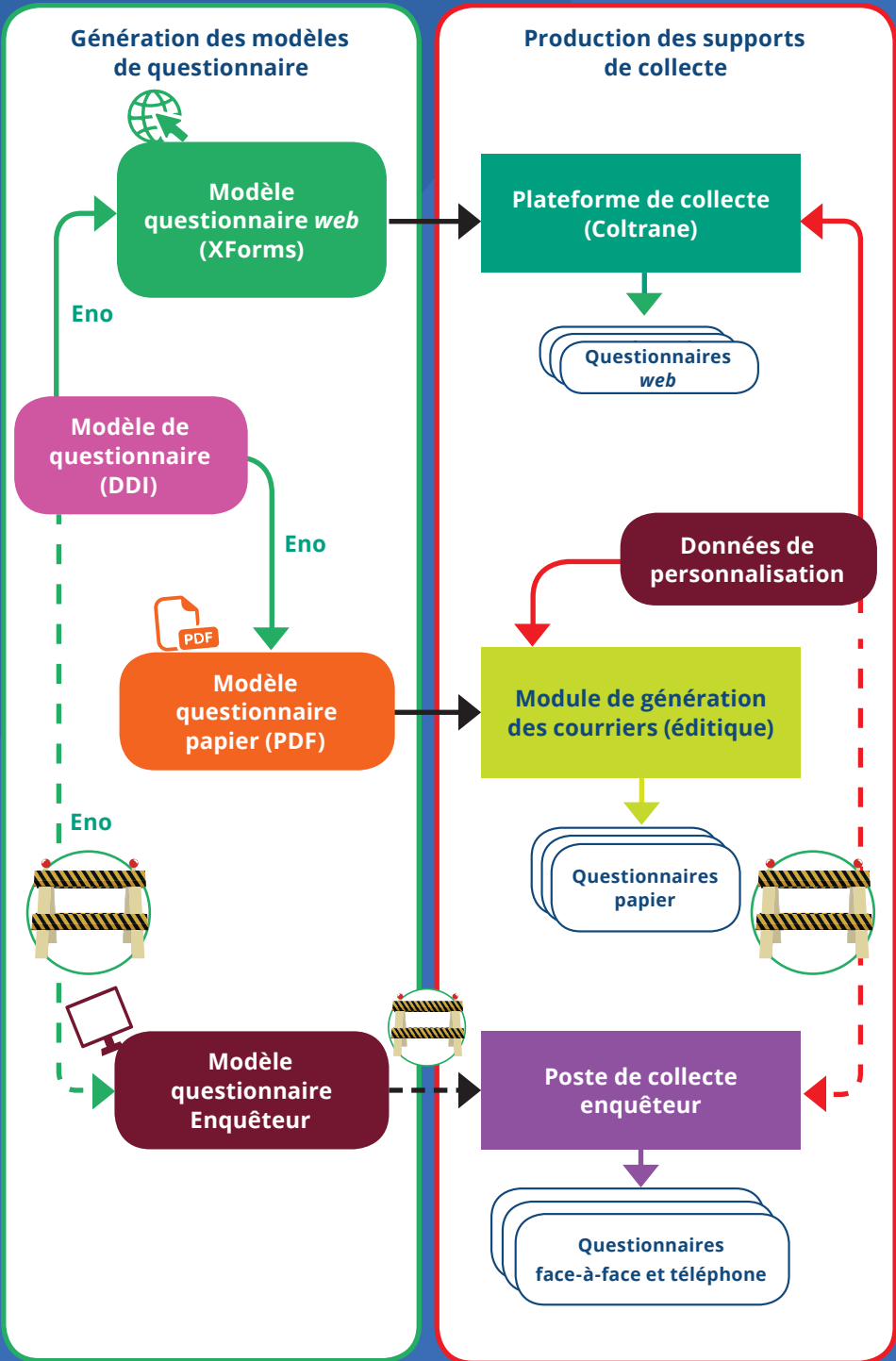
“ Le questionnaire est vu comme un ensemble de composants graphiques modulaires, réutilisables et mutualisables. ”

hors-ligne, quand jusqu'à présent Eno générerait des questionnaires pour une plateforme *web* classique. Ce changement de paradigme technique entraîne nécessairement un changement de technologies et constitue donc l'occasion de proposer une évolution ambitieuse : le questionnaire est vu comme un ensemble de composants graphiques modulaires, réutilisables et mutualisables. Ces composants hébergés dans une application dédiée

constitueront le questionnaire affiché sur le poste de l'enquêteur. Cette approche offre la possibilité de partager les composants avec d'autres outils ou même de les utiliser pour afficher autre chose qu'un questionnaire de collecte à proprement parler (par exemple les questionnaires dans une forme adaptée à l'activité de reprise par un gestionnaire).

16. L'Insee a lancé le programme Métallica (MÉTadonnées Actives, Logiciels Libres et Infrastructure pour une Collecte Assistée) afin d'optimiser le processus de conception des enquêtes ménages, et en particulier des questionnaires.

Figure 3. Eno dans la construction des supports de collecte



Les développements correspondants s'appuient sur une bibliothèque¹⁷ de composants *web* (par exemple un libellé de question, une liste déroulante, un champ de saisie pour les dates, etc.), conforme à l'état de l'art en matière de développement *web*¹⁸, et qui peut être utilisée dans d'autres applications. Une telle réutilisation est déjà engagée dans deux projets :

- ❶ la refonte de l'indice des prix, qui mettra en œuvre le partage des composants avec une autre application ;
- ❷ le développement d'un poste générique de traitement des enquêtes : les composants seront utilisés pour afficher des écrans de contrôle de questionnaires.

Ainsi, comme ce dernier cas l'illustre, la description des modèles de questionnaires en DDI, déjà utilisée pour produire les supports de collecte, servira également à générer les écrans relatifs au questionnaire dans l'application de reprise (*figure 4*).

Ces évolutions illustrent la faculté d'Eno à s'adapter à des architectures radicalement différentes, qu'elles soient centralisées (serveur de questionnaires) ou complètement distribuées (poste enquêteur embarqué), à des supports variés (tablettes des enquêteurs prix) et à des contextes applicatifs différents (collecte, traitements aval, etc.).

Enfin, des évolutions techniques sont par ailleurs prévues pour rendre le moteur Eno plus puissant et plus facilement paramétrable.

❶ UNE GOUVERNANCE TRANSVERSALE POUR FACILITER LA RÉUTILISATION

Une innovation n'émerge pas toujours dans le cadre le plus adapté à son bon fonctionnement en régime courant : le contexte « métier » initial est particulier, il a ses contraintes propres, ses objectifs et son savoir-faire, au sein duquel l'innovation s'apparente d'ailleurs au lancement de projet avec ressources dédiées sans l'être formellement.

Eno a été développé à l'Insee sous la gouvernance de la direction des Statistiques d'entreprises (DSE). Les conditions étaient alors réunies pour permettre le développement d'un générateur de questionnaires destiné en premier lieu à satisfaire un besoin pour des enquêtes auprès des entreprises.

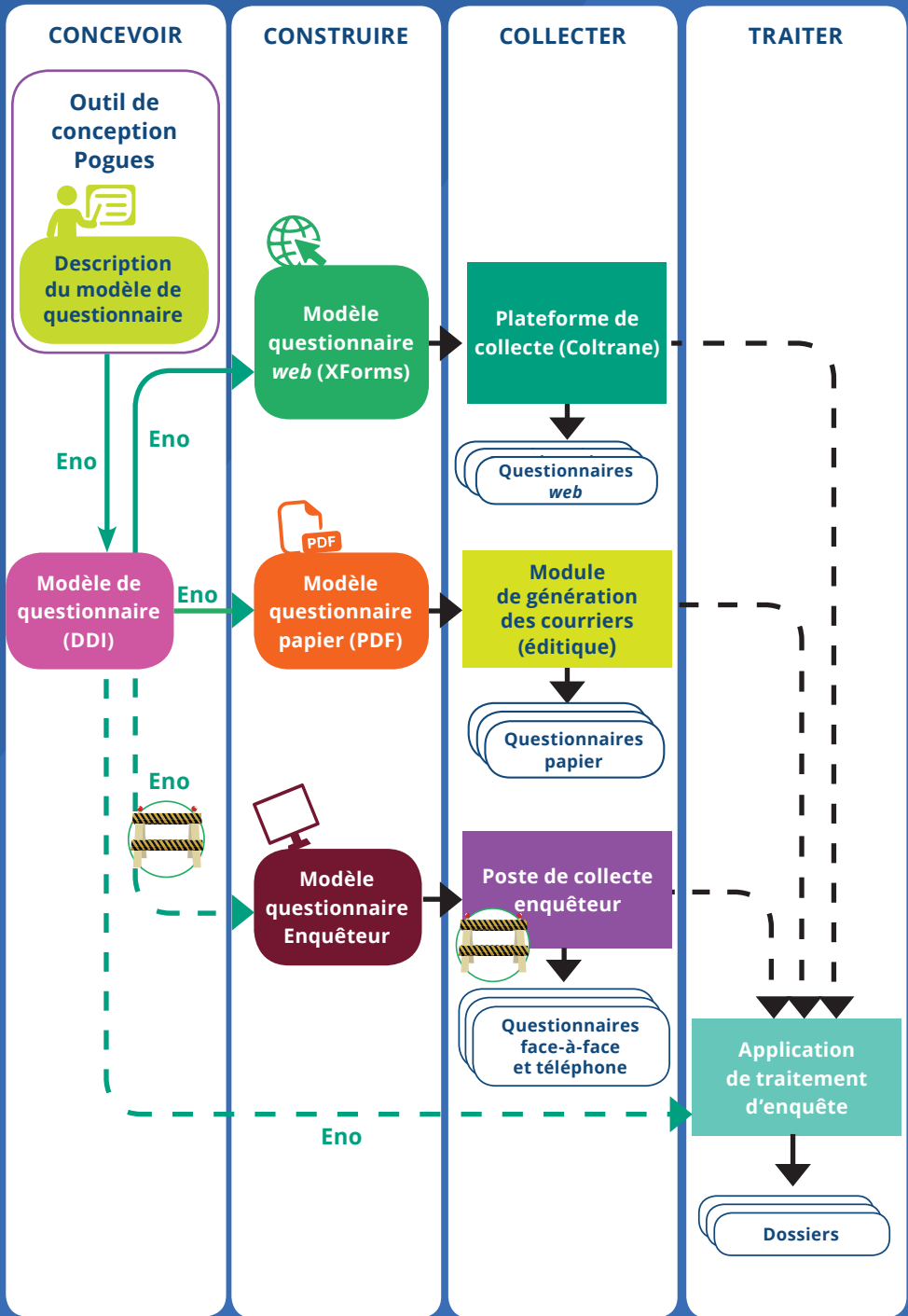
Mais dès le départ, Eno avait été conçu pour être générique, réutilisable et extensible et s'appuyant sur un standard. La question de la gouvernance d'un tel outil, à vocation transversale, s'est donc rapidement posée.

Les conditions initiales n'étaient plus adaptées pour développer un outil transversal dont la portée pouvait aller au-delà du périmètre de la direction métier « mère ». Assez naturellement, la gouvernance d'Eno est donc passée d'une direction métier à une direction plus orientée « conseil/expertise » et donc plus transversale : la direction de la Méthodologie et de la coordination statistique et internationale (DMCSI), déjà maître d'ouvrage du référentiel de métadonnées statistiques RMÉS.

17. La bibliothèque Lunatic est publiée en *open source*, voir <https://github.com/InseeFr/Lunatic>.

18. Par état de l'art, on parle ici des exigences liées à JavaScript, en termes d'architecture et de code.

Figure 4. Eno demain, de la Conception au Traitement ?



Note de lecture : aujourd'hui Eno contribue aux phases de Conception et de Construction (au sens du GSBPM) des supports. Son principe de base le rend éligible à contribuer au-delà, à la phase de Traitement.

Ce changement de gouvernance, assez inédit dans le contexte de l'institut, peut être vu comme une innovation en soi. C'est également une réussite, car il offre aujourd'hui de nouvelles perspectives d'évolutions à l'outil Eno. Ainsi, tout en continuant d'étendre l'offre

« Eno devient peu à peu le moteur de génération des modèles de questionnaires pour la collecte d'enquêtes auprès des ménages voire la collecte des prix. »

de services de la filière des enquêtes auprès des entreprises (par exemple l'ajout du format de sortie pour la collecte papier), Eno devient peu à peu le moteur de génération des modèles de questionnaires pour la collecte d'enquêtes auprès des ménages voire la collecte des prix. Il est aussi utilisé en aval de la collecte, pour des représentations de questionnaires de reprise de données post-collecte (voir *supra*). Mais, même confiée à une unité transversale, la gouvernance d'outils transverses pose de redoutables questions dont les réponses, encore empiriques aujourd'hui, sont à consolider :

comment arbitrer entre des demandes d'évolution émises par des utilisateurs différents ? Comment gérer les montées de version de ces composants transverses, lorsqu'elles ont un impact sur les applications clientes ? Comment financer la maintenance ? On voit que l'innovation technique amène à questionner nos organisations.

🌐 OUVERTURE À L'INTERNATIONAL

Eno a été pensé pour acquérir une dimension internationale, dès l'origine, comme en témoignent les échanges avec les experts de l'office statistique australien ABS, qui ont fourni les premiers éléments de sa conception, puis au fil de ses développements successifs.

Le choix du standard DDI a donné l'opportunité d'effectuer de multiples présentations du générateur dans les colloques consacrés à DDI.

Eno fait partie des outils choisis pour être mutualisés dans le cadre d'un projet européen (ESSnet *Sharing Common Functionalities*). Il a ainsi été réutilisé par SURS, l'INS slovène : cela reste un cas d'école qui n'a pas été jusqu'à une mise en production, mais des contacts existent avec d'autres utilisateurs potentiels.

En parallèle, des travaux d'internationalisation du code se sont poursuivis (traduction en anglais des éléments de commentaires et de documentation de ce code). Il est également publié et partagé en *open source* sur le dépôt GitHub¹⁹ de l'Insee. La communauté des utilisateurs d'Eno devra encore se développer pour que l'institut tire véritablement profit de cette ouverture. Un certain nombre de travaux reste d'ailleurs à conduire pour faciliter la réutilisation d'Eno, indépendamment d'un environnement technique : amélioration du *packaging*, développement de services simplifiant son utilisation, etc.

19. GitHub est une plateforme qui héberge le code d'un logiciel, d'une application, etc. Ce code peut être partagé, s'il est *open source*, et enrichi par des contributions volontaires, les auteurs du code source conservant les droits de validation. Voir <https://github.com/InseeFr> et l'article dans ce même numéro de Franck Cotton et Thomas Dubois sur Pogues.

📍 EVERYTHING THAT HAPPENS WILL HAPPEN TODAY²⁰ (AND FURTHER...)

Eno est désormais la solution de référence à l'Insee pour la création de tous les supports de collecte. Fin 2019, Eno produit presque tous les questionnaires auprès des entreprises, ainsi qu'un premier questionnaire ménages. La mise en œuvre du générateur dans la plate-forme Coltrane a contribué à une industrialisation poussée de la collecte des enquêtes d'entreprises. Cette industrialisation s'est accompagnée d'une large unification des pratiques et d'une certaine « normalisation » des outils. On citera en particulier l'ergonomie des supports de collecte (et plus généralement des documents échangés avec les enquêtés, par exemple les mails ou les courriers) : celle-ci a connu une harmonisation qui a pu conduire parfois à brider la créativité des concepteurs d'enquête²¹, mais qui a largement contribué à professionnaliser la démarche de l'Insee auprès des entreprises.

Eno s'engage maintenant, avec son utilisation dans les enquêtes en multimode auprès des ménages, dans une nouvelle période qui permettra d'enrichir encore ses fonctionnalités. S'il reste une insatisfaction à ce stade, elle se situe au niveau de la réutilisation d'Eno en dehors de l'Insee. Malgré de nombreuses manifestations d'intérêt (différents services statistiques ministériels, la Banque de France, des instituts statistiques étrangers), et malgré les efforts faits pour documenter l'outil, le conditionner sous des formes simplement réutilisables²² et le présenter en de nombreuses occasions, aucun cas de réutilisation n'a été poussé jusqu'en production. On touche là les limites d'une communication peu développée et encore très centrée sur la ré-utilisation interne de nos innovations et nos applications : il n'y a pas à l'Insee d'approche professionnelle en matière de « marketing » des développements informatiques, contrairement à ce que l'on peut observer dans d'autres organisations statistiques.

20. « Tout ce qui peut arriver arrivera aujourd'hui », dernier clin d'œil des auteurs à Brian Eno.

21. Voir sur ce point l'article d'Olivier Haag et Anne Husseini-Skalitz dans ce même numéro.

22. Eno fut le premier développement de l'Insee (et reste à ce jour le seul) à être publié sur le répertoire global de librairies Java (voir <https://mvnrepository.com/artifact/fr.insee.eno>).

▣ BIBLIOGRAPHIE

AROFAN, Gregory et HEUS, Pascal, 2007. DDI and SDMX : Complementary, Not Competing, Standards. In : *site de l'Open Data Foundation*. [en ligne]. Juillet 2007. [Consulté le 16 octobre 2019]. Disponible à l'adresse : http://www.opendatafoundation.org/papers/DDI_and_SDMX.pdf

BONNANS, Dominique, 2019. RMÉS, le référentiel de métadonnées statistiques de l'Insee. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp.46-55. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168396/courstat-2-6.pdf>

CHRISTIAN, Leah M., DILLMAN, Don A. et MORRISON, Rebecca, 2010. Questionnaire Design Guidelines for Establishment Surveys. In : *Journal of Official Statistics*. [en ligne]. Vol. 26, N°1, pp. 43-85. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/questionnaire-design-guidelines-for-establishment-surveys.pdf>

CHRISTIAN, Leah M., DILLMAN, Don A. et SMYTH, Jolene D., 2014. *Internet, Phone, Mail, and Mixed-Mode surveys – The tailored design method – 4th edition*. Août 2014. Wiley Online Edition. ISBN 978-1-11-845614-9

COTTON, Franck, SIGAUD, Éric, TAILHURAT, Romain et VAN DER VLIST, Éric, 2013. *XForms generation: a real world example*. [en ligne]. 5 août 2013. International Symposium on Native XML User Interfaces. Balisage Series on Markup Technologies, Vol. 11. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.balisage.net/Proceedings/vol11/html/Cotton01/BalisageVol11-Cotton01.html>

EUROSTAT, UNECE, 2002. *Common Open standards for the Exchange and Sharing of Socio-economic Data and Metadata : the SDMX Initiative*. [en ligne]. 6-8 mars 2002. Work Session on Statistical Metadata, Working Paper N°11. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://sdmx.org/wp-content/uploads/wp111.pdf>

HARALDSEN, Gustav, JONES, Jacqui, SNIJKERS, Ger et WILLIMACK, Diane K., 2013. *Designing and conducting business surveys*. 26 juillet 2013. Wiley Online Edition. ISBN 978-0-47-090304-9


HEUS Pascal, THOMAS, Wendy et VARDIGAN, Mary, 2008. Data Documentation Initiative : Toward a Standard for the Social Sciences. In : *The International Journal of Digital Curation*. [en ligne]. N°1, Vol. 3, pp. 108-113. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <http://www.ijdc.net/article/view/66/45>

COLLECTE PAR INTERNET DES ENQUÊTES AUPRÈS DES ENTREPRISES

L'INSEE ENTRE DANS L'ÈRE COLTRANE

Olivier Haag*, Anne Husseini-Skalitz**

La collecte des enquêtes auprès des entreprises accorde une place croissante à internet. Or dans le système statistique public français, l'offre était protéiforme. Le projet Coltrane a été développé de façon modulaire dans le but d'apporter de la cohérence technique et fonctionnelle à ce qui était épars. La plateforme de collecte résultant du projet a ainsi non seulement pris la place de quatre dispositifs de collecte antérieurs, mais elle a aussi été le premier maillon d'une nouvelle filière de l'Insee basée sur des métadonnées actives. Ainsi, à partir de la description du questionnaire par le concepteur, différents supports de collecte sont générés automatiquement et intégrés dans une infrastructure de collecte par internet. Plus qu'une simple plateforme de collecte, Coltrane, aujourd'hui, c'est une offre de service complète qui permet aux concepteurs d'enquête : de collecter les réponses par internet, de disposer d'un référentiel de contacts au sein des entreprises et de le gérer, d'envoyer des courriers, des mails et des questionnaires papier aux entreprises qui le souhaitent, le tout en disposant d'un système d'assistance dédié. Coltrane constitue aussi une avancée notable pour les entreprises qui peuvent maintenant mieux identifier et répondre à la plupart des enquêtes qui leur sont adressées par la statistique publique grâce à un unique couple « identifiant – mot de passe », selon un cadre ergonomique commun.

 *The collection of business surveys is increasingly focusing on the Internet. However, in the French Official Statistical System, the offer was protean. The Coltrane project was developed in a modular way in order to bring technical and functional coherence to what was scattered. The data collection platform resulting from the project thus not only replaced four previous collection systems, but was also the first link in a new INSEE network based on active metadata. Thus, based on the designer's description of the questionnaire, different collection media are automatically generated and integrated into an Internet collection infrastructure. Coltrane is today more than just a simple collection platform: it is a complete service offer that allows survey designers to collect responses via the Internet, to access and manage a repository of contacts within businesses, to send letters, emails and paper questionnaires to companies that are interested, while having a dedicated*

* Coordinateur des projets structurants de la direction des Statistiques d'entreprises, Insee, olivier.haag@insee.fr

** Cheffe de la section Coltrane et animation du réseau d'enquêteurs entreprises, Insee, anne.husseini-skalitz@insee.fr

support system. Coltrane also represents a significant step forward for enterprises, which can now better identify and respond to most of the surveys sent to them by official statistics thanks to a single "login/password" couple, according to a common ergonomic framework.

① SIMPLIFIER LES ENQUÊTES POUR LES ENTREPRISES ET HARMONISER LES OUTILS DES CONCEPTEURS

En 2000, le Sessi¹, alors service statistique ministériel de l'Industrie, initiait la collecte par internet auprès des entreprises avec son enquête mensuelle de branche, avant de l'étendre aux autres enquêtes du secteur de l'industrie. Avec la fusion d'une partie du Sessi et de l'Insee en 2008, l'infrastructure de collecte est devenue propriété de l'Insee. Depuis 2004, l'Insee disposait déjà de son propre dispositif² pour une douzaine d'enquêtes fréquentes et régulières³ auprès des entreprises. Peu à peu, la diffusion rapide d'internet au sein des entreprises et la volonté de réduire les coûts de collecte ont encouragé d'autres services producteurs, de l'Insee ou des services statistiques ministériels, à dématérialiser leurs questionnaires. En cherchant à répondre au plus vite à l'aspiration des producteurs et à la demande pressante des entreprises, on aboutissait à une juxtaposition d'outils hétérogènes. Cette situation devait conduire l'Insee à chercher à réduire le nombre d'infrastructures de collecte *web*⁴ afin de faciliter la réponse des entreprises, tout en réduisant les coûts de maintenance informatique des différentes plateformes d'autre part.

En parallèle, lors des assises nationales de la simplification administrative d'avril 2011, les entreprises avaient fait remonter leurs difficultés face à la charge statistique⁵. Parmi les 80 mesures issues de ces assises, figurait l'objectif de « Dématérialiser à 100 % les enquêtes de la statistique publique »⁶ et de centraliser les questionnaires sur un site unique à partir de 2013. Tout le monde s'accordait à penser – et les interviews des entreprises le confirmaient – que ce mode de réponse était moins coûteux pour les entreprises que des questionnaires papier. À cette époque, la couverture de la collecte par internet restait insuffisante : les enquêtes pour lesquelles c'était techniquement possible représentaient 73 % du volume des questionnaires envoyés par l'Insee aux entreprises, mais seulement 51 % des enquêtes de l'ensemble du Service statistique public (SSP). L'ambition était donc moins d'alléger la charge statistique au sens classique du terme (« diminuer le volume d'enquêtes »), que d'alléger la charge de réponse en simplifiant les dispositifs de collecte pour les entreprises.

1. Sessi : Service des statistiques industrielles. La plateforme de collecte s'appelait sessi.fr.

2. CRPI : Collecte et Retour Par Internet.

3. Essentiellement des enquêtes auprès des entreprises de la direction des Études et synthèses économiques (enquêtes de Conjoncture) et de la direction des Statistiques d'entreprises (enquêtes mensuelles de fréquentation touristique).

4. Voir l'article de Jean-Marc Béguin sur la naissance d'une innovation en production statistique, dans ce même numéro.

5. La charge statistique, en d'autres termes la charge de réponse aux enquêtes, figurait au deuxième rang des difficultés jugées les plus coûteuses pour les entreprises petites et moyennes (une autre étant l'instabilité et la complexité de la réglementation).

6. Voir <https://archives.entreprises.gouv.fr/2012/www.pme.gouv.fr/simplification/80-mesures.html> et notamment la mesure de la fiche n°12 « Dématérialiser les enquêtes de la statistique publique ».

C'est dans ce contexte que le projet Coltrane (COLlecte TRANsversale d'Enquêtes) a démarré en 2010.

Il avait pour objectifs principaux de :

- ❶ mettre au point « une plateforme technique » de collecte multimodale ;
- ❷ mettre à disposition des enquêtés et des responsables d'enquête de l'Insee, voire du Système statistique public, un ensemble de fonctionnalités répondant à leurs besoins (authentification des entreprises sur la plateforme, poste de pilotage pour les responsables d'enquêtes, etc.) ;
- ❸ fournir un mode d'emploi d'intégration des enquêtes sur cette plateforme ;
- ❹ offrir une infrastructure de gestion des contacts.

Le projet avait un périmètre déjà large au regard de son calendrier. Il n'entrait pas dans ses objectifs de départ de pouvoir satisfaire les besoins des enquêtes auprès des ménages (ou du recensement de la population). Toutefois, l'idée subsistait que toutes les fonctions potentiellement communes entre les deux domaines de collecte et développées pendant le projet devraient à terme être partagées.

❶ UNE NOUVELLE FILIÈRE À LA LOGIQUE BASÉE SUR LES MÉTADONNÉES ACTIVES

Une particularité distingue Coltrane des plateformes existantes⁷ : ce sont ses liens étroits, dès l'origine, avec le projet de référentiel des métadonnées statistiques RMÉS (Bonnans, 2019) et notamment l'objectif affiché d'utiliser les métadonnées qui décrivent les questionnaires d'enquêtes⁸ pour générer directement les outils de collecte, d'où l'expression de métadonnées « actives » pour décrire la filière.

L'Insee a adopté le GSBPM⁹ pour modéliser ses processus statistiques et le standard DDI¹⁰ pour décrire formellement le cycle de vie des données, en particulier les objets « questionnaires ». Ce choix structurant a rendu impossible le recours à un des progiciels de collecte sur le marché, car ceux-ci n'étaient pas encore adaptés à ces standards.

« L'Insee a adopté le GSBPM pour modéliser ses processus statistiques et le standard DDI pour décrire formellement le cycle de vie des données. »

La génération automatique des supports de collecte, c'est-à-dire des écrans et des programmes qui permettent de récupérer les réponses, à partir de leur description en DDI, est à la racine des gains de productivité escomptés. Elle a

d'abord été développée dans le cadre d'un sous-projet de Coltrane baptisé Eno¹¹. Le générateur Eno permet, à partir de la spécification du questionnaire, d'automatiser des tâches auparavant réalisées manuellement : développement du support de collecte par un informaticien puis test et recette par le statisticien ; il permet de garantir que plusieurs modes

7. Voir (Saraiva et alii, 2016 ; Fazio et alii, 2015 ; Thieme et alii, 2015).

8. Liste et enchaînement des questions, modalités possibles des réponses, contrôles à la saisie, filtres, etc.

9. *Generic Statistical Business Process Model* ou modèle générique de description des processus de production statistique.

10. *Data Documentation Initiative*, voir <https://ddialliance.org/>.

11. Voir l'article sur le générateur Eno de Heidi Koumarianos et Éric Sigaud dans ce même numéro.

de collecte différents (incluant le papier) correspondent à la même description générale. Ainsi, contrairement au passé, le concepteur ne fait plus autant de spécifications qu'il y a de supports de collecte. Une seule spécification suffit et elle lui garantit l'homogénéité du questionnement quel qu'en soit le mode.

Entamée avec Coltrane et son sous-projet Eno, l'activation des métadonnées pour la collecte s'est enrichie de l'outil Pogues¹² qui permet au statisticien de concevoir son modèle de questionnaire, et de générer automatiquement le modèle en DDI, sans recourir à un expert de ce standard.

UN DÉVELOPPEMENT PROGRESSIF, APPUYÉ SUR DES OPÉRATIONS EN VRAIE GRANDEUR

Aujourd'hui, l'ensemble de la filière Pogues /Eno / Coltrane vise à mettre l'utilisateur-concepteur au centre du jeu. Pour Coltrane, il avait été décidé dès le départ d'opter pour un développement modulaire, avec la mise en production successive « d'opérations ».

« *L'atterrissage en production d'une opération devait se matérialiser par la bascule sur internet d'une enquête pilote et d'un ensemble de fonctionnalités (parfois versionnées) du projet Coltrane, qui étaient ensuite utilisables pour les enquêtes pilotes suivantes.* »

Une opération désignait un ensemble fonctionnel complet et cohérent, pour lequel une application était livrée en production et faisait l'objet d'une recette complète par un groupe d'utilisateurs. L'atterrissage en production d'une opération devait se matérialiser par la bascule sur internet d'une enquête pilote et d'un ensemble de fonctionnalités (parfois versionnées) du projet Coltrane, qui étaient ensuite utilisables pour les enquêtes pilotes suivantes.

Trois opérations ont été lancées au cours de la première phase du projet, phase pendant laquelle on prévoyait de développer toutes les fonctionnalités prévues à l'origine.

- ❶ En février 2013, la première étape a été de réaliser la collecte en ligne pour l'Enquête Sectorielle Annuelle (ESA), *via* un système de dépôt-retrait d'un questionnaire au format « PDF remplissable ». Cette opération, non prévue à l'origine, a été développée en priorité pour répondre aux attentes des assises de la simplification (voir *supra*). Il a permis d'envoyer un signal fort en dématérialisant l'enquête ayant l'échantillon le plus volumineux ; c'était aussi l'enquête qui, du fait de sa taille, touchait le plus de PME, lesquelles étaient les plus sensibles à la charge de collecte.
- ❷ La deuxième étape, en juin 2013, a été la mise en ligne de l'enquête Innovation (CIS). Il s'agissait, de fait, de la première utilisation de l'infrastructure technique Coltrane sous le format visé (questionnaire « en ligne » alimenté par et alimentant un fichier au format XML).
- ❸ On restait cependant dans une logique purement Insee, et dans un troisième temps (mars 2015), l'objectif a été d'outiller une enquête d'un service statistique hors Insee, en l'occurrence la Dares¹³ et l'enquête trimestrielle Acemo (jusqu'alors tout papier), ainsi que l'ajout de fonctionnalités nouvelles par rapport à l'opération précédente.

12. Voir l'article sur l'outil Pogues de Franck Cotton et Thomas Dubois dans ce même numéro.

13. Direction de l'animation de la recherche, des études et des statistiques du ministère du Travail.

La seconde phase appelée aussi phase de généralisation ou « d'intégration de masse » a débuté en septembre 2016. Elle consistait, une fois la plupart des composants de l'offre de service Coltrane développés, à basculer progressivement la trentaine d'enquêtes de l'Insee dans le dispositif, notamment celles qui utilisaient l'ancienne infrastructure du Sessi (comme les enquêtes mensuelles de branche (EMB)).

Le démarrage de cette seconde phase a été plus difficile que prévu initialement. L'intégration des premières enquêtes a en effet mis en évidence des fragilités dans la plateforme de collecte. Afin de les résoudre, il a été nécessaire d'optimiser certains développements et de renforcer l'infrastructure de collecte en production afin d'être sûr de pouvoir supporter la charge prévue en régime de croisière. L'intégration massive des enquêtes dans le dispositif a donc été réalisée dans le calendrier prévu, mais la plateforme n'offrait pas à l'époque toutes les fonctionnalités imaginées au cours du projet.

📍 COLTRANE AUJOURD'HUI ? À LA CHARNIÈRE ENTRE LES ENTREPRISES ET LES STATISTICIENS

La plateforme offre des services aux concepteurs d'enquêtes lors de la phase de Construction¹⁴ des supports de collecte. Coltrane transmet les réponses collectées *via* la plateforme, dans un format standard, aux applications de gestion des enquêtes (*figure 1*). Afin de ne pas « décourager » les répondants, peu de contrôles de cohérence ou de non-réponse partielle sont implémentés dans Coltrane. Cette absence de contrôle bloquant permet en outre de faire l'hypothèse que Coltrane n'induit à ce jour aucun effet de mode dû à la collecte par internet par rapport à une collecte « papier ».

Les applications de gestion des enquêtes, qui sont en dehors du périmètre fonctionnel de Coltrane, permettent de suivre la collecte, notamment d'identifier les unités non-répondantes à relancer, de contrôler la qualité des réponses et de les redresser le cas échéant. Coltrane offre donc aussi, par l'intermédiaire des applications de gestion, des services aux équipes de gestionnaires d'enquêtes, lors des phases de Collecte. Nous allons voir plus loin, comment Coltrane outille directement les gestionnaires en contact avec les entreprises.

Mais les premiers utilisateurs de Coltrane sont, de fait, les entreprises répondant aux enquêtes que la plateforme abrite.

📍 L'OFFRE DE SERVICE AUX ENTREPRISES ENQUÊTÉES

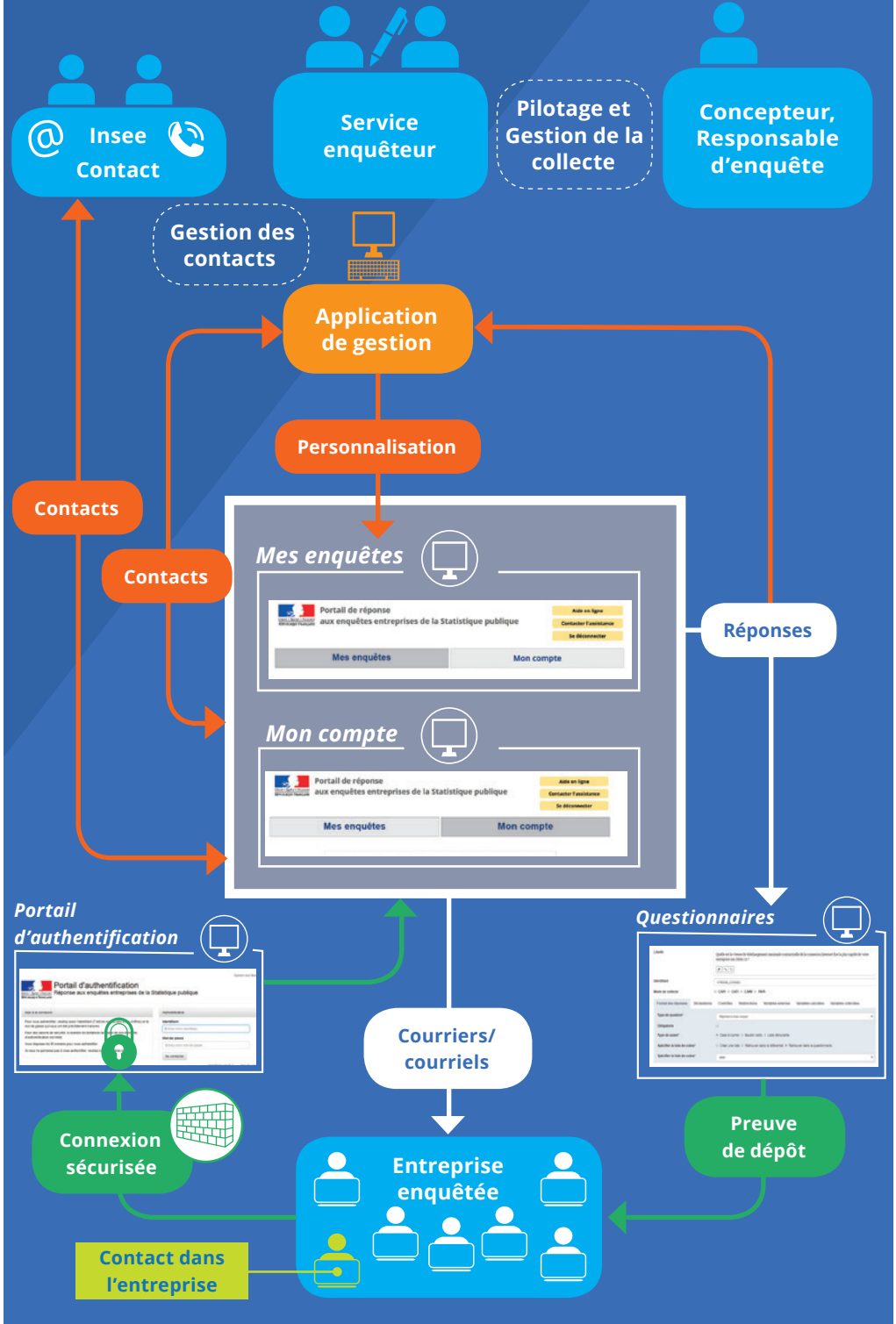
Quand elle est échantillonnée, l'entreprise enquêtée reçoit un courrier (voire un courriel) lui annonçant qu'elle a été sélectionnée pour répondre à un questionnaire. Ce courrier, en plus de présenter succinctement l'enquête, contient des éléments pour se connecter à la plateforme : lien à ouvrir¹⁵, identifiant et mot de passe, attribués à la personne physique qui va répondre à l'enquête et qu'on appelle « contact », de sorte que :

- 📍 une unité enquêtée a au moins un contact (dont les noms et prénoms peuvent ne pas être connus au lancement de l'enquête) et peut en avoir plusieurs ;
- 📍 un même contact peut être associé à plusieurs unités enquêtées ;
- 📍 le contact peut ne pas appartenir à l'entreprise enquêtée mais être mandaté pour le faire, par exemple un comptable qui répond aux enquêtes pour ses clients.

14. Au sens du GSBPM.

15. Voir la page du site de l'Insee consacrée aux enquêtes auprès des entreprises <https://www.insee.fr/fr/information/1401986> et celle spécifiquement dédiée à l'accueil sur la plateforme Coltrane : <https://entreprises.stat-publique.fr/>.

Figure 1. Coltrane, à la charnière entre entreprises et statisticiens



À la première connexion sur le **portail d'authentification (figure 2)**, le contact doit personnaliser son mot de passe. En cas de perte du mot de passe, il peut demander son renouvellement en ligne. Le portail d'authentification fait appel à un service mutualisé d'identification, développé en dehors du projet.

Une fois authentifié, le contact est dirigé sur un portail de réponse (**figure 2**), et plus spécifiquement sur un onglet nommé « Mes Enquêtes », point d'entrée unique pour répondre à toutes les enquêtes gérées par Coltrane (enquête de l'Insee ou de tout service statistique ministériel qui a adopté Coltrane) :

- ❶ il y visualise les enquêtes en cours de collecte pour lesquelles il a les droits de réponse, ainsi que les dates de retour souhaité, voire les dates de réponse si le questionnaire a déjà été renseigné ;
- ❶ il accède au questionnaire de son choix parmi la liste des enquêtes, soit pour répondre à l'enquête, soit pour consulter les questionnaires ou réponses déjà saisis tant que la collecte n'est pas fermée.

« *Tout au long de son chemin sur la plateforme de collecte, le contact dans l'entreprise peut accéder à une foire aux questions et à une assistance en ligne, via un formulaire.* »

Lorsqu'un contact navigue dans un **questionnaire**, ses réponses sont automatiquement enregistrées à chaque changement de page. Il peut ainsi quitter le questionnaire à tout moment et y revenir quand il le souhaite : il lui est alors proposé de retourner directement sur la dernière page consultée ou de recommencer depuis le début. En bout de course, il peut choisir d'envoyer sa réponse ou de simplement l'enregistrer pour y revenir ultérieurement.

Lors de l'envoi, un téléchargement au format « PDF » récapitulant sa réponse lui est proposé¹⁶.

Depuis le portail de réponse, le contact a bien sûr la possibilité de renseigner ou de mettre à jour ses informations personnelles (nom, prénom, adresse mail, numéro de téléphone, fonction, adresse postale, etc.) par le biais de l'onglet « **Mon Compte** ». Il peut également personnaliser son mot de passe et surtout, dans le cas où il aurait reçu plusieurs identifiants de connexion (un par enquête par exemple), il peut les « regrouper » et disposer d'un unique compte pour accéder et répondre aux questionnaires de plusieurs unités enquêtées ou de plusieurs enquêtes. Sur un plan pratique, c'est un grand atout du portail.

Tout au long de son chemin sur la plateforme de collecte, le contact dans l'entreprise peut accéder à une foire aux questions et à une assistance en ligne, *via* un formulaire. Cette assistance est contextualisée, c'est-à-dire que selon l'endroit d'où le formulaire est appelé, il est plus ou moins prérempli et peut être transmis à deux types d'acteurs :

- ❶ les pôles Insee-Contact¹⁷, pour les questions « techniques » (modalité d'accès à la plateforme de collecte, réalité de l'enquête, objectifs, perte de code d'accès, etc.) ;
- ❶ les équipes des gestionnaires d'enquête, pour les questions « métier » directement en lien avec l'enquête.

16. Pour certaines enquêtes, comme l'EAP ou l'EMB, le contact a aussi la possibilité de corriger et renvoyer, pendant la période de collecte, un questionnaire déjà transmis, sur demande du gestionnaire de la collecte.

17. Service de l'Insee qui répond aux demandes d'information, qu'il s'agisse des produits de diffusion ou des opérations statistiques en cours, réparti sur deux pôles situés à Rouen et Strasbourg. Les demandes arrivent par mail, courrier ou téléphone.

Cet ensemble de services offert par l'intermédiaire de Coltrane permet *in fine* d'améliorer l'image de la statistique publique, notamment vis-à-vis des grandes entreprises qui reçoivent un nombre important de questionnaires d'enquêtes statistiques¹⁸. Ces dernières voient d'une part se réduire le coût d'appropriation du remplissage des questionnaires et de la navigation sur la plateforme de collecte. D'autre part, elles peuvent être incitées à répondre aux enquêtes qui figurent dans leurs en-cours et dont elles avaient oublié l'existence.

Si, depuis la mise en œuvre de Coltrane, on n'a pas encore observé d'accroissement significatif des taux de réponse, au moins le système permet-il de maintenir les taux existants à des niveaux plutôt élevés.

En plus des portails destinés aux entreprises enquêtées, le projet Coltrane a mis en place plusieurs outils à destination des concepteurs d'enquête pour faciliter la création des supports de collecte d'une part et pour gérer la collecte proprement dite d'autre part.

Figure2. Ce que l'entreprise enquêtée visualise avec la plateforme Coltrane

Une assistance et une aide en ligne

Un point d'entrée unique et sécurisé pour la personne qui répond au nom de l'entreprise

Toutes les enquêtes à traiter pour ce contact sont accessibles

Portail d'authentification
Réponse aux enquêtes entreprises de la Statistique publique

Portail de réponse
aux enquêtes entreprises de la Statistique publique

Référence enquête	Nom de l'enquête	Statut	Accès
999 999 081	Technologies de l'information et de la communication et commerce électronique (TIC) 2019	Réponse attendue pour le 29 janvier 2019	→
999 999 081 00001	Technologies de l'information et de la communication et commerce électronique (TIC) 2019	Répondu le 28 octobre 2019	→
999 999 081	Système d'information sur les nouvelles entreprises 2020	Réponse attendue pour le 07 novembre 2020	→
AN1 234 001	Situation des associations en 2018	Réponse attendue pour le 07 novembre 2019	→
AN1 234 002	Situation des associations en 2018	Réponse attendue pour le 07 novembre 2019	→

Début Précédent 1 2 3 4 5 8 Suivant Fin

INSEE Institut national de la statistique et des études économiques
INSEE Nouvelle par entreprise
DARES
Version: 3-1-5

18. L'identité visuelle a également bénéficié d'un effort d'harmonisation de la présentation des questionnaires. Voir aussi à ce sujet l'article de Heidi Koumarianos et Éric Sigaud sur Eno dans ce même numéro.

DES OUTILS INNOVANTS POUR LES CONCEPTEURS D'ENQUÊTES

C'est dans le cadre du projet Coltrane que le générateur de supports de collecte Eno a été développé. L'équipe de projet a en outre fortement contribué à la mise en place de Pogues¹⁹, outil de conception du modèle de questionnaire en DDI. Grâce à ces outils, le statisticien conçoit le modèle de questionnaire, le visualise dans une version « vierge » et non personnalisée, et valide sa version définitive.

Il doit ensuite utiliser l'application de gestion de son enquête²⁰ pour fournir à la plateforme Coltrane l'échantillon de l'enquête avec ses caractéristiques (au format XML). Ce fichier dit « de personnalisation » contient pour chaque unité enquêtée :

- ❶ l'ensemble des données d'identification des entreprises sélectionnées (identifiant de l'unité, raison sociale, etc.) ;
- ❷ éventuellement les contacts : identifiant du contact s'il existe déjà dans Coltrane et si l'on veut le partager avec d'autres responsables d'enquêtes, adresse, nom-prénom, fonction, téléphone, adresse mail, etc.) ;
- ❸ et d'autres données utiles à la collecte : réponses de l'unité enquêtée lors d'une précédente campagne de collecte, si elles existent.

Coltrane génère et héberge tous les questionnaires *web* des unités à interroger, à partir du modèle de questionnaire vierge et non personnalisé fourni par Eno, et du fichier de personnalisation fourni par l'application externe de gestion d'enquête. Au moment du lancement de la collecte, Coltrane a donc généré autant de questionnaires vierges (mais personnalisés) qu'il y a d'unités à interroger.

« Au moment du lancement de la collecte, Coltrane a donc généré autant de questionnaires vierges (mais personnalisés) qu'il y a d'unités à interroger. »

Une fois la collecte d'une enquête ouverte, Coltrane extrait les réponses qui ont été collectées depuis la dernière extraction, au rythme choisi par le responsable de l'enquête (en pratique jusqu'à 4 fois par jour) et les transmet

(au format XML) à l'application de gestion de l'enquête. Certains responsables d'enquête ont également choisi de récupérer les PDF « preuves de dépôt ».

En revanche, Coltrane n'offre pas de service d'archivage des réponses post-collecte. Les réponses des unités enquêtées ne sont conservées sur la plateforme que trois mois après la fermeture de la collecte ; cela permet aux gestionnaires d'accéder aux données brutes durant toute leur phase de validation des données. Elles sont ensuite supprimées, à charge au responsable d'enquête de les sauvegarder dans son environnement propre, si besoin était.

19. Voir l'article de Franck Cotton et Thomas Dubois dans ce même numéro.

20. Cette application est en dehors du projet Coltrane et différente pour chaque enquête aujourd'hui. Elle est de la responsabilité de chaque maîtrise d'ouvrage d'enquête et permet de gérer la collecte et de contrôler les réponses aux questionnaires.

UN MODULE POUR GÉRER LES COURRIERS AUX ENTREPRISES

Pour mener à bien la collecte d'une enquête, il est indispensable de contacter régulièrement les entreprises, par exemple pour leur rappeler leur obligation de répondre aux enquêtes labellisées de la statistique publique. Usuellement, quatre modèles de courriers sont prévus : ouverture de la collecte, rappel, mise en demeure, constat de non-réponse. Coltrane propose également un modèle de courrier pour les entrées dans un panel et un autre pour remercier les sortants de panel.

Au cours du projet Coltrane, l'Insee a développé un module pour créer des courriers « prêts à être imprimés ». Ce « **Module Courrier** » est aujourd'hui mis à disposition de toute autre application ayant des besoins similaires. Pour les enquêtes collectées grâce à Coltrane :

- ❶ la partie générique de ces courriers a été standardisée puis validée par le comité du Label de la statistique publique²¹ ;
- ❷ le contenu spécifique à chaque enquête (nom de l'enquête, objectif, etc.) est composé à l'aide de métadonnées stockées dans les bases Coltrane et alimentées par le concepteur d'enquête, selon le principe de métadonnées actives déjà évoqué.

Comme précédemment, Coltrane s'appuie sur un fichier XML de personnalisation (liste et éléments de personnalisation des unités pour lesquelles le courrier doit être envoyé, type de courrier) fourni par l'application de gestion de l'enquête, pour produire un fichier qu'il transmet ensuite au Module Courrier.

Le Module Courrier utilise les modèles (ouverture de la collecte par exemple) et le fichier de Coltrane pour éditer des courriers personnalisés au format PDF, ainsi que des informations utiles au pilotage de l'expédition des courriers²².

Coltrane n'est cependant pas un outil générique de gestion de la collecte. Ainsi :

- ❶ c'est toujours l'application de gestion d'enquête qui indique le type de courrier à créer et la liste des unités concernées ;
- ❷ l'offre Coltrane ne concerne que les courriers dits de masse. Elle ne permet pas l'envoi d'un courrier spécifique ou ponctuel d'un gestionnaire à une unité enquêtée particulière.

Les courriers peuvent être accompagnés d'un questionnaire au format papier, selon la stratégie de collecte adoptée. L'expérience montre par exemple que pour augmenter le taux de réponse par internet, il est préférable de ne pas joindre de questionnaire papier à l'ouverture d'une collecte : le taux de réponse par internet pour l'enquête Acemo est ainsi passé de 20 à 80 % l'année où la Dares a décidé de ne plus envoyer de questionnaire « papier » lors de l'envoi initial. En revanche, il est préférable d'envoyer un questionnaire papier au moment de la mise en demeure pour éviter toute contestation ultérieure de l'entreprise. Il faut aussi pouvoir effectuer un tel envoi lorsque l'enquêté en fait la demande. Pour cela, la procédure est la même que pour un courrier classique à ceci près que le modèle instancié dans le Module Courrier est généré par Eno, comme l'était le support de collecte *web*. Coltrane viendra en quelque sorte « publier » le modèle de questionnaire.

21. Le comité propose ainsi sur son site <https://www.comite-du-label.fr/> des modèles de cartouche à utiliser sur les questionnaires pour informer les entreprises enquêtées des conditions réglementaires de l'obligation de réponse.

22. À l'Insee, l'équipe « éditique » est en charge de l'impression et de l'envoi des documents papier en masse.

ENCOURAGER LA RÉPONSE DES ENTREPRISES VIA UN COURRIEL PRÉALABLE

Coltrane permet aussi, lors des ouvertures de collecte ou des rappels, de doubler l'envoi du courrier par un message électronique aux contacts disposant d'une adresse mail, qui les invite à répondre à l'enquête.

La logique en est la même que pour l'envoi des courriers : à partir de modèles de mail pré-définis et d'un fichier XML de personnalisation fourni par l'application de gestion de l'enquête, Coltrane génère un message qui est diffusé *via* un outil de l'Insee d'envoi de courriels automatique. Dans ce message figurent des informations sur l'enquête concernée (nom, objectifs, date de retour attendu, etc.), des extraits des textes légaux relatifs aux enquêtes mais surtout le lien pour accéder au portail d'authentification, un rappel de l'identifiant du contact et un lien en cas d'oubli de mot de passe, afin de réduire les demandes d'assistance.

Par convention, l'émetteur de ces mails est le service collecteur de l'enquête, ce qui permet à l'enquêté d'avoir une assistance facile à contacter.

En outre, dans moins d'un an il sera également possible d'envoyer le questionnaire au format PDF par mail à une entreprise qui le souhaite.

Cette offre de services sera bien entendu d'autant plus efficace que la maîtrise d'ouvrage de l'enquête disposera d'une base de contacts à jour, contenant un maximum d'adresses mail valides (les mails invalides sont d'ailleurs signalés aux concepteurs des enquêtes qui utilisent l'offre « envoi de mails »).

UNE APPLICATION INTERNE POUR GÉRER LES CONTACTS

Coltrane est plus qu'une plateforme de collecte conviviale pour les répondants, facilitant le travail des concepteurs d'enquêtes. C'est aussi un référentiel de contacts qui va servir aux gestionnaires d'enquêtes qui traitent les réponses et sont en lien avec les entreprises enquêtées.

Pour ce faire, lorsqu'une enquête intègre le dispositif, Coltrane importe massivement les contacts déjà connus de l'enquête (en particulier lorsqu'elle recourait déjà à la collecte par internet). Ensuite, afin de faire vivre ce référentiel, les gestionnaires d'enquête utilisent une application de « **gestion des contacts** » : pour créer un contact, modifier ses caractéristiques, lui donner ou lui enlever le droit de répondre à une enquête, etc. En outre, les mises à jour de ses coordonnées, faites directement par les répondants (*via* la partie « mon compte » de l'interface Coltrane) sont enregistrées en temps réel dans ce référentiel.

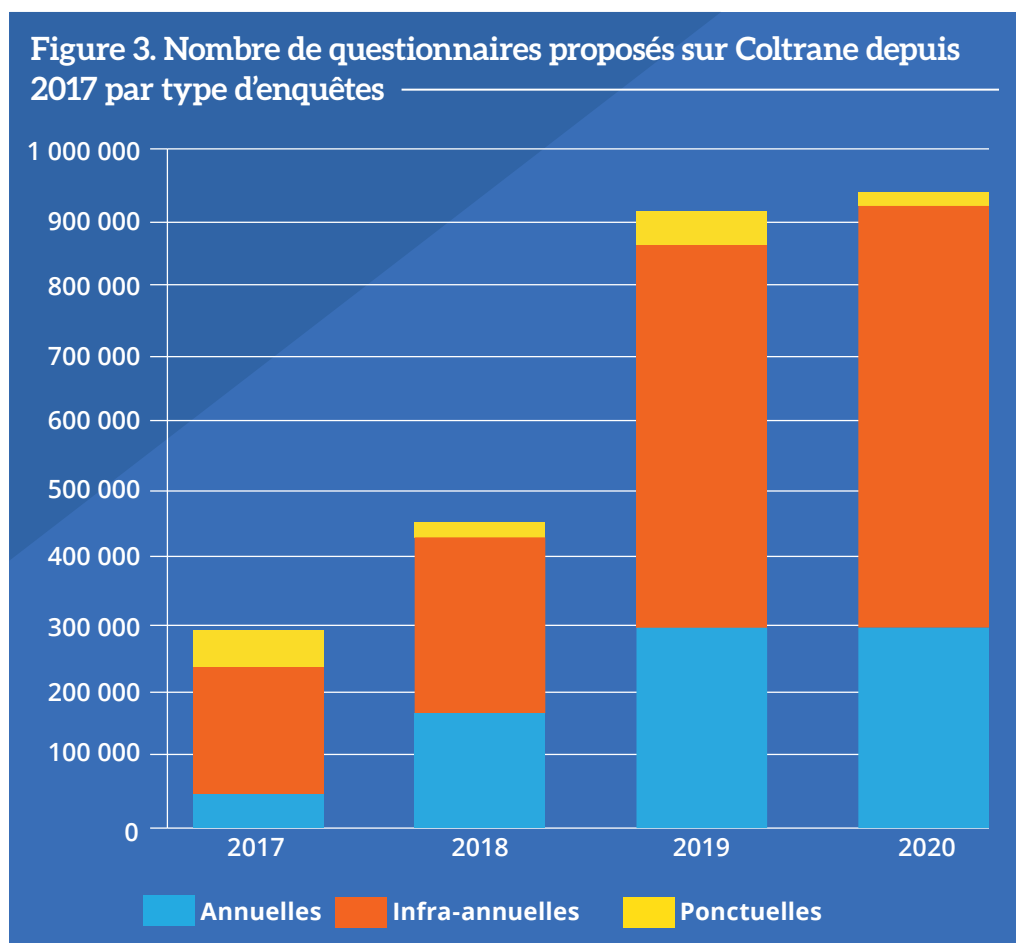
Lorsqu'un gestionnaire modifie une caractéristique d'un contact répondant à plusieurs enquêtes, les responsables des autres enquêtes sont prévenus par mail. L'application est également accessible par les pôles Insee-Contact pour assister l'enquêté (perte de code d'accès par exemple).

Par ailleurs, les gestionnaires d'une enquête peuvent visualiser les questionnaires de leurs enquêtés quel que soit leur statut (déjà transmis à l'Insee ou pas). Ceci permet notamment d'aider les entreprises qui ont des difficultés à répondre, ou de vérifier l'envoi ou le simple enregistrement d'un questionnaire.

LE MILLION DE QUESTIONNAIRES BIENTÔT ATTEINT

Coltrane a atteint en 2019 son régime de croisière avec l'intégration de toutes les enquêtes auprès des entreprises conduites par l'Insee (à l'exception de celles collectées par le biais d'un dépôt/retrait de questionnaires²³) et par la Dares, proposant une collecte par internet. Le nombre d'enquêtes menées selon ce dispositif ne cesse de croître au fil des ans.

L'objectif est de maximiser l'usage de Coltrane pour les enquêtes du Service statistique public, afin que les entreprises qui répondent à plusieurs enquêtes n'aient à utiliser qu'une seule infrastructure. S'il est atteint, le million annuel de questionnaires²⁴ sous Coltrane sera dépassé. En 2019, 900 000 questionnaires auront été accessibles sur la plateforme (figure 3) et on devrait franchir pour la première fois la barre des 500 000 questionnaires retournés. Par ailleurs, depuis 2015, plus de 615 000 contacts ont été enregistrés dans le référentiel.



23. Il s'agit également d'une collecte dématérialisée mais la logique est différente : le dépôt retrait permet à l'entreprise d'envoyer à l'Insee de façon sécurisée des données déjà existantes sous forme de fichiers dans l'entreprise sans avoir à les ressaisir.

24. En raison de la fréquence de certaines enquêtes.

Ce volume conséquent génère des exigences de robustesse sur le site, dont l'activité n'est pas uniforme au cours de l'année, ni même au cours d'une journée (on constate dans la **figure 4** des pics de collecte quotidiens en fin de matinée et en fin d'après-midi).

🌐 LA COLLECTE AUJOURD'HUI : UN MÉTIER À PART ENTIÈRE —

La mise en œuvre de Coltrane a conduit à industrialiser, mutualiser et professionnaliser la collecte.

Le générateur Eno offre des supports standardisés, mettant en œuvre les règles de bonnes pratiques du questionnement par internet (Morrison, 2010 ; Snijkers, 2013 ; Dillman 2014). Ces règles sont gérées à l'Insee par une section spécialisée au sein de la direction de la Méthodologie qui assure une veille sur ces sujets au niveau international. Ce travail permet de disposer de supports de collecte répondant aux standards en vigueur aussi bien au niveau des supports papier que *web*. Il conduit par conséquent à une homogénéité des pratiques et de l'ergonomie des supports pour les utilisateurs de cette filière.

La collecte gagne ainsi en robustesse, car cette spécialisation permet d'obtenir plus de retours d'expérience et de mieux appréhender les effets de mode de collecte, ou les impacts éventuels de l'ergonomie sur les réponses.

Cependant, cette standardisation est parfois mal vécue par certains concepteurs qui vivent cette exigence comme une réduction de leur marge de manœuvre, en termes de maquettage des questionnaires ou de formulation des questions. En effet, la standardisation imposée par cette nouvelle filière conduit souvent à abandonner des choix de mise en forme anciens, qui permettaient des gains de place et une réduction des coûts liés à l'envoi massif de papier, mais n'étaient pas toujours accompagnés de la lisibilité requise.

« La standardisation imposée par cette nouvelle filière conduit souvent à abandonner des choix de mise en forme anciens. »

Pour les aider à entrer dans cette démarche nouvelle, les concepteurs sont guidés par une documentation présentant les possibilités du générateur Eno, les bonnes pratiques d'interrogation *web*, un protocole de création d'un questionnaire.

En outre, ils sont accompagnés dans leur parcours par des experts spécialisés dans le métier qu'est la collecte. Grâce à l'offre du Module Courrier, ils évitent également la conception, souvent fastidieuse, des courriers d'accompagnement.

In fine, cela permet de proposer des supports de collecte plus homogènes, des courriers cohérents, bref une offre qui rend les questionnaires plus accessibles par les enquêtés. Le choix d'une ergonomie commune est également un atout majeur pour identifier plus facilement les questionnaires comme provenant de la statistique publique.

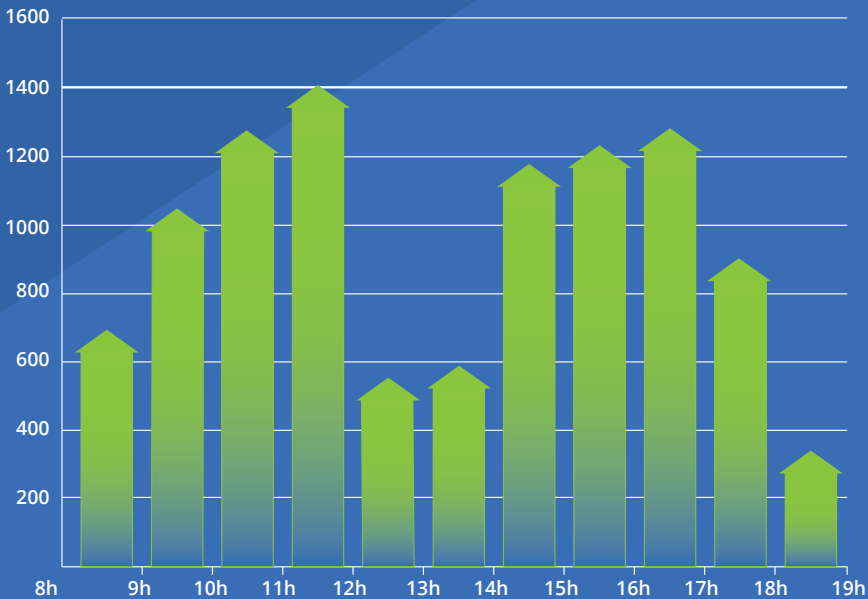
Cette mutualisation n'est qu'un premier pas. Un autre projet est en cours, visant à uniformiser les applications de gestion des enquêtes thématiques et à mutualiser les outils de contrôle post-collecte. Il se place directement dans la filière Pagues/Eno/Coltrane et achèvera de limiter, pour le concepteur, la charge de description du questionnaire : la même occurrence sera commune à toutes les phases de l'enquête, depuis la Collecte jusqu'au Traitement, au sens du GSBPM.

VERS UNE APPLICATION AUX ENQUÊTES MÉNAGES ?

La plateforme Coltrane est dédiée aux enquêtes auprès des entreprises, qui ont leurs spécificités. La généralisation aux enquêtes ménages ne va cependant pas de soi²⁵. Néanmoins, le caractère modulaire de Coltrane et le fait que ces modules soient désormais éprouvés par des années de pratique permet d'imaginer une forme de réutilisation, fût-elle partielle. Ainsi, dans le cadre des travaux de l'Insee sur la collecte multimode des enquêtes ménages, on envisage que la partie « collecte par internet » passe par un portail dédié, et que ce portail réutilise un nombre significatif d'investissements déjà réalisés pour Coltrane. Cela permettra non seulement de mutualiser, donc de réduire, les moyens de maintenance des deux applications et ainsi de gagner en flexibilité, mais aussi, vu l'importance stratégique, de spécialiser des personnes sur les technologies utilisées et de sécuriser ainsi l'ensemble du processus de collecte dématérialisée de l'Insee.

Coltrane pourrait également bénéficier, de façon symétrique, de développements réalisés pour ce portail « ménages » et augmenter ainsi son offre de service. On peut notamment évoquer les travaux sur le *responsive design*²⁶ qui ont été pris en compte par Coltrane courant 2019, ou ceux sur le site public de promotion des enquêtes auprès des ménages dont le principe pourrait être adapté aux besoins des enquêtes auprès des entreprises.

Figure 4. Nombre maximal de connexions par heure entre 8 h et 19 h en 2019 sur le site Coltrane



Note de lecture : en 2019, on a enregistré jusqu'à 1 400 connexions entre 11 h et midi sur le portail de réponse Coltrane. C'est aussi le maximum jamais atteint sur une heure dans une journée en 2019.

25. Voir l'article de Jean-Marc Béguin dans ce même numéro sur la naissance d'une innovation en production statistique.
26. Le *responsive design* ajuste automatiquement l'affichage d'une page *web* à la taille d'écran du terminal utilisé. Cette technique de conception de site *web*, ou d'interface digitale, répond à un besoin des utilisateurs, toujours plus nombreux à se connecter sur le *web* depuis un appareil mobile (téléphone, tablette, etc.).

DES ÉVOLUTIONS NÉCESSAIRES

Au-delà de mutualisations possibles, à terme, avec les enquêtes ménages, d'autres fonctionnalités sont à l'étude afin d'améliorer encore les services offerts aux répondants et aux concepteurs.

Pour les entreprises, l'équipe étudie un outil d'aide à la codification ou d'auto-complétion. Cela concerne les questions posées aux entreprises qui doivent conduire au choix d'un code dans une nomenclature, ayant parfois de très nombreuses modalités (comme la NAF par exemple). Aujourd'hui, le contact de l'unité enquêtée saisit un libellé, lequel est ensuite codé dans l'application de gestion de l'enquête (donc hors Coltrane). Un outil d'aide à la codification en ligne serait à même de faciliter la réponse à ce type de questions.

Dans un premier temps, on envisage de mettre en œuvre une auto-complétion pour saisir le code de la commune dans les adresses du contact. Pour l'appliquer à un questionnaire, il faudra auparavant mesurer l'effet « mode de collecte » qu'elle engendre afin de pouvoir le corriger. À l'heure actuelle, en l'absence d'aide à la saisie et de contrôle bloquant dans le portail de réponse, l'hypothèse d'absence d'impact du mode de collecte entre un questionnaire *web* et un questionnaire papier paraît plausible (même si elle n'a pas été formellement testée). Mais avec la mise en œuvre de processus d'auto-complétion, cette hypothèse tombe puisqu'on développe cet outil avec l'espoir d'améliorer (donc de modifier) la qualité de la réponse.

Une seconde piste d'aide à la saisie est envisagée pour la réponse à des tableaux contenant un nombre important de cases (dont la valeur est au demeurant tirée du système d'information de l'entreprise). Dans ce cas, l'idée serait de proposer d'importer un fichier plutôt que de saisir chaque case du tableau. Pour les entreprises de grande taille qui sont systématiquement interrogées et pour les enquêtes répétitives, l'écriture d'un petit programme permettant de fabriquer le fichier en question serait très probablement rentable.

Pour les concepteurs, de nouvelles fonctionnalités associées au pilotage de la collecte seront offertes aux responsables d'enquête afin de les rendre encore plus autonomes. Ainsi, une nouvelle interface sera développée pour qu'ils puissent saisir eux-mêmes, sans avoir à recourir aux équipes de maintenance de Coltrane, les métadonnées nécessaires à la génération des supports de collecte et à leur utilisation (nom et objectifs de l'enquête, dates d'ouverture, de retour attendu, de fermeture de la campagne), métadonnées qui pilotent l'affichage des données sur le portail de réponse. Ces fonctionnalités seront ouvertes aux concepteurs en dehors de l'Insee.

Ces évolutions sont d'autant plus importantes que la plateforme Coltrane est désormais la norme pour les enquêtes internet auprès des entreprises. Elle s'impose ainsi aux concepteurs, il est donc logique qu'en contrepartie de cette contrainte, Coltrane leur facilite le travail *via* une palette de services suffisamment large.

BIBLIOGRAPHIE

BONNANS, Dominique, 2019. RMÉS, le référentiel de métadonnées statistiques de l'Insee. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 46-55. [Consulté le 5 novembre 2019]. Disponible à l'adresse :

<https://www.insee.fr/fr/statistiques/fichier/4168396/courstat-2-6.pdf1>

DILLMAN, Don A., SMYTH, Jolene D., CHRISTIAN Leah M., 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys : The Tailored Design Method*. Août 2014. Éditions Wiley, collection Evaluation & Research Methods, 4^e édition. ISBN 978-1-118-45614-9

FAZIO, Natale Renato et MOTTURA, Giulia, 2015. The Business Portal – an integrated system to collect data and communicate. *Workshop on Statistical Data Collection : Riding the Data Deluge*. [en ligne]. 29 avril–1^{er} mai 2015. Washington, USA. [Consulté le 5 novembre 2019]. Disponible à l'adresse :

https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2015/mtg1/WP4-Fazio_Mottura_AD.pdf

MOREIRA, Almiro et SARAIVA, Paulo, 2016. Managing an integrated respondent communication: statistics Portugal experience. *Conférence européenne sur la qualité en statistique officielle (Q2016)*. [en ligne]. 31 mai-3 juin 2016. Madrid, Espagne. [Consulté le 25 octobre 2019]. Disponible à l'adresse : <http://www.ine.es/q2016/docs/q2016Final00146.pdf>

MORRISON, Rebecca L., DILLMAN, Don A. et CHRISTIAN, Leah M., 2010. Questionnaire Design Guidelines for Establishment Surveys. In : *Journal of Official Statistics*. [en ligne]. Vol. 26, n° 1, pp. 43-85. [Consulté le 5 novembre 2019]. Disponible à l'adresse :

<https://www.scb.se/contentassets/ff271eeeca694f47ae99b942de61df83/questionnaire-design-guidelines-for-establishment-surveys.pdf>

SNIJKERS, Ger, HERALDSEN, Gustav, JONES, Jacqui et WILLIMACK, Diane, 2013. *Designing and conducting business surveys*. Août 2013. Éditions Wiley, collection Survey Research Methods & Sampling. ISBN 978-0-470-90304-9


THIEME, Michael T. et MATHUR, Anup, 2015. *Designing and Architecting a Shared Platform for Adaptive Data Collection in Surveys and Censuses*. [en ligne]. 4 mars 2015. US Census. [Consulté le 5 novembre 2019]. Disponible à l'adresse :

https://www.census.gov/fedcasic/fc2015/ppt/16_thieme.pdf

UTILISER LES DONNÉES DE CAISSES POUR LE CALCUL DE L'INDICE DES PRIX À LA CONSOMMATION

Marie Leclair*

Les données de caisses sont des données recueillies par les enseignes de la grande distribution quand le consommateur passe à la caisse des magasins. Ces données très volumineuses seront utilisées pour calculer l'indice des prix à la consommation (IPC) à partir de janvier 2020. La couverture exhaustive du champ par ces données, la connaissance fine de la consommation des ménages qu'elles apportent sont des avancées majeures pour la production de statistiques des prix : elles en améliorent, notamment, la précision et permettent à terme de produire de nouvelles statistiques (indices plus détaillés, indices régionaux, prix moyens, comparaisons spatiales des prix). Cependant, elles obligent également à trouver un certain nombre de solutions nouvelles, en particulier en termes d'automatisation des traitements. Le choix d'une solution informatique big data, le recours à un référentiel des codes-barres contribuent à pouvoir exploiter ces données, tout en conservant les concepts de l'IPC, notamment le recours à un panier fixe de produits.

 *Scanner data are data gathered by large retailers when consumers go to pay for their goods in store. From January 2020 onwards, these enormous volumes of data will be used to calculate the consumer price index (CPI). The comprehensive coverage of the field by these data and the detailed knowledge of household consumption that they provide are major steps forward when it comes to producing price statistics: they improve, in particular, their accuracy and ultimately mean that new statistics can be produced (more detailed indexes, regional indexes, average prices and spatial price comparisons). However, they also require that a number of new solutions be found, in particular in terms of automated processing. Choosing a big data IT solution and using a barcode dictionary help enable these data to be exploited whilst preserving the concepts of the CPI, in particular the reference to a fixed basket of goods.*

L'indice des prix à la consommation (IPC) français est calculé à partir de 200 000 relevés mensuels effectués par des enquêteurs dans des points de vente physiques. Cette collecte de terrain a été progressivement complétée par d'autres sources, dématérialisées : relevés sur internet et données administratives totalisent 190 000 prix additionnels chaque mois. À partir de janvier 2020, une nouvelle source, d'une toute autre ampleur sera utilisée : les données de caisses.

*Cheffe de la division des Prix à la consommation, Insee,
marie.leclair@insee.fr

📊 UNE NOUVELLE SOURCE DE DONNÉES, LES DONNÉES PRIVÉES

Les données de caisses sont les informations sur les prix payés et les produits achetés, recueillies par les enseignes au moment où le consommateur passe à la caisse des magasins.

« Les données de caisses sont les informations sur les prix payés et les produits achetés, recueillies par les enseignes au moment où le consommateur passe à la caisse des magasins. »

Beaucoup plus volumineuses que les données mobilisées jusqu'à présent pour le calcul de l'IPC (1,7 milliard d'enregistrements reçus chaque mois), ces données privées sont une réelle opportunité pour le calcul de statistiques de prix. Mais ces données privées posent également de nouvelles questions, sur leur accès tout d'abord, sur leur fiabilité ensuite, et enfin sur la capacité de l'Insee à les exploiter à des fins statistiques et d'un point de vue informatique.

Les données de caisses, qu'on appelle plus génériquement données de transaction, sont disponibles pour bien des magasins, dès lors que les transactions sont enregistrées. Si la centralisation en favorise la mise à disposition, le traitement statistique de ces fichiers n'est pas pour autant aisé pour tous les produits de la consommation : à ce stade, l'Insee n'exploite « que » les données des enseignes de la grande distribution (super et hypermarchés) de France métropolitaine, pour les produits alimentaires industriels, d'entretien et d'hygiène-beauté.

📊 UNE OPPORTUNITÉ POUR LE CALCUL DE STATISTIQUES DE PRIX

Les données de caisses sont mobilisées depuis de nombreuses années par des panélistes à des fins d'études de marché. Certains instituts nationaux statistiques ont vu assez tôt l'intérêt de mobiliser de telles sources pour le calcul de leur IPC : les Pays-Bas les utilisent ainsi depuis 2002, suivis par la Norvège en 2005, la Suisse (2008), la Suède (2012), la Belgique (2015), le Danemark (2016), l'Islande (2016), le Luxembourg et l'Italie (2018). Eurostat a contribué à l'extension de leur utilisation, *via* des *grants*, des *workshops* et un manuel (Eurostat, 2017). L'utilisation des données de caisses est également un sujet récurrent du groupe des experts prix de l'UNECE¹ et du BIT², réunissant des universitaires et des statisticiens du monde entier (UNECE, 2018).

Différentes motivations expliquent cet intérêt : le souhait de faire entrer les instituts statistiques dans l'ère du big data, le recours à des données privées produites « gratuitement » (même si leur exploitation peut être coûteuse), l'exhaustivité des données mais également l'apport de nouvelles informations sur les produits consommés, non disponibles jusqu'à présent et qui ouvrent de nombreuses opportunités pour les statistiques de prix, comme on le verra par la suite.

1. Commission économique pour l'Europe des Nations unies (CEE-ONU), en anglais *United Nations Economic Commission for Europe* (UNECE).

2. Bureau international du travail.

Dans le cas français, les données de caisses sont des données exhaustives sur leur champ ; elles sont collectées en continu et centralisées selon un pas journalier. Ainsi pour chaque code-barres de produits³ (*figure 1*), sont enregistrés les quantités vendues du produit ainsi que le prix ou le chiffre d'affaires associé, par points de vente et jour de vente (*figure 2*).

DES STATISTIQUES PLUS PRÉCISES ET PLUS DÉTAILLÉES

Selon Tassi (2019), les données *big data* se caractérisent par la quantité d'information (pouvant aller jusqu'à l'exhaustivité sur un domaine donné) et la fréquence d'acquisition de cette information. Dans le cas des statistiques de prix, la disponibilité des données selon un pas quotidien et en continu⁴ est importante pour produire un indice mensuel comme l'IPC, qui plus est selon des contraintes temporelles fortes⁵. Il n'est cependant pas prévu de produire un IPC infra-mensuel.

L'exhaustivité de la source permet, en revanche, de produire des statistiques plus précises et d'envisager de produire des indices plus détaillés, par exemple sur des segments de consommation spécifiques. L'IPC est déjà produit mensuellement pour plus

Figure 1. La structure d'un code-barres

Tout article destiné à la vente au consommateur dans un magasin de détail doit être identifié avec un code à barres. Le code-barres est la transcription graphique d'un code GTIN (*Global Trade Item Number*). Le code GTIN et les barres forment un tout indissociable pour sa reconnaissance en caisse. Les barres permettent la saisie automatique des données à chaque lecture de l'article.

Un code-barres est donc *un identifiant unique de produit, respectant une clé et donnant une information sur le pays et l'entreprise de production, mais il ne permet pas de donner une information sur la nature du produit suivi.*

Le GTIN (nom actuel et officiel), l'EAN (ancienne dénomination Européenne) ou le GENCOD (ancienne dénomination française), désignent le code qui identifie une unité commerciale c'est-à-dire une unité dont le prix peut être fixé, qui peut être commandée, livrée ou facturée

aux fins d'échanges commerciaux. Ces unités commerciales peuvent être des unités de vente au consommateur, des unités logistiques (cartons, palettes, box, bacs...), des unités d'expédition. Le code GTIN est inscrit le plus souvent en chiffres au-dessous des barres. Il en existe différents formats, dont le plus fréquemment utilisé est le GTIN 13, soit 13 chiffres structurés de la manière suivante :



Préfixe entreprise : 6, 7, 8, 9, 10 ou 11 chiffres selon les besoins de codification de l'entreprise.

Code produit : attribué par le propriétaire de la marque commerciale qui dispose de 6, 5, 4, 3, 2 ou 1 chiffres selon la longueur du préfixe entreprise. C'est la seule partie variable du code, incrémentée par l'entreprise.

Clé de contrôle : 1 caractère, calculé sur la base des chiffres précédents et permet d'éviter toute erreur de saisie.

3. Appelé également GTIN pour *Global Trade Item Number* ou EAN pour *European Article Numbering*.

4. Les données sont envoyées à l'Insee avec deux jours de décalage.

5. Une estimation provisoire de l'IPC est publiée dès le dernier jour ouvré du mois.

de 250 sous-classes et annuellement pour plus de 360 postes⁶. Une demande de la part du public d'information toujours plus détaillée existe, mais ce n'est pas forcément le principal apport des données de caisses.

C'est la dimension géographique de l'exhaustivité des données de caisses qui est particulièrement intéressante. Jusqu'à présent, les prix sont relevés par l'IPC dans un échantillon de communes de plus de 2 000 habitants, représentatif au niveau national. Or l'Insee rencontre une difficulté pratique dans les plus petites communes car, du fait d'un tissu commercial moins dense, les enquêteurs doivent parcourir beaucoup plus de kilomètres pour relever des prix.

Outre la fiabilisation de la collecte dans les zones plus rurales, l'exhaustivité permet d'être représentatif au niveau de chaque région et d'envisager de produire, à terme, des indices de prix régionaux (sur le champ des données de caisses) : jusqu'à présent, seuls un indice métropolitain et pour chaque département d'Outre-mer étaient publiés. Par ailleurs, la représentativité par territoire fin et le détail de l'information sur les produits suivis permettent d'envisager à terme des comparaisons spatiales de niveau de prix : cet exercice n'est réalisé actuellement que tous les 5-6 ans par l'Insee et, en France métropolitaine, uniquement entre l'agglomération parisienne, la Corse, et la province (Clé et *alii*, 2016). Des travaux expérimentaux (Léonard et *alii*, 2019) montrent que ces données de caisses peuvent être avantageusement mobilisées pour des comparaisons spatiales de prix. Elles sont d'ailleurs déjà utilisées par certains pays pour des comparaisons européennes de niveaux de prix, les parités de pouvoir d'achat.

Figure 2. Un échantillon des données de caisses

Identifiant du point de vente	EAN	Description de l'article	Date des ventes	Quantités vendues	Prix de vente (en €)	Chiffre d'affaires (en €)
723	3275770004817	██████████ 150G	20140108	10	1.89	18.90
723	3155230040286	██████ BACON 150G	20140108	7	2.38	16.66
986	3185670001080	██████ STRAINED SOFT 6%MG 1KG	20140128	25	2.59	64.75

Permet de faire le lien avec le référentiel des points de vente
(localisation, surface)

Permet de faire le lien avec le référentiel des articles
(marque, description détaillée, volume)

6. L'indice des prix à la consommation est publié selon la nomenclature Coicop (*Classification of Individual Consumption by Purpose*) ; la sous-classe et le poste sont les deux derniers niveaux de cette nomenclature.

1 CONNAÎTRE LES QUANTITÉS CONSOMMÉES ET DISPOSER D'UNE BASE DE SONDAGE

Si les statisticiens des prix se sont autant intéressés aux données de caisses, c'est aussi parce qu'elles leur donnent accès à une information dont ils avaient jusqu'à présent une connaissance très grossière, ancienne et agrégée : le type de produits consommés par les ménages et les quantités consommées par ceux-ci.

Par exemple, la comptabilité nationale donne aujourd'hui une information sur le poids des céréales pour petit déjeuner dans la consommation sur le territoire français. Mais elle ne la détaille pas par variété de céréales ou encore moins par marques ou par points de vente.

On ne peut donc pas s'appuyer sur une base de sondage pour tirer un échantillon aléatoire de produits dont on suit le prix. Faute d'information, le panier de l'IPC est aujourd'hui *de facto* défini par une méthode de quotas : les unités urbaines dans lesquelles se rend l'enquêteur sont certes tirées aléatoirement en fonction de la population qui y vit et des habitudes de consommation qu'on lui impute (Jaluzot et Sillard, 2016). Mais le choix des points de vente précis et des produits suivis est effectué par l'enquêteur en fonction de quelques contraintes ou quotas (forme de vente, variétés de produits suivies). Dans le cas des céréales de petit déjeuner, on demandera à un enquêteur de se rendre dans une unité urbaine donnée, tirée aléatoirement, et d'y trouver par exemple 4 paquets de céréales de type muesli, dont un en hypermarché, deux en supermarché et un en supérette. C'est l'enquêteur qui choisira les points de vente, et dans ceux-ci la boîte de céréales qu'il retiendra.

« Avec les données de caisses, on dispose enfin d'une base de sondage : l'ensemble des articles vendus par points de vente, avec le poids de chacun de ces articles dans le chiffre d'affaires des points de vente. »

Avec les données de caisses, on dispose enfin d'une base de sondage : l'ensemble des articles vendus par points de vente, avec le poids de chacun de ces articles dans le chiffre d'affaires des points de vente. L'existence d'une base de sondage permet de recourir à un tirage aléatoire des produits d'une part, et d'autre part de maîtriser le biais d'échantillonnage. Elle permet également de repérer rapidement les produits nouveaux à introduire dans le panier IPC ou les produits en perte de vitesse qu'il faut retirer, afin que le panier soit toujours à jour et représentatif de la consommation des ménages.

Au final, à l'Insee, compte tenu des possibilités informatiques et d'automatisation (voir *infra*), c'est l'exhaustivité des données de caisses qui a été retenue, sans recourir à un échantillonnage.

1 MIEUX TRAITER LES SUBSTITUTIONS DU CONSOMMATEUR ENTRE LES PRODUITS

La connaissance des quantités précises vendues pour chaque article permet également de progresser dans la pratique des indices : pour calculer un indice synthétique des prix, les relevés sont agrégés *via* un certain nombre de formules (*encadré*). La théorie des indices définit les propriétés de ces formules et les indices qu'il convient de retenir d'un point de vue théorique (Sillard, 2016 ; FMI, 2004). Dans la pratique, l'absence de connaissance sur les quantités courantes consommées par les consommateurs à un niveau détaillé contraint le choix.

Pour l'IPC, on utilise :

- ① au niveau le plus agrégé (par exemple pour agréger les boîtes de céréales avec les pâtes alimentaires) un indice de type Laspeyres, avec une pondération utilisant les dépenses passées de consommation ;
- ① au niveau le plus fin (par exemple pour agréger les différentes boîtes de céréales entre elles), en l'absence d'information sur les quantités consommées, même passées, un indice de Dutot ou Jevons, impliquant une équipondération des prix relevés.

Par le détail de l'information qu'elles apportent, les données de caisses permettent de choisir des formules d'indice qui tiennent mieux compte au niveau le plus fin des substitutions effectuées par le consommateur entre deux produits (Leclair et *alii*, 2019) : lorsque le prix d'un produit augmente, l'impact sur l'utilité du consommateur peut être plus ou moins fort selon qu'il peut ou non reporter sa consommation vers un autre produit, plus ou moins substituable au premier. La capacité de traiter ces substitutions avait fait l'objet de débats, dans les années quatre-vingt-dix, sur l'éventualité d'un biais, à la baisse, des indices de prix qui les auraient insuffisamment prises en compte (Boskin, 1996). À l'époque, le recours aux données de caisses et à l'information fine sur les quantités vendues, donc

Encadré. Comment produit-on l'IPC ?

L'IPC mesure l'évolution des prix des produits consommés par les ménages. Les prix d'un panier fixe de produits sont suivis chaque mois de manière à mesurer une évolution pure de prix, à qualité constante. L'indice est un indice de type Laspeyres*, les différentes variétés de produits sont pondérées par leur poids passé dans la consommation des ménages. À un niveau plus fin que la variété des produits, les pondérations ne sont plus connues et des hypothèses sont effectuées pour agréger les prix élémentaires : les formules de Dutot** et de Jevons*** sont utilisées par l'IPC.

Afin de demeurer représentatifs de la consommation des ménages, les poids et le panier de produits suivis sont renouvelés chaque année : l'IPC est un indice chaîné annuellement. Par ailleurs, en cas de disparition d'un produit en cours d'année, celui-ci est remplacé par un produit proche et un ajustement qualité est effectué afin de corriger de l'écart de qualité entre le produit remplacé et remplaçant.

L'IPC est publié à un rythme mensuel, dès le dernier jour ouvré du mois pour l'indice provisoire, quinze jours environ après la fin du mois pour l'indice définitif. Cet indice définitif n'est par la suite plus révisé. Ces délais très courts et l'absence de révision imposent des contraintes très fortes au processus de production de l'IPC.

Outre les données de transaction, l'IPC utilise deux types de sources : des relevés de prix effectués par des enquêteurs de l'Insee sur le terrain (avant l'utilisation des données de caisses, de l'ordre de 200 000 relevés chaque mois dans des unités urbaines représentatives du territoire français) dans diverses formes de vente (y compris internet) ; des relevés collectés de manière centralisée soit que le prix de ces produits soit unique sur tout le territoire (service de télécommunication, électricité, tabac...), soit que des bases de données puissent être mobilisées pour calculer les évolutions de prix (données de la Caisse nationale de l'assurance maladie pour les services de santé, par exemple).

* L'indice de Laspeyres est un indice à panier fixe rapportant la moyenne des prix de la période courante à la moyenne des prix de la période de référence, en pondérant les prix courants et de la période de référence par les quantités consommées au cours de la période de référence.

** Un indice de Dutot est un indice à panier fixe rapportant la moyenne arithmétique des prix de la période courante à la moyenne arithmétique des prix de la période de référence. Tous les prix du panier sont équipondérés.

*** Un indice de Jevons est un indice à panier fixe rapportant la moyenne géométrique des prix de la période courante à la moyenne géométrique des prix de la période de référence. Tous les prix du panier sont équipondérés.

aux ajustements effectués par les consommateurs sur leur panier, avait déjà été présenté comme une solution prometteuse (Lequiller, 1997).

❶ RÉPONDRE À DES POLÉMIQUES RÉCURRENTES AUTOUR DE L'IPC

Les données de caisses permettent également d'apporter des éléments de réponse au débat sur une sous-estimation possible de l'inflation par l'IPC : après le passage à l'euro, l'écart entre l'inflation ressentie par les ménages et l'inflation mesurée par l'Insee s'est accru (Leclair et Passeron, 2017). Une des explications de cet écart est que les ménages ont tendance à ne pas considérer certaines améliorations de la qualité des produits qu'ils consomment, alors que l'IPC les neutralise : une amélioration de la qualité d'un produit à prix inchangé se traduit dans l'IPC comme une baisse de prix. Or, si la norme sociale est modifiée vers des produits de meilleure qualité, et en conséquence plus onéreux, il est probable que le consommateur se sentira contraint à cette dépense de qualité supérieure et ressentira ce déplacement de la norme sociale comme une hausse du coût de la vie.

Une réponse possible était de construire des statistiques plus proches du ressenti, tout en continuant à produire l'IPC, qui est pertinent pour mesurer, notamment, la croissance du PIB en volume et le pouvoir d'achat des ménages. L'idée est de calculer des prix moyens d'ensemble de produits, qui prennent en compte les évolutions des habitudes de consommation et ne neutralisent pas ces changements de qualité (Moati et Rochefort, 2008). Par exemple, le développement de la consommation de riz de type basmati ou thaï, en général plus cher que du riz ordinaire, peut être considéré comme un changement de qualité du produit et ne se traduira pas par une augmentation de l'indice des prix du riz dans l'IPC. En revanche, le prix moyen du riz (toutes catégories confondues) augmentera du fait du poids croissant de ces riz parfumés dans les achats.

Les données de caisses permettent de calculer de tels prix moyens en donnant une information précise sur les quantités consommées, en plus des prix pratiqués.

❷ UN MEILLEUR SUIVI DES PRIX EFFECTIVEMENT PAYÉS PAR LE CONSOMMATEUR

De manière peut-être un peu surprenante, un autre avantage des données de caisses est qu'elles permettent de mieux suivre le concept de prix que l'on souhaite mesurer avec l'IPC, par rapport à une enquête spécifique classique.

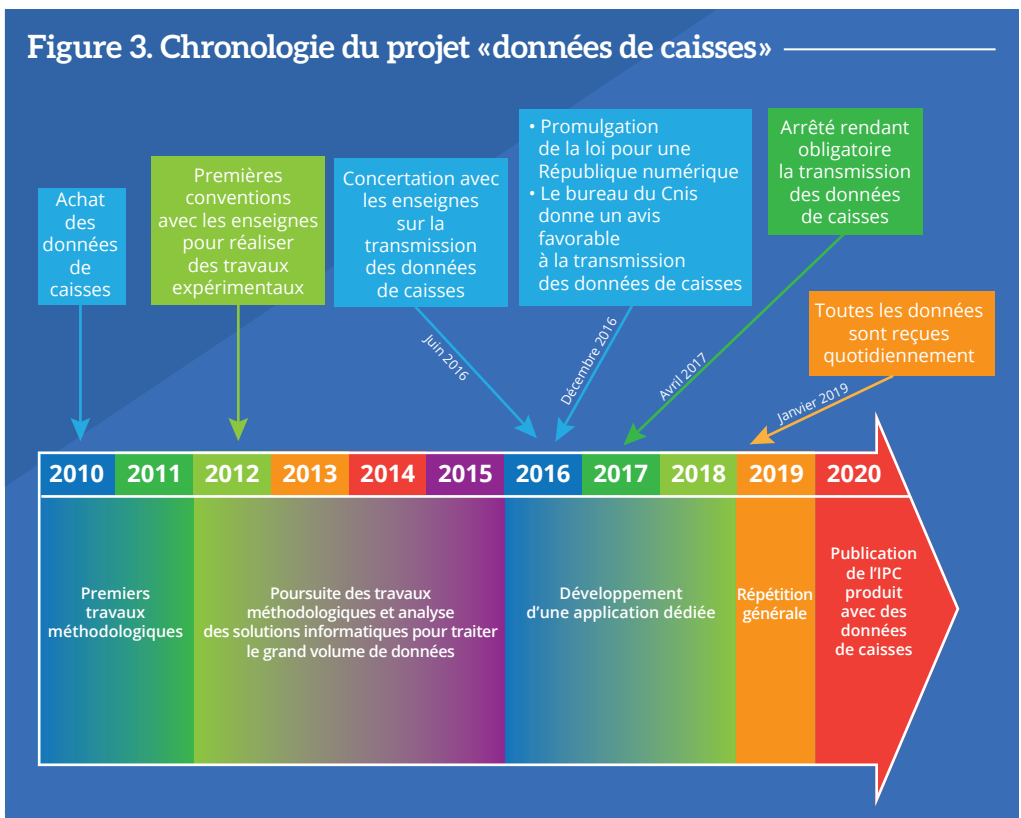
En effet, les enquêteurs envoyés par l'Insee sur le terrain ne peuvent relever que les prix affichés dans les points de vente. Or, ceux-ci peuvent diverger des prix réellement payés par le consommateur : ces différences s'expliquent en partie par des erreurs d'affichage, mais surtout par la mise en œuvre de certaines promotions. Actuellement, l'IPC ne mesure les promotions que lorsqu'elles s'appliquent à l'ensemble des consommateurs. Cette méthode, conforme à la réglementation européenne sur les indices de prix à la consommation, est la conséquence d'un manque d'informations sur le nombre d'acheteurs bénéficiant effectivement des promotions « discriminantes », par exemple liées au fait d'être porteur d'une carte du magasin.

Dans les données de caisses, ce sont les prix effectivement payés qui sont enregistrés. Elles intègrent ainsi de nombreuses promotions, même si certaines pratiques commerciales leur échappent encore, comme le « cagnottage » qui consiste à donner des points en vue d'une consommation future, en contrepartie de l'achat d'un produit spécifique.

LE LONG CHEMIN POUR ACCÉDER AUX DONNÉES PRIVÉES

Si l'apport des données de caisses pour la production de statistiques de prix est incontestable, leur utilisation par l'Insee n'est pas sans poser un certain nombre de difficultés.

La première, dans l'ordre chronologique, est tout simplement de pouvoir y accéder : les données de caisses sont des actifs incorporels des entreprises qui les produisent, et celles-ci n'avaient donc pas d'obligation à en donner un accès à l'Insee, même à des fins d'intérêt général pour la production de statistiques publiques. Dans un premier temps, des contacts ont été pris avec certaines enseignes afin de les convaincre de les transmettre à l'Insee. Depuis 2012, quatre enseignes (40 % environ du marché de la grande distribution) fournissaient ainsi les données à titre expérimental et dans le cadre de conventions (figure 3). Pour d'une part obtenir les données sur l'ensemble du champ des super et hypermarchés, et d'autre part pérenniser la mise à disposition, la loi de 1951 sur l'obligation, la coordination et le secret en matière statistique a été amendée. La loi⁷ prévoit désormais la possibilité de rendre obligatoire la transmission de certaines données privées, après concertation des acteurs et uniquement pour remplacer des enquêtes statistiques obligatoires. Cet amendement, outre l'accès aux données de caisses, pourrait permettre de faciliter, à terme, l'accès à d'autres données privées.



7. Article 19 de la loi du 7 octobre 2016 pour une République numérique.

Après une concertation avec les enseignes de la grande distribution en juin 2016, une étude préalable de faisabilité et d'opportunité de l'utilisation des données de caisses pour l'IPC a été présentée fin 2016 au Conseil national de l'information statistique (Cnis). Après avoir reçu un avis favorable, un arrêté⁸ a été signé par le ministre le 13 avril 2017, rendant obligatoire la transmission des données de caisses par les commerces de détail en magasin non spécialisé à prédominance alimentaire de plus de 400 m². Depuis janvier 2019, l'ensemble des données de caisses de la grande distribution alimentaire, hors hard discount, sont ainsi reçues quotidiennement par l'Insee.

❶ DES STATISTIQUES À PARTIR DE DONNÉES PRODUITES À D'AUTRES FINS ?

Une deuxième difficulté dans l'utilisation des données de caisses est qu'elles n'ont pas été produites à des fins de production statistique. Cela pose deux questions : les statistiques produites à partir de ces données le sont-elles en toute indépendance et impartialité ? Les informations collectées sont-elles en adéquation avec l'objectif statistique recherché ?

Concernant le premier point, l'importance de l'IPC dans le débat public et l'existence jusqu'aux années soixante-dix de polémiques sur la manipulation de l'indice (Jany-Catrice, 2018) posent la question de la confiance que l'on peut avoir dans des données produites par des acteurs privés. Bien qu'il paraisse difficile que les enseignes puissent manipuler des données si volumineuses, l'Insee a souhaité garantir la qualité des données utilisées pour le calcul de l'IPC en organisant des enquêtes de contrôle. À l'avenir, chaque mois, un certain nombre de prix enregistrés dans les données de caisses seront contrôlés par des enquêteurs

dans les points de vente. Dès 2019, une double collecte des prix, par les enquêteurs d'une part pour le calcul de l'IPC, dans les données de caisses d'autre part, permet de s'assurer de l'absence de divergence.

« Bien qu'il paraisse difficile que les enseignes puissent manipuler des données si volumineuses, l'Insee a souhaité garantir la qualité des données utilisées pour le calcul de l'IPC en organisant des enquêtes de contrôle. »

La seconde question porte sur le fait que ces données n'ont pas été produites initialement pour l'établissement de statistiques. Comme pour les données administratives (Rivière,

2018), il est possible que les données recueillies, les définitions retenues, le champ, ne correspondent pas exactement à ce que le statisticien souhaite mesurer. Ces faiblesses sont encore plus marquées pour le *big data* (Blanchet et Givord, 2017), caractérisé, non seulement par leur « Volume » et par la « Vitesse » de leur mise à disposition, mais également par un troisième « V », la Variété, soulignant l'aspect bien souvent peu structuré de ces données.

Dans ce paysage, les données de caisses semblent un cas un peu à part. Car pour commencer, l'information recueillie peut se révéler plus pertinente que dans le cas d'une enquête dédiée à la collecte des prix : le concept de prix est par exemple mieux suivi dans les données de caisses que par l'enquêteur qui ne peut collecter que le prix affiché ; l'information sur les quantités consommées est difficilement mesurable par une enquête, en tout cas, pas avec le

8. Arrêté du 13 avril 2017 rendant obligatoire la transmission des données par voie électronique à des fins de statistiques publiques.

degré de finesse nécessaire ; le champ des données de caisses peut être aisément complété par des données d'enquêtes (dans les autres formes de vente, pour les produits frais vendus en super et hypermarché, par exemple) pour couvrir l'ensemble de la consommation des ménages ; il couvre même dans certains cas mieux la consommation des ménages, en intégrant les données sur le *drive* par exemple, non couvert par l'IPC jusqu'à présent.

Par ailleurs, les données de caisses se distinguent au sein du *big data* parce qu'il s'agit en réalité de données très structurées, qui n'ont du *big data* que les deux premiers « V » du volume et de la vélocité.

Au total, alors que d'autres pistes d'utilisation du *big data* visent plutôt à produire de nouvelles statistiques, complémentaires mais non substituts de statistiques publiques existantes, les données de caisses ont cela de spécifique qu'elles peuvent réellement se substituer à des données d'enquête sans avoir à modifier les concepts ou le cadre méthodologique de ce que l'on veut mesurer.

1,7 MILLIARD D'ENREGISTREMENTS : NOUVELLE ARCHITECTURE INFORMATIQUE...

Si le choix de l'Insee est donc de considérer que les données de caisses peuvent remplacer les relevés effectués par les enquêteurs sans que cela nécessite d'adaptation des concepts utilisés pour l'IPC, la volumétrie des données à traiter nécessite un certain nombre de solutions, informatiques tout d'abord mais également statistiques ensuite, pour automatiser des traitements préalablement effectués manuellement.

L'Insee réceptionne chaque mois 1,7 milliard d'enregistrements, qui correspondent aux lignes dans les données de caisses, c'est-à-dire les ventes associées dans un point de vente pour un code-barres donné et un jour donné (*figures 1 et 2*). Autant d'enregistrements⁹ à traiter mensuellement n'auraient pu être gérés par des bases de données classiques, dites relationnelles. Des technologies adaptées au *big data* (en l'occurrence le système Hadoop) ont été retenues : elles permettent de répartir les données et les traitements sur plusieurs serveurs afin d'améliorer les performances des traitements et de rendre le système robuste à la panne d'un ou plusieurs des serveurs.

... ET NÉCESSAIRE AUTOMATISATION DES PROCESSUS STATISTIQUES

D'un point de vue statistique, le volume des données ne permet plus les traitements manuels qui pouvaient être effectués, en général par les enquêteurs, et qui doivent donc être automatisés. Trois exemples peuvent être donnés : être capable de classer un produit dans une nomenclature détaillée, identifier les relances commerciales et remplacer un produit lorsqu'il disparaît.

- ① Dans les données de caisses, **les produits sont identifiés par leur code-barres** (*figure 1*) et celui-ci ne donne pas d'information sur la nature du produit suivi. Compte tenu du nombre de codes-barres présents dans les données de caisses (près de 9 millions), il est impensable de rechercher manuellement à quel produit correspond chacun d'eux. Pour résoudre ce problème, l'Insee achète à un panéliste un dictionnaire de codes-barres, décrivant très précisément les caractéristiques du produit associé à chaque code-barres. **Classer les produits** revient alors à construire une simple table de passage entre ce dictionnaire et la nomenclature Coicop.

9. Plus précisément, seul 1,3 milliard est effectivement exploité pour l'IPC (du fait de l'exclusion de certains produits). Ces données sont ensuite consolidées par mois et par article de classes équivalentes.

- ① Ce dictionnaire de codes-barres permet également de traiter correctement le cas des « **relances commerciales** ». Ces relances consistent en une modification marginale du *packaging* du produit avec souvent un prix en hausse, ou un prix stable mais avec un volume de produit vendu plus faible. Ces relances commerciales masquent en général une hausse de prix, il est donc fondamental de pouvoir les repérer. Dans une collecte sur le terrain, c'est l'enquêteur qui identifie la relance ; avec les données de caisses, le code-barres change en général avec le *packaging* et il faut être capable, de manière automatique, de relier le produit initial à sa relance. En Suède, Tongur (2019) estime à 0,1 point le biais qui pourrait s'ensuivre si la relance n'était pas identifiée dans les données de caisses. Dans le cas français, l'existence d'un dictionnaire de codes-barres permet de faire le lien nécessaire entre un produit et sa relance commerciale et de bien enregistrer la hausse de prix associée.
- ① Le **remplacement des produits** appartenant au panier de l'IPC et disparaissant en cours d'année est également une opération impliquant fortement les enquêteurs. Ils choisissent le produit remplaçant de manière à être le plus proche possible du produit disparu et décident s'il est nécessaire ou non d'effectuer un ajustement qualité¹⁰. Cette opération stratégique pour l'IPC peut elle aussi être automatisée (Léonard *et alii*, 2017). Le produit remplaçant est tiré aléatoirement parmi les produits de la même variété et un ajustement qualité est réalisé systématiquement, en comparant le prix du produit remplacé et du produit remplaçant au cours d'une même période : la différence de qualité est estimée égale à la différence de prix. Cette méthode d'ajustement de la qualité est couramment utilisée dans l'IPC, mais les prix comparés sont presque toujours imputés car lorsqu'on recourt à des relevés de prix, il n'est en général pas possible de comparer au cours d'une même période le prix du produit disparu et remplaçant. Par définition, on n'a pas anticipé que le produit allait disparaître et on n'a pas relevé le prix du produit remplaçant avant même de savoir que le produit remplacé allait disparaître. Puisque les données de caisses sont exhaustives, on peut y rechercher *a posteriori* le prix passé d'un produit.

① UN CHOIX UN PEU DIFFÉRENT DE NOS PARTENAIRES EUROPÉENS

L'achat d'un dictionnaire de codes-barres¹¹ et le recours à des technologies de *big data* permettent ainsi de conserver les concepts de l'IPC actuel (en particulier l'idée d'un panier fixe annuellement) tout en exploitant les données de caisses dans leur exhaustivité. Cette situation est assez singulière dans le paysage des instituts statistiques recourant aux données de caisses.

Historiquement, et c'est encore la solution retenue par un certain nombre de pays comme la Suède et l'Italie, les pays qui ont utilisé les données de caisses pour calculer l'IPC ont tiré un échantillon de produits de manière à pouvoir réaliser manuellement les trois traitements précédemment décrits : la classification, l'identification des relances, le remplacement des produits. Les concepts de l'IPC sont alors strictement conservés et on utilise les données de caisses comme base de sondage, afin de repérer le plus rapidement possible l'apparition ou la suppression de produits ou pour connaître précisément les poids associés à chaque produit.

10. Pour neutraliser le fait que le produit remplaçant peut être de qualité légèrement différente du produit disparu.

11. Les INS des autres pays européens n'ont pas un tel référentiel ; ils s'appuient en général sur une description (plutôt courte) du produit que l'on peut retrouver sur les tickets de caisses et à l'aide de méthodes de *machine learning* réussissent à classer les codes-barres dans la nomenclature de fonction utilisée pour le calcul de l'IPC.

Les développements suivants ont visé à bénéficier de toute la précision apportée par le caractère exhaustif des données de caisses, et à limiter les traitements manuels liés aux remplacements, qui deviennent très vite importants dès qu'on augmente la taille de l'échantillon. La méthode précédente a été ainsi progressivement remplacée par une exploitation des données de caisses qui renonçait à la fixité annuelle du panier de produits suivis : l'évolution mensuelle des prix a été mesurée sur un panel cylindré de produits présents au cours de deux mois consécutifs ; ces évolutions mensuelles étaient ensuite chaînées les unes aux autres. Ce chaînage d'indices à fréquence trop élevée est connu des statisticiens des prix pour créer des dérives d'indice. Pour éviter cela, il a fallu renoncer, au niveau le plus fin, aux pondérations¹², une innovation pourtant importante des données de caisses.

Une dernière génération d'indices enfin a émergé : les indices multilatéraux, méthode GEKS (Diewert, Fox Ivancic, 2009) ou Geary-Khamis (Chessa 2015), s'inspirent des méthodes utilisées pour la comparaison spatiale des prix et permettent de traiter le fait que les paniers puissent être différents chaque mois. Celles-ci sont néanmoins moins intuitives et plus difficiles à expliquer au grand public.

🎯 AU FINAL, QUEL IMPACT DES DONNÉES DE CAISSES ?

Les données de caisses seront utilisées pour la première fois pour calculer l'IPC publié en janvier 2020. 30 000 prix relevés mensuellement par les enquêteurs seront remplacés par environ 77 millions de produits présents dans les données de caisses. En dehors de ce champ, la collecte actuelle sera conservée.

“ Les données de caisses seront utilisées pour la première fois pour calculer l'IPC publié en janvier 2020. ”

Avant d'utiliser cette nouvelle source de données dans la production d'une statistique aussi importante pour le débat public que l'IPC, il a été nécessaire de faire la preuve de sa fiabilité, de sa pérennité et de son apport à la mesure de l'inflation. Une expérimentation et

des travaux méthodologiques ont permis de définir le traitement des données de caisses et leur intégration dans l'IPC (Leclair *et alii*, 2019). Une application informatique spécifique a été développée, en technologie *big data*, pour s'assurer de la réception des données, des contrôles statistiques et des traitements à réaliser : l'IPC est produit en effet dans des délais très restreints qui nécessitent de s'appuyer sur des traitements informatiques robustes. La mise à disposition des données de caisses a été fiabilisée par un arrêté (voir *supra*), doublé, pour certaines enseignes, de conventions.

Enfin, avant l'utilisation effective des données de caisses dans l'IPC, l'Insee a souhaité réaliser pendant une année une répétition générale : alors que l'IPC publié chaque mois en 2019 s'appuyait sur les relevés de prix des enquêteurs (et des autres sources traditionnelles utilisées par l'IPC), en parallèle, un IPC a été produit en utilisant les données de caisses.

12. Ce sont en effet les pondérations qui créent la dérive de l'indice : le poids dans la consommation dépendant en général du niveau de prix, une période de promotion lie un poids élevé à une baisse de prix alors que le retour au prix normal s'accompagne d'un poids dans la consommation faible. L'indice chaîné mensuellement et utilisant des pondérations ne revient pas ainsi à son niveau d'origine après une période de promotion, car il ne pondère pas de la même manière les hausses et les baisses de prix.

Cette répétition générale permet de roder le processus de production ; elle permet surtout de comparer les résultats. Quel impact alors de cette nouvelle source de données ? L'impact sur l'inflation d'ensemble est peu visible car au final, les données de caisses ne représentent qu'un poids faible de la consommation d'ensemble (11 % environ) : beaucoup de produits ne peuvent pas être suivis par des données de caisses (les services, les produits frais, en l'absence de codes-barres) ou ne le sont pas pour des raisons méthodologiques (l'habillement

« L'impact sur l'inflation d'ensemble est peu visible car au final, les données de caisses ne représentent qu'un poids faible de la consommation d'ensemble (11 % environ). »

et les biens durables, pour lesquels la rotation des produits est importante et la méthodologie utilisée pour les ajustements qualité est spécifique) et enfin, la consommation des ménages se fait aussi dans d'autres formes de vente que les super et hypermarchés.

Mais à un niveau plus fin, pour des postes où les données de caisses sont plus fortement utilisées, des différences peuvent être notées. L'analyse fine des écarts montre qu'elles s'expliquent essentiellement par trois facteurs :

- 1 une meilleure représentativité des produits suivis : la connaissance fine des quantités dans les données de caisses amène à suivre des variétés de produits qui ne l'étaient pas jusqu'à présent faute d'avoir pu identifier leur importance ; or, ces variétés de produits ont des dynamiques de prix propres qui n'étaient pas prises en compte préalablement ;
- 2 une meilleure précision de l'indice : sur les mêmes variétés de produits, des différences d'évolution de prix peuvent exister du fait de l'imprécision de l'échantillonnage,
- 3 un meilleur suivi des prix ; la meilleure intégration des promotions dans les données de caisses (voir *supra*) permet de mettre en avant des évolutions de prix qui ne peuvent l'être dans la collecte traditionnelle.

Les données de caisses sont donc une source prometteuse pour le calcul de statistiques de prix. En janvier 2020, seules les statistiques de prix publiées actuellement (l'IPC, l'indice des prix dans la grande distribution) seront produites. Mais dans un avenir plus lointain, de nouvelles statistiques pourront être diffusées : des prix moyens pour de nombreux produits, des comparaisons spatiales de prix ou des indices régionaux. Des études méthodologiques se poursuivront pour exploiter les données de caisses sur des champs pour lesquels elles ne sont pas encore mobilisées, l'habillement ou les biens durables par exemple. Des travaux seront menés pour accéder à de nouvelles données de caisses, celles des *hard-discounters* ou de la grande distribution spécialisée.

■ BIBLIOGRAPHIE

BLANCHET, Didier et GIVORD, Pauline, 2017. Données massives, statistique publique et mesure de l'économie. In : *L'Économie française, édition 2017*. [en ligne]. Insee Références, pp. 59-77. [Consulté le 7 octobre 2019]. Disponible à l'adresse :

https://www.insee.fr/fr/statistiques/fichier/2894010/ECOFRA17b_D1_big-data.pdf

BOSKIN, Michael J., 1996. Toward a More Accurate Measure of the Cost of Living : Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. In : *site de l'administration de la Sécurité Sociale des États-Unis*. [en ligne]. [Consulté le 7 octobre 2019]. Disponible à l'adresse : <http://www.ssa.gov/history/reports/boskinrpt.html>

CHESSA, Antonio, 2015. Towards a generic price index method for scanner data in the Dutch CPI. In : *Fourteenth Meeting of the Ottawa Group (International Working Group On Price Indices)*. [en ligne]. 20-22 mai 2015. Tokyo, Japon. [Consulté le 7 octobre 2019]. Disponible à l'adresse : <https://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s1room2.pdf>

CLÉ, Émeline, JALUZOT, Laurence, MALAVAL, Fabien, RATEAU, Guillaume, SAUVADET, Luc, 2016. *En 2015, les prix en région parisienne dépassent de 9 % ceux de la province* [en ligne]. 14 avril 2016. Insee Première, n°1590. [Consulté le 7 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/version-html/1908158/ip1590.pdf>

DIEWERT, Erwin, FOX, Kevin J., IVANCIC, Lorraine, 2009. Scanner Data, Time Aggregation and the Construction of Price Indexes. In : *Eleventh Meeting of the Ottawa Group (International Working Group On Price Indices)*. [en ligne]. 27-29 mai 2009. Neuchâtel, Suisse. [Consulté le 8 octobre 2019]. Disponible à l'adresse :

<http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/a49bc2a164b232c4ca2576a100773522?OpenDocument>

EUROSTAT, 2017. *Practical Guide for Processing Supermarket Scanner Data*. [en ligne]. Septembre 2017. [Consulté le 8 octobre 2019]. Disponible à l'adresse :

<https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf>

FMI, 2004. *Manuel des prix à la consommation. Théorie et pratique*. OIT/FMI/OCDE/CEE-ONU/Eurostat/Banque mondiale. Genève, Organisation internationale du travail. ISBN 1-58906-330-9

JALUZOT, Laurence et SILLARD, Patrick, 2016. *Échantillonnage des agglomérations de l'IPC pour la base 2015*. [en ligne]. Janvier 2016. Insee, Document de travail, N°F1601. [Consulté le 8 octobre 2019]. Disponible à l'adresse :

<https://www.insee.fr/fr/statistiques/fichier/2022137/F1601.pdf>

JANY-CATRICE, Florence, 2019. *L'indice des prix à la consommation*. Édition La Découverte. Collection Repères, N°717. Janvier 2019. ISBN 978-2-7071-9931-7

LECLAIR, Marie, LÉONARD, Isabelle, RATEAU, Guillaume, SILLARD, Patrick, VARLET, Gaëtan et VERNÉDAL, Pierre, 2019. Les données de caisses : avancées méthodologiques et nouveaux enjeux pour le calcul d'un indice des prix à la consommation. In : *Économie et statistique*. [en ligne] 17 septembre 2019. N°509, pp. 13-31. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/4203515/509_Leclair-Leonard-Rateau-Sillard-Varlet-Vernedal-FR.pdf


- LECLAIR, Marie, et PASSERON, Vladimir, 2017. *Une inflation modérée depuis le passage à l'euro*. [en ligne]. 24 mai 2017. Insee Focus, N°87. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/2854085>
- LÉONARD, Isabelle, SILLARD, Patrick, VARLET, Gaëtan, ZOYEM, Jean-Paul, 2017. *Scanner data and quality adjustment*. [en ligne]. Juin 2017. Insee, Document de travail, N°F1704. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/2912650/F1704.pdf>
- LÉONARD, Isabelle, SILLARD, Patrick, VARLET, Gaëtan, ZOYEM, Jean-Paul, 2019. Écarts spatiaux de niveaux de prix entre régions et villes françaises avec des données de caisses. In : *Économie et statistique*. [en ligne]. 17 septembre 2019. N°509, pp. 73-87. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/4203527/509_Leonard-Sillard-Varlet-Zoyem-FR.pdf
- LEQUILLER, François, 1997. L'indice des prix à la consommation surestime-t-il l'inflation ? In : *Économie et statistique*. [en ligne]. Mars 1997. N° 303, pp. 3-32. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.epsilon.insee.fr/jspui/bitstream/1/21417/1/estat_1997_303_1.pdf
- MOATI, Philippe et ROCHEFORT, Robert, 2008. *Mesurer le pouvoir d'achat*. [en ligne]. Janvier 2008. Édition La Documentation française, Collection Les Rapports du Conseil d'analyse économique. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://www.vie-publique.fr/sites/default/files/rapport/pdf/084000050.pdf>
- RIVIÈRE, Pascal, 2018. Utiliser les déclarations administratives à des fins statistiques. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. N°N1, pp. 14-24. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/3647013/courstat-1-5.pdf>
- SILLARD, Patrick, 2017. *Indices de prix à la consommation*. [en ligne]. 7 août 2017. Insee, Document de travail, N°F1706. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/version-html/2964204/F1706.pdf>
- TASSI, Philippe, 2018. Les apports des Big Data. In : *Économie et statistique*. [en ligne]. N°505-506, pp. 5-15. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/3705956/505-506_Tassi-FR.pdf
- TONGUR, Can, 2019. Inflation Measurement with Scanner Data and an Ever-Changing Fixed Basket. In : *Économie et statistique*. [en ligne]. 17 septembre 2019. N°509, pp. 33-50. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/4203519/509_Tongur-FR.pdf
- UNECE, 2018. *Report of the Group of Experts on Consumer Price Indices*. [en ligne]. 7-9 mai 2018. Fourteenth session, Genève, Suisse. ECE/CES/GE.22/2018/2. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Report_of_CPI_expert_group_meeting_7-9_May_2018.pdf

LE CENTRE D'ACCÈS SÉCURISÉ AUX DONNÉES (CASD)

UN SERVICE POUR LA DATA SCIENCE ET LA RECHERCHE SCIENTIFIQUE

Kamel Gadouche*

L'accès des chercheurs aux données individuelles collectées par le Service statistique public constitue un enjeu scientifique majeur. Ces informations très détaillées exigent un très haut niveau de sécurité pour éviter toute dissémination préjudiciable au citoyen, ou toute utilisation par un tiers non autorisé. Pour répondre à ce besoin de sécurité, l'Insee a créé en 2010 le CASD, Centre d'accès sécurisé aux données : les équipes du centre ont conçu un équipement sécurisé, permettant un accès à distance, tout en garantissant une authentification forte des utilisateurs et un confinement des fichiers. Le CASD, désormais autonome, s'est développé au fil du temps, élargissant son champ à d'autres producteurs et à d'autres types de données très détaillées et sensibles telles les données de santé ou les données administratives. Ce service propose de nouvelles solutions à la problématique des appariements et de la reproductibilité des études sur données confidentielles. Le CASD est de plus en plus utilisé par la communauté de la recherche en France, et son expérience originale, bien que relativement récente comparé à celle de ses partenaires étrangers, lui permet de s'ouvrir à l'international.

 *Giving researchers access to individual data collected by the Official Statistical System constitutes a major scientific challenge. This very detailed information requires a very high level of security in order to avoid any disclosure which would be prejudicial to the citizen, or any use by an unauthorised third party. To meet this security requirement, INSEE created in 2010 the Secure Data Access Centre (Centre d'accès sécurisé aux données, or CASD) whose teams have designed a secure device, allowing remote access whilst ensuring strong user authentication and confinement of the files. The CASD, now autonomous, has developed over time, extending its perimeter to other data producers and other types of highly detailed, sensitive data such as health data or administrative data. This service provides new solutions to the issue of record linkage and reproducibility of research work based on confidential data. The research community in France uses increasingly CASD's services. The experience is original, although relatively recent compared with that of foreign partners : it enables now CASD to expand on an international level.*

* Directeur du CASD,
kamel.gadouche@casd.eu

Le premier ouvrage entièrement consacré au secret statistique est paru cette année (Le Gléau, 2019). Il présente un panorama complet pour la France, avec quelques aperçus pour des pays étrangers, des mesures mises en place pour encadrer l'utilisation des données collectées pour établir des statistiques. L'accent est particulièrement mis sur la dualité entre la nécessité de collecter des données individuelles et le besoin de sécurité pour en assurer la confidentialité, que ce soit du point de vue juridique ou technique. Pour les statisticiens du Service statistique public, les garanties sont apportées par leur statut et par leur intégration au sein de l'Insee ou des ministères. Pour les chercheurs, dont les besoins en données très détaillées de la statistique publique sont toujours croissants, se pose la question de l'encadrement juridique de l'accès mais aussi la question des garanties techniques de sécurité pour préserver la confidentialité des données. Nous verrons comment le Centre d'accès sécurisé aux données (CASD) apporte une réponse à ces questions et favorise ainsi le développement de l'accès aux données pour la recherche scientifique. Nous verrons également comment les relations avec les producteurs comme avec les chercheurs se sont développées au fil du temps et ont permis la mise en place de nouveaux usages liés aux nouvelles technologies de *data science*, d'appariement ou encore de certification des résultats.

📍 DONNÉES INDIVIDUELLES ET CONFIDENTIALITÉ

Un grand nombre de données individuelles sur les personnes ou les entreprises sont aujourd'hui collectées par l'Insee et les services statistiques ministériels à des fins de statistique publique, par les administrations dans l'exercice de leurs missions, par les entreprises pour leurs besoins de gestion, par des universités à des fins de recherche dans différents domaines, comme celui de la santé. S'y ajoutent de façon croissante, les données individuelles liées à l'utilisation des moyens électroniques (paiement par carte de crédit...) et donc collectées automatiquement. Toutes ces informations couvrent un grand nombre de domaines particulièrement intéressants pour la recherche : revenus, patrimoine, santé, données comptables des entreprises, informations de localisation géographique, parcours scolaires, trajectoires professionnelles, etc.

Même si elles ne sont pas toutes directement identifiantes (nom, ou identifiants tels que le numéro de sécurité sociale, adresse), un très grand nombre de ces données le sont indirectement du fait de leur précision. Certaines, sensibles au sens de la loi, font peser de par leur nombre un risque d'autant plus grand en cas d'identification pour les personnes ou les entreprises concernées. Pour les entreprises, peu d'informations suffisent le plus souvent à les identifier.

📍 LES BESOINS SPÉCIFIQUES DES CHERCHEURS

Différents secrets inscrits dans des règlements ou des lois couvrent ainsi les données selon les domaines : le secret fiscal, le secret médical, le secret pénal, le secret des affaires, le secret professionnel plus généralement. Dans le cas de la statistique publique, le législateur a inscrit dans la loi l'obligation de confidentialité pour les statisticiens. On appelle cette obligation de confidentialité « secret statistique », une version spécialisée du secret professionnel.

Ces différentes dispositions n'ont généralement pas pris en compte initialement la finalité de recherche. Elles ont été progressivement modifiées pour l'intégrer afin de permettre aux chercheurs d'utiliser ces données très riches pour leurs analyses quantitatives.

Les premières avancées en matière d'accès des chercheurs à des données anonymisées de la statistique publique avec le Réseau Quetelet¹ ont en effet montré très vite que ces fichiers moins détaillés, directement transmis aux chercheurs, s'ils marquaient un progrès important, ne permettaient cependant pas de répondre aux besoins de nombre de projets de recherche. La préoccupation grandissante de la Cnil² en matière de données personnelles a conduit à davantage restreindre le niveau de détail de ces données, au point parfois de rendre impossible certains travaux en démographie ou en sociologie urbaine par exemple. Parallèlement, de nouvelles méthodes statistiques, associées à des moyens de calcul plus performants, requéraient notamment chez les économistes des données très détaillées à un moment où il devenait possible de mobiliser pour l'analyse de plus en plus de données administratives utiles à l'évaluation des politiques publiques.

Pour répondre à ces besoins, la loi sur le secret statistique, dite loi de 1951³, déjà plusieurs fois modifiée, notamment pour permettre l'utilisation par les chercheurs des données d'entreprises, a pris en compte en 2008 la finalité de recherche pour les données des personnes et des ménages. Une modification en ce sens pour l'ensemble des données personnelles avait été réalisée dès 2004 dans la loi Informatique et Libertés⁴. Des modifications d'autres dispositions ont également suivi, notamment dans le domaine des données fiscales.

Encadré 1. Quelques repères

Le CASD accueille aujourd'hui environ 500 projets de recherche menés en France (Amiens, Lyon, Marseille, Dijon, Paris...) et à l'étranger (Royaume-Uni, Allemagne, Pays-Bas, Pologne, Espagne, Italie...), ce qui représente environ 350 sites déployés. Ce sont en tout près de 1 500 utilisateurs qui s'appuient sur le CASD pour l'accès aux données confidentielles.

Juin 1999 – Rapport de Roxane Silberman sur l'accès aux données pour la recherche.

Octobre 2007 – L'Insee initie un projet de pilote de CASD à la suite de la revue par les pairs de la mise en œuvre du code de bonnes pratiques de la statistique européenne.

Juillet 2008 – Modification de la loi 51-711 pour permettre l'accès des chercheurs aux données sur les individus et les ménages.

Octobre 2009 – Premier Comité du secret statistique concernant les données sur les individus et les ménages – Annonce de la mise en place de la facturation à l'usage pour couvrir les coûts.

Février 2010 – Mise en production du CASD pour 30 projets.

Janvier 2011 – Le CASD est lauréat de l'appel à projet Equipex (Équipement d'Excellence du programme Investissements d'avenir) et obtient un financement de 4 M€ pour se développer.

Mars 2012 – Création de l'entité CASD au sein du Genes.

Septembre 2014 – Mise à disposition des données fiscales.

Octobre 2016 – Loi pour une République numérique dont l'article 36 étend les compétences du Comité du secret statistique aux données administratives.

Décembre 2018 – Création du Groupe d'Intérêt Public CASD.

1. Réseau français des centres de données pour les sciences sociales, « Quetelet Progedo Diffusion » lui a succédé pour diffuser les données françaises en sciences humaines et sociales auprès de la communauté de recherche.
2. Commission nationale de l'informatique et des libertés.
3. Loi n° 51-711 du 7 juin 1951 modifiée sur l'Obligation, la coordination et le secret en matière de statistiques.
4. Loi n° 78-17 du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux libertés.

Les aménagements juridiques ne suffisent pas à eux seuls à garantir la confidentialité des données. L'accès effectif doit s'accompagner de mesures de sécurité appropriées permettant d'apporter des garanties techniques additionnelles. Ces mesures requièrent un niveau d'exigence élevé, d'emblée plus facile à appliquer au sein de la statistique publique de par sa fonction qu'à l'extérieur de celle-ci. D'où des difficultés qu'il fallait prendre en compte pour garantir l'application de telles mesures dans des établissements, comme les universités et les centres de recherche, avant de leur transmettre les données. C'est dans le but d'étendre le dispositif de sécurité à l'extérieur du Service statistique public, et d'apporter ainsi une réponse au besoin d'accès aux données des chercheurs, que l'Insee et le Genes⁵ ont créé en 2010 le Centre d'accès sécurisé aux données (CASD), initialement centré sur l'accès aux données issues de la statistique publique (voir (Le Gléau et Royer, 2011) et **encadré 1**).

🔒 CONFIDENTIALITÉ VS OUVERTURE : PREMIÈRES PISTES

Afin de résoudre le conflit entre confidentialité des données et volonté d'utilisation plus large de ces données par les chercheurs, dès les années quatre-vingt-dix, certains pays comme les États-Unis, le Canada, la Grande-Bretagne ou l'Allemagne ont mis en place des centres d'accès sécurisé sous forme de locaux isolés : les utilisateurs doivent s'y rendre *physiquement* pour y travailler avec des contrôles très stricts à l'entrée comme à la sortie des locaux. Toute sortie de résultats ne peut notamment être récupérée qu'après une vérification effectuée par des opérateurs pour garantir le respect du secret statistique. Même si l'accès aux données était possible, il était cependant très contraignant pour les chercheurs de devoir se déplacer, parfois loin, pour accéder aux données.

Pour pallier les inconvénients de tels dispositifs, des expérimentations ont commencé à la fin des années quatre-vingt-dix pour développer des systèmes qui permettraient un accès à la fois distant et sécurisé. Dès 1999, un rapport rendu au ministre en charge de l'Enseignement supérieur et de la recherche (Silberman, 1999) soulignait la nécessité pour les chercheurs d'utiliser des données très détaillées. Il indiquait l'existence de telles expérimentations à l'étranger, notamment à Statistique Québec. À partir des années deux-mille, on trouvait déjà ce type de système aux États-Unis (le *NORC⁶ Data Enclave* à Chicago) et dans plusieurs pays européens. Le cas du Danemark était souvent cité par des chercheurs du Crest⁷ au moment où la question de création d'un tel dispositif commençait à être discutée en France.

Ces systèmes différaient dans leurs implémentations techniques et en raison des différentes législations nationales de protection des données, mais leurs caractéristiques restaient assez similaires. En particulier, ils s'appuyaient sur des logiciels dédiés d'accès distant (*remote access* en anglais). De tels systèmes nécessitent l'installation de logiciels sur des postes de travail non maîtrisés. Ainsi, ils n'apportent pas les garanties de sécurité suffisantes et sont complexes à mettre en œuvre : ils génèrent souvent des problèmes de compatibilité, de conflits d'installation ou de maintenance.

5. Le Groupe des Écoles Nationales d'Économie et Statistique (Genes) est un établissement public d'enseignement supérieur et de recherche rattaché au ministère de l'économie et des finances, dont l'Insee assure la tutelle technique.

6. NORC : *National Opinion Research Center* (Lane et Shipp, 2007).

7. Le Centre de recherche en économie et statistique (Crest) est un centre de recherche dépendant du Genes.

UNE TECHNOLOGIE POUR OUVRIR LE CHAMP DES POSSIBLES

Ces contraintes techniques ont conduit l'Insee et le Genes, poussés également en ce sens par les moyens restreints dont ils disposaient initialement, à concevoir un équipement spécifique, permettant de répondre au besoin d'accès tout en évitant les inconvénients précédemment cités. C'est ainsi que la France a développé son propre système d'accès à distance pour permettre aux chercheurs d'accéder et d'exploiter les données confidentielles principalement issues de la statistique publique.

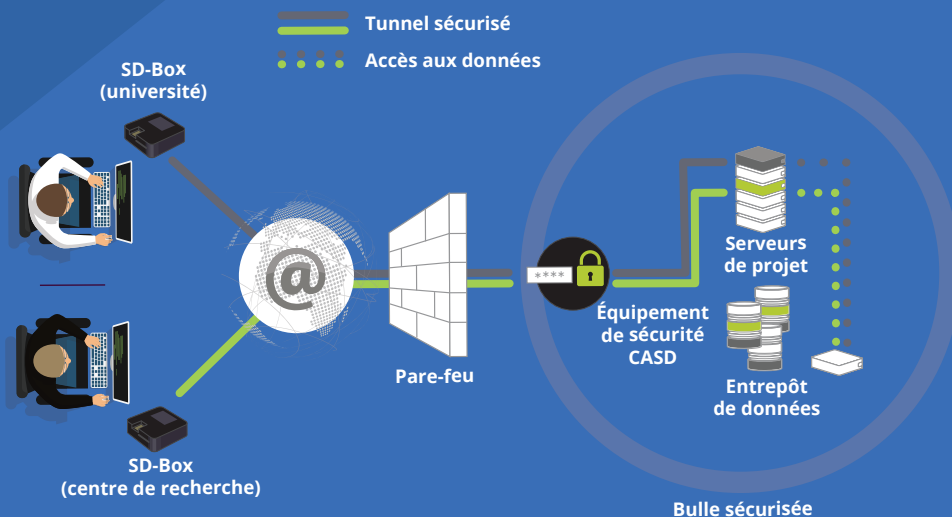
Au lieu d'utiliser des solutions logicielles tierces, l'équipe projet du CASD a conçu un boîtier informatique spécifiquement pour cet usage particulier d'accès distant sécurisé à des données confidentielles : **la SD-Box**. Une fois en possession d'une SD-Box, l'utilisateur n'a plus qu'à se connecter pour avoir accès à distance à des moyens de traitement sur les données confidentielles confinées au sein de locaux techniques sécurisés. Cet endroit de stockage et de traitement des données est appelé **bulle sécurisée** (figure 1). Le principe de cette bulle est qu'aucune donnée ne peut en sortir sans une procédure de contrôle adaptée. L'authentification de l'utilisateur est réalisée à l'aide d'un dispositif s'appuyant sur une carte à puce, contenant un certificat de sécurité, et sur un lecteur biométrique d'empreintes digitales. Conformément à la loi, ce traitement a fait l'objet d'une autorisation

Figure 1. La SD-Box et la bulle sécurisée : un dispositif breveté, conçu par l'Insee et le Genes



Le dispositif technologique du CASD a fait l'objet d'un brevet d'invention déposé par l'Insee le 12 octobre 2009 auprès de l'Institut national de la propriété industrielle (Inpi) sous le numéro FR0957127 (Système informatique d'accès à des données confidentielles par au moins un boîtier distant) et d'une demande d'extensions internationales déposée le 11 octobre 2010, sous le n° WO2011045516A1.

Il a été transféré par l'Insee au Genes à sa création en 2011 et fait désormais l'objet d'une licence exclusive accordée au CASD.



de la Cnil. Le système de bulle sécurisée crée une isolation totale du boîtier, le tout fonctionnant en circuit fermé, sans contact avec l'extérieur : cela permet de garantir une sécurité élevée de bout en bout.

🔒 UNE SÉCURISATION EXTRÊMEMENT POUSSÉE ET CERTIFIÉE...

La technologie développée présente l'avantage de pouvoir réaliser une certification de sécurité ISO 27 001, référence internationale en la matière⁸, extrêmement poussée parce que chaque composant de la chaîne d'accès est entièrement maîtrisé. Par exemple pour établir une connexion, il faut être localisé dans un établissement lié contractuellement avec le CASD, disposer d'une SD-Box à jour et authentifiée, mais également disposer d'une carte à puce biométrique et d'un compte utilisateur valide. Le pilotage des risques peut être réalisé de manière efficace grâce à un dispositif technique et organisationnel entièrement contrôlé. Ce modèle d'architecture unifiée a été validé par plusieurs audits de sécurité réalisés par des sociétés spécialisées qui ont tous souligné le très haut niveau de sécurité du dispositif.

In fine, le fait d'avoir une technologie qui assure l'authentification, le confinement et la traçabilité des données apporte des garanties essentielles pour la diffusion sécurisée de données confidentielles. Le confinement est un prérequis technique primordial pour garantir la traçabilité des données : une fois les données à l'air libre, elles peuvent être copiées sans limite, à un coût marginal quasi nul. Il devient dès lors impossible de les tracer.

🔒 ... QUI N'ENTRAVE PAS LES USAGES DES CHERCHEURS

Le confinement des données ne doit pas créer des conditions d'utilisation si restrictives qu'elles compliqueraient considérablement, voire empêcheraient, la réalisation de certains travaux. Les chercheurs doivent pouvoir disposer de tous les outils nécessaires ainsi que d'une puissance de calcul adaptée. Ce dernier point a été une véritable préoccupation du CASD dès la conception de l'architecture et continue de l'être dans sa gestion courante. Contrairement à un service classique de mise à disposition d'environnement de calcul, comme on peut en voir par exemple sur le *cloud*, l'exigence de confinement du CASD n'offre pas la possibilité à l'utilisateur d'installer par lui-même des logiciels. C'est une contrainte forte pour les chercheurs et c'est pour cela qu'elle doit être compensée par une offre large de logiciels scientifiques mis à leur disposition et la possibilité d'en ajouter si nécessaire et dans des délais assez courts. Il en est de même pour la puissance de calcul qui doit être paramétrable en fonction des besoins, du type de traitement et du volume des données.

À la sécurité technique, s'ajoute une sécurité juridique. Avant de pouvoir accéder à l'infrastructure technique, les chercheurs doivent procéder à un ensemble de démarches qui engagent par ailleurs leur responsabilité personnelle.

🔒 ACCÈS AUX DONNÉES : PROCÉDURE ADMINISTRATIVE...

Lorsqu'il s'agit de données issues de la statistique publique ou de données fiscales, un projet de recherche doit d'abord être soumis au Comité du secret statistique⁹ pour obtenir la levée du secret statistique ou fiscal pour les membres de ce projet. Pour les autres

8. La norme ISO 27 001 permet aux entreprises et aux administrations d'obtenir une certification qui atteste de la mise en place effective d'un système de management de la sécurité de l'information.

9. <https://www.comite-du-secret.fr>

« Vérifier que le projet soumis peut être qualifié de projet de recherche scientifique et que les porteurs du projet sont des chercheurs. »

données administratives, le comité peut également être saisi par le producteur de ces données.

À cet égard, il faut rappeler que le statut de chercheur couvre certes des institutions variées (universités, instituts, etc.), mais qu'à la différence de celui de statisticien public, il n'est pas défini par les textes (loi ou décrets). C'est pour cette raison que l'Insee avait décidé d'instaurer

le Comité du secret statistique et de lui confier la tâche de vérifier que le projet soumis peut être qualifié de projet de recherche scientifique et que les porteurs du projet sont des chercheurs.

Ce comité, où siègent notamment les producteurs des données et des représentants des chercheurs, examine un ensemble de critères défini dans les textes, dont la finalité de la recherche proposée, la pertinence des données pour lesquelles l'accès a été demandé et la qualification de chercheur du demandeur.

Après cette instruction, le comité émet un avis, lequel est suivi d'une décision de l'administration des Archives nationales, ou du ministre du Budget s'il s'agit de données fiscales. La Cnil intervient également s'il s'agit de données personnelles. *Quelle que soit la procédure, l'accord du service producteur des données est requis.*

📍 ... PUIS SÉANCE D'ENRÔLEMENT

Avant tout accès au CASD, les chercheurs, une fois habilités, doivent suivre dans les locaux du CASD à Palaiseau, une session de formation et de sensibilisation dite *séance d'enrôlement* durant laquelle ils sont sensibilisés aux lois de protection de la confidentialité et au respect des règles du secret statistique, à l'image de séances de formation telles qu'il en existe dans plusieurs centres à l'étranger.

Il s'ensuit une présentation des règles de sécurité informatique figurant dans les conditions d'utilisation du CASD signées par chaque utilisateur : accès strictement personnel, obligation de retirer sa carte d'authentification lorsque l'on s'absente (même un court instant) du poste SD-Box...

On y présente aussi les conditions d'hébergement du boîtier, bien que ces conditions fassent l'objet d'un contrat entre le CASD et l'établissement où sera installée la SD-Box. Par exemple, on demande que la SD-Box soit installée dans un local fermant à clé, que l'écran ne soit visible que par son utilisateur, etc. À la fin de la séance, le chercheur obtient sa carte d'accès et il lui appose, suivant une procédure encadrée par les ingénieurs du CASD, ses empreintes digitales (*figure 2*).

📍 UNE AUTONOMIE QUI N'EMPÊCHE PAS LE CONTRÔLE

Après l'enrôlement, les chercheurs reçoivent dans leur établissement un boîtier SD-Box. Il leur suffit alors de lui brancher un écran et un clavier et de le connecter au réseau. Ils peuvent immédiatement commencer à travailler et faire leurs analyses avec une réelle autonomie. À cela près que les utilisateurs ne peuvent techniquement récupérer aucun fichier à partir

de leur boîtier (**figure 3**). Celui-ci est isolé de tout autre dispositif. L'impression de fichiers, le transfert de fichiers, les opérations de copier-coller sont impossibles. Lorsque leur travail est suffisamment avancé et qu'ils désirent récupérer des fichiers de résultats, les chercheurs utilisent un programme du CASD. Celui-ci, entre autres, poste les résultats dans un espace du serveur réservé à cet usage :

- ❶ Dans le cas d'une procédure dite de *contrôle a priori*, les gestionnaires du CASD vérifient que l'utilisateur a bien réalisé le nécessaire pour garantir que les fichiers de résultats satisfont aux règles de confidentialité définies par le producteur des données, et si c'est bien le cas, les lui transmettent.
- ❷ Il existe une procédure automatique, sans vérifications manuelles, pour certaines catégories de données, comme les données de santé. Les utilisateurs doivent renseigner un formulaire électronique où ils s'engagent avoir respecté les règles de confidentialité. Le fichier leur est alors transmis automatiquement par notification de messagerie, accompagnée d'un lien de téléchargement sécurisé. Une copie de ces fichiers est conservée par le CASD pendant une durée de cinq ans pour permettre de réaliser des *contrôles a posteriori*.

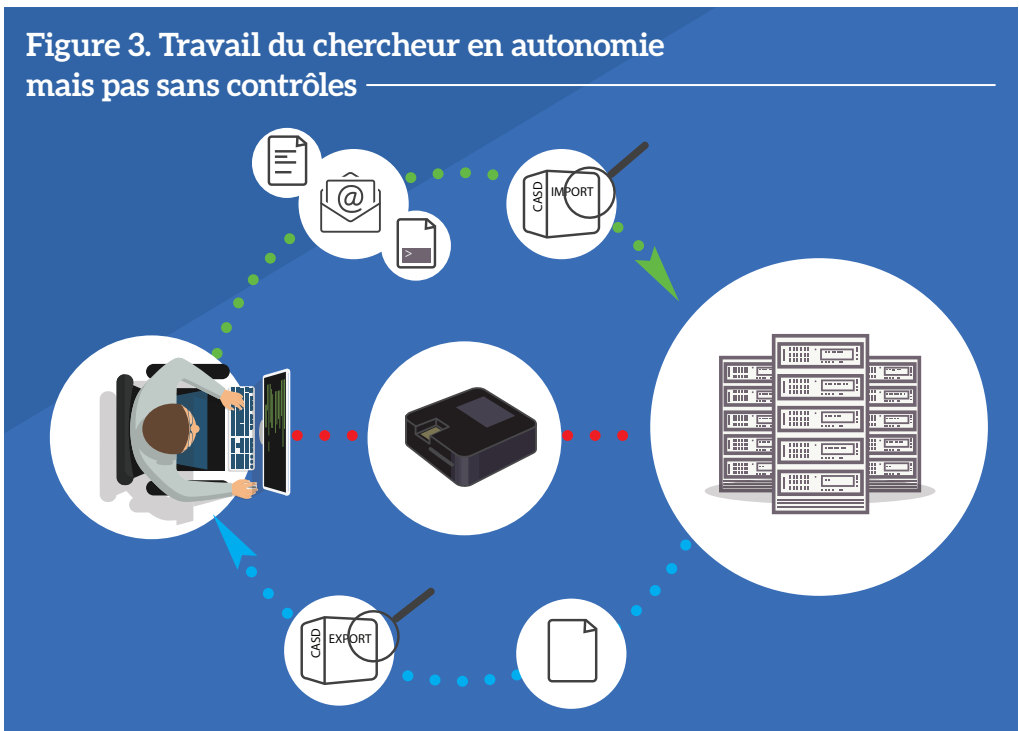
Dans les deux cas, les contrôles portent exclusivement sur la confidentialité des données, et en aucun cas sur la qualité ou la pertinence scientifique des travaux.



LES AVANTAGES POUR LES PRODUCTEURS DE DONNÉES...

Le fait que la diffusion sécurisée des données soit assurée par un tiers comme le CASD permet au producteur, dont ce n'est en général qu'une mission auxiliaire, de ne pas avoir à investir trop lourdement dans une infrastructure pour offrir ce service. Cela permet aussi de mutualiser ce service entre plusieurs producteurs de données afin d'en minimiser les coûts d'investissement et de fonctionnement ; il est à noter qu'il n'y a pas de coût d'entrée, ni même de coût d'exploitation pour un producteur de données souhaitant mettre à disposition sur le CASD ses données à des fins de recherche scientifique. Le CASD se charge également de contractualiser avec les chercheurs (*figure 4*). Dans bien des cas il n'est plus nécessaire pour les producteurs de conclure une convention avec les chercheurs pour permettre l'accès à leurs données. Et pour les producteurs qui souhaitent tout de même avoir une convention avec les chercheurs, celle-ci s'en trouve très nettement allégée car ne contenant plus de clauses spécifiques détaillées sur la technique et la sécurité.

Depuis la mise en application du RGPD¹⁰, les garanties de sécurité sont devenues des exigences juridiques fortes. Un modèle standardisé permet de réduire considérablement ces formalités et facilite ainsi la mise en conformité aux exigences des registres de traitements ou des études d'impact.



10. Règlement général de protection des données.

🌐 ... ET POUR LES CHERCHEURS

Cette configuration offre également le très grand avantage de rendre possible l'utilisation de données provenant de plusieurs producteurs par utilisation conjointe ou par appariement au sein d'un même environnement de travail.

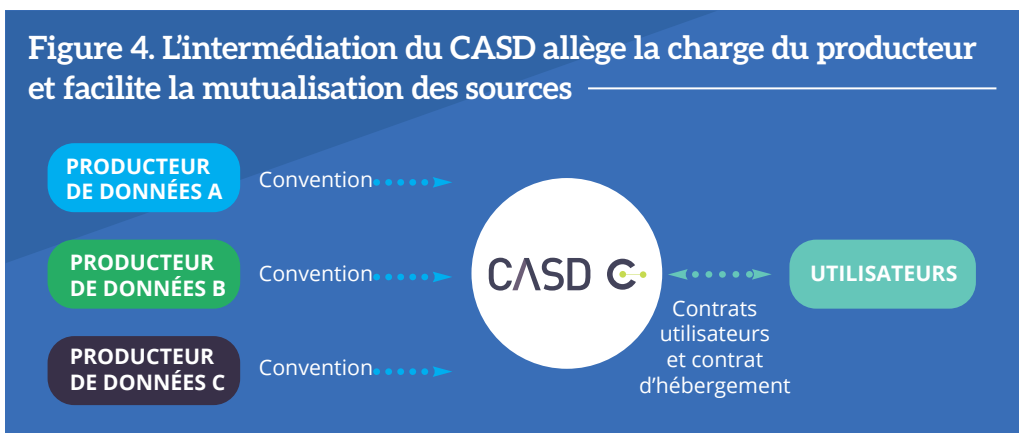
Les chercheurs n'ont d'ailleurs pas attendu longtemps pour en profiter : en 2013, déjà 16 % des projets utilisaient les sources de deux ou trois producteurs, à un moment où celles de l'Insee constituaient encore l'essentiel des données déposées au CASD. Depuis, sur un nombre de projets presque quatre fois plus important, la proportion de ceux utilisant les sources de plusieurs producteurs est montée à 52 % avec désormais des projets incluant les sources de quatre voire cinq producteurs de données. Actuellement, 171 projets utilisent conjointement les données de l'Insee et de la DGFIP.

Cette possibilité constitue un avantage pour les chercheurs, comparé à une situation comme celle du Royaume-Uni où les dispositifs d'accès aux données fiscales et aux données de la statistique publique sont différents. En effet, à l'étranger, dans la plupart des pays comparables, le développement de plusieurs centres d'accès sécurisés en silos a conduit à rendre complexe pour les chercheurs la réalisation d'appariements ou l'utilisation conjointe de plusieurs sources de données provenant de plusieurs producteurs. Cette difficulté s'explique notamment parce qu'à l'étranger, historiquement, les premiers centres d'accès sécurisés ont été créés sous forme de centres physiques avant de devenir des centres d'accès à distance.

Ce modèle de spécialisation et de mutualisation présente un avantage indéniable pour l'utilisation des données françaises. À notre connaissance, cette offre est unique au monde à cette échelle. Cela explique que de plus en plus de chercheurs européens demandent désormais l'accès aux données françaises.

🌐 DES APPARIEMENTS POUR DÉMULTIPLIER LES POSSIBILITÉS...

Les données disponibles sur le CASD sont en soi des sources très riches d'information pour les études et la recherche. Cependant, leur puissance d'information et d'explication se trouve démultipliée en les appariant, c'est-à-dire en enrichissant les informations recueillies pour un individu dans un fichier par celles disponibles pour ce même individu dans un



autre fichier. Certaines études ou évaluations ne sont possibles qu'à condition d'effectuer préalablement un tel appariement. C'est le cas par exemple lorsqu'il s'agit d'étudier les liens entre les revenus salariaux et les revenus de remplacement (chômage, indemnités journalières, d'assurance maladie, retraites), entre trajectoire scolaire et trajectoire professionnelle d'un individu ou entre la santé et le travail.

« C'est le cas par exemple lorsqu'il s'agit d'étudier les liens entre les revenus salariaux et les revenus de remplacement (chômage, indemnités journalières, d'assurance maladie, retraites...). »

De tels appariements de fichiers individuels sont particulièrement utiles et même nécessaires pour concevoir, mettre en place et évaluer des politiques publiques dans de nombreux domaines. Ils offrent des avantages par rapport à des enquêtes qui seraient spécifiquement conçues pour répondre à ces questions prédéfinies. De telles enquêtes seraient en effet très coûteuses au regard des moyens dont disposent les chercheurs et ne pourraient

évidemment porter que sur un échantillon beaucoup plus restreint que les fichiers administratifs, dont une des caractéristiques est d'être souvent exhaustifs sur la population concernée.

Jusqu'alors, il n'existait cependant que très peu d'études ou de recherches en France fondées sur les appariements de tels fichiers. En réalité, l'appariement qui permet réellement de faire correspondre des individus figurant dans deux fichiers, se fait en général sur un numéro d'identification comme le NIR (Numéro d'Inscription au Répertoire national d'identification des personnes physiques). Celui-ci figure dans un grand nombre de fichiers. Cependant son utilisation était jusqu'en 2017 très restreinte, car elle nécessitait préalablement la publication d'un décret en Conseil d'État autorisant le traitement.

📍 ... EN UTILISANT LE « NIR HACHÉ »

Depuis la loi pour une République numérique adoptée en 2016 et son décret d'application paru un an plus tard¹¹, il est devenu juridiquement possible de réaliser des traitements portant sur des données utilisant un dérivé du NIR. En effet, le NIR est un indicateur partiellement signifiant (sexe, âge, lieu de naissance) et, lors de sa création, il n'était prévu que pour des usages dans le domaine social.

La loi sur la Santé¹² a élargi ce domaine à celui d'un identifiant de santé des personnes pour leur prise en charge à des fins sanitaires et médico-sociales. La Cnil reste néanmoins toujours vigilante et préfère limiter son usage. Mais il existe des techniques permettant de faire correspondre à un NIR un autre indicateur, appelé « NIR haché » selon un processus asymétrique qui permet à chaque NIR d'avoir un correspondant unique, mais qui ne permet pas de recalculer le NIR d'origine à partir du « NIR haché ». Le hachage du NIR permet de rendre les mêmes services que l'usage du NIR lui-même, mais avec un risque considérablement atténué d'identifier une personne.

11. Loi n° 2016-1321 du 7 octobre 2016 (article 34) modifiant les articles 22 et 25 de la loi Informatique et libertés en vigueur, et décret n° 2016-1930 du 28 décembre 2016, portant simplification des formalités préalables relatives à des traitements à finalité statistique ou de recherche.

12. Loi n° 2016-41 du 26 janvier 2016 de Modernisation de notre système de santé (article 147).

Pour utiliser le « NIR haché », il faut recourir à un tiers de confiance pour la gestion des clés secrètes nécessaire aux opérations cryptographiques de *hachage*. Une clé secrète différente est créée pour chaque projet de recherche, aboutissant à des NIR chiffrés différents pour chaque appariement. Le résultat de l'appariement de deux fichiers est davantage ré-identifiant que les fichiers initiaux pris séparément. Cela oblige à prendre des précautions particulières quant à sa diffusion. C'est pour cette raison que la loi prévoit qu'un second tiers, comme le CASD, soit sollicité pour la réalisation effective de l'appariement à partir du NIR haché ainsi que pour la mise à disposition sécurisée des données une fois celles-ci appariées (*figure 5*).

🔗 INNOVER POUR CERTIFIER LES RECHERCHES FONDÉES SUR DES DONNÉES CONFIDENTIELLES...

Depuis l'ouverture sécurisée des données confidentielles, la question de la reproductibilité des calculs, garante de la scientificité de la démarche, s'est posée. Les journaux scientifiques demandent en effet que les chercheurs déposent données et code pour que les résultats publiés puissent être soumis à vérification par des tiers, ce qui pose bien évidemment des difficultés s'agissant de données confidentielles (Pérignon *et alii*, 2019).

Jusqu'ici la vérification des résultats devait se faire après soumission aux relecteurs (*reviewers*) désignés par l'éditeur. Or pour que le *reviewer* puisse faire ces vérifications, il doit suivre la procédure d'accréditation par le Comité du secret statistique et se déplacer au CASD pour se faire enrôler. Ce processus peut prendre des mois. En pratique, aucun *reviewer* n'a eu le temps ni les ressources nécessaires à consacrer à ce type de vérification.

Le CASD a conclu un partenariat avec une agence de certification, Cascad¹³, pour mettre en place une solution de certification de la reproductibilité d'une recherche qui s'appuie sur des données confidentielles.

L'agence spécialisée et accréditée vérifie la conformité des résultats, en amont de leur soumission à la revue scientifique. La certification est attribuée à l'issue d'un processus d'évaluation mené par un spécialiste du langage de programmation utilisé par le chercheur, à partir des données sources présentes sur le CASD et de l'ensemble des codes informatiques mis à disposition par le chercheur. Elle doit permettre d'augmenter les chances de publication d'un article dans les revues académiques.

Grâce au soutien du Comité du secret statistique et de l'ensemble des producteurs, un pilote a pu débuter en avril 2019 pour une durée d'un an. Le principe de ce pilote est qu'après habilitation du certificateur, le CASD lui ouvre un accès à un environnement sécurisé pour chaque demande sur la durée nécessaire à la certification. Les environnements ainsi créés sont fermés à la fin de chaque certification. Les programmes et les données associées sont estampillés et archivés de manière chiffrée sur une période de cinq ans.

Au moment où la question de la reproductibilité des résultats de la recherche est souvent mise en avant dans nombres de domaines, la systématisation de la certification offerte par ce service permettrait une avancée considérable¹⁴.

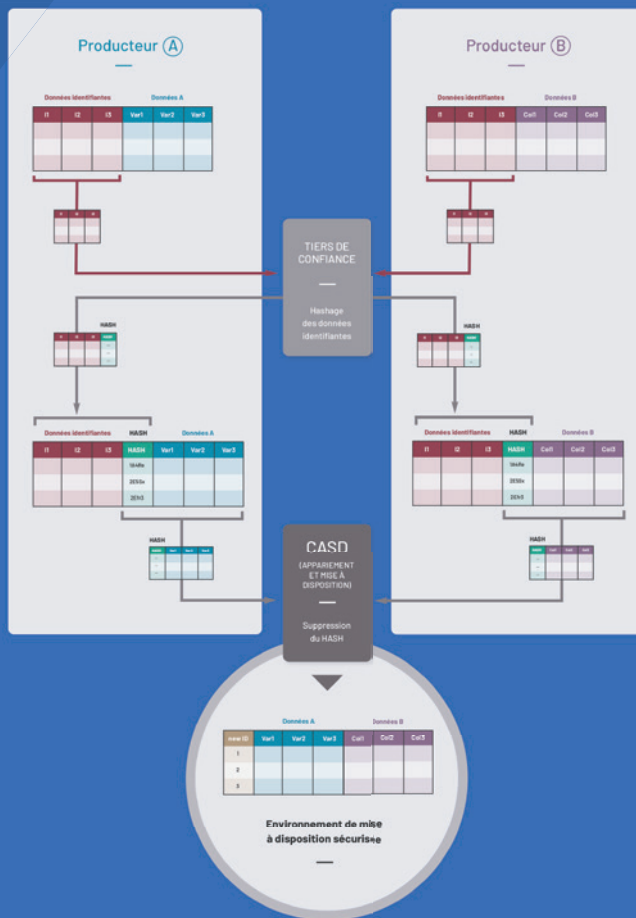
13. Cascad (*Certification Agency for Scientific Code and Data*) est une structure d'appui à la recherche, à but non lucratif, financée par différentes institutions françaises dont le CNRS, HEC Paris et l'Université d'Orléans.

14. Voir également la consultation publique lancée en juillet 2019 par la Cnil auprès des chercheurs sur les traitements de données à des fins de recherche scientifique (Cnil, 2019).

Figure 5. Un premier tiers de confiance produit une clé secrète et le CASD procède à l'appariement

Lorsque le CASD est acteur de la mise à disposition sécurisée des données appariées, la procédure est la suivante :

- pour chaque étude, le premier tiers de confiance génère une clé de hachage dédiée à l'appariement ;
- chaque producteur de données attribue un numéro aléatoire unique à chaque enregistrement afin d'obtenir un identifiant unique qui ne se rapporte à aucune autre information spécifique, appelé identificateur « neutre » ;
- chaque producteur de données envoie une table qui ne contient que l'identifiant neutre et le NIR au tiers de confiance. En parallèle, ils envoient l'identifiant neutre (sans le NIR) et les micro-données au CASD ;
- le tiers de confiance hache les NIR des fichiers de chaque producteur de données avec la clé de hachage et envoie le NIR haché accompagné de l'identifiant neutre au CASD ;
- le CASD a alors reçu toutes les tables nécessaires pour effectuer la mise en correspondance des données sans à aucun moment avoir connaissance du NIR et peut les mettre à disposition des chercheurs du projet.



OUVERTURES INTERNATIONALES

Sur le plan international, les producteurs de données ont donné leur accord pour un accès à distance transnational aux chercheurs de l'Union européenne et des pays associés de l'AELE¹⁵. Les chercheurs ont récemment été autorisés à travailler depuis les États-Unis et le Canada, avec quelques conditions supplémentaires¹⁶, sur les données de l'Insee et du ministère de l'Agriculture.

En bonne position de ce point de vue sur le plan international, après avoir participé au projet européen *Data without Boundaries (DwB)*¹⁷, le CASD coordonne la mise en place d'une collaboration entre centres d'accès sécurisé français, britanniques, allemands et hollandais (**figure 6**) : l'objectif d'IDAN (*International Data Access Network*) est de faciliter l'accès aux données sécurisées de ces pays pour les chercheurs, leur évitant des déplacements et leur permettant de mobiliser plus facilement les données de plusieurs pays à partir du site de chacun des partenaires du réseau. Ainsi, il devrait être possible d'ici la fin de l'année 2019, à partir de chacun des centres, d'accéder aux données de tous les autres centres.

Créé tardivement par rapport à d'autres grands pays en Europe et en Amérique du Nord, le CASD bénéficie aujourd'hui d'une position de pointe à la fois par sa technologie et par le nombre de données qui y sont déposées. Il a ainsi fait partie des quatre centres sécurisés auditionnés par la Commission américaine en charge du rapport sur l'ouverture des données administratives pour l'évaluation des politiques publiques et la recherche (American Congress, 2017). Le nombre croissant de données qui sont disponibles au CASD témoigne de la confiance créée du côté des producteurs, gage de futurs développements notamment dans le domaine de

Figure 6. Le CASD coordonne un réseau entre les centres français, britanniques, allemands et hollandais



15. AELE : Association européenne de libre-échange.

16. Répondre à au moins une des deux conditions suivantes : soit être citoyen européen, soit avoir un établissement européen impliqué dans le projet.

17. Il s'agit d'un projet dans le cadre du septième programme-cadre pour 2007-2013 de l'Union européenne (FP7), pour la recherche et le développement technologique (Alvheim *et alii*, 2012).

la santé à un moment où l'articulation entre ce domaine et celui des sciences économiques et sociales est de plus en plus important. Du côté des chercheurs, les craintes initialement exprimées devant les contraintes à passer par un système sécurisé ont été très largement compensées par le nombre et la richesse des nouvelles données ouvertes permettant la réalisation de projets importants utilisant les données françaises y compris à l'international.

Encadré 2. Statut et financement du CASD

Le CASD est un équipement labellisé en 2011 Équipex (Équipement d'Excellence du programme Investissements d'avenir) et a bénéficié à ce titre d'un financement jusqu'en 2019 pour se développer.

Le règlement de l'appel à projet Équipex exigeait notamment la mise en place d'un dispositif de facturation pour assurer l'autofinancement du service au-delà de 2019. C'est ce qu'a fait le CASD à partir de 2012 en facturant ses services. La facturation annuelle moyenne aujourd'hui est d'un peu plus d'un millier d'euros par utilisateur. Cette facturation sert à couvrir partiellement les frais d'exploitation, l'autre partie étant couverte par les contributions des partenaires du projet à savoir l'Insee, le Genes, le CNRS, l'École polytechnique et HEC Paris.

À la fin de l'année 2018, les partenaires du projet ont décidé de créer pour le CASD une structure dédiée, en constituant un Groupement d'intérêt public (GIP), personne morale de droit public à but non lucratif dotée de l'autonomie administrative et financière. La transformation du CASD en GIP dans la continuité du consortium Équipex lui permet de disposer d'un mode de fonctionnement plus souple et adapté tout en apportant des garanties aux partenaires publics en tant que personne publique.

La création d'une telle structure a permis d'inscrire dans les textes les missions du CASD, à savoir organiser et mettre en œuvre des services d'accès sécurisé pour les données confidentielles à des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation, activités qualifiées de services à la recherche, principalement publiques, et valoriser la technologie développée pour sécuriser l'accès aux données dans le secteur privé.

La mise en application récente du nouveau règlement européen de protection des données (RGPD) laisse présager un besoin accru de sécurisation pour tous ceux qui souhaitent travailler sur des données très détaillées. Le CASD en se dotant d'une structure autonome sera ainsi plus agile pour fournir le service nécessaire à la recherche.

BIBLIOGRAPHIE

ALVHEIM, Atle, BOND, Steve, GADOUCHE, Kamel, GÜRKE, Christopher et SCHILLER, David, 2012. *Report on the state of the art of current SC in Europe*. Septembre 2012. Project DwB, funded under : FP7-Infrastructures

AMERICAN CONGRESS, 2017. *The Promise of Evidence-Based Policymaking*. [en ligne]. Septembre 2017. Rapport de la Commission sur l'Élaboration de politiques fondées sur des données probantes. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://www.cep.gov/report/cep-final-report.pdf>

CAPELLE-BLANCARD, Günther et BELLANDO, Raphaëlle, 2015. *L'accès aux données bancaires et financières : une mission de service public*. [en ligne]. Juillet 2015. Rapport du groupe de travail du CNIS. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.cnis.fr/wp-content/uploads/2017/10/RAP_2015_140_acces-aux-donnees-bancaires.pdf

COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS (CNIL), 2016. *Communication cadre relative au Big Data*. [en ligne]. 18 février 2016. [Consulté le 8 octobre 2019]. Disponible à l'adresse : https://www.casd.eu/wp/wp-content/uploads/2016-02_Communication_cadre_relative_au_big_data.pdf

COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS (CNIL), 2019. *Régime juridique applicable aux traitements poursuivant une finalité de recherche scientifique (hors santé)*. [en ligne]. [Consulté le 26 septembre 2019]. Disponible à l'adresse : https://www.cnil.fr/sites/default/files/atoms/files/consultation_publicque_-_presentation_du_regime_juridique_applicable_aux_traitements_a_des_fins_de_recherche.pdf

GUESDON, Maxence, BENZENINE, Eric, GADOUCHE, Kamel, et QUANTIN, Catherine, 2016. *Securizing data linkage in french public statistics*. [en ligne]. 6 octobre 2016. BMC Medical Informatics and Decision Making. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://hal.inria.fr/hal-01377334/document>

LANE, Julia et SHIPP, Stephanie, 2007. Using a Remote Access Data Enclave for Data Dissemination. In : *The International Journal of Digital Curation*. [en ligne]. 27 juillet 2007. N°1, Volume 2 | 2007, pp. 128 134. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://www.ijdc.net/article/download/31/20/>

LE GLÉAU, Jean-Pierre, 2019. *Le secret statistique*. 2 mai 2019. EDP Sciences, Collection Le monde des données. ISBN 978-2-75982-342-0

LE GLÉAU, Jean-Pierre et ROYER, Jean-François, 2011. Le centre d'accès sécurisé aux données de la statistique publique française : un nouvel outil pour les chercheurs. In : *Courrier des statistiques*. [en ligne]. Mai 2011. N°130, pp. 1-5. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <https://www.epsilon.insee.fr/jspui/bitstream/1/8288/1/cs130e.pdf>

LOTH, André et alii, 2015. *Données de santé : anonymat et risque de ré-identification*. [en ligne]. 6 juillet 2015. Dossiers solidarité et santé, Drees, N° 64. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <https://drees.solidarites-sante.gouv.fr/IMG/pdf/dss64-2.pdf>

DE MONTJOYE, Yves-Alexandre et alii, 2018. *On the privacy conscientious use of mobile phone data*. [en ligne]. 11 décembre 2018. Nature, Scientific Data, Comment. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <https://www.nature.com/articles/sdata2018286.pdf>

MOREL-À-L'HUISSIER, Pierre et PETIT, Valérie, 2018. *Rapport d'information sur l'évaluation sur l'évaluation des dispositifs d'évaluation des politiques publiques*. [en ligne]. 15 mars 2018. Assemblée Nationale. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://www.assemblee-nationale.fr/15/pdf/rap-info/i0771.pdf>

PÉRIGNON, Christophe, GADOUCHE, Kamel, HURLIN, Christophe, SILBERMAN, Roxane et DEBONNEL, Éric, 2019. Certify reproducibility with confidential data. In : *Science*. [en ligne]. Juillet 2019. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://science.sciencemag.org/content/365/6449/127/tab-pdf>

SILBERMAN, Roxane, 1999. *Les sciences sociales et leurs données*. [en ligne]. Juin 1999. Rapport, Ministère de l'éducation nationale, de la recherche et de la technologie. [Consulté le 14 octobre 2019]. Disponible à l'adresse : <http://media.education.gouv.fr/file/96/5/5965.pdf>

SILBERMAN, Roxane, 2013. Transnational Access to Official Micro-data : The Data without Boundaries European Network. In : KLEINER, Brian, RENSCHLER, Isabelle, WERNLI, Boris, FARAGO, Peter, JOYE, Dominique, 2013. *Understanding Research Infrastructures in the Social Sciences*. Seismo Press, Social Sciences and Social Issues AG, Zurich, pp. 47-66. ISBN 978-3-03777-133-4

❶ FONDEMENTS JURIDIQUES

Loi n° 51-711 du 7 juin 1951 sur l'Obligation, la coordination et le secret en matière de statistiques

Loi n° 78-17 du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux libertés

Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la Protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la Protection des données)


Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique

Loi n° 2016-41 du 26 janvier 2016 de Modernisation de notre système de santé

UN NOUVEAU TRIPTYQUE JURIDIQUE POUR LES STATISTIQUES EUROPÉENNES

Hervé Piffeteau*

L'approfondissement de l'intégration européenne nécessite de recourir à un nombre croissant de statistiques pour le suivi et le pilotage des politiques européennes. Ces politiques, dont le spectre est de plus en plus étendu, sont fondées sur des enjeux partagés et complétés d'objectifs chiffrés. Basé initialement sur une forme de coopération, le système statistique européen repose de plus en plus sur des fondements juridiques qu'il convient d'adapter au contexte et aux besoins de l'approfondissement de l'intégration européenne. Les règlements-cadres, qui sont appelés à couvrir tous les grands domaines des statistiques européennes, constituent le nouveau socle juridique pour la production de celles-ci. Trois règlements-cadres ont déjà été adoptés en 2018-2019, qui couvrent les statistiques sociales (IESS), d'entreprises (FRIBS) et agricoles (IFS). Leur élaboration a permis de rationaliser la base législative actuellement en vigueur dans leur domaine. Le développement de ce nouveau socle a été quelque peu laborieux compte tenu d'une part du nouveau contexte institutionnel induit par le traité de Lisbonne et des attentes parfois divergentes des co-législateurs. Les changements apportés par ces règlements-cadres ont été âprement contestés par certains États membres. Il en ressort néanmoins un nouveau triptyque législatif pour le domaine de la statistique, susceptible de répondre de manière plus efficace au cours des prochaines années aux évolutions attendues.

 *The deepening of European integration requires the use of an increasing number of statistics for monitoring and steering European policies. These policies, which are increasingly broad in scope, are based on common objectives complemented by quantified targets. Initially based on a form of cooperation, the European statistical system is increasingly based on legal foundations that need to be adapted to the context and needs of deepening European integration. The Framework Regulations, which are intended to cover all the main areas of European statistics, constitute the new legal basis for their production. Three framework regulations have already been adopted in 2018-2019, covering social (IESS), enterprise (FRIBS) and agricultural (IFS) statistics. Their preparation has made it possible to rationalise the legal basis currently in force in their field. The development of this new basis has been somewhat laborious, considering the new institutional context induced by the Lisbon Treaty and sometimes divergent expectations of the co-legislators. The changes made by these framework regulations have been strongly contested by some Member States. However, it does suggest a new legislative triptych for statistics that can respond more effectively to expected developments in the coming years.*

*Chef de la division Coordination internationale, Insee,
herve.piffeteau@insee.fr

Nota : l'auteur remercie J-P Cling pour ses commentaires et suggestions sur cet article.

Il est communément admis que la construction européenne est avant tout une construction juridique : nul ne pourrait contester que le fondement juridique est essentiel pour une production harmonisée des statistiques européennes. Le traité de Lisbonne en 2009 a bouleversé le paysage juridique européen et son fonctionnement, en réformant d'une part les relations entre le Parlement européen, le Conseil de l'Union européenne et la Commission, et en modifiant d'autre part la comitologie¹. La statistique n'a pas échappé au mouvement. Face à une accumulation, au fil des années, de textes législatifs définissant ces statistiques, il était nécessaire de rationaliser et simplifier le socle juridique dans ce domaine, d'autant que l'accélération de l'intégration européenne s'est accompagnée d'une demande grandissante de nouveaux indicateurs (Scherrer, 2019). Face à ce nouvel équilibre institutionnel, les co-législateurs européens² et la Commission (Eurostat) ont élaboré un règlement-cadre respectivement pour les statistiques sociales (IESS), d'entreprises (FRIBS) et agricoles (IFS)³, pierres angulaires d'un nouveau triptyque juridique (règlement-cadre /acte délégué/acte d'exécution) indispensable à la production des statistiques européennes.

📊 UNE STATISTIQUE ORGANISÉE POUR APPUYER LA CONSTRUCTION EUROPÉENNE

Le pilotage de l'Union européenne (UE) nécessite des statistiques fiables et harmonisées entre les pays. Les besoins d'information sont liés à la mise en œuvre et au suivi des politiques européennes, lesquelles concernent des domaines variés : finances publiques, politiques macroéconomique, régionale, environnementale, sociétale, etc.

Le terme de « statistiques européennes » recouvre les informations quantitatives nécessaires à l'accomplissement des activités européennes. Elles sont régies par le programme statistique européen, dont les modalités sont définies par le règlement statistique du Conseil et du Parlement européen 223/2009 modifié en 2015 (2015/759). D'emblée, en

matière de statistique européenne, le cadre juridique apparaît au premier plan (Mordant, 2009).

« D'emblée, en matière de statistique européenne, le cadre juridique apparaît au premier plan. »

Afin d'assurer la comparabilité et la cohérence des données de tous les États membres, celles-ci doivent se conformer à des cadres communs de référence. Ces normes sont très majoritairement inscrites

dans un texte adopté par le Parlement européen et le Conseil de l'Union européenne. Le cadre législatif leur confère un statut juridique fort et une priorité dans l'allocation des ressources : cela vaut pour les instituts nationaux de statistique (INS), qui ont une obligation de réponse, mais aussi pour Eurostat qui peut allouer des subventions aux États membres.

Malgré des organisations et des périmètres variables, tous les INS mettent en œuvre la législation européenne relative aux statistiques. Ils participent ainsi à des travaux et des opérations communs sur la base des orientations stratégiques du système statistique européen (SSE) (**encadré 1**).

1. Ce terme désigne le processus d'élaboration et d'encadrement des compétences d'exécution ou des compétences déléguées que la Commission peut se voir confier par un règlement législatif.

2. Il s'agit du Parlement européen et du Conseil de l'Union européenne.

3. En anglais, IESS : *Integrated European Social Statistics*, FRIBS : *Framework Regulation Integrating Business Statistics*, IFS : *Integrated Farming Statistics*.

Encadré 1. Le système statistique européen, du politique à l'opérationnel

Le système statistique européen (SSE) comprend :

- l'autorité statistique de la Commission, à savoir la direction générale Eurostat,
- les instituts nationaux de statistique (INS) et les autres autorités statistiques nationales (ONAs pour *Other National Authorities*), responsables dans chaque État membre du développement, de la production et de la diffusion des statistiques européennes.

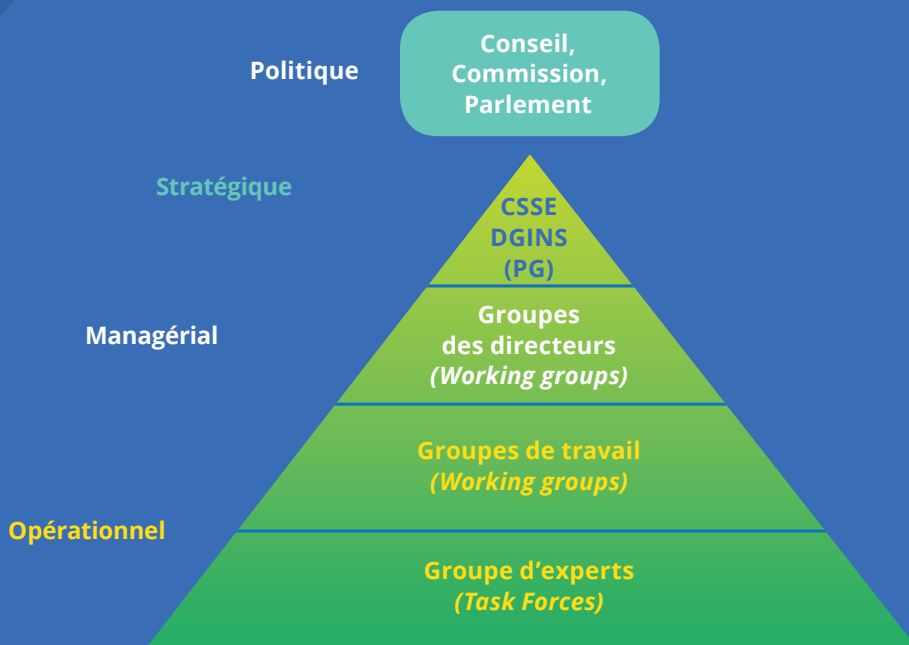
En France, ces missions incombent à l'Insee et à certains services statistiques ministériels.

Le SSE fonctionne comme un réseau, au sein duquel Eurostat est chargé de promouvoir la production de statistiques comparables, en étroite collaboration avec les autorités statistiques nationales.

Le comité du SSE (CSSE), qui est présidé par Eurostat et regroupe les directeurs généraux des INS (DGINS), conseille le SSE pour le développement, la production et la diffusion des statistiques européennes selon les principes statistiques énoncés à l'article 338 du traité sur le fonctionnement de l'UE, et précisés dans le Code de bonnes pratiques de la statistique européenne.

Le comité du SSE est épaulé dans ses missions par les groupes de directeurs établis par grands domaines d'activité (méthodologie, informatique, statistiques sociales, d'entreprises, comptes nationaux...) eux-mêmes secondés par des groupes de travail ou des groupes d'experts.

Cet ensemble dessine quatre niveaux : politique, stratégique, managérial et opérationnel.



La production repose sur un même modèle : l'institut national de statistique de chaque État membre produit des données soit détaillées soit agrégées, à partir desquelles il élabore ses propres statistiques « européennes ». Celles-ci sont transmises à Eurostat qui les centralise et les exploite sous forme d'indicateurs et de tableaux diffusés sur son site internet⁴. En parallèle, la diffusion nationale des statistiques européennes est assurée par chaque INS.

Au nom du principe de subsidiarité, selon lequel une autorité centrale ne peut effectuer que les tâches qui ne peuvent pas être entreprises à l'échelon inférieur, les statistiques européennes sont essentiellement produites par les systèmes statistiques nationaux, selon les méthodes que ceux-ci jugent les plus adaptées : enquêtes, exploitation de sources administratives, ou de données massives (*big data*), etc. Ils ont aussi le choix des modes d'enquête, des techniques d'échantillonnage, des méthodes de traitement des données, etc. L'essentiel réside dans l'obligation qui leur est faite d'aboutir à des statistiques conformes aux définitions et aux normes de qualité adoptées dans le cadre législatif européen.

📊 UNE DEMANDE DE STATISTIQUES EUROPÉENNES EN AUGMENTATION CONSTANTE

L'évolution de la construction européenne a conduit ces dernières années à une accélération et une amplification des demandes de statistiques communes et comparables. Le spectre « législatif » des statistiques européennes s'est en conséquence progressivement étendu, fruit d'une demande grandissante d'indicateurs mais aussi d'un besoin d'informations à des niveaux géographiques de plus en plus fins.

“ Pas moins d'une soixantaine de textes encadrent la production de statistiques. ”

Ce contexte a conduit à un foisonnement de textes législatifs, relevant plus de l'« empilement » que d'une organisation structurée : pas moins d'une soixantaine de textes encadrent la production de statistiques, dans de nombreux

domaines, tels que l'agriculture, l'industrie, l'énergie, les transports, l'emploi ou le revenu des ménages.

📊 UNE ADAPTATION RENDUE NÉCESSAIRE SUR LE PLAN LÉGISLATIF

Dans un souci de mieux satisfaire les attentes des décideurs et du public notamment en termes de comparabilité et de disponibilité des statistiques, tout en veillant à améliorer le rapport coût-efficacité de leur production, la Commission européenne a jugé opportun d'adapter son approche législative.

Elle a d'abord cherché à rationaliser, harmoniser et simplifier, à travers des règlements-cadres, la base législative pour les statistiques européennes, en intégrant des règlements sectoriels jusqu'alors autonomes. La rationalisation de l'appareil législatif a aussi donné une base législative à la production de certaines statistiques, qui étaient jusqu'alors fondées sur des accords entre Eurostat et les États membres.

4. Voir <https://ec.europa.eu/eurostat/fr/home>.

Dans ce contexte, trois règlements-cadres ont été mis en chantier en restructurant la statistique européenne dans trois grands domaines (statistiques agricoles, statistiques d'entreprises, statistiques sociales) couverts chacun par un règlement-cadre :

- ❶ IESS : *Integrated European Social Statistics*, ou Statistiques sociales européennes intégrées ;
- ❷ FRIBS : *Framework Regulation Integrating Business Statistics*, ou règlement cadre pour les statistiques d'entreprises intégrées ;
- ❸ IFS *Integrated Farming Statistics*, ou Statistiques agricoles intégrées.

Leur entrée en application devrait permettre de réduire de près de 20 % la masse des textes législatifs régissant la production des statistiques européennes⁵.

Par exemple, le règlement-cadre sur les statistiques sociales IESS⁶ a permis d'abroger et donc d'intégrer les deux principaux règlements législatifs qui constituaient les piliers des enquêtes sociales : celui sur l'enquête Emploi (*Labour Force Survey*) et celui sur l'enquête relative aux Ressources et conditions de vie des ménages (*SILC*). FRIBS, le règlement-cadre relatif aux statistiques d'entreprises, remplace à lui seul dix règlements législatifs sectoriels⁷.

❶ RENFORCER LA COMPARABILITÉ DES STATISTIQUES EUROPÉENNES

Au premier rang des améliorations attendues, un objectif de ces règlements-cadres est de permettre une meilleure comparabilité des statistiques européennes entre pays mais aussi entre les sources de chacun des domaines.

Par exemple, IESS doit permettre d'améliorer la comparabilité des statistiques sur le statut d'activité de la population (au sens du Bureau International du Travail) par la mise en place d'une structure commune et obligatoire du questionnaire de l'enquête Emploi. En parallèle, dans le cadre de la refonte de la législation pour les statistiques sociales, un travail important de standardisation des variables, ainsi qu'une harmonisation des concepts (introduction du concept de ménage économique dans l'enquête Emploi et dans l'enquête Ressources et conditions de vie des ménages) ont été entrepris.

Dans le domaine des statistiques d'entreprises c'est la définition d'une Entreprise qui a été rendue plus cohérente dans le règlement-cadre FRIBS.

Toujours dans le but de renforcer la comparabilité, cette nouvelle approche législative vise également :

- ❶ à promouvoir des échanges de données individuelles entre pays lorsque cela est nécessaire ;
- ❷ à exploiter des sources multiples (données administratives, enquêtes, mégadonnées, etc.) ;
- ❸ à développer des méthodes innovantes et la mise à disposition d'informations à des niveaux géographiques très fins.

FRIBS apportera par exemple des simplifications et des adaptations au nouveau contexte de la mondialisation, notamment pour l'échange de micro-données sur les flux d'exportations intracommunautaires et par la production d'un répertoire communautaire des groupes d'entreprises qui est sensiblement enrichi.

5. Les règlements FRIBS, IESS et IFS ne couvrent pas la totalité de la production statistique européenne.

6. Voir l'article de Chantal Cases sur IESS dans ce même numéro.

7. Voir l'article de Christel Colin sur FRIBS dans ce même numéro.

Dans le domaine des statistiques agricoles, IFS vise à simplifier et mieux intégrer les processus de collecte par la fusion et la refonte du règlement relatif aux structures des exploitations avec celui consacré aux différentes cultures développées sur ces exploitations.

📍 ÉLARGIR LES THÉMATIQUES...

Eurostat a souhaité également établir des normes dans des domaines plus étendus que ceux couverts jusqu'à présent. De nouveaux thèmes apparaissent, comme la mondialisation, le changement climatique, le vieillissement de la population.

IFS, le premier règlement-cadre adopté en août 2018, avait par exemple pour but d'intégrer les nouvelles demandes induites par les évolutions de la politique agricole commune, les nouvelles exigences agro-environnementales, ainsi que les attentes sociétales et économiques.

Sur le sujet de la mondialisation, on vise à mieux en mesurer les effets par un meilleur profilage des entreprises. Les registres nationaux d'entreprises et le répertoire européen des groupes seront ainsi davantage exploités sur cette thématique. La mise en œuvre de FRIBS permettra également d'élargir la couverture à des secteurs en forte croissance comme les services et le commerce. Un indice mensuel de production dans les services sera notamment produit (Scherrer, 2019).

Enfin, IESS fournit désormais un cadre légal pour le calcul de l'indicateur mensuel européen de chômage même si les modalités de son calcul par Eurostat à partir de données transmises par les États membres restent quasi inchangées.

“ Ces règlements-cadres sont aussi l'occasion d'établir une base législative et donc contraignante à des enquêtes qui étaient jusqu'à présent menées sur une base volontaire. ”

Ces règlements-cadres sont aussi l'occasion d'établir une base législative et donc contraignante à des enquêtes qui étaient jusqu'à présent menées sur une base volontaire. Ainsi, le champ du règlement-cadre IESS inclut désormais les enquêtes Budget de famille et Emploi du temps, et même si cette dernière demeure pour le moment optionnelle, elle devrait à terme être rendue obligatoire⁸.

📍 ... POUR RÉPONDRE AUX BESOINS ACCRUS DES DÉCIDEURS PUBLICS

Si les règlements-cadres prévoient une extension du champ des statistiques européennes, c'est également pour mieux prendre en compte les demandes des utilisateurs particuliers que sont les décideurs politiques : pour le suivi des politiques européennes, de nouveaux indicateurs sont toujours nécessaires ; la tendance étant à l'amplification, il convenait donc d'anticiper, autant que possible, les demandes futures. Cette volonté s'illustre au moins de trois manières.

Ainsi, IESS précise déjà les concepts et les définitions pour l'enquête Emploi du temps alors que celle-ci demeure facultative pour le moment. L'élaboration d'un programme pluriannuel pour les enquêtes sociales annuelles, couplé avec l'introduction de sujets *ad hoc*

8. Aucune date n'est toutefois précisée dans le règlement-cadre. Eurostat avait initialement souhaité rendre cette enquête obligatoire mais a dû se résigner face à la forte opposition des États membres (voir *infra*).

complémentaires, permettra également de satisfaire plus aisément les futures demandes en matière de statistiques sociales. Deux sujets *ad hoc* sur une période de huit ans doivent compléter l'enquête Emploi et pas moins de quatre pour l'enquête relative aux Ressources et conditions de vie des ménages sont inscrits dans ce programme.

« La flexibilité est renforcée par la mise en place d'études pilotes, ou de faisabilité, qui permettront non seulement d'incorporer les nouvelles demandes des utilisateurs mais aussi de les tester. »

La flexibilité est renforcée par la mise en place d'études pilotes, ou de faisabilité, qui permettront non seulement d'incorporer les nouvelles demandes des utilisateurs mais aussi de les tester avant un éventuel passage en production de manière pérenne. De nouvelles thématiques testées dans le cadre d'études pilotes sont déjà propo-

sées pour FRIBS tels que les échanges internationaux de services, l'immobilier commercial et résidentiel, ou encore l'environnement et le climat.

Les délais de diffusion des statistiques et des données à Eurostat ont parfois été raccourcis. Les données relatives aux Ressources et conditions de vie des ménages devront ainsi être transmises avant la fin de l'année de référence, et à un niveau géographique plus fin, dans le but de mieux satisfaire les attentes des décideurs pour le suivi des politiques régionales. L'indicateur de pauvreté et d'exclusion sociale (AROPE) sera ainsi désormais calculé par Eurostat au niveau NUTS 2⁹ sur la base de données détaillées transmises par les États membres. Dans le domaine des statistiques d'entreprises, certains indicateurs conjoncturels tels que les indices de chiffre d'affaires seront désormais publiés sur une base mensuelle et non plus trimestrielle.

🕒 UNE NÉGOCIATION LONGUE ET LABORIEUSE

Cette refonte de la législation européenne a des implications majeures pour la production statistique. En France, elle a nécessité une forte mobilisation de l'Insee et des services statistiques des ministères, ainsi qu'une étroite concertation entre les acteurs pour mener à bien les négociations (**encadré 2**).

Le premier règlement-cadre relatif aux statistiques sur les exploitations agricoles européennes a été adopté après 18 mois de négociations, ce qui est raisonnable au regard de la durée moyenne de la procédure législative (**figure 1**). La durée des négociations a été bien plus longue en ce qui concerne les deux autres règlements-cadres : près de trois ans de négociations ont été nécessaires pour trouver un accord de compromis par les deux co-législateurs européens pour IESS et près de deux ans pour FRIBS. Pour IESS, quinze réunions du groupe statistique du Conseil, échelonnées entre septembre 2016 et janvier 2019 ont été nécessaires avant l'ouverture des discussions avec le Parlement européen (contre seulement dix réunions entre mars 2018 et janvier 2019 pour FRIBS).

Au final, il aura fallu près de dix ans entre la phase de préparation et l'entrée en application de IESS, et sept ans de préparation et de négociations pour FRIBS.

9. Nomenclature des unités territoriales statistiques d'Eurostat, dont le niveau 2 (NUTS 2), est destiné à comparer les régions européennes entre elles. En France, ce niveau correspond aux régions avant la réforme territoriale. Voir également <https://insee.fr/fr/metadonnees/definition/c2112>.

Encadré 2. Processus de négociations d'un règlement-cadre : l'exemple de IESS

(*Étape 1*) Contrairement à l'approche traditionnelle de l'élaboration des règlements européens, les membres du système statistique européen (SSE) ont été étroitement associés à la préparation du projet de règlement, en participant à différents ateliers, groupes de travail, où chacun a pu exposer ses propositions et meilleures pratiques.

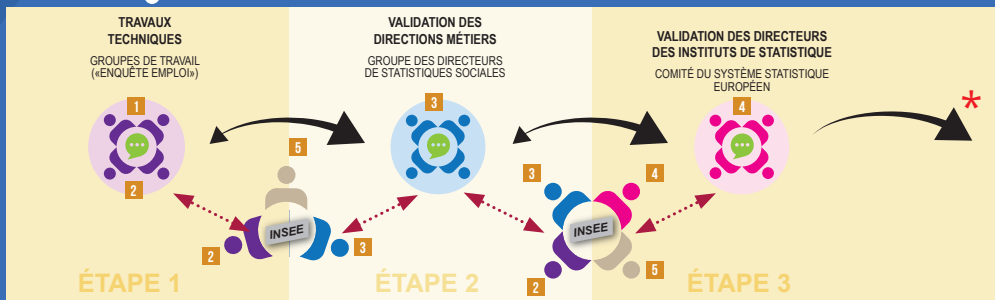
(*Étape 2*) Puis les propositions de chaque groupe de travail ont été transmises au groupe des directeurs concerné qui a veillé à la cohérence d'ensemble.

(*Étape 3*) Avant soumission au comité du système statistique européen (CSSE), ces propositions ont été examinées dans chacune des capitales européennes pour en étudier l'impact et la charge sur les systèmes nationaux de statistique. Cette étape marque la fin de la phase « technique » de l'ébauche d'un règlement.

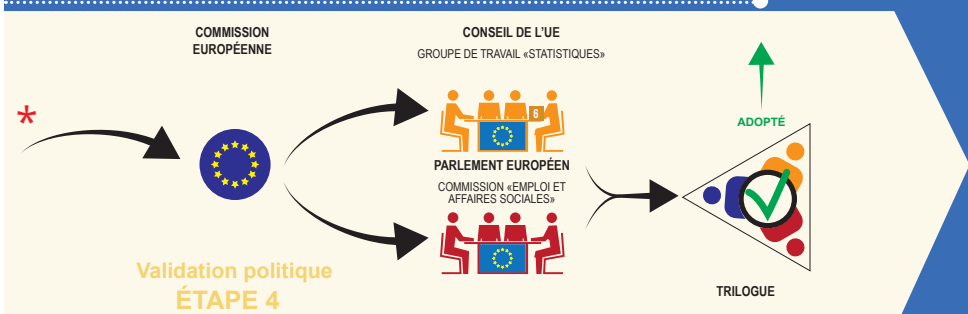
(*Étape 4*) La Commission (Eurostat) a ensuite transmis son projet de loi aux deux co-législateurs, le Parlement européen et le Conseil de l'Union européenne ; ceux-ci, après examen, ont chacun proposé des amendements à Eurostat. Il s'agit donc de l'étape « politique ».

Malgré le travail technique mené en amont, certains pays peuvent revenir sur des accords passés ou s'opposer fortement aux amendements introduits par le Parlement. Le processus de négociation peut s'avérer fastidieux : il s'agit de réconcilier les exigences politiques de la Commission, celles du Parlement et la nécessité de disposer de statistiques fiables pour le suivi des politiques européennes. Le compromis recherché doit définir une production de données respectant des normes de qualité, dans des délais et à un coût raisonnables pour les INS. Sur ce dernier point, Eurostat accorde des subventions aux États membres, afin de faciliter la mise en œuvre des nouvelles spécifications induites par ces règlements (délais, précisions, nouvelles données...), mais aussi dans la perspective de nouveaux développements.

2010



Mise en œuvre 2021



DES DIFFICULTÉS LIÉES AU NOUVEAU CONTEXTE INSTITUTIONNEL...

Les difficultés rencontrées pour faire aboutir les négociations concernant ces règlements européens sont principalement liées à la mise en œuvre des nouvelles règles du traité de Lisbonne sur le fonctionnement de l'Union européenne (2009). En effet, une nouvelle architecture légale a émergé au fil de ces négociations, un triptyque constitué du règlement-cadre, des actes délégués et des actes d'exécution (*figure 2*). Selon les termes du traité¹⁰ :

- ❶ « un acte législatif peut déléguer à la Commission le pouvoir d'adopter des actes non législatifs de portée générale qui complètent ou modifient certains éléments non essentiels de l'acte législatif » ;
- ❷ par ailleurs¹¹, des actes d'exécution peuvent être adoptés « lorsque des conditions uniformes d'exécution des actes juridiquement contraignants de l'Union sont nécessaires... ».

Une difficulté importante rencontrée dans la négociation de ces différents règlements tient aux incertitudes que génèrent les actes délégués. Les critiques suscitées par les actes délégués résultent pour partie du caractère hybride de ces actes : non législatifs de par leur forme, car non adoptés selon la procédure législative européenne, ils le sont de par leur portée puisqu'ils permettent de modifier des textes législatifs (Guilloud, 2011). De plus, à la différence des actes d'exécution, les actes délégués ne sont pas votés par le comité du système statistique européen (CSSE) : ils entraînent de ce fait une perte d'influence des INS, tout en renforçant en contrepartie celle de la Commission (qui reçoit la délégation), ainsi que le pouvoir de contrôle du Parlement européen.

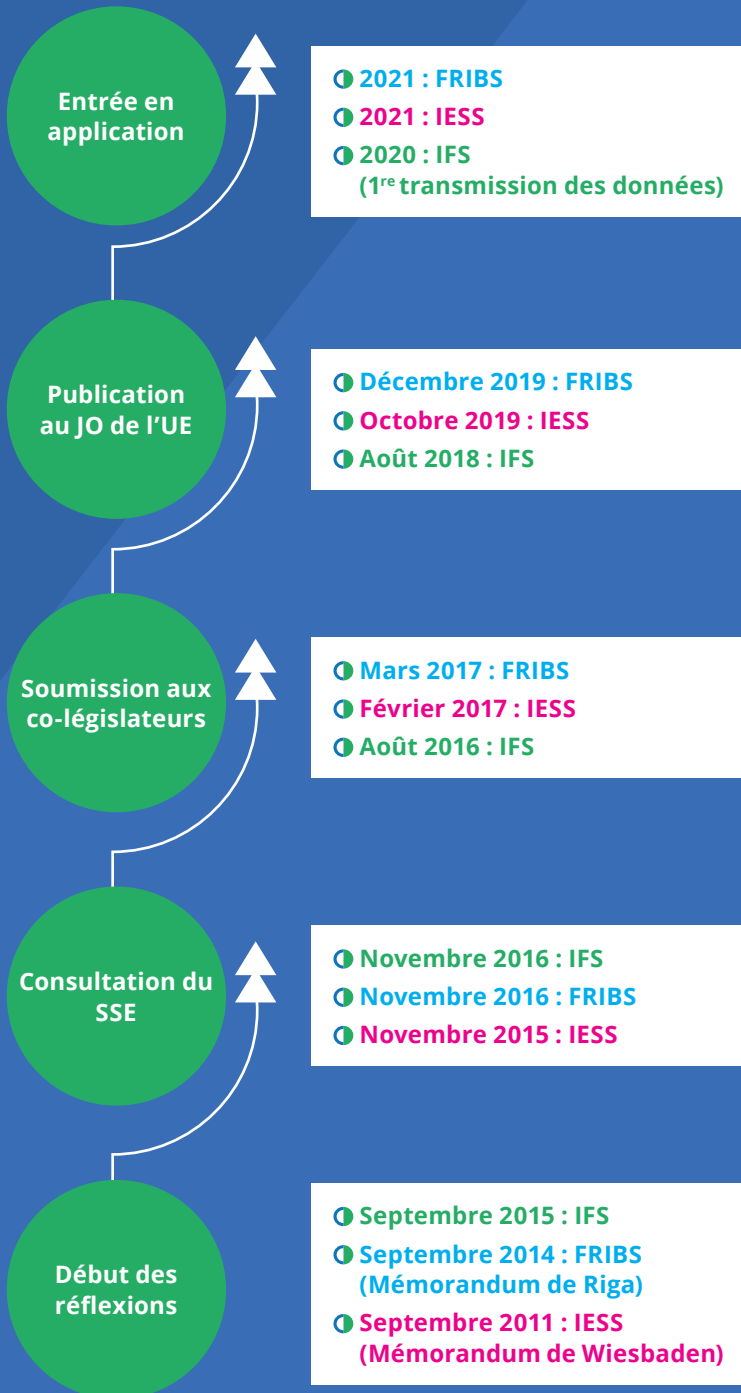
La Commission a pour objectif de couvrir par des règlements-cadres tous les grands domaines des statistiques européennes ; elle aura tendance à recourir à des actes délégués pour compléter les actes de base, qui ont par définition un contenu de nature assez générale. Cet objectif est légitime et l'intérêt des règlements-cadres est reconnu de façon unanime au sein du CSSE. Bien qu'elle suscite des inquiétudes chez quelques États membres, la délégation est également perçue comme un instrument permettant d'adapter autant que de besoin la production statistique des INS sans recourir à des procédures législatives répétées et longues.

Une autre nouveauté introduite par le traité de Lisbonne pose également problème : il s'agit des trilogues, réunions tripartites entre le Parlement européen, le Conseil de l'Union européenne et la Commission. Les INS se plaignent notamment que les États membres sont souvent contraints de choisir entre le blocage du processus législatif (exemple sur IESS) ou l'acceptation de compromis difficiles. Si les INS sont étroitement associés en amont, à l'élaboration du projet du règlement (*encadré 2*), le CSSE n'est pas toujours bien informé de l'avancée des négociations menées en trilogue. Pour y remédier, le CSSE avait suggéré de renforcer les liens avec les rapporteurs du Parlement européen en charge des règlements statistiques afin de faciliter la recherche d'un compromis tout en préservant les fondements de la statistique tels que décrits dans le code de bonnes pratiques.

10. Article 290 du Traité de Fonctionnement de l'Union européenne.

11. Article 291 du Traité de Fonctionnement de l'Union européenne.

Figure 1. Genèse d'un règlement-cadre (FRIBS, IESS, IFS)

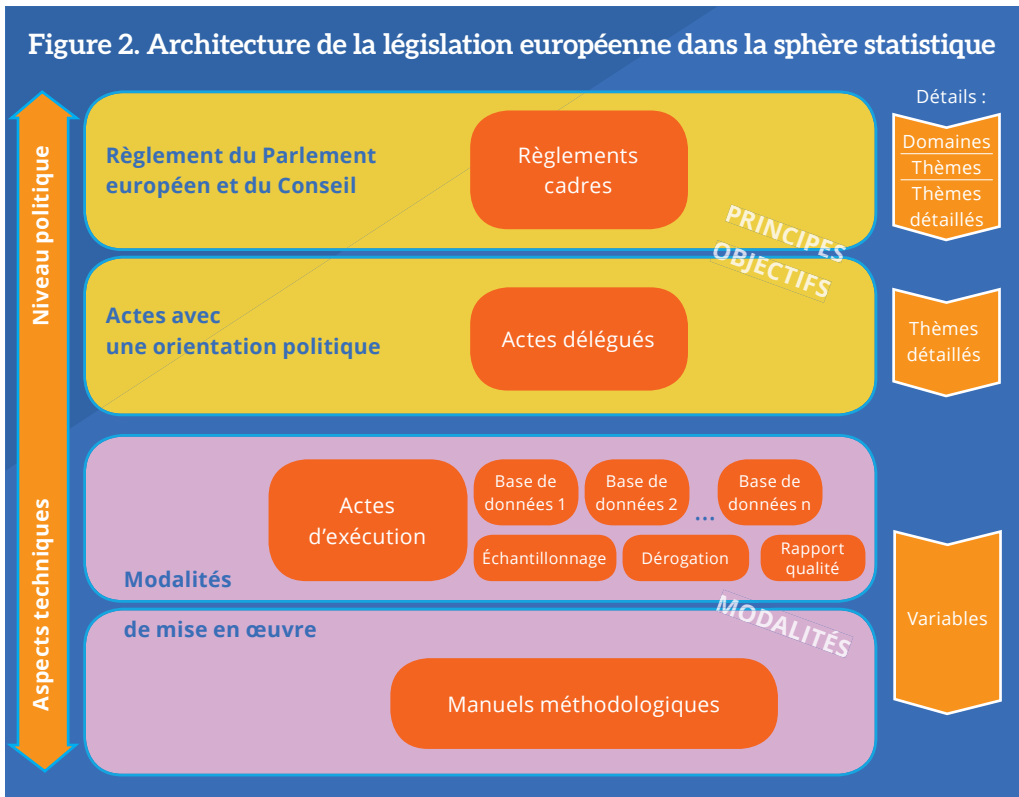


IFS : Integrated Farm Statistics
 IESS : Integrated European Social Statistics
 FRIBS : Framework Regulation Integrating Business Statistics

... MAIS AUSSI LIÉES À LA FORTE INFLUENCE DU PARLEMENT EUROPÉEN...

Au-delà du renforcement de ses pouvoirs suite au traité de Lisbonne, le Parlement européen voit s'exprimer à travers lui l'influence de divers groupes de pression ou d'agences.

Parmi ses propositions d'amendements pour IESS, le Parlement avait souhaité autoriser les agences européennes, qui n'appartiennent pas au SSE, à produire des statistiques en obligeant les États à leur fournir des échantillons ou de l'information pour le tirage d'échantillons¹². Le Parlement avait également exprimé sa volonté de confier à des agences hors SSE, la mission de contribuer à la production de statistiques européennes alors que ce rôle était jusqu'ici dévolu aux seuls acteurs du SSE, selon les dispositions de la loi statistique de 2009 et le code des bonnes pratiques¹³. L'introduction d'un tel amendement avait alors constitué un frein majeur à l'obtention d'un texte de compromis pour IESS. La rédaction en a finalement été ajustée pour éviter qu'à terme, le SSE soit élargi à certaines agences de l'Union pour la production de statistiques européennes. Toutefois, ce risque n'est pas complètement écarté puisque cette assimilation entre agences de l'Union européenne et le système statistique européen transparait, au moment de la rédaction de cet article, dans les amendements introduits par le Parlement dans le projet de programme statistique pluriannuel¹⁴.



12. Cette disposition a été considérablement amoindrie dans le compromis final : la Commission est simplement invitée à contribuer à l'expertise des agences qui mèneraient des enquêtes statistiques dans le domaine social ne relevant pas du SSE pour l'élaboration de nouveaux indicateurs.

13. Certaines enquêtes auprès des ménages au niveau européen s'appuient déjà sur des instituts de sondage privés (EUROFRA par exemple sur l'enquête Discriminations).

14. Le règlement relatif au programme pluriannuel des statistiques européennes est désormais intégré au programme plus vaste en faveur du marché unique, de la compétitivité des entreprises et des statistiques européennes.

L'extension de couverture de ces règlements-cadres a été appuyée par le Parlement qui a ainsi proposé que le commerce international de services, ainsi que le calcul d'indicateurs financiers relatifs aux entreprises puissent être mieux intégrés à FRIBS. De la même façon, le Parlement avait souhaité que l'enquête Emploi du temps puisse être rendue obligatoire. Ces dispositions figuraient, entre autres, parmi les points les plus clivants âprement discutés lors des trilogues.

... ET AUX FORTES RÉTICENCES DES ÉTATS MEMBRES

Dans un environnement externe mouvant et de plus en plus concurrentiel, la réduction des ressources nationales dédiées à la statistique publique accroît les difficultés des INS face à une demande croissante de statistiques. Si la délégation de pouvoirs consentie à la Commission permet d'adapter de manière flexible la production statistique des INS, elle a néanmoins suscité de fortes inquiétudes au cours des négociations parmi les États membres. Une majorité d'entre eux ont alors souhaité introduire des clauses de sauvegarde qui restreignent le nombre de thèmes et de variables qui peuvent être ajoutés ou modifiés, en vue de limiter les charges additionnelles. L'institut de statistique danois avait mené la

Encadré 3. L'Insee et les services statistiques ministériels concernés, favorables à la mise en œuvre des règlements-cadres

L'Insee a soutenu fortement les deux projets de règlements-cadres IESS et FRIBS dès leur préparation. Ceux-ci visaient en effet à moderniser et harmoniser les définitions et concepts qui avaient émergé au fil de l'élaboration des règlements sectoriels. L'Insee était d'autant plus favorable à la mise en œuvre de ces nouveaux règlements que l'institut avait déjà largement anticipé la rénovation de son système d'enquêtes sociales avec le lancement d'un programme d'enquêtes en multimode. Cette approche intégrée était également soutenue du côté des statistiques d'entreprises : cela permettait de gommer les incohérences et de réduire les redondances, notamment en ce qui concerne les concepts relatifs à l'unité statistique. L'Insee et le service statistique en charge du commerce extérieur ont également activement contribué à l'élaboration de FRIBS en participant dès le lancement à des projets pilotes tels que ceux relatifs au profilage des grands groupes multinationaux et aux échanges de biens intra-UE.

L'Insee avait cependant quelques réserves sur certaines des dispositions introduites par ces deux règlements, notamment en ce qui concerne les délais de transmission à Eurostat des données SILC* et les contraintes de précision introduites dans IESS. L'assouplissement de certaines des mesures lors des négociations au Conseil (recours possible aux estimations sur petits domaines), l'ajustement des échantillons d'enquêtes, voire le recours à des mesures dérogatoires permettront finalement à l'Insee de se conformer à ces nouvelles exigences. Pour FRIBS, les principales inquiétudes portaient sur l'arborescence juridique de ce texte qui dans sa phase initiale accordait une place prépondérante aux actes délégués.

Enfin, en ce qui concerne IFS, il s'agissait de simplifier et de rendre plus flexible le processus de collecte des données relatives aux exploitations agricoles, avec des variables de base à collecter, de manière exhaustive *via* un recensement exhaustif, puis sur la base d'un échantillon pour les enquêtes intermédiaires. L'Insee et le service statistique du ministère de l'Agriculture ont veillé à préserver cette flexibilité et à la renforcer notamment par l'ajout de modules et de variables.

* L'enquête statistique sur les Ressources et conditions de vie (Dispositif SRCV) est la partie française du système communautaire EU-SILC (*European union-Statistics on income and living conditions*).

« Dans un environnement externe mouvant et de plus en plus concurrentiel, la réduction des ressources nationales dédiées à la statistique publique accroît les difficultés des INS face à une demande croissante de statistiques. »

fronde avec les instituts néerlandais et polonais : ces trois INS ont justifié la mise en place de tels garde-fous par leur souhait de mieux prioriser les demandes toujours croissantes de la Commission européenne. L'introduction de telles clauses restrictives vient par conséquent amoindrir les bénéfices attendus des règlements-cadres, d'autant que leurs modalités de mise en œuvre sont complexes, voire impraticables.

L'harmonisation de la production des statistiques sociales, tant au niveau des concepts que des méthodes employées, a parfois été vivement contestée par quelques États membres qui jugeaient sans doute cette harmonisation par les inputs comme une entrave à la production nationale de statistiques sur ces mêmes thèmes. L'Allemagne s'est ainsi longtemps opposée à l'instauration d'une structure commune et obligatoire pour le questionnaire de l'enquête Emploi, ce qui avait contribué fortement à l'enlisement des négociations.

L'échange de micro-données confidentielles sur les répertoires d'entreprises et sur le commerce international de biens, l'une des innovations majeures de FRIBS, a également été fortement contesté, notamment par l'Irlande qui le considère comme une entorse au respect du secret statistique.

➊ APRÈS DE LONGUES NÉGOCIATIONS, UNE NOUVELLE ARCHITECTURE LÉGALE ÉMERGE...

L'analyse des trois règlements-cadres adoptés en 2018 et 2019 (**encadré 4**) permet finalement de dégager quelques grandes lignes de l'architecture légale des nouveaux règlements-cadres statistiques.

L'acte de base décrit, selon une approche modulaire (domaines/thèmes/thèmes détaillés/modules), les principes généraux pour la production des statistiques européennes, tandis que les actes délégués et les actes d'exécution en détaillent les modalités. Au fil des discussions au sein du Conseil, une référence collective informelle a été peu à peu établie :

- ➊ la population statistique couverte, la fréquence des statistiques, le thème traité et les exigences de précision sont généralement considérés par les co-législateurs comme des éléments essentiels. Ils sont donc présents dans l'acte de base ;
- ➋ au-delà de ce « noyau dur », des différences liées à la nature des statistiques apparaissent entre IFS, IESS et FRIBS sur ce qui figure ou non dans l'acte législatif, notamment en ce qui concerne le niveau de ventilation des résultats ou leur date limite de transmission à Eurostat.

Le Parlement européen, le Conseil et la Commission ont signé une convention d'entente¹⁵ afin d'établir des critères non contraignants pour distinguer les actes délégués et les actes d'exécution. Ces critères, ainsi que la jurisprudence de la cour de justice européenne, servent

15. Accords entre le Parlement européen, le Conseil de l'Union européenne et la Commission européenne du 18 juin 2019 (JO 2019/C 223/01).

Encadré 4. Les architectures légales de FRIBS, IESS et IFS sont partiellement différentes

Le règlement-cadre sur les statistiques sociales (IESS) renforcera la cohérence entre les enquêtes ménages pour les concepts employés et les statistiques produites. Les enquêtes couvrent sept domaines principaux. Chaque domaine est composé de plusieurs thèmes qui peuvent être décomposés en thèmes détaillés. Cette approche modulaire, que l'on retrouvait déjà en partie dans le règlement-cadre des statistiques agricoles, est fixée dans le texte législatif, avec un tronc commun de variables couvrant tous les domaines conformément à une approche dite « d'harmonisation par les inputs ». Cette architecture thématique peut être amendable par acte délégué selon des dispositions fortement encadrées et précisées dans le texte législatif.

Pour IESS, le planning pluriannuel des enquêtes est également adopté par acte délégué alors que la périodicité des enquêtes est fixée par le règlement de base. La Commission (Eurostat) décide donc simplement du millésime de chacune des enquêtes à périodicité supra-annuelle. Cet exemple est assez éloquent de la porosité qui existe entre les éléments essentiels de l'acte législatif et ce qui peut être modifié ou complété par un acte délégué.

Par ailleurs, en vue d'assurer une forte comparabilité des données en matière d'emploi et de chômage et donc la possibilité de renforcer l'harmonisation par les inputs, les co-législateurs ont cherché, par le recours à un acte d'exécution, à fixer non seulement les variables à mesurer mais la conception précise des questionnaires dans le cas spécifique de l'enquête Emploi (architecture des questions dans le module BIT). Cette disposition a en particulier été soutenue par l'Insee.

Le règlement-cadre sur les statistiques d'entreprises (FRIBS) va remplacer dix règlements sectoriels et apportera des simplifications et des adaptations au nouveau contexte de la mondialisation (échange de micro données sur les flux d'exportations intracommunautaires, production d'un répertoire communautaire des groupes d'entreprises, etc.). L'articulation globale entre la structure thématique (domaines\ thèmes\ thèmes détaillés) et l'architecture légale du texte est assez proche de celle qui prévaut pour les statistiques sociales, avec toutefois certaines différences (par exemple le niveau de ventilation des résultats qui ne figurerait pas dans l'acte de base). L'architecture légale sous-jacente (actes d'exécution vs acte délégué) est toutefois différente en raison de l'exploitation de sources de données et de processus d'élaboration des statistiques distincts de ceux utilisés pour les statistiques sociales.

Le règlement-cadre sur les exploitations agricoles (IFS) vise à refondre les deux règlements législatifs l'un sur les exploitations agricoles et l'autre sur leur production afin de réduire la charge et le coût de ces statistiques. L'approche retenue dans ce règlement est une approche modulaire, avec des variables de base à collecter de manière exhaustive *via* un recensement en 2020, puis sur la base d'un échantillon pour les enquêtes intermédiaires en 2023 et en 2026. Dans la perspective de la refonte de la politique agricole commune, les co-législateurs ont habilité Eurostat à adopter des actes délégués pour compléter les données à collecter dans ces modules si des besoins supplémentaires émergent.

de lignes directrices aux co-législateurs. Si la discussion sur les contours des actes délégués et des actes d'exécution est parfois difficile tant au niveau du Conseil qu'en trilogue avec le Parlement, une certaine « jurisprudence » tend toutefois à se dessiner au fil de la rédaction des règlements-cadres dans la sphère statistique :

- ❶ lorsqu'il s'agit de modifier une production statistique suffisamment décrite dans le texte de base, le Conseil tend à borner très fortement la délégation, par exemple par des clauses qui imposent de ne pas accroître de manière significative la charge et les coûts tant pour les INS que pour les répondants ;
- ❶ lorsque le sujet n'est pas suffisamment détaillé voire pas du tout défini dans le corps du texte, et que la délégation a pour objet principal de compléter le règlement, les États membres ont tendance à exiger que l'ajout de dispositions par acte délégué soit explicitement motivé par la Commission.

Les négociations des règlements-cadres ont aussi conduit à faire évoluer progressivement les positions des États membres. Une majorité d'entre eux ont ainsi accepté que les thèmes et les thèmes détaillés soient décrits dans le règlement de base, et que la liste des variables au sein de chacun des thèmes – dont le nombre peut être limité par le règlement de base

« Les négociations des règlements-cadres ont aussi conduit à faire évoluer progressivement les positions des États membres. »

– soit définie par un acte délégué. Il est aussi désormais communément admis que le thème et le contenu des enquêtes *ad hoc* supra annuelles greffées sur les enquêtes annuelles soient définis par un acte délégué¹⁶.

Le comité du SSE peut par ailleurs adopter des actes d'exécution, principalement pour définir les spécifications techniques des variables statistiques, le contenu des rapports qualité, les formats de transmission, éventuellement d'autres standards techniques (dans le cas de l'enquête Emploi, on va plus loin puisqu'on fixe aussi par acte d'exécution le détail et l'ordre des questions du module BIT, voir **encadré 2** sur IESS et **encadré 4**).

❶ ... POUR ABOUTIR À UN SYSTÈME LÉGISLATIF DÉJÀ SENSIBLEMENT RESTRUCTURÉ POUR LES STATISTIQUES

La mutation en cours du cadre législatif pour le système statistique européen s'est donc traduite par l'adoption de trois règlements-cadres couvrant les statistiques sociales, d'entreprises et agricoles. Les négociations ont fait émerger peu à peu l'ossature d'un triptyque juridique, qui constitue le socle à partir duquel la Commission (Eurostat) pourra continuer à élaborer ses futurs règlements-cadres. Les textes sont examinés dans différentes enceintes juridiques : les actes d'exécution par le CSSE, les actes délégués par les experts des INS et le texte législatif par le Conseil et le Parlement. La transmission simultanée par Eurostat aux États membres de tous ces projets de textes, constitutifs du nouveau triptyque juridique, en amont des discussions au Conseil, aurait sans doute permis d'améliorer le processus de négociation. Pour le règlement FRIBS, les États membres avaient ainsi pu disposer des différentes mesures d'application, ce qui avait facilité l'obtention d'un compromis comme en atteste la durée plus courte des négociations au Conseil.

16. Exemple des données *ad hoc* (article 9 du règlement-cadre sur les exploitations agricoles 2018/1091) dont l'ajout doit être motivé et ne pas entraîner une charge disproportionnée et injustifiée sur les INS.

Deux autres règlements-cadres sont en préparation, l'un concernant les statistiques de production agricole¹⁷ et un second relatif aux statistiques de population. Ils viendront enrichir le socle légal du SSE à partir duquel seront désormais produites la majorité des statistiques européennes. Cette adaptation du système statistique et de son socle législatif contribue à mieux structurer la demande croissante de statistiques par grands domaines, dans un contexte par ailleurs marqué par la réduction des ressources. Cette mutation du système statistique européen constitue un levier fort pour l'Insee, pour s'adapter d'une part aux évolutions économiques et sociales et d'autre part aux défis soulevés par les évolutions du contexte de la statistique aux niveaux national et européen.

17. Le règlement-cadre sur la production agricole (produits et matières utilisées pour la production agricole) complètera les statistiques du recensement agricole définies par le règlement IFS.

BIBLIOGRAPHIE

EUROSTAT, 2009. *Communication de la Commission au Parlement européen et au Conseil concernant la méthode de production des statistiques de l'Union européenne : une vision de la prochaine décennie*. [en ligne]. 10 août 2009. N° COM/2009/0404. [Consulté le 19 novembre 2019]. Disponible à l'adresse :

<https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52009DC0404&from=FR>

EUROSTAT, 2010. *Cadre juridique pour les statistiques européennes, la Loi statistique*. [en ligne]. Compact guides, édition 2010. [Consulté le 19 novembre 2019]. Disponible à l'adresse :

<https://ec.europa.eu/eurostat/documents/3217494/5719345/KS-31-09-254-FR.PDF/f2bb628d-5d6b-49a2-bc5c-3292d5840597?version=1.0>

GUILLOUD, Laetitia, 2011. La nouvelle nomenclature des actes dans le Traité de Lisbonne. In : *Revue interdisciplinaire d'études juridiques*. [en ligne]. 2011/1, Vol. 66, pp. 85-108. [Consulté le 19 novembre 2019]. Disponible à l'adresse :

<https://www.cairn.info/revue-interdisciplinaire-d-etudes-juridiques-2011-1-page-85.htm>

MORDANT, Guillaume, 2009. La nouvelle loi statistique européenne. In : *Courrier des statistiques*. [en ligne]. Septembre-décembre 2009. N°128, pp. 25-32. [Consulté le 19 novembre 2019]. Disponible à l'adresse :

<https://www.epsilon.insee.fr/jspui/bitstream/1/8551/1/cs128g.pdf>

SCHERRER, Philippe, 2019. Statistiques conjoncturelles d'entreprises – Des impulsions européennes, des avancées françaises. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 72-85. [Consulté le 19 novembre 2019]. Disponible à l'adresse :

<https://www.insee.fr/fr/statistiques/fichier/4168406/courstat-2-8.pdf>

FONDEMENTS JURIDIQUES

(IFS) Règlement (UE) 2018/1091 du Parlement européen et du Conseil du 18 juillet 2018 concernant les statistiques intégrées sur les exploitations [en ligne]. [Consulté le 13 décembre 2019]. Disponible à l'adresse :

<https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32018R1091&from=FR>

(IESS) Règlement (UE) 2019/1700 du Parlement européen et du Conseil du 10 octobre 2019 établissant un cadre commun pour des statistiques européennes relatives aux personnes et aux ménages fondées sur des données au niveau individuel collectées à partir d'échantillons. [en ligne]. [Consulté le 13 décembre 2019]. Disponible à l'adresse :

<https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019R1700&from=FR>

(FRIBS) La publication au Journal Officiel de l'Union Européenne du règlement-cadre FRIBS est en cours à la date de publication de cet article.


Critères non contraignants pour l'application des articles 290 et 291 du traité sur le fonctionnement de l'Union européenne. [en ligne]. 18 juin 2019. [Consulté le 19 novembre 2019]. Disponible à l'adresse :

[https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019Q0703\(01\)&from=FR](https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019Q0703(01)&from=FR)

FRIBS : UN NOUVEAU CADRE COMMUN POUR LES STATISTIQUES D'ENTREPRISES EUROPÉENNES

Christel Colin*

Le règlement-cadre sur les statistiques d'entreprises, nommé FRIBS (Framework Regulation Integrating Business Statistics) et adopté en 2019, est le résultat d'un long processus de préparation et de négociation au niveau européen. Trois principes fondateurs le guident : s'appuyer sur des bases méthodologiques, des définitions et des critères de qualité communs ; permettre une plus grande flexibilité pour s'adapter aux changements économiques et aux besoins émergents de statistiques ; rechercher des synergies entre domaines et des simplifications dans les données demandées afin de réduire la charge sur les entreprises. De fait, FRIBS remplace dix règlements ou décisions sectoriels actuellement en vigueur. Il crée un cadre juridique commun pour les statistiques d'entreprises, réduisant les incohérences entre domaines et les redondances, sans les supprimer totalement. Au-delà de cette approche intégrée, plusieurs évolutions et innovations vont permettre de mieux couvrir certains secteurs du tertiaire et certains thèmes comme la mondialisation ou les groupes. FRIBS s'appliquera pour l'essentiel à partir de l'année de référence 2021: une bonne partie des évolutions demandées dans ce cadre ont déjà été anticipées et mises en œuvre par la France. Mais la partie restante demandera encore des adaptations conséquentes.

 *The Framework Regulation Integrating Business Statistics (FRIBS), adopted in 2019, is the result of a long process of preparation and negotiation at European level. Three fundamental principles guide it: common methodological basis, definitions and quality criteria; greater flexibility to adapt to economic changes and emerging statistical needs; synergies between domains and simplifications in the data requested in order to reduce the burden on businesses. In fact, FRIBS replaces ten sectoral regulations or decisions currently in force. It creates a common legal framework for business statistics, reducing inconsistencies between domains and redundancies, without removing them completely. Beyond this integrated approach, it introduces several developments and innovations, the main ones aimed at better covering certain sectors in the tertiary sector and certain themes such as globalisation or groups. FRIBS will apply mainly from the 2021 reference year: many of the changes requested in this context have already been anticipated and implemented by France. But the remaining part will still require significant adjustments.*

*Directrice des Statistiques d'entreprises, Insee,
christel.colin@insee.fr

Les statistiques d'entreprises produites par la statistique publique française sont depuis de nombreuses années très largement régies par des règlements européens. Les institutions de l'Union européenne (UE) ont en effet eu besoin d'indicateurs économiques, notamment sur les entreprises, afin d'accompagner la construction européenne et son approfondissement. L'enquête communautaire sur la production industrielle a fait l'objet d'un règlement dès 1991, le premier règlement européen relatif aux statistiques structurelles date de 1996 et le cadre européen pour les indicateurs conjoncturels a été posé en 1998. Historiquement toutefois les statistiques européennes d'entreprises étaient définies dans des règlements dits « sectoriels », par domaine, selon une approche en tuyaux d'orgue. Afin de rationaliser cette architecture législative dans l'esprit du Traité de Lisbonne (2009), de la rendre plus aisée à adapter aux évolutions futures, et de gagner en cohérence et en efficacité, Eurostat a lancé au début des années 2010 plusieurs projets de règlements dits « cadres », relatifs aux statistiques sociales, aux statistiques agricoles, aux statistiques d'entreprises¹.

🌐 LA PRÉPARATION DE FRIBS, UN TRAVAIL DE LONGUE HALEINE

Le règlement-cadre sur les statistiques d'entreprises, nommé FRIBS pour *Framework regulation integrating business statistics* et adopté en 2019, est le résultat d'un long processus de préparation et de négociation au niveau européen.

Dès 2009, Eurostat avait posé les bases d'une stratégie² pour la décennie suivante (Eurostat, 2009) fondée sur une nouvelle génération d'actes législatifs trans-domaines, dans lesquels les sujets techniques seraient renvoyés aux textes d'application (actes d'exécution et actes délégués). Une telle architecture permettrait de réduire les redondances et incohérences entre domaines, sources d'inefficacité pour les instituts nationaux de statistique (INS) et Eurostat, et de charge de réponse pour les enquêtés.

Le projet FRIBS lancé en 2011 affiche ainsi ses ambitions :

« Le règlement-cadre a pour objectif d'intégrer les statistiques d'entreprises au sein de l'UE, de les rationaliser, de les rendre plus flexibles et de réduire la charge induite par la réponse aux questionnaires statistiques. Il fournira des outils d'infrastructure communs pour la collecte et la production de statistiques d'entreprises tout en améliorant leur pertinence et leur qualité générale. L'objectif opérationnel de ce projet est de mettre en place un cadre juridique commun pour la collecte, l'établissement, la transmission et la diffusion systématiques de statistiques européennes liées à la structure, à l'activité économique, à la compétitivité, aux transactions mondiales et aux performances économiques du secteur des entreprises. Il ne s'agit toutefois pas d'une simple consolidation des textes juridiques existants dans les divers domaines relatifs aux statistiques d'entreprises, mais bien d'un projet ambitieux reposant sur une révision critique des exigences en matière de données, des dispositions méthodologiques et des dispositions de mise en œuvre existantes ainsi que sur une analyse des besoins des utilisateurs » (Eurostat, 2012).

1. Voir également les articles d'Hervé Piffeteau sur le nouveau triptyque juridique de la statistique européenne, et de Chantal Cases sur IESS, dans ce même numéro.

2. Ceci ne concernait pas que le domaine des statistiques d'entreprises et s'est appliqué à d'autres domaines (voir l'article de Chantal Cases sur IESS dans ce même numéro).

Aux origines de FRIBS, on trouve donc l'idée d'un règlement-cadre qui remplace plusieurs règlements sectoriels existants. Il doit s'appuyer sur des principes méthodologiques, des définitions et critères de qualité communs. Il vise à permettre d'accroître la flexibilité pour s'adapter aux changements économiques et aux besoins émergents de nouvelles statistiques. Par ailleurs, il recherche des synergies entre domaines et des simplifications dans les données demandées afin de réduire la charge de réponse des entreprises.

Le projet FRIBS a également été alimenté par les conclusions du programme MEETS (*Programme for the Modernisation of European Enterprise and Trade Statistics*) qui a couvert la période 2009-2013. Basé sur des travaux d'investissements méthodologiques et conceptuels au sein des ESSnets³, ce programme a notamment traité des questions :

- ❶ de cohérence des concepts et méthodes ;
- ❷ de méthodologie du répertoire européen des groupes (l'EGR) ;
- ❸ de profilage des grands groupes complexes multinationaux⁴ ;
- ❹ de modernisation d'Intrastat, le système de statistiques sur le commerce international de biens intra-communautaire.

Plusieurs projets de la « Vision 2020 » d'Eurostat⁵ ont ensuite pris le relais. Le projet ESBRS (*European System of Interoperable Statistical Business Registers*) a permis d'approfondir les sujets liés au cadre européen commun pour les répertoires statistiques d'entreprises et

« Tous ces travaux ont été des *inputs* importants dans la préparation du règlement FRIBS. »

à l'EGR : meilleure harmonisation des répertoires nationaux, améliorations de la qualité de l'EGR, et donc des statistiques sur les groupes et la mondialisation, partage de services (d'identification par exemple) entre INS. Les projets SIMSTAT et REDESIGN ont quant à eux permis de tester la faisabilité de l'échange de micro-données sur le commerce

intra-Union Européenne (intra-UE) de biens et de mener une analyse coûts-avantages des différentes options pour une refonte de la production des statistiques sur le commerce intracommunautaire des biens. Tous ces travaux ont été des *inputs* importants dans la préparation du règlement FRIBS.

❶ UN RÈGLEMENT-CADRE QUI REMPLACE DIX RÈGLEMENTS SECTORIELS EXISTANTS..

FRIBS remplace dix règlements ou décisions sectoriels du Parlement européen et du Conseil actuellement en vigueur (**encadré 1**). Ces textes couvrent un large spectre d'opérations statistiques, comme en témoigne la liste suivante (**figure 1**) :

- ❶ les statistiques conjoncturelles (sur l'activité et la production, les prix de production et d'importation, l'emploi, les heures travaillées, les permis de construire, etc.) ;
- ❷ les statistiques structurelles sur la démographie et les résultats économiques des entreprises ;

3. *European Statistical System Network* ou réseaux collaboratifs entre instituts nationaux de statistique.

4. Voir à ce sujet (Haag, 2019).

5. Orientation stratégique d'Eurostat, adoptée en 2014, voir <https://ec.europa.eu/eurostat/web/ess/about-us/ess-vision-2020>.

- ① les répertoires statistiques d'entreprises ;
- ① les statistiques sur la production industrielle (Prodcum) ;
- ① les statistiques sur les filiales étrangères (IFATS et OFATS) ;
- ① les statistiques sur la société de l'information (usage des technologies de l'information et communication (TIC) et commerce électronique) ;
- ① les statistiques sur la science et la technologie (R&D et Innovation) ;
- ① les statistiques sur le commerce extérieur de biens.

Sans abroger de règlement existant, le règlement-cadre FRIBS couvre également :

- ① en partie, les statistiques sur les échanges de services : celles-ci sont actuellement régies par le règlement 184/2005 du Parlement européen et du Conseil du 12 janvier 2005 relatif aux statistiques communautaires de la balance des paiements, du commerce international des services et des investissements directs étrangers. FRIBS inclut des demandes de statistiques primaires sur les échanges de services, sur la base desquelles des imputations ultérieures pourront être effectuées pour produire des données et des agrégats conformes aux normes européennes et internationales pour la balance des paiements. Cela fait suite à un subtil compromis trouvé en 2016, après de longues négociations, entre la Commission européenne (Eurostat) et le Comité sur les statistiques monétaires, financières et de balance des paiements (CMFB), compromis formalisé dans un avis du CMFB du 12 septembre 2016 (CMFB, 2016) ;

Encadré 1. Dix règlements et décisions abrogés par FRIBS

- ① Création d'une enquête communautaire sur la production industrielle (*Règlement (CEE) n°3924/91 du Conseil du 19 décembre 1991*).
- ① Statistiques conjoncturelles (*Règlement (CE) n°1165/98 du Conseil du 19 mai 1998*).
- ① Production et développement de statistiques communautaires de la science et de la technologie (*Décision n°1608/2003/CE du Parlement européen et du Conseil du 22 juillet 2003*).
- ① Production de statistiques communautaires annuelles de l'industrie sidérurgique pour les années de référence 2003-2009 (*Règlement (CE) n°48/2004 du Parlement européen et du Conseil du 5 décembre 2003*).
- ① Statistiques communautaires sur la société de l'information (*Règlement (CE) n°808/2004 du Parlement européen et du Conseil du 21 avril 2004*).
- ① Statistiques communautaires des échanges de biens entre États membres (*Règlement (CE) n°638/2004 du Parlement européen et du Conseil du 31 mars 2004 abrogeant le règlement (CEE) n°3330/91 du Conseil*).
- ① Statistiques communautaires sur la structure et l'activité des filiales étrangères (*Règlement (CE) n°716/2007 du Parlement européen et du Conseil du 20 juin 2007*).
- ① Cadre commun pour le développement de répertoires d'entreprises utilisés à des fins statistiques (*Règlement (CE) n°177/2008 du Parlement européen et du Conseil du 20 février 2008, abrogeant le règlement (CEE) n°2186/93 du Conseil*).
- ① Refonte des statistiques structurelles sur les entreprises (*Règlement (CE) n°295/2008 du Parlement européen et du Conseil du 11 mars 2008*).
- ① Statistiques communautaires relatives au commerce extérieur avec les pays tiers (*Règlement (CE) n°471/2009 du Parlement européen et du Conseil du 6 mai 2009 abrogeant le règlement (CE) n°1172/95 du Conseil*).

● les statistiques sur les chaînes d'activité (ou de valeur) mondiales, c'est-à-dire sur les externalisations, délocalisations et créations d'activité à l'étranger et relocalisations d'activité. Ce domaine, important pour l'analyse des questions de mondialisation, fait jusqu'à présent l'objet d'enquêtes pilotes, réalisées dans certains pays de l'Union européenne sur une base volontaire.

Dans ses premières versions, FRIBS couvrait aussi les statistiques sur les investissements directs étrangers, mais ce domaine a été sorti du champ suite à l'avis du CMFB sus-cité.

La question s'est également posée d'inclure les statistiques sur le tourisme. Ce dernier est finalement resté hors champ et demeure régi par un règlement de 2011 qui couvre à la fois les statistiques d'offre et de fréquentation des hébergements touristiques, et celles sur les comportements de départ des ménages.

Le règlement FRIBS ne couvre pas non plus les « statistiques sociales fondées sur les entreprises » comme celles sur la formation continue dans les entreprises ou sur le coût de la main-d'œuvre et la structure des salaires. La possibilité d'un règlement FRIBS2 portant sur ces domaines a été un temps envisagée, mais ne semble plus d'actualité.

Compte tenu des nombreux domaines couverts et du caractère très transversal du règlement FRIBS, ce dernier concerne plusieurs services statistiques en France et dépasse largement le périmètre de l'Insee. Sont ainsi concernés, outre l'Insee, plusieurs services statistiques ministériels : agriculture⁶, commerce extérieur⁷, développement durable⁸, enseignement supérieur, recherche et innovation⁹, travail¹⁰, ainsi que la Banque de France.

● ... ET VISE À UNE PLUS GRANDE COHÉRENCE ENTRE DOMAINES

L'approche actuelle en tuyaux d'orgue (*figure 1*) posait notamment des questions de cohérence entre domaines et de redondances. Par exemple, l'unité statistique n'était pas la même selon les domaines : elle peut être principalement l'Unité d'activité économique ou l'Entreprise (*encadré 2*) ; la définition d'un concept (par exemple, le chiffre d'affaires) n'était pas forcément la même d'un domaine à l'autre, ni le champ couvert ou les ventilations demandées (par activité économique ou tranche de taille). Une même donnée pouvait être demandée dans deux règlements différents, et conduire à des réponses différentes de la part des États membres.

Un exemple typique est celui du nombre d'entreprises, demandé d'une part dans le cadre des statistiques structurelles et d'autre part dans le cadre de la démographie d'entreprises. Ces incohérences limitaient également les possibilités de *data linking*.

Davantage qu'une harmonisation entre pays ou une harmonisation par les *inputs*, le règlement FRIBS vise ainsi surtout à accroître la cohérence entre domaines et le rôle des référentiels communs que sont les nomenclatures et les répertoires d'entreprises, ainsi qu'à développer les statistiques sur certains sujets ou secteurs : la mondialisation, dans la suite du Memorandum de Riga adopté par le Comité du système statistique européen (CSSE, 2014), les groupes, le secteur tertiaire (commerce et services).

6. Service de la Statistique et de la prospective (SSP).

7. Département des Statistiques et des études du commerce extérieur (DSECE).

8. Service de la Donnée et des études statistiques (SDES).

9. Sous-direction des Systèmes d'information et des études statistiques (Sies).

10. Direction de l'animation, de la recherche et des études statistiques (Dares).

Le règlement FRIBS et ses mesures d'application définissent des données à transmettre, des *outputs* (variables agrégées, indices...). Pour les calculer, les instituts nationaux de statistique sont libres de mobiliser plusieurs types de sources de données, précisés dans le règlement-cadre, sous réserve que les résultats remplissent des critères de qualité : des enquêtes, des fichiers administratifs, des échanges de micro-données, ainsi que toute autre source, méthode ou approche innovante pertinente. Par exemple, aucun questionnaire d'enquête n'est fixé par FRIBS, même si pour les sujets tels que les TIC, l'innovation ou les chaînes de valeur mondiales, les variables demandées dans les textes d'application sont si précises qu'elles se

« Le règlement FRIBS et ses mesures d'application définissent des données à transmettre, des outputs. »

rapprochent de questions d'enquêtes¹¹. Mais ce cas est minoritaire. Il n'est pas non plus exigé que telle ou telle information soit recueillie par enquête directe auprès des entreprises, ni que les enquêtes soient, si enquête il y a, obligatoires. Les données à transmettre sont, sauf exception, des agrégats. Les seules exceptions concernent les échanges de micro-données obligatoires, qui portent d'une part sur les répertoires d'entreprises et le répertoire européen des groupes, d'autre part sur le commerce international de biens au sein de l'Union européenne.

Une disposition a également été introduite pour autoriser des échanges de données confidentielles, collectées ou calculées dans le cadre de FRIBS, entre les autorités statistiques nationales des États membres, leurs banques centrales, la Banque centrale européenne et Eurostat, « afin de préserver la qualité et la comparabilité des statistiques d'entreprises européennes ou des comptes nationaux ». Les modalités d'usage effectif de cette disposition, introduite au départ pour autoriser des échanges de données confidentielles sur les multinationales et qualifiée à ce stade de « clause d'habilitation », restent à préciser.

Le règlement-cadre crée donc un cadre juridique commun pour les statistiques d'entreprises. Au-delà, il introduit plusieurs évolutions et innovations, les principales figurant dans les **encadrés 3 et 4**.

CE QUE FRIBS VA CHANGER CONCRÈTEMENT EN FRANCE

La France a largement soutenu le projet FRIBS au fil de sa préparation¹². L'Insee et le SSM Commerce extérieur ont aussi été très actifs dans les projets et les ESSnet qui en ont construit certaines briques, notamment sur le profilage des grands groupes multinationaux, sur le cadre européen des répertoires, sur les échanges de biens intra-UE.

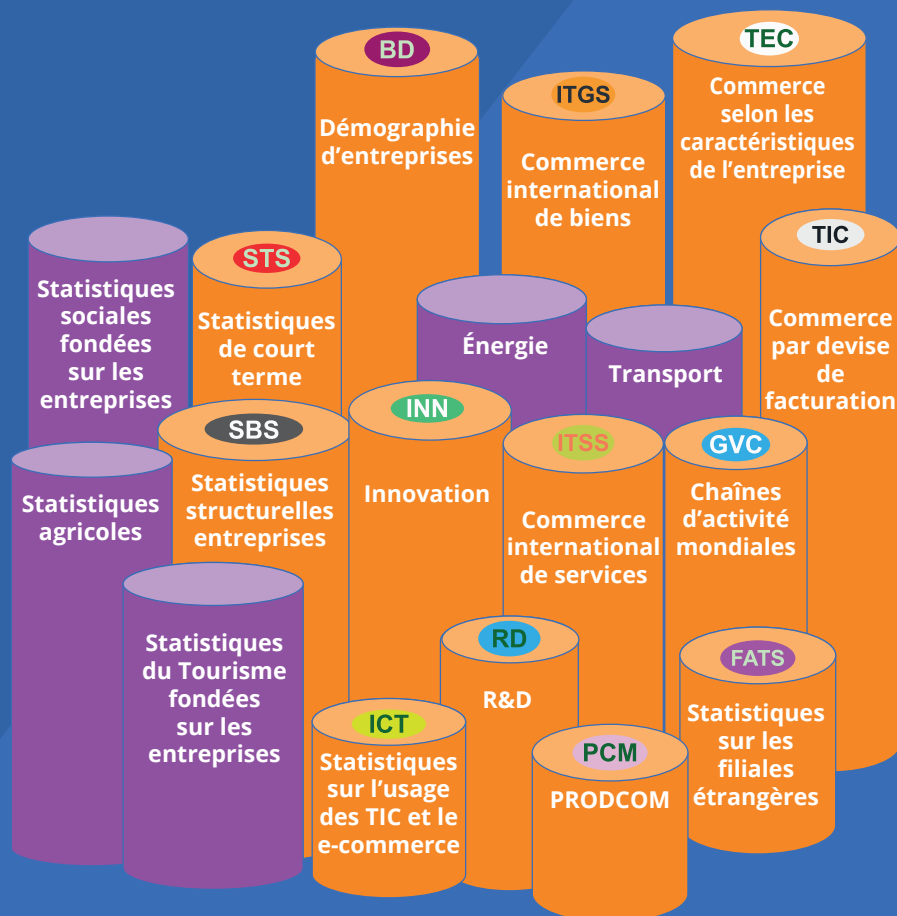
De nombreuses évolutions demandées dans le cadre de FRIBS ont par ailleurs pu être anticipées par la statistique publique française et mises en œuvre au cours des années récentes. Si dans certains cas le système d'observation français permettait déjà de répondre aux

11. Sur ces sujets, les groupes de travail européens établissent également des « questionnaires harmonisés » qui servent de fil directeur pour la réalisation des enquêtes.

12. Voir à ce sujet l'article d'Hervé Piffeteau dans ce même numéro.

Figure. 1 Les domaines couverts par FRIBS

D'une approche en tuyaux d'orgue...



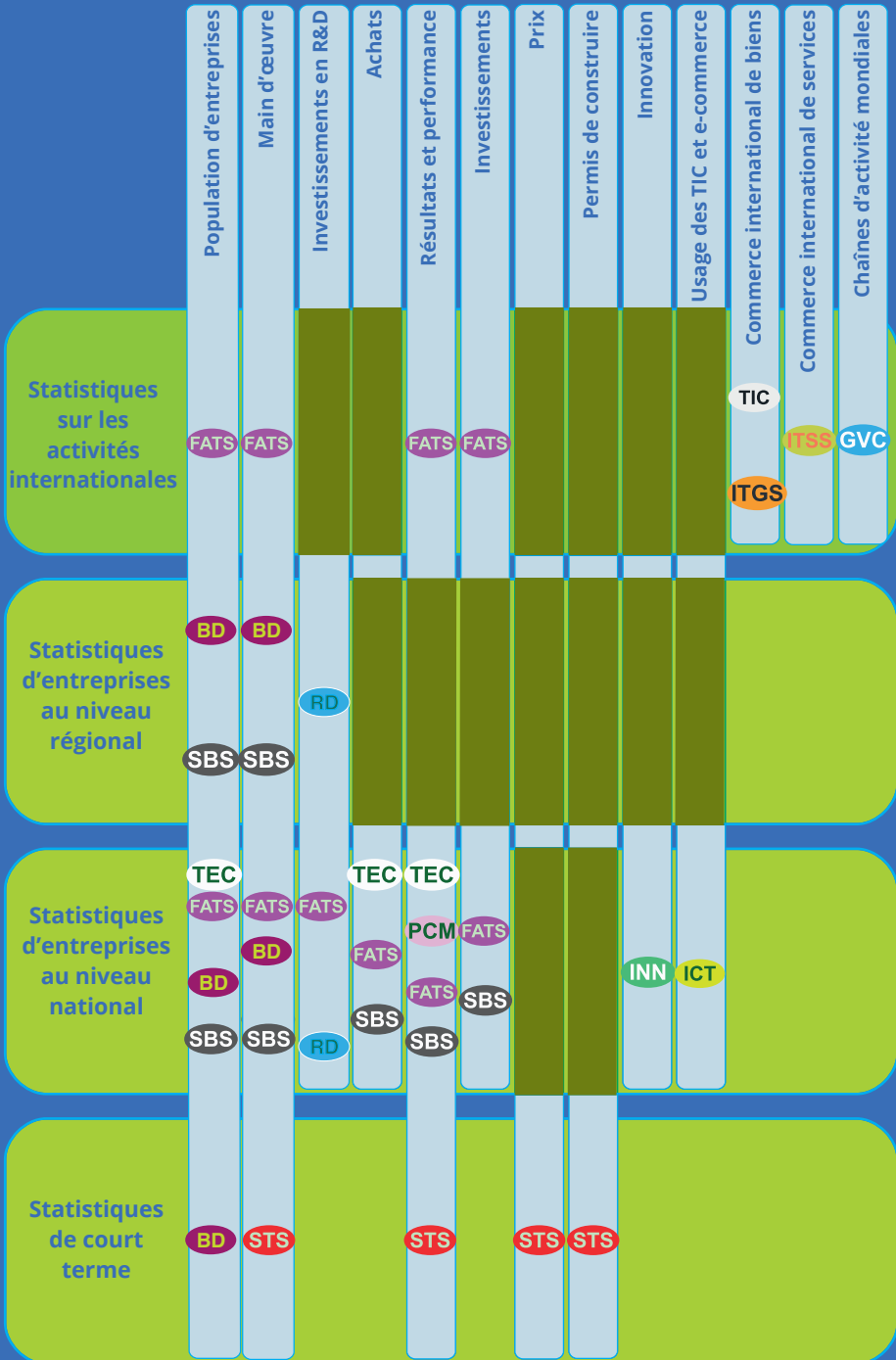
Légende :



Source : Eurostat

- BD : *Business demography* ou Démographie d'entreprises
- FATS : *Foreign Affiliates Statistics* ou Statistiques sur les filières étrangères
- GVC : *Global Value Chains Statistics* ou Statistiques sur les chaînes d'activité mondiales
- ICT : *ICT Statistics* ou Statistiques sur l'usage des TIC et le e-commerce
- INN : *Innovation Statistics* ou Statistiques sur l'innovation
- ITGS : *International Trade in Goods Statistics* ou Statistiques sur le commerce international de biens
- ITSS : *International Trade in Services Statistics* ou statistiques sur le commerce international de services
- PCM : *Prodcom* ou Production communautaire
- RD : *Research & Development Statistics* ou Statistiques sur la R&D
- SBS : *Structural Business Statistics* ou Statistiques structurelles sur les entreprises
- STS : *Short Term Statistics* ou Statistiques de court terme
- TEC : *Trade by enterprise characteristics* ou Statistiques sur le commerce selon les caractéristiques de l'entreprise
- TIC : *Trade by invoicing currency* ou Statistiques sur le commerce par devise de facturation

... à une approche intégrée



Encadré 2. Les unités statistiques à utiliser au niveau européen sont en partie modifiées

FRIBS retient l'Unité d'Activité Économique dans tous les secteurs pour les statistiques conjoncturelles sur les prix, la production, les chiffres d'affaires et l'emploi (ainsi que pour celles de production industrielle)^(a) ; dans le règlement de 1998 sur les statistiques conjoncturelles, l'unité statistique était soit l'Unité d'Activité Économique, soit l'Entreprise, selon les secteurs d'activité.

L'Entreprise est l'unité statistique pour les statistiques structurelles, y compris la démographie structurelle des entreprises, ou celles sur les TIC, la R&D (pour le secteur privé), l'innovation, les chaînes d'activités mondiales. Formellement, c'est déjà le cas dans les règlements sectoriels actuels. Toutefois, la mise en œuvre n'était pas effective, au sens où les INS considéraient que l'Entreprise était l'unité légale, ce qui est de moins en moins pertinent avec le développement des organisations en groupes de sociétés.

En France, c'est à partir de l'exercice 2017 que les statistiques structurelles (issues du dispositif Ésane) seront diffusées en Entreprises sur l'ensemble du champ^(b) (Haag, 2019). La plupart des autres pays^(c) le feront sur l'exercice 2018. Pour la démographie structurelle et les enquêtes « thématiques » (TIC, Innovation...), la mise en œuvre interviendra dans le cadre du règlement FRIBS.

Les unités statistiques pour produire les résultats restent définies par le règlement de 1993^(d).

L'Unité d'activité économique (UAE) regroupe au sein d'une entreprise l'ensemble des parties qui concourent à l'exercice d'une activité du niveau classe (quatre chiffres) de la Nace. Cette entité correspond à une ou plusieurs subdivisions opérationnelles de l'entreprise, qui doit pouvoir fournir pour chaque UAE au moins la valeur de la production, des consommations intermédiaires, des frais de personnel, de l'excédent brut d'exploitation, ainsi que l'emploi et la formation brute de capital fixe.

L'Entreprise correspond à la plus petite combinaison d'unités légales qui constitue une unité organisationnelle de production de biens et de services, jouissant d'une certaine autonomie de décision notamment pour l'affectation de ses ressources courantes. Une Entreprise exerce une ou plusieurs activités dans un ou plusieurs lieux. Une Entreprise peut correspondre à une seule unité légale.

L'Unité institutionnelle^(e) est un centre élémentaire de décision économique, *i.e.* une unité qui jouit d'une autonomie de décision dans l'exercice de sa fonction principale (elle est responsable et redevable des décisions et des actions qu'elle prend) et dispose d'une comptabilité complète (à la fois de documents comptables où apparaît la totalité de ses opérations, économiques et financières, effectuées au cours de la période de référence des comptes, et d'un bilan de ses actifs et de ses passifs).

Le Groupe d'entreprises rassemble des entreprises tenues par des liens juridico-financiers. Il peut comporter une pluralité de centres de décision, notamment en ce qui concerne la politique de production, de vente et des bénéfices ; il peut unifier certains aspects de la gestion financière et de la fiscalité. Il constitue une entité économique qui peut effectuer des choix qui concernent les unités alliées qui le composent.

L'Unité Locale correspond à une Entreprise ou une partie d'Entreprise (atelier, usine, magasin, bureau, mine, entrepôt) sise en un lieu topographiquement identifié. Dans ce lieu, ou à partir de ce lieu, sont exercées des activités économiques pour lesquelles, sauf exception, une ou plusieurs personnes travaillent pour le compte d'une même Entreprise.

(a) Les statistiques conjoncturelles sur les créations (immatriculations) et défaillances d'entreprises, nouvelles dans FRIBS, sont en revanche demandées en unité légale (unité juridique).

(b) Comme l'Italie.

(c) Comme l'Allemagne.

(d) Ces unités peuvent ou non être identiques aux unités de collecte. Voir le Règlement (CEE) n° 696/93 du Conseil, du 15 mars 1993, relatif aux unités statistiques d'observation et d'analyse du système productif dans la Communauté (http://admi.net/eur/loi/leg_euro/fr_393R0696.html).

(e) L'Unité Institutionnelle est utilisée en comptabilité nationale. Elle ne doit pas être confondue avec l'Unité légale constituée toujours, seule ou parfois avec d'autres unités légales, le support juridique de l'unité statistique « Entreprise ».

nouvelles demandes, les évolutions au niveau européen ont également constitué une impulsion décisive pour enrichir et compléter la gamme d'indicateurs au niveau national, en avance de phase sur la mise en œuvre du règlement-cadre. Tel est notamment le cas pour les nouveaux indicateurs conjoncturels de volume des ventes dans le commerce et de production dans les services, ou pour les extensions de champ et changements de périodicité des indicateurs conjoncturels (Scherrer, 2019).

Certaines adaptations restent cependant à mener au niveau français. Parmi les plus structurantes, on peut citer le passage à une diffusion effective selon le concept d'Entreprise pour les statistiques sur les TIC, l'innovation, la R&D, les chaînes d'activité mondiales, et la mise en place de l'échange de micro-données sur le commerce extérieur de biens intra-communautaire.

La diffusion en Entreprise nécessite encore des analyses sur le protocole le plus adapté : faut-il collecter directement à ce niveau et dans ce cas, quelle(s) unité(s) interroger pour recueillir une information pertinente et de qualité ? Faut-il interroger des sociétés au sein des Entreprises et agréger ensuite leurs réponses, dans un contexte où les données recueillies sont principalement des variables qualitatives¹³ ? Des travaux sont en cours à l'Insee et au Sies, en s'appuyant notamment sur des expérimentations menées au cours des dernières années.

Sur le second sujet, de nombreux points sont encore en cours de définition ou de réglage dans le cadre des actes délégués ou d'exécution complémentaires au règlement-cadre ; une infrastructure d'échanges sécurisée doit être mise en place, s'agissant de données confidentielles très détaillées sur les exportations intra-UE¹⁴, et chaque pays souhaitant recevoir les données individuelles des autres États-membres doit être certifié ; la qualité des statistiques basées sur l'échange de micro-données n'a pas encore été évaluée de manière approfondie. De cette qualité dépendra l'usage qui pourra en être fait effectivement, pour éventuellement compléter ou remplacer la collecte nationale des statistiques d'importations intra-communautaires. Sur les échanges de services également, le détail des demandes de données dans le cadre des textes d'application de FRIBS n'est pas encore acté. Selon ce qui sera finalement retenu, des travaux complémentaires conséquents pourraient être nécessaires côté Banque de France. Certains modes de collaboration entre producteurs de données pourraient aussi être amenés à évoluer. Par exemple, de nouveaux agrégats sur les résultats économiques des entreprises cumulant des données produites par l'Insee d'une part sur les secteurs marchands non financiers et par la Banque de France d'autre part sur le secteur financier, seront à transmettre, nécessitant une coopération accrue entre ces services.

UNE MISE EN ŒUVRE PROGRESSIVE, NÉCESSITANT DES TEXTES COMPLÉMENTAIRES

Le règlement-cadre fixe le cadre général et les éléments essentiels. De nombreux points seront toutefois précisés dans des actes délégués et actes d'exécution¹⁵. Ainsi, le règlement-cadre définit pour les statistiques de niveau national, régional ou sur les activités

13. C'est ce qui est fait pour l'enquête sectorielle annuelle, hors groupes faisant l'objet d'un profilage en face-à-face, mais cette enquête recueille des informations quantitatives.

14. Ces échanges de micro-données ont bénéficié du projet ESDEN (*European Statistical Data Exchange Network de la Vision 2020*), qui modernise les services d'échanges de données au sein du SSE, en étend le volume et améliore l'interopérabilité et la sécurité.

15. Voir l'article d'Hervé Piffeteau dans ce même numéro.

internationales, une liste de thèmes généraux (par exemple pour les statistiques de court terme : les prix, la production et la performance, la main-d'œuvre, etc.) et de thèmes détaillés (les prix à la production ; les prix à l'importation, etc.). Mais l'architecture légale retenue prévoit la possibilité que les variables, les demandes de données puissent être définies (et modifiées) à un niveau encore plus détaillé par des actes d'exécution. De même, la périodicité des demandes de données est fixée par le règlement-cadre, mais le champ ou les ventilations demandées le seront dans d'autres textes.

Dans le cas de FRIBS, tous les textes complétant le règlement-cadre ont été préparés en parallèle à celui-ci. Ils constituent le « *FRIBS package* », qui a déjà été largement examiné dans les différents groupes de travail thématiques et par le groupe des directeurs de statistiques d'entreprises, même si certains points, principalement sur le commerce international, restent à préciser.

Encadré 3. Au-delà de l'approche intégrée, quelques nouveautés introduites par FRIBS

- **De nouvelles demandes de données et de nouveaux indicateurs** : un indice mensuel de la production dans les services, un indice mensuel de volume des ventes pour l'ensemble du commerce (Scherrer, 2019), un enrichissement de la démographie d'entreprises avec des ventilations par région, des données trimestrielles sur les créations et défaillances d'entreprises (démographie conjoncturelle), des données sur les gazelles* et entreprises à forte croissance, désormais obligatoires, des données provisoires à dix mois sur les PME pour la démographie d'entreprises et les principales caractéristiques structurelles, de nouvelles variables obligatoires pour les statistiques sur les filiales des groupes français à l'étranger (frais de personnel et investissements corporels).
- **Le rôle des répertoires statistiques d'entreprises est renforcé**, qu'il s'agisse des répertoires nationaux ou du répertoire européen des groupes, l'EGR. Ces répertoires sont reconnus comme « source faisant autorité » (« *authoritative source* ») pour la définition de populations (à des fins d'échantillonnage par exemple). Le contenu des échanges de micro-données entre les autorités statistiques nationales des différents pays, *via* Eurostat, et entre les pays et Eurostat est enrichi, par exemple de données démographiques sur les groupes et entreprises.
- **Un échange obligatoire de micro-données sur le commerce intra-Union européenne de biens est mis en place** : chaque État membre doit transmettre à l'ensemble des autres États membres des données individuelles d'exportations, qui pourront (mais il n'y a pas d'obligation) être utilisées comme « données miroir » par chaque pays partenaire pour établir ses statistiques d'importations. Ce point a été tranché tardivement dans la préparation du règlement FRIBS, car il fallait attendre les résultats des projets SIMSTAT et REDESIGN. C'est au CSSE de novembre 2016 que les directeurs généraux des INS ont finalement validé l'inclusion de ces échanges dans FRIBS, y compris de la transmission de l'identifiant TVA de l'exportateur et du pays d'origine des biens. Ces derniers points ont fait l'objet de vives réactions de la part de certains États membres particulièrement soucieux de ne pas alourdir la charge de déclaration des entreprises. Les travaux de l'ESSnet sur SIMSTAT avaient toutefois montré que ces variables étaient indispensables pour la qualité des statistiques pouvant être construites sur la base des micro-données échangées.

* Le terme de gazelles a été utilisé par David Birch, l'inventeur de la démographie des entreprises, chercheur au MIT, autour des années 1980 pour désigner les jeunes entreprises à croissance rapide.

« La périodicité des demandes de données est fixée par le règlement-cadre, mais le champ ou les ventilations demandées le seront dans d'autres textes. »

Le règlement-cadre prévoit une mise en œuvre des dispositions de FRIBS à compter de l'année de référence 2021, sauf pour le commerce international de biens intra-communautaire (2022). Pour les statistiques conjoncturelles, si la mise en œuvre démarre sur l'année 2021, elle ne sera complète qu'en 2024, un recul minimum étant nécessaire pour

calculer des données CVS pour les nouveaux indicateurs. Comme pour les statistiques sociales, plusieurs appels à projets lancés par Eurostat ont vocation à soutenir financièrement les pays dans leur mise en œuvre des nouveautés induites par FRIBS. La France n'a pour l'instant pas répondu positivement à ceux qui ont été lancés, car ils concernaient des nouveautés qu'elle avait déjà mises en œuvre.

📍 IN FINE, FRIBS VA-T-IL PERMETTRE PLUS DE COHÉRENCE ?

Si FRIBS permet de compléter l'offre statistique, un de ses principaux objectifs était aussi d'accroître la cohérence entre domaines. De fait, lorsque trois définitions différentes sont remplacées par une seule, ou bien lorsque les pays doivent fournir un seul « nombre d'entreprises », ou bien lorsque les tranches de taille sont harmonisées selon les domaines, la cohérence s'en trouve accrue. Les possibilités de lier des données relatives à différents domaines seront aussi élargies.

Des progrès restent cependant possibles :

- ❶ par exemple, les unités statistiques ne sont pas complètement harmonisées entre domaines. Et si tous les pays vont effectivement diffuser des données sur les résultats économiques selon le concept d'Entreprise, la manière de constituer ces Entreprises, leur granularité et la façon d'établir leurs comptes consolidés peuvent varier selon les pays, malgré un manuel opérationnel de recommandations¹⁶ ;
- ❶ pour les domaines couverts le plus souvent par voie d'enquêtes, comme l'innovation, les caractéristiques de celles-ci (taux de sondage, caractère obligatoire ou non, critères d'arrêt, etc.) peuvent varier sensiblement et affecter la comparabilité des résultats (Duc, Ralle, 2019) ;
- ❶ la cohérence entre les statistiques sur les échanges extérieurs, de biens et de services, et les autres statistiques d'entreprises est également perfectible. Par exemple, les échanges de services continuent de s'appuyer sur la nomenclature EBoPS¹⁷ utilisée pour la balance des paiements, qui n'est pas bien articulée avec la nomenclature européenne de produits, la CPA¹⁸. Des ventilations en CPA devraient toutefois être demandées dans le cadre de FRIBS sur une base volontaire ;
- ❶ du côté des échanges de biens, le champ des opérateurs n'est pas complètement cohérent avec ce qui figure dans les répertoires statistiques d'entreprises.

Ainsi, la cohérence et la comparabilité seront accrues, mais FRIBS ne règle pas l'ensemble de ces questions.

16. Par exemple, la méthode italienne conduit à davantage découper les groupes pour constituer les entreprises que la méthode française (Alonzi, Migliardo, 2018 ; Haag, 2019). Eurostat considère cependant que les pratiques de profilage sont assez homogènes entre pays (Eurostat, 2019).

17. EBoPS : *Extended Balance of Payments Services*.

18. CPA : Classification statistique des Produits associée aux Activités.

FRIBS VA-T-IL PERMETTRE UNE MEILLEURE RÉPONSE AUX BESOINS DES UTILISATEURS ?

FRIBS intègre déjà des demandes qui avaient été émises par les utilisateurs, principalement les directions politiques de la Commission européenne ou la Banque centrale européenne (BCE), notamment sur la mondialisation et les groupes, sur les indicateurs conjoncturels.

Pour les évolutions futures, le règlement-cadre comporte des clauses de sauvegarde (ou garde-fous) visant à contenir et encadrer le nombre de changements possibles, notamment sur les thèmes détaillés et les variables, sur des périodes glissantes de cinq ans.

Ces clauses de sauvegarde, ainsi que les instruments juridiques permettant de faire évoluer les demandes de données, ont été des points durs de négociation entre Eurostat et les INS tout au long du processus de préparation de FRIBS. Le Parlement européen et la Commission poussaient vers un élargissement du champ des actes délégués, qui aurait pu avoir pour conséquence un accroissement des coûts pour les INS et les entreprises. Un compromis a donc été trouvé, afin de tenter de concilier les objectifs de souplesse et d'adaptabilité aux évolutions de la demande et de garantie de maîtrise des coûts.

In fine des évolutions des variables et des thèmes détaillés resteront possibles, mais encadrées. Elles seront plus faciles pour certains thèmes. La recherche d'un compromis a en effet conduit Eurostat et les directeurs de statistiques d'entreprises à proposer en 2015 de distinguer des

Encadré 4. Au-delà de l'approche intégrée, les principales évolutions introduites par FRIBS sur la périodicité et le champ

- **Une périodicité accrue et un raccourcissement des délais** de fourniture des données pour certains indicateurs. Les indices de chiffre d'affaires dans le commerce de gros et le commerce et réparation automobile, ainsi que dans les services, deviennent mensuels et non plus trimestriels.
- **Les statistiques sur les chaînes d'activités mondiales sont désormais demandées à titre obligatoire** pour tous les pays et tous les trois ans.
- **Des extensions de champ, notamment dans le secteur tertiaire** : les statistiques structurelles, de démographie et sur les filiales en France de groupes étrangers sont étendues aux activités marchandes de l'enseignement et de la santé-action sociale, ainsi qu'aux services aux ménages. La couverture de l'indice trimestriel des prix de production dans les services ou bien des données sur la main-d'œuvre est également étendue dans les services. Ces extensions visent d'une part une plus grande cohérence entre les champs couverts selon les domaines, d'autre part à mieux couvrir le secteur tertiaire qui représente une part croissante de la valeur ajoutée des économies européennes.
- **Des simplifications et alignements de champ et de ventilations** : certaines demandes de données sont supprimées, par exemple la ventilation du chiffre d'affaires par produit dans le commerce pour les statistiques structurelles, ou bien les dépenses pour protéger l'environnement. Ces dernières restent toutefois nécessaires pour d'autres usages (comptes de l'environnement) et sont demandées dans d'autres cadres. Afin d'accroître la cohérence entre domaines et simplifier les demandes de données, les ventilations sont par ailleurs rationalisées : par exemple, alignement des statistiques sur la R&D et l'innovation sur les ventilations des autres domaines en termes de taille et de secteur d'activité ainsi que sur les concepts d'effectifs, alignement des indicateurs conjoncturels du secteur de la construction sur les postes de la nomenclature européenne d'activités, la Nace.

« *dynamic statistics* » (TIC, Innovation, Chaînes de valeur mondiales) et des « *stable statistics* ». Les premières ont des clauses de sauvegarde moins strictes que les secondes et offrent donc une plus grande flexibilité. C'est pour les répertoires que les possibilités d'évolution sont les plus limitées, les pays ayant dans leur grande majorité soutenu le caractère de colonne vertébrale de ces répertoires, qui demande une grande stabilité.

« *Un compromis a donc été trouvé, afin de tenter de concilier les objectifs de souplesse et d'adaptabilité aux évolutions de la demande et de garantie de maîtrise des coûts.* »

Les évolutions à venir passeront aussi par des études pilotes. Des thèmes sont d'ores et déjà identifiés dans le règlement-cadre. Ils correspondent à des demandes fortes des utilisateurs : l'immobilier, avec le sujet nouveau de l'immobilier commercial, les échanges internationaux de services (demandes de

la Commission ou de la BCE, des indicateurs financiers sur les entreprises, le climat et l'environnement (demande du Parlement européen). On peut donc s'attendre à ce que des études pilotes sur ces thèmes soient lancées et débouchent sur de nouvelles demandes de données dans les années à venir. Le processus est d'ailleurs déjà lancé sur l'immobilier commercial et sur les échanges de services par caractéristiques des entreprises et par mode d'offre.

FRIBS VA-T-IL PERMETTRE DE RÉDUIRE LA CHARGE ?

FRIBS encourage l'usage de sources de données autres que les enquêtes, et supprime les demandes multiples de données identiques. Dans le même temps, de nouvelles demandes de données ont été intégrées. Au final, la baisse de charge pour les INS comme pour les entreprises ne paraît pas flagrante. Un des axes sur lesquels une baisse de charge significative était annoncée est celui du commerce international de biens intra-UE. Une légère baisse du taux de couverture a été décidée, mais la réduction de charge ne sera significative que si la mise en place des échanges de micro-données permet d'alléger la collecte nationale. Sur ce sujet, tout reste encore à prouver...

BIBLIOGRAPHIE

ALONZI, Francesca et MIGLIARDO, Serena, 2018. The implementation of ISTAT automatic profiling to support the delineation of complex enterprises. *26th Meeting of the Wiesbaden Group on Business Registers – Profiling and Globalisation*. [en ligne]. 24-27 septembre 2018. Neuchâtel. [Consulté le 4 novembre 2019]. Disponible à l'adresse : http://www.wiesbaden2018.bfs.admin.ch/wp-content/uploads/2018/08/Paper_Alonzi2.pdf

COMMITTEE ON MONETARY, FINANCIAL AND BALANCE OF PAYMENTS STATISTICS (CMFB), 2016. CMFB opinion on re-shaping the connection between business statistics and BOP-IIP via the FRIBS draft Regulation. [en ligne]. 12 septembre 2016. [Consulté le 4 novembre 2019]. Disponible à l'adresse : https://circabc.europa.eu/sd/a/1fbb5839-ee7d-442c-a216-1cdb1e098962/2016-09-12_%20-CMFB%20opinion%20on%20re-shaping%20the%20connection%20between%20business%20statistics%20and%20BOP-IIP.pdf

CSSE, 2014. *Riga Memorandum towards better measurement of the globalised economy*. [en ligne]. 26 septembre 2014. [Consulté le 4 novembre 2019]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/documents/7330775/7339365/DGINS+memorandum+2014/b425bd92-e61b-44e9-bf20-c6a400656f5e>

DUC, Cindy et RALLE, Pierre, 2019. Une certaine convergence de l'innovation dans les entreprises en Europe. In : *Les entreprises en France*, édition 2019. [en ligne]. 3 décembre 2019. Insee Références, Dossiers, pp. 85-100. [Consulté le 3 décembre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/4255795/ENTFRA19_D3.pdf

EUROSTAT, 2009. *Communication de la Commission au Parlement européen et au Conseil concernant la méthode de production des statistiques de l'Union européenne : une vision de la prochaine décennie*. [en ligne]. 10 août 2009. N° COM/2009/0404. [Consulté le 4 novembre 2019]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52009DC0404&from=FR>

EUROSTAT, 2012. Projet en vue de l'établissement d'un règlement-cadre relatif à l'intégration des statistiques d'entreprises : rapport d'avancement. In : *Rapport pour le CSSE de novembre 2012*. 15 novembre 2012. Luxembourg, 15^e réunion du Comité du système statistique européen, point 11 de l'ordre du jour

EUROSTAT, 2019. European Profiling to better measure Multinational Enterprise Groups' activities. *Conference of European Statisticians – Group of Experts on National Accounts – Eighteenth session*. [en ligne]. 10-12 avril 2019. Genève. [Consulté le 4 novembre 2019]. Disponible à l'adresse : https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2019/mtg1/Eurostat_new.pdf


HAAG, Olivier, 2019. Le profilage à l'Insee – Une identification plus pertinente des acteurs économiques. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 86-102. [Consulté le 4 novembre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168409/courstat-2-9.pdf>

SCHERRER, Philippe, 2019. Statistiques conjoncturelles d'entreprises – Des impulsions européennes, des avancées françaises. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 72-85. [Consulté le 4 novembre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168406/courstat-2-8.pdf>

IESS: L'EUROPE HARMONISE SES STATISTIQUES SOCIALES POUR MIEUX ÉCLAIRER LES POLITIQUES

Chantal Cases*

Le règlement-cadre IESS (Integrated European Social Statistics), adopté en avril 2019, est issu d'un long processus de modernisation des statistiques européennes. Pour soutenir des politiques sociales ambitieuses, les statistiques doivent toujours mieux répondre aux besoins des décideurs et du public, alliant comparabilité, précision, flexibilité, délai de diffusion, et un bon rapport coût-efficacité. Auparavant, chaque enquête découlait d'un règlement ou d'un accord informel spécifique. IESS impose un cadre cohérent et intégré pour l'ensemble des enquêtes, qui améliorera la comparabilité entre États membres. L'Insee en tire les conséquences, avec un programme de rénovation progressive de ses enquêtes sociales, dont une part plus importante sera désormais sous règlement européen. Au-delà des changements induits par IESS (programmation pluriannuelle, concepts communs, échantillonnage, questionnaires), d'autres transformations sont également à l'œuvre, dont le développement du multimode. Après investissement, les coûts de production de ces données devraient être mieux maîtrisés, pour peu que soient développés des outils adaptés. Enfin, il faudra être attentif à un public plus large que les seuls décideurs, autant en termes d'utilité collective que de charge de réponse aux enquêtes.

 *The Integrated European Social Statistics (IESS) regulation, adopted in April 2019, results from a long process of modernising European social statistics. In order to support challenging social policies, statistics must better meet the needs of decision-makers and the public, combining comparability, accuracy, flexibility, timeliness and cost-effectiveness. Previously, each social household surveys was based on a specific regulation or gentlemen's agreement. IESS imposes a coherent and integrated framework, which will improve comparability between Member States. As a result, INSEE is gradually implementing a programme to renovate its social surveys. A larger proportion of them will now be subject to European regulations. In addition to the changes induced by IESS (multi-annual programming, common concepts, sampling, questionnaires), other transformations are also at work, including the development of multimode surveys. After the investment period, the costs of producing this data should be better controlled, provided that appropriate tools are developed. Finally, it will be necessary to be attentive to the expectations of a wider public than just decision-makers, both in terms of collective utility and the response burden to surveys.*

* Directrice des statistiques démographiques et sociales, Insee, chantal.cases@insee.fr

Pauvreté, jeunes sortant du système éducatif sans diplôme, taux de chômage... les indicateurs sociaux sont largement utilisés aussi bien par les décideurs que les chercheurs ou le grand public, pour comprendre et analyser l'évolution de la société européenne dans le temps et dans l'espace, pour fixer des cibles pour l'action. Leur pertinence et leur qualité, notamment en termes de comparabilité entre pays, mais aussi de cohérence, de précision ou de délai de mise à disposition sont donc d'importance. Le système statistique européen de statistiques sociales qui est la base de leur production s'est construit progressivement, thème par thème, et en fixant d'abord des objectifs en termes de variables à produire dans le cadre de règlements ou d'accords informels¹ spécifiques à chaque enquête. L'enjeu du nouveau règlement-cadre sur les statistiques sociales intégrées était d'harmoniser davantage la production de statistiques sociales en amont, entre les différentes enquêtes sociales collectées par les instituts nationaux de statistique, au niveau des concepts et parfois des méthodes employées.

UN PEU D'HISTOIRE, OU LE LONG CHEMIN D'UN RÈGLEMENT-CADRE

Le règlement-cadre IESS (*Integrated European Social Statistics*, ou statistiques sociales européennes intégrées), adopté en avril 2019, est le premier résultat d'un long processus de modernisation des statistiques sociales européennes entamé à la fin des années 2000 dans le cadre de la « Vision 2020 », la stratégie européenne pour 2020.

« *Le premier résultat d'un long processus de modernisation des statistiques sociales européennes entamé à la fin des années 2000.* »

En 2009, l'adoption de la « loi statistique européenne » fixait les principes de base et les règles de fonctionnement du système statistique européen (Eurostat, 2010) ; un premier essai d'outillage de la programmation des enquêtes sociales avait alors également été élaboré. La voie était ouverte à la préparation des règlements-cadres

devant structurer la production de pans entiers du système statistique européen rénové². La réflexion sur une mise en cohérence renforcée du contenu et des méthodes des deux principaux « piliers » des enquêtes sociales (l'enquête sur les Forces de travail³ et le panel sur les Ressources et les conditions de vie⁴), amorce de la vision « intégrée » qui allait prévaloir dans IESS, démarrait dès 2010.

En 2011, les directeurs généraux des instituts nationaux de statistique adoptaient, sur proposition d'Eurostat, un projet de modernisation des statistiques sociales destiné à soutenir des politiques sociales ambitieuses (Eurostat, 2011). Ce projet visait également à présenter un bon rapport coût-efficacité et une bonne flexibilité, afin de s'adapter rapidement aux nouveaux besoins en maintenant une production statistique de qualité dans un contexte général de tensions sur les moyens budgétaires et de concurrence accrue dans la production de données.

1. Ces *gentlemen's agreements* font l'objet d'un texte écrit mais n'engagent que les pays volontaires.

2. Voir l'article d'Hervé Piffeteau sur le triptyque juridique de la statistique européenne dans ce même numéro.

3. Dénommée « enquête Emploi » en France. Dans la suite du papier, on utilisera indifféremment les deux dénominations.

4. Dénommé « statistiques sur les Ressources et les conditions de vie » (SRCV) en France et *European Union Statistics on Income and Living Conditions* (EU-SILC) au niveau européen.

Ce mémorandum dit « de Wiesbaden » fixait les bases de IESS : une architecture commune et standardisée pour les statistiques sociales, autour d'une série d'enquêtes menées périodiquement auprès d'échantillons de personnes ou de ménages, complétées par un meilleur usage de données administratives. La nécessité de disposer de données de qualité sur les conditions de vie de sous-populations particulières (enfants, migrants, personnes âgées), d'enquêtes Emploi du temps et Budget de famille, le renforcement des liens avec les Comptes nationaux étaient spécifiquement évoqués. Ainsi que la nécessité de disposer de financements adéquats, à la hauteur des investissements nécessaires en ces temps de budgets contraints...

Le programme de modernisation dans son ensemble comportait également un projet de règlement portant sur les recensements et les statistiques démographiques. Par ailleurs, à plus long terme, au-delà de 2020, un troisième règlement-cadre était envisagé pour

« Le processus de co-législation, structurellement long puisqu'il doit atteindre un consensus entre le Parlement et le Conseil sur proposition de la Commission, aura duré finalement plusieurs années. »

fournir une base légale à la production statistique issue de données administratives dans des domaines variés (emploi, santé, protection sociale, éducation, criminalité etc.) ainsi que les comptes sectoriels.

Ces bases du programme de modernisation furent adoptées par le comité des directeurs généraux des instituts nationaux de statistique fin 2014, et le processus de préparation formelle du règlement lancé. Il incluait notam-

ment une consultation publique sur internet, menée en 2015. Le processus de co-législation, structurellement long puisqu'il doit atteindre un consensus entre le Parlement et le Conseil sur proposition de la Commission, aura duré finalement plusieurs années, et le règlement fut adopté dans les derniers jours du mandat du Parlement Européen, en avril 2019, en vue d'une mise en œuvre effective à partir de 2021 (Parlement européen, 2019).

🌐 LES NOUVEAUTÉS DE IESS : UNE VISION GLOBALE DES STATISTIQUES SOCIALES...

Antérieurement à IESS, les statistiques sociales collectées auprès des ménages étaient organisées en tuyaux d'orgue : une enquête, un règlement. Il n'existait d'ailleurs pas toujours de texte législatif : certaines enquêtes étaient transmises à Eurostat sur une base volontaire par ceux des États membres qui les collectaient.

C'était par exemple le cas des enquêtes sur les Budgets des familles ou sur l'Emploi du temps, mais aussi de la première édition des enquêtes Santé, collectée entre 2006 et 2009 selon les pays. De plus, le système d'enquêtes s'est construit progressivement, le plus ancien règlement concernant l'enquête Emploi (1998). La cohérence entre les différents textes ne pouvait qu'en souffrir.

Finalement, le règlement-cadre remplacera ou amènera les règlements législatifs sur :

- 📍 les enquêtes Emploi (*Labour Force Survey*), ainsi que l'ensemble des mesures d'application (acte d'exécution, acte délégué) concernant chacun des modules thématiques complémentaires à l'enquête, (dits modules *ad hoc*) ;
- 📍 le panel sur les Ressources et les conditions de vie (SILC – *Statistics on Income and living conditions*) ;

- ① les enquêtes sur la formation des adultes (AES – *Adult education survey*) ;
- ① les enquêtes Santé ;
- ① et les enquêtes sur l’usage des technologies de l’information et de la communication (TIC).

Chaque enquête continuera néanmoins à faire l’objet de mesures d’application : un acte d’exécution spécifique, ainsi qu’un acte délégué⁵. La principale nouveauté de IESS est en effet de fournir un cadre cohérent et intégré pour l’ensemble des enquêtes, avec une programmation pluriannuelle globale et des concepts communs.

Au sein de ce cadre commun, le champ des thématiques incluses dans le règlement a été étendu. Il porte désormais sur sept domaines : le marché du travail, les ressources et conditions de vie, la santé, l’éducation et la formation, l’utilisation des technologies de l’information et de la communication, l’emploi du temps et la consommation. Les deux derniers n’avaient jamais fait l’objet d’actes réglementaires (ou de règlements) et, au sein du marché du travail, est inclus le support juridique pour la production et la publication de taux de chômage mensuels agrégés, pour lesquels les données étaient jusqu’alors transmises sur la base d’accords informels⁶. Même si certains thèmes importants ne sont pas couverts, comme le marché du logement ou la sécurité, et même si les enquêtes sur l’Emploi du temps sont encore incluses de manière optionnelle, le champ des statistiques sociales européennes s’en trouve notablement élargi.

Un programme pluriannuel d’enquêtes tournant sur huit ans a été défini. C’était une demande importante des États membres que de pouvoir disposer d’une vision pluriannuelle de leurs engagements et de s’assurer de pouvoir mettre en place en temps utile les organisations et les moyens nécessaires. Les collectes pluriannuelles (anciens modules *ad hoc* de l’enquête Emploi et du panel sur les Ressources et les conditions de vie) sont intégrées au programme et une grande partie de leurs thèmes a été définie à l’avance.

① ... UNE MISE EN COMMUN DES CONCEPTS..

Un vaste travail de définition et de standardisation des variables communes à différentes enquêtes a été réalisé. Il couvre une large liste de variables, qu’elles soient collectées dans toutes les enquêtes (variables de base) ou dans certaines d’entre elles seulement. Des manuels techniques précisent et détaillent le contenu de chacune d’entre elles. C’est une véritable avancée dans le sens d’une meilleure comparabilité entre pays et entre enquêtes, qui a fait l’objet d’un investissement collectif considérable en amont avec l’ensemble des pays dans le cadre des groupes de travail thématiques.

Le nouveau règlement et ses mesures d’application (ou textes complémentaires) introduisent également le concept de ménage économique (*housekeeping unit*), au sens de personnes vivant sous le même toit et partageant leurs ressources ou leurs dépenses. Les colocataires ne mettant pas en commun leurs ressources seront considérés comme des ménages séparés. Cette définition était déjà effective pour les enquêtes portant sur les revenus et la consommation, mais ne l’était pas dans d’autres, comme l’enquête Emploi. Elle est désormais présentée comme une cible pour l’ensemble des enquêtes, dont le champ reste celui des ménages ordinaires.

5. Pour plus de détails sur la différence entre les deux types d’actes, se reporter à l’article d’Hervé Piffeteau dans ce même numéro.

6. La France transmet les Demandeurs d’emplois en fin de mois. Eurostat produit sous sa propre responsabilité un taux de chômage mensuel sur la base d’une modélisation économétrique utilisant les inscrits à Pôle emploi et l’enquête Emploi. Ce système sera identique dans le nouveau règlement-cadre.

Pour ce qui concerne les contenus des enquêtes, les actes d'exécution définissent précisément les variables à fournir, leur périodicité ; les documents méthodologiques complémentaires aux textes législatifs en explicitent les définitions. Rappelons que ces enquêtes ne donnent pas seulement lieu à des transmissions d'agrégats : ce sont les micro-données individuelles, anonymisées, qui sont envoyées à Eurostat, et réutilisables par des chercheurs habilités, dans les conditions définies par la « loi statistique ». Dans la majorité des cas, le questionnement pour parvenir à ces variables reste, comme auparavant, de la responsabilité des États membres, de même que le mode de collecte de l'information – enquête directe par téléphone, face-à-face ou encore internet, données administratives ou même estimations sur petits domaines pour les données localisées – sous réserve de conformité aux critères de qualité définis dans la « loi statistique » (Règlement EC/223/2009, article 12) et détaillés dans le code de bonnes pratiques de la statistique européenne (Eurostat, CSSE, 2017).

... ET PLUSIEURS AUTRES HARMONISATIONS

Une enquête fait exception et va plus loin dans l'harmonisation par les inputs : l'enquête Emploi. Dans son cas, une structure commune et obligatoire de questionnaire (*flowchart*)

« Une enquête fait exception et va plus loin dans l'harmonisation par les inputs : l'enquête Emploi. »

a été élaborée pour le cœur de l'enquête, à savoir le module permettant de déterminer le statut d'activité au sens du Bureau International du Travail (BIT). L'utilisation de données administratives est ici proscrite. Auparavant, il était déjà nécessaire de collecter un « module BIT » par enquête, mais son contenu était de la responsabilité des États membres⁷. Cette construction est essentielle pour assurer une

meilleure comparabilité entre les différents pays, dans un domaine où les données administratives diffèrent conceptuellement des critères retenus par le BIT (Coder *et alii*, 2019).

L'annexe II du règlement-cadre définit de manière unifiée les contraintes de précision auxquelles doivent souscrire les différentes enquêtes, pour certains indicateurs phares (*encadré*). Certaines contraintes de précision ont été définies au niveau régional⁸. Des spécifications techniques précises sont données sur les bases de sondage à utiliser afin de produire des échantillons aléatoires sur l'ensemble du champ des ménages, et l'annexe III précise l'échantillonnage rotatif requis pour l'enquête sur les Forces de travail. La durée de rotation du panel sur les Ressources et les conditions de vie est également fixée, à 4 ans minimum.

L'annexe V du règlement-cadre fige dans la loi les **délais de transmission des données** à Eurostat pour chacun des domaines, et le *reporting* sur la qualité des enquêtes est basé sur la structure standardisée des rapports qualité et la norme de documentation des métadonnées définies en lien avec les experts qualité des différents pays (Eurostat, 2015).

Enfin, une clause de flexibilité permet à Eurostat de modifier partiellement les contenus des enquêtes par actes délégués pour répondre à de nouveaux besoins. Cette clause, très

7. Commencer le module par une question du type « avez-vous travaillé ne serait ce qu'une heure rémunérée ? » était un principe retenu, mais il n'était pas appliqué par tous les pays.

8. Au niveau des anciennes régions administratives en France (nomenclature NUTS 2).

débatte en amont avec les États membres qui y voyaient un risque de dérive de leurs coûts, est toutefois encadrée par la contrainte de ne pas ajouter de charge significative, la nécessité de réaliser une étude de faisabilité et surtout par un seuil en termes de contenu modifié (pas plus de 5 % des thématiques détaillées sur une période de 4 ans pour les enquêtes Emploi et le panel SILC, pas plus de 10 % entre deux éditions pour les autres enquêtes). Des contraintes du même type portent sur le nombre de variables. Les instituts statistiques par l'intermédiaire des directeurs généraux, en leur qualité d'experts, sont consultés par Eurostat sur les actes délégués avant leur adoption⁹.

📌 CONSÉQUENCES : UN POIDS ACCRU DES ENQUÊTES «EUROPÉENNES»...

L'Insee a largement soutenu le projet de règlement-cadre dans ses objectifs d'intégration et de standardisation, suivant d'ailleurs en cela les recommandations du Conseil national de l'information statistique en matière de comparabilité internationale des données. Notre système d'enquêtes préexistant était déjà largement adapté, même si, comme on le verra ci-dessous, les investissements induits par les nouveaux textes ne sont pas négligeables.

Avec l'extension du champ du règlement-cadre et la définition précise d'un programme pluriannuel assorti des délais de transmission stricts, le programme d'enquêtes de l'Insee¹⁰ est ainsi de plus en plus structuré autour des enquêtes sous règlement européen, autour desquelles doivent s'organiser les autres enquêtes collectées par son réseau d'enquêtrices et d'enquêteurs : enquêtes sur le logement, les conditions de travail, les mobilités, voire sur les sans-domicile. Mécaniquement, les ajustements de dates porteront sur ces dernières dans un contexte de moyens contraints et de demandes croissantes des utilisateurs pour des collectes de données nouvelles ou renouvelées (CNIS, 2019). En matière de statistiques sociales, l'utilisation croissante de données administratives, même si elle est centrale sur certains domaines comme les revenus, ne suffit pas à répondre aux besoins de connaissance : situations détaillées¹¹, comportements ou opinions des personnes, pauvreté en conditions de vie, état de santé perçu¹². Les enquêtes restent indispensables, et le resteront sans doute longtemps, même dans la perspective d'utilisation de données innovantes relevant du *big data* (Desrosières, 2004 ; Tassi, 2018). Il faut donc pouvoir continuer à les mener efficacement, en maintenant leur qualité. C'est pourquoi l'Insee s'est lancé, à l'instar d'autres pays européens (De Leeuw, 2018), dans le développement d'un programme d'enquêtes multimodes, construit autour d'outils mutualisés¹³. La rénovation de l'enquête Emploi, nécessitée par IESS, sera ainsi l'occasion d'introduire internet comme mode de réponse supplémentaire en réinterrogation.

9. Voir article d'Hervé Piffeteau sur le triptyque juridique des statistiques européennes dans ce même numéro.

10. Et plus largement du système statistique public, les enquêtes santé étant de la responsabilité du service statistique ministériel de la santé, la Drees.

11. Le statut de chômage au sens du BIT, déjà évoqué, en est un exemple typique.

12. IESS va ainsi étendre l'utilisation du « mini-module » standardisé de trois questions sur l'état de santé perçu, les gênes et difficultés dans la vie quotidienne et la présence de maladies chroniques.

13. Voir les articles sur Pogues, Eno et Coltrane dans ce même numéro.

📍 ... UN AJUSTEMENT DE L'ÉCHANTILLONNAGE ...

Les contraintes de précision portées par le règlement (*encadré*) ont également nécessité de revoir l'échantillonnage, notamment pour assurer la précision demandée au niveau régional pour la part de chômeurs dans la population en ce qui concerne l'enquête Emploi¹⁴. Le calendrier, déjà fixé pour le renouvellement périodique des échantillons des enquêtes ménages de l'Insee, a facilité le travail. En effet, l'échantillon de l'enquête Emploi devait être renouvelé en 2019. L'échantillon-maître des autres enquêtes auprès des ménages devait l'être progressivement à partir de 2020. La nouvelle base de sondage (basée sur les données fiscales et non plus sur les recensements de population), ainsi que des méthodes de tirage plus efficaces, permettaient d'améliorer la précision des données collectées. Les deux tirages ont alors pu être coordonnés au niveau des zones géographiques qui en constituent

« L'ensemble se traduit pour l'enquête Emploi par un ajustement de la répartition régionale des zones de collecte pour satisfaire le mieux possible aux contraintes de précision régionales imposées par IESS. »

les unités primaires pour faciliter l'organisation des collectes sur le terrain. L'ensemble se traduit pour l'enquête Emploi par un ajustement de la répartition régionale¹⁵ des zones de collecte pour satisfaire le mieux possible aux contraintes de précision régionales imposées par IESS, la taille finale d'échantillon étant finalement réduite de 15 %. Les conséquences sont plus lourdes pour le panel sur les Ressources et les conditions de vie : les contraintes de précision mais surtout le raccourcissement de la durée du panel à 4 ans¹⁶ vont conduire à modifier l'échantillonnage sur plusieurs années. Les

contraintes au niveau régional, trop coûteuses à réaliser par augmentation de la taille de l'échantillon pour cette enquête, y seront traitées en produisant la variable permettant de construire l'indicateur sur le risque de pauvreté par estimations sur petits domaines, qui devront donc être produites avec la même régularité que l'enquête.

Les changements les plus visibles et les plus rapides pour les utilisateurs des données concerneront la rénovation de l'enquête Emploi et du panel sur les Ressources et les conditions de vie : en premier lieu, la modification des questionnaires afin qu'ils puissent s'adapter aux concepts définis en commun et fournir précisément la liste des variables prescrite dans les actes d'exécution et délégués concernés. Pour l'enquête Emploi, l'ampleur du projet, associée au recours d'une collecte par internet pour les réinterrogations¹⁷, nécessite la mise en place d'un pilote dès 2020. C'est un délai très court après l'adoption du règlement, notamment pour pouvoir évaluer les ruptures de séries et « rétropoler » les données passées afin d'obtenir des séries statistiques cohérentes dans le temps sur les périodes antérieures. Eurostat a fortement soutenu ces évolutions en accordant des financements pour accompagner ces transformations.

14. Soulignons que IESS porte sur l'ensemble du territoire, y compris Mayotte, sauf pour le panel sur les Ressources et les conditions de vie. Il faudra donc réaliser à Mayotte une enquête Emploi en continu, alors qu'elle était jusqu'à présent annuelle. Ce projet sera mené dans un second temps, Mayotte devant mettre en place, à la même période, le recensement de la population en continu. Le système de dérogations inclus dans le règlement permet de donner au système statistique local ce temps d'adaptation supplémentaire.

15. Les petites régions voient leur nombre de zones de collecte augmenter, alors que celles de l'Île-de-France diminuent.

16. Durée minimale obligatoire du panel dans le règlement.

17. Les logements échantillonnés sont interrogés six trimestres de suite.

Encadré. Les seuils de précision des enquêtes fixés par les règlements européens dans le cas de la France

Le règlement IESS fixe des seuils de précision à respecter par les pays dans l'élaboration d'indicateurs notamment relatifs au marché du travail et de pauvreté. Ils concernent six indicateurs.

Il s'agit d'abord de trois estimations trimestrielles rapportées dans chaque cas à la population des 15-74 ans : nombre d'actifs occupés, nombre de chômeurs, et nombre de chômeurs au niveau régional (NUTS 2).

S'y ajoutent trois indicateurs relatifs au revenu et aux conditions de vie : indicateur AROPE* des personnes en risque de pauvreté ou d'exclusion sociale au niveau national, au niveau régional, et enfin taux de pauvreté persistante à 4 ans au niveau national.

Les seuils de précision définis en annexe du règlement-cadre sont exprimés à l'aide d'une formule qui fixe un maximum pour l'écart-type de l'indicateur. Elle s'écrit :

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{f(N)}}$$

Dans cette expression, « \hat{p} » est la valeur de l'indicateur et $f(N) = a\sqrt{N} + b$, où N est la taille du domaine (pays ou NUTS 2) et a et b des paramètres fixés pour chaque indicateur.

Cette formulation peut paraître étrange aux statisticiens : le dénominateur de la fraction devrait être non pas une fonction de la taille de la population d'intérêt (N), mais de la taille de l'échantillon sélectionné n.

En adoptant pour f une forme liée à la racine carrée la taille de la population d'intérêt (N), on limite toutefois, pour les grandes valeurs de N, la dépendance à cette taille par rapport à ce qu'elle serait si le principe d'une allocation proportionnelle était appliqué. Par ailleurs une formule spécifique est prévue dans le cas de la part de chômeurs, pour les régions de moins de 300 000 habitants. Mais les valeurs des paramètres a et b, fixées isolément selon qu'ils s'appliquent au niveau national ou au niveau NUTS2, peuvent conduire à une incohérence entre les exigences de précision portant sur chacun des différents niveaux.

Ainsi, avec les seuils fixés dans le règlement selon la formule précédente, le nombre de répondants requis pour respecter le seuil national pour la part de chômeurs s'établit, dans le cas de la France métropolitaine, à 109 400. Or la somme des répondants requis pour respecter l'ensemble des seuils régionaux s'établit à 83 000, soit 31 % d'écart. Pour l'indicateur AROPE, l'écart est de 48 %... mais dans le sens inverse. Il est apparu judicieux, dans ces circonstances, de recourir pour les estimations régionales aux méthodes moins coûteuses « d'estimation de petits domaines » fondées sur l'enquête nationale et sur une modélisation complémentaire (Rao et Molina, 2015 ; Sautory, 2018). Cette possibilité a finalement été introduite dans le règlement-cadre à la demande de la France.

* Cet indicateur, connu sous son sigle anglais (ARPE, *At-risk of poverty or social exclusion*), résume le nombre de personnes qui est menacé par le risque de pauvreté ou vivent dans la privation matérielle aiguë et/ou vivent dans un ménage à faible niveau d'intensité de travail. Il s'agit de l'indicateur principal utilisé pour suivre l'objectif de réduction de la pauvreté inclus dans la stratégie Europe 2020.

Nota : l'analyse détaillée qui a permis de rédiger cet encadré a été conduite par Patrick Sillard et Martin Chevalier (Insee, département des méthodes statistiques).

📍 ... ET UNE RÉDUCTION DES DÉLAIS

Enfin, et ce n'est pas le moindre sujet, un investissement conséquent sera nécessaire pour la production plus rapide de données de revenus dans le panel sur les Ressources et les conditions de vie. Le règlement-cadre prévoit la transmission des données à la fin de l'année de référence, et « dans des cas exceptionnels » l'envoi de données provisoires à la fin de l'année de référence et de données révisées fin février de l'année suivante.

C'est un pari difficile pour la France en raison des délais de mise à disposition des données fiscales et sociales utilisées pour produire ces informations : à ce jour, les revenus qui en sont issus ne peuvent être transmis à Eurostat qu'en septembre de l'année suivante. Le projet est d'utiliser, pour produire des données révisées, des émissions plus précoces (mais moins complètes) des données fiscales, et de combler ce décalage temporel sans trop détériorer la qualité des données transmises. Cela nécessitera plusieurs années d'adaptation, d'autant plus que dans le même temps, les changements de systèmes d'information liés à la mise en place du prélèvement à la source¹⁸ doivent être pris en compte. Là encore, le système de dérogations inclus dans le règlement devrait permettre de gérer cette transition.

Au bout du compte, les transformations liées au nouveau règlement-cadre se réaliseront de manière étalée dans le temps, au rythme du programme d'enquêtes et de la préparation des actes d'exécution et délégués associés à chacune d'entre elles¹⁹. Eurostat assurera un soutien financier pour adapter les enquêtes²⁰ et accordera des dérogations si nécessaire pour permettre des délais d'ajustements. Néanmoins les délais sont serrés pour l'enquête Emploi et le panel sur les Ressources et les conditions de vie²¹, premières enquêtes concernées dans le programme. Le choix du mode de collecte est pour l'instant laissé entièrement à l'initiative de chaque pays, même si Eurostat soutient une réflexion autour

« Le choix du mode de collecte est pour l'instant laissé entièrement à l'initiative de chaque pays, même si Eurostat soutient une réflexion autour de l'établissement de recommandations sur ces aspects. »

de l'établissement de recommandations sur ces aspects. En France, le passage progressif au multimode et la standardisation des outils d'enquête seront conduits en parallèle, avec des ajustements successifs ; par exemple, contrairement à l'enquête Emploi, la collecte par internet d'une partie du panel sur les Ressources et les conditions de vie n'est pas encore instruite et le sera dans un second temps. L'enquête Emploi pour sa part n'est pas concernée à court terme par les nouveaux outils destinés à produire les enquêtes en multimode, en cours de réalisation et qu'elle adoptera dans un second temps.

18. Et peut-être à une future suppression des déclarations de revenus...

19. Ces textes doivent être définitifs au plus tard un an avant la collecte concernée.

20. Ces financements sont explicitement prévus dans le règlement-cadre.

21. Pour le panel, la rénovation a été organisée dès 2020 en raison du besoin de renouvellement de l'ensemble de la chaîne informatique de l'enquête, programmée par ailleurs.

VERS PLUS DE COHÉRENCE ET DE COMPARABILITÉ ENTRE PAYS ?

La standardisation des concepts et des variables mise en place par IESS est évidemment un outil essentiel pour améliorer la comparabilité des données entre pays. C'est par exemple le cas sur la durée de travail effective : les problèmes de comparabilité entre pays (Körner et Wolff, 2016) seront notablement améliorés lorsque tous les pays adopteront une semaine de référence fixe comme le prévoit le règlement-cadre. Encore faudra-t-il que les interprétations des définitions et des contenus soient aisées et qu'elles ne laissent pas trop prise à des ajustements locaux.

La discussion des variables de l'enquête sur les Forces de travail dans le cadre de la préparation des textes d'application a par exemple montré que les différences n'étaient pas mineures entre les pays sur la manière de classer en emploi ou non les parents en congés parentaux, eux-mêmes organisés de manière fort différente d'un pays à l'autre. Auparavant en France, comme dans d'autres pays, l'interprétation des critères du BIT faisait prévaloir un critère de durée : au-delà de trois mois, les parents absents de leur emploi pour congé parental n'étaient pas considérés en emploi. Après discussion, a prévalu dans IESS un critère plus souple d'existence de lien avec l'emploi : les parents seront considérés en emploi sans limite de durée d'absence, pour peu qu'ils déclarent être en emploi et percevoir (ou être éligibles à) un revenu lié à l'emploi, même si la rémunération du congé parental n'est pas proportionnelle au salaire antérieur. À l'occasion de cette discussion, il est apparu que de nombreux pays procédaient déjà de cette façon, bien que la consigne en ce sens n'ait pas été donnée, biaisant les comparaisons de taux d'emploi, notamment des femmes. Au bout du compte, c'est cette définition plus floue qui prévaudra pour l'indicateur de taux d'emploi européen, même s'il peut paraître étrange qu'un éloignement durable des mères du marché du travail pour élever leurs enfants n'affecte pas le taux d'emploi féminin... Il reste encore dans cet exemple une marge d'interprétation. Un travail précis d'expertise des dispositions nationales en matière de congés parentaux et de classement des situations d'emploi concernées serait nécessaire pour assurer une parfaite comparabilité. Il n'a pas été fait jusqu'à présent.

Un autre exemple est la manière de positionner les étudiants vivant seuls par rapport au ménage de leurs parents, souvent largement mobilisés pour les soutenir financièrement et dans lequel ils reviennent très régulièrement. Si l'on suit les concepts du recensement, les étudiants vivant séparément dans un logement ordinaire pendant l'année universitaire doivent être interrogés à part. Dans les enquêtes sur les revenus, en cohérence avec la définition du ménage « économique », on peut souhaiter les inclure dans le ménage de leurs parents, afin de calculer des niveaux de vie qui tiennent mieux compte du partage familial des ressources. Mais dans ce cas, ce sont le plus souvent les parents qui répondent pour le compte de leurs enfants²² lors du passage de l'enquêtrice ou de l'enquêteur. Finalement le règlement laisse les deux possibilités aux États membres pour chacune des enquêtes, sous réserve de documenter précisément l'option choisie.

Un arbitrage majeur laissé à chaque pays est celui de l'amplitude d'utilisation des données administratives en lieu et place de données d'enquêtes (à l'exception du module d'activité

22. On parle alors d'enquête par « proxy ».

de l'enquête sur les Forces de travail²³), ainsi que le choix du mode de collecte dans les enquêtes. Il aurait été irréaliste de légiférer plus avant sur ce sujet, et à vrai dire aucun État membre ne le souhaitait, tant les systèmes statistiques diffèrent entre pays : une bonne partie des données sociales sont issues de registres administratifs dans les pays du Nord, et certains pays collectent encore des enquêtes sur la base de questionnaires papier, alors que d'autres sont résolument passés à une collecte par internet ou téléphonique en première intention. Cependant, les travaux de recherche réalisés sur les effets

de mode de collecte laissent penser que les comportements de réponse des enquêtés ne sont pas indépendants du mode, même pour des questionnements identiques (Vinceneux, 2018 ; Razafindranovona, 2015).

« IESS laisse aussi la place à l'innovation en prévoyant de soutenir des études pilotes sur de nouvelles méthodes et de nouveaux sujets. »

Finalement, IESS va sans aucun doute contribuer à l'amélioration de la comparabilité, mais il ne s'agira pas d'un *big bang* ; c'est plutôt une avancée sensible sur la voie d'un processus

de convergence, incomplet mais avec davantage d'harmonisation par les « *inputs* » (les variables, les contenus) que par le passé.

🌐 VERS UNE MEILLEURE RÉPONSE AUX BESOINS ?

Les besoins d'information des utilisateurs, notamment des décideurs publics, évoluent au fil du temps. IESS, on l'a vu, s'autorise une certaine flexibilité et des marges d'évolution. La mise en place d'une enquête pluriannuelle sur l'Emploi du temps, centrale notamment pour quantifier la répartition des rôles des hommes et des femmes dans les activités de la vie quotidienne, ou analyser la conciliation entre vie familiale et vie professionnelle, est optionnelle mais reste une cible à atteindre pour tous, dans des délais non précisés. IESS laisse aussi la place à l'innovation en prévoyant de soutenir des études pilotes sur de nouvelles méthodes et de nouveaux sujets : une étude de ce type est par exemple en cours sur le thème des violences de genre. Des études pilotes pourront aussi être financées pour étendre la collecte de données sociales à des personnes vivant hors logement ordinaire, ou des populations difficiles à joindre. Les migrants sont explicitement cités²⁴. Des études de faisabilité sont aussi prévues et financées pour moderniser la collecte des données de consommation ou d'emploi du temps. Des groupes de travail rassemblant des pays volontaires commencent d'ailleurs à explorer l'utilisation de données anonymisées de cartes de crédit ou d'agendas électroniques à cet effet. On le voit, IESS accorde un espace à l'innovation, et c'est une très bonne chose. Mais au bout du compte, tout dépendra des résultats effectifs de ces expérimentations et des moyens que les États membres pourront effectivement consacrer à ces innovations (au-delà du très utile soutien financier d'Eurostat) et surtout à leur passage en production, éventuellement en réalisant des gains de productivité sur les autres sujets.

Par ailleurs, le nouveau système des modules *ad hoc* intégrés dans les enquêtes à des périodicités variables sur les huit années du programme glissant permettent aussi d'introduire une certaine souplesse : certains thèmes (deux sur huit par exemple pour l'enquête

23. Et de certains modules du panel sur les Ressources et les conditions de vie comme la santé, les privations matérielles ou le bien-être.

24. On pourrait aussi penser aux populations sans domicile, mais ce sujet n'a pas été abordé à ce jour au niveau européen.

sur les Forces de travail) sont laissés libres pour permettre de s'adapter aux besoins de la Commission, et le rythme de collecte est adapté à des besoins moins fréquents.

Une autre question est de savoir si les utilisateurs des données se trouveront satisfaits au bout du compte. Il faut bien reconnaître que si les besoins des autres directions de la Commission sont régulièrement explorés et mis en avant par Eurostat, il est moins évident de pouvoir faire le lien avec les utilisateurs nationaux ou même les chercheurs. Les procédures et les structures de consultation des utilisateurs restent assez confidentielles au niveau européen²⁵, et l'articulation avec des commissions nationales comme le Conseil national de l'information statistique en France peine encore à s'organiser. Le développement de la littératie statistique est également une condition essentielle de dialogue avec les utilisateurs non experts, qui comprennent difficilement les contraintes de la production de données de qualité, tant ils sont abreuvés de (mauvaises) « données » diverses et variées sur les réseaux sociaux. Ce point est heureusement un des axes centraux de communication d'Eurostat comme d'un grand nombre d'États membres, y compris la France.

🕒 UNE AMÉLIORATION DU RAPPORT COÛT-EFFICACITÉ DE LA STATISTIQUE PUBLIQUE ?

IESS, en mettant en avant la standardisation et l'intégration des enquêtes, en permettant une meilleure prévisibilité des engagements en matière de collecte et de transmission de données, permettra sans doute à terme une meilleure organisation des collectes de données nationales. Mais, s'il est facilitateur en la matière, il ne constitue pas un gage d'amélioration du rapport coût-efficacité de la statistique publique à court terme. L'investissement à

« L'investissement à réaliser pour la rénovation des enquêtes, les quelques extensions géographiques nécessaires et l'accélération de certains délais de production est en effet important, même dans un pays comme la France, assez proche du modèle cible. »

réaliser pour la rénovation des enquêtes, les quelques extensions géographiques nécessaires et l'accélération de certains délais de production est en effet important, même dans un pays comme la France, assez proche du modèle cible. Pour améliorer ce rapport coût-efficacité au niveau national, il faut se saisir des possibilités offertes par IESS pour développer la mutualisation des outils de collecte et réaliser certains pans de la collecte par internet.

C'est bien ce qui est fait en France, avec le développement d'un programme d'enquêtes multimodes et ses outils partagés, qui permettra de concentrer le temps des enquêtrices et enquêteurs sur les collectes de données pour lesquels elles ou ils sont indispensables et irremplaçables : les questionnaires complexes, les populations particulières ou peu connectées (Beck, De Peretti, 2017), etc. C'est aussi ce qui est fait avec la mise en place d'un référentiel de documentation des enquêtes, qui permettra de réutiliser simplement des

25. La consultation électronique des utilisateurs réalisée en 2015 par Eurostat n'a ainsi donné lieu qu'à une centaine de contributions, dont environ 1/3 de chercheurs, nombre qui paraît bien modeste au regard, par exemple, des 21 330 contributeurs à la consultation sur la loi République Numérique menée en France à la même période et sur une durée plus brève.

parties de questionnaires ou de produire plus simplement les métadonnées des enquêtes, et par exemple les rapports qualité dans le format standard développé en lien avec Eurostat (Bonnans, 2019). Mais tout cela prendra encore du temps.

Par ailleurs, la flexibilité prévue dans IESS aura un coût. Changer une partie du contenu d'un questionnaire, développer de nouveaux sujets est un investissement. C'est pourquoi les États membres ont souhaité fixer des limites à ces ajustements. Ce n'est pas la seule raison : les statistiques sociales ont besoin de comparabilité dans l'espace, mais aussi dans le temps. Faire évoluer un questionnement est toujours un arbitrage entre besoin de s'ajuster à une

situation changeante tout en préservant une utile vue comparative sur le passé. Le surcoût passe alors, par exemple, par la nécessité de mener une enquête pilote qui permette de mesurer l'ampleur des ruptures de séries.

« *Changer une partie du contenu d'un questionnaire, développer de nouveaux sujets est un investissement.* »

Enfin, la maîtrise des coûts n'est pas seulement une nécessité budgétaire pour les instituts nationaux. La charge de réponse des enquêtés doit aussi faire l'objet d'une attention particulière afin

de ne pas décourager les ménages. Le développement du questionnement par internet répond ainsi à une demande récurrente d'une partie des enquêtés, ce qui devrait permettre de limiter l'érosion du taux de réponse, notamment par téléphone. Le développement de l'utilisation conjointe de données complémentaires aux enquêtes, qu'elles soient issues de fichiers administratifs ou d'autres sources, doit être poursuivi, et le développement de nouveaux questionnements doit être raisonnable. C'est aussi une des préoccupations mises en avant dans la discussion sur IESS, et elle ne devra pas être abandonnée.

BIBLIOGRAPHIE

BECK, François et DEPERETTI, Gaël, 2017. *La collecte multimode à l'Insee: d'un effet de mode à un effet sur nos travaux*. [en ligne]. 22 juin 2017. Séminaire de Méthodologie statistique du département des méthodes statistiques. Paris. [Consulté le 25 octobre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/3555272/presentation_SMS_22%20juin_beck_deperetti.pdf

BONNANS, Dominique, 2019. RMÉS, le référentiel de métadonnées statistiques de l'Insee. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 46-55. [Consulté le 24 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168396/courstat-2-6.pdf>

CNIS, 2019. *Avis du Moyen-terme 2019-2023*. [en ligne]. 1er février 2019. N°8/H030. [Consulté le 5 novembre 2019]. Disponible à l'adresse : https://www.cnis.fr/wp-content/uploads/2018/11/Avis-Moyen-Terme-2019-2023-du-Cnis_AP-valid%C3%A9s-1-1.pdf

CODER, Yohan, DIXTE, Christophe, HAMEAU, Alexis, HAMMAN, Sophie, LARRIEU, Sylvain, MARRAKCHI, Anis et MONTAUT, Alexis, 2019. Les chômeurs au sens du BIT et les demandeurs d'emploi inscrits à Pôle emploi : une divergence de mesure du chômage aux causes multiples. In : *Insee Références – Emploi, chômage, revenus du travail*. [en ligne]. Édition 2019, pp. 71-85. [Consulté le 25 octobre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/4183055/ECRT19e_D4_chomeurs.pdf

DE LEEUW, Edith D., 2018. Mixed-Mode: Past, Present, and Future. In : *Survey Research Methods – Journal of the European Survey Research Association*. [en ligne]. 13 août 2018. Vol 12, n°2, pp. 75-86. [Consulté le 28 octobre 2019]. Disponible à l'adresse : <https://ojs.ub.uni-konstanz.de/srm/article/view/7402/6582>

DESROSIERES, Alain, 2004. Enquêtes versus registres administratifs : réflexions sur la dualité des sources statistiques. In : *Courrier des statistiques*. [en ligne]. Septembre 2004. N° 111, pp. 3-16. [Consulté le 29 octobre 2019]. Disponible à l'adresse : <https://www.epsilon.insee.fr/jspui/bitstream/1/8400/1/cs111b.pdf>

EUROSTAT, 2010. *Cadre juridique pour les statistiques européennes, la loi statistique*. [en ligne]. 13 juillet 2010. [Consulté le 28 octobre 2019]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/documents/3217494/5719345/KS-31-09-254-FR.PDF>

EUROSTAT, 2011. *Mémoire de Wiesbaden*. [en ligne]. 28 septembre 2011. DGINS. [Consulté en ligne le 5 novembre 2019]. Disponible à l'adresse : https://circabc.europa.eu/webdav/CircaBC/ESTAT/dss/Library/DSS%20meetings/meetings_2012/DSS%20Meeting%2020-21%20September%202012/Item_3.3.%20Modernisation%20of%20Social%20statistics_annex.pdf

EUROSTAT, 2015. *Single Integrated Metadata Structure v2.0 (SIMS v2.0) and its underlying reporting structures – The ESS Quality and Reference Metadata Reporting Standards – ESMS 2.0 and ESQRS 2.0*. [en ligne]. Décembre 2015. Luxembourg. [Consulté le 13 novembre 2019]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/documents/64157/4373903/SIMS-2-0-Revised-standards-November-2015-ESSC-final.pdf>

EUROSTAT, CSSE, 2017. *Code de bonnes pratiques de la statistique européenne*. [en ligne]. 16 novembre 2017. [Consulté le 28 octobre 2019]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/documents/4031688/9332274/KS-02-18-142-FR-N.pdf>

KÖRNER, Thomas et WOLFF, Loup, 2016. La fragile comparabilité des durées de travail en France et en Allemagne. In : *Insee Analyses*. [en ligne]. Juin 2016. N° 26. [Consulté le 29 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/version-html/2019603/IA26.pdf>

RAO, J.N.K. et MOLINA, Isabel, 2015. *Small Area Estimation*. Août 2015. Éditions Wiley, Series in Survey Methodology. Seconde édition. ISBN 978-1-118-73578-7

RAZAFINDRANOVARA, Tiaray, 2015. La collecte multimode et le paradigme de l'erreur d'enquête totale. In : *Série des documents de travail – Méthodologie statistique – de la Direction de la méthodologie et de la coordination statistique et internationale*. [en ligne]. N° M 2015/01. Insee. [Consulté le 29 octobre 2019]. Disponible à l'adresse : <https://www.epsilon.insee.fr/jspui/bitstream/1/28875/1/m1501.pdf>

SAUTORY, Olivier, 2018. Méthodes d'estimation sur petits domaines pour l'indicateur AROPE (At-risk of poverty or social exclusion) au niveau régional à partir de l'enquête SRCV. *13^e Journées de méthodologie statistique de l'Insee*. [en ligne]. 12-14 juin 2018, Paris. [Consulté le 28 octobre 2019]. Disponible à l'adresse : http://www.jms-insee.fr/2018/S11_4_ACTE_SAUTORY_JMS2018.pdf

TASSI, Philippe, 2018. Introduction – Les apports des Big Data. In : *Économie et Statistique*. [en ligne]. N°505-506, pp. 5-16. [Consulté le 13 novembre 2019]. Disponible à l'adresse : https://www.insee.fr/fr/statistiques/fichier/3705956/505-506_Tassi-FR.pdf

VINCENEUX, Klara, 2018. Mode de collecte et questionnaire, quels impacts sur les indicateurs européens de l'enquête Emploi ? In : *Documents de travail – Insee*. [en ligne]. N° F1804. [Consulté le 29 octobre 2019]. Disponible à l'adresse : <https://www.epsilon.insee.fr/jspui/bitstream/1/82474/1/f1804.pdf>

❶ FONDEMENTS JURIDIQUES

Règlement (CE) n°223/2009 du Parlement européen et du Conseil du 11 mars 2009. In : *Journal officiel de l'Union européenne*. [en ligne]. 31 mars 2009. L87, Article 12. [Consulté le 13 novembre 2019]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=OJ:L:2009:087:FULL&from=FR>

Parlement européen, 2019. *Résolution législative du Parlement européen du 16 avril 2019 sur la proposition de règlement du Parlement européen et du Conseil établissant un cadre commun pour des statistiques européennes relatives aux personnes et aux ménages fondées sur des données au niveau individuel collectées à partir d'échantillons*. [en ligne]. 16 avril 2019. [Consulté le 19 novembre 2019]. Disponible à l'adresse : http://www.europarl.europa.eu/doceo/document/TA-8-2019-0387_FR.html

Règlement (UE) 2019/1700 du Parlement européen et du Conseil du 10 octobre 2019 établissant un cadre commun pour des statistiques européennes relatives aux personnes et aux ménages fondées sur des données au niveau individuel collectées à partir d'échantillons. [en ligne]. [Consulté le 13 décembre 2019]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019R1700&from=FR>


LE SYSTÈME STATISTIQUE ALLEMAND

UNE ÉVOLUTION FACILITÉE PAR LA RÉGLEMENTATION EUROPÉENNE

Véronique Alexandre* et Jean-Pierre Cling**

L'organisation du système statistique public en Allemagne, adaptée à une république fédérale, présente trois caractéristiques principales. En premier lieu, un légalisme contraignant qui exige que toute enquête statistique soit régie par une loi spécifique (ou dans certains cas par un acte législatif délégué). En second lieu, une décentralisation très marquée avec un office statistique fédéral, Destatis, et quatorze offices statistiques des Länder qui en sont indépendants. Enfin, une forte concentration fonctionnelle et l'absence de services statistiques ministériels au sens français du terme.

Les évolutions de la réglementation européenne ont donné des marges de liberté nouvelles au système statistique allemand, notamment des possibilités de recours plus large aux données administratives et d'ouverture accrue des données statistiques au monde de la recherche. Elles ont également renforcé le rôle de coordination de Destatis.

 *There are three main characteristics defining the organisation of Germany's Official Statistical System, a system which has been tailored to suit a federal republic. Firstly, mandatory legalism which requires that any statistical survey be governed by a specific law (or, in certain cases, by a delegated legislative act). Secondly, very pronounced decentralisation, with a Federal office of statistics, DESTATIS, and fourteen Landër offices of statistics which are independent thereof. Finally, high functional concentration and the absence of Ministerial Statistical Departments in the French sense of the term.*

Changes in European regulations have given the German Statistical System greater freedom, in particular the ability to make broader use of administrative data and to open up more statistical data in the world of research. They have also reinforced the role played by DESTATIS as a coordinator.

* En fonction au département de la Coordination statistique et internationale, Insee, veronique.alexandre@insee.fr

** Chef du département de la Coordination statistique et internationale, Insee, jean-pierre.cling@insee.fr

Nota : les auteurs remercient Thomas Riede, Kay Sommer et Bernd Stoertzbach de Destatis, pour leurs commentaires précieux sur une première version de cet article. Les auteurs demeurent entièrement responsables des opinions exprimées et des erreurs qui pourraient subsister.

LE SYSTÈME STATISTIQUE ALLEMAND EST ENCADRÉ, DÉCENTRALISÉ ET COORDONNÉ PAR CONSENSUS

Vue de France¹, la statistique publique en Allemagne apparaît comme très encadrée par la loi. Façonnée par l'organisation fédérale du pays, elle est fortement décentralisée et coordonnée essentiellement par consensus. En particulier, les composantes du système statistique public, l'office statistique fédéral, Destatis, et ceux des *Länder*, jouissent chacune d'une totale indépendance administrative et financière vis-à-vis des autres. L'accès aux données administratives est par ailleurs très encadré.

Mutatis mutandis, cette organisation s'apparente à bien des égards à celle de la statistique européenne. En effet, la relation d'harmonisation par la négociation entre Destatis et les offices statistiques des *Länder* ressemble à celle qui prévaut entre Eurostat et les Instituts nationaux de statistique (INS). De plus, la régulation d'ensemble est largement portée par la loi en Allemagne, par les règlements au sein de l'Union européenne. En revanche, Destatis concentre l'essentiel des statistiques publiques au niveau fédéral et il n'existe pas, comme en France, de services statistiques ministériels à part entière.

UN ENCADREMENT LÉGISLATIF FORT

Le cadre juridique de la statistique publique est inscrit dans le texte même de la constitution allemande du 23 mai 1949. L'article 73 de celle-ci (alinéa 11) attribue la compétence législative pour les statistiques fédérales au niveau fédéral.

La loi sur les statistiques fédérales (*Bundesstatistikgesetz*) de 1987 (remplaçant la première loi datant de 1953), dont la dernière modification substantielle remonte à 2016, est basée sur cet article. Elle régit notamment l'organisation et les missions de Destatis (voir ci-après) et du Conseil de la statistique. Elle précise que toute statistique doit avoir pour fondement une loi² dans laquelle sont définies la couverture, la périodicité et les variables d'enquête. Ce principe s'applique également pour les règlements statistiques européens.

Certaines adaptations sont cependant possibles sous certaines conditions à travers un acte législatif délégué pris par le gouvernement (équivalent d'une ordonnance) avec l'accord de l'assemblée qui réunit les représentants des *Länder* (*Bundesrat*). Trois cas de figure autorisent le suivi de cette procédure : un alourdissement marginal des enquêtes (ajout de variables, augmentation de la taille de l'échantillon, etc.), dont la validité peut atteindre jusqu'à trois ans ; en sens inverse, des allègements marginaux (par exemple report d'une enquête ou réduction du nombre de répondants), dont la validité peut atteindre jusqu'à quatre ans ; enfin, l'application des nouveaux règlements statistiques européens.

La loi sur les statistiques fédérales ne concerne pas les *Länder* (art. 83 de la constitution) : chacun d'entre eux a également sa propre loi statistique³ qui, sans contredire la loi fédérale (art. 87, alinéa 3), indique comment sont remplies les obligations fédérales et la complète pour prendre en compte ses propres spécificités.

1 Le système statistique allemand a été antérieurement présenté, d'une part à l'occasion d'une comparaison France-Allemagne (Alexandre, de Gimel et Heinlein, 1997) et d'autre part, dans le cadre de présentation des systèmes statistiques des grands pays européens (Lefèvre, 1998, 1999).

2 En France, la statistique ne fait pas partie des domaines relevant de la loi (cf. article 34 de la Constitution de la République française qui liste les domaines relevant de la loi).

3 La Rhénanie du Nord-Westphalie est le seul *Land* à ne pas avoir de loi statistique.

Il en résulte que le processus de toute nouvelle opération statistique ou toute modification substantielle d'une statistique existante est long, car il nécessite un vote identique des deux chambres, le parlement fédéral (*Bundestag*) et l'assemblée qui réunit les représentants des *Länder* (*Bundesrat*).

« Cet encadrement oblige les statisticiens à une grande transparence vis-à-vis du législateur et du public. »

Cet encadrement oblige les statisticiens à une grande transparence vis-à-vis du législateur et du public. Ceux-ci doivent notamment évaluer l'impact des changements qu'ils proposent, les envisager dans la durée, les expliciter et les justifier. Le vote des lois par les deux chambres introduit une forte dimension de négociation et de concertation que l'on retrouve dans le cadre des relations entre l'office statistique fédéral et ceux des *Länder*.

Le Conseil de la statistique (*Statistischer Beirat*) représente, lui, les utilisateurs, les fournisseurs et les producteurs de données qui peuvent y exprimer leurs besoins et avis (**encadré 1**).

La loi sur les statistiques fédérales de 1987 (et ses révisions ultérieures) a renforcé l'encadrement juridique de la statistique publique en Allemagne. Il s'agissait de tirer les conclusions d'un jugement de la Cour fédérale constitutionnelle relatif au recensement de la population prévu en 1983⁴. Ce jugement répondait à l'inquiétude de l'opinion publique et étendait la protection des libertés individuelles face à l'utilisation croissante des fichiers informatiques et à l'abondance des enquêtes statistiques. En conséquence, la loi statistique et des lois spécifiques ont restreint en particulier l'interconnexion des fichiers et fixé des exigences particulières pour les offices statistiques.

Encadré1. Le Conseil de la statistique représente les utilisateurs, les producteurs et les fournisseurs de données statistiques

Le Conseil de la statistique en Allemagne est l'équivalent du Conseil national de l'information statistique en France. Son rôle est de conseiller Destatis sur les questions statistiques. Son secrétariat est assuré par ce dernier, tandis que son président est élu parmi ses membres. Le Conseil de la statistique recueille les besoins des utilisateurs : représentants des différents secteurs économiques, des professions libérales, des syndicats de salariés et d'employeurs, des associations environnementales, des communes, du monde de l'éducation et de la communauté scientifique. Il tient également compte des avis des répondants aux enquêtes statistiques et des producteurs de données eux-mêmes (Destatis, les offices statistiques des *Länder* et Eurostat). Le niveau fédéral y est représenté par les ministères, la Cour des comptes, la *Deutsche Bundesbank* et un délégué au niveau fédéral pour la protection des données. Il existe aussi une commission statistique régionale dans chaque office de *Land*, équivalente au Conseil de la statistique, qui doit être consultée sur la législation, les modifications ou le maintien de statistiques ; elle peut proposer la mise en place, la modification ou des compléments à des travaux statistiques.

⁴ La Cour constitutionnelle fédérale a mis fin au recensement prévu pour 1983 avec son « jugement de recensement » du 15 décembre 1983 et a développé le droit à l'autodétermination informationnelle. La collecte et le traitement des données à caractère personnel ont été soumis à des exigences accrues en matière de statistiques publiques.

Au final, l'utilisation des fichiers administratifs à des fins statistiques (compte tenu des difficultés d'appariement, de partage de données entre administrations, etc.) est restée moins développée en Allemagne que dans beaucoup de pays développés et notamment en France, où la statistique publique s'est de plus en plus appuyée sur les sources administratives.

UNE DÉCENTRALISATION MARQUÉE

Le système statistique public allemand s'inscrit dans l'organisation fédérale de l'Allemagne, qui diffère radicalement de l'organisation très centralisée de la France. Celle-ci distingue un niveau fédéral qui dispose de compétences exclusives (défense, politique extérieure, nationalité notamment) et seize *Länder* qui ont chacun leurs propres institutions politiques (un gouvernement, un parlement, une administration, une cour constitutionnelle). Ces derniers bénéficient d'une dévolution totale dans plusieurs domaines (éducation, culture, police, etc.), ou partagent les compétences avec le niveau fédéral (droit du commerce par exemple).

« La statistique publique est une compétence partagée. »

La statistique publique est une compétence partagée. Elle repose ainsi sur une institution fédérale chargée des statistiques, le *Statistisches Bundesamt* (Destatis), et sur les offices statistiques des *Länder*, qui jouissent d'une totale indépendance par rapport à Destatis (article 83 de la constitution) mais qui exécutent la loi fédérale. Très schématiquement, ces derniers collectent et traitent les données tant pour le niveau fédéral et européen que pour le niveau régional. Ils fournissent à Destatis principalement des données agrégées pour lui permettre d'établir les statistiques fédérales et des données détaillées (micro-données). Les données individuelles sur les ménages sont anonymisées avant leur transmission à l'office statistique fédéral. Destatis est, pour sa part, responsable de la conception des opérations statistiques, du calcul et de la diffusion des résultats au niveau de l'Allemagne, ainsi que de la coordination d'ensemble du système statistique public.

LE RÔLE DE COORDINATION D'ENSEMBLE DE DESTATIS

Destatis est une administration fédérale autonome, rattachée au ministère de l'Intérieur. Elle employait 2 300 personnes⁵ en 2018 (Destatis, 2018) et disposait en 2019 d'un budget de 280 millions d'euros. Son siège est à Wiesbaden (capitale du *Land* de Hesse), sachant qu'il existe un bureau de taille importante à Bonn et une antenne à Berlin. Son président (Georg Thiel depuis octobre 2017) est nommé par le Président de la République sur proposition du gouvernement fédéral.

Destatis représente le système statistique allemand dans les instances internationales. Aux réunions du Comité du système statistique européen, le Président de Destatis est accompagné d'un représentant des offices des *Länder*.

Dans le contexte fédéral allemand, Destatis ne peut pas transmettre des directives aux offices statistiques des *Länder* et agit en partenariat avec eux. Ce partenariat opère de manière informelle, avec des comptes rendus de réunions qui actent des décisions prises (de

5. À titre de comparaison, l'Insee employait 5 370 personnes fin 2018, dont environ 1 500 étaient affectées à la direction générale (Montrouge et Metz). Le budget de l'Insee était de 442 millions d'euros en 2019.

« Aux réunions du Comité du système statistique européen, le Président de Destatis est accompagné d'un représentant des offices des Länder. »

manière analogue au fonctionnement des réunions du comité du système statistique européen, hors comitologie). Destatis exerce une coordination des travaux de production et veille à la qualité homogène des résultats au niveau national. Il bénéficie également de la concertation étroite qui existe avec les offices des *Länder* au sein de la Conférence des dirigeants des offices statistiques, qui les réunit, mais aussi via des échanges permanents entre eux.

Destatis est également responsable, au niveau fédéral, de la préparation du programme statistique, des normes et études méthodologiques et techniques, des publications ainsi que du traitement et de la fourniture de données à Eurostat et aux autres institutions internationales.

Il produit en propre notamment les statistiques du commerce extérieur et une enquête sur « le coût de la bureaucratie⁶ », collectant les données et les traitant au niveau central. Il a la responsabilité de la conduite des opérations de niveau fédéral et des enquêtes européennes, sans avoir la possibilité de diriger toutes les phases des travaux. En particulier, Destatis tire les échantillons pour les enquêtes de niveau fédéral telles que le *Mikrozensus* (Rivière, 2019). Les *Länder* publient les résultats des enquêtes pour leur *Land*, contrairement à la France⁷.

LES OFFICES STATISTIQUES DES LÄNDER, PILIERS DE LA STATISTIQUE PUBLIQUE

Les quatorze offices statistiques régionaux (pour seize *Länder*, du fait de deux regroupements) ne dépendent pas de l'office central comme en France, mais en général du ministère de l'Intérieur de leur *Land*, quelquefois de celui des finances. La définition de leur programme statistique tient compte des demandes de tous les ministères de leur *Land*, avec lesquels sont négociés les travaux à effectuer.

Comparés à ceux de Destatis, le budget et le personnel global des offices statistiques des *Länder* (environ 4 700 employés au total) en représentent plus du double. À l'Insee, le ratio est du même ordre de grandeur (environ 2,5) entre les effectifs des directions régionales et ceux de la direction générale.

Ces offices régionaux collectent et traitent les données tant pour le niveau fédéral et celui de l'Union européenne que pour le niveau régional. Pour l'office statistique de l'Union européenne Eurostat, ce sont d'ailleurs tous des ONAs (*Other National Authorities*), même s'ils ne transmettent pas les statistiques européennes directement à Eurostat. En effet, si les offices statistiques des *Länder* contribuent à produire les statistiques européennes pour leur territoire, c'est Destatis qui se charge de les transmettre à Eurostat.

6. Coût des démarches et formalités administratives pesant sur les personnes et les entreprises. Cette enquête n'entre pas dans le champ des enquêtes soumises à la loi statistique.

7. Après l'entrée en vigueur au début des années 2020 du règlement-cadre européen sur les statistiques sociales, l'enquête sur les revenus et les conditions de vie des ménages devra fournir, comme le fait déjà l'enquête emploi, des résultats régionaux pour certains indicateurs dans tous les pays de l'Union européenne.

Les offices régionaux de statistique ont développé une coopération entre eux, grâce au Comité de pilotage sur la coopération optimisée, La Optiko (*Lenkungsausschuss* « *Optimierte Kooperation* »), où siègent les chefs des offices statistiques (du niveau fédéral et des *Länder*). Ils y travaillent notamment à la mutualisation du développement et de la maintenance de logiciels ainsi qu'à la standardisation des processus de traitement et d'apurement des données.

« Si les offices statistiques des *Länder* contribuent à produire les statistiques européennes pour leur territoire, c'est Destatis qui se charge de les transmettre à Eurostat. »

Chacun d'entre eux publie ses propres résultats et analyses : mensuellement, pour l'indice des prix à la consommation (élaboré selon une méthodologie différente de celle de l'indice des prix harmonisé au niveau européen) et l'indice de la production industrielle ; trimestriellement, pour les salaires dans l'industrie. Ils diffusent également les résultats des enquêtes qu'ils ont collectées, soit au total l'équivalent des deux tiers des statistiques publiées au niveau fédéral⁸.

Ces offices régionaux peuvent par ailleurs réaliser différents types d'analyses et de recherches, prévisions et modélisations en relation avec les administrations régionales compétentes ou le niveau fédéral. À la demande de leur ministère de tutelle, ils peuvent aussi participer à la préparation de règlements juridiques et administratifs concernant le niveau fédéral ou régional.

Coopération et mutualisation sont aussi très développées pour les comptes régionaux. Les comptes nationaux de l'Allemagne relèvent de la responsabilité de Destatis. Une fois ceux-ci publiés, les comptes régionaux des *Länder* (et ceux établis à un niveau géographique encore plus fin) sont établis et publiés par un groupe de travail commun qui existe depuis 1954 (Frey et Thalheimer, 2010). Tous les offices statistiques des *Länder* en sont membres avec droit de vote, alors que Destatis y participe sans droit de vote⁹. Celui du Bade-Wurtemberg en assure la présidence et la direction. L'organisation retenue est originale : au lieu d'établir les comptes annuels pour son *Land*, chaque autorité régionale est responsable du calcul d'un agrégat économique ou d'un secteur économique pour l'ensemble des seize *Länder*. Cela garantit une cohérence d'ensemble des comptes des *Länder*, fondés sur les mêmes sources et des méthodologies identiques.

UN PRINCIPE DE CONCENTRATION FONCTIONNELLE DU SYSTÈME STATISTIQUE PUBLIC...

En Allemagne, Destatis et les offices statistiques des *Länder* sont les principaux producteurs de statistiques fédérales et européennes. À la différence de la France où la loi statistique affirme dès son article 1 que l'Insee et les services statistiques ministériels sont les deux composantes du service statistique public, la loi statistique allemande n'évoque à aucun endroit l'existence de services statistiques ministériels (ou d'administrations ayant une mission statistique, hors offices statistiques)¹⁰. Cela implique que les ministères n'ont pas

8. En mai 2019, la répartition est de 128 statistiques fédérales centralisées et 246 décentralisées.

9. L'office public local de la statistique et des élections de Francfort, en tant que représentant du Congrès des villes allemandes, est également membre du groupe de travail sans droit de vote.

10. Une seule mention d'agences ayant une mission de production statistique figure dans la loi - article 3(2) - mais seulement à propos de leur obligation de fourniture de données individuelles à l'office statistique fédéral à sa demande.

de légitimité à disposer d'un service statistique, à la différence de ce qui existe en France. Le système statistique public allemand est par conséquent caractérisé par une nette concentration fonctionnelle.

Ainsi, Destatis (avec l'aide des offices de *Länder*) est notamment responsable au niveau fédéral des statistiques économiques, démographiques et sociales et des comptes nationaux comme l'est l'Insee en France mais également de celles de l'agriculture, de l'environnement, du commerce extérieur, de la santé, de la protection sociale et de l'éducation, qui en France relèvent des services statistiques ministériels. Cette organisation présente certains avantages. Destatis et les offices statistiques des *Länder* peuvent ainsi potentiellement centraliser toutes les demandes des utilisateurs¹¹, si tant est qu'elles s'expriment effectivement, ce qui donne une plus grande cohérence aux travaux. Les ministères demeurent cependant l'instance de décision politique et budgétaire pour le travail statistique. Ils en sont les principaux utilisateurs et, à ce titre, réalisent aussi des analyses.

De la même manière, chaque office statistique régional répond aux demandes des ministères de son *Land* avec des priorités qui peuvent varier suivant les régions. Il en résulte certaines contraintes pour Destatis qui a un rôle de négociation, d'harmonisation et de coordination des productions statistiques. En effet, il doit publier, sur commande fédérale, les résultats fédéraux à partir de données qui doivent être également commandées par les ministères des *Länder*, et collectées par les offices régionaux dans la majorité des cas. La plupart d'entre elles ne sont transmises que sous forme anonyme et parfois seulement sous des formes agrégées.

📍 ... AVEC CEPENDANT QUINZE AUTRES PRODUCTEURS DE STATISTIQUE EUROPÉENNE

Des missions de statistique européenne sont assumées par quinze autres entités que Destatis au niveau fédéral. Au sein de ces entités, la mission statistique est assurée par des équipes de taille réduite. Si la concentration fonctionnelle affichée comme principe juridique est réelle en pratique, elle est moindre qu'aux Pays-Bas par exemple, où elle est poussée à l'extrême (il n'existe pas d'ONA).

« Outre les offices statistiques des *Länder*, les entités définies comme ONAs sont la Banque centrale (*Bundesbank*), qui produit les statistiques monétaires et celles de la balance des paiements, et des administrations fédérales aux statuts très variés. »

Outre les offices statistiques des *Länder*, les entités définies comme ONAs sont la Banque centrale (*Bundesbank*), qui produit les statistiques monétaires et celles de la balance des paiements, et des administrations fédérales aux statuts très variés. Il s'agit surtout d'offices, d'agences ou d'autorités fédérales dans des domaines très divers (emploi, migrations et réfugiés, automobile, environnement, agriculture et alimentaire, protection des consommateurs)¹², d'instituts spécialisés (emploi, santé, etc.) et enfin de deux ministères (Travail et affaires

11. On notera que les demandes de statistiques sont principalement passées aux offices statistiques, mais elles peuvent l'être à des universités ou à des instituts sans but lucratif, notamment les instituts dits de conjoncture..

12. La plupart des offices ou agences travaillent dans la sphère d'un ministère. Par exemple, l'agence fédérale de l'Agriculture et de l'alimentation, en tant que prestataire de services central du ministère fédéral de l'Agriculture, prépare des données sur divers sujets agricoles et nutritionnels. C'est une autorité centrale dont le siège est à Bonn. La base de données provient de diverses sources : les statistiques internes, les chiffres des *Länder*, du ministère fédéral de l'Agriculture, de l'office statistique fédéral et d'Eurostat au niveau international.

sociales, Éducation et recherche). Ce ne sont toutefois pas des services statistiques ministériels à part entière, au sens français du terme¹³ mais des ONAs qui produisent et diffusent à Eurostat des statistiques européennes dans leur domaine de compétence. Ceci diffère des offices statistiques des *Länder*, qui envoient leurs chiffres à Destatis qui les agrège ensuite et les transmet à Eurostat. Ces statistiques ne proviennent généralement pas d'enquêtes, mais de sources administratives.

📌 LES ÉVOLUTIONS DE LA RÉGLEMENTATION EUROPÉENNE ONT LEVÉ PLUSIEURS CONTRAINTES

La réglementation européenne a beaucoup évolué au cours des trente dernières années (Voir l'article d'Hervé Piffeteau dans ce même numéro). L'introduction du code de bonnes pratiques de la statistique publique en 2005 et les règlements statistiques de l'Union européenne ont permis aux autorités statistiques allemandes d'introduire une certaine souplesse dans leur dispositif législatif.

En particulier, la révision en 2015 du règlement européen n°223/2009 relatif aux statistiques européennes a donné aux Instituts nationaux de statistique (Destatis en Allemagne et l'Insee en France) un rôle de coordination de l'ensemble des activités en lien avec les statistiques européennes produites au niveau national. Elle a conféré également à ces mêmes instituts et aux autres autorités statistiques nationales (en particulier les offices statistiques des *Länder* en Allemagne) « un droit d'accès gratuit et immédiat à l'ensemble des fichiers administratifs de même qu'un droit d'utilisation et d'intégration de ces fichiers aux statistiques, etc. ». Ce règlement ayant force de loi dans les États membres pour ce qui a trait aux statistiques européennes, il a donné des marges de liberté nouvelles tout particulièrement au système statistique allemand ainsi qu'à Destatis pour y faire évoluer son rôle et son positionnement¹⁴.

Encadré 2. Les missions autres que statistiques de Destatis

Les missions de Destatis se limitent à la production statistique et ne couvrent pas les études économiques, ce qui est le cas de la quasi-totalité des INS dans le monde (à l'exception notable de l'Insee). Pour autant, ceci n'empêche pas celui-ci de participer à d'autres activités.

Son président est en charge de l'organisation des élections législatives et de celles au Parlement européen, ainsi que de la publication de leurs résultats. Cette particularité allemande est unique au monde à notre connaissance.

Par ailleurs, Destatis apporte son soutien matériel et technique à un Conseil d'experts en économie*, dénommé « les Cinq Sages de l'économie », qui est abrité dans ses locaux. Ce Conseil a été créé en 1963 dans le but de conseiller le gouvernement fédéral et le *Bundestag* (Parlement fédéral) sur les questions économiques, en particulier en réalisant des analyses et prévisions conjoncturelles.

* Der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung.

13. Il existe cependant des unités ministérielles qui ont le terme « statistique » dans leur intitulé, par exemple : la division M5 (statistiques de migration) du ministère de l'Intérieur, la division I-C3 (Direction I, département 3) du ministère de l'Économie et l'énergie ou encore les divisions I-1b-4 et 5 et la division II-a4 du ministère du Travail et des affaires sociales. Le ministère de l'Agriculture a un site dédié aux statistiques : <http://www.bmel-statistik.de/>.

14. Pour la France, l'impact de ce règlement est faible dans ces deux domaines, puisque l'Insee accède aux données administratives et coordonne le système statistique français de longue date (Décret n° 46-1432 du 14 juin 1946, article 1).

Le règlement statistique de 2009 avait, lui, donné la possibilité à Eurostat et aux Instituts nationaux de statistique d'ouvrir l'accès des données statistiques confidentielles (indirectement nominatives) à des chercheurs, à des fins scientifiques. Là encore, le règlement ne portait que sur les données intervenant dans la statistique européenne, mais son impact juridique et pratique était fort en Allemagne du fait du développement de la statistique européenne au cours des dernières décennies.

En outre, la revue par les pairs, menée en 2015 dans le cadre européen, avait adressé au système statistique public et aux pouvoirs publics allemands plusieurs recommandations pour une meilleure mise en œuvre du code de bonnes pratiques, qui concernaient notamment l'accès aux données administratives et le rôle de Destatis (Fenwick, Hackl et Potisepp, 2015).

Ainsi la dernière révision de la loi statistique, votée le 27 juillet 2016, comportait plusieurs orientations :

- ① une plus grande utilisation des fichiers administratifs, notamment dans l'objectif de réduire la charge de réponse aux enquêtes ;
- ① une plus grande ouverture des données statistiques aux chercheurs (voir *infra*) ;
- ① un assouplissement des obligations législatives pour la production des statistiques fédérales, en particulier, pour les enquêtes *ad hoc* qui pourraient être décidées par le biais d'actes réglementaires.

① LE RECOURS AUX SOURCES ADMINISTRATIVES EST ÉLARGI MAIS RESTE TRÈS ENCADRÉ

“ Les offices des Länder ne peuvent recourir à des fichiers d'adresses que pour un objectif statistique particulier mentionné par une loi spécifique. ”

L'usage des données administratives reste très encadré en Allemagne et doit être réalisé dans le respect des règles de confidentialité¹⁵. En particulier, dans un contexte de forte protection des données, leur récupération est conditionnée à leur anonymisation par la suppression des identifiants individuels (pour les données sur les ménages). Ceci interdit de fait les appariements avec d'autres sources pour ces données, sachant que les appariements sont possibles pour une liste limitative de données qui sont surtout économiques (cf. le répertoire d'entreprises). Destatis et les offices des

Länder ne peuvent recourir à des fichiers d'adresses que pour un objectif statistique particulier mentionné par une loi spécifique.

La révision de la loi statistique de 2016 autorise l'office statistique fédéral à demander aux administrations concernées des métadonnées et des données de test. Ces premières ouvertures devraient se renforcer sensiblement dans les toutes prochaines années. En effet, le gouvernement fédéral allemand souhaite développer l'administration numérique, notamment par la mise en commun des données entre services ministériels grâce à des registres administratifs interconnectés. Ces évolutions ouvriraient largement les sources

15. Une nouvelle loi fédérale sur la protection des données a été adoptée en 2018 pour prendre en compte l'entrée en application dans tous les pays de l'Union européenne du règlement général sur la protection des données (RGPD).

administratives au système statistique public allemand et le mettraient en conformité avec le règlement statistique européen 223/2009, qui prévoit une large utilisation de ces sources (Bens et Schukraft, 2019).

Par ailleurs, Destatis a lancé le projet de remplacer les recensements de population par l'utilisation de registres administratifs à l'horizon 2031.

🌐 UN ACCÈS PLUS LARGE POUR LE MONDE DE LA RECHERCHE —

La création par l'office statistique fédéral d'un centre de données pour mettre à disposition des données statistiques aux chercheurs remonte à 2001. Elle a été suivie dès 2002 par la création d'un centre de données par les offices des *Länder*. Aujourd'hui, les données individuelles sont stockées sur un site centralisé et accessibles via un portail unique des offices statistiques (fédéral et régionaux).

Du fait d'une prudence juridique de la part du système statistique public, et de l'absence d'organismes équivalents à ceux qui existent en France (Commission nationale pour l'informatique et les libertés, Comité du secret statistique), la communication de données individuelles à l'extérieur des offices statistiques était assez restreinte jusqu'à la révision de la loi statistique de 2016. Celle-ci a élargi l'accès des chercheurs aux micro-données statistiques, avec l'ajout d'un article précisant les conditions d'accès à des données individuelles « formellement anonymisées ».

🌐 LE RÔLE DE COORDINATION DE DESTATIS A ÉTÉ RENFORCÉ DANS LE DROIT

Le rôle de coordination de Destatis a été renforcé par la révision en 2015 du règlement européen n°223/2009. Celui-ci a conféré aux instituts nationaux statistiques un rôle de coordination du système statistique national pour les statistiques européennes. Depuis 2016, Destatis organise ainsi avec tous les ONAs (hors offices de *Länder*) une réunion annuelle d'un groupe de travail sur les statistiques européennes, sachant qu'un accord-cadre a été signé avec chacun de ces ONAs.

Avec les offices statistiques des *Länder*, Destatis met en œuvre cette coordination renforcée par la voie d'un « pilotage par la qualité », qui consiste à s'assurer de la qualité des processus statistiques et de celle des données produites.

Le rapport de la revue par les pairs, conduite par Eurostat en 2015, va dans ce sens. Il met l'accent sur le fait que Destatis est en

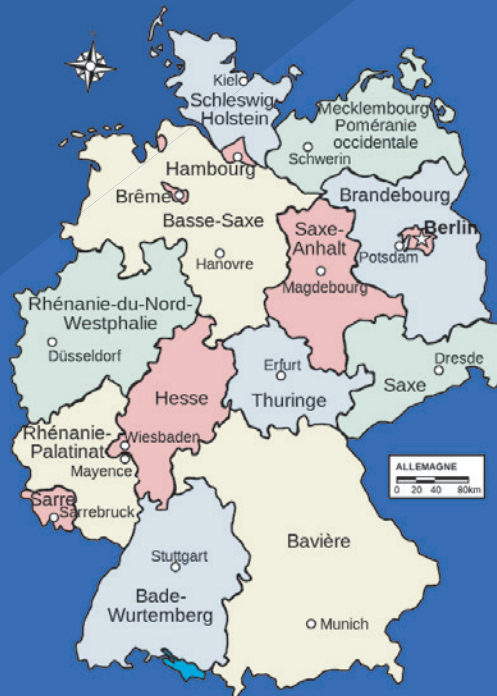
« Destatis est en charge de veiller au niveau élevé de conformité au code de bonnes pratiques de l'ensemble des statistiques fédérales, y compris celles produites par les ONAs. »

charge de veiller au niveau élevé de conformité au code de bonnes pratiques de l'ensemble des statistiques fédérales, y compris celles produites par les ONAs.

Au moment de la rédaction de cet article, Destatis était en train de lancer une stratégie de suivi de la qualité des statistiques décentralisées (produites en coopération avec les offices des *Länder*), complétant la stratégie de suivi de la qualité des statistiques produites par Destatis seul

(statistiques centralisées) adoptée en 2018. Par ailleurs, la préparation d'un programme statistique national consolidé (qui était aussi une recommandation de la revue par les pairs) a démarré et pourrait être achevée prochainement.

Figure 1. Carte de l'Allemagne et des 16 Länder



Note : il existe 14 offices statistiques pour les 16 Länder administratifs, sachant que l'office de Berlin-Brandebourg (basé à Berlin) et celui de Hambourg-Schleswig-Holstein (basé à Hambourg) couvrent chacun deux Länder.

ALEXANDRE, Véronique, DE GIMEL, Louis et HEINLEIN, Dirk, 1997. Environnement institutionnel de la statistique – Comparaison France-Allemagne. In : *Courrier des statistiques*. [en ligne]. Décembre 1997. N°83-84, pp. 35-64. [Consulté le 24 septembre 2019]. Disponible à l'adresse : <https://www.epsilon.insee.fr/jspui/bitstream/1/14351/1/cs83-84.pdf>

BENS, Arno et SCHUKRAFT, Stefan, 2019. Modernisation des registres administratifs en Allemagne : développements actuels et enjeux pour la statistique publique. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp. 10-20. [Consulté le 24 septembre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168390/courstat-2-3.pdf>

DESTATIS, 2018. Ihr Nutzen. Unser Auftrag. In : *site web de Destatis*. [en ligne]. Wiesbaden. [Consulté le 24 septembre 2019]. Disponible à l'adresse : <https://www.destatis.de/DE/UEber-uns/Ziele-Strategie/ihr-nutzen-unser-auftrag-2018.pdf>

FENWICK, David, HACKL, Peter G. et POTISEPP, Priit, 2015. Peer Review Report on compliance with the Code of Practice and the coordination role of the National Statistical Institute : Germany. In : *site web d'Eurostat*. [en ligne]. Février 2015. Luxembourg. [Consulté le 24 septembre 2019]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/documents/64157/4372828/2015-DE-Report>

FREY, Jeanette et THALHEIMER, Frank, 2010. Germany Regional GVA Inventory. In : *site web du Baden-Württemberg Statistisches Landesamt*. [en ligne]. Juin 2010. Stuttgart. [Consulté le 24 septembre 2019]. Disponible à l'adresse : https://www.statistik-bw.de/VGRdL/MethDef/DE_GVA-NUTS2_Inventory_en-EN.pdf

LEFÈVRE, Cécile, 1998. *La statistique en Allemagne*. [en ligne]. Octobre 2018. Insee. Regards à l'étranger, N° 2. [Consulté le 27 septembre 2019]. Disponible à l'adresse : https://www.epsilon.insee.fr/jspui/bitstream/1/64635/1/Regards_stat_allemande.pdf

LEFÈVRE, Cécile, 1999. Le système statistique public allemand – Un homologue proche, mais en partie méconnu. In : *Courrier des statistiques*. [en ligne]. Décembre 1999. N° 91-92, pp. 5-14. [Consulté le 24 septembre 2019]. Disponible à l'adresse : <https://www.epsilon.insee.fr/jspui/bitstream/1/8525/1/cs91-92b.pdf>

RIVIÈRE, Pascal M., 2019. Le nouveau Mikrozensus en Allemagne ; un dispositif intégré des enquêtes auprès des ménages à l'horizon 2020. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N° N2, pp. 10-20. [Consulté le 24 septembre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168392/courstat-2-4.pdf>



Présentation du numéro N3

Le numéro N3 est placé sous le signe de l'innovation dans la statistique publique. Ces innovations se présentent sous les aspects les plus divers. Pour concrétiser l'activation des métadonnées dans le processus statistique, l'Insee a construit un écosystème autour de sa plateforme de collecte par internet ; il a conçu un générateur automatique de supports de collecte (*web*, papier, face-à-face, etc.), développé l'outil du concepteur pour décrire son questionnaire dans les standards internationaux de métadonnées, et bâti ainsi une offre de service pour les enquêtes entreprises. La méthodologie de l'indice des prix à la consommation (IPC) change à partir de janvier 2020, utilisant désormais des *big data* particulières : les données de caisses. Le Centre d'accès sécurisé aux données (CASD), déjà novateur en soi, innove en 2019 avec une technique de certification des recherches fondées sur données confidentielles. Enfin, les deux nouveaux règlements européens sur les statistiques d'entreprise (FRIBS) et sociales (IESS) vont rationaliser le cadre commun aux instituts nationaux de statistique, qui s'inscrit lui-même dans un socle juridique précis. Ils vont avoir des conséquences concrètes pour la production tout en assurant plus de cohérence. Loin de contraindre uniquement, la réglementation européenne peut jouer un rôle facilitateur : c'est le cas avec le système statistique allemand.

ISSN 2107-0903
ISBN 978-2-11-1151253-5

