



Conseil national
de l'information statistique

Montrouge, le 12 janvier 2021 – n° 5/H030

COMMISSION
SYSTÈME FINANCIER ET FINANCEMENT DE L'ÉCONOMIE

Réunion du 18 novembre 2020

PROJET D'AVIS	2
COMPTE RENDU DE LA RÉUNION	3
LISTE DES PARTICIPANTS	4
LISTE DES DOCUMENTS PRÉPARATOIRES ET COMPLÉMENTAIRES.....	31

PROJET D'AVIS

COMMISSION « SYSTÈME FINANCIER ET FINANCEMENT DE L'ÉCONOMIE » du 18 novembre 2020

Avis n° 1

L'exploitation statistique du gisement d'informations individuelles bancaires et financières produit à des fins de régulation et d'évaluation des politiques a constitué une « révolution » dans les années post-crise de 2008-2009. Ces données granulaires au niveau des ménages ou des firmes sont devenues un outil clé dans l'analyse empirique, car elles permettent d'identifier les mécanismes sous-jacents aux grandes relations économiques et de mieux cerner la distribution des risques afin d'appréhender les enjeux de stabilité financière. Depuis quelques années, des travaux sont en cours tant au niveau national qu'au niveau européen pour développer le recours à ces données et la commission a estimé qu'il était pertinent de faire le point sur ce sujet.

Les interventions de la Banque de France et du CASD (centre d'accès sécurisé aux données) mettent en lumière deux grands enjeux pour la statistique publique dans la mise à disposition de ces données. Le premier est celui de leur accès. En France, les deux sources principales sont les bases de la Banque de France, mises à disposition via l'*Open Data Room* et celles de l'Insee, disponibles via le CASD. Dans la continuité des recommandations de la réunion du 23 mai 2019, la commission souligne qu'un projet de rapprochement des accès est en cours de discussion, dont l'objectif serait de permettre aux chercheurs de disposer au sein d'un même environnement des données granulaires financières avec des données socio-économiques. Le second enjeu de disponibilité concerne le cloisonnement national des données, qui rendait difficile jusqu'à récemment l'appréhension globale des enjeux internationaux. Il existe depuis peu plusieurs projets visant à rapprocher des sources nationales et à les agréger au niveau européen (AnaCredit sur l'enrichissement des données sur le crédit, SHS, RIAD, EMIR).

La commission soutient fortement ces initiatives qui visent à faciliter un accès aux données transdisciplinaires intégrant les aspects financiers et à encourager l'ouverture aux données européennes, compte tenu de leur qualité, pour tous les travaux à finalité de recherche scientifique, contribuant ainsi à une meilleure compréhension des enjeux économiques internationaux. Elle souhaite être informée de la progression des travaux sur ce sujet, notamment sur l'instruction des questions juridiques relativement à l'accès aux données par les chercheurs et à la protection de leur confidentialité dans leur usage.

Référence : avis généraux de moyen terme n° 1 et 8, avis de moyen terme n° 4 et 8 de la commission « Système financier et financement de l'économie »

COMPTES RENDUS DE LA RÉUNION DE
LA COMMISSION « SYSTÈME FINANCIER ET FINANCEMENT DE L'ÉCONOMIE »
- 18 novembre 2020 -

Président : Anne-Laure DELATTE, chargée de recherche au CNRS, rattachée au Leda, Paris Dauphine

Rapporteurs : Guillaume HOURIEZ, Chef du département des comptes nationaux à l'Insee
Cécile GOLFIER, chargée de mission à la Direction générale des statistiques, des études et de l'international (DGSEI) à la Banque de France

Responsable de la commission : Stéphane TAGNANI (01 87 69 57 03)

RAPPEL DE L'ORDRE DU JOUR

INTRODUCTION.....	5
I. INFORMATION SUR LES CONSÉQUENCES DE LA CRISE SANITAIRE DU PRINTEMPS : ÉVOLUTIONS DU PROGRAMME DE TRAVAIL 2020 (PAT 2020) ET GESTION DES AVIS D'OPPORTUNITÉ (AO)	5
II. L'UTILISATION DES DONNÉES MASSIVES POUR LE SUIVI CONJONCTUREL LORS DU CONFINEMENT DÙ À LA PANDÉMIE	8
III. ACCÈS AUX DONNÉES GRANULAIRES FINANCIÈRES : ÉVOLUTIONS RÉCENTES	11
CONCLUSION	30

Liste des participants

ADAM	David	Banque de France (BdF)
ANXIONNAZ	Isabelle	Conseil national de l'information statistique (Cnis)
BAGEIN	Guillaume	Ministère des Solidarités et de la santé - Direction de la recherche, des études, de l'évaluation et des statistiques (Drees)
BERTIAUX	Juliette	Chambre de commerce et d'industrie de Paris-Ile-de-France
BRIDENNE	Isabelle	Caisse des dépôts et consignations (CdC)
CAPELLE BLANCARD	Gunther	Université Paris 1 Panthéon-Sorbonne
DELATTE	Anne-Laure	Université Paris Dauphine
DIETSCH	Michel	Université Strasbourg
DUBOIS	Marie- Michèle	Conseil national de l'information statistique (Cnis)
DUPRAT	Vincent	Banque de France (BdF)
GADOUCHE	Kamel	Centre d'accès sécurisé distant aux données (CASD)
GAZANIOL	Alexandre	Banque publique d'investissement France (Bpifrance)
GOLFIER	Cécile	Banque de France (BdF)
HADJIBEYLI	Benjamin	Ministère de l'Economie, des finances et de la relance - Direction générale du trésor (DGT)
HEYDEMANN	Pascale	Institut français du cheval et de l'équitation (IFCE)
HOURIEZ	Guillaume	Institut National de la statistique et des études économiques (Insee) - Direction des études et des synthèses économiques (Dese)
ISRAËL	Jean-Marc	Banque de France (BdF)
KREMP	Elisabeth	Institut National de la statistique et des études économiques (Insee) - Direction des statistiques d'entreprises (DSE)
MAUREL	Françoise	Conseil national de l'information statistique (Cnis)
MOREAU	Delphine	Banque de France (BdF)
MOREAU	Sylvain	Institut National de la statistique et des études économiques (Insee) - Direction des statistiques d'entreprises (DSE)
PEZZIN	Guillaume	Ministère de l'Economie, des finances et de la relance - Direction générale des finances publiques (DGFIP)
POUGET	Julien	Institut National de la statistique et des études économiques (Insee) – Direction des études et des synthèses économiques (Dese)
ROBIN	Benoît	Confédération générale du travail - Force ouvrière (CGT-FO)
SILBERMAN	Roxane	Centre national de la recherche scientifique (CNRS)
SIMON	Olivier	Institut National de la statistique et des études économiques (Insee) – Direction des études et des synthèses économiques (Dese)
TAGNANI	Stéphane	Conseil national de l'information statistique (Cnis)
TALLET	Frédéric	Institut National de la statistique et des études économiques (Insee) - Direction des études et des synthèses économiques (Dese)
THORNARY	Baptiste	Bpifrance

INTRODUCTION

Anne-Laure DELATTE

Bonjour à tous. Nous avons un programme très chargé, mais passionnant. Trois points sont prévus. Nous évoquerons tout d'abord les conséquences de la crise sanitaire du printemps sur le programme de travail et la gestion des avis d'opportunité. L'Insee reviendra ensuite sur l'utilisation des données massives pour le suivi conjoncturel lors du confinement dû à la pandémie, c'est-à-dire l'utilisation des données de téléphonie, de Google, etc. Enfin, nous aborderons le sujet principal de notre commission qui fait partie de mes sujets de prédilection et de ceux de mon prédécesseur, qui est parmi nous aujourd'hui, sur l'accès aux données granulaires financières. Plusieurs présentations nous apporteront des éléments sur l'évolution de l'accès aux données individuelles, notamment l'appariement des données granulaires et les démarches conduites au niveau transnational.

I. INFORMATION SUR LES CONSÉQUENCES DE LA CRISE SANITAIRE DU PRINTEMPS : ÉVOLUTIONS DU PROGRAMME DE TRAVAIL 2020 (PAT 2020) ET GESTION DES AVIS D'OPPORTUNITÉ (AO)

I.1 Présentation du PAT 2020

Cécile GOLFIER, Banque de France

Nous allons vous présenter les adaptations du programme de travail de l'Insee, la DGFIP, la Banque de France et l'ACPR (l'Autorité de contrôle prudentiel et de résolution) face à cette situation de crise.

Face à la crise sanitaire de la Covid-19, la Banque de France et l'ACPR ont parallèlement intensifié l'usage des statistiques financières existantes et innové.

Côté Banque de France, la direction des statistiques monétaires et financières a immédiatement pris conscience de l'attente des utilisateurs de données pour disposer de données en temps utile sur tout le champ de la statistique. La direction s'interrogeait sur la capacité des fournisseurs à livrer les données. Elle s'est donc rapprochée des déclarants pour engager un dialogue de proximité, ce qui a permis le respect des échéances de remise. Collectivement, nous avons bénéficié du fait que les banques avaient intégré dans leur plan de continuité les obligations de *reporting*. Les données ayant été remises comme attendu, la qualité des statistiques monétaires et financières a pu être maintenue.

Deux problématiques particulières étaient identifiées : la distribution du crédit et le comportement d'accumulation de liquidités par les ménages. La crise a montré la pertinence des statistiques monétaires et financières de la Banque de France pour faire face à une situation totalement inédite. Ces statistiques ont permis de produire une nouvelle publication portée sur la page « Suivi de crise Covid-19 » qu'a créée la Banque de France sur son site internet. Une note mensuelle a notamment été publiée sur l'impact de la crise de la Covid-19 sur la situation financière des ménages et des entreprises.

La direction des entreprises tenait quant à elle à maintenir la fiabilité de la cotation des entreprises. En juillet 2020, elle a interrogé les entreprises sur l'impact de la crise sur l'évolution des agrégats financiers et de la liquidité, les mesures de soutien utilisées et autres questions relatives au financement. Cette enquête a permis de produire une synthèse, portée sur la même page « Suivi de crise Covid-19 », montrant notamment que 44 % des entreprises répondantes ont eu recours à un prêt garanti par l'État ou que 21 % des PME ont renégocié leur prêt en conséquence de la crise.

Les publications préexistantes ont été adaptées, avec des focus sur la crise financière. En termes d'actualité financière, des billets de blog ont été consacrés aux correspondants TPE-PME, à la médiation du crédit ou au lien entre liquidités, solvabilité bancaire et crise financière. Le *Bulletin* de la Banque de France a, pour sa part, proposé des articles permettant de faire un point sur la situation d'avant-crise, que ce soit sur les ménages, comme l'étude sur le patrimoine et l'endettement des ménages en France, en Allemagne et en Italie avant la crise, ou sur les entreprises, avec une étude de la structure financière des PME.

L'Autorité de contrôle prudentiel et de résolution (ACPR) a elle aussi réagi à la crise sanitaire selon deux axes. Elle a intensifié l'exploitation des bases de données existantes et noué des échanges rapprochés avec les principaux groupes et établissements pour analyser l'impact de la crise sur les banques et les assurances. Parallèlement, lors du 1^{er} confinement, l'ACPR et les autorités de supervision européennes, l'EBA (*European Banking Authority*) et l'EIOPA (*European Insurance and Occupational Pensions Authority*) ont assoupli les conditions de *reporting* des remises prudentielles.

Guillaume HOURIEZ, Insee

Avant d'évoquer les adaptations des comptes trimestriels, j'ai relevé deux points importants dans les programmes de travail de l'Insee et du bureau GF-3C de la DGFIP.

L'Insee a ouvert le code du modèle Avionic qui mobilise des tableaux des entrées et sorties (TES) afin de modéliser les interactions entre les différents secteurs d'activité ou les impacts sur l'emploi. Ce modèle permet par exemple de mesurer les conséquences à court terme sur l'ensemble de l'économie d'un choc sur des secteurs d'activité particuliers. Il incorpore aussi des contenus des imports qui, couplés avec les TES internationaux de l'OCDE et d'Eurostat, permettent une analyse des chaînes de valeur. De son côté, la DGFIP a acquis et mis en production de nouvelles bases pour répondre à des demandes des services statistiques ministériels, de l'Insee et du monde académique. Quatre nouveaux fichiers ont ainsi été produits : les mouvements sur les comptes bancaires (Ficoba), les droits de mutation à titre gratuit (DMTG) les déclarations de revenus des utilisateurs de plates-formes et les déclarations des dons reçus par les associations.

Face à la crise sanitaire, l'ensemble du système statistique public a dû s'adapter. Il l'a fait de deux façons : d'une part, en publiant des indicateurs nouveaux pour répondre à une demande nouvelle et d'autre part, en adaptant la méthodologie des indicateurs existants pour faire face à des problèmes de remontée d'informations.

S'agissant des comptes trimestriels, le défi se révélait particulièrement redoutable. Il a nécessité de nombreuses adaptations et différents changements d'indicateurs. Je ne les commenterai pas tous, vous invitant plutôt à vous reporter aux notes méthodologiques qui ont accompagné la publication des comptes trimestriels. Dans le cadre des premières estimations à 30 jours pour lesquelles il manque un mois sur toute la partie « production de services », nous avons pu nous appuyer sur les travaux du département de la conjoncture de l'Insee qui a mesuré la perte instantanée d'activité liée au confinement. Ces travaux vous seront présentés par Julien Pouget, le chef du département de la conjoncture. Pour les comptes d'agents, nous avons mobilisé un grand nombre de données administratives, en part des données de la DGFIP, notamment pour estimer l'ensemble des aides aux ménages et aux entreprises.

Les activités non marchandes ont posé un double problème. Ces activités regroupent l'ensemble des services produits notamment par les administrations publiques et fournis gratuitement ou presque aux ménages et à toute la collectivité (santé, éducation, administration générale, etc.). Leur mesure est conventionnelle alors que ces activités représentent environ un cinquième du PIB. Lors des crises précédentes, ces activités n'avaient jamais été impactées. Les comptes trimestriels n'ont donc pas développé d'indicateurs infra-annuels de suivi, se contentant de retracer l'activité tendancielle de ces services non marchands.

Or durant la crise sanitaire, le confinement a affecté les fonctionnaires de la même façon ou presque que les salariés du privé. Nous pouvons donc penser que le choc d'offre subi par l'activité marchande a touché aussi les activités non marchandes. Pour rendre compte de ce choc, les comptes trimestriels français ont mobilisé des données, en particulier les données de l'enquête Emploi et les données d'activité des hôpitaux complétées par les coûts de réorganisation et les coûts salariaux, pour mesurer la perte d'activité dans les services non marchands.

Même si les modalités de confinement ont différé d'un pays à l'autre et les administrations avaient certainement des capacités de réaction différentes, nous pouvons penser malgré tout que le confinement a dû engendrer une baisse d'activité à peu près similaire ou du moins proche entre les activités marchandes et les activités non marchandes. Pour trois pays parmi les cinq que nous avons retenus, l'Allemagne, la France et le Royaume-Uni, l'activité non marchande évolue de façon proche du reste de l'activité. En Espagne, l'activité non marchande est restée constante alors que l'activité a chuté en glissement annuel de plus de 20 points. Ainsi, malgré la coordination d'Eurostat qui a notamment produit un guide

méthodologique, il reste sans aucun doute des problèmes de comparabilité entre les mesures de perte d'activité d'un pays à l'autre. Il faudra y travailler à l'avenir et des révisions pourraient intervenir dans un certain nombre de pays.

Pour plus d'informations, je vous renvoie vers les différents points de conjoncture qui ont eu lieu pendant la crise sanitaire et après, rassemblés dans un ouvrage, mais aussi vers les publications des comptes trimestriels qui intègrent des notes méthodologiques expliquant nos différents traitements et les notes de blog de l'Insee qui ont évoqué notamment la statistique publique à l'épreuve de la crise sanitaire, la mesure de l'activité pendant la crise, les nouvelles données pour suivre la conjoncture économique et le modèle Avionic.

I.2 Présentation de la gestion des avis d'opportunité

Isabelle ANXIONNAZ, Cnis

Cette présentation a pour objectif de vous informer des adaptations du Cnis en termes de consultation et d'attribution des avis d'opportunité dans le contexte de la crise sanitaire, autour de la première période de confinement et de celle-ci. Deux grands principes ont présidé à ces adaptations : d'une part la poursuite de la concertation malgré l'annulation des commissions avec la mise en place d'un nouveau mode de consultation, et d'autre part, la volonté d'accompagner les producteurs dans la modification du système d'information.

S'agissant du maintien de la concertation, suite à l'annulation des commissions de printemps et pour limiter les retards dans les programmes de travail, nous avons mis en place une consultation électronique sur les avis d'opportunité et les demandes d'accès aux données administratives régies par l'article 7bis de la loi de 1951. Les demandes ont été mises en ligne sur le site du Cnis. Nous avons envoyé un message de consultation aux utilisateurs de la commission concernée, leur expliquant les modalités de la consultation, puis nous avons compilé les questions et les réponses des utilisateurs et des producteurs dans un document de synthèse qui a été envoyé au président de la commission qui a pu donner son avis d'opportunité en toute connaissance de cause. La consultation s'est terminée par la mise en ligne du compte rendu de consultation. Pour la commission « Système financier et financement de l'économie », la consultation a porté sur une seule enquête, l'enquête Crédit-bail de l'Insee, à laquelle un avis d'opportunité favorable a été attribué. Nous avons par ailleurs mis en ligne les programmes de travail des services producteurs qui n'avaient pas pu être présentés aux commissions de printemps, décidant que les modifications de ces programmes seraient présentées lors des commissions d'automne.

Nous avons également accompagné les modifications du système d'information pour que les producteurs puissent apporter rapidement un éclairage sur les conséquences sanitaires, sociales et économiques de la pandémie. Nous avons ainsi été conduits à attribuer des avis d'opportunité en urgence selon une procédure prévue dans les textes relatifs à l'organisation et au fonctionnement du Cnis. L'article 2 précise en effet que : *« en cas d'urgence constatée par le président du Conseil national de l'information statistique et après avis du président de la commission thématique concernée, une enquête peut être réalisée, même si elle n'a pas été inscrite au programme de l'année. Elle est néanmoins soumise au comité du label en vue de l'obtention du visa prévu à l'article 20 »*.

Cette mesure a été mise en œuvre pendant le 1^{er} confinement pour quatre enquêtes, notamment l'enquête Impact de la crise sanitaire sur l'organisation et l'économie des entreprises. Pour ces enquêtes, compte tenu du contexte, une consultation classique en commission n'était pas envisageable, même de manière électronique, dans les délais très courts qui étaient prévus pour la collecte sur le terrain. Sur la base d'une fiche descriptive de l'enquête, d'un projet de questionnaire et après avoir pris l'avis des présidents des commissions concernées, le président du Cnis a constaté l'urgence de ces enquêtes et émis un avis favorable à leur réalisation. Les projets, après avoir été instruits par le comité du label, ont pu obtenir un visa ministériel de la statistique publique et être inscrits au programme statistique publié au Journal officiel.

Toutes ces étapes ont été franchies très rapidement, puisqu'il ne s'est écoulé qu'une semaine ou deux entre les premiers contacts avec les services producteurs et le lancement de l'enquête sur le terrain. Dans l'intervalle, les descriptifs des enquêtes, les décisions d'opportunité et les avis d'examen délivrés par le comité du label ont été mis en ligne sur le site du Cnis sous la forme usuelle. Pour assurer la transparence, chaque procédure en urgence a donné lieu à une communication particulière sur le site du Cnis, dans la

rubrique des actualités. En parallèle, une lettre d'actualité reprenant l'ensemble des procédures a été adressée à tous nos contacts.

Parallèlement aux enquêtes nouvelles, élaborées et réalisées dans l'urgence, d'autres enquêtes ont été modifiées. Nous avons ainsi été amenés à mettre à jour des avis d'opportunité en cours afin d'intégrer des modules d'interrogation Covid, comme nous l'avons fait pour l'enquête Emploi. Nous avons également été conduits à prolonger des avis d'opportunité pour accompagner des décalages de réalisation d'enquêtes. Pour la présente commission, nous l'avons fait pour l'enquête complémentaire sur les échanges internationaux de service (ECEIS) de la Banque de France.

Dans cette période un peu particulière où la souplesse apparaît nécessaire, ces adaptations nous ont permis de continuer à assurer la mission essentielle de concertation du Cnis tout en permettant d'accompagner les producteurs. Nous allons capitaliser sur certaines de ces expériences, non pas pour accroître les décisions d'opportunité en urgence, qui ont vocation à rester exceptionnelles, mais plutôt pour diversifier les modes de consultation en ayant recours plus régulièrement à des consultations écrites en ligne, à l'instar de la pratique développée pour les commissions annulées au printemps, et à des consultations restreintes et courtes, réunissant des membres du Cnis, en amont et en aval des commissions, en prévoyant bien entendu un retour systématique lors des commissions formelles.

II. L'UTILISATION DES DONNÉES MASSIVES POUR LE SUIVI CONJONCTUREL LORS DU CONFINEMENT DÛ À LA PANDÉMIE

Julien POUGET, Insee

Je vous présenterai succinctement notre démarche depuis le début de la crise. Le contexte n'a pas rendu muettes les données traditionnelles, mais il a fortement impacté les enquêtes de conjoncture. Dans ces enquêtes, nous demandons aux chefs d'entreprise leur opinion sur l'activité récente et leurs perspectives. Ces enquêtes sont mensuelles. Néanmoins, pendant la première quinzaine de mars, les réponses reflétaient assez peu l'état du monde à la fin du mois. Nous calculons par ailleurs des indicateurs synthétiques qui agrègent les opinions sur le passé récent et le futur proche, dont l'utilisation devient difficile quand ils agrègent des informations sur des états du monde très différents.

Durant cette période, s'agissant des enquêtes de conjoncture nous avons donc privilégié la lecture directe des données brutes. Nous avons aussi abondamment utilisé l'enquête Acemo-Covid menée par la Dares en lien avec l'Insee. Les dernières enquêtes de conjoncture font apparaître, dans l'industrie, un très fort rebond, suivi d'une inquiétude sur les perspectives personnelles de production selon les données recueillies début octobre. Ces inquiétudes sont encore plus marquées dans les services et notamment les plus touchés, comme l'hébergement et la restauration.

En parallèle, nous avons utilisé de manière accrue les données à haute fréquence qui résultent de l'accroissement des traces numériques des activités humaines. Ces données sont multiformes : données de caisse, transactions par cartes bancaires, activation des réseaux de téléphonie mobile, requêtes sur les moteurs de recherche, services de géolocalisation, etc. Ces données sont disponibles rapidement, souvent de manière quasi instantanée. Elles sont, du moins en apparence, homogènes d'un pays à l'autre. Leur coût marginal est proche de zéro et la charge de collecte est faible, voire insensible pour les particuliers ou les entreprises concernées.

Ces données se sont retrouvées sur le devant de la scène avec la crise, car elles sont bien adaptées aux circonstances extrêmes que nous avons connues : un choc très soudain et de très grande ampleur. L'utilisation de certaines de ces sources n'est pour autant pas totalement nouvelle, que ce soit pour la production statistique ou la prévision économique. Par exemple, depuis plusieurs trimestres, les données de caisse de la grande distribution sont utilisées par l'Insee pour calculer l'indice des prix à la consommation. Pour le diagnostic conjoncturel, certaines de ces données figurent depuis quelque temps dans la boîte à outils des conjoncturistes. Elles constituent en temps normal un substitut assez utile aux données traditionnelles tant que ces dernières ne sont pas disponibles. En revanche, leur attrait diminue lorsque les données habituelles sont calculables.

Depuis le début de la crise, nous nous sommes concentrés sur deux agrégats macroéconomiques : l'activité économique au sens du produit intérieur brut et la consommation des ménages. Sur ces deux agrégats, les

données massives ont contribué en partie à l'élaboration du diagnostic conjoncturel, à côté des données d'enquêtes, des remontées de fédérations professionnelles ou de ce que nous savions sur la réglementation en vigueur et la fermeture des activités dites non essentielles.

Ces données permettent d'établir un certain nombre de diagnostics, comme le montre le dixième point de conjoncture de l'année 2020 paru hier. Les données de consommation d'électricité par les entreprises directement raccordées au réseau, c'est-à-dire les plus gros consommateurs, font apparaître, malgré un gros creux en avril, un maintien dans de nombreux secteurs sauf le transport et l'entreposage. L'automobile était quasiment à l'arrêt au mois d'avril, mais aujourd'hui les usines continuent de tourner. À l'inverse, dans les services, les contrastes sectoriels sont très marqués et découlent du degré d'exposition aux mesures de protection sanitaire. Les fréquences de recherche de certains mots-clés (train, restaurant, cinéma, etc.) sont quant à elles retombées à un niveau proche du niveau très bas atteint durant le premier confinement.

Ces indicateurs retracent bien les contrastes sectoriels et offrent la possibilité de réaliser des comparaisons entre les pays, même si la documentation n'est pas forcément exhaustive et si l'homogénéité n'est pas toujours pleinement assurée. Sur le mois de novembre, nous pouvons noter malgré tout la chute de l'indicateur de recherche d'itinéraires en transports publics, beaucoup plus nette en France que dans les autres pays. La fréquentation des commerces de détail hors alimentaire et des lieux de loisirs a également fortement reculé selon l'indicateur tiré de *Google Mobility Report*. Dans ces domaines, la baisse d'activité serait actuellement plus marquée en France que dans les pays principaux européens voisins. Autre indicateur, celui de *Google Mobility* sur le temps passé chez soi peut certes refléter en partie l'essor du télétravail, au-delà du confinement des populations, mais il paraît en tout cas très corrélé avec les pertes d'activité économique estimées au mois le mois.

Toutes ces données, conjuguées aux autres informations dont nous disposons, nous ont permis tout au long de la crise de proposer des estimations « instantanées » (« nowcasting ») de l'activité économique. Pour le mois de novembre, nous estimons actuellement à 13 % la baisse par rapport à la situation d'avant-crise et nous avons établi différents scénarios pour le mois de décembre.

Par ailleurs, l'utilisation, rendue possible dès la mi-mars, des données agrégées de transactions par carte bancaire nous a offert la possibilité de réaliser un diagnostic assez fin de la consommation des ménages. Ces données montrent que la baisse de la consommation a été, en partie seulement, rattrapée à l'issue du premier confinement. L'équipement du foyer a ainsi fait l'objet de reports d'achat. À l'inverse, dans la restauration, les pertes sont sèches et se révèlent beaucoup plus importantes. Des différences apparaissent aussi entre le premier et le deuxième confinement, au cours duquel la chute s'est avérée moins rude. Les consommateurs ont adopté des comportements assez différents. La vente en ligne s'est ainsi envolée dans les jours qui ont suivi ce deuxième confinement alors que, du fait de la sidération, ces achats étaient plutôt en berne au début du premier. Nous observons également une adaptation dans la restauration, où la chute est moins rude (avec l'essor de la livraison à domicile ou du « *click & collect* »), même si le secteur reste très fortement touché. De la même manière, nous notons des comportements d'achats de précaution moins marqués pour ce deuxième confinement. Ces achats de précaution sont d'ailleurs d'une nature quelque peu différente : juste avant ce deuxième confinement, les données font apparaître des achats d'anticipation dans l'alimentation, mais dans des proportions largement moindres que les jours qui ont précédé le 17 mars. Inversement, les achats d'anticipation ont été beaucoup plus importants fin octobre pour l'équipement du foyer alors que ces dépenses avaient chuté en mars, quand la priorité avait été donnée à l'alimentation.

Mais les données « massives » ne sont pas uniquement issues des opérateurs privés. Il peut s'agir aussi de données administratives disponibles sur un pas mensuel comme la déclaration sociale nominative (DSN), le formulaire que toutes les entreprises remplissent auprès d'un certain nombre d'organismes. Ces données permettent d'expliquer dans quelle mesure l'activité partielle a joué le rôle d'amortisseur pour l'emploi. L'étude parue hier dans le point de conjoncture présente un indicateur de volume des heures rémunérées et fait la décomposition de la baisse. 70 % s'expliquent par l'essor exceptionnel de l'activité partielle. La DSN offre aussi la possibilité de rendre compte de l'hétérogénéité. Ces données peuvent en effet se décliner selon la profession, l'âge ou le type de contrat de travail. Elles peuvent également être déclinées par région ou département. L'évolution du volume de travail rémunéré dans l'hébergement-restauration montre par exemple que les grandes métropoles, notamment parisiennes, sont plus fortement affectées que les régions de l'ouest ou du sud de la France en septembre. Toujours en matière de déclinaison territoriale, l'utilisation des données de téléphonie mobile a permis de rendre compte des déplacements de la population.

Si certains outils nous ont aidés à parer à l'urgence, d'autres présentent sans doute un potentiel plus pérenne. Certaines données se sont surtout révélées utiles pour donner un ordre de grandeur dès lors que le choc était soudain et de très grande ampleur, comme en mars ou en novembre. Quand les variations deviennent plus faibles, ces outils s'avèrent en revanche moins informatifs. Nous nous interrogeons aussi sur la représentativité de certaines données et sur la documentation relative au processus qui les génère. Nous retirerons certainement beaucoup de cette expérience : nous espérons par exemple continuer à utiliser les données sur les transactions par carte bancaire, pour aider à suivre la consommation des ménages au niveau macro mais aussi en en rendant compte de l'hétérogénéité.

Anne-Laure DELATTE

En dehors des transactions par carte bancaire, vous n'avez finalement pas accès à des données massives financières. Il nous a néanmoins semblé intéressant de revenir sur cette nouvelle façon de procéder, même si vous avez souligné que cette démarche n'était pas totalement nouvelle.

Julien POUGET

Nous commençons à travailler avec des données bancaires, comme l'a fait le Conseil d'analyse économique (CAE), dont les dernières publications reprenaient des données du Crédit Mutuel. Nous envisageons de coupler les données du Crédit Mutuel avec l'étude des transactions par carte bancaire afin d'avoir une vision plus fine des impacts financiers de la crise sur les ménages.

Anne-Laure DELATTE

Cette étude a permis de mettre en évidence des inégalités assez fortes durant la crise sur l'accumulation de l'épargne qui s'est concentrée sur les 20 % des ménages les plus riches.

Julien POUGET

Ces résultats ont effectivement été publiés par le CAE voilà un mois.

Kamel GADOUCHE, CASD

Avez-vous connaissance de travaux menés dans les universités, en France ou à l'étranger, sur la même thématique de l'impact des crises ?

Julien POUGET

L'Insee a été l'un des tout premiers instituts de statistique à quantifier la chute d'activité liée à la crise, mais les travaux sont maintenant extrêmement nombreux. Je pense notamment aux travaux de Raj Chetty aux Etats-Unis, qui visent justement à rassembler un maximum de données à haute fréquence disponibles en temps réel rendant compte de l'hétérogénéité des ménages, des territoires.

Anne-Laure DELATTE

Raj Chetty travaille avec une équipe dénommée « The Opportunity Insights Team » qui a produit un document de travail en juin 2020 basé sur des données massives, notamment des données issues de plates-formes d'éducation.

David ADAM, Banque de France

Je souhaitais confirmer tout l'intérêt de suivre les travaux américains qui visent à précalculer les données de comptabilité nationale américaine à partir de ces données de consommation des ménages et de toutes les informations qu'ils peuvent collecter. De nombreuses données concernant les grandes entreprises américaines ne sont pas accessibles unitairement, par entreprise, mais les entreprises américaines acceptent de contribuer à des pools d'informations qui permettent d'anonymiser les données en vue de réaliser ces prévisions. Cette démarche, qui n'est pas encore menée en Europe, me semble intéressante.

III. ACCÈS AUX DONNÉES GRANULAIRES FINANCIÈRES : ÉVOLUTIONS RÉCENTES

Anne-Laure DELATTE

Je tiens à remercier les équipes du secrétariat du Cnis qui ont coordonné l'ensemble de ces présentations. Après l'introduction de David Adam, de la Banque de France, sur les enjeux des données granulaires, nous aurons une série de présentations autour de deux enjeux : d'une part, l'accès aux données individuelles et leur appariement pour élargir le champ de travail et d'autre part, la centralisation dans une même base de données de différents pays pour étendre le champ géographique.

Introduction : les enjeux de la collecte granulaire de données financières et bancaires

David ADAM, Banque de France

Globalement, nous avons pour habitude de travailler avec des données sur le crédit, les marchés, les dérivés dans des bases de données très riches au cours des dix dernières années. Nous nous apercevons aujourd'hui que c'est dans la mise en commun de ces données à l'échelle la plus granulaire que nous obtenons les meilleurs résultats et les meilleures informations. Je vous renvoie notamment aux travaux au Haut Conseil de stabilité financière (HCSF) qui a réalisé une simulation sur l'ensemble des 8 000 acteurs financiers français, à la fois les compagnies d'assurance, les banques et les fonds d'investissement, pour tenter d'observer les effets de domino qui peuvent exister entre chacune de ces entités prises une à une. Ce travail, présenté à l'échelle française et du Fonds monétaire international, a suscité un grand intérêt. Lorsque nous sommes capables de suivre les effets de domino entité par entité, nous avons une meilleure appréhension des crises financières et des crises économiques et des liens entre les deux. Par la capacité à obtenir des données individuelles, nous pouvons mettre en place des modèles assez riches sur les mécanismes et les canaux de propagation des crises.

En Europe, nous avons la chance de disposer d'un cadre à peu près intégré. La base des données de crédit va être constituée au niveau européen dans le cadre du projet AnaCredit que nous évoquerons tout à l'heure. La France compte aussi, parmi ses banques, des acteurs majeurs des instruments dérivés et nous avons la possibilité aujourd'hui de suivre ce qu'il advient en cas de faillite ou de dégradation de la note d'un grand acteur sur la base des indicateurs de marché dérivé. Enfin, le marché des titres est très significatif. Les crises de liquidités qui peuvent apparaître sur certains fonds ou certains compartiments de titres méritent tout l'intérêt des superviseurs et de tous ceux qui suivent la stabilité financière. Toutefois, nous éprouvons des difficultés pour identifier des règles générales. Il apparaît absolument nécessaire de descendre au niveau de chaque individu, de chaque structure juridique.

Ce va-et-vient entre l'analyse macroéconomique et le besoin de données microéconomiques existe un peu sur tous les compartiments et je crois que la décennie à venir permettra de mettre en commun les différentes informations. L'Eurosystème s'est quand même organisé de manière à ouvrir de plus en plus les données granulaires qu'il met à la disposition de tous les chercheurs. Le Comité européen du risque systémique fait de même. Au niveau international, le *Financial Stability Board* vient encore de produire des analyses où il appelle à l'utilisation des données granulaires. Nous nous inscrivons donc clairement dans un mouvement extrêmement fort. Il nous revient d'y consacrer les moyens et de ne pas nous montrer timides. Les chercheurs français doivent directement demander l'information lorsqu'ils n'en disposent pas. Il existe des procédures pour le faire. Nous ne faisons pas partie des juridictions les plus actives pour l'instant. La Banque de France souhaite jouer le rôle de courroie de transmission pour permettre à un grand nombre d'analyses d'exister.

Je vous renvoie au site du Haut Conseil de stabilité financière pour voir tout ce qui peut être fait aujourd'hui avec les données granulaires. Ces analyses exigent beaucoup de travail et un appariement manuel de bases de données qui ne sont pas conçues pour l'instant pour être connectées les unes avec les autres, mais qui le seront demain. Nous allons donc vers la bonne direction.

Anne-Laure DELATTE

Vous avez mis l'accent sur l'un des grands intérêts de ces données granulaires en matière de régulation et de stabilité financière. Pour les régulateurs, cet intérêt apparaît évident. Pour les académiques, ces données permettront de bien identifier les mécanismes sous-jacents aux relations économiques.

III.1. Les données financières granulaires françaises

III.1.1 Innovations dans l'accès aux données de la Banque de France via l'Open Data Room

Vincent DUPRAT, Banque de France

Les données se révèlent importantes en matière de stabilité financière. Notre système permet de mettre à disposition les données à la fois en interne et en externe, et nous avons été extrêmement sollicités en interne depuis le mois de mars pour mettre à disposition un certain nombre de données. Aujourd'hui, nous nous concentrerons cependant sur notre dispositif de mise à disposition externe, l'Open Data Room.

L'Open Data Room dite physique date de 2016. Elle consistait en une mise à disposition de données dans une salle du 37 rue du Louvre, ce qui présentait un certain nombre d'avantages, mais aussi un certain nombre de limites. Elle représentait quand même un premier pas vers une ouverture des données de la Banque de France. Elle avait fait l'objet d'une inauguration avec le gouverneur et le directeur général des statistiques de l'époque. La Banque de France avait alors la volonté de démontrer qu'il s'agissait d'une démarche d'ouverture et de modernisation. Avec quatre ans de recul, cette démarche peut apparaître comme une évidence. Néanmoins, à l'origine, elle s'est révélée assez novatrice bien que l'accès reste relatif, puisque circonscrit aux locaux de la Banque de France par opposition à la démarche menée par le CASD depuis un certain temps déjà.

L'Open Data Room met à la disposition des chercheurs externes à la Banque de France des données de la Banque de France, mais il ne s'agit que de la partie émergée de l'iceberg. Nous possédons par ailleurs un dispositif interne plus conséquent à la fois en termes techniques et en termes de données, Muses.

Muses représente la première application sur une infrastructure *big data* de la Banque de France. Elle met à disposition un ensemble des données de la Banque de France de manière très transverse, y compris des données de l'ACPR. À l'inverse de l'Open Data Room qui met des données à disposition en fonction de la demande des chercheurs, elle propose des données. Enfin, elle mobilise des technologies aujourd'hui assez classiques, mais qui, à l'époque, étaient relativement innovantes : du stockage de données sur des infrastructures *big data HADOOP*, des moteurs de recherche *Elastic Search*. Muses constitue l'antichambre de l'Open Data Room, puisque toutes les données mises à disposition des chercheurs dans cette dernière doivent au préalable être présentes dans notre outil de partage interne et sont transférées de celui-ci vers l'outil de partage externe en passant au travers d'un algorithme qui pseudonymise les données. Toutes les données présentes sur l'Open Data Room sont en effet pseudonymisées. Les données déjà présentes dans Muses affichent un niveau de disponibilité plus rapide que des données pour lesquelles nous sommes obligés de solliciter les services producteurs, de leur demander de préparer des jeux de données adaptés à leur intégration dans Muses et un export vers l'Open Data Room.

Pour mettre à disposition des données aux utilisateurs, au-delà de l'aspect opérationnel, nous sommes confrontés à une problématique de gouvernance des données. Les données posent-elles des problèmes de confidentialité ? Sous quel régime réglementaire peuvent-elles être mises à disposition ? Nous avons mis en place une double gouvernance pour l'accès interne et l'accès externe, puisque des impératifs différents s'imposent dans les décisions d'accès.

Delphine MOREAU, Banque de France

Je vais vous présenter l'articulation pour que vous compreniez, dans le traitement d'une demande d'accès aux données, les interlocuteurs auxquels vous aurez à faire face. En interne, la solution d'accès à ces données est totalement imbriquée à la solution externe. Les équipes de Vincent Duprat seront vos interlocuteurs pour répondre à des questions plutôt techniques sur cet accès aux données. Il gère également le comité d'accès aux données interne, puisque des économistes, chercheurs et statisticiens de la Banque de France qui souhaiteraient accéder à des données présentes dans Muses doivent aussi en faire la demande pour voir leur accès autorisé.

Ce comité d'accès aux données sera pris en charge par la direction générale des statistiques, des études et de l'international (DGSEI) à compter de 2021 afin d'améliorer autant que possible le traitement des demandes, assister davantage les chercheurs pour déterminer les données qui peuvent répondre au mieux à leurs besoins d'étude et de recherche. Nous aurons alors une articulation entre un secrétariat au comité d'accès aux données géré de façon centralisée au sein de la DGSEI, les services producteurs responsables

d'analyser la capacité à partager à l'extérieur les données qu'ils produisent avec l'appui de la direction juridique de la Banque de France.

Dans l'évolution que nous mettons actuellement en production, nous allons nous aligner sur le même type d'accès aux données que le CASD, conformément à la recommandation du rapport élaboré par Gunther Capelle-Blancard dans le cadre du Cnis, consistant à permettre un accès virtualisé aux données bancaires et financières. Cette offre de service simplifiera grandement le travail du chercheur qui n'aura plus besoin de se déplacer dans les locaux de la Banque de France. Ainsi, les chercheurs hors Île-de-France pourront plus facilement faire des demandes d'accès aux informations de la Banque de France avec une telle solution.

Nous aurons également la capacité d'accueillir des données que le chercheur souhaiterait croiser avec les données de la Banque de France. Une réflexion est en cours sur le pseudonymisation des données et des modalités de facturation de cet accès aux données. Nous envisageons de revisiter chacun des cadres juridiques par catégorie de données pour déterminer si nous pourrions, comme le CASD, lever l'anonymisation pour un certain nombre d'informations. Nous devrions par ailleurs proposer un niveau de facturation assez proche de celui que vous pouvez connaître dans l'accès au CASD. Nous devrions vous redonner des précisions sur ces aspects d'ici début 2021. Nous en ferons également la communication aux équipes de recherche intéressées.

Dans ce projet, nous avons mis en œuvre des objectifs forts de la Banque de France, puisque cette démarche s'inscrit pleinement dans le plan stratégique de notre entreprise. L'action dédiée permet de vous offrir aujourd'hui cette Open Data Room virtuelle et nous menons une collaboration qui fera l'objet de la prochaine présentation avec le CASD afin de faciliter la mise à disposition de données présentes dans chacune des bases.

Vincent DUPRAT

Nous devons respecter des impératifs de sécurité des données assez forts, que ce soit l'impossibilité de sortir des données ou la vérification de la nature des données que les chercheurs souhaitent sortir. Nous pouvons maîtriser certains aspects du point de vue technologique. D'autres nécessitent une connaissance des données et font plutôt l'objet d'un processus organisationnel de vérification du contenu des données. Ce processus vaut à la fois pour les données qu'il faut intégrer dans l'Open Data Room et pour les données qu'il s'agit d'en faire sortir.

Le chercheur pourra se connecter avec son propre poste. Il aura simplement besoin d'installer un logiciel sur son PC pour interagir avec l'Open Data Room et un logiciel sur son smartphone pour générer automatiquement des codes de connexion permettant de se connecter à l'Open Data Room. Ces deux logiciels pourront être téléchargés sur des appareils banalisés. Une fois ces barrières techniques de sécurité informatique franchies, le chercheur pourra avoir accès aux outils statistiques installés sur son environnement, qu'il s'agisse d'outils gratuits ou payants. Avec la DGSEI, nous travaillons actuellement à des modèles de facturation de ces logiciels payants.

Les travaux seront stockés sur ces machines virtuelles à distance. Nous travaillons aussi sur une offre de mise en veille en cas d'interruption des travaux de recherche pour une remise à disposition selon des conditions identiques. Il existe des environnements dédiés par chercheur, avec la possibilité d'échanger des données entre des chercheurs d'une même équipe par l'intermédiaire de répertoires d'échange partagés. L'infrastructure technique qui héberge cette Open Data Room est totalement étanche au système d'information de la Banque de France. Le seul lien qui existe entre les deux réside dans une passerelle technique qui permet de faire passer les données de Muses vers l'Open Data Room virtuelle. Du point de vue du réseau, les deux systèmes sont totalement étrangers. De fait, les deux réseaux ne peuvent pas collaborer *via* un simple mail par exemple. Toutes les communications entre les chercheurs et l'équipe de support mise en place, que ce soit pour le support fonctionnel sur les données ou pour le support technique, s'effectueront par le biais d'une messagerie externe. Les chercheurs ne seront pas intégrés au réseau de la Banque de France. De la même manière, les mises à disposition des travaux de recherche se feront *via* une solution de transfert de fichiers. Le contenu devra aussi faire l'objet d'un examen pour vérifier qu'il est bien éligible à une sortie de l'Open Data Room.

III.1.2 Appariement des données : le projet de collaboration Banque de France – CASD

Kamel GADOUCHE, CASD

Une collaboration est engagée depuis quelques années maintenant entre le Centre d'accès sécurisé aux données (CASD) et la Banque de France. Le CASD est, depuis 2019, un groupement d'intérêt public à but non lucratif qui regroupe l'État représenté par l'Insee, le groupe des écoles nationales d'économie et de statistique (Genes), le CNRS, l'École polytechnique et HEC Paris.

Aujourd'hui, 25 personnes travaillent pour ce service de mise à disposition sécurisée de données à des fins non lucratives de recherche, d'étude, d'évaluation et d'innovation, mission principale inscrite dans les statuts de ce groupement d'intérêt public. Nous jouons le rôle de tiers de confiance dans cette mise à disposition sécurisée. Les données nous sont confiées contractuellement pour que nous les mettions à disposition. En aucun cas, nous n'avons la propriété de ces données. Nous n'avons aucune prise non plus sur les personnes qui peuvent accéder aux données, la durée ou le cadre de cet accès.

Le CASD a intégré dès sa conception un niveau de sécurité élevé pour la mise à disposition sécurisée des données. Aujourd'hui, le centre est certifié sur les standards internationaux de sécurité de l'information, notamment ISO 27001, et plus récemment ISO 27701 qui prend en compte la protection des données à caractère personnel et les exigences du RGPD. Nous sommes également certifiés en tant qu'hébergeur de données de santé et nous avons obtenu les autorisations de la Cnil pour les traitements sur les données à caractère personnel, notamment les accès biométriques. Enfin, notre certification auprès de l'Anssi (Agence nationale de la sécurité des systèmes d'information) est en cours. Régulièrement, nous faisons l'objet d'audits techniques réalisés par des « ethical hackers » qui essaient de pénétrer dans le système ou de récupérer les données selon différents scénarios. Ces tests se révèlent essentiels pour garantir un niveau de sécurité élevé.

En termes de gouvernance, le CASD comprend un conseil scientifique présidé par Lars Vilhuber, de Cornell University, un comité des producteurs qui s'est réuni pour la première fois la semaine dernière avec l'ensemble des producteurs ayant signé au moins une convention avec le CASD, y compris la Banque de France, et un comité de suivi de la sécurité de l'information qui veille à l'application de toutes les mesures et à l'amélioration continue de la sécurité au sein du CASD. Nous mettons à disposition plus de 300 sources de données, représentant 2 000 fichiers, soit un volume de près de 500 To. Nous recensons 500 institutions utilisatrices en France, en Europe, aux États-Unis et au Canada et plus de 1 500 chercheurs utilisant le CASD.

Grâce à la structuration du CASD, les chercheurs qui accèdent aux données par ce biais ne sont pas limités par le nombre de sources qui sont mises à disposition et peuvent accéder à des données individuelles provenant de plusieurs producteurs et donc de plusieurs domaines. Ils peuvent ainsi combiner des données de santé avec des données économiques, des données du ministère de la Justice avec celles du ministère de l'Éducation nationale, etc.

Vincent DUPRAT

L'Open Data Room (ODR) physique de la Banque de France présente un certain nombre de différences structurelles avec le CASD que nous essayons de traiter les unes avec les autres pour arriver à collaborer effectivement. Les deux systèmes affichent trois différences majeures : le CASD propose un accès distant, payant et sans hachage de l'identifiant alors que l'ODR ne permet qu'un accès sur site, de manière gratuite, à des données pseudonymisées. Dans ces conditions, nous ne parvenons pas pour l'instant à répondre à la demande de chercheurs qui ont besoin à la fois de données du CASD et de données de la Banque de France, ce qui constitue un vrai problème. Certains accèdent aux données d'un côté et de l'autre et réalisent des analyses sur chacun des périmètres sans possibilité de croisement. D'autres s'autocensurent, se limitant aux données du CASD faute de pouvoir apparier les données de la Banque de France. Nous avons pleinement conscience de ce besoin de travailler de manière transversale sur les données.

Cette année, nous avons reçu la demande de l'Inspection générale des finances qui est assimilée, au sein de la Banque de France, au ministère des Finances, ce qui lève un certain nombre de barrières juridiques à la fois sur la transmission des données et sur la nécessité de pseudonymiser les données. Dans ce cadre extrêmement favorable, nous avons pu échanger des données avec le CASD, éprouvant ainsi les

mécanismes d'échange. Nous avons pu constater la faisabilité technique de l'export de nos données au sein du CASD. Nous nous heurtons plutôt à un blocage juridique.

Depuis le mois de mars, nous avons par ailleurs imaginé des pilotes avec des chercheurs ayant des besoins transversaux à nos deux entités, basés sur des échanges croisés de données. Nous avons envisagé de transmettre des données anonymisées de la Banque de France au CASD, qui aurait dû également anonymiser le périmètre de données que le chercheur souhaitait apparier avec nos données. Dès qu'une demande nous sera présentée, nous procéderons de cette manière. Même si ce processus est lourd, il reste conforme aux contraintes juridiques qui nous imposent d'anonymiser nos données.

Kamel GADOUCHE

Une convention plus générale est également à l'étude au niveau de la Direction juridique de la Banque de France pour couvrir ces demandes de chercheurs.

Vincent DUPRAT

Chaque source de données est un cas juridique spécifique. Les données EFI, ECO et ECEIS sont régies par le comité du secret. Les sortir ne présente pas la plus grande complexité sur le plan juridique. Le dossier est aujourd'hui à l'étude par la direction juridique en vue de signer une convention d'échange sur ce périmètre, ce qui devrait répondre à une partie des demandes.

Les données des « *significant institutions* » relèvent de la supervision européenne et présentent des contraintes de partage au niveau de la BCE. Pour partager ces données, nous devons donc demander l'avis de la BCE, y compris en interne à la Banque de France, contrairement aux données bancaires sur lesquelles nous conservons la maîtrise. Chaque jeu de données dépend potentiellement d'un corpus de données qui lui est propre et doit faire l'objet d'une instruction spécifique. La direction juridique est donc systématiquement impliquée, que ce soit dans les décisions de partage interne ou dans les décisions de partage externe.

Pour l'ODR physique, nous avons adopté une position conservatrice et un peu protectionniste, en nous alignant sur le régime le plus contraignant et en pseudonymisant tout par défaut pour des raisons de simplification et de protection. Cette démarche constitue cependant un vrai frein à l'intérêt des données proposées et aux possibilités de comparaison et d'appariement des données. Nous remettons donc en question ce dogme. Certaines données nécessiteront une anonymisation, mais nous espérons que le cercle de ces données sera relativement restreint et que nous pourrions proposer des données non anonymisées sur un périmètre large afin de faciliter la collaboration avec le CASD et ajouter de la valeur pour les chercheurs.

Les difficultés juridiques semblent les plus élevées dans ce projet de collaboration. Nous souhaitons donc tout d'abord tester des échanges croisés. Nous n'avons pas encore envisagé de projet de plate-forme d'interconnexion de nos systèmes. La question se posera à un moment donné, mais il faudra d'abord lever certaines barrières. Nous n'avons pas non plus évoqué la tarification, mais si les données sont gratuites, le service est payant et nous devons nous mettre d'accord sur des protocoles et des partages de coûts associés entre institutions.

Kamel GADOUCHE

Le CASD n'est pas propriétaire des données. Pour tout transfert à l'ODR, la décision reviendrait donc au propriétaire. Par ailleurs, il faut rappeler que les données fiscales, le régime fiscal sont uniquement disponibles *via* le CASD à la demande de la DGFIP. L'ODR 2.0 présentera une plus grande convergence avec le CASD, ce qui ouvre des perspectives de collaboration plus étroite entre nos deux institutions.

III.2. Utilisation des données financières granulaires en Europe

Introduction : les données individuelles disponibles au niveau européen

Roxane SILBERMAN, conseillère scientifique au CASD

Je ne suis pas une spécialiste des données bancaires, mais je m'intéresse depuis très longtemps aux problèmes d'accès aux données pour les chercheurs, notamment au niveau transnational.

Deux projets vous seront présentés aujourd'hui dans cette session que l'on m'a demandé d'introduire : Inexda, un réseau de banques centrales et AnaCredit qui propose l'intégration de données en amont. Cette introduction vise à vous donner rapidement une vue d'ensemble sur l'évolution générale de la mise à disposition de données confidentielles pour la recherche, y compris les données dites granulaires des banques, ainsi que sur les problèmes posés pour l'utilisation des données de plusieurs pays. Je vous donnerai sur ce point un exemple d'initiative dans le domaine des données de la statistique publique et des données administratives qui peut être intéressant pour contribuer à la discussion.

Depuis vingt ans, on peut constater une ouverture massive à la recherche des données confidentielles des instituts nationaux de statistiques, enquêtes mais aussi données administratives (pour ces dernières, depuis longtemps dans les pays dits à registres). On observe aussi une évolution très importante dans les modes d'accès, qui ont d'abord été sur site puis de plus en plus à distance. Les principes d'accès présentent de très fortes similitudes, avec des contraintes de sécurité modulées en fonction du niveau de risque. Néanmoins, de nombreuses différences persistent dans les modalités d'accès et les procédures, ce qui constitue un problème dès lors que l'on veut rapprocher les données au niveau national et transnational. Il existe ainsi plusieurs modes d'accès : *remote access* comme pour le CASD, *remote execution* ou *job submission*, où le chercheur n'a pas directement accès aux données.

Cette ouverture a concerné plus récemment les données financières et bancaires. Les accès sur site dans ce domaine sont majoritaires, même si quelques projets émergent pour passer à un accès à distance, comme à la Banque de France. Nous voyons aussi émerger des coopérations pour permettre de combiner les données de la statistique publique et les données des banques centrales. Aux Pays-Bas par exemple, la banque centrale néerlandaise (DND) a engagé une coopération très forte avec CBS, l'institut national statistique. Les données sont mises à disposition par le centre d'accès sécurisé du CBS ce qui permet d'apparier ou de combiner les données pour la recherche par cet intermédiaire.

L'accès transnational est aussi une question cruciale pour les approches comparatives entre plusieurs pays comme pour accompagner la mobilité croissante des chercheurs et les collaborations entre les équipes de différents pays. Les aspects juridiques de l'accès transnational aux micro-données de la statistique publique ont fait l'objet de nombreuses discussions au sein de l'OCDE et d'Eurostat. Ces discussions ont montré qu'en général, la loi n'interdit pas formellement le transfert de données d'un pays à l'autre. Elle est le plus souvent silencieuse. Comme au niveau national, il importe cependant d'assurer la sécurité. La question se pose dès lors des équivalences, notamment en ce qui concerne les sanctions en cas de rupture de confidentialité. Il faut également mettre en place des clauses contractuelles pour permettre le transfert entre les pays. L'idée générale avancée dans ces discussions à l'OCDE est celle d'un cercle de confiance. Dans l'Union européenne, le cadre mis en place pour les données personnelles avec le RGPD conforte désormais la circulation au sein de l'Union et même en dehors de l'Union quand des équivalences existent.

Dans les faits, on observe une croissance des accès transnationaux pour la recherche aux données confidentielles de la statistique publique, que le chercheur ait encore à se déplacer dans le pays ou, de plus en plus, qu'il puisse utiliser un accès distant comme au Danemark, aux Pays-Bas ou en France. Pour l'instant, cette démarche reste limitée pour les banques centrales, sauf lorsqu'il s'agit de données agrégées en amont.

Les difficultés demeurent néanmoins encore importantes dans tous les domaines pour l'utilisation de données de plusieurs pays, ce qui est nécessaire pour nombre de travaux de recherche, si ces données ne sont pas agrégées et harmonisées en amont comme ce peut être le cas pour les microdonnées européennes pour l'instant accessibles uniquement sur site à Luxembourg. Pour un chercheur qui a besoin d'utiliser les données de plusieurs pays, il peut être ainsi nécessaire de se rendre successivement dans chaque pays. Il a également affaire à des équipements d'accès et des procédures qui diffèrent d'un pays à

l'autre, et il se heurte ainsi à des silos qui ne permettent pas le traitement conjoint des données dans un même environnement de recherche.

Pour faciliter l'utilisation pour la recherche de données confidentielles de plusieurs pays, le réseau IDAN (*International Data Access Network*) a été créé en 2018 avec six centres sécurisés de quatre pays : la France, l'Allemagne, le Royaume-Uni et les Pays-Bas. Le CASD en est le coordinateur. Ce réseau prend appui sur le travail mené précédemment dans un projet européen *Data without Boundaries* (DwB) qui avait impliqué 14 pays avec 29 partenaires dont les instituts statistiques, et dans un autre projet NORDMAN mené dans les pays nordiques sur le transfert de données entre instituts statistiques pour les chercheurs qui avaient besoin de combiner les données.

L'objectif final du réseau IDAN est de permettre à un chercheur d'accéder d'un seul point aux données des six centres partenaires et de travailler dans un même environnement de recherche. Idéalement, un chercheur souhaiterait accéder de son université à toutes les données de la même façon, selon les mêmes procédures et dans le même environnement de recherche. C'est un objectif difficile à atteindre d'emblée, les six centres partenaires présentant de grandes différences en la matière. Pour avancer plus rapidement, ce réseau a adopté une stratégie pragmatique centrée sur l'accès. L'accréditation, la formation et le contrôle des résultats restent de la responsabilité des partenaires. Les équipements d'accès de chacun des centres ont été installés chez tous les partenaires. Depuis la *Safe Room* de l'un des partenaires, le chercheur peut ainsi plus facilement travailler sur les données de chacun des partenaires, avec tous les équipements d'accès présents dans une seule salle.

Cette première étape a déjà nécessité des évolutions et des harmonisations entre partenaires. Ainsi, l'IAB, le centre du ministère fédéral de l'Emploi en Allemagne a dû accepter de proposer des données d'un même niveau de détail que celles accessibles sur site à Nuremberg. L'institut national statistique au Royaume-Uni, l'ONS, devra ouvrir à l'international l'accès à distance qu'il est en train de mettre en place à l'intérieur du pays. Enfin, pour le CASD en France et CBS aux Pays-Bas, qui proposaient déjà l'accès à distance, il a fallu créer une *Safe Room* pour héberger les équipements d'accès des partenaires. Il a également été nécessaire d'harmoniser et d'adopter des accords sur les réquisits de sécurité pour chaque *Safe Room*. Sauf au Royaume-Uni pour l'instant, les points d'accès sont désormais implantés dans chaque pays partenaire et des chercheurs peuvent d'ores et déjà travailler sur les données.

Le site web IDAN propose par ailleurs une vue d'ensemble rapide sur les procédures d'accréditation, les procédures d'accès, les catalogues avec les liens sur les sites des partenaires. Un travail a également commencé sur la comparabilité des données, notamment entre l'Allemagne et la France. Les chercheurs sont au centre de cette démarche. Nous espérons que leurs projets seront moteurs pour l'établissement d'accords de transfert de données lorsque nécessaire pour leur traitement conjoint dans un même environnement de recherche. Les premiers accords entre des partenaires qui ont déjà convenu des conditions de sécurité et qui se connaissent bien sont une étape pour aboutir à terme à un modèle plus intégré permettant un accès aux données des différents pays concernés.

III.2.1 Le projet Inexda (*International Network for Exchanging Experience on Statistical Handling of Granular Data*)

Vincent DUPRAT

Inexda est une organisation internationale d'échange entre des banques centrales, Eurostat et des instituts nationaux de statistique sur les bonnes pratiques de manipulation de données granulaires à des fins de recherche. Il représente un réseau de bonnes pratiques et non un réseau d'échange de données. Contrairement aux initiatives précédemment présentées, l'échange de données ne constitue pas son objectif, du moins à ce stade.

Dans le mandat d'Inexda, sa mission se limite aux fins de recherche. Dans les faits, nous abordons aussi les bonnes pratiques de partage interne et nous débattons sur les notions de chercheur ou d'analyste interne/externe, car les procédures d'accès diffèrent selon le statut. Inexda n'est donc ni un réseau de partage de données entre les membres ni une offre commune d'accès aux données pour les chercheurs. Chaque pays reste responsable de son offre de données pour la recherche. Inexda n'a pas pour volonté de proposer une offre groupée en rassemblant toutes les données offertes par les pays qui le constituent. En revanche, nous souhaitons mettre en visibilité, au travers d'Inexda, le fait que des membres du réseau offrent des services de données pour la recherche. Chaque membre doit ainsi faire la publicité des autres.

Inexda a été fondé en décembre 2016 par les banques centrales nationales d'Italie, du Portugal, du Royaume-Uni, de la France et de l'Allemagne. Depuis, le réseau a été rejoint par les banques centrales nationales d'Espagne, du Chili, de la Russie, du Mexique, ainsi que par l'institut statistique britannique l'ONS, la BCE et Eurostat. La présidence a été assurée par la Bundesbank entre 2017 et 2018, puis la Banque de France, dont le mandat s'achève fin 2020. Un appel à candidature a été lancé et la prochaine présidence devrait être annoncée lors de la réunion plénière du 11 décembre prochain. L'institution présidente est également chargée du secrétariat d'Inexda durant son mandat, ce qui représente une charge non négligeable.

Des groupes de travail thématiques ont été constitués sur un certain nombre de sujets. La base de nos échanges concerne le processus de partage des données. Nous abordons le processus d'approbation pour l'accès aux données, qui se traduit en général par la signature d'un contrat ou au moins d'un certain nombre de documents (code de bonne conduite, engagement de confidentialité, etc.), le processus de mise à disposition des données (sécurité, anonymisation par des outils développés par les institutions ou des outils de marché). Nous nous interrogeons aussi sur les outils statistiques les plus couramment utilisés. Nous échangeons sur la manière de contrôler les outputs d'un point de vue technique comme d'un point de vue fonctionnel afin de s'assurer que les données qui sortent de la *data room* peuvent effectivement en sortir. Nous nous intéressons également aux processus à mettre en place pour suivre les publications. Enfin, nous souhaiterions connaître l'impact de l'ouverture de ces données sur la recherche en général. Sur ce sujet, cependant, nos travaux sont encore à un stade exploratoire.

La Banque de France a par ailleurs piloté un groupe de travail sur le site web d'Inexda. Nous avons également rassemblé les offres de données de tous les membres. Or nous avons constaté qu'un concept, même traduit en anglais, peut être appelé différemment d'un pays à l'autre, ce qui soulève un enjeu d'harmonisation des terminologies pour assurer la visibilité du catalogue de l'offre. Nous travaillons aussi sur des modes d'accès aux données, les processus d'autorisation en fonction de la nature des données, de leur usage ou de la qualité des usagers, avec la volonté de développer l'automatisation de ces procédures pour les simplifier et en renforcer l'efficacité. L'anonymisation figure aussi au centre de l'un des groupes de travail, de même que le contrôle des sorties. Eurostat, par exemple, a développé un outil qui automatise dans une certaine mesure ce contrôle des sorties et doit nous le présenter le 11 décembre. Nous espérons pouvoir capitaliser sur cette initiative et élaborer un dispositif un peu similaire en France et dans les autres institutions membres.

III.2.2 Le projet AnaCredit de l'Union européenne

Jean-Marc ISRAËL, Banque de France

Le projet AnaCredit a vu le jour dans les années 2010, dans la vague de la première crise. Il est né de l'idée que les données sur le crédit sont nombreuses, et certaines étaient granulaires dans les registres de crédit nationaux, mais qu'elles existaient dans des formats différents et donc difficiles à comparer.

Les *reportings* à la Commission européenne ou toute autre institution donnaient des résultats différents selon qu'ils étaient basés sur le *reporting* prudentiel, le *reporting* statistique, des enquêtes, etc. Les analyses ne différaient pas sur le fond ; elles étaient seulement basées sur des données différentes. Nous avons alors proposé de mettre en commun les informations des registres de crédit, qui étaient jusqu'alors totalement cloisonnées et fragmentaires. Nous avons pris contact avec les opérateurs des registres de crédit dans les différentes banques centrales. Ces opérateurs étaient confrontés aux mêmes problématiques : ils ne parvenaient en effet à échanger les données que sur quatre attributs par prêt, compte tenu des différences.

À l'origine, la banque centrale était chargée de décider de la politique monétaire et de sa mise en œuvre en intervenant sur les marchés. Jusqu'à la crise de 2007, nous étions sur une tendance lourde. Avec la première crise et, un an plus tard, la faillite de Lehmann, les prévisions sont devenues de plus en plus difficiles. Le conseil des gouverneurs était gêné par l'absence de données rapides et fiables sur les marchés. Les marchés donnaient des signaux difficiles à interpréter, souvent contradictoires et perturbants. Pour la conduite de la politique monétaire, la demande a alors émergé de renforcer la connaissance des marchés monétaires et du crédit, d'autant que les politiques monétaires se décidaient un peu moins du côté du passif des banques, c'est-à-dire sur la masse monétaire, et de plus en plus sur les actifs des banques, c'est-à-dire sur le crédit. Le crédit est devenu un enjeu absolument majeur.

Avec la crise s'est développée aussi la fonction macro-prudentielle. En 2011, le Comité européen du risque systémique (*European Systemic Risk Board* en anglais) a été créé et son secrétariat est hébergé par la BCE, et la BCE a également été investie d'une fonction macro-prudentielle pour la zone euro. En parallèle se sont constituées les autorités de supervision pour la banque, les assurances et les marchés qui participent elles aussi à la fonction macro-prudentielle. Même si ce concept existait de longue date, puisqu'il est né en 1966 dans un document de la BRI, il était resté en second rang dans les préoccupations des banques centrales et des décideurs économiques. Or la crise économique a mis en avant le fait que même des entités qui *a priori* au niveau micro-prudentiel étaient relativement bien gérées et solvables pouvaient être confrontées à un risque de contagion. Ce pôle s'est alors développé.

Plus récemment, en novembre 2014, la BCE a pris la responsabilité de la supervision bancaire. Ces trois fonctions fondamentales de la BCE et du système européen de banques centrales : la politique monétaire, la politique macro-prudentielle et la supervision bancaire, doivent être servies par la fonction statistique. Or elles ont des angles de vue différents sur le crédit et le risque de crédit. Ainsi, les mêmes données doivent être utilisées pour des besoins différents. Il ne s'agit plus seulement d'utiliser des agrégats. La crise a en effet montré l'importance des phénomènes de queue de distribution pour l'analyse monétaire, l'analyse prudentielle et les analyses de marché. L'analyse de cet aspect distributif exige d'examiner ce qui se passe dans le détail au niveau des agents, voire au niveau des instruments et des marchés, c'est-à-dire au niveau granulaire.

La réponse a consisté à étudier les registres de crédit. Or ces registres étaient trop différents. Il convenait donc de les remplacer par un registre de crédit commun à la zone euro. Le projet AnaCredit a été initié en 2013 pour aboutir en 2018. Depuis novembre 2018 sont ainsi rassemblées les données de 2 500 banques à travers la zone euro sur les prêts aux personnes morales. Le registre recouvre actuellement 25 millions de prêts et 10 millions de personnes morales sont incluses dans la base RIAD, un registre d'entreprises. Les pays d'Europe hors zone euro se sont également investis dans le projet AnaCredit et la plupart sont arrivés à maturité. Nous sommes passés de 4 à 88 variables de prêt communes à travers toute l'UE. Outre la description du prêt, nous disposons d'informations sur le taux d'intérêt qui n'existaient que dans le registre de crédit du Portugal avant la crise, de variables qui permettent d'analyser les prêts et surtout de données sur les entreprises prêteuses, emprunteuses ou garantes, qui offrent la possibilité de réaliser différentes classifications pour des besoins statistiques, prudentiels ou d'analyse de marché. Les données sont mensuelles et couvrent 98 % des prêts en application du principe de proportionnalité.

Le projet visait à obtenir le même outil d'analyse pour tous. Dans le domaine de la politique monétaire, par exemple, si nous craignons un « *credit crunch* », nous devons identifier s'il résulte de l'offre ou de la demande. En 2007-2008, il était absolument essentiel de déterminer si les banques ou les pays éprouvaient des difficultés pour offrir le crédit ou si le problème venait de la demande de crédit des entreprises. Avec la fragmentation des marchés et encore plus avec la crise du Covid, ces questions deviennent cruciales. Entre le secteur de l'hébergement-restauration et le secteur du numérique, nous voyons bien que les comportements sont totalement différents. Il faut pouvoir analyser ces comportements secteur par secteur, voire entreprise par entreprise. Toujours en matière de politique monétaire, il est important de savoir comment se répand la liquidité, si celle-ci est également répartie entre les acteurs, si les acteurs qui ont vraiment besoin de ces liquidités y ont réellement accès, etc. Comprendre les vulnérabilités au 1^{er} comme au 2^e rang apparaît également important pour la gestion des risques. Il s'agit aussi de déterminer si ces risques restent domestiques ou s'ils sont beaucoup plus larges et peuvent présenter des impacts systémiques. Quant à la stabilité financière, ces données sont utiles pour étudier toutes les interactions entre les filiales des grands groupes, les effets de contagion dans la distribution du crédit ou les contraintes dans cette distribution. Ces données peuvent aussi être utilisées pour évaluer dans quelle mesure le crédit accompagne le développement durable.

Enfin, ces données servent à la supervision micro-prudentielle. Plus nous disposons d'informations sur le portefeuille, plus nous pouvons effectuer un contrôle précis des établissements et de la manière dont ils distribuent le crédit. Nous pouvons aussi voir si les notes relatives à la probabilité de défaut sont alignées ou si certains établissements se montrent plus optimistes ou plus pessimistes que la moyenne, ce qui permet de réaliser une analyse très fine de l'analyse de crédit des banques. Pour les statisticiens et les économistes, ces données offrent la possibilité de reconstituer des agrégats et de les utiliser dans le passé.

Ce projet a fait l'objet d'une intensive communication auprès des 2 500 banques qui font du *reporting* au travers de la zone euro afin d'obtenir les bonnes classifications, les bons contrôles de validité, etc. Un dictionnaire de données a également été établi conjointement avec les banques. Mis à la disposition de celles-ci, il contient les définitions des données, mais aussi les règles de transformation qui peuvent aider

les banques à réaliser une bonne programmation. Ce métalangage permet une communication mieux comprise entre les établissements et les banques centrales. En retour, les banques comprennent aussi mieux les données et les concepts sous-jacents.

Nous avons également besoin d'un registre des entreprises. Nous n'avons pas pu avoir accès au registre EGR d'Eurostat (le registre des groupes d'entreprises), car celui-ci ne pouvait être utilisé qu'à des fins statistiques. Or AnaCredit n'est pas seulement conçu à des fins statistiques ; il sert aussi pour l'analyse. Nous avons donc besoin d'un registre utilisable pour tous les métiers des banques centrales. Nous l'avons bâti selon le même modèle de données que l'EGR. Dès lors que les obstacles juridiques auront été levés, nous pourrions donc effectuer des échanges de données entre RIAD et l'EGR. Ce registre recouvre 10 millions d'entreprises et 25 000 groupes, avec une historisation complète des données. Le fait d'avoir une vue centralisée et commune des entreprises et des groupes permet de réconcilier de l'information qui serait à défaut fragmentaire, ce qui se révèle essentiel pour assurer la cohérence et la complétude de l'information transfrontière. Nous n'avons pas seulement accès aux crédits d'un pays ; nous pouvons connaître les crédits de tous les pays pour les mêmes entreprises ou les mêmes groupes d'entreprises.

Depuis 2017, nous travaillons sur l'exposition des banques sur les entreprises. Sur ce sujet, nous avons besoin des données de crédit, mais aussi des données de titres. Avec RIAD, un seul identifiant sert à la fois pour les données sur les émetteurs et les détenteurs de titres comme pour les parties aux prêts, y compris les engagements hors bilan. Nous sommes ainsi capables de faire une structure renforcée de l'exposition au risque et de l'endettement réel des entreprises individuellement ou par secteur.

L'accès aux données européennes peut se faire *via* la Banque de France. Nous sommes actuellement confrontés à une problématique que nous espérons voir levée rapidement. Des données individuelles de personnes physiques auraient été intégrées dans AnaCredit alors qu'elles n'étaient pas attendues. De ce fait, le *Data Protection Officer* examine de près ces données pour vérifier que le RGPD ne s'appliquerait pas. En principe, AnaCredit ne couvre en effet que les prêts aux personnes morales qui ne sont pas couverts par le RGPD.

Anne-Laure DELATTE

Merci pour cette présentation qui montre que nous avons réalisé de nombreux progrès. Je vous propose d'ouvrir les échanges.

III.3 Échanges

Benjamin HADJIBEYLI, DG Trésor

Je travaille au pôle d'analyse économique du secteur financier et je suis un fervent utilisateur du CASD. Nous travaillons énormément avec l'Inspection générale des finances (IGF) sur l'impact de la crise. Je crois d'ailleurs que des discussions sont nouées à un haut niveau entre le Trésor et la Banque de France pour nous ouvrir l'accès aux données. Dans la présentation, vous évoquiez un contexte juridique particulier pour le ministère des Finances. Pouvez-vous nous donner plus de précisions sur le sujet ?

Sur AnaCredit, pouvez-vous en dire plus sur les modalités d'accès aux données, notamment *via* la Banque de France ? Les données seront-elles également accessibles par le biais du CASD ?

Roxane SILBERMAN

Les chercheurs de la Banque de France travaillent-ils dans Muses ou dans l'ODR ? Par ailleurs, des travaux de comparaison ont-ils été menés sur les données pseudonymisées et les données de Muses ?

Kamel GADOUCHE

La présentation d'AnaCredit était très intéressante et instructive. J'ai été impressionné par le travail réalisé et la portée des données présentées. J'ai éprouvé cependant une certaine difficulté à voir l'articulation entre la Banque de France et la BCE sur ce sujet. Qui porte la coordination du projet ? La base de données pourrait-elle être accessible à des fins de recherche ? Dans quelle mesure ces données pourraient-elles être appariées avec d'autres sources de données dans d'autres environnements sécurisés ?

Anne-Laure DELATTE

J'ajoute une question sur le contexte juridique. Nous comprenons bien que les freins à l'appariement entre le CASD et l'ODR sont de nature technique et juridique. Quels sont précisément ces freins juridiques ? Pour quelle raison le ministère des Finances bénéficie-t-il d'un statut lui permettant d'accéder plus facilement à ces données ?

Vincent DUPRAT

Ce sujet a fait l'objet d'une instruction spécifique. Je n'ai pas les textes en tête. Je vous les communiquerai hors séance. Dans la convention que nous avons signée avec le CASD, seules les personnes de l'IGF ont le droit d'accéder aux données. Même l'accès aux données par France Stratégie posait problème. L'IGF appartenant au ministère des Finances a le droit de manipuler des données individuelles, d'en tirer un certain nombre d'analyses et de transmettre ses conclusions à France Stratégique qui, à l'inverse, n'a pas le droit d'accéder à ces données individuelles.

Anne-Laure DELATTE

Les deux barrières juridiques sont liées à l'autorisation de transfert et à la nécessité de pseudonymiser les données. Pour France Stratégie, le problème semblait plutôt tenir au transfert.

Vincent DUPRAT

Certaines collaborations entre le ministère des Finances et la Banque de France permettent de s'échanger librement des données concourant à l'analyse économique. Je retrouverai le texte précis. En l'occurrence, l'instruction juridique a porté sur un usager membre du ministère de l'Economie pour des données du registre des crédits. Elle a finalement donné une réponse positive et nous avons obtenu assez facilement un accord pour que ces données ne soient pas anonymisées. À l'inverse, nous avons reçu une demande du ministère de la Défense qui souhaitait savoir si ses sous-traitants présentaient des anomalies de paiement et une probabilité de défaut élevée. Nous leur avons refusé l'accès, puisqu'il n'existe pas de base juridique avec le ministère de la Défense. Ces informations ne visent pas à contribuer à de la recherche ou de l'analyse économique, mais servent les propres besoins du ministère. Nous avons la volonté de coopérer. Pour autant, nous n'avons pas encore trouvé de base juridique et nous ne sommes pas capables de lui donner des indications sur la solidité de ses sous-traitants.

Anne-Laure DELATTE

Nous nous interrogeons davantage sur l'accès pour les chercheurs extérieurs aux données de la Banque de France et de la statistique publique. Il conviendrait d'identifier, par type de données, celles qui sont transférables, celles qui doivent être pseudonymisées, etc. Or nous avons l'impression qu'il faudrait instruire la question juridique de façon plus globale.

Vincent DUPRAT

Nous en sommes d'accord. La doctrine de base est en cours de révision. Delphine Moreau a présenté le projet devant le secrétaire général ce matin et le présentera lundi devant le CGR, avec le gouvernement de la Banque. Nous nous inscrivons dans cette volonté de n'anonymiser que les données qui doivent nécessairement l'être. Pour l'instant, cependant, nous anonymisons tout par défaut et chaque jeu de données fait l'objet d'une instruction spécifique. Il n'existe pas de procédure d'instruction *a priori* qui permettrait de mettre à disposition les données plus rapidement.

Anne-Laure DELATTE

Ce procédé est un peu décevant. Nous observons une avancée vraiment importante sur l'enjeu d'appariement entre les données administratives du CASD et les données financières. Nous sommes heureux de savoir que quelques chercheurs pourront y avoir accès. Néanmoins, pour éviter l'effet d'opportunité, il faudrait donner accès à ces données sur une plate-forme commune. Pour cela, il conviendrait sans doute de mener une instruction juridique plus globale qui nous permettrait d'envisager de systématiser l'accès à ces données hors ODR. Je vous propose d'évoquer le sujet dans l'avis.

Vincent DUPRAT

Les chercheurs internes à la Banque de France accèdent à Muses tandis que les chercheurs externes accèdent à l'ODR. La Bundesbank n'a pas retenu la même segmentation, préférant mettre en place un système pour les analystes et un autre pour les chercheurs.

Roxane SILBERMAN

Nous avons connu cette situation pour la statistique publique qui distinguait au départ les chercheurs et les analystes internes au SSP. Je constate que l'on retrouve un peu le même système du côté des données bancaires. Il faut instruire les contraintes juridiques, car il va devenir indispensable de changer le cadre. Du côté de la statistique publique, pour ce qui est de l'accès aux données, on a vu la frontière s'effacer peu à peu entre les chercheurs et les analystes. Au départ, les modalités d'accès étaient strictement séparées. Ainsi., même les fichiers de production et de recherche (FPR) anonymisés de l'Insee ne pouvaient pas être communiqués par Quetelet, le centre en charge de cette diffusion, aux chercheurs de la Banque de France, parce qu'ils étaient considérés comme analystes interne à la BDF. Aujourd'hui, cette frontière s'efface dans un monde où le contexte des données est bien différent.

Kamel GADOUCHE

Les textes relatifs à l'accès aux données pour le ministère des Finances prévoyaient déjà des possibilités de transfert des données. C'est d'ailleurs dans ce cadre que nous avons étudié la possibilité de mettre à la disposition des inspecteurs des finances les données au sein de leur environnement de calcul. Pour la DG Trésor, je pense que le *fast-track* est tout aussi valable. Les vecteurs juridiques existent en tout cas.

Je tenais aussi à apporter une précision sur la pseudonymisation. Le hachage de l'identifiant est réalisé à la Banque de France, mais il ne l'est pas au CASD pour plusieurs raisons. Pour autant, il est possible techniquement de réaliser, sur les sources présentes au CASD, le même type de hachage de l'identifiant pour permettre les appariements. D'un point de vue théorique, il n'est donc pas impossible de réaliser les appariements, même si une pseudonymisation est opérée. L'exercice peut néanmoins nécessiter un travail extrêmement conséquent, puisqu'il s'agit de pseudonymiser les identifiants sur toutes les sources. Or un projet de recherche peut mobiliser plusieurs dizaines de sources. Maintenir le hachage des identifiants sur toute la durée d'un projet de recherche se révélerait donc très coûteux en ressources.

Jean-Marc ISRAËL

L'accès aux données s'effectuera pour l'instant sous l'ODR. Il est évident que nous examinons de très près l'initiative Inexda qui permettrait de définir des règles communes pour l'ensemble des chercheurs en Europe et au-delà. Cette évolution entraînerait une extension et une certaine égalité entre les chercheurs.

Quant à l'articulation entre la BCE et la Banque de France, la BCE joue évidemment le rôle d'aiguillon dans ce projet, mais le travail a été réalisé de manière concertée et collective par les membres de l'Eurosystème et même de l'Union européenne et la BCE. Nous travaillons effectivement dans le Comité des statistiques qui rapporte au Conseil des Gouverneurs. De fait, la BCE et les banques centrales nationales ne peuvent qu'avancer effectivement au même pas.

Ce méta-projet a nécessité de coordonner 2 500 banques et 27 banques centrales et de mettre en place des systèmes informatiques, un cadre juridique contraignant tout en restant flexible et un échange de données, avec deux projets : RIAD, le registre des entreprises étendu aux entreprises non financières, et AnaCredit pour les prêts. La Banque de France a accusé un léger retard, lié à des problèmes de cohérence et de priorité. Le projet est cependant arrivé à son terme à la bonne date, en septembre 2018. Nous disposons donc aujourd'hui d'un certain recul.

Quant à l'accès aux données, plus la coopération se développera à plusieurs niveaux, plus l'accès aux données sera facilité. Sur AnaCredit, il existe deux niveaux de confidentialité : d'une part, du côté des prêteurs et des garants des prêts et d'autre part du côté des emprunteurs. *A priori*, nous pourrions très bien imaginer une pseudonymisation avant ou après, à la condition de définir un cadre juridique et une relation de confiance permettant d'utiliser les données dans l'esprit pour lequel elles sont mises à disposition.

Anne-Laure DELATTE

Avec RIAD, AnaCredit et SHS, avons-nous fait le tour des bases européennes de données individuelles ? Ces trois bases sont-elles disponibles au niveau de l'ODR ?

Jean-Marc ISRAËL

Nous n'avons pas fait le tour des bases européennes. Il ne faut pas oublier non plus EMIR, une base non statistique, mais très importante pour les produits dérivés, qui existe depuis 2014, ou *Money Market Statistical Reporting*, une base de données de toutes les transactions du marché monétaire à travers la zone euro (environ 50 000 par jour) qui existe depuis 2016. Avec SHS, AnaCredit et RIAD, nous couvrons quand même une bonne partie du bilan des institutions financières dans le financement de l'économie, mais les données sont très nombreuses et volumineuses.

Anne-Laure DELATTE

Les données sont nombreuses, mais SHS, AnaCredit et RIAD permettent d'avoir une bonne vision du financement de l'économie. Ces trois bases sont-elles disponibles dans l'ODR ?

Vincent DUPRAT

Non.

Anne-Laure DELATTE

Sauf AnaCredit.

Vincent DUPRAT

AnaCredit n'est pas disponible non plus. L'ODR constituerait la meilleure solution de mise à disposition, mais cette mise à disposition n'est pas effective pour l'instant.

Roxane SILBERMAN

Existe-t-il un contrôle de la confidentialité des résultats sur ces bases ?

Vincent DUPRAT

Oui.

Roxane SILBERMAN

S'agissant d'une base européenne, chaque institution doit-elle donner son aval ? Nous rencontrons des problèmes de ce type sur les micro-données européennes avec Eurostat. Des accords ont-ils été conclus sur le sujet ?

Vincent DUPRAT

Nous avons eu un *call* avec les Allemands hier matin pour essayer d'interpréter les signaux envoyés par la BCE sur la possibilité de partager les données d'AnaCredit. Dans les données d'un pays, il faut tenir compte de la problématique de « home »/« host ». Dans un même groupe, la connexion des données n'est pas forcément homogène. Ainsi, les données de BNP Paribas Italie pourraient être collectées par la banque centrale italienne alors que celles de BNP Paribas Espagne seraient collectées par la Banque de France. Les accords bilatéraux de complémentarité entre les banques centrales sont en effet différents. En outre, le processus de partage des données est de toute façon bloqué, puisque nous nous heurtons à un problème de confidentialité de certaines données. Des données potentiellement soumises au RGPD ont effectivement été collectées à tort dans la base.

Aujourd'hui, le sujet de partage des données européennes n'est pas clair pour moi. Les données européennes ont été considérées comme pouvant être partagées, en raison de la crise Covid et à titre

temporaire, avec un certain niveau d'agrégation. Nous avons par ailleurs la liberté de partager les données collectées nationalement à un niveau granulaire. La collecte d'AnaCredit reste assez jeune. Les textes de partage de ces données ne sont pas encore totalement établis. La démarche soulève également des enjeux de qualité de données. La BCE ne souhaitait pas partager des données tant qu'elle estimait que le niveau de qualité n'était pas suffisant.

Anne-Laure DELATTE

La situation est un peu frustrante pour les chercheurs. Altavilla et Pedro ont publié un document voilà un an sur les données d'AnaCredit. Certes, ces deux chercheurs ont toute légitimité, puisqu'ils sont à l'origine de la littérature sur l'utilisation de ces données individuelles pour l'examen des questions de politique monétaire. Cela démontre toutefois que la qualité est au rendez-vous et que ces données peuvent être utilisées, mais que leur accès n'est pas assuré de manière égale pour tous.

Vincent DUPRAT

Je ne connais pas la qualité de ce document. J'ai vu passer des *reportings* de la BCE sur la qualité des données pays par pays et les indicateurs n'étaient pas tous au vert par rapport au seuil d'acceptabilité. Ils ont pris une photographie des données telles qu'elles existaient. J'ignore s'ils ont redressé certaines anomalies, mais il en existait forcément.

Jean-Marc ISRAËL

L'accès aux utilisateurs internes à l'Eurosystème n'a été accordé par le Comité des statistiques qu'en avril dernier. Je ne sais pas sur quelle base ils ont réalisé leur papier. Il s'agissait très certainement de données de test qui n'affichaient pas le niveau de stabilité requis.

Anne-Laure DELATTE

Cela signifie qu'il faut être chercheur dans l'une des banques centrales de l'Eurosystème pour avoir accès à AnaCredit, SHS ou RIAD ?

Jean-Marc ISRAËL

Pas forcément. Pendant longtemps, des organismes de recherche au sein de l'Eurosystème se réunissaient en *pool* avec des universitaires pour réaliser des papiers. Cette solution existe encore. Aujourd'hui, nous nous dirigeons vers une ouverture plus large aux chercheurs par des moyens modernes, en particulier sous l'initiative d'Inexda, mais cette ouverture s'effectuera par étape et il faudra un peu de temps pour parvenir à la stabilité et la confiance sur l'utilisation des données et le respect de la confidentialité. La volonté existe d'apparier des données différentes. Il faut cependant arriver à bien définir les conditions dans lesquelles cet accès peut se faire. Travailler sur des données granulaires est compliqué. Un accompagnement méthodologique apparaît donc nécessaire pour assurer leur bonne interprétation.

III.4 Avis de la commission

Isabelle ANXIONNAZ

Nous vous proposons l'avis suivant :

L'exploitation statistique du gisement d'informations individuelles bancaires et financières produit à des fins de régulation et d'évaluation des politiques a constitué une « révolution » dans les années post-crise de 2008-2009. Ces données granulaires au niveau des ménages ou des firmes sont devenues un outil clé dans l'analyse empirique, car elles permettent d'identifier les mécanismes sous-jacents aux grandes relations économiques. Depuis quelques années, des travaux sont en cours tant au niveau national qu'au niveau européen pour développer le recours à ces données et la commission a estimé qu'il était pertinent de faire le point sur ce sujet.

Les interventions de la Banque de France et du CASD (Centre d'accès sécurisé aux données) mettent en lumière deux grands enjeux pour la statistique publique dans la mise à disposition de ces données. Le premier est celui de leur accès : en France, les deux sources principales sont les bases

de la Banque de France, mises à disposition via l'*Open data Room* et celles de l'Insee, disponibles via le CASD. Dans la continuité des recommandations de la réunion du 23 mai 2019, la commission souligne qu'un projet de rapprochement des accès est en cours de discussion, dont l'objectif serait de permettre aux chercheurs de disposer au sein d'un même environnement des données granulaires financières avec des données socioéconomiques. Le second enjeu de disponibilité concerne le cloisonnement national des données, qui rendait difficile jusqu'à récemment l'appréhension globale des enjeux internationaux. Il existe depuis peu plusieurs projets visant à rapprocher des sources nationales et à les agréger au niveau européen (projets Inexda sur des échanges d'expériences et AnaCredit sur l'enrichissement des données sur le crédit).

La commission soutient fortement ces initiatives qui visent à faciliter un accès aux données transdisciplinaires intégrant les aspects financiers et à encourager les recherches sur les données européennes, contribuant ainsi à une meilleure compréhension des enjeux économiques internationaux.

La commission souhaite être informée de la progression des travaux sur ce sujet, notamment de l'instruction des questions juridiques.

Je proposerais d'ajouter un paragraphe *ad hoc* sur ces questions juridiques :

« La commission note que les questions juridiques constituent encore le frein principal au rapprochement des données, notamment par une plate-forme commune ».

Vincent DUPRAT

Nous n'avons pas de projet de plate-forme commune et les problèmes juridiques se posent aussi pour nos échanges croisés. Je ne lierais donc pas les deux. Nous pouvons très bien imaginer des modalités qui rendraient le service sans la construction d'une plate-forme commune.

J'émetts également des réserves sur l'idée d'un même environnement. Littéralement, l'expression est vraie, puisqu'il s'agira de l'environnement du CASD ou de celui de la Banque de France, mais la formulation laisse entendre la création d'une plate-forme tierce commune. Or ce projet n'est pas d'actualité aujourd'hui.

Anne-Laure DELATTE

Il serait cependant souhaitable. Nous pouvons trouver une autre formulation.

Vincent DUPRAT

Nous pourrions indiquer : « *de disposer de manière transversale...* »

Kamel GADOUCHE

Je pense au contraire que l'expression « *au sein d'un même environnement* » est particulièrement juste, puisque ce sera soit l'environnement du CASD, soit celui de l'ODR.

Vincent DUPRAT

Mais sera-t-elle comprise ainsi ?

Kamel GADOUCHE

Il est vrai qu'il n'est pas du tout prévu de bâtir une nouvelle plate-forme.

Roxane SILBERMAN

Il n'existe pas non plus d'exemple à l'étranger. Pour moi, un même environnement de recherche ne signifie pas forcément une plate-forme. Il importe avant tout pour le chercheur que ces données soient accessibles dans le même environnement de recherche. Il ne s'agit pas forcément de construire une nouvelle plate-forme.

Kamel GADOUCHE

L'objectif consiste à mettre les données à disposition non pas en parallèle, mais dans un seul environnement pour pouvoir les appairer et travailler conjointement sur les données.

Anne-Laure DELATTE

L'expression a été utilisée dans la présentation de Roxane Silberman : « *le traitement conjoint de données dans un même environnement de recherche* ».

Vincent DUPRAT

Très bien.

Isabelle ANXIONNAZ

Ajoutons-nous un petit paragraphe sur les questions juridiques ?

Anne-Laure DELATTE

Nous souhaitons être informés. Nous pourrions même avancer l'idée d'un groupe de travail sur ces questions juridiques.

Roxane SILBERMAN

Je crois comprendre que les questions juridiques sont particulièrement importantes s'agissant de la distinction entre chercheurs et analystes. La question de la pseudonymisation se pose aussi, mais j'ignore si elle est de nature juridique.

Vincent DUPRAT

Elle peut l'être, car certaines données peuvent être partagées à condition qu'elles soient pseudonymisées. Des textes le prévoient explicitement.

Roxane SILBERMAN

Il serait bon que le paragraphe précise ces deux points.

Anne-Laure DELATTE

Transfert des données et pseudonymisation constituent bien les deux enjeux juridiques que nous devons éclaircir.

Isabelle ANXIONNAZ

Je propose que nous ajoutons : « *questions juridiques relativement au transfert des données et à leur pseudonymisation* ».

Vincent DUPRAT

Vous parlez de transfert alors que nous parlons plutôt de partage.

Kamel GADOUCHE

J'émettrais aussi une réserve sur le terme de transfert. Il faudrait plutôt parler d'accès aux données par les chercheurs.

Vincent DUPRAT

Le terme d'accès me convient aussi.

Kamel GADOUCHE

Je formulerais également une petite réserve sur la pseudonymisation. Ce sujet est très important d'un point de vue pratique, mais dans l'absolu, il reste assez secondaire. Il existe des pis-aller techniques. Il suffit que la pseudonymisation soit la même des deux côtés pour que l'appariement se déroule dans de bonnes conditions. Faites-vous référence à la nécessité de leur pseudonymisation ?

Isabelle ANXIONNAZ

Nous évoquons le traitement de cette pseudonymisation, le rapprochement des données pseudonymisées et non pseudonymisées.

Vincent DUPRAT

Ce sujet doit être abordé sous l'angle juridique : ces données doivent-elles être pseudonymisées ou non ?

Jean-Marc ISRAËL

Le but ne réside pas dans la pseudonymisation, mais dans la protection de la confidentialité. Ne pourrions-nous donc pas indiquer « *et à la protection de leur confidentialité dans l'usage* » ?

Anne-Laure DELATTE

Cela me semble effectivement préférable.

Vincent DUPRAT

La pseudonymisation constitue en effet un moyen parmi d'autres. Certains textes exigent de faire disparaître la manière d'identifier soit en ne donnant que des données agrégées, soit en pseudonymisant des données individuelles. La finalité reste la garantie de la confidentialité.

Jean-Marc ISRAËL

J'ajouterais « *dans leur usage* ».

Vincent DUPRAT

Ceci explique par exemple que l'IGF a le droit d'accéder à des données granulaires, mais ne peut sortir que des données agrégées pour les confier à France Stratégie.

Anne-Laure DELATTE

Pouvons-nous lancer l'idée d'un groupe de travail ?

Isabelle ANXIONNAZ

Nous ne formulons aucune recommandation sur l'avancée pratique de ces questions juridiques. Il faudrait faire un état des lieux et des préconisations.

Anne-Laure DELATTE

Cette phrase lance l'idée d'un travail sur le sujet. Nous préconisons quand même une instruction plus systématique de ces enjeux juridiques qui apparaissent comme le frein principal. J'aurais ajouté SHS et RIAD dans les parenthèses à la fin du 2^e paragraphe, puisque ces bases ont été évoquées dans les présentations. Inexda ne se place pas en revanche sur le même niveau.

Vincent DUPRAT

À ce stade, Inexda ne recouvre que des échanges de pratiques et ne représente aucunement un rapprochement de données.

Anne-Laure DELATTE

Nous l'avons bien noté dans la présentation. Il conviendrait de le préciser. Nous pourrions citer uniquement AnaCredit, SHS et Riad. Ces données sont maintenant disponibles aux chercheurs de l'Eurosystème, mais pas encore l'extérieur. Ne faudrait-il pas évoquer ce point ?

David ADAM

Je pense qu'il faut ajouter les données EMIR sur les dérivés, car elles sont très importantes pour réaliser des analyses complètes de la situation, notamment pour la stabilité financière. Comme mes collègues l'ont montré, la volonté de partage existe. Je ne pense pas qu'il y ait une volonté d'ostraciser les chercheurs externes par rapport aux chercheurs internes, même si, de fait, les modèles de données sous-jacents sont plus connus en interne, en particulier pour ceux qui ont contribué à la construction des outils. L'un des problèmes que nous avons rencontrés sur AnaCredit tient au fait qu'il s'agit d'un projet de l'Eurosystème qui implique dans sa construction certaines banques centrales, mais pas toutes. Si toutes les banques centrales participent à la définition des besoins, seules certaines prennent en charge la fabrication du produit lui-même. Or la Banque de France n'était pas au cœur de la fabrication d'AnaCredit. C'est peut-être la raison pour laquelle les Italiens ou les Portugais avaient une longueur d'avance dans l'utilisation des données, puisqu'ils connaissent très bien les modèles de construction sous-jacents. En théorie et juridiquement, il n'existe pas de blocage pour l'accès aux informations dès lors que celles-ci présentent la qualité requise, ce qui n'était pas le cas aujourd'hui. Cet avis doit rendre l'intérêt pour la communauté des chercheurs français à accéder aux données européennes qui sont en voie de constitution.

Anne-Laure DELATTE

Dans le paragraphe suivant, après « *la commission soutient fortement ces initiatives qui visent à faciliter un accès aux données transdisciplinaires intégrant les aspects financiers* », j'indiquerais « *et à encourager l'accès des chercheurs externes aux données européennes, une fois que la qualité de ces données sera confirmée par les producteurs eux-mêmes* ».

David ADAM

Nous progressons sur le sujet en vue d'une ouverture progressive de ces données. Je voudrais balayer l'idée qu'au sein de la Banque, les chercheurs sont un peu privilégiés par rapport aux chercheurs externes. Lorsque le FMI est venu réaliser son audit du système financier français, il a utilisé l'ODR. Nous avons dû bricoler des bases de données pour lui permettre d'accéder aux informations que nous ne possédions pas nous-mêmes. Au sein même de la Banque, nous rencontrons des difficultés pour combiner certaines bases de données entre elles, liées à l'insuffisante qualité des données statistiques à ce stade. L'effort de la dernière décennie a porté davantage sur la collecte des informations que sur leur utilisation. Nous avons vu des chantiers de collecte comme nous n'en avons jamais vu auparavant. Le chantier sur l'utilisation et la combinaison reste devant nous.

Kamel GADOUCHE

Je préciserais la notion de chercheurs externes. Ne pourrions-nous pas évoquer les chercheurs « *académiques* » ?

Anne-Laure DELATTE

Un chercheur peut être un académique tout en étant associé à une banque centrale.

David ADAM

Nous n'avons pas forcément besoin de préciser. Nous pourrions indiquer « *pour l'intégralité des chercheurs* ». Je ne souhaite pas perpétuer l'idée qu'il existe des chercheurs privilégiés au sein des banques centrales, car ce n'est pas le cas.

Roxane SILBERMAN

J'en suis d'accord.

David ADAM

Par rapport au cadrage juridique, l'un des textes fondateurs émane du Haut Conseil de stabilité financière qui regroupe à la fois l'AMF (l'Autorité des marchés financiers) pour les marchés, l'ACPR pour les banques et les assurances, le gouverneur de la Banque de France et trois académiques. Dans le Code monétaire et financier, les textes qui fondent ce haut conseil soulignent que celui-ci doit veiller à la circulation des informations pour réaliser toutes les analyses nécessaires à l'examen de la stabilité financière. Il existe donc bien une volonté politique très forte de partager les informations chaque fois que cela est possible et que cela ne risque pas d'entraîner de fausses analyses, du fait de données insuffisamment sécurisées ou de mauvaise qualité.

Jean-Marc ISRAËL

Je suis gêné par l'expression « *l'ensemble des chercheurs* ». Dans les domaines de la santé, il existe des chercheurs de très haut niveau et des chercheurs de moindre niveau, qui ne font pas que de la recherche. Ne pourrions-nous pas trouver une formulation plus orientée vers la qualité de la recherche, en distinguant les chercheurs selon qu'ils poursuivent véritablement un but de recherche ou un objectif tout autre ?

Roxane SILBERMAN

Je pense que le sujet relève des modalités d'accréditation mises en place. Il me paraît compliqué d'entrer dans une telle distinction.

Jean-Marc ISRAËL

J'en suis d'accord. Pour autant, le fait d'évoquer l'ensemble des chercheurs induit quand même une ouverture extrêmement large par rapport à un objectif qui doit rester scientifique et de vraie recherche.

David ADAM

Nous pourrions évoquer « *l'ensemble des chercheurs accrédités* », ce qui laisse entendre qu'une autorité est en charge de donner un accord.

Roxane SILBERMAN

Dans d'autres domaines, nous avons toujours indiqué l'ensemble des chercheurs et cela n'a jamais empêché d'opérer un filtrage dans les modalités d'accréditation. Les discussions sur les bonnes pratiques dans le cadre d'Inexda portent bien sur les accréditations. Il existe d'ailleurs des pratiques différentes en la matière.

Vincent DUPRAT

Ne pourrions-nous pas indiquer « *pour la communauté des chercheurs* » pour réduire le caractère un exhaustif ?

Kamel GADOUCHE

Dans les textes, nous ne précisons pas la qualité des personnes, puisqu'il n'existe pas de définition des chercheurs. Nous évoquons plutôt les usages « *à des fins de recherche scientifique* ».

Jean-Marc ISRAËL

C'est mieux.

Roxane SILBERMAN

Nous pouvons aussi évoquer « *l'ensemble des chercheurs une fois accrédités* », ce qui montre qu'il existe un processus d'accréditation.

Kamel GADOUCHE

Cette expression me semble encore plus limitative. Dans les textes, en général, nous définissons la finalité.

Anne-Laure DELATTE

J'ai peur que nous ayons dilué l'idée principale d'ouvrir l'accès aux chercheurs internes et externes.

Roxane SILBERMAN

La loi Informatique et Libertés et le Règlement général pour la protection des données (RGPD) ont toujours fait référence aux chercheurs.

Kamel GADOUCHE

Ces deux textes évoquent des finalités de recherche. Il en est de même dans la loi de 1951. J'avais proposé « *à tous les usages à des fins de recherche scientifique* », mais l'expression est peut-être trop large.

Roxane SILBERMAN

C'est la finalité qui définit, pas le statut.

Kamel GADOUCHE

Tout à fait. Il n'existe pas de définition du chercheur. Nous pourrions revenir à l'expression « *chercheurs externes au système bancaire* ».

Anne-Laure DELATTE

L'expression « *les chercheurs externes à l'Eurosystème* » ne répond pas aux commentaires de M. Israël.

Roxane SILBERMAN

Nous pourrions indiquer « *pour tous les travaux à des fins de recherche scientifique* ».

Anne-Laure DELATTE

Tout à fait

Dans le 1^{er} paragraphe, j'aurais ajouté ce que David Adam nous a montré dans son intervention sur la stabilité financière et l'effet de distribution. Il faudrait indiquer « *car elles permettent d'identifier les mécanismes sous-jacents aux grandes relations économiques et de mieux cerner la distribution des risques afin d'appréhender les enjeux de stabilité financière* ».

David ADAM

Nous devons effectivement mentionner la stabilité financière de manière explicite. Nous pensions que le système financier pouvait perturber l'économie réelle, mais nous constatons aujourd'hui que les interactions sont beaucoup plus riches entre les deux sphères et que nous avons besoin d'avoir une bonne compréhension de l'ensemble des mécanismes. Or sans données granulaires, l'exercice est impossible à réaliser.

L'avis tel que présenté et amendé en séance est adopté.

CONCLUSION

Anne-Laure DELATTE

Merci pour cette séance très riche. Je pense que nous avons bien avancé. Il nous faudrait un agenda un peu plus clair sur la façon dont les questions d'appariement et les enjeux juridiques progressent. Nous aurons sans doute l'occasion d'en reparler dans les commissions à venir. Nous avons quand même beaucoup appris aujourd'hui.

Roxane SILBERMAN

Nous pourrions également en discuter lors de la réunion du Bureau de décembre en présence de tous les présidents de commissions.

Isabelle ANXIONNAZ

Le Bureau se tiendra le 9 décembre. Nous faisons toujours un point sur le bilan de l'année.

Anne-Laure DELATTE

Merci à tous.

La séance est levée à 18 heures.

Tous les documents sont disponibles sur le site cnis.fr à la [page de la commission](#).

DOCUMENTS PRÉPARATOIRES

DOCUMENTS COMPLÉMENTAIRES

[Le programme de travail 2020 dans le contexte de la crise sanitaire - Banque de France - Insee - Diaporama \(pdf - 1 Mo\)](#)

[La gestion des avis d'opportunité dans le contexte de la crise sanitaire - Cnis - Diaporama \(pdf - 516 Ko\)](#)

[L'utilisation des données massives pour le suivi conjoncturel lors du confinement dû à la pandémie - Insee - Diaporama \(pdf - 2 Mo\)](#)

[Innovation dans l'accès aux données de la Banque de France via l'Open Data Room - Banque de France - Diaporama \(pdf - 1 Mo\)](#)

[Appariement des données : point sur le projet de collaboration Banque de France - CASD - Diaporama \(pdf - 558 Ko\)](#)

[Les données individuelles disponibles au niveau européen - CASD - Diaporama \(pdf - 3 Mo\)](#)

[Le projet Inexda - Banque de France - Diaporama \(pdf - 720 Ko\)](#)

[Le projet AnaCredit de l'Union européenne - Banque de France - Diaporama \(pdf - 2 Mo\)](#)