

L'accès aux données bancaires et financières : une mission de service public

Gunther Capelle-Blancard, Raphaëlle Bellando

▶ **To cite this version:**

Gunther Capelle-Blancard, Raphaëlle Bellando. L'accès aux données bancaires et financières : une mission de service public. [Rapport de recherche] Conseil national de l'information statistique. 2015. hal-01296626

HAL Id: hal-01296626
https://hal.science/hal-01296626

Submitted on 1 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conseil national
de l'information statistique

L’accès aux données bancaires et financières : une mission de service public

Rapport du groupe de travail du Cnis

Juillet 2015

Gunther CAPELLE-BLANCARD

Raphaëlle BELLANDO

Avant-propos

Ce rapport part du constat suivant : le principal problème, aujourd'hui, en matière de données sur le secteur bancaire et financier ne réside pas tant dans le manque de données que dans les difficultés d'accès à ces données.

Les données bancaires et financières forment un ensemble vaste et hétérogène. Dans aucun autre domaine d'activité on ne retrouve une telle variété de données, pas un secteur ne répond à autant de demandes de *reporting*, aucune industrie ne produit, ni ne consomme une telle masse de données. Et pourtant, ces données sont aujourd'hui largement inaccessibles pour un grand nombre d'utilisateurs, qu'ils soient chercheurs, économistes, conseillers financiers, journalistes, citoyens engagés ou simplement consommateurs de services bancaires et financiers.

L'accès aux données bancaires et financières se heurte à la dispersion des informations, à leurs coûts d'accès, à leur plus ou moins bonne qualité et à d'éventuelles contraintes réglementaires et juridiques. Aucun de ces obstacles n'est toutefois insurmontable ; surtout, ils ne sauraient tenir lieu de prétexte. Après plusieurs mois d'analyse et d'enquête, il m'apparaît que le principal obstacle à un meilleur accès – et donc à une plus large utilisation des données bancaires et financières en France – est d'ordre pratique, voire culturel : il faut passer d'une culture du secret à une culture de l'ouverture aux données.

L'accès le plus large possible aux données doit être vu comme une mission de service public. Les bénéfices seraient nombreux, qu'il s'agisse de la recherche, de la mise en œuvre et du suivi des politiques publiques, de la stabilité financière, de l'information des consommateurs et de l'efficacité du marché. Plus fondamentalement encore, il s'agit de répondre à une demande sociale de transparence, et ce faisant de restaurer la confiance dans notre système bancaire et financier.

L'idée d'un tel rapport a été évoquée dès 2013 dans le cadre de la Commission « Système financier et financement de l'économie » du Cnis. En mars 2014, le bureau du Cnis a validé cette proposition (*cf*. le mandat du groupe de travail ci-joint). Le groupe de travail s'est depuis réuni à six reprises sous ma présidence. J'ai également mené en parallèle de nombreuses discussions avec des collègues chercheurs, français et étrangers, avec des économistes, des juristes, des banquiers, des journalistes, des analystes, des gérants de fonds, des membres de la société civile, etc.

Ce rapport est organisé en deux parties. La première présente la problématique d'ensemble, décrit les solutions envisageables en distinguant les données non-confidentielles et les données confidentielles ; cette partie se conclut par plusieurs recommandations. La seconde partie est organisée sous formes de compléments et donne libre champ aux parties prenantes, producteurs de bases de données (à savoir principalement les autorités de supervision

financière, la Banque de France, l'ACPR et l'AMF) et utilisateurs (chercheurs, ou membres de la société civile).[1] En tant que président du groupe de travail, je me suis chargé de la rédaction de la première partie, mais les recommandations sont formulées au nom du groupe de travail. Elles ont parfois suscité des débats au sein du groupe et n'ont pas toutes le soutien unanime de ses membres. Elles ont, en revanche, toutes été rédigées dans un esprit constructif, avec le souci constant d'un équilibre entre la diffusion des données et la protection du secret statistique. Elles s'articulent principalement autour de deux axes :

- rendre public ce qui peut l'être ;
- améliorer les procédures d'accès aux données.

Dans le domaine de l'accès aux données, nous sommes à un vrai tournant. Partout les données deviennent de plus en plus disponibles. Des données autrefois inaccessibles – car relevant de la vie privée notamment, telles que les données fiscales ou de santé – s'ouvrent au(x) public(s). L'accès aux données est aujourd'hui un vrai enjeu démocratique et stratégique. C'est d'ailleurs le sens de la démarche initiée par le G8 sur l'ouverture des données publiques, matérialisée en France par la mission Etalab. Il n'y a aucune raison pour que le secteur bancaire et financier reste, en France, en marge de ce mouvement.

Sur certains aspects, la France est en pointe en matière d'accès aux données bancaires et financières : c'est le cas par exemple des *reporting* « pays par pays » qui représentent une avancée majeure dans la lutte contre les paradis fiscaux et réglementaires. Dans d'autres, il faut bien le reconnaître, la France est plutôt à la traîne. C'est le cas, par exemple, pour ce qui est des données publiques sur les banques ou les incidents de paiements ; c'est le cas aussi pour ce qui est de la gestion des procédures d'accès aux données confidentielles. Sur ces différents points, la France gagnerait beaucoup à s'inspirer de ce qui se fait aux États-Unis, en Grande Bretagne ou en Allemagne pour l'accès aux données bancaires et financières. Et sans même évoquer les expériences étrangères, la procédure mise en place en France dans le cadre du Comité du secret statistique pour les données de la statistique publique fournit un remarquable exemple de ce qui peut être fait.

Je souhaite enfin insister sur le rôle fondamental qu'ont à jouer les autorités en charge du secteur bancaire et financier, principalement la Banque de France, l'Autorité de Contrôle Prudentiel et de Résolution (ACPR), et l'Autorité des Marchés Financiers (AMF). Ces institutions sont au cœur de la collecte de données bancaires et financières. Ce faisant, elles sont les mieux placées pour répondre aux besoins d'accès aux données. Il ne s'agit évidemment pas de les détourner de leur mission de régulation, ni de réduire leur indépendance – bien au contraire. J'ai la conviction que favoriser l'accès aux données participe à la régulation : les chercheurs, les journalistes, les associations, ainsi que les acteurs de la finance eux-mêmes forment autant de maillons essentiels dans la prévention des dérives et de l'instabilité financière. La transparence ne nuit pas à la stabilité financière. L'accès large aux données est aussi une condition nécessaire à l'exercice de contre-pouvoir. Il en va de la

[1] Selon la formule consacrée, les propos et opinions exprimées dans ce rapport n'engagent que leurs auteurs et ne reflètent pas nécessairement le point de vue des institutions auxquelles ils sont rattachés.

crédibilité du dispositif de régulation et de sa légitimité aux yeux de la société. Les autorités de régulation sont indépendantes, mais ce sont des agences publiques au service du public.

Ce rapport prône le libre accès aux données publiques et la transparence des procédures d'accès aux données confidentielles. Cela nécessite un vrai changement culturel et j'ai pu mesurer dans le cadre de la tenue de ce groupe de travail qu'il se heurtait, encore aujourd'hui, à de nombreuses réticences. Cette demande de transparence, transparence face au pouvoir – celui du secteur bancaire et financier, mais aussi celui des autorités de contrôle – doit pourtant être soutenue. Ces recommandations recevront donc, je l'espère, un accueil favorable du législateur, des autorités de supervision, ainsi que du secteur bancaire et financier lui-même. Nous avons tous à gagner, collectivement, à une plus grande transparence du secteur bancaire et financier.

Gunther Capelle-Blancard

Professeur à l'Université Paris 1 *Panthéon-Sorbonne*
Président de la Commission « Système financier et financement de l'économie » du Cnis

Mandat du Groupe de travail du Cnis rattaché à la Commission *Système financier et financement de l'économie* concernant l'accessibilité aux données relatives au secteur bancaire et financier

Version adoptée par le bureau du Cnis du 20 mars 2014

La demande sociale en faveur d'une plus grande transparence du secteur bancaire et financier est très forte aujourd'hui, même si ces préoccupations ne datent pas de la crise. Ainsi, en 1998, le Comité de Bâle publiait un rapport intitulé « Renforcement de la transparence bancaire ». On pouvait y lire en conclusion : « Le Comité de Bâle considère que la transparence bancaire revêt la plus haute importance (…). À cette fin, banques et autorités de contrôle doivent s'assurer que les éléments appropriés sont rendus publics. ».

Ce groupe de travail a deux objectifs. Il s'agit d'abord de répertorier les données publiques relatives au secteur bancaire et financier. Le second objectif est de faire le point sur l'accessibilité aux données de bilan des banques, eu égard aux besoins, et en particulier de clairement définir les conditions d'accès des chercheurs, des journalistes, des associations et entreprises.

Les données relatives aux personnes physiques qui ont recours au secteur bancaire et à leurs opérations avec celui-ci n'entrent pas dans le champ de réflexion du groupe.

S'agissant du premier objectif, le groupe ne cherchera pas à faire un inventaire exhaustif des informations statistiques portant sur le secteur financier français, mais il identifiera les principales bases de données existantes et en dressera une première typologie dans la perspective de mettre à disposition un répertoire documenté des principales sources utilisables. Le champ pour ce premier objectif couvrira les banques, les assurances, les autres intermédiaires financiers et les marchés. Les sources prises en compte pourront être des données agrégées ou des données détaillées, voire des données propres à chacun des établissements du champ. Les conditions d'accès à ces sources seront précisées.

Le second objectif, restreint au champ des banques, visera spécifiquement l'accès aux données de bilan et d'exploitation de chaque banque. Le groupe de travail examinera notamment les conditions juridiques d'accès à ces données. Il les rapprochera des pratiques en vigueur en la matière à l'étranger (principalement aux États-Unis et en Allemagne).

Les autorités de supervision (Banque de France et Autorité de contrôle prudentiel et de résolution), les producteurs de données (Fédération bancaire française) et les utilisateurs (chercheurs, journalistes, ONG) seront associés aux travaux du groupe. Ce dernier proposera des mesures susceptibles d'améliorer l'accès aux données, sans porter atteinte aux exigences de confidentialité.

Sommaire

Avant-propos **3**
Rapport général **11**
1. Faciliter l'accès aux données bancaires et financières : une mission de service public **12**
1.1 Pourquoi favoriser l'accès aux données bancaires et financières ? 12
1.2 Le champ d'étude du rapport 15
1.3 Le plan du rapport 16
2. Les données macro-financières **17**
3. Les données micro-financières non-confidentielles **18**
3.1 Un gisement de données insuffisamment exploité 18
3.2 Les autorités de supervision ont un rôle important à jouer 20
3.3 Une grande diversité des procédures 21
3.4 La charte du G8 : des données publiques librement accessibles, gratuites et réutilisables 24
3.5 Aux États-Unis, la création d'une agence des données bancaires et financières 27
3.6 Au Royaume Uni, un plan stratégique d'ouverture aux données bancaires et financières 29
3.7 En Europe : une nécessaire harmonisation 30
4. Les données micro-financières confidentielles **32**
4.1 Diffusion des données après instruction des demandes 32
4.2 Accès restreint sur site 36
4.3 Programme de chercheurs invités 38
4.4 Centre d'accès sécurisé à distance (CASD) 39
4.5 Une transparence des procédures est indispensable 40
5. Quels sont les obstacles à lever ? **41**
5.1 Les contraintes juridiques : secret statistique et secret professionnel 41
5.2 Les contraintes budgétaires : une question de priorités 44
6. Recommandations : Passer d'une culture du secret à une culture de l'accès aux données **45**
6.1 Rendre public ce qui peut l'être 46
6.2 Une nécessaire transparence des procédures d'accès aux données confidentielles 50
Compléments **52**
A. Les dispositifs d'accès aux données bancaires et financières en France **54**
B. Résultats de l'enquête sur l'accès aux données financières pour les chercheurs **86**
C. Bref historique des recommandations du Cnis sur l'accès aux données bancaires **88**
D. Charte du G8 pour l'ouverture des données publiques (extraits) **91**
E. Les bases de données commerciales **95**
F. Cas spécifiques d'accès aux données **100**
G. Annexes **131**

Rapport général

1. Faciliter l'accès aux données bancaires et financières : une mission de service public

1.1 Pourquoi favoriser l'accès aux données bancaires et financières ?

Une très forte demande sociale de transparence du secteur bancaire et financier

La crise financière et économique a montré à quel point une bonne connaissance du secteur bancaire et financier, de ses pratiques et de ses acteurs, est indispensable. Dans le même temps, la crise a fait apparaître de larges zones d'ombre sur l'activité des intermédiaires financiers et des marchés. **Le secteur financier est indispensable au bon fonctionnement de l'économie ; sa maîtrise et donc sa compréhension le sont tout autant. Et ceci passe en grande partie par la production, la centralisation et la diffusion de données sur son activité et son fonctionnement.**

L'objectif est bien de favoriser l'accès aux données pour faciliter les recherches, les études, les analyses et améliorer ainsi la transparence du secteur bancaire et financier. Et comme le Comité de Bâle le soulignait déjà en 1998, « *la transparence constitue un élément clé d'un système bancaire sûr, sain et contrôlé de manière efficace* ». Cette demande de transparence répond à une demande sociale accrue par la crise. La défiance du public (déposants, épargnants, consommateurs de services financiers ou citoyens) vis-à-vis des banques et du secteur financier est en effet très importante.[2] La qualité des débats sur les politiques publiques, notamment la régulation financière, dépend en grande partie de la disponibilité de statistiques et de données pertinentes et fiables.

Favoriser l'accès aux données bancaires et financières, c'est :

- **pour les chercheurs, la possibilité d'examiner précisément les questions relatives aux financements de l'économie et à l'instabilité financière ;**
- **pour les citoyens, journalistes, associations, etc., un enjeu démocratique ;**
- **pour les consommateurs de services bancaires et financiers, l'opportunité de faire jouer la concurrence, *via* la création de plateformes d'informations financières innovantes, permettant d'optimiser les décisions en matière de crédit et d'épargne ;**
- **pour les établissements bancaires et financiers eux-mêmes, une condition indispensable à la pratique de la discipline de marché.**

[2] « La confiance des Français envers les banques n'a jamais été aussi basse », par A. Michel, Le Monde, 11 avril 2012. Cet article s'appuie sur l'enquête annuelle Deloitte sur les relations banques/clients.

Faciliter l'accès aux données bancaires et financières s'inscrit dans une tendance de fond qui s'exprime dans de nombreux domaines et partout dans le monde. En 2013, les pays du G8 ont ainsi publié une Charte pour l'ouverture des données publiques : « ***Les citoyens exigent aujourd'hui de pouvoir accéder à l'information et aux services par voie électronique au moment et selon les modalités qui répondent à leurs attentes. C'est de plus en plus souvent le cas pour les données publiques elles-mêmes. Nous arrivons à un tournant qui marque le début d'une nouvelle ère au cours de laquelle il sera possible de se servir des données ouvertes pour acquérir des connaissances, générer des idées et produire des services afin de créer un monde meilleur pour tous*** ». Ce mouvement est également bien ancré en France, où l'État s'est lancé en 2011 dans une politique ambitieuse d'ouverture et de partage des données publiques, reconnaissant explicitement que la diffusion de données relève d'une mission de service public. Depuis lors, la mission Etalab a mis en place le portail data.gouv.fr qui vise à rassembler et à mettre à disposition librement l'ensemble des informations publiques de l'État. Au printemps 2015, en France, le Parlement a voté l'ouverture des données de santé qui seront regroupées dans un système national des données de santé. Et il est également question d'ouvrir les données de transport. La France est ainsi au 4ème rang de l'*Open Data Barometer* qui classe les pays en fonction de leur degré d'ouverture aux données publiques.[3]

Faciliter l'accès aux données : une tendance lourde

Pour le moment, le secteur bancaire et financier est resté relativement en marge de ce phénomène d'*open data*, particulièrement en France. Des initiatives intéressantes existent toutefois à l'étranger. Ainsi, depuis janvier 2015, le FMI met ses données économiques en ligne à la disposition de tous gratuitement. Plus intéressant encore, aux États-Unis, les données sur les états financiers des banques (les *call reports*) sont librement disponibles, sans coût, dans un format facilement exploitable. La loi Dodd Frank aux États-Unis, votée en réponse à la crise, a aussi permis la mise en place d'une agence indépendante, l'*Office of Financial Research*, chargée de centraliser l'information financière et de faciliter son accès. La Banque d'Angleterre a récemment fait de l'ouverture aux données le principal axe de son plan stratégique. L'Allemagne a elle adopté depuis plusieurs années un programme d'accès sécurisé aux données bancaires. Ces exemples, et d'autres, sont détaillés dans le rapport.

De toute évidence, les intermédiaires financiers sont tenus de préserver la confidentialité de certaines données relatives à leur clientèle ou dont la divulgation pourrait porter atteinte à leur compétitivité. Il en va de même pour les autorités de contrôle chargées de collecter ces données. Pour ces raisons d'ailleurs, la loi française et les règles européennes restreignent la diffusion de ces données.

[3] http://barometer.opendataresearch.org/report/analysis/rankings.html

Etablir un équilibre entre transparence et secret statistique

Les règles de confidentialité ou de secret professionnel ne doivent pas empêcher tout accès aux données individuelles sur le système bancaire et financier, ni tenir systématiquement lieu de prétexte. Toutes les données ne relèvent pas de la vie privée des personnes, et ne portent pas atteinte au secret des affaires. Qui plus est, des procédures existent pour l'exploitation des données confidentielles (avis du Comité du secret statistique, centre d'accès sécurisé, anonymisation des données, etc.), comme le montrent les nombreuses initiatives menées dans d'autres secteurs ou d'autres pays.

L'objectif de ce rapport est de favoriser l'accès aux données relatives au secteur bancaire et financier tout en respectant scrupuleusement les règles du secret statistique et du secret professionnel. Il s'agit bien de rechercher un équilibre entre ces deux impératifs, équilibre qui peut différer selon la nature des données et le public souhaitant y avoir accès.

La logique est, au fond, la même que pour les données de santé, par exemple. Les contraintes en termes de confidentialité ne sont, en effet, pas moins importantes que pour le secteur bancaire et financier. Et pourtant, cette ouverture aux données de santé est considérée quasi-unanimement comme une opportunité majeure pour la recherche en santé publique, ainsi que pour développer une meilleure information du public.

Pour ce qui est de la recherche, la commission Système financier et financement de l'économie du Cnis a mené une enquête en 2013 sur l'accessibilité des données auprès d'une centaine de chercheurs (universitaires ou chercheurs du CNRS en poste en France) en économie bancaire et financière (voir complément B). Cette enquête révèle que parmi les répondants, environ 9 sur 10 considèrent « indispensable » l'utilisation de données détaillées sur les banques pour leurs recherches ; 7 sur 10 estiment que l'accès à ces données est « difficile » ou « très difficile », particulièrement pour les données confidentielles (9 sur 10) ; 9 sur 10 évaluent l'accès aux données bancaires confidentielles plus restrictif en France que dans les autres grands pays de l'OCDE. Il existe, d'ailleurs, un décalage saisissant avec l'étranger. On dénombre en effet plusieurs milliers d'articles académiques sur le secteur bancaire américain qui s'appuient sur des données publiques, quelques dizaines sur le secteur bancaire allemand, contre une poignée à peine d'études réalisées spécifiquement sur le secteur bancaire en France. En la matière, encourager la recherche consiste à rendre accessibles les données relatives au secteur bancaire et financier à partir de fichiers bruts, centralisés, et diffusés dans un format facilement exploitable.

L'accès aux données : une opportunité majeure pour la recherche

L'accès aux données : un enjeu stratégique

Les difficultés d'accès aux données sont un handicap pour la recherche en France. Mais c'est surtout un handicap pour comprendre notre propre système bancaire et financier ; or il en va de la stabilité financière et de la croissance économique. Il y a d'ailleurs aujourd'hui un très large consensus sur ce point. Citons par exemple, Christian Noyer, Gouverneur de la Banque de France : « *À une époque où l'innovation financière connaît un développement rapide, la recherche analytique fondée sur les données individuelles est particulièrement utile pour aider les banquiers centraux à relever les défis du maintien de la stabilité des prix et de la stabilité financière* ».[4]

L'accès aux données bancaires et financières ne concerne, ainsi, pas seulement les chercheurs, mais plus largement tous ceux qui s'intéressent aux développements de la sphère financière : journalistes, enseignants, associations, ONG, responsables politiques, etc. Cela concerne aussi le secteur bancaire et financier lui-même, dans le cadre de la discipline de marché, ainsi que les régulateurs. Bien sûr, les conditions d'accès ne sauraient être les mêmes pour chaque catégorie d'utilisateurs. Et ce rapport a notamment pour objectif de clarifier ce point.

Ce rapport propose treize recommandations pour faciliter l'accès aux données bancaires et financières, sans augmenter la charge des déclarations pour les répondants, sans compromettre la confidentialité des données recueillies par les autorités de supervision, et sans augmenter de manière significative les coûts de collecte. Ces recommandations visent essentiellement à promouvoir une véritable culture de l'accès aux données et une meilleure gouvernance des bases de données existantes.

1.2 Le champ d'étude du rapport

Le champ des données bancaires et financières est potentiellement très large. Le domaine couvre en effet les données relatives aux établissements de crédits (états financiers, ratios prudentiels, structures organisationnelles, flux financiers avec l'étranger, etc.), aux autres intermédiaires financiers (assureurs, fonds de placement, *shadow banking*, etc.), aux marchés (marchés boursiers, marchés dérivés, marchés de la dette, marchés de matières premières et de commodités, etc.), et au financement de l'économie (crédit bancaire aux entreprises et aux ménages, patrimoine des ménages, capital-risque, etc.).

Au-delà des spécificités propres à chacun de ces domaines, il apparaît un besoin commun d'ouverture et de transparence. Ce rapport propose donc une réflexion générale sur les modalités d'accès aux données bancaires et financières.

[4] « Innovation financière, politique monétaire et stabilité financière », Avril 2007 : http://inbdf/instit/discours/discours_2007.htm

L'approche retenue dans ce rapport ne se veut pas exhaustive. Deux limites sont à préciser :

i) **Ce rapport ne vise pas la production de nouvelles données, mais cherche à tirer le meilleur parti de l'exploitation des bases de données existantes.** Les problématiques spécifiques de production de données relatives au système financier et au financement de l'économie sont abordées régulièrement par la commission éponyme du Cnis.

ii) **Ce rapport ne traite pas de l'accès aux données collectées par les entreprises du secteur bancaire et financier sur leurs clients.** Ces données à caractère personnel[5] sont pour la plupart soumises au secret bancaire (article L.511-33 du Code monétaire et financier) – précisons toutefois que la loi permet aux personnes physiques d'accéder aux données les concernant (article 39 de la Loi informatique et libertés 78-17 du 6 janvier 1978 modifiée).

Ce rapport traite essentiellement des données sur le secteur bancaire et financier collectées par les autorités de supervision. Celles-ci sont en effet au cœur de la collecte et de la production de données bancaire et financière. Parmi ces données, il convient de distinguer :

- les données macroéconomiques qui sont des données ou des statistiques agrégées ;
- les données microéconomiques qui sont des données individuelles. Précisons, pour éviter toute confusion, qu'il ne s'agit pas là forcément de données sur les personnes physiques. Les « individus » sont ici des « unités statistiques ». Dans le cadre de ce rapport, il s'agit essentiellement de données sur les banques, les intermédiaires et les marchés financiers.

Dans ce rapport, nous avons choisi de mettre l'accent sur l'accès aux données microéconomiques. En effet, les données macro-financières sont, de manière générale, facilement disponibles aujourd'hui. En revanche, les données microéconomiques relatives au secteur bancaire et financier sont, en France, difficilement accessibles. Cela tient à des problèmes de coût, de dispersion et de qualité de l'information et, dans certains cas, à des contraintes juridiques liées à la confidentialité et au secret professionnel.

1.3 Le plan du rapport

Le rapport général est organisé, outre cette première section introductive, en cinq autres sections.

- La section 2 présente, brièvement, les données macro-financières.

[5] « Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement. »

- La section 3 s'intéresse aux données micro-financières qui peuvent être rendues publiques sans porter atteinte à la vie privée ou au secret des affaires.
- La section 4 porte sur les données micro-financières pour lesquelles il y a véritablement des exigences en termes de confidentialité.
- La section 5 examine spécifiquement les obstacles juridiques et budgétaires qui restreignent l'accès aux données bancaires et financières en France.
- La section 6 regroupe les recommandations.

2. Les données macro-financières

Les données macro-financières regroupent essentiellement les données sur : i) les agrégats monétaires ; ii) les crédits, les placements et les dépôts ; iii) les taux d'intérêt, les taux de change, les indices boursiers ; iv) les investissements directs étrangers et de portefeuille de la balance des paiements ; v) l'actif ou le passif financier des différents secteurs institutionnels. Il s'agit dans tous les cas de données agrégées ou de statistiques.

La production et la diffusion de ces données agrégées relèvent pour l'essentiel des banques centrales et des autorités de supervision. Ainsi, en France, la partie financière des comptes nationaux, les statistiques monétaires et financières, ainsi que les données sur la balance des paiements sont du ressort de la Banque de France (Article L141-6) et de la Banque Centrale Européenne. La Banque mondiale, le Fonds monétaire international et l'OCDE disposent également de bases de données agrégées très complètes sur le secteur financier, qui couvrent un champ très vaste et permettent des comparaisons historiques et internationales.

En France, les principales sources de données agrégées sur le secteur bancaire et financier sont :

- La Banque de France qui a ouvert en 2011 un portail Internet, baptisé Webstat, dédié à la diffusion de données macroéconomiques. Ce portail met aujourd'hui à disposition du public plus de 20 000 séries issues de la Banque de France et de grands organismes internationaux partenaires (voir complément A3).
- L'ACPR publie également tous les ans les « Chiffres du marché français de la banque et de l'assurance »[6] qui rassemble dans, un fichier Excel, une quarantaine de tableaux regroupant les chiffres clés (nombre d'établissements, répartition, bilans et compte de résultat agrégés, etc.).
- L'Insee, qui regroupe les données financières sur son site dans un onglet spécifique « Finances – Monnaie – Bourse » et donc facilement identifiables (voir complément

[6] https://acpr.banque-france.fr/publications/rapports-annuels/chiffres-du-marche-francais-de-la-banque-et-de-lassurance.html

A2). L'Insee gère également, pour la France, le dispositif d'identifiant unique des intervenants sur les marchés financiers et attribue ainsi les *Legal entity identifier* (LEI) aux établissements, essentiellement financiers, qui en font la demande (voir complément F10).

Ces données s'inscrivent de plus en plus dans une logique d'*open data*. **Les données macro-financières sont aujourd'hui facilement disponibles en ligne, souvent gratuitement, avec un long historique.** Notons que le niveau d'agrégation de ces données macroéconomiques peut varier, celles-ci pouvant faire l'objet d'une déclinaison institutionnelle, sectorielle ou géographique (par type de clients, par type d'intermédiaires financiers, par régions, etc.).

L'accès aux données macroéconomiques ne soulève, en général, pas d'obstacles particuliers. Les difficultés d'accès portent bien davantage sur les données microéconomiques, y compris lorsque ces données ne posent pas de problème de confidentialité.

3. Les données micro-financières non-confidentielles

3.1 Un gisement de données insuffisamment exploité

Il existe un très large gisement de données microéconomiques sur le secteur bancaire et financier, et beaucoup de ces informations sont non-confidentielles. Malheureusement, elles sont très dispersées.

Ces données microéconomiques sont souvent qualifiées de données individuelles, par opposition aux données agrégées. Il ne s'agit toutefois pas forcément de données sur les personnes physiques ; l'unité statistique est le plus souvent, dans le cadre de ce rapport, la banque ou l'entreprise. Précisons aussi que par données non-confidentielles nous entendons les données qui ne portent atteinte ni à la vie privée des personnes, ni au secret des affaires. Il est question, en particulier, de ces nombreuses données qui sont rendues publiques par ailleurs ; celles répondent bien souvent à des exigences réglementaires.

- La plupart des intermédiaires financiers, comme les entreprises non financières d'ailleurs, sont tenus de publier régulièrement et dans un format approprié des informations concernant leur situation financière (bilan, compte de résultat, tableaux des flux de trésorerie, annexes). Il s'agit de données publiques disponibles dans leurs rapports annuels.

- Il existe également des obligations de publication spécifiques, du fait des particularités du secteur bancaire et financier – à la fois indispensable au bon fonctionnement de l'économie mais susceptible de conduire à de graves crises. Certaines sont confidentielles, mais beaucoup ne le sont pas. Et il ne s'agit pas seulement des données collectées par les autorités dans le cadre de leur mission de supervision prudentielle. C'est le cas, par exemple, de l'obligation qui est faite aux banques en

France de publier un certain nombre d'informations par juridiction, dans le cadre de la lutte contre les paradis fiscaux (voir complément F8).

- Il existe enfin d'autres données, collectées par des entités publiques ou privées, qui ne relèvent pas de la supervision. C'est le cas, par exemple, des données de marché (cours boursiers, volumes de transactions, etc.) ou des données liées aux activités de financement et aux faillites qui sont consignées par les Greffes des Tribunaux de Commerce (voir complément F3).

L'exploitation de ces informations à des fins d'études et de recherches suppose de pouvoir détenir ces données pour un large ensemble d'entités, dans un format comparable, et sur une longue période. Or, l'absence de standard de présentation rend les comparaisons et l'exploitation statistique difficiles sans un très long travail de reconstitution d'une base de données exploitable.

Des fournisseurs privés commercialisent des données sur les banques (Bureau van Dijk, etc.) ou les marchés financiers (Bloomberg, Thomson Reuters, etc.), à des tarifs toutefois élevés. La plupart de ces bases produites à usage commercial ont comme premier débouché le secteur financier lui-même, gros consommateur de données (voir complément E).

Le coût élevé de ces bases de données pour les utilisateurs limite considérablement leur exploitation, notamment à des fins de recherche. Or, les recherches menées sur le secteur bancaire (et notamment l'instabilité financière) sont d'utilité publique. On peut aussi considérer que **la diffusion des données bancaires et financières doit relever d'une mission de service public**. D'autant qu'une grande part de ces données est collectée par la Banque de France et l'ACPR dans le cadre du système de remise Surfi (*Système unifié de rapport financier* – voir complément A1), sans qu'il soit possible toutefois de distinguer les données confidentielles des données non-confidentielles.

Les données microéconomiques, mais non-confidentielles, sont essentielles pour analyser le fonctionnement du secteur bancaire et des marchés financiers ; elles sont toutefois difficilement accessibles pour les chercheurs, les associations, les journalistes, etc. en raison de leur coût élevé. Le problème n'est pas, ici, lié à la protection de la vie privée ou au secret des affaires, mais tient à la dispersion des données. La difficulté est de pouvoir disposer de ces données de manière centralisée, gratuitement et dans un format homogène et réutilisable.

Dans la mesure où les autorités publiques en charge du secteur bancaire et financier collectent et centralisent un grand nombre d'informations dans le cadre de leur mission de supervision, il est très regrettable qu'elles ne les mettent pas à la disposition du public. Ce rapport contient plusieurs recommandations qui pourraient contribuer à une large diffusion de données individuelles non-confidentielles à moindre coût pour les utilisateurs en s'appuyant sur la collecte déjà réalisée par les autorités de supervision.

3.2 Les autorités de supervision ont un rôle important à jouer

En octobre 2014, la Banque Centrale Européenne (BCE) a organisé la septième conférence sur les statistiques.[7] Mario Draghi, Président de la BCE, et Danièle Nouy, présidente du conseil de supervision de la BCE, s'y sont exprimés en personne.

> « *Le grand défi de la statistique dans les prochaines années n'est pas tant ''d'avoir beaucoup plus de chiffres'', mais peut être bien davantage la réconciliation des informations statistiques collectées*. (…) *Les défis sont multiples – juridiques, culturels, technologiques, ainsi qu'organisationnels – mais ils peuvent être surmontés.* »[8]
>
> Mario Draghi, Président de la BCE.

> « *Il existe déjà des informations accessibles au public qui devraient aider à améliorer la transparence. Néanmoins, actuellement, il n'est pas encore possible d'accéder à cette information d'une manière rapide et harmonisée. (…) Les données sont des informations et les informations existent pour être communiquées. (…)*
>
> *Je ne peux pas promettre que la BCE peut, une fois pour toutes, éliminer le risque d'une prochaine crise financière. Mais la BCE est équipée pour minimiser ce risque. Et les données statistiques jouent un rôle crucial ici. Rappelons que l'incapacité à mesurer et à analyser correctement les risques liés à l'activité bancaire a été l'une des raisons de la crise financière. Développer et communiquer des données statistiques précises et opportunes est essentiel pour éviter que cet échec ne se répète. Pour cette raison, nous partageons tous – individus et institutions impliquées dans le processus statistique bancaire, institutions remettantes, régulateurs, statisticiens et superviseurs – une responsabilité commune envers la société. Continuons à travailler à la construction de fondations plus solides pour l'avenir du système financier*. »[9]
>
> Danièle Nouy, présidente du Conseil de supervision de la BCE.

---

[7] 7th ECB Statistics Conference "*Towards the banking Union. Opportunities and challenges for statistics*", Frankfurt am Main, 15 octobre 2014.

[8] M. Draghi, "The big challenge for Statistics in the coming years is not only "many more numbers", but perhaps much more so, the reconciliation of statistical information collected. (…) The challenges are manifold – legal, cultural, technological, as well as organisational – but they can be overcome." 7th ECB Statistics Conference (*op. cit.*).

[9] D. Nouy, "There is hence already information publicly available that should help to improve transparency. Nevertheless, currently it is not yet possible to access this information in a quick and harmonised way. (…) Data is information and information exists to be communicated. (…)
I cannot promise that the ECB can once and for all eliminate the risk of another financial crisis. But ECB is equipped to minimise this risk. And statistics play a crucial role here. Remember that the inability to correctly measure and analyse the risks associated to banking activity was one of the reasons of the current financial crisis. Developing and communicating accurate and timely statistics is essential for avoiding the repetition of this failure in the future. For that reason, we all, persons and institutions involved in the banking statistical process, reporters, regulators, statisticians and supervisors, share a common responsibility towards society. Let's keep on working in the construction of a more solid ground for the future of the financial system." 7th ECB Statistics Conference (*op. cit.*).

La feuille de route est claire : il faut centraliser et harmoniser les données pour les rendre accessibles. Bien souvent, l'accès aux données bancaires et financières se heurte à des problèmes de dispersion de l'information. Les données existent et ne posent pas forcément de problème de confidentialité. Pourtant, **les données brutes sur le secteur bancaire et financier restent très difficilement accessibles en France et sont donc sous-utilisées.**

La collecte des données individuelles sur le secteur bancaire et financier est principalement du ressort des autorités de supervision, Banque de France, ACPR et AMF. De manière générale, celles-ci n'ont, pour l'heure, pas une politique d'ouverture des données individuelles – contrairement à ce qui peut se passer dans d'autres pays.

3.3 Une grande diversité des procédures

De nombreux pays mettent à disposition du public des données détaillées sur le secteur bancaire et financier. La diffusion de ces données incombe, la plupart du temps, à la banque centrale et, dans quelques cas, aux autorités de supervision ou à une agence indépendante spécifiquement désignée. Le degré d'ouverture varie toutefois largement selon les pays.

- Certains pays ont une procédure très ouverte, c'est le cas notamment aux États Unis, où les standards en termes de diffusion des données sont très élevés. De très nombreuses bases sur le secteur bancaire et financier sont ainsi accessibles[10], dont les « *reports of condition and income* » connus sous le nom de « *call reports* » (voir complément F1). Ces données trimestrielles recensent les états financiers individuels des banques, bilan et compte de résultat, ainsi que des informations détaillées sur la qualité de leurs actifs, les prêts non-performants, leurs ratios de liquidité et de solvabilité. Ces données sont harmonisées entre les établissements et les informations sont décomposées par type d'actifs, y compris domestique/étranger. Ces données sont disponibles depuis 1976 et directement téléchargeables sous différents formats, et sans aucune formalité, sur le site du *Federal Financial Institutions Examination Council* (https://cdr.ffiec.gov/public/). De manière générale, les États-Unis sont très en avance en ce qui concerne l'accès aux données – l'exemple américain fera d'ailleurs l'objet de la section 3.5.

- L'Espagne et le Portugal, ainsi que les Pays baltes, ont également une politique assez ouverte et offrent un accès libre aux données bancaires individuelles par exemple.

- D'autres pays accordent un accès plus restreint après instruction des dossiers, comme c'est le cas pour la Bundesbank (http://www.bundesbank.de/Navigation/EN/Bundesbank/Research/RDSC/Data_access/data_access.html) qui détaille toutefois sur son site la procédure d'accès aux données et édite même un guide de l'utilisateur.

[10] Pour une description détaillée de la procédure de collecte et de diffusion des données bancaires aux États-Unis, voir Mark D. Flood, Simon Kwan et Irina S. Leonova, Data for Microprudential Supervision of U.S. Banks, Handbook of Financial Data and Risk Information, *Cambridge University Press*, 2014.

- Plusieurs banques centrales fonctionnent par appel à projets. C'est le cas notamment de la Banque de Finlande, de la Banque Nationale de Pologne, de la Banque Nationale Tchèque ou de la Banque Centrale de Hongrie qui offrent un accès à toute l'information disponible, y compris aux données confidentielles (toujours dans le respect des procédures du secret statistique). C'est le cas aussi de la Banque Populaire de Chine qui a récemment lancé un appel à projet pour les chercheurs étrangers.

- La Banque d'Angleterre n'avait, jusqu'à récemment, pas de procédure formelle, mais permettait un accès au cas par cas à ses données confidentielles (dans le respect des procédures du secret statistique) aux chercheurs invités. Toutefois, depuis l'année dernière, elle s'est engagée dans un ambitieux programme de développement de la recherche, « *One Bank Research* », présenté dans la section 3.6 et qui promeut notamment les collaborations entre la Banque d'Angleterre et les chercheurs extérieurs et prévoit l'ouverture des données au public (http://www.bankofengland.co.uk/research/Pages/onebank/agenda.aspx).

  « *À travers l'initiative ''One Bank Research'', la Banque d'Angleterre adopte une approche intégrée pour accroître la quantité, la qualité et l'impact de la recherche dans tous les domaines couverts par la banque centrale.*

  Cette approche vise à améliorer la coordination et l'ouverture de nos travaux de recherche (…), à veiller à ce que la Banque fasse le meilleur usage de ses données, et à cultiver une communauté de recherche tant à l'intérieur qu'à l'extérieur de la Banque, en ouvrant davantage nos recherches et l'ensemble des données au public. »[11]

- Enfin, autre exemple, la Russie, dont la banque centrale rend public des informations, moyennant toutefois un abonnement payant à l'agence d'information financière Interfax. Ces données sur les banques russes sont très détaillées, avec notamment une décomposition par type de déposants et par type de prêts (aux ménages, aux entreprises et à l'État).

Le tableau 1 résume les conditions d'accès aux données bancaires individuelles par les banques centrales. En excluant la France, sur les 27 autres pays de l'UE et les États-Unis :

- 12 offrent un accès libre aux données bancaires individuelles, dont les États-Unis ;
- 9 offrent un accès restreint, dont l'Allemagne ;
- 7 n'offrent aucun accès ;

[11] "Through our One Bank Research initiative, the Bank of England is taking an integrated approach to increase the quantity, quality and impact of the Bank's research across all aspects of central banking.
This approach aims to improve the coordination and openness of our research across all policy areas, to ensure the Bank makes the best use of our data, and to cultivate an extensive research community both within and outside of the Bank by opening up more of our research and data sets to the public."

Tableau 1. Accès à des données bancaires individuelles dans l'UE et aux États-Unis

Pays	Banque centrale ou régulateur	Accès
Allemagne	Bundesbank	Accès restreint
Autriche	Oesterreichische Nationalbank (OeNB)	Accès restreint
Belgique	Nationale Bank van België/Banque Nationale de Belgique	Accès restreint
Bulgarie	Bulgarian National Bank	Accès libre*)
Chypre	Central Bank of Cyprus	Aucun accès
Croatie	Croatian National Bank	Accès libre*)
Danemark	Danmarks Nationalbank	Accès restreint
Espagne	Banco de España	Accès libre*)
Estonie	Finantsinspektsioon	Accès libre*)
Finlande	Finanssivalvonta	Accès libre*)
France	*Banque de France*	*Accès restreint*
Grèce	Bank of Greece	Accès restreint
Hongrie	Central Bank of Hungary	Accès libre*)
Irlande	Central Bank of Ireland	Accès restreint
Italie	Banca d´Italia	Aucun accès
Lettonie	Finanšu un kapit la tirgus komisija	Accès libre*)
Lituanie	Bank of Lithuania	Accès libre*)
Luxembourg	Commission de Surveillance du Secteur Financier	Accès restreint
Malte	Malta Financial Services Authority	Accès restreint
Pays-Bas	De Nederlandsche Bank	Aucun accès
Pologne	Narodowy Bank Polski	Aucun accès
Portugal	Banco de Portugal	Accès libre*)
Rep. Tchèque	Czech National Bank	Accès libre*)
Roumanie	National Bank of Romania	Accès restreint
Royaume-Uni	Bank of England	Accès restreint
Slovaquie	Národná banka Slovenska	Aucun accès
Slovénie	Banka Slovenije	Aucun accès
Suède	Sveriges Riksbank	Aucun accès
États-Unis	FFIEC (Réserve fédérale / OCC / FDIC)	Accès libre

*) Voir accès ci-dessous.

Bulgarie: www.bnb.bg/BankSupervision/BSCreditInstitution/BSCIFinansReports/BSCIFRBankingSystem/BS_201503_EN ; Croatie : www.hnb.hr/supervizija/eindex.html ; Estonie : www.fi.ee/index.php?id=1174 ; États-Unis : https://cdr.ffiec.gov/public/ ; Espagne : www.aebanca.es/en/EstadosFinancieros/index.htm (banques commerciales), www.cajasdeahorros.es/#/PAG|estados (caisses d'épargne) ; Finlande : www.finanssivalvonta.fi/en/Statistics/keyfinancialfigures/Pages/Keyfinancialfigures.aspx ; Hongrie : felugyelet.mnb.hu/en/left_menu/pszafen_publication/golden_books.html ; Lettonie : http://www.fktk.lv/en/statistics/credit-institutions/quarterly-reports.html ; Lituanie : www.lb.lt/stat_pub/statbrowser.aspx?group=8070&lang=en ; Portugal : www.bportugal.pt/pt-PT/Supervisao/Paginas/Contasdasinstituicoes.aspx ; République Tchèque : www.cnb.cz/en/supervision_financial_market/information_published_issuers/

La France, elle, est dans une situation assez singulière, comme nous le détaillerons dans la section suivante, dans la mesure où la Banque de France a depuis quelques années mis en place un accès restreint aux données (que celles-ci soient confidentielles ou non), sans en communiquer ouvertement la procédure.

Les difficultés d'accès aux données ont évidemment des conséquences sur les possibilités de recherche qui peuvent être menées. On dénombre ainsi :

- **plus de 2 000 articles académiques qui s'appuient sur les « *call reports* » américains**[12] **;**
- **près d'une centaine d'articles académiques qui utilisent les données bancaires individuelles collectées par la Bundesbank ;**
- **à peine cinq articles de recherche ont pu être réalisés à partir des données bancaires individuelles collectées par la Banque de France.**

Il est clair que les difficultés d'accès aux données handicapent la recherche française, relativement aux autres pays.

3.4 La charte du G8 : vers des données publiques librement accessibles, gratuites et réutilisables

L'accès aux données bancaires et financières doit s'inscrire dans une démarche globale. La France a d'ailleurs pris des engagements généraux en ce sens, dans le cadre de **la Charte du G8 pour l'ouverture des données publiques** (voir complément D). L'objectif est d'offrir un accès « *au plus grand nombre d'utilisateurs possible et pour le plus grand nombre de fins possible* ».

Ces engagements ne concernent pas uniquement la statistique publique[13], mais bien l'ensemble des données collectées par les autorités. C'est même le principe n°1 de la Charte sur l'**ouverture par défaut des données publiques :** « *l'expression « données publiques » doit être entendue dans son sens le plus large possible. Elle peut recouvrir les données produites par des organismes gouvernementaux nationaux, fédéraux, locaux ou internationaux, ou par le reste du secteur public.* »

Précisons également que la Charte vise en premier lieu les données brutes individuelles. En vertu du principe n°2, **les données doivent être disponibles en qualité et en quantité** : « *les données seront disponibles sous leur forme initiale non modifiée, et présenteront le meilleur degré de granularité possible* ».

[12] Estimation réalisée à partir de Google Scholar.
[13] En France, rappelons que dès 2011, la mission Etalab a lancé le chantier d'ouverture des données publiques.

En pratique, le principe n°3 exige que les données soient **accessibles et réutilisables par tous**. Cela doit se traduire par un accès :

- gratuit ;
- pratique (« *sans faire l'objet d'obstacles d'ordre bureaucratique ou administratif, comme des exigences d'enregistrement* (…) *qui peuvent dissuader les usagers d'y accéder* ») ;
- dans un (ou plusieurs) format(s) ouvert(s) et réutilisable(s).

Nous sommes aujourd'hui loin de cet idéal. En effet, rappelons-le, la plupart des bases de données existantes sont coûteuses car soumises à une logique de marché.[14] Or les autorités de supervision réalisent un large travail de collecte de ces données. En centralisant l'information et en rendant accessibles ces données publiques, dans un format facilement exploitable, les autorités réduiraient considérablement le coût d'accès aux bases pour les utilisateurs. Qui plus est, les données collectées par les superviseurs sont de meilleure qualité que celles issues des bases commerciales (voir complément E).

Cette démarche en faveur d'une meilleure mise à disposition des données publiques peut facilement s'appliquer à de nombreuses données bancaires et financières non-confidentielles, parmi lesquelles :

- *Les données sur les états financiers des banques*. Nous avons vu, qu'aux États-Unis, la procédure de collecte des données sur les établissements bancaires américains (*call reports*) sert à alimenter la base de données mise en ligne avec un accès libre et gratuit sur Internet. En France, la Banque de France et l'ACPR collectent leurs données de façon centralisée *via* le portail Surfi (voir complément A1). Mais elles ne publient, pour l'essentiel, que des données agrégées. La plupart des données ainsi collectées ne portent pas atteinte à la vie privée, ni au secret des affaires ; il n'y a pas donc foncièrement de raisons pour que ces données ne soient pas mises à disposition du public.
- *Les données sur les entreprises non-financières, leurs modes de financement, les équipes dirigeantes*. Ces données contenues dans la base FIBEN sont commercialisées par la Banque de France. Elles représentent aussi une source d'information précieuse

[14] L'information financière représente un marché de plusieurs dizaines de milliards de dollars. Bloomberg (États-Unis, chiffre d'affaires 2012 : $7,9 mds) et Thomson Reuters (Canada, chiffre d'affaires 2012 : $7,5 mds) représentent chacun environ un tiers de ce marché. D'autres entreprises, telles que FactSet Research Systems (États-Unis, chiffre d'affaires 2014 : $0,9 mds) ou SIX Financial Information (Suisse), fournissent également des données sur les marchés boursiers (cotation en temps réels, opérations sur titre, etc.). D'autres sont spécialisées sur les données financières d'entreprise : parmi les plus notables, S&P Capital IQ (États-Unis, chiffre d'affaires 2012 : environ $1 mds), Dow Jones and Company, Moody's Analytics, et Bureau van Dijk (Belgique) qui publient plusieurs bases de données très utilisées par les chercheurs en économie et qui portent sur les données financières sur les entreprises (Amadeus, Diane), les banques (Bankscope) et les assurances (Isis).

pour les chercheurs, pour lesquels toutefois aucun accès n'est actuellement prévu (voir complément F2).[15]

- *Les réponses aux enquêtes sur le crédit aux entreprises*. Les résultats de ces enquêtes menées auprès des entreprises françaises ne sont disponibles que sous forme très agrégée (voir complément F3).

- *Les données des greffes des Tribunaux de commerce.* Ces données représentent une source très importante de renseignements sur le financement de l'économie et les faillites d'entreprises. Ces données publiques sont toutefois éparses et demandent un long travail de retraitement qui pourrait être centralisé (voir complément F3).

- *Les données relatives à l'information règlementée*. Depuis 2009, les « mécanismes officiellement désignés » (OAM, *Officially Appointed Mechanism*) assurent au niveau de chaque État-membre de l'Union européenne, le stockage centralisé de l'information règlementée, en application de l'article 21 de la directive 2004/109/CE du Parlement européen (dite Directive Transparence). L'OAM pour la France est la DILA – Direction de l'Information Légale et Administrative. Elle garantit la transparence économique et financière par la publication au niveau national de l'ensemble des informations légales, économiques et financières relatives à la vie des entreprises et au milieu associatif (notamment grâce au Bulletin des annonces légales obligatoires ou au Bulletin officiel des annonces des marchés publics). À titre d'exemple, le site http://www.info-financiere.fr/ permet de consulter des informations réglementées des sociétés cotées.[16]

- *Les données sur les implantations à l'étranger*. Depuis 2014, les établissements financiers sont tenus de fournir publiquement un reporting « pays par pays » pour quelques éléments clés de leur activité. L'exploitation de ces données reste toutefois difficile du fait de leur dispersion et d'un manque d'homogénéité dans leur présentation (voir complément F9).

- *Les données sur les produits d'épargne-retraite supplémentaire*. Il existe en France de nombreux produits facultatifs, souscrits à titre individuel : i) les produits d'assurance retraite qui relèvent du Code de l'assurance (PERP, contrats « Madelin », etc.) ; ii) les produits d'épargne salariale régis par le Code du travail (PERCO) ; iii) les produits spécifiques réservés à des catégories particulières (COREM, Préfon, COREL,…). Bien que certains de ces produits soient substituables, ils sont soumis à des règles différentes en termes de transmission d'informations. Une information homogène, transparente et facile d'accès permettrait l'émergence de nouveaux intermédiaires, qui

[15] On peut souligner que cela va dans le sens du projet ESEF (*European Single Electronic Format*) qui, au nom de la directive transparence (novembre 2013) impose aux sociétés cotées de mettre en place un format électronique de reporting harmonisé pour les rapports financiers annuels au 1er janvier 2020.

[16] En octobre 2013, la Directive Transparence a été modifiée, entraînant la nécessité de mettre en place, au plus tard le 1er janvier 2018, un portail d'accès électronique au niveau européen (EEAP : *European Electronic Access Point*) qui faciliterait l'accès à l'information réglementée des sociétés cotées dans l'ensemble des État-membres.

pourraient apporter aux épargnants des conseils à bas coût, comme c'est le cas aux États-Unis (voir complément F7).

- *Les données sur les positions bancaires bilatérales*. Il s'agit là de données agrégées par pays : par exemple, les dépôts bancaires des résidents français en Suisse, des résidents allemands à Singapour, des prêts accordés par les banques de Hong Kong à des résidents américains, etc. Ces données sont collectées par la Banque des Règlements Internationaux (BRI) auprès des banques centrales nationales. Mais la BRI ne rend pas public ces données, alors qu'il s'agit de données agrégées qui ne présentent aucun problème de confidentialité. Notons qu'il y a une forte pression de la société civile internationale pour la communication de ces données par la BRI : voir la pétition http://act.one.org/sign/follow_the_money.

- *Les données de marché*. Un partenariat existe entre l'AMF et le CNRS (institut de recherche Eurofidai) pour la centralisation et l'harmonisation des données boursières, notamment des données haute-fréquence (voir compléments A6, E2 et F5). Les bases de données ainsi constituées mériteraient d'être mises en ligne avec un accès direct et gratuit. Elles sont notamment fondamentales pour apprécier certaines sources de risque systémique (voir complément F8).

De manière générale, pour les nouvelles données sur le secteur bancaire et financière qui pourraient être collectées, il sera nécessaire d'anticiper en amont la mise à disposition par défaut des données.

3.5 Aux États-Unis, la création d'une agence des données bancaires et financières

L'accès aux données bancaires et financières aux États-Unis ne s'arrête pas à la publication des « *call reports* ». Bien qu'ayant déjà les standards de transparence les plus élevés, les États-Unis continuent de progresser dans ce domaine. En particulier, la loi Dodd-Frank, votée en réponse à la crise financière américaine, instaure l'*Office of Financial Research* (OFR). Cette agence indépendante, créée en 2011, a explicitement pour but de « *d'offrir des données financières de qualité, des standards, et des analyses afin de promouvoir la stabilité financière* ».[17]

L'OFR a notamment pour rôle d'établir des standards en termes de construction de la base de données. À ce titre, il joue un rôle essentiel dans le développement du LEI (*Legal Entity Identifier*, voir aussi complément F10) aux États-Unis.

Parmi les récentes initiatives de l'OFR, citons également la création d'un inventaire général inter-agences des données bancaires et financière, l'*Interagency Data Inventory*. L'idée est de constituer un catalogue des données que les autorités américaines de supervision collectent

[17] "The *Office of Financial Research* (OFR) delivers high-quality financial data, standards, and analysis to promote financial stability."

auprès de l'industrie ou achètent auprès de fournisseurs. Cet inventaire permettrait d'identifier les éventuelles zones d'ombre dans l'accès aux données et, *in fine,* de combler certaines lacunes pour mieux comprendre les vulnérabilités du système financier. L'objectif affiché est aussi de faciliter la recherche de données bancaires et financières pour le public.

Les données sont classées par organisation et par type de données. Un bref descriptif est également joint. Il est toutefois regrettable que le répertoire ne renseigne pas le lien vers les données. En février 2015, ce répertoire recense plus de 400 bases de données différentes provenant de dix autorités de supervision différentes. Le paysage français en matière de supervision financière est certes moins morcelé qu'aux États-Unis, mais ce serait une avancée majeure que de disposer d'un tel répertoire. Au niveau européen, l'avantage d'avoir accès à une telle centralisation de l'information serait évidemment encore plus grand. La Banque Centrale Européenne ou l'Autorité bancaire européenne pourrait s'inspirer de ce dispositif (voir section suivante).

De manière générale, les États-Unis sont donc à la pointe pour ce qui relève de la mise à disposition des données – pas uniquement des données relatives au secteur bancaire et financier d'ailleurs. Rappelons qu'en 2009, le Président Obama a lancé, dès son premier jour à la Maison Blanche, l'« *Open Government Initiative* » qui vise à améliorer foncièrement l'accès aux données. Dans son mémorandum, il s'engage « *à créer un niveau sans précédent d'ouverture au sein du gouvernement. Nous allons travailler ensemble pour assurer la confiance du public et établir un système de transparence, la participation du public et la collaboration. L'ouverture permettra de renforcer notre démocratie et de promouvoir l'efficience et l'efficacité au sein du gouvernement* ».[18]

Aux États-Unis, cette initiative engage fermement les agences publiques à améliorer l'accès aux données. C'est dans cette perspective que s'inscrit d'ailleurs la création au sein de la Réserve fédérale du poste de Directeur général des données (*Chief data officer*), indépendant de la direction générale des études et des statistiques. Pour Micheline Casey, qui occupe ce poste depuis sa création en 2013, il s'agit « *d'aider l'agence [la Fed] et les économistes qui travaillent sur la politique monétaire et la régulation financière à obtenir un accès sans précédent au trésor que représentent les données monétaires et financières* ».[19] Cela représente une vraie révolution culturelle vis-à-vis des données, que l'on cherchait jusque-là davantage à protéger qu'à partager.

L'accès aux données constitue un réel enjeu démocratique, mais cela représente aussi un véritable marché outre-Atlantique où se développent de nouveaux métiers et de nouveaux services associés. C'est d'ailleurs explicitement un des enjeux soulevés par le G8 dans sa Charte pour l'ouverture des données publiques : « *Des données publiques librement*

[18] B. Obama, "My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government." https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/

[19] M. Casey "Nailing down those two foundational issues will help grant the agency and the economists who work on monetary policies and regulations unprecedented access to the treasure trove of financial and statistical data." Juin 2013, http://fedscoop.com/meet-federal-reserve-boards-new-chief-data-officer/

accessibles et gratuitement réutilisables peuvent être à la source de services et de produits innovants susceptibles d'aider les individus à faire face plus facilement à la vie moderne. Utilisées de la sorte, ces données amènent le secteur privé à innover, contribuant à la création de nouveaux marchés, de nouvelles entreprises et de nouveaux emplois. » Dans un article de la *Harvard Business Review*, Richard H. Thaler[20] et Will Tucker montrent combien l'accès aux données GPS notamment a permis le développement de nombreux services pour appareils mobiles, services aujourd'hui considérés comme indispensables.[21] Ils prennent également l'exemple, dans le domaine financier, de Brightscope qui permet de comparer facilement les coûts des plans d'épargne retraite (voir complément F6).

3.6 Au Royaume Uni, un plan stratégique d'ouverture aux données bancaires et financières

En 2014, la Banque d'Angleterre a lancé, dans le cadre de son Plan stratégique « *One Bank Research* », un ambitieux programme d'ouverture aux données avec comme but explicite d'apporter des solutions participatives à des questions clés de la recherche[22] : « *Cette initiative vise à transformer la façon dont la recherche est effectuée à la Banque [d'Angleterre], par l'amélioration de l'intégration de notre recherche dans nos différents domaines de responsabilités et par l'ouverture de notre programme de recherche aux contributions externes*. »[23]

Ce programme couvre de nombreux champs :

- la promotion du *big data* ; c'est ainsi tenu, en juillet 2014, une conférence organisée par la Banque d'Angleterre sur le thème « *Big data and central banks* »[24] ;
- l'ouverture des données publiques ; sont ainsi désormais disponibles de nombreuses bases de données très riches, comme les réponses individuelles aux enquêtes, ou données macroéconomiques sur très longue période (trois siècles) ;
- la mise en place de collaborations avec les chercheurs pour l'exploitation de données confidentielles.

La Banque d'Angleterre insiste ainsi sur l'aspect stratégique de l'ouverture aux données, dans une démarche qui se veut collaborative. La logique est véritablement celle du *crowdsourcing*,

[20] Richard H. Thaler est Professeur à l'Université de Chicago et l'un des experts les plus renommés en finance comportementale. Il a notamment joué un rôle pionnier dans ce mouvement d'ouverture aux données.
[21] R.H. Thaler et W. Tucker, Smarter Information, Smarter Consumers, *Harvard Business Review*, 2013.
[22] "The Bank announced in its Strategic Plan our commitment to opening up more of our data to the public in order to crowdsource answers to key research questions and support collaborative research with Bank staff." http://www.bankofengland.co.uk/research/Pages/onebank/datasets.aspx
[23] "This initiative aims to transform the way research is done at the Bank, by enhancing the integration of our research across different areas of our policy responsibilities and opening up our research agenda to external contributions." http://www.bankofengland.co.uk/research/Pages/onebank/conference.aspx
[24] http://www.bankofengland.co.uk/publications/Documents/quarterlybulletin/2015/q108.pdf.

à savoir l'appropriation et l'exploitation par le plus grand nombre, pour pouvoir répondre à des défis communs.

3.7 En Europe : une nécessaire harmonisation

Comme nous l'avons vu Mario Draghi, Président de la BCE, et Danièle Nouy, Présidente du Conseil de supervision, ont insisté sur le nécessaire changement de culture et l'importance qu'il y a à utiliser au mieux les données. Pour ce qui relève des aspects pratiques, ils ont surtout proposé de mettre en place un système harmonisé.

Ainsi, pour Mario Draghi, « *les avantages de l'intégration progressive des systèmes d'information existants dans un système européen harmonisé de l'information ne sauraient être surestimés* ».[25] De même, pour Danièle Nouy, « *il ne fait aucun doute que l'intégration au sein d'une seule institution, la BCE, des différents domaines liés au système bancaire de la zone euro ouvre la voie à l'amélioration du dispositif statistique concernant le système bancaire européen*. (…)

L'intégration, l'harmonisation et la normalisation sont des conditions nécessaires, mais pas suffisantes pour atteindre un degré de transparence pleinement satisfaisant pour le système bancaire. Nous devons également diffuser et communiquer correctement les données. En ce sens, la création d'un référentiel commun (''un hub européen'') pour les données accessibles au public pourrait être une tâche relativement simple avec un impact très important et positif.

En ce moment, le monde de la statistique est face à plusieurs défis et il y a encore beaucoup de choses à faire. L'Union bancaire doit servir de catalyseur pour obtenir les améliorations que nous espérons tous. »[26]

[25] M. Draghi, "The benefits of gradually integrating the existing information systems into a harmonised European information system cannot be overstated. First, it ensures consistency in the information received by different policy makers for different purposes. Second, ideally asking every piece of information only once will help limit the burden of reporting agents. Third, it allows the exploitation of synergies across information domains and permits further rationalising of the data production processes both at the reporting banks and at the relevant Authorities. Fourth, it enriches the basis for policy decision making, respecting the separation of different policy domains. 7th ECB Statistics Conference (*op. cit.*).
There are however several constraints to data integration and they will need to be overcome in a stepwise approach because of their sometimes strong historical and cultural nature. Discrepancies due to different accounting standards and legal barriers preventing data sharing among institutions and even within the same institution must be overcome with proper transition periods. Also, integrating flows of micro-data from many sources across the world and analysing them effectively in a timely fashion must be enabled through the broad adoption, at global level, of data standards such as the Global Legal Entity Identifer, which is now operational."

[26] D. Nouy, "There is no doubt that the integration in one institution, the ECB, of different domains related to the euro area banking system paves the way for enhancing the statistical function within the European banking system. (…)
Integration, harmonisation and standardisation are necessary conditions, although not sufficient for achieving a fully satisfactory degree of transparency for the banking system: We also need to properly disseminate and communicate the data. In that sense, creating a common repository ("European Hub") for publicly available data could be a relatively simple task with a very important and positive impact.

Lors de la conférence précédente, Andrea Enria, Président de l'Autorité bancaire européenne, avait également tenu des propos similaires : « *Les données sont un élément clé de [la stabilité financière], et cette ressource rare et chère mérite une large discussion. L'évaluation des risques – à l'échelle micro ou macro-prudentielle – nécessite une quantité importante de données de bonne qualité et fiables, avec une granularité différente et différents niveaux d'agrégation. En effet, le manque d'information limite la capacité des autorités à identifier les vulnérabilités au cœur de la crise financière. Le fait de n'avoir accès qu'à des données limitées sur les institutions et les marchés financiers est un premier problème. Mais les problèmes principaux sont surtout liés à l'utilisation inadéquate des ressources existantes et de l'information, entravée par la fragmentation et la non-harmonisation de certaines données macro- et micro-financières. Par conséquent, le défi est de trouver le juste équilibre entre la nécessité d'améliorer la quantité et la qualité des données financières sans surcharger les institutions financières avec les obligations de déclaration superflues. Dans cet article, je soutiens que la voie à suivre est de faire un bon usage des données statistiques que nous avons déjà ou qui sont en cours d'élaboration* ».[27]

Le constat est ainsi largement partagé. À terme – souhaitons le plus rapidement possible – **l'Union bancaire doit déboucher sur une centralisation et une harmonisation de la collecte et de la diffusion des données bancaires**.[28]

À noter qu'il existe déjà un cadre standard pour le *reporting* des états financiers des établissements de crédits en Europe (FINREP) sur la base des normes IFRS. Il existe également un cadre standard pour le *reporting* des exigences prudentielles (COREP). FINREP et COREP sont deux bons exemples d'harmonisation au niveau européen, même si, comme le souligne Andrea Enria, il y a encore des efforts à réaliser. Mentionnons également que, dans le champ des sociétés d'assurances, la directive Solvabilité II, qui entrera en vigueur au 1er janvier 2016, s'accompagne de la mise en place d'une réglementation et d'un *reporting* harmonisés à travers l'Europe.

L'enjeu d'une centralisation et d'une harmonisation des données bancaires et financières en Europe n'est pas simplement de pouvoir mener des comparaisons entre pays. Une des difficultés majeures de la supervision bancaire et financière tient aux aspects transnationaux, alors que le dispositif de supervision reste relativement fragmenté – c'est bien ce qui motive

---

In this moment, the statistical world is facing many challenges and there are still many things to do. The Banking Union shall serve as a catalyst for the very relevant improvements we all expect." 7th ECB Statistics Conference (*op. cit.*).

[27] A. Enria, "Data is a key component of these activities [the financial stability tasks], and this scarce and expensive resource merits wider discussion. Risk assessment – either micro- or macro-prudential – requires a significant amount of good quality and reliable data, with different granularity and at different levels of aggregation and analytical information. Indeed, information gaps limited the ability of authorities to identify the building-up of vulnerabilities at the core of the financial crisis. Limited data on specific institutions and markets was a first shortcoming. But most gaps were mainly linked to the inadequate use of existing resources and information, hindered by the fragmentation and non-harmonisation of certain macro- and micro-financial data across jurisdictions. Therefore, the challenge is to strike the right balance between the need to improve the quantity and quality of financial data and the reluctance to over-burden financial institutions with unnecessary reporting obligations. In this paper I argue that the way forward is to make good use of the statistics we already have or that are being developed (…)". 6th ECB Statistics Conference

[28] Voir également Christopher Gandrud et Mark Hallerberg, Supervisory Transparency in the European Banking Union, *Bruegel Policy Contribution*, 2014.

l'Union bancaire et ce qui doit, par conséquent, motiver l'élaboration d'une stratégie commune en termes de production et d'accès aux données.

Comme nous l'avons vu, la centralisation et l'harmonisation des données bancaires et financières aux États-Unis échoient à une agence indépendante, nouvellement créée. En Europe, on peut imaginer que cette mission revienne, dans un avenir proche, à la BCE ou l'Autorité bancaire européenne. À noter que, dans le cadre des *stress tests*, l'Autorité bancaire européenne a adopté – fait novateur en Europe – une attitude très ouverte en publiant un jeu important de données sur les banques participant à l'exercice. Cette transparence avait bien pour objectif de renforcer la confiance dans le secteur bancaire.

4. Les données micro-financières confidentielles

Parmi les données collectées par les autorités de tutelles des intermédiaires et des marchés financiers, certaines sont confidentielles et relèvent du secret statistique. Le secret statistique – défini par la loi de 1951 – vise à protéger, d'une part la vie privée, d'autre part les intérêts économiques des agents. Il s'applique à tous les renseignements individuels[29] collectés par la statistique publique, mais ses dispositions varient très largement selon les cas. **Il existe de nombreuses possibilités d'accès (contrôlé) aux données confidentielles. En France, généralement, les demandes d'accès aux données confidentielles transitent par le Comité du secret statistique. La mise à disposition de ces données peut ensuite s'appuyer sur l'utilisation d'un centre d'accès sécurisé à distance ou d'autres procédures de partage adaptées, accompagnées de contrats de confidentialité, de la sélection des données utiles voire de l'anonymisation des éléments directement identifiants. Des solutions existent donc. Elles sont régulièrement appliquées, en France, comme à l'étranger, pour permettre l'accès à des bases de données très riches en informations, notamment pour les besoins de la recherche.**

4.1 Diffusion des données après instruction des demandes

C'est la pratique la plus courante. En France, c'est le Comité du secret statistique qui instruit toutes les demandes concernant des données individuelles collectées par voie d'enquête statistique ou transmises au service statistique public.[30] « *Son principe général est d'apporter aux personnes qui fournissent à l'administration, ou aux personnes chargées d'une mission de service public, des informations utilisées pour l'établissement de statistiques publiques,*

[29] Le secret statistique s'applique également aux données agrégées dès lors qu'il serait possible i) d'isoler un individu à partir de l'ensemble de données, (ii) de relier entre eux des ensembles de données distincts concernant un même individu, (iii) de procéder par déduction pour obtenir de l'information sur un individu. *Cf.* avis du G29 de février 2014.

[30] La composition et le fonctionnement du comité du secret statistique sont précisés dans le décret n° 2009-318 du 20 mars 2009 relatif au Conseil national de l'information statistique et au comité du secret statistique.

l'assurance que ces informations ne seront pas utilisées d'une façon susceptible de leur porter tort. »

Les données collectées par les autorités en charge du secteur bancaire et financier ne rentrent pas dans le cadre de la statistique publique.[31]

La Banque de France a recours, depuis 2012, à une procédure spécifique. Celle-ci a été présentée dans le cadre de la commission « Système financier et financement de l'économie » du Cnis mais reste très confidentielle (voir complément A4). L'accès aux données se fait en adressant au préalable une demande à l'adresse de messagerie électronique suivante : DGS-DIMOS-acces-donnees-ut@banque-france.fr. Le formulaire de demande figure dans le complément A4 joint au rapport.

Cette procédure pose principalement deux problèmes :

- **La procédure est la même pour toutes les données, qu'elles soient confidentielles ou non**. Autant cela se justifie dans le premier cas, autant la procédure semble disproportionnée dans le second cas dans la mesure où les données dont il est question ne portent atteintes ni à la vie privée, ni au secret des affaires. Pourquoi donc chercher à en restreindre l'accès ?

- **La démarche à suivre est très peu documentée.** Seule une mention de quelques lignes figure sur la page d'accueil du portail Webstat (en bas et en très petits caractères). Jusqu'en mai 2015, l'adresse électronique de contact n'était renseignée que dans un compte-rendu d'une réunion de la commission « Système financier et financement de l'économie » du Cnis datant d'octobre 2013. Comme cela avait d'ailleurs déjà été souligné, la procédure d'accès mériterait davantage de publicité, avec notamment un onglet dédié sur le site de la Banque de France.

L'ACPR collecte également des données individuelles auprès des banques, des assurances et des autres intermédiaires financiers, dans le cadre de ses missions prudentielles. Ces données servent principalement à des fins de contrôles individuels et transversaux, mais sont aussi utilisées par les économistes de l'ACPR à des fins d'études et de recherche. Quelques chercheurs extérieurs ont également pu participer à l'analyse de ces données, dans le cadre de missions réalisées pour et à l'ACPR (voir complément A5).

Les données collectées par la Banque de France ou l'ACPR ne relèvent pas, *a priori*, du Comité du secret statistique. Pour autant, rien n'empêche à l'avenir de s'inspirer des procédures mises en œuvre dans ce cadre. La procédure du Comité du secret statistique a le mérite d'être très transparente et de bénéficier d'une longue expérience en matière d'accès aux données confidentielles.

[31] Article 7 ter de la Loi n° 51-711 du 7 juin 1951 – Modifié par la Loi n°2009-526 du 12 mai 2009 - Art. 19. « La formation plénière du comité du secret statistique est compétente pour émettre, après avis facultatif de l'administration ou de la personne morale ayant procédé à la collecte des données concernées, des recommandations relatives à l'accès pour des besoins de recherche scientifique ou pour la réalisation d'études économiques aux données individuelles transmises à l'Institut national de la statistique et des études économiques et aux services statistiques ministériels en application de l'article 7 bis de la présente loi. »

Sur le site du Cnis (voir ci-dessous), les demandeurs disposent facilement d'information détaillée sur la procédure, la composition du comité, le calendrier des réunions, les critères de sélection, etc. Il existe d'ailleurs un guide du secret statistique, directement téléchargeable.[32]

Pour obtenir une dérogation au secret statistique, le demandeur doit présenter un dossier à l'appui de sa demande. Celui-ci comprend l'objet détaillé de l'étude qui motive la demande de dérogation, la description précise des informations demandées et le nom des personnes qui auront accès à ces informations. Un guide qui détaille la procédure est d'ailleurs publié. Les critères eux-mêmes sont très clairs. Le Comité du secret statistique examine la demande à partir des critères suivants :

- le demandeur présente-t-il toutes les garanties pour la protection des données confidentielles qui pourraient lui être transmises ?
- la communication de ces informations ne risque-t-elle pas de porter atteinte aux règles de la concurrence que la loi sur le secret statistique a entendu protéger ?
- ces informations ne risquent-elles pas d'être utilisées à des fins de contrôle fiscal ou de répression économique, usages interdits par l'article 6 de la loi de 1951 ?
- l'objet de l'étude qui justifie la demande présente-t-il un intérêt suffisant ?
- les informations demandées ne sont-elles pas excessives par rapport à l'étude qui les justifie ?

[32] http://www.cnis.fr/files/content/sites/Cnis/files/Fichiers/comite_du_secret/COMITE_DU_SECRET_guide.PDF

Dispositif d'accès aux données confidentielles du Comité du secret statistique

Comité du secret statistique

5 juin 2015

Dernière réunion

Comité du secret statistique - 2015 - 3e réunion

Procédure actuelle de transmission

- Afin de fluidifier le travail du secrétariat du comité et de diminuer les délais, il est recommandé aux demandeurs d'envoyer au secrétariat le formulaire (disponible sur le site) dès que celui-ci aura été validé par le(s) service(s) producteur(s).
- Le formulaire complété sera adressé aux services producteurs concernés (voir le vademecum sur le site) pour recueillir leur avis.
- En ce qui concerne les données ménages, le demandeur aura vérifié auparavant auprès du réseau Quételet que les données présentes sur ce réseau ne sont pas suffisantes pour le projet mentionné.
- Le secrétariat vérifiera que le dossier est complet et que les services producteurs ont été consultés.
- Le secrétariat informera le demandeur de la date à laquelle sera rendu l'avis du comité et indiquera les démarches à suivre (engagements à signer, présence ou non à la session du comité…)
- En cas d'avis favorable, le secrétariat avertira le demandeur de la date probable d'accès aux données.

Le secret statistique permet d'apporter aux personnes qui fournissent des informations utilisées pour l'établissement de statistiques l'assurance que ces informations ne seront pas utilisées d'une façon susceptible de leur porter tort.

Il permet ainsi d'assurer :

- aux personnes physiques que la confidentialité des informations relatives à leur vie personnelle et familiale est garantie

- aux entreprises que le secret commercial sera respecté : les informations transmises ne seront pas accessibles à leurs concurrents.

Le comité du secret statistique, présentation et historique, article paru dans le **courrier des statistiques** de septembre 2009.

Le comité du secret statistique veille à la préservation de ces garanties. Les compétences du comité du secret statistique sont fixées par la **loi n° 51-711 du 7 juin 1951 modifiée** sur l'obligation, la coordination et le secret en matière de statistiques. Il est appelé à se prononcer sur toute question relative au secret en matière de statistiques, et donne son avis sur les demandes de communication de données individuelles collectées par voie d'enquête statistique ou transmises au service statistique public, à des fins d'établissement des statistiques.

Le comité du secret statistique est présidé par un conseiller d'État. La composition et le fonctionnement du comité du secret statistique sont précisés dans le **décret n° 2009-318 du 20 mars 2009** relatif au Conseil national de l'information statistique et au comité du secret statistique. Le comité comprend deux sections et une formation plénière, qui réunit les membres des deux sections.

La première section est compétente pour les renseignements individuels ayant trait à la vie personnelle et familiale et, d'une manière générale, aux faits et comportements d'ordre privé.

La seconde section est compétente pour les renseignements individuels d'ordre économique ou financier. Afin de répondre aux besoins d'informations ponctuelles le comité du secret peut proposer des dérogations individuelles ou collectives. Pour obtenir cette dérogation , le demandeur doit présenter un dossier à l'appui de sa demande. Celui-ci comprend l'objet détaillé de l'étude qui motive la demande de dérogation, la description précise des informations demandées et le nom des personnes qui auront accès à ces informations (**voir le guide du secret statistique**).

Source : http://www.cnis.fr/cms/Accueil/activites/_trois_comites/Comite_du_secret_statistique

Notons que le Comité du secret statistique n'est pas le seul organisme en France en charge d'instruire les demandes d'accès à des données confidentielles. C'est le cas également de la Commission nationale de l'informatique et des libertés (Cnil) qui défend notamment le droit d'accès des particuliers aux données qu'un organisme détient sur eux. C'est le cas aussi du Conseil scientifique du Comité des données du Réseau Quetelet (www.reseau-quetelet.cnrs.fr/), qui coordonne les activités d'archivage, de documentation et de diffusion des données en sciences humaines et sociales. On peut également citer l'Institut des données de santé (http://www.institut-des-donnees-de-sante.fr/)[33] qui se donne comme triple objectif de favoriser l'utilisation des bases de données, de favoriser la mise en commun des données et d'élaborer des standards pour l'accès aux données confidentielles (*cf*. encadré). Dans les deux cas, la procédure d'accès aux données, qui diffère selon le statut des utilisateurs, est tout à fait transparente.

Les missions de l'Institut des données de santé
http://www.institut-des-donnees-de-sante.fr/missions/78-les-missions.html

La recherche d'une meilleure gouvernance du système de santé repose sur une connaissance partagée de son fonctionnement et de son financement par tous les acteurs. L'Institut des données de santé a pour mission de réunir les conditions permettant cette connaissance. Sa mission consiste notamment à :

- favoriser l'utilisation des bases de données de santé par les membres de l'IDS et les organismes à des fins d'études et de recherches ;
- favoriser la mise en commun de données de santé provenant de sources différentes et le partage de ces données enrichies ;
- élaborer un cadre de confiance : respect du secret médical et des libertés individuelles, de l'éthique et de la déontologie pour l'utilisation des données de santé.

4.2 Accès restreint sur site

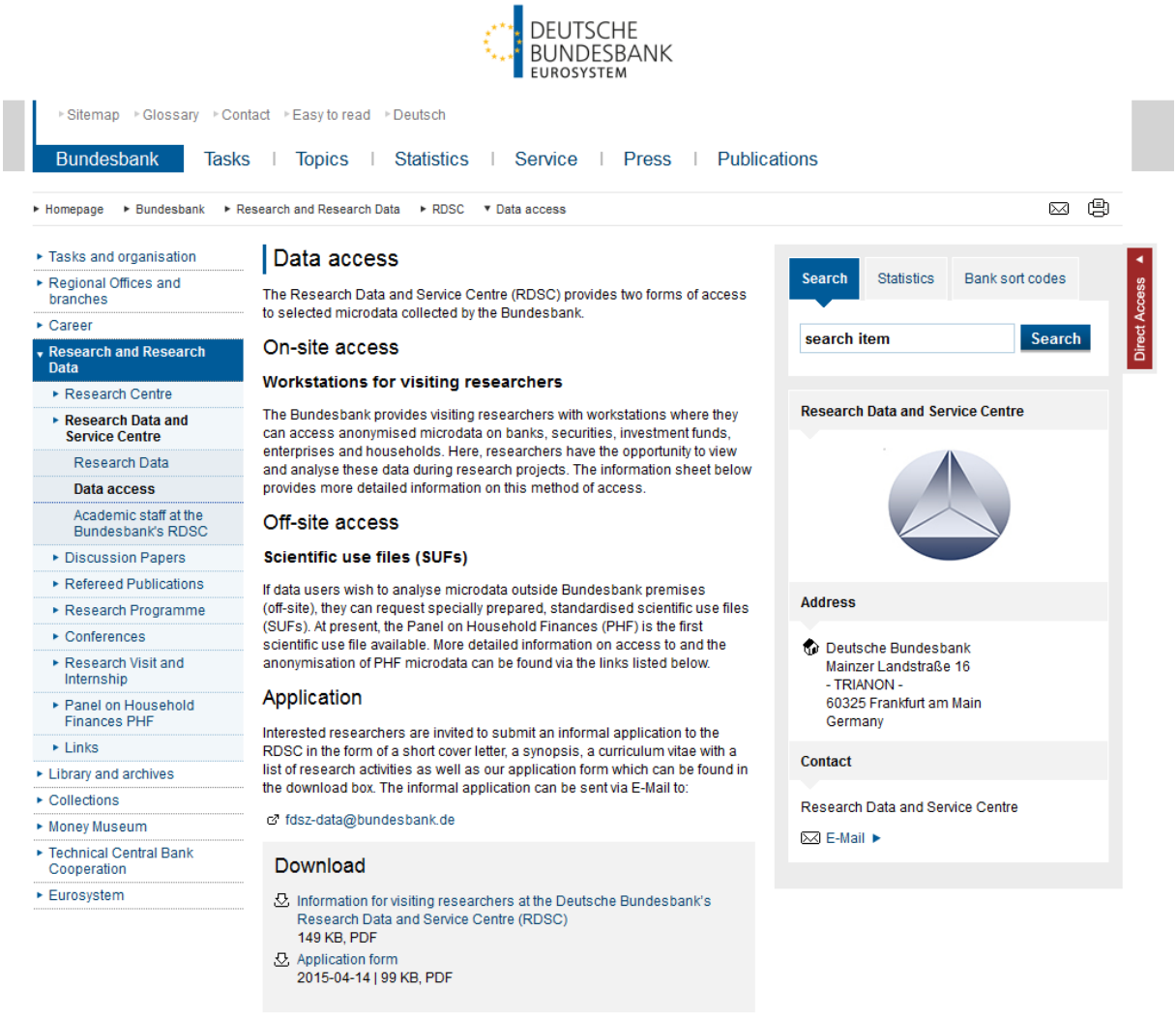
Pour les données confidentielles, il existe des possibilités d'accès sur site, généralement sur des stations de travail dédiées, ce qui permet un contrôle très strict.

C'est ce que propose notamment la Bundesbank. Le dispositif d'accès est remarquablement transparent, avec des possibilités d'accès sur site ou à distance selon la nature des données. Des centaines de travaux ont ainsi pu être menés grâce à ce dispositif, sans jamais porter atteinte au secret statistique. La procédure est accessible à tout chercheur, allemand ou étranger. Le dossier de demande d'accès est très facilement disponible, directement en ligne

[33] L'article 47 de la Loi de santé, prévoit que l'Institut des données de santé qui était en place depuis 2007, soit remplacé par un Institut national des données de santé (INDS).

(voir ci-dessous). Ce dispositif est au fond très similaire à celui du Comité du secret statistique.

Dispositif d'accès aux données confidentielles de la Bundesbank



Source : http://www.bundesbank.de/Navigation/EN/Bundesbank/Research/RDSC/Data_access/data_access.html

Certaines données sont ainsi accessibles à distance. C'est le cas, par exemple, du *Panel of Household Finance* (PHF) qui renseigne sur la situation financière des ménages (leurs revenus, leur patrimoine, leurs habitudes de consommation, les transferts intergénérationnels, etc.).[34] Pour accéder aux données, il faut demander une accréditation au *Research Centre of*

[34] Le PHF constitue le volet allemand de l'enquête harmonisé au sein de l'Eurosystème HFCS (*Household Finance and Consumption Survey*). Il s'agit d'un panel intégral, interrogeant les répondants sur le cycle de vie. L'échantillonnage est fait sur la base des données de recensement, avec des strates définies géographiquement et une surreprésentation des ménages riches (repérés *via* les statistiques de revenu au niveau local ou des données géographiques sur la qualité des logements). Plus de 3 500 ménages ont répondu à la première vague, en 2011.

the Deutsche Bundesbank. L'acte de candidature contient le CV des requérants, une description du projet de recherche et la signature d'une lettre d'engagement. Les données sont anonymisées (suppression des identifiants personnels, regroupements de secteurs NACE, arrondis sur les données continues, ...). La mise à disposition des données se fait par email, *via* des fichiers zippés et protégés par mot de passe.

D'autres données sont uniquement accessibles sur site, en accès local au *Research Data and Service Centre (RDSC)*. Les bases de données concernent les banques (bilans mensuels, position extérieure, données individuelles sur les taux d'intérêt bancaires, données prudentielles, lignes de crédit supérieures à 1,5 millions d'euros…), la détention d'obligations, les entreprises (investissement direct, commerce de services, bilans) et l'enquête PHF mentionnée précédemment.

L'acte de candidature contient le CV des requérants, une description du projet de recherche et la signature d'une lettre d'engagement. L'utilisation de données externes (pour un appariement) est autorisée, mais ces données doivent être contrôlées au préalable et la procédure d'appariement doit être validée. L'accès aux données se fait exclusivement à la Bundesbank. L'usage d'ordinateur personnel ou d'outils de stockage est prohibé. Le RDSC met à disposition les logiciels de traitement statistique Stat et SAS. L'utilisation d'autres programmes est autorisée au cas par cas. Les programmes et résultats sont soumis au staff de la Bundesbank, qui contrôle, entre autres, le respect des engagements de confidentialité.

Ce dispositif est comparable à celui mis en place par le *Census Bureau* et le *Système de Réserve Fédérale* aux États-Unis.

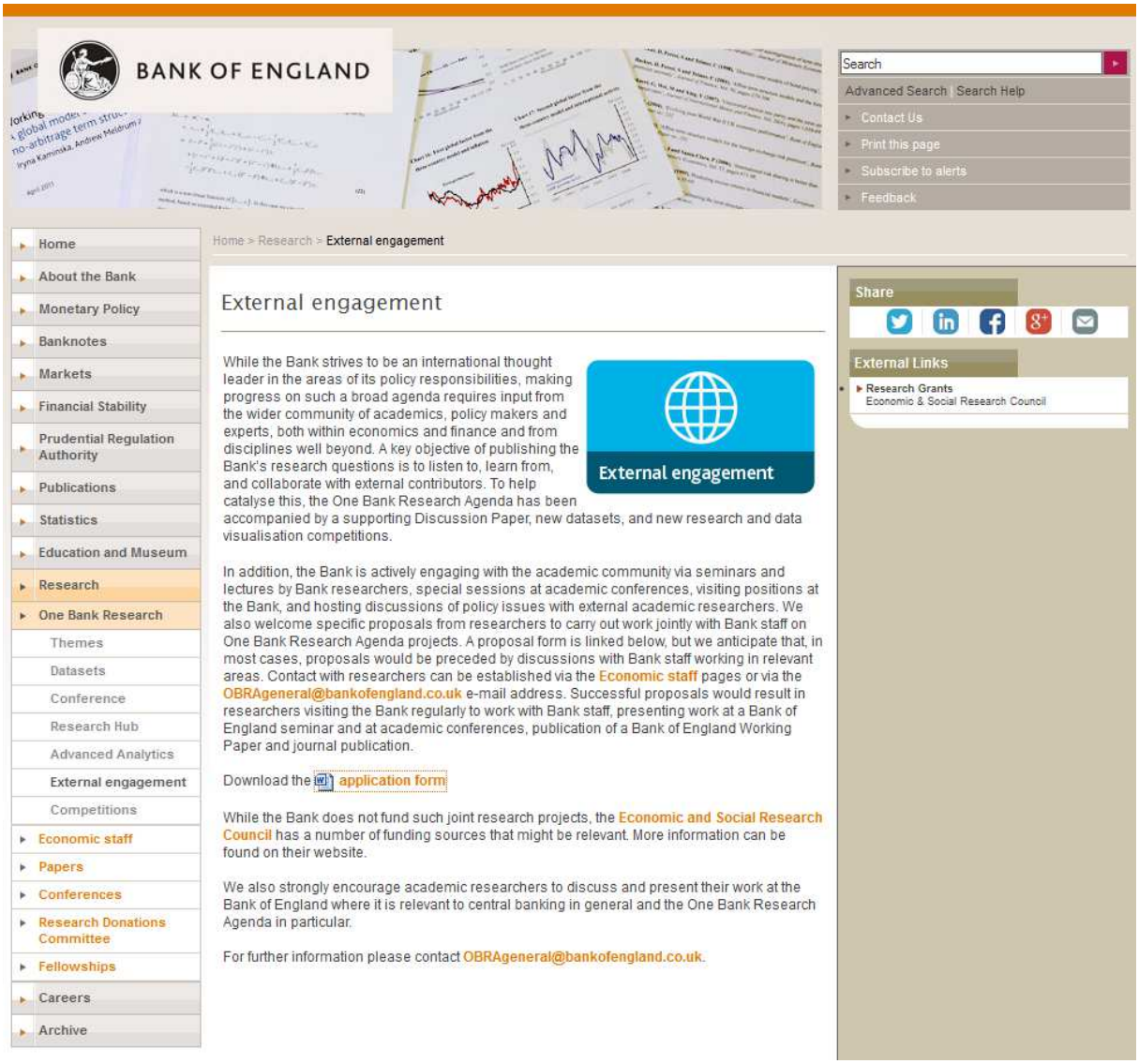
4.3 Programme de chercheurs invités

Dans certains pays, l'accès aux données confidentielles se fait *via* le programme plus général des chercheurs invités. C'est ainsi, par exemple, que procède la Banque d'Angleterre. Jusque récemment, la procédure n'était guère formalisée. Mais depuis le lancement du Plan stratégique « *One Bank Research* », la Banque d'Angleterre fait preuve d'une réelle volonté d'ouvrir ses données et de collaborer avec des chercheurs externes. Lors du lancement du programme, Andrew G. Haldane, l'économiste en chef de la Banque d'Angleterre, en appelle directement aux chercheurs extérieurs pour l'exploitation des données : « *La promesse de nouvelles données ... requiert votre aide* ».[35]

Comme dans le cas de la Bundesbank, ou de la procédure du Comité du secret statistique, la démarche pour les chercheurs extérieurs est parfaitement transparente.

[35] "The promise of new data ...will require your help" 25 février 2015. Discours à l'occasion de la conférence de lancement One Bank Research Agenda.

Dispositif d'accès aux données confidentielles de la Banque d'Angleterre



Source : http://www.bankofengland.co.uk/research/Pages/onebank/externaleng.aspx

En France, quelques chercheurs ont pu participer à l'analyse des données confidentielles de la Banque de France et à l'ACPR. Il n'existe toutefois aucune procédure formalisée, ce qui limite considérablement l'exploitation de ces données : seules quelques études empiriques existent sur les données bancaires et financières en France contre des dizaines chaque année pour l'Allemagne.

4.4 Centre d'accès sécurisé à distance (CASD)

Une solution technique innovante pour la diffusion de données confidentielles consiste à mettre en place un Centre d'accès sécurisé à distance (CASD). La France est vraiment à la

pointe dans le monde pour cette offre de services. En 2009, le Génés (Groupe des écoles nationales d'économie et statistique) a ainsi créé, avec l'Insee, un équipement conçu pour permettre aux chercheurs de travailler à distance, depuis leurs lieux de recherche, sur des données individuelles très détaillées, et donc soumises à la confidentialité, dans des conditions de sécurité élevées. Aujourd'hui, plus de 500 chercheurs participant à plus de 200 projets de recherche bénéficient de ce service (voir complément A7).

4.5 Une transparence des procédures est indispensable

Récemment, deux champs d'études ont bénéficié d'une ouverture des données confidentielles aux chercheurs : celui de la santé et celui de la fiscalité. Les données de santé comme les données fiscales relèvent de la sphère privée et sont éminemment confidentielles. **Une des recommandations de ce rapport vise, comme cela a été fait récemment dans de nombreux autres domaines, à améliorer et à formaliser la procédure permettant d'avoir accès, sous certaines conditions, aux données confidentielles relatives au secteur bancaire et financier.**

Là encore, soulignons que **la France est en retard dans ce domaine par rapport à de nombreux autres pays qui ont formalisé depuis plusieurs années leurs procédures d'accès, facilitant ainsi les recherches menées sur leur secteur bancaire et financier**.

La mise en place d'un onglet spécifique sur le site des régulateurs, la publication d'un guide détaillé sur les modalités du dispositif d'accès, l'énoncé de critères transparents, ou la nomination d'un référent unique pour l'accès aux données sont autant de pistes à suivre.

Concernant le référent, il convient que ce dernier soit indépendant du service en charge de la production de données. La présidence du Comité du secret statistique n'est, par exemple, jamais confiée à un représentant de l'Insee.[36] Sans aller jusqu'à imaginer que la présidence du Comité en charge de l'accès aux données bancaires individuelles échappe à un agent de la Banque de France et/ou de l'ACPR, il conviendrait que sa composition soit publique, qu'elle intègre des membres extérieurs, et qu'elle soit présidée par un agent indépendant des services en charge de la production des données.

[36] Le Comité du secret statistique est composé, outre le Président, d'un député, d'un sénateur, d'un membre du Conseil économique et social, d'un représentant de l'Insee, de représentants du Garde des Sceaux, de représentants du service interministériel des Archives de France (Siaf), d'un représentant du service producteur concerné, désigné par le ministre dont relève ce service, de représentants de la Commission nationale de l'informatique et des libertés (Cnil), de représentants parmi les organisations syndicales, d'un représentant de l'Institut national d'études démographiques (Ined), d'un membre du Conseil scientifique du comité de concertation pour les données en sciences humaines et sociales et d'un élu local.

5. Quels sont les obstacles à lever ?

Ce rapport prône une plus grande ouverture et un meilleur accès aux données sur le secteur bancaire et financier. Aujourd'hui, en France, de nombreuses données non-confidentielles ne sont disponibles que de façon dispersée, peu harmonisée ou à un coût élevé, tandis que les procédures pour l'accès aux données confidentielles manquent de transparence. Les sections précédentes montrent qu'il n'en est pas ainsi dans tous les pays. Qu'est ce qui, en France, empêche donc l'accès aux données bancaires et financières ?

Dans le cadre des discussions du groupe de travail, deux arguments sont revenus de façon récurrente : les contraintes juridiques et les contraintes budgétaires.

5.1 Les contraintes juridiques : secret statistique et secret professionnel

Plusieurs textes de loi, français et européens, viennent encadrer la diffusion de données bancaires et financières.

- Un ensemble de lois générales sur la diffusion des données :
    - La Loi n°51-711 du 7 juin 1951 (modifiée) sur le secret statistique ;
    - La Loi n°78-17 du 6 janvier 1978 (modifiée) relative à l'informatique, aux fichiers et aux libertés ;
    - La Loi n°78-753 du 17 juillet 1978 (modifiée) relative à l'accès aux documents administratifs et à la réutilisation des informations publiques (loi « Cada ») ;
- Les textes de lois qui régissent l'activité des régulateurs :
    - L'Article L141-6 du Code monétaire et financier qui spécifie les missions statistiques de la Banque de France ;
    - L'Article L612-17 du Code monétaire et financier qui mentionne l'activité de collecte des données par l'ACPR ;
    - L'Article 5 sur les statuts du Système Européen de Banques Centrales et de la BCE au sujet de la collecte d'informations statistiques ;
    - L'Articles 53 et suivants de la directive 2013/36/UE qui soumettent à une obligation de confidentialité les données et encadrent strictement l'utilisation et l'échange d'informations prudentielles que les organismes ont l'obligation de remettre aux autorités de surveillance prudentielle compétentes.

L'ensemble de ces textes visent à protéger l'accès aux données, mais mettent en avant deux notions distinctes : le secret statistique d'une part, le secret professionnel d'autre part.

Les trois premiers textes, qui régissent **le secret statistique**, visent à protéger tout particulièrement les libertés individuelles et le secret des affaires.[37] Le second ensemble de textes met en avant **le secret professionnel** qui incombe aux autorités chargées de la collecte des données sur le secteur bancaire et financier.[38]

Le secret statistique n'apparaît pas véritablement comme un obstacle à une meilleure diffusion des données sur le secteur bancaire et financier. Comme cela a été mentionné à maintes reprises, il existe un très large gisement de données non-confidentielles qui pourrait être mis à la disposition du public ; et s'agissant des données confidentielles, les dispositifs techniques existent pour garantir leur confidentialité.

Ce sont les contraintes imposées par le respect du secret professionnel qui ont suscité le plus de discussions au sein du groupe de travail. Les membres du groupe de travail se sont souvent opposés sur la lecture des textes juridiques, sans pouvoir dégager un avis qui fasse consensus. Deux logiques s'opposent : celles des autorités de supervision qui ont une interprétation stricte des dispositions relatives au secret professionnel et celle des utilisateurs pour qui l'ouverture des données ne viole pas nécessairement ces dispositions.

La Banque de France et l'ACPR mettent en avant que toutes les données collectées dans le cadre de leur mission de supervision sont soumises au secret professionnel. Le secret professionnel leur ferait interdiction de divulguer les données collectées. La Banque de France et l'ACPR en déduisent que même les données qui ne mettent en cause ni la vie privée des individus, ni le secret des affaires – y compris celles qui sont diffusées publiquement par ailleurs – ne peuvent être communiquées. Cette interprétation stricte du secret professionnel repose implicitement sur l'idée que ce n'est pas la nature de la donnée qui importe, mais le mode de collecte : **pour la Banque de France ou l'ACPR, toute donnée collectée dans le cadre de leurs missions de supervision deviendrait automatiquement confidentielle**.

Certes, comme le rappelle le complément A5 sur la collecte et la diffusion de données par l'ACPR, « *la violation du secret professionnel constitue un délit, puni des peines prévues à l'article 226-13 du Code pénal, soit un an d'emprisonnement et 15 000 euros d'amende* ». Mais l'article en question précise bien que cela ne concerne que « *la révélation d'une information à caractère secret* ».

Une grande partie de ce rapport porte justement, répétons-le, sur des données qui ne portent pas atteinte à la vie privée ni au secret des affaires et qui sont pour la plupart, publiées par

[37] Aux trois textes précités, on aurait également pu ajouter le Code du patrimoine (qui précise les délais de libre communicabilité des archives et les dérogations possible à des fins de recherche scientifique ou historique) ainsi que l'article 16 A de la loi DCRA du 12 avril 2000 (qui permet de ne pas opposer le secret professionnel en cas de données utiles à une seconde démarche administrative et partagées sous la maîtrise de l'individu concerné). Ces deux derniers textes relayent l'objectif de partager les données relevant de « patrimoines ou de besoins communs », cet objectif étant d'ailleurs susceptible d'être consacré de façon plus impérative et élargie pour l'ensemble des données utiles aux « travaux de recherche scientifiques », qu'ils poursuivent des finalités analytiques, statistiques ou historiques.

[38] Conformément au mandat du groupe de travail, les données relatives aux personnes physiques qui ont recours au secteur bancaire et à leurs opérations avec celui-ci n'entrent pas dans le champ de réflexion du groupe. Il n'est donc pas ici question du secret bancaire qui s'impose aux dirigeants et aux employés des établissements de crédits (article L.511-33 du Code monétaire et financier).

ailleurs, quoique de façon dispersée et dans un format non-réutilisable. Les sections précédentes de ce rapport font état d'un très large gisement de données non-confidentielles sur le secteur bancaire et financier. **La divulgation des données non-confidentielles par les autorités de supervision ne relève pas du secret professionnel.**

Non seulement il n'y a pas d'interdiction spécifique à la divulgation de données non-confidentielles, mais plusieurs textes de loi favorisent explicitement l'ouverture à ces données.

Pour ce qui est de l'ACPR en particulier, l'Article L.612-1 du Code monétaire et financier stipule qu'elle peut « *porter à la connaissance du public toute information qu'elle estime nécessaire à l'accomplissement de ses missions, sans que lui soit opposable le secret professionnel mentionné à l'article L.612-17* ». Rappelons que la mission de l'ACPR est de « *veille[r] à la préservation de la stabilité du système financier et à la protection des clients, assurés, adhérents et bénéficiaires des personnes soumises à son contrôle* ». Cela laisse donc une grande latitude à l'ACPR dans la mesure où la diffusion de données participe directement à la fois à la stabilité financière et à la protection des clients de services bancaires et financiers.

De manière plus générale, **la loi française autorise et encourage l'accès aux documents administratifs** par la Loi n°78-753 du 17 juillet 1978 (modifiée). Le premier chapitre de cette loi considère comme documents administratifs, « *quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission*. » Sont donc *a priori* concernées les données collectées par les autorités de supervision du secteur bancaire et financier – à moins de considérer que leurs missions ne relèvent pas d'une mission de service public. Cette loi précise évidemment certaines restrictions, notamment pour les informations « *dont la consultation ou la communication porterait atteinte (…) à la monnaie et au crédit public (…), à la protection de la vie privée (…) et au secret en matière commerciale et industrielle* ». Pour toutes les autres données, il revient aux autorités de supervision de divulguer les informations collectées.

Cette loi créée, par ailleurs, la Commission d'accès aux documents administratifs (Cada), une autorité administrative indépendante qui a pour objectif de faciliter et contrôler l'accès aux documents administratifs. La Cada veille non seulement au respect de la liberté d'accès, mais aussi aux conditions de réutilisation. Notons en effet que, depuis 2005, « ***la loi prévoit la possibilité de réutiliser les informations publiques à d'autres fins que celles pour lesquelles elles sont détenues ou élaborées*** ».

Il n'apparaît donc pas nécessaire, a priori, de modifier la loi pour permettre aujourd'hui un meilleur accès aux données bancaires et financières détenues par les organismes en charge d'une mission de service public (principalement Banque de France, ACPR et AMF) : il conviendrait peut-être, plus simplement, d'affirmer le périmètre des données bancaires et financières relevant de l'accessibilité organisée par la Loi « Cada ».

L'accès aux données de santé et aux données de l'administration fiscale a nécessité l'introduction de nouveaux textes de loi (respectivement Article 47 du projet de loi de santé et Décret n° 2014-950 du 21 août 2014 relatif à l'accès aux informations fiscales à des fins de recherche scientifique). Mais il s'agissait là de données à caractère personnel. Les données dont il est question dans ce rapport sont dans leur grande majorité des données non-confidentielles, ou alors déjà soumises à des procédures garantissant le respect du secret statistique.

Au final, compte tenu des nombreux textes qui encadrent la diffusion de données bancaires et financières, le groupe de travail s'est accordé sur l'idée de solliciter la Cada afin d'obtenir des éclaircissements sur ces aspects juridiques.

5.2 Les contraintes budgétaires : une question de priorités

La production, la collecte et l'harmonisation des données ont un coût. La mise à disposition de ces données également. Le groupe de travail a bien conscience de l'effort financier que cela représente, sans toutefois avoir pu le chiffrer. Plusieurs remarques sont néanmoins nécessaires :

- Le groupe de travail a, d'emblée, circonscrit son analyse aux bases de données existantes, pour limiter l'impact sur les coûts.

- Les coûts additionnels liés à la diffusion des données ne sont pas négligeables mais sont relativement faibles comparés aux investissements nécessaires pour produire ces données ; or la diffusion et l'exploitation des données sont bien des objectifs qui justifient les dépenses consenties dès les étapes de collecte et de production.

- Les coûts de diffusion des données sont à considérer globalement ; aujourd'hui, il en coûte aux utilisateurs de devoir centraliser eux-mêmes certaines données ou de devoir faire appel à des bases de données commerciales aux tarifs particulièrement élevés.

- Le coût de diffusion des données est évidemment à mettre au regard des bénéfices attendus qui pourraient se mesurer, d'après certaines études, à plus d'une centaine de milliards d'euros par an.[39] Plus encore que les coûts, les bénéfices sont particulièrement difficiles à quantifier mais il est très probable que ces derniers l'emportent très largement.

[39] Une étude de McKinsey (2013) évalue entre 210 et 280 milliards le gain annuel, au niveau mondial, qui résulterait d'une plus grande ouverture des données bancaires et financières (et jusqu'à 3 000 milliards en cas de généralisation de l'*open data* à tous les domaines). http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

6. Recommandations : Passer d'une culture du secret à une culture de l'accès aux données

Au vu des éléments présentés ci-dessus, l'accès aux données bancaires individuelles pourrait être largement facilité en France, sans augmenter la charge des déclarations pour les répondants, sans compromettre la confidentialité des données recueillies par les autorités de supervision, et sans augmenter les coûts de collecte. Cela vaut pour les données non-confidentielles, comme pour les données confidentielles. Un éventail de solutions existe en effet pour favoriser l'accès aux données du secteur bancaire et financier. Ces solutions sont plus ou moins contraignantes et plus ou moins coûteuses, ce qui permet de s'adapter à chaque type de données et à chaque type de public.

L'essentiel de nos recommandations s'adresse aux autorités de régulation du secteur bancaire et financier qui ont un rôle clé à jouer pour améliorer l'accès aux données. Ce rapport s'inscrit dans le prolongement des recommandations déjà formulées par le Comité de Bâle, « *à l'adresse des autorités chargées du contrôle et de la réglementation bancaires, pour la définition et l'amélioration des cadres réglementaires de communication financière et de notification prudentielle (…) pour les éléments clés qui devraient être portés à la connaissance du public*. (…) *Le Comité de Bâle considère que la transparence bancaire revêt la plus haute importance. (…) À cette fin, banques et autorités de contrôle doivent s'assurer que les éléments appropriés sont rendus publics*. »

Rappelons que favoriser la diffusion des données relève d'une mission de service public et que plusieurs pays, dont les États-Unis, se sont résolument engagés dans cette voie. La France s'est elle-même engagée à respecter la Charte du G8 sur l'ouverture des données publiques.

Une meilleure diffusion des données bancaires et financières suppose un changement de paradigme : il convient de passer d'une culture du secret à une culture de l'accès aux données dans le domaine bancaire et financier.

En s'inspirant de ce qui se fait dans d'autres domaines ou dans d'autres pays, ce rapport propose donc treize recommandations classées en deux groupes :

- la première série de recommandations vise à promouvoir l'accès par défaut aux données non-confidentielles ;
- la seconde série de recommandations propose de formaliser les procédures d'accès aux données confidentielles.

Ces propositions sont formulées au nom du groupe de travail, même si elles n'ont pas toujours suscité l'unanimité de ses membres.

6.1 Rendre public ce qui peut l'être

Pour un grand nombre de données, la question de la confidentialité ne se pose pas, puisqu'il s'agit de données qui ne mettent en cause ni la vie privée des individus, ni le secret des affaires et qui pour la plupart sont diffusées publiquement par ailleurs. En revanche, ces données sont éparses et les autorités publiques ont un rôle à jouer en centralisant et en harmonisant ces informations (ce qu'elles font déjà) et en les mettant à disposition des différents utilisateurs. Ce faisant, les autorités réduiraient considérablement le coût d'accès aux bases, aujourd'hui élevé ; il y a là un important gisement de données inexploitées car trop dispersées. Il s'agit là, au fond, d'appliquer le principe d'ouverture par défaut des données publiques reconnu dans la Charte du G8.

- Les états financiers des banques représentent une source d'information très riche et indispensable à la bonne connaissance du secteur, que ce soit pour les chercheurs, les analystes, les journalistes, ou les consommateurs. Il est dommage que ces données, pourtant non-confidentielles, ne soient disponibles que de façon dispersée et/ou à un coût élevé, alors même que la Banque de France centralise ces données. Les « call reports » qui recensent les états financiers des banques aux États-Unis sont un très bon exemple en la matière.

  Recommandation n°1. Le groupe de travail recommande que la Banque de France mette librement à la disposition du public, dans un format uniformisé et facilement exploitable (principe n°3 de la Charte du G8 sur l'ouverture des données publiques), les données publiées par les banques sur leur situation financière, à l'instar de ce qui existe aux États-Unis avec les « call reports ».

- La Banque de France recense un très grand nombre d'informations sur les entreprises françaises : documents comptables, informations sur les dirigeants, indicateurs de risque de défaut, etc. Ces données FIBEN sont commercialisées par la Banque de France auprès d'adhérents (établissements de crédit, des sociétés d'assurance-crédit et d'assurance-caution). Toutes ces informations ne sont pas forcément confidentielles : certaines d'entre elles devraient être mises à la disposition du public ; d'autres, pourraient être accessibles sous conditions, pour les besoins de la recherche notamment. Notons que cela rejoint une des recommandations du rapport du Conseil d'Analyse Economique (note n°18 – décembre 2014) : « *cette information [contenue dans les fichiers FIBEN] pourrait soit être vendue à tous les acteurs économiques, soit être mise dans le domaine public au titre de l'open data* »

Recommandation n°2. Le groupe de travail recommande que certaines données FIBEN sur les entreprises françaises, commercialisées par la Banque de France, puissent être plus largement accessibles, sous réserve des règles de confidentialité qui s'imposent.

- La Banque de France réalise plusieurs enquêtes auprès des ménages et des entreprises françaises, notamment sur les conditions d'accès aux crédits. Le résultat de ces enquêtes est publié de façon agrégée, mais les données individuelles restent inaccessibles, contrairement aux enquêtes de même type réalisées par la Réserve fédérale américaine ou la Banque d'Angleterre, par exemple.

Recommandation n°3. Le groupe de travail recommande que les données individuelles (anonymes) associées aux enquêtes menées par la Banque de France sur l'accès au crédit soient mises à la disposition du public, comme cela se fait dans de nombreux pays sans contrevenir aux règles de confidentialité.

- Les données sur les incidents de paiement et le risque de faillite des entreprises collectées par les tribunaux de commerce sont des données publiques particulièrement riches en information. En ce sens, les greffes de ces tribunaux font un travail essentiel et facilitent l'accès aux données, dans les limites des règles de confidentialité qui s'imposent. Ces informations sont toutefois sous-valorisées car elles sont fournies dans un format qui limite considérablement les possibilités d'étude systématique. Le groupe de travail a conscience des coûts qu'impliquerait ce changement de format. Ces coûts peuvent cependant être contrebalancés par i) une gestion facilitée des dossiers pour les greffes eux-mêmes ; ii) une mise à disposition payante pour certains acteurs, à l'instar de ce que pratique aujourd'hui la *Companies house* au Royaume-Uni.

Recommandation n°4. Le groupe de travail recommande qu'une réflexion s'engage avec les greffes des tribunaux de commerce pour que les données qu'ils stockent sur les procédures collectives puissent être mises à la disposition du public, de manière centralisée et informatisée, dans un format facilement exploitable, comme c'est le cas par exemple des Companies House au Royaume-Uni.

- Le reporting « pays par pays » représente une avancée majeure vers davantage de transparence en permettant d'avoir une image plus précise des activités des établissements financiers dans les paradis fiscaux. À noter que la France a joué un moteur dans cette initiative au niveau européen. Les données individuelles sont mises à disposition du public gratuitement, ce qui a déjà permis à la société civile de s'emparer de ce sujet. Néanmoins, des limites subsistent : i) les formats utilisés par les banques pour publier ces données rendent leur exploitation difficile ; ii) les données ne sont pas publiées dans un format directement exploitable ; iii) la typologie utilisée pour qualifier les activités des banques n'est pas harmonisée ; iv) la liste complète des

banques soumises à cette obligation est difficile à établir ; v) les paramètres de consolidation ne sont pas clairement définis et laissés à discrétion des répondants.

Recommandation n°5. Le groupe de travail recommande que les reportings « pays par pays » des banques soient établies selon des formats uniformisés et facilement exploitables (principe n°3 de la Charte du G8 sur l'ouverture des données publiques). Il convient également d'établir une typologie des activités commune pour l'ensemble des répondants et de clarifier les paramètres de consolidation.

- La Banque des Règlements Internationaux collecte auprès des banques centrales nationales des données bilatérales sur les positions bancaires agrégées au niveau de chaque pays. Ces données ne sont pas diffusées publiquement par la BRI. Elles sont pourtant fondamentales pour comprendre le rôle joué par les places financières *offshore* dans l'activité bancaire internationale, pour surveiller la stabilité financière, ou pour mesurer les progrès faits dans la lutte contre les paradis fiscaux. Il y a d'ailleurs une forte demande sociale pour l'ouverture de ces données. Et il n'y a, *a priori*, aucune raison de ne pas les rendre publiques, puisqu'il s'agit de données agrégées qui ne violent pas les lois sur le secret bancaire.

  Recommandation n°6. Le groupe de travail recommande que les données bilatérales sur les positions bancaires agrégées, qui sont collectées par les banques centrales pour le compte de la BRI, soient rendues publiques.

- Il existe en France une grande hétérogénéité de la réglementation s'appliquant aux produits d'épargne-retraite supplémentaire, notamment en matière de transparence et d'information (plus élevées sur le PERP, que pour les « Madelin », « Article 28 » ou PERCO). Cette situation ne favorise pas un fonctionnement efficace du marché de l'épargne-retraite. En particulier, il est essentiel que les épargnants aient accès aux données relatives à la performance et au montant des frais.

  Recommandation n°7. Le groupe de travail recommande que soit engagée une uniformisation des règles de transparence des produits d'épargne-retraite en s'alignant sur le fonctionnement des PERP/GERP. Ces données devraient être mises à la disposition du public, avec une centralisation et une homogénéisation de l'information vers l'ACPR (assurance retraite) et l'AMF ou le Ministère du Travail (épargne salariale), et sous la forme d'un prospectus uniformisé, à l'image de celui existant pour les OPCVM.

- La plupart des bases de données de marché (sur les transactions boursières, les produits dérivés, les événements de marché, etc.) sont le fait d'entreprises privées de services et d'informations financières internationales ; elles fournissent notamment aux établissements financiers (banques, *brokers*, *hedge funds*, etc.) des flux de données « en temps réel ». Le CNRS, *via* l'institut de recherche Eurofidai, a construit plusieurs bases de données historiques sur les marchés boursiers et les fonds de

placement collectifs. Il a également le projet, en collaboration avec l'AMF, de développer une base de données sur les transactions à très haute fréquence.

Recommandation n°8. Le groupe de travail recommande que le CNRS facilite l'accès à ces bases de données historiques (journalières et infra-journalières) sur les actions, les indices et les fonds de placement collectifs.

- Les données bancaires et financières représentent un ensemble vaste et varié : les données portent sur un large éventail de produits, avec une fréquence qui peut aller de la nanoseconde à l'année, et un grand nombre d'acteurs ; ces données sont collectées indépendamment par différentes entités, financières ou non financières, publiques ou privées. Il serait donc utile de disposer d'un catalogue des données détenues par les autorités publiques[40], que celles-ci soient achetées auprès de prestataires externes, ou collectées dans le cadre de leur mission de supervision.

  Recommandation n°9. Le groupe de travail recommande que soit mis en place un large inventaire des bases de données sur le secteur bancaire et financier, semblable par exemple à celui qui existe aux États-Unis (Interagency Data Inventory).

  Par ailleurs, à l'heure où se met en place l'Union bancaire, il est particulièrement regrettable que l'on ne dispose pas de données centralisées sur le secteur bancaire et financier à l'échelle européenne. La France et l'Allemagne pourraient jouer un rôle moteur sur ce sujet.

  Le groupe de travail soutient aussi vivement l'initiative d'une plate-forme commune aux différentes autorités de supervision de l'Union européenne pour la diffusion de bases de données harmonisées et la création d'un répertoire collectif. Cette mission pourrait notamment être confiée à la Banque centrale européenne, ou à une agence indépendante, à l'instar de l'Office of Financial Research aux États-Unis.

- Le *Global LEI System* (GLEIS) est un dispositif mondial d'identification unique des intervenants sur les marchés financiers mis en place par le *Financial Stability Board* à la demande du G20. L'objectif est, à terme, de constituer une cartographie des groupes et des liaisons entre les entités qui soit librement accessible à tous.

  Recommandation n°10. Le groupe de travail réitère son vif soutien à l'initiative LEI. L'initiative est déjà bien engagée. Il serait souhaitable, comme initialement prévu, de compléter les données de référencement des entités par des données sur les relations intra- et inter-unités juridiques, ainsi que d'étendre, par le biais des décisions des autorités de supervision, le périmètre des unités concernées par l'identification.

[40] L'article 17 de la loi du 17 juillet 1978 prévoit que « les administrations qui produisent ou détiennent des informations publiques tiennent à la disposition des usagers un répertoire des principaux documents dans lesquels ces informations figurent ».

6.2 Une nécessaire transparence des procédures d'accès aux données confidentielles

Pour ce qui est des données confidentielles, il ne s'agit évidemment pas de contrevenir au secret statistique. Cela ne signifie pas, pour autant, qu'il faille systématiquement en interdire l'accès, quel que soit le public et l'utilisation qui est faite de ces informations. Il existe, comme le démontre la pratique dans d'autres pays ou dans d'autres domaines, de nombreuses procédures qui permettent l'utilisation de ces données pour des études ou des recherches, tout en en contrôlant strictement l'accès. Il conviendrait donc que les autorités de supervision (essentiellement Banque de France, ACPR et AMF) formalisent et communiquent leurs procédures d'accès.

- La Banque de France s'est engagée depuis plusieurs années à favoriser l'accès des chercheurs aux données confidentielles. Il est regrettable toutefois que la procédure mise en place reste largement ignorée. Il serait ainsi souhaitable que la Banque de France prévoit un onglet dédié sur son site internet et édite un guide d'accès, comme le font plusieurs banques centrales étrangères ou comme le fait l'Insee dans le cadre du Comité du secret statistique. Il serait enfin utile que les travaux de recherche engagés à partir de ces données soient référencés au sein d'un répertoire unique pour une meilleure circulation de l'information.

  Recommandation n°11. Le groupe de travail recommande que la Banque de France formalise sa procédure d'accès aux données confidentielles et qu'elle diffuse sur son site un guide d'accès, comme le font par exemple la Bundesbank et la Banque d'Angleterre.

- L'Autorité de contrôle prudentiel et de résolution (ACPR) collecte des données individuelles dans le cadre de sa mission de supervision prudentielle. L'Autorité des marchés financiers (AMF) collecte aussi quotidiennement des données de marché et en centralise un grand nombre. L'exploitation de ces données confidentielles représente un enjeu majeur pour la recherche comme pour la conduite de la politique économique. Il est dommage que ces données ne puissent pas être davantage exploitées à des fins de recherche, alors que des dispositifs existent (contrat de confidentialité, centre d'accès sécurisé à distance, données anonymisées, etc.) pour que soient respectés le secret statistique et le secret professionnel.

Recommandation n°12. Le groupe de travail recommande que l'ACPR et l'AMF développent et formalisent leur programme de chercheurs invités pour permettre une plus large exploitation des données confidentielles dont elles disposent, sans contrevenir aux règles du secret statistique et professionnel.

Le groupe de travail suggère également que l'ACPR, l'AMF et la Banque de France réfléchissent à la possibilité de recourir à des outils d'accès à distance, du type CASD du Génés.

- Le principal problème concernant l'accès aux données bancaires et financières individuelles est d'ordre pratique, voire culturel. Pour faciliter les démarches des utilisateurs et promouvoir une culture de l'accès aux données, il est souhaitable que les superviseurs désignent un référent et mettent en place un Comité en charge du suivi des demandes d'accès, sur le même modèle que le Comité du secret statistique.

  Recommandation n°13. Le groupe de travail recommande que la Banque de France et l'ACPR désignent un référent pour l'accès aux données bancaires et financières à des fins d'études et de recherches. Ce référent, ou Chief Data Officer, aurait notamment pour rôle de présider le Comité en charge d'instruire les demandes d'accès aux données bancaires et financières ; la composition de ce comité serait publique et ferait place à des membres extérieurs représentant les différentes parties prenantes.

Compléments

Compléments 52
A. Les dispositifs d'accès aux données bancaires et financières en France 54
1. Présentation d'e-Surfi 54
2. Les données financières disponibles à l'Insee 56
3. Les données macro-financières de la Banque de France : le portail Webstat 57
4. Le dispositif de la Banque de France pour l'accès aux données individuelles 58
5. Les données prudentielles collectées par l'ACPR dans le cadre de ses missions 70
6. L'accès aux données de marché par l'AMF 78
7. Le Centre d'accès sécurisé à distance (CASD) 81
B. Résultats de l'enquête sur l'accès aux données bancaires financières pour les chercheurs. 86
C. Bref historique des recommandations du Cnis sur l'accès aux données bancaires 88
D. Charte du G8 pour l'ouverture des données publiques (extraits) 91
E. Les bases de données commerciales 95
1. Bankscope 95
2. Eurofidai 98
F. Cas spécifiques d'accès aux données 100
1. Les données bancaires individuelles consolidées aux États-Unis : les « Call reports » 100
2. Accès aux données FIBEN – Extrait de la note du conseil d'analyse économique n°18 103
3. Les données sur les crédits aux entreprises et les incidents de paiements 104
4. Les données sur les crédits aux particuliers et le surendettement 109
5. Les données de marché 111
6. Les données sur les fonds de placement 114
7. Epargne retraite supplémentaire en France : quelles données, quelle transparence ? 117
8. Les données nécessaires au calcul du risque systémique 124
9. Le reporting pays par pays des établissements financiers 126
10. L'initiative LEI (*Legal entity identifier*) : enjeux et mise en œuvre 129
G. Annexes 131
Composition du groupe de travail 131
Auditions 131
Remerciements 131

A. Les dispositifs d'accès aux données bancaires et financières en France

1. Présentation d'e-Surfi

Renaud Lacroix, Direction Générale des Statistiques, Banque de France

Depuis 2010, les établissements bancaires effectuent leurs remises d'états prudentiels, comptables et financiers au SGACPR ainsi que l'essentiel de leurs déclarations statistiques à la Banque de France au travers d'un outil unifié : le portail de collecte OneGate et le « Système unifié de rapport financier » (Surfi). La réglementation définissant ces reportings figure sur les sites internet de la Banque de France et du SGACPR :

- Pour les déclarations statistiques, la réglementation est définie par des décisions du Gouverneur de la Banque de France.[41]
- Pour les déclarations prudentielles, la réglementation est définie par des instructions de l'ACPR.

L'ensemble du reporting est organisé sous forme de tableaux. Les données collectées à des fins statistiques sont identifiées par le nom du tableau auquel elles sont rattachées : celui-ci comporte en effet le préfixe « M_ ».

Un site internet documentaire, e-Surfi, présente l'ensemble des tableaux du dispositif par thème (comptable, prudentiel, blanchiment, statistique, outremer, garantie des dépôts), les règles d'assujettissement, les nomenclatures, et les dictionnaires de données (dénommés taxonomies dans le formalisme XBRL) utilisées.

Essentiel au bon remplissage des déclarations, ce site fournit également au grand public un aperçu très détaillé des informations reçues par la Banque de France et par le SGACPR, ainsi que des contextes réglementaires dans lesquels ces remises interviennent. En offrant une vision précise de l'information récoltée au niveau le plus fin, ce site permet de cibler les demandes d'accès aux données.

Il est important cependant de noter que les informations demandées aux déclarants dans la collecte ne correspond pas exactement aux informations transmises à la BCE : ces dernières sont généralement beaucoup plus agrégées, avec pour conséquence que seules les données diffusées à la BCE bénéficient du niveau de qualité conforme aux standards en vigueur au sein de l'Eurosystème. Le niveau de qualité des données individuelles de collecte est variable d'un poste à l'autre et le passage des micro-données aux agrégats publiés résulte d'une séquence complexe réalisée dans des délais très contraints (cinq jours ouvrés en général) comportant des phases d'estimation de variables non collectées, de valeurs manquantes ou anormales et de calcul d'agrégats.

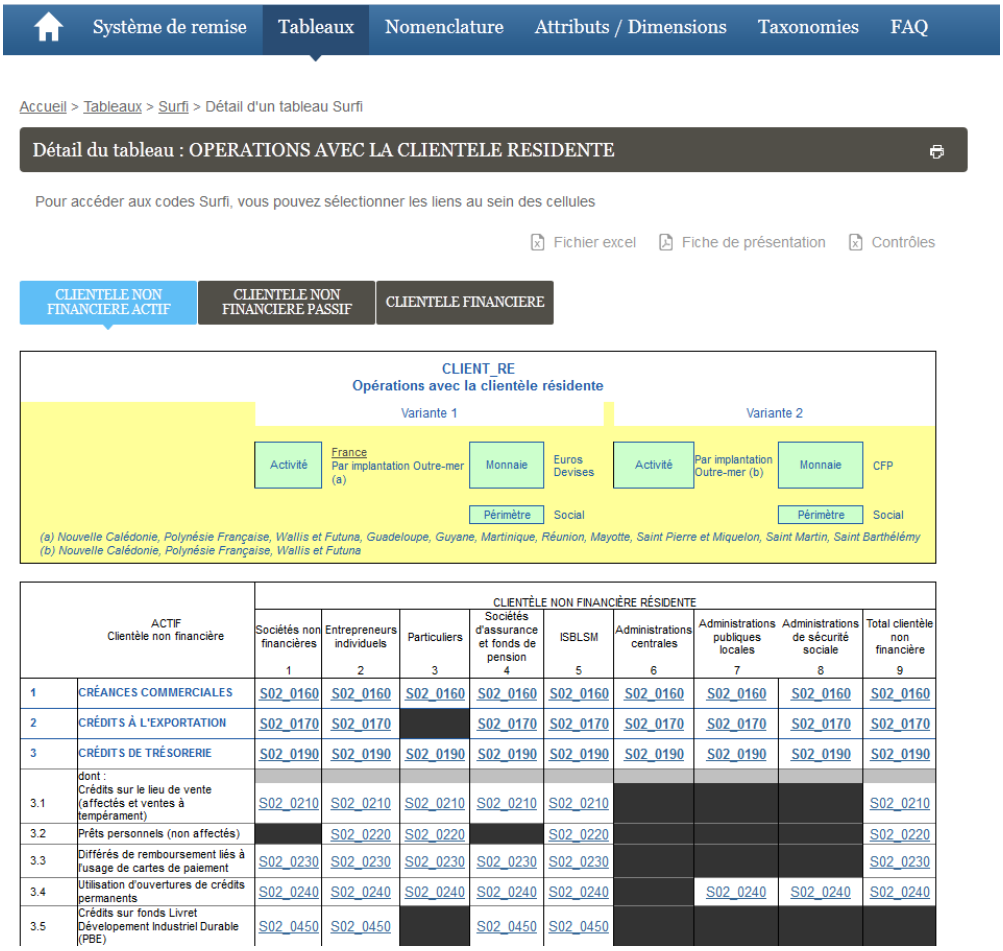
[41] https://www.banque-france.fr/economie-et-statistiques/espace-declarants/reglementation-des-statistiques-monetaires-et-financieres/dispositif-reglementaire-de-la-banque-de-france.html

Pour chaque rapport, le site propose :

- une fiche de présentation ;
- une vision de la remise sous la forme de tableau (exportable au format .xls) avec plusieurs onglets le cas échéant et la liste des dimensions sur lesquelles chaque onglet est ventilé[42] : la figure suivante en présente un exemple ;
- la liste des contrôles de cohérences qui doivent être vérifiés ;
- le détail des concepts mobilisés.

Les concepts sont potentiellement communs à plusieurs dimensions et à plusieurs tableaux : il est possible d'accéder à leur description soit en cliquant directement sur la cellule d'un tableau (ce qui ouvre une fenêtre « pop-up »), soit en les recherchant dans la nomenclature. Les attributs et les dimensions sur lesquels sont déclinés ces concepts sont également décrits, ainsi que la succession des versions de la taxonomie.

Figure 1. Exemple de gabarit disponible, pour les opérations avec la clientèle résidente



	ACTIF Clientèle non financière	CLIENTÈLE NON FINANCIÈRE RÉSIDENTE								
		Sociétés non financières 1	Entrepreneurs individuels 2	Particuliers 3	Sociétés d'assurance et fonds de pension 4	ISBLSM 5	Administrations centrales 6	Administrations publiques locales 7	Administrations de sécurité sociale 8	Total clientèle non financière 9
1	CRÉANCES COMMERCIALES	S02_0160	S02_0160	S02_0160	S02_0160	S02_0160	S02_0160	S02_0160	S02_0160	S02_0160
2	CRÉDITS À L'EXPORTATION	S02_0170	S02_0170		S02_0170	S02_0170	S02_0170	S02_0170	S02_0170	S02_0170
3	CRÉDITS DE TRÉSORERIE	S02_0190	S02_0190	S02_0190	S02_0190	S02_0190	S02_0190	S02_0190	S02_0190	S02_0190
3.1	dont : Crédits sur le lieu de vente (affectés et ventes à tempérament)	S02_0210	S02_0210	S02_0210	S02_0210	S02_0210				S02_0210
3.2	Prêts personnels (non affectés)		S02_0220	S02_0220		S02_0220				S02_0220
3.3	Différés de remboursement liés à l'usage de cartes de paiement	S02_0230	S02_0230	S02_0230	S02_0230	S02_0230				S02_0230
3.4	Utilisation d'ouvertures de crédits permanents	S02_0240	S02_0240	S02_0240	S02_0240	S02_0240		S02_0240	S02_0240	S02_0240
3.5	Crédits sur fonds Livret Développement Industriel Durable (PBE)	S02_0450	S02_0450		S02_0450	S02_0450				

[42] Par exemple, si les mêmes informations sont demandées respectivement pour les opérations en Euros et en devises étrangères, comme c'est le cas dans la Figure 1, alors l'onglet peut n'être présenté qu'une fois, en précisant qu'il est ventilé par monnaie.

2. Les données financières disponibles à l'Insee

Clément Guillo, Insee

L'Insee est en charge des comptes annuels nationaux qui incluent le compte des institutions financières et celui des assurances. Les données utilisées proviennent de la Banque de France, du SESOF pour les institutions financières et du SEFT pour les assureurs et de la balance des paiements pour les échanges avec le reste du monde. Les comptes trimestriels calés sur les comptes annuels construisent ces mêmes comptes pour un passé plus proche.

L'Insee diffuse également le tableau des opérations financières (TOF), partie financière des comptes nationaux dans lequel sont présents les encours en actif ou passif détenus par les différents secteurs institutionnels ainsi que, sous la forme de flux, les différentes opérations financières. Le solde au niveau des flux donne la capacité de financement.

Les enquêtes de conjoncture fournissent des données sur le compte des institutions financières et monétaires, le compte de la banque centrale, le compte des autres sociétés financières ainsi que des indicateurs tel l'indicateur de solidité financière et l'indice de marché des actions ainsi que des données sur les titres de dette et les taux d'intérêt.

Le FMI fournit à l'Insee, pour le secteur extérieur, une enquête sur les investissements de portefeuille et l'investissement direct.

Par le biais du dispositif Ésane, l'Insee diffuse des données sur les sociétés du secteur financier, sociétés captées en amont par l'ACPR.

L'enquête LIFI sur les liaisons financières entre entreprises a pour but d'identifier les groupes de sociétés en France et leur contour. Cette enquête s'inscrit dans la dynamique de passage de la notion d'unité légale à celle de groupes, passage qui aura un impact sur la construction des comptes de secteurs. Cette enquête a pour but d'obtenir une ventilation de l'actionnariat et de connaître les principaux actionnaires des sociétés privées.

Ainsi, modulo l'ensemble des traitements effectués, les diffusions effectuées par l'Insee s'appuient le plus souvent sur des données financières provenant d'organismes extérieurs.

3. Les données macro-financières de la Banque de France : le portail Webstat

Renaud Lacroix, Direction Générale des Statistiques, Banque de France

Les données macroéconomiques sont pour leur part diffusées dans le portail statistique Webstat disponible sur le site de la Banque de France. Ouvert en 2011, ce portail contient 21 000 séries en avril 2015. Il fait l'objet d'un enrichissement permanent. À l'horizon de la fin de l'année 2015, le nombre de séries pourrait atteindre 26 000.

Webstat permet de :

- visualiser les valeurs des séries sous forme de tables ;
- représenter les séries sous forme graphique (différents types de graphiques interactifs sont disponibles, allant des courbes jusqu'à des cartes géographiques) ;
- exporter des séries (format csv, SDMX-ML).

Dans Webstat sont diffusées principalement des séries produites par la Banque de France – monnaie, crédit, conjoncture, Balance des paiements, entreprises, marchés financiers, comptes financiers– mais aussi des données de prix ou encore de chômage.

Sont en cours d'introduction des séries relatives aux différents pays membres de la zone euro, diffusées selon un mode harmonisé par les Banques centrales du SEBC (*Joint dissemination framework*, JDF).

Webstat est le support de diffusion utilisé dans le cadre de la norme SDDS Plus pour les séries produites par la Banque de France, à laquelle la France a adhéré mi-février 2015.

L'accès aux séries est possible selon différents axes : thème économique, support de publication des séries (communiqué de presse statistique, tableau de bord, rapport…), ensemble d'appartenance des séries, zone géographique, source, date de mise à jour ou encore panier de données personnel préalablement défini.

4. Le dispositif de la Banque de France pour l'accès aux données individuelles

Renaud Lacroix, Direction Générale des Statistiques, Banque de France

1. La procédure d'accès aux données individuelles

Une nouvelle procédure a été mise en place à l'initiative de la Direction générale des statistiques afin de fournir des informations aux chercheurs.

Les données micro-économiques sont nécessaires aux travaux des chercheurs et aux réflexions sur le système financier, plus encore depuis la crise financière. La Banque de France s'inscrit dans cette tendance et entend développer la mise à disposition des données individuelles qu'elle produit, dans le respect du cadre juridique qui entoure ces données, et sous contrainte budgétaire. Les moyens financiers qui sont alloués à la Banque de France sont, en effet, chèrement comptés et il s'agit d'un aspect important pour les aspects opérationnels de la procédure d'accès.

L'approche retenue par la Banque de France vise à concilier trois objectifs :

- s'ouvrir davantage aux besoins des chercheurs ;
- ne pas prendre de risques juridiques et/ou de réputation ;
- limiter les coûts.

Avant 2012, il n'existait aucune procédure formalisée pour l'accès aux données. Les demandes parvenaient directement aux services producteurs et faisaient l'objet de traitements au cas par cas. Depuis 2012, à l'initiative de la Direction générale des statistiques (DGS), une procédure unifiée a été mise en place pour instruire les demandes d'accès aux données individuelles de la Banque (ménages, entreprises non financières, institutions financières). Cette procédure vise à instaurer une égalité de traitement entre les demandes des chercheurs et à garantir la transparence du processus. Les chercheurs internes de la Banque de France, logés en majorité au sein de la Direction générale des études et des relations internationales, doivent à ce titre justifier leurs demandes d'accès lorsqu'ils souhaitent utiliser des données individuelles produites à des fins statistiques. Cette procédure n'est pas applicable aux données gérées par le Secrétariat de l'Autorité de contrôle prudentiel et de résolution (SGACPR) qui relèvent d'un régime juridique spécifique. À ce propos, la position officielle du SGACPR est de considérer que les données prudentielles ne sont pas utilisables dans le cadre de travaux de recherche conduits par des équipes de recherche externes.

Du point de vue de la procédure elle-même, l'examen des demandes est collégial. Il s'opère dans le cadre d'une instance (le Secrétariat des demandes d'accès aux données statistiques) mise en place par les Gouverneurs en 2012 et présidée par le Directeur général des

statistiques. Y siègent également des représentants de la Direction générale des études et des relations internationales, de la Direction générale des activités fiduciaires et de place ainsi que le Directeur des services juridiques.

Les consultations du Secrétariat d'accès aux données interviennent au fil de l'eau en fonction des besoins formulés par les chercheurs. La procédure standard consiste en une consultation par voie électronique des membres du Secrétariat. En cas d'avis négatif du Directeur des services juridiques, la demande est rejetée. Dans le cas contraire, en cas d'avis partagé des membres du Secrétariat sur le bien-fondé d'une demande, une réunion physique est programmée afin de débattre de façon approfondie. La décision est prise en appliquant la règle de la majorité, étant entendu qu'un membre du Secrétariat peut toujours faire appel auprès du Gouvernement de la Banque pour arbitrage final.

Dans un souci de transparence, dès lors que le Secrétariat a été saisi, une information est communiquée au demandeur pour lui indiquer le délai d'instruction de la demande. En cas d'avis favorable par le Secrétariat, un délai maximum de communication des données est notifié au demandeur.

Au sein de la Direction générale des statistiques, la Direction de l'ingénierie et de la maitrise d'ouvrage statistiques (DIMOS) est en charge de l'instruction des demandes et de leur suivi ainsi que de la mise à disposition des données auprès des demandeurs. Le point d'entrée est constitué par une adresse électronique (email : DGS-DIMOS-acces-donnees-ut@banque-france.fr) qui permet de déposer des demandes et d'enclencher le cycle d'instruction de ces demandes. L'instruction d'une demande prend du temps, puisque des échanges préalables entre le demandeur et le(s) service(s) producteur(s) afin de préciser et bien délimiter le périmètre de la demande. Puis un formulaire doit être rempli, décrivant avec précision l'objet de la recherche et la composition de l'équipe de recherche. Ce formulaire doit permettre ensuite au Secrétariat de juger du sérieux du projet, de l'adéquation des données demandées avec ce projet ainsi que de la solidité de l'équipe de recherche. Le Secrétariat s'assure également que les données ne risquent pas d'être utilisées à des fins commerciales et définit les conditions de mise à disposition des informations (anonymisation, identification indirecte, *cf. infra*).

Lorsqu'un accord du Secrétariat pour l'accès à des jeux de données est notifié à une équipe de recherche, un engagement de confidentialité doit être signé par chacun de ses membres, que ceux-ci soient internes ou externes à la Banque de France. Ce document stipule notamment que les chercheurs s'engagent à ne publier que des résultats qui ne permettent pas une indentification, même indirecte, des déclarants et que toute publication en dehors du cadre de la recherche scientifique doit être préalablement validée par la Banque. Un formulaire doit être complété pour toute requête, y compris les demandes d'actualisation ou d'extension du périmètre de données communiquées antérieurement.

La réception de ces engagements de confidentialité signés de la main des demandeurs constitue un préalable à la mise à disposition effective des données.

Sur le plan juridique, les données collectées par la Banque de France relèvent généralement d'un règlement européen qui autorise la transmission de données confidentielles anonymisées pour des besoins de recherche (*cf. infra*). L'identification des déclarants ne devant pas être possible, les codes identifiants des banques (et de toute personne morale pour laquelle des données figurent dans la base) sont rendus anonymes.

Il existe une autre problématique spécifique liée à la forte concentration du secteur bancaire : dans certains cas, des croisements de données permettent d'identifier un établissement de façon indirecte. Le remplacement des identifiants par la procédure décrite dans la section précédente ne peut néanmoins empêcher le risque d'identifications indirecte, c'est-à-dire la possibilité d'identifier dans une base de données une personne morale ou physique alors que l'on ne dispose pas de l'identifiant explicite. Une identification indirecte peut se faire à l'aide de données dont on dispose par ailleurs, en particulier de données publiques.

L'identification indirecte d'un déclarant est en théorie presque toujours envisageable, mais sa mise en œuvre effective nécessiterait de mobiliser des moyens considérables ainsi que des processus complexes et coûteux de rapprochements de sources et d'études de similarité. C'est pourquoi l'anonymisation des données vise seulement à empêcher les situations les plus flagrantes d'identification indirecte.

Cette procédure est récente (2012). La prise en charge est effectuée par la Direction générale des statistiques qui assure l'instruction puis le traitement de ces demandes, en coordination avec les différents services producteurs de la Banque. À ce jour, 23 demandes de chercheurs internes et externes ont été traitées depuis le début de l'année 2012. Le délai de traitement varie entre 4 et 7 mois, même si certaines demandes ont pu nécessiter davantage de temps.

Communication de données *via* le Secrétariat d'accès aux données statistiques

Thème	demandes reçues	demandes validées	Études publiées	Remarques
Économie internationale	12	6	1	3 études parues
Économie monétaire et bancaire	12	9	5	
Entreprises	6	5		1 conférence
Économie financière	1	1		
Études microéconomiques et structurelles	1	1	1	
Innovation	1	1		
Ensemble des projets reçus par le secrétariat	**33**	**23**	**7**	

Les coûts constituent un facteur important dans le contexte budgétaire actuel. Il faut instruire les dossiers, c'est-à-dire entamer un dialogue avec les chercheurs afin de circonscrire leur demande, qui doit être aussi complète et aussi intelligible que possible pour permettre au Secrétariat de délibérer efficacement. Puis s'ouvre une phase d'échanges avec les services producteurs, qui devront, dans certains cas, s'assurer de la possibilité de réponse à la demande. Ces échanges peuvent prendre du temps lorsque les données individuelles

demandées n'ont pas été produites par le service producteur, pour diverses raisons. Il est alors nécessaire d'évaluer avec le service producteur la capacité à produire les données demandées afin de satisfaire, autant que faire se peut, la demande des chercheurs.

Enfin, l'anonymisation des données peut s'avérer complexe. Pour des raisons de maîtrise de la sécurité, la Banque de France ne gère plus l'anonymisation par des tables de passage et a investi dans des techniques plus efficaces (fonctions de « hachage », *cf. supra*) pour la gestion des identifiants cryptés. Ensuite, l'identification indirecte peut être difficile à mettre en œuvre car les techniques utilisées sont contingentes aux données traitées et aux possibilités, par recoupement d'informations, de parvenir à l'identification indirecte d'une banque ou d'une entreprise.

Jusqu'à présent, le service était gratuit : la Banque de France ne facturait aucun coût aux chercheurs. Cette situation pourrait évoluer à l'avenir, compte tenu de l'obligation de facturation de prestations qui a été rappelée par la Cour des Comptes. Ce sujet doit donc être étudié avec le plus grand soin.

2. Aspects juridiques et mise en œuvre pratique

L'article 8 paragraphe 5 du règlement n°2533/98 prévoit la transmission, pour des besoins de recherche académique, de données individuelles collectées par le SEBC à des fins statistiques sous réserve d'une anonymisation des codes identifiants des personnes morales au titre desquelles les données ont été collectées.

"*The national central banks shall use the confidential statistical information collected to fulfil the ECB's statistical reporting requirements exclusively for the exercise of the tasks of the ESCB except […]:*

(d) for granting scientific research bodies access to confidential statistical information which does not allow direct identification".

L'article 5 de l'orientation (ECB/1998/NP28) du 22 décembre 1998 concernant les règles communes et les normes minimales pour la protection de la confidentialité des informations statistique individuelles collectées par la BCE assistée par les BCN précise en outre les modalités de mise à disposition des informations en indiquant que "*where allowed by Article 8 of Regulation (EC)No 2533/98, confidential statistical information shall be transmitted extra muros electronically, following encryption*".

En pratique, le remplacement des identifiants natifs d'une base de données individuelles (nom, SIREN, CIB, etc.) par un nouvel identifiant anonyme est constant pour toutes les échéances figurant dans la base. Ces identifiants sont générés à l'aide d'une fonction de hachage (technique utilisée en cryptographie). Une telle fonction garantit plus facilement que d'autres techniques l'unicité dans le temps du lien entre le code originel et le code anonyme. Cette propriété permet l'appariement de bases anonymisées constituées à des dates différentes et facilite les mises à jour d'une base anonymisée existante. En contrepartie, la confidentialité

de la fonction de hachage est cruciale puisque sa divulgation permettrait d'identifier tous les individus présents dans une base.

À l'issue du changement des identifiants par la fonction de hachage, il convient de trier la base selon les nouveaux identifiants afin d'effacer l'éventuelle information contenue dans l'ordre d'apparition d'un élément dans la base.

La notion d'identification indirecte est introduite par l'article 8 paragraphe 9 du règlement n°2533/98 indique :

"*The ECB and the national central banks shall take all the necessary regulatory, administrative, technical and organisational measures to ensure the protection of confidential statistical information. The ECB shall define common rules and minimum standards to prevent unlawful disclosure and unauthorised use. The protection measures shall apply to all confidential statistical information (...)*"

L'article 2 de l'orientation (ECB/1998/NP28) du 22 décembre 1998 précise quant à lui ces « *common rules* » dans le contexte de données agrégées publiques, ces règles fournissant de fait une interprétation du concept d'identification indirecte :

"*All appropriate measures shall be taken to ensure that confidential statistical information is arranged in such a way that any published data covers at least three economic agents. Where one or two economic agents make up a sufficiently large proportion of any observation to make them indirectly identifiable, published data shall be arranged in such a way as to prevent their indirect identification. These rules shall not apply if the reporting agents or the other legal persons, natural persons, entities or branches that can be identified have explicitly given their consent to the disclosure*."

Comme l'a confirmé l'analyse des services juridiques de la Banque de France, la définition de l'identification indirecte semble bien être plutôt une question d'opportunité qu'une question strictement juridique, les textes laissant ici une certaine marge d'appréciation :

- dans le réglement n°223/209, article 3 (7) : "*To determine whether a statistical unit is identifiable, account shall be taken of all relevant means that might reasonably be used by a third party to identify the statistical unit*" ;

- dans le réglement n°223/209, article 3 (10) : "*indirect identification' means the identification of a statistical unit by any other means than by way of direct identification*".

Il en résulte que les possibilités d'identification indirecte peuvent être étudiées au cas par cas, et il est procédé le cas échéant à une anonymisation additionnelle destinée à empêcher cette identification dans le cadre d'une mobilisation de moyens « raisonnables ».

En pratique, on se contente de considérer que l'identification indirecte est effective dans les cas suivants :

- l'une des entités n'est pas présente sur toute la période[43] ;
- l'échantillon compte une ou deux entités ;
- l'une des entités représente au moins 85% du total pour une des variables numériques.

Afin de rendre difficile une identification indirecte dans le sens indiqué précédemment, il apparaît nécessaire de respecter les conditions suivantes :

- C1 : l'échantillon doit compter au moins trois entités par strate constituée à partir de variables publiques disponibles dans la base ;
- C2 : aucune des entités ne doit représenter 85% ou plus du total pour l'une des variables numériques de la base au sein d'une strate constituée à partir de variables publiques disponibles dans la base ;
- C3 : les entités qui ne sont pas présentes sur toute la période couverte par la base de données sont consolidées pour former une entité fictive dont le code identifiant est communiqué aux chercheurs afin de ne pas les induire en erreur.

La condition C3 est particulièrement restrictive et son application risquerait d'appauvrir considérablement la base de données effectivement mise à disposition. De plus, la date de création et/ou de disparition d'une entreprise peuvent constituer des informations indispensables aux chercheurs dans la conduite de leur projet, par exemple s'ils s'intéressent aux déterminants de la faillite d'une entreprise. Enfin, la date d'apparition ou de disparition statistique d'un élément dans une base ne correspond pas forcément à la réalité juridique. En conséquence, le niveau d'application de la condition C3 est laissé à l'appréciation des services producteurs.

Par ailleurs, les conditions C1 et C2 sont très contraignantes si elles doivent s'appliquer à toutes les strates susceptibles d'être constituées à partir de la base de données. Dans le cas d'une base sur le secteur bancaire, on pourrait par exemple s'intéresser aux banques mutualistes selon un code postal précis dans une zone de faible activité. Il est alors probable que les conditions C1 et/ou C2 ne soient pas respectées et il deviendrait impossible de transmettre quoi que ce soit. Pour cette raison, on se contente de vérifier en pratique que les conditions décrites ci-dessus sont bien respectées pour un nombre limité de croisements issus de variables centrales telles que :

- le statut juridique ;
- le secteur d'activité ;
- la localisation géographique.

La définition de ces variables et la vérification des conditions sont à la charge du service producteur de la base de données qui communique ces informations à la DGS-DIMOS. Si ces contrôles révèlent qu'une identification indirecte est possible, la DIMOS contactera alors les chercheurs demandeurs afin de discuter avec eux de la (ou des) variable(s) qui sera(seront)

[43] Par exemple, une entreprise qui a été créée à une certaine date ou une entreprise qui fusionne avec une autre et disparaît de la base de données. Si on sait à quelle date une entreprise a été créée ou à quelle date elle a fusionné, il peut être facile de l'identifier dans la base de données.

supprimée(s) de façon à se prémunir contre ce risque tout en limitant la perte d'informations utiles à leur recherche. Le service producteur extraira à nouveau la base de données avec des restrictions supplémentaires (par exemple une variable en moins) et vérifiera que plus aucune entité n'est identifiable sur la base des contrôles simplement grâce à l'application.

Cette proposition est un compromis tenant compte d'une part de la grande complexité qu'il y aurait à réaliser une anonymisation effective pour tous les croisements de variables, et d'autre part du souhait de la Banque de France d'encourager la recherche scientifique et donc de transmettre des bases de données exploitables par les chercheurs demandeurs.

Demande d'accès à des données confidentielles de la Banque de France

Volet n° 1 : Identification du demandeur

1 ***Organisme demandeur***

❶ *Identification de l'organisme*

Nom de l'organisme (en toutes lettres) :

Sigle correspondant (s'il existe) :

Adresse (lieu où les données seront traitées) :

Site internet :

❷ *Statut de l'organisme [44] et rattachement administratif*

☐ Unité de recherche (UMR, Institut, Centre ou Laboratoire)
- ☐ CNRS
- ☐ INRA. Dans quel département scientifique ?
- ☐ CEMAGREF. Dans quel département scientifique ?
- ☐ Unité de recherche propre à l'université. Dans quelle université ?

☐ Établissement d'enseignement supérieur

☐ Autre établissement public
- ☐ à caractère administratif (EPA)
- ☐ à caractère scientifique et technologique (EPST)
- ☐ à caractère scientifique, culturel et professionnel (EPSCP)
- ☐ à caractère industriel et commercial (EPIC)

☐ Administration de l'État. Dans quel ministère?

☐ Collectivité territoriale

☐ Association régie par la loi de 1901

☐ Organisme professionnel

☐ Organisme d'un état membre de l'Union européenne

☐ Banque de France. Précisez l'unité administrative de rattachement

☐ Autre. Précisez :

[44] Cochez la case correspondante

2 *L'équipe de recherche*[45]

Nom, prénom	Titre ou fonction	Lieu d'affectation	Courriel
⓪ *Le responsable administratif pour la demande*			
❶ *Le responsable scientifique de l'équipe*			
❷ *Le correspondant de l'équipe* *(si autre que le responsable scientifique)*			
❸ *Autre(s) membre(s) associé(s) à la recherche*			

[45] Toutes les personnes ayant accès aux données doivent être nommées ici et signeront un engagement de confidentialité préalablement à toute transmission de données.

Volet n° 2 :

Présentation du projet de recherche

---

1 ***Domaines d'études et de recherche***

❶ Quel est le domaine d'études et de recherche habituellement couvert par l'organisme demandeur et plus particulièrement par l'équipe de recherche ?

❷ Existe-il des relations régulières entre l'organisme demandeur ou l'équipe de recherche et les services de la Banque de France ?

2 ***Destination finale de la recherche***

Les travaux seront-ils effectués dans le cadre d'un contrat de recherche ? Si tel est le cas, précisez le nom et l'adresse de l'organisme commanditaire.

3 ***Quel est l'intitulé de votre projet de recherche ?***

4 ***Décrivez le projet de recherche***

5 ***Quel devrait être l'apport des données statistiques confidentielles que vous demandez pour votre recherche ?***

6 ***Existe-t-il des travaux de recherche, en France ou à l'étranger, pour lesquels le recours à des données similaires à celles que vous demandez a été nécessaire ?***

7 ***Type de publication***

Sous quelle forme sera publié le résultat des travaux (thèse, document de travail, article dans une revue de recherche) ?

Volet n° 3 :

Renseignements sur les demandes

---

1 ***Point sur les données confidentielles précédemment obtenues***

❶ Avez-vous déjà eu accès à des données confidentielles sur les entreprises ? Oui ☐ Non ☐

Avez-vous conservé ces données ?

☐ Oui totalement

☐ Oui partiellement

Quelles données avez-vous conservées ?

☐ Non

❷ Avez-vous déjà eu accès à des données confidentielles sur les ménages ? Oui ☐ Non ☐

Avez-vous conservé ces données ?

☐ Oui totalement

☐ Oui partiellement

Quelles données avez-vous conservées ?

☐ Non

2 ***Nature de la demande d'habilitation***

❶ Votre demande de données confidentielles concerne-t-elle un nouveau projet de recherche ? Oui ☐ Non ☐

❷ Souhaitez-vous élargir une précédente habilitation avec des données confidentielles supplémentaires ? Oui ☐ Non ☐

Si oui, souhaitez-vous disposez de

- ☐ plages temporelles élargies
- ☐ variables supplémentaires
- ☐ observations supplémentaires
- ☐ autre. Précisez :

3 *Variables demandées*

❶ ***Quelles sont les variables dont vous souhaitez disposer ? Justifiez le besoin pour chacune d'entre elles. Classez les dans l'ordre de priorité de la plus importante à la moins importante (les contraintes d'anonymisation ne nous permettent pas systématiquement de fournir toutes les variables demandées, il est donc nécessaire que vous précisiez les variables qui sont prioritaires pour l'aboutissement de votre recherche).***

❷ ***De quelle plage temporelle souhaitez-vous disposer ?***

4 *Conservation et sécurisation des données au sein de l'organisme demandeur*

❶ ***Quels seront les moyens mis en œuvre pour contrôler l'accès aux informations fournies ?***

① Concernant la sécurité du bâtiment et des bureaux de l'organisme demandeur

- ☐ Accès des locaux sur présentation d'un badge ou d'une carte magnétique
- ☐ Accès des locaux sous vidéo-surveillance
- ☐ Bureaux, armoires et tiroirs fermés à clé
- ☐ Autre. Précisez :

② Concernant la sécurité du système d'information

- ☐ Les données seront stockées sur un serveur sécurisé au sein de l'organisme demandeur.
- ☐ Les données seront stockées sur un micro qui ne sera pas connecté au réseau de l'organisme demandeur.
- ☐ Autre. Précisez : Un login et un mot de passe spécifiques sont créés pour l'accès du demandeur aux données sur le serveur sécurisé.

5. Les données prudentielles collectées par l'ACPR dans le cadre de ses missions

Denis Marionnet, Chef du service des études statistiques et veille documentaire, Autorité de contrôle prudentiel et de résolution (ACPR)

1. L'ACPR, qui n'a pas de mission statistique, collecte des données dans le cadre de ses missions prudentielles

Le législateur n'a pas confié de mission statistique à l'ACPR. Ses missions, définies à l'article L.612-1 du Code monétaire et financier sont avant tout prudentielles : « L'autorité de contrôle prudentiel et de résolution, autorité administrative indépendante, veille à la préservation de la stabilité du système financier et à la protection des clients, assurés, adhérents et bénéficiaires des personnes soumises à son contrôle. »

Pour l'exercice de ses missions, l'ACPR collecte un grand nombre de données, de nature prudentielle, auprès des organismes soumis à son contrôle.[46] Les données à remettre par ces derniers à l'ACPR pour l'exercice du contrôle prudentiel découlent désormais pour la plus grande part d'obligations harmonisées à l'échelle européenne par les autorités européennes de surveillance : Autorité bancaire européenne (ABE) pour les établissements de crédit et entreprises d'investissement et Autorité européenne des assurances et pensions professionnelles (AEAPP) pour les organismes du secteur des assurances. Ainsi, avec la directive CRD4 et le règlement CRR, l'Union européenne a mis en place un livre de règles prudentielles unique (*Single RuleBook*) qui s'applique directement à l'ensemble des établissements de crédit de l'Union européenne et qui s'étend au reporting associé (Normes techniques de l'ABE). De même, la directive Solvabilité II, qui entrera en vigueur au 1er janvier 2016, s'accompagne de la mise en place d'une réglementation et d'un reporting harmonisés à travers l'Europe pour les organismes d'assurance.

Comme la Banque de France, l'ACPR utilise le site Internet d'information technique et réglementaire sur le *reporting* « e-Surfi », qui comporte :

- une partie Banque (https://esurfi-banque.banque-france.fr/current/accueil/)
- une partie Assurance (https://esurfi-assurance.banque-france.fr/current/accueil/).

Les personnes soumises au contrôle de l'ACPR remettent les informations exigées par la réglementation à l'ACPR *via* le Guichet unique de collecte *OneGate*. Les remises sont réalisées au format XBRL.

[46] Pour davantage de précisions sur les personnes soumises au contrôle de l'ACPR, le lecteur pourra se référer l'article L612-1 ou à la page suivante : http://acpr.banque-france.fr/controle-prudentiel/les-assujettis-au-controle.html.

Il convient de noter, en outre, que l'ACPR participe au Mécanisme de Supervision Unique (MSU) des établissements de crédit de la zone euro et que la BCE, responsable de la surveillance directe des établissements et groupes importants français depuis le 4 novembre 2014, est l'autorité compétente pour recevoir les données prudentielles des entités susmentionnées (voir aussi point 3 ci-dessous).

2. **La publication d'études et de données agrégées vise à informer le public et les organismes sur la perception par l'ACPR des risques auxquels est soumis le système financier**

Les données prudentielles collectées par l'ACPR servent en premier lieu à l'exercice du contrôle individuel des organismes et établissements assujettis : vérification du respect de la règlementation et des exigences réglementaires par les services du contrôle bancaire et du contrôle des assurances. **Elles servent également à des fins de contrôle transversal**, afin par exemple de vérifier les contributions individuelles aux agrégats et de comparer les établissements entre eux afin de détecter des situations atypiques. Ces contrôles transversaux sont principalement réalisés par la direction des études qui produit des rapports, des notes ou des études et statistiques dont certaines font l'objet d'une publication, après accord du Collège de l'ACPR.

L'ACPR met ainsi à disposition du public des rapports ainsi que de nombreuses études, analyses et données associées à ces publications. **Ces publications ont pour objectif d'informer le public et les établissements assujettis sur la perception par l'ACPR des principaux risques auxquels est soumis le système financier.**

- Rapport « *Les chiffres du marché français de la banque et de l'assurance* » (https://acpr.banque-france.fr/publications/rapports-annuels/chiffres-du-marche-francais-de-la-banque-et-de-lassurance.html) : ce rapport vient compléter chaque année le rapport d'activité de l'ACPR qui est disponible un peu plus tôt dans l'année. Le rapport, qui est habituellement publié à l'automne, analyse en une centaine de pages la situation du système bancaire français ainsi que celle du secteur de l'assurance à travers un peu plus de 80 tableaux ou graphiques. Deux fichiers Excel reprenant l'ensemble des données sous-jacentes à ces tableaux et graphiques sont mis à disposition du public.

- Publication « *Analyses et Synthèses* » (http://acpr.banque-france.fr/etudes/analyses-et-syntheses.html) : cette publication regroupe des documents d'analyse statistique ou commentaires d'enquêtes menées par le SG-ACPR sur les risques dans le secteur bancaire et de l'assurance. Certains numéros sont accompagnés de statistiques ou données agrégées sous la forme d'un fichier Excel. Depuis 2013, cette collection rassemble entre une dizaine et une quinzaine d'études récurrentes ou ponctuelles sur des sujets variés des secteurs banque et assurance. C'est notamment dans une étude

récurrente de cette collection[47], consacrée à la situation consolidée des principaux groupes bancaires français qu'ont été intégrés les éléments enrichis sur les résultats par ligne de métier bancaire, qui faisait suite aux recommandations du rapport du groupe de travail du Cnis sur les « lignes de métiers bancaires » rendu public en mai 2012.

- Publication « Débats économiques et financiers » (http://acpr.banque-france.fr/etudes/debats-economiques-et-financiers.html) : cette publication, lancée en 2013, rassemble des articles invitant à lancer une réflexion sur des questions d'économie bancaire ou d'assurance, de réglementation ou de politique prudentielle. Tout en restant en parfaite cohérence avec les grandes lignes politiques de l'ACPR, ces documents, destinés à être publiés dans des revues professionnelles, n'engagent que leurs auteurs.

- Une page consacrée aux travaux et séminaires de recherche (http://acpr.banque-france.fr/etudes/travaux-de-recherche.html) est également régulièrement mise à jour.

L'ACPR publie également des informations et des données agrégées dans le cadre de la transparence des autorités de supervision (*Supervisory disclosure*). Il s'agit de règles européennes précisées dans la directive 2013/36/UE concernant l'accès à l'activité des établissements de crédit et la surveillance prudentielle des établissements de crédit et des entreprises d'investissement (*Capital Requirement Directive*, CRDIV) et qui imposent aux autorités compétentes de publier certaines informations afin que le marché intérieur bancaire puisse fonctionner avec une efficacité croissante et que les citoyens de l'Union bénéficient de niveaux adéquats de transparence. Les informations publiées doivent être suffisantes pour permettre une comparaison utile des approches adoptées par les autorités compétentes des différents États membres. L'article 143(1) de la CRDIV précise les différentes catégories d'information que les superviseurs doivent publier :

- texte des dispositions législatives, réglementaires et administratives et orientations générales adoptées dans chaque État membre en matière de régulation prudentielle ;
- modalités d'exercice des options et facultés nationales prévues par la législation communautaire ;
- critères généraux et méthodes que les superviseurs appliquent aux fins du contrôle et de l'évaluation prudentielle des établissements ;
- données statistiques agrégées sur les principaux aspects de la mise en œuvre du cadre prudentiel dans chaque État membre.

Ces informations et ces données sont disponibles sur cette page du site de l'ACPR : http://acpr.banque-france.fr/international/supervisory-disclosure.html

[47] Cf. Analyse et Synthèse n°46 « La situation des grands groupes bancaires français à fin 2014 », section 1.3 (page 16 à 20).

En 2014, l'ACPR a également rendu public l'ensemble des résultats, agrégés et individuels, de l'évaluation complète des bilans bancaires des 13 groupes bancaires français ayant participé à cet exercice avant l'entrée en vigueur du MSU le 4 novembre 2014. Les informations et données associées sont accessibles depuis cette page : http://acpr.banque-france.fr/international/les-grands-enjeux/stress-tests.html

Par ailleurs, l'ACPR publie des registres et des listes des établissements assujettis qu'elle met régulièrement à jour. Pour le secteur bancaire, ces informations sont disponibles sur cette page : http://acpr.banque-france.fr/agrements-et-autorisations/registres-et-listes/regafi-et-liste-des-agents-financiers.html. Pour le secteur assurance, les listes sont disponibles sur cette page : http://acpr.banque-france.fr/agrements-et-autorisations/registres-et-listes/registre-des-organismes-dassurance.html.

L'ACPR met également à disposition un registre des agents financiers -REGAFI- (https://www.regafi.fr) qui liste les entreprises autorisées à exercer une activité bancaire, financière, de monnaie électronique ou de services de paiement, réglementée conformément au code monétaire et financier.

Enfin, s'agissant des données prudentielles devant être publiées par les établissements, il convient de signaler que l'Autorité bancaire européenne a publié des normes techniques d'exécution et des orientations homogénéisant tant dans leur granularité que dans leur format les informations prudentielles à communiquer au public par les établissements en matière notamment de fonds propres et d'actifs grevés. D'une manière plus générale, le Comité de Bâle a finalisé en janvier 2015 une revue globale des informations prudentielles à publier par les établissements de crédit allant vers une plus grande standardisation. Ces modifications, introduites après consultation des utilisateurs, permettent au public de disposer d'une information adaptée à leur besoin, plus homogène et comparable.

3. Le cadre législatif encadrant l'accès aux données bancaires et financières est très strict

Le Règlement du Conseil (UE) 1024/2013 du 15 octobre 2013 (RMSU) a établi le Mécanisme de Surveillance Unique (MSU) des établissements de crédit (EC) dans la zone euro. L'ACPR participe en tant qu'autorité compétente nationale au MSU et siège au Conseil de surveillance prudentielle institué au sein de la BCE. Depuis le 4 novembre 2014, la BCE assure la **surveillance directe** des établissements de crédit (et groupes) français importants (*Significant Institutions ou SI)* et de la zone euro dont la liste est publiée par la BCE sur son site internet. À ce titre, **c'est désormais la BCE qui est propriétaire des données prudentielles transmises par les établissements de crédit importants. La publication de données prudentielles individuelles devrait être effectuée par la BCE elle-même ou avec son accord.** L'ACPR exerce la **surveillance** des établissements et groupes moins importants (*Less Significant Institutions ou LSI*), la BCE ayant reçu par ailleurs, la mission de surveiller l'ensemble du MSU ce qui inclut les activités de surveillance des autorités compétentes nationales, dont l'ACPR, au titre du MSU.

3.1 Les données prudentielles collectées par l'ACPR le sont pour l'exercice de ses missions

Outre les reportings désormais prévus par les textes européens directement applicables (*cf.* point 1), le cas échéant complétés par la BCE en ce qui concerne les établissements de crédit, l'ACPR dispose d'un pouvoir général, en application de l'article L612-24 du code monétaire et financier dans le cadre de l'exercice du contrôle, de requérir des informations auprès des personnes soumises à son contrôle pour l'exercice de ses missions de surveillance prudentielle telles que listées à l'article L612-1 et L612-2 du code monétaire et financier. En ce qui concerne les établissements de crédit, ce pouvoir s'exerce désormais dans les conditions définies par le Règlement MSU, précisé par le Règlement 468/2014 de la BCE du 16 avril 2014.

Les données prudentielles collectées ont ainsi pour seule finalité le contrôle prudentiel des personnes soumises au contrôle de l'ACPR (ou de la BCE, selon le cas). Le code monétaire et financier transpose les dispositions des directives européennes. En ce qui concerne les établissements de crédit, les articles 53 et suivants de la directive 2013/36/UE concernant l'accès à l'activité des établissements de crédit et la surveillance prudentielle des établissements de crédit et des entreprises d'investissement (CRD IV) **soumettent à une obligation de confidentialité les données et encadrent strictement l'utilisation et l'échange d'informations prudentielles que les organismes ont l'obligation de remettre aux autorités de surveillance prudentielle compétentes**. La BCE, conformément au RMSU, applique le droit de l'Union, CRD4-CRR et la transposition nationale de la directive.

Pour ces raisons, l'ACPR ne peut communiquer des données individuelles sur les établissements soumis à son contrôle. Toutefois, dans le cadre de ses pouvoirs de contrôle, de police administrative et de sanction, et dans des circonstances exceptionnelles qui doivent faire l'objet d'une décision ad hoc du Collège de supervision, l'ACPR « peut porter à la connaissance du public toute information qu'elle estime nécessaire à l'accomplissement de ses missions, sans que lui soit opposable le secret professionnel mentionné à l'article L. 612-17 ». Ces dispositions permettent notamment de publier, au registre officiel de l'ACPR, des mesures de police prises à l'encontre d'entreprises assujetties. En revanche, **ces dispositions ne permettent pas de donner un accès général, permanent ou régulier à des données confidentielles soumises au secret professionnel**.

3.2 Les agents de l'ACPR sont astreints à une obligation spécifique de confidentialité concernant les données collectées : le secret professionnel

En application de l'article L. 612-17 du Code monétaire et financier, les données collectées auprès des organismes soumis au contrôle ou reçues d'autres superviseurs sont soumises au secret professionnel.

La violation du secret professionnel constitue un délit, puni des peines prévues à l'article 226-13 du Code pénal, soit un an d'emprisonnement et 15 000 euros d'amende. Elle peut également constituer une faute disciplinaire, laquelle est autonome de la faute pénale.

En conséquence, les informations détenues par l'ACPR ne peuvent pas être communiquées à des tiers à moins qu'une disposition législative spécifique ne prévoie une exception, comme c'est le cas à l'article L.612-17 ou encore à l'article L.631-1 (entre autorités nationales), ou encore entre autorités compétentes au sein de l'Union européenne ou dans les pays tiers, à certaines conditions, prévues par la loi.

Dans le cadre du MSU, l'article 27 du RMSU assujettit, de même, la BCE, ses organes de décision et son personnel au secret professionnel.

3.3 Les hypothèses de levée du secret professionnel sont strictement limitées

C.1 Non opposabilité du secret professionnel de l'ACPR

L'article L.612-17 liste limitativement les hypothèses dans lesquelles le secret professionnel est levé. Elles concernent :

- l'autorité judiciaire agissant dans le cadre soit d'une procédure de liquidation judiciaire ouverte à l'égard d'une personne soumise au contrôle de l'ACPR, soit d'une procédure pénale ;
- les juridictions administratives saisies d'un contentieux relatif à l'activité de l'ACPR ;
- le Parlement dans le cadre d'une audition par une commission d'enquête ;
- la commission chargée des finances de l'Assemblée nationale et du Sénat ;
- la Cour des comptes dans le cadre de ses contrôles.

C.2 Transmission d'informations à d'autres autorités de supervision

L'ACPR peut, par dérogation aux dispositions générales sur le secret professionnel, échanger des informations nécessaires à l'accomplissement de leurs missions avec les autorités françaises de supervision du secteur financier[48], les autorités de supervision de pays de l'Union européenne[49], voire de pays tiers sous réserves entre autres de la conclusion d'un

[48] Article L.631-1 du code monétaire et financier
[49] Article L.632-1 et L.632-12-1 du code monétaire et financier

accord d'échange d'informations (MoU – *Memorandum of Understanding*)[50], mais également avec les autorités européennes de supervision.[51]

Les renseignements ainsi communiqués doivent être soumis à des règles du secret professionnel au moins équivalentes à celles de l'ACPR.

L'article 27 du RMSU prévoit, dans son paragraphe 2, que : « *Aux fins de l'accomplissement des missions que lui confie le présent règlement, la BCE est autorisée, dans les limites et dans les conditions prévues par les dispositions pertinentes du droit de l'Union, à échanger des informations avec des autorités et organes nationaux ou de l'Union lorsque les dispositions pertinentes du droit de l'Union autorisent les autorités compétentes nationales à communiquer ces informations à ces entités, ou lorsque les États membres autorisent une telle communication en vertu des dispositions pertinentes du droit de l'Union.* »

4. La recherche et l'accueil des chercheurs à l'ACPR

Afin de favoriser les échanges avec le milieu académique et de développer un centre de réflexion et de propositions en ce qui concerne la gestion du risque systémique, l'ACPR a lancé l'Initiative de Recherche Risques, Régulation et Risques Systémiques, dite Chaire ACPR « Régulation et Risque Systémique ». Cette Chaire ACPR a pour missions principales d'organiser des activités de recherche et de faciliter les contacts entre le monde académique et l'ACPR. La Chaire organise autour de ces thèmes un séminaire mensuel académique qui a lieu dans les locaux de l'ACPR et est ouvert au public. Les séminaires sont annoncés sur le site à la page travaux de recherche. Au-delà des conférences académiques qui ont été co-sponsorisées par la Chaire ACPR (conférence jointe avec SoFiE et la Banque de France par exemple), la Direction des études de l'ACPR organise aussi tous les deux ans une conférence académique internationale (*cf*. la page conférences du site de l'ACPR).

Comme la Banque de France, l'ACPR accueille régulièrement des chercheurs pour contribuer à l'exploitation des données prudentielles qu'elle collecte dans le cadre de ses missions. Pour les raisons juridiques mentionnées supra, ces chercheurs ne peuvent directement avoir accès aux données de l'autorité : ils interviennent à travers un statut de consultant qui leur permet d'apporter leur expertise sur un domaine donné en collaborant étroitement avec les chercheurs et économistes de la direction des études de l'ACPR. Dans ce cas, le chercheur consultant conseille les équipes de l'ACPR sur des aspects méthodologiques, techniques ou ayant trait à la présentation des résultats en vue de leur publication dans des revues externes. Les travaux de recherche sont disponibles sur le site internet de l'ACPR, dans la collection « Débats Économiques et Financiers » à fin mai 2015, sur les 16 numéros de cette collection, huit études ont été réalisées à partir de données bancaires individuelles. En outre, des données sont également publiées dans la collections « Analyses et synthèses ». Depuis la création de la Direction des études de l'ACPR à l'été 2011, cinq chercheurs extérieurs ont été accueillis. À

[50] Article L.632-7 et L.632-13 du code monétaire et financier
[51] Article L.632-4 et article L.632-6-1 du code monétaire et financier

la date de rédaction de ce rapport, trois chercheurs collaboraient avec l'ACPR en tant que consultants.

Pour contribuer en tant que chercheur consultant sur les travaux d'intérêt pour l'ACPR dans le domaine bancaire ou assurantiel, il convient de s'adresser à la direction des études de l'autorité. Il convient toutefois de noter que la direction des études de l'ACPR a des capacités d'accueil limitées et que toutes les demandes de collaboration ne peuvent être satisfaites.

6. L'accès aux données de marché par l'AMF

Carine Romey, AMF

Les régulateurs ont mis l'accent depuis la crise sur l'amélioration des outils de détection et sur la promotion d'une meilleure transparence des marchés pour déceler plus rapidement des tendances potentiellement dangereuses.

L'AMF n'a pas de mission statistique, ses collaborateurs sont soumis à une obligation stricte de confidentialité concernant les données collectées au titre des *reporting* règlementaires ; ces données ne peuvent être transmises qu'à d'autres régulateurs et dans des conditions juridiques strictes.

L'AMF n'a ainsi pas mis en place de procédures d'accès aux données proprement dites. Elle participe néanmoins à favoriser l'accessibilité de plusieurs manières :

- en améliorant la qualité des données de marché notamment ;
- en diffusant sur son site des informations publiques et réglementaires au niveau individuel ;
- en favorisant l'accès aux chercheurs à certaines données.

1. Améliorer la transparence des données de marchés

Après avoir suivi attentivement, en 2013, les négociations pour la révision de la directive Marchés d'instruments financiers (MIF II), l'AMF a participé, en 2014, aux différents groupes de travail lancés par l'ESMA – le régulateur des marchés financiers européens – pour proposer des mesures d'application. Parmi les sujets d'importance, figurent le régime harmonisé des pas de cotation qui vient compléter l'encadrement du trading à haute fréquence ou encore la notion de liquidité qui déterminera les obligations de transparence pour les instruments financiers non actions (dérivés, obligations).

Dans le cadre d'EMIR**,** l'article 9 prévoit que tout contrat dérivé doit être déclaré à un référentiel central de données (*Trade repositories*) dans un délai d'un jour ouvrable suivant sa conclusion, modification ou terminaison. Les *Trade repositories* doivent donc collecter les données des transactions réalisées sur des produits (OTC et listées). Ils doivent de plus rendre publiques des données sous formes agrégés concernant les positions ouvertes, les volumes de transactions et les montants agrégés par grandes classes de produits dérivés (actions, crédit matière première, changes et taux). Ces informations devront être actualisées au moins chaque semaine.

MIF2 introduit, par ailleurs, des entités règlementaires spécifiques jouant un rôle essentiel dans la publication ou la transmission de données de marché, les *Data Reporting Services*.[52]

2. Relayer les données publiques et /ou réglementaires sur son site

L'AMF publie sur son site deux bases de données :

- La base des décisions et informations financières(BDIF). Cette base contient des données sur les émetteurs : Prospectus & autres documents d'informations, Offres publiques d'acquisition, Seuils, pactes, dérogations & examens, Déclaration des dirigeants.

- La base GECO. Il s'agit des données sur les produits d'épargne et les sociétés de gestion agréés. L'AMF y publie, notamment, des statistiques mensuelles, la liste des sociétés de gestion en activité, celle des conseillers en investissements financiers (CIF) ainsi que la liste des produits autorisés à la commercialisation en France, etc.

Par ailleurs, l'ESMA a constitué une base de données relatives à l'activité des agences de notation, la base CEREP, visant à :

- collecter des données auprès des agences certifiées ou enregistrées dans l'UE sur les notes qu'elles émettent ;

- produire des statistiques de manière harmonisée sur l'activité de notation des agences (nombre de notes publiées, changements de notes, ampleur des changements, taux de défauts, matrices de transition) pour des périodes prédéfinies, avec une segmentation possible selon différents critères (géographique, type de produits notés, …) ;

- rendre ces statistiques et des informations qualitatives accessibles au public *via* le site Internet du CESR (http://cerep.esma.europa.eu/) et une interface dédiée.

[52] Cette catégorie recouvrant en fait trois catégories : *Approved Publication Arrangements* (APA). Ces acteurs sont chargés d'assurer au marché la publication des transactions de gré à gré réalisées par les entreprises d'investissement, notamment les prix de ces transactions. Les APA peuvent diffuser moyennant « un coût commercialement raisonnable » ces données en temps réel, mais les rendre publiques gratuitement dans les 15 minutes. *Approved Reporting Mechanism* (ARM). Sous MIF1 et MIF2, il est obligatoire de communiquer aux régulateurs les données de transactions en direct ou à travers un ARM. En Angleterre, cet intermédiaire est obligatoire alors qu'il ne l'est pas en France. Ces ARM sont en général des sociétés de services, filiales de Bourses (par exemple de Deutsche Börse). La France ne compte pas encore d'ARM. *Consolidated Tape Providers* (CTP). Ces acteurs pourront fournir des éléments consolidés des transactions, de gré à gré et celles négociées sur des plateformes. Ils ne sont pas tenus à une diffusion gratuite après un délai, à la différence des APA. Les CTP pourront aussi offrir des services complémentaires comme la publication de référentiels ou de données historiques. L'optique choisie par le régulateur européen a été de laisser le marché se former et faire émerger un *consolidated tape* par le jeu de la concurrence – un tel acteur n'ayant pas vu le jour,

3. Favoriser les travaux de recherche

L'AMF noue de nombreux liens avec les chercheurs, notamment dans le cadre de son conseil scientifique. Celui-ci a vocation à :

- améliorer l'information du régulateur sur les réflexions académiques en cours dans le domaine financier et favoriser ainsi les synergies potentielles entre la recherche académique, les évolutions de l'industrie financière et les préoccupations du régulateur ;

- identifier de façon prospective les évolutions et les sujets susceptibles d'impacter les champs d'activité de l'AMF ;

- lancer ou participer à des travaux de recherche en ligne avec les préoccupations de l'AMF, le partage de l'expérience acquise dans la régulation des marchés avec les chercheurs pouvant enrichir mutuellement les approches théoriques et empiriques.

En outre, a été lancée récemment une revue dédiée, la Revue du Conseil scientifique, destinée à faire connaître des travaux originaux. Cette publication souligne ainsi la volonté de l'AMF de diffuser la recherche académique et la réflexion autour des questions d'actualité du régulateur, sur la Place de Paris et bien au-delà de ses frontières, en encourageant les débats de nature à renforcer la qualité de la réglementation. S'ajoute à cela un prix AMF du « Jeune chercheur » ; l'AMF attribuera, à partir de 2015, un prix au profit d'un jeune chercheur travaillant sur des thèmes économiques et financiers présentant un intérêt pour le régulateur.

Pour ce qui est de l'accès aux données proprement dit, en particulier aux données haute-fréquence, l'AMF a conclu un contrat de coopération avec Eurofidai (CNRS) pour alimenter le projet BEDOFIH (une base de données financière européenne à haute fréquence). Grâce à ce projet les chercheurs pourront s'appuyer sur des historiques de données européennes à haute fréquence, et ainsi travailler sur la conception de modèles plus fiables et la définition de nouvelles modalités de régulation financière.

Ces données couvrent les données d'ordres, de transactions et d'évènements de cotation. Elles sont établies à partir des données anonymisées reçues par l'AMF et communiquées par Euronext pour le marché réglementé, complétée par le *reporting* réglementaire des PSI (Prestataires de services d'investissement) français et étrangers sur l'ensemble de leurs transactions y compris les transactions réalisées de gré à gré (OTC) du moment que l'AMF a été désignée comme l'autorité compétente sur le titre négocié. Ces données sont complétées par celles reçues d'Euroclear France (données Règlement livraison) et de LCH-Clearnet (pour les dérivés Matif et Monep). Elles seront diffusées par Eurofidai auprès des chercheurs notamment, après conclusion d'accords.

Il convient de souligner que le traitement de l'anonymisation de ces données transmises engendre des coûts financiers importants en termes de ressources à mobiliser pour l'AMF.

7. Le Centre d'accès sécurisé à distance (CASD)

Kamel Gadouche, Directeur du CASD

1. Bref historique

L'Inse a souhaité au milieu des années 2000 permettre la diffusion auprès des chercheurs de ses données brutes couvertes par le secret statistique. Il s'agissait alors de lever à la fois un **verrou juridique** et trouver **une solution technologique** qui garantisse la non-dissémination des données adaptée aux besoins des chercheurs.

Le **verrou juridique** été levé *via* un changement législatif en 2008, adossé au Comité du secret statistique qui est donc, pour les données administratives de nature économique et sociale, l'équivalent du CCTIRS pour les données médico-administratives.

Dans le même ordre d'idée, un verrou juridique de même nature, concernant les données fiscales, a été levé récemment, en 2013, et pourra peut-être servir de modèle pour les données de santé prochainement.

S'agissant du verrou technologique, il a été levé ces dernières années par les équipes du Groupe des Ecoles Nationales d'Economie et Statistique (GENES) grâce à la mise au point de la technologie d'accès sécurisé à distance du CASD.

Cette solution technologique a été développée par les informaticiens du GENES en collaboration étroite avec les utilisateurs de la plateforme (i.e. les chercheurs), avec un cahier des charges très strict :

- Sécurité maximale empêchant toute dissémination incontrôlée des données, ce qui excluait les solutions classiques *via* des logiciels de sécurité directement installés sur les postes de travail des utilisateurs.
- Coût minimal, afin de ne pas avoir à faire supporter aux utilisateurs des coûts prohibitifs comme ceux constatés dans la plupart des pays étrangers.
- Adaptation maximale aux besoins des utilisateurs, ce qui excluait des systèmes de centre physique (dans lesquels les utilisateurs sont obligés de se déplacer) ou les dispositifs de *remote execution*.

La solution à ce cahier des charges a conduit à la construction de l'architecture CASD *via* les SD Box, boîtier « tout-en-un » qui permet au chercheur de se connecter et de travailler directement sur les bases de données confidentielles.

La levée simultanée du verrou juridique et du verrou technologique a permis à la plateforme CASD de prendre son essor et d'être au service de toute la communauté des chercheurs en économie, sciences humaines et sociales.

Le CASD aujourd'hui en quelques chiffres :

- 1 000 utilisateurs en France et en Europe ;
- Près de 350 projets de recherche ;
- 300 SD-Box installées en France et en Europe ;
- 120 sources de données mises à disposition ;
- Une équipe mobilisant environ 15 ingénieurs statisticiens et informaticiens de très haut niveau.

2. Une puissance de calcul élevée et évolutive

Le Genes a une expérience de plus de 15 ans dans la fourniture de service de calcul pour la recherche scientifique. Le milieu de la recherche scientifique est un milieu très concurrentiel, chaque jour compte pour les chercheurs. Certains, par exemple, peuvent lancer des calculs qui durent plusieurs semaines à chaque itération. Un gain de performance de 20% peut ainsi leur faire gagner plusieurs semaines. Le Genes dispose d'un cluster de calcul dédié aux traitements complexes des chercheurs sur de gros volumes de données.

L'expérience du Genes a été entièrement mise à profit pour la conception des serveurs de calculs hébergés au sein du CASD. C'est ainsi que chaque composant de l'infrastructure des serveurs de calcul est déterminé avec le plus grand soin pour une adéquation optimale avec le système et les logiciels scientifiques de traitement de données :

- les disques durs (intégration de SSD) et leurs configurations ;
- les cartes de liaison pour les attachements directs de baies ;
- les liaisons réseaux (10Gb/s) ;
- la mémoire vive (RAM) ;
- les processeurs avec une évolution vers l'ajout de GPU (*Graphics Processing Unit*) ;
- les cartes mères et les cartes d'accélération.

Le système d'exploitation est optimisé pour le traitement de gros volumes de données avec une configuration permettant de favoriser les temps d'accès disque et les volumes de données transférés des disques aux processeurs.

Depuis quelques années, les progrès de la virtualisation ont permis l'intégration de cette technologie au sein des serveurs de calcul du CASD, renforçant ainsi, lorsque c'est nécessaire, les possibilités d'extension et d'allocation de puissance personnalisée.

L'architecture de la plateforme est évolutive, et permet d'adapter la puissance de calcul aux besoins du projet. Elle permet également de gérer de gros volumes de données, en particulier grâce à la mise en place d'un ensemble de serveurs Bigdata (Hadoop).

Enfin, il est important de noter que l'infrastructure CASD peut être implémentée :

- soit en mode **« hébergé »** où les bases de données sont hébergées sur les serveurs sécurisés du GENES, dans les locaux de l'INSEE ;
- soit en mode **« on-site »** où les bases de données restent hébergées chez le producteur (ou son prestataire d'hébergement) et où la technologie du CASD est gérée à distance par les équipes du GENES. Dans le cas d'Eurostat, c'est ce mode qui serait choisi, les données étant hébergées dans les locaux d'Eurostat au Luxembourg.

3. De bons retours par les utilisateurs actuels

L'enquête de satisfaction menée en mars 2014 avait pour but de collecter les opinions des utilisateurs du CASD sur différents critères, de la procédure d'accès au support informatique, en passant bien sûr par les dimensions statistiques. L'enquête a été envoyée à l'ensemble des utilisateurs du CASD, 73 d'entre eux ont répondu au questionnaire.

Il est à noter que c'est la première enquête réalisée depuis que le CASD est passé en mode payant. Certains répondants ont donc l'expérience des deux modes (payant et non payant).

Les utilisateurs ont plébiscité la qualité du support informatique avec une moyenne de 8,2/10 et une médiane de 9/10.

L'adéquation des logiciels disponibles était un critère dont nous attendions particulièrement les retours. En effet, l'environnement de travail du CASD, hermétique pour satisfaire les exigences de sécurité, oblige le CASD à proposer des packages complets de logiciels installés par défaut pour chaque projet. Il était important de savoir si ces logiciels correspondent aux besoins de nos utilisateurs et leur permettent de travailler dans des conditions proches de celles de leur environnement de travail habituel. L'objectif de l'étude était également d'identifier les logiciels qui pourraient enrichir l'offre actuelle.

4. Une offre de logiciels scientifiques large

À travers la plateforme CASD, les utilisateurs évoluent dans un environnement de travail convivial et disposent des logiciels nécessaires à la réalisation de leurs travaux de recherche : SAS, STATA, R et RStudio, SPAD, MATLAB, GAUSS, PYTHON, QGis...

Le CASD fournit un service d'assistance technique informatique sur l'utilisation des logiciels ainsi que pour certains, des conseils d'optimisation d'utilisation des ressources (interface entre le logiciel, le système et le matériel).

En complément de cette configuration de base, les utilisateurs peuvent s'ils le souhaitent insérer leurs propres données ou d'autres logiciels par l'intermédiaire d'une procédure sécurisée dédiée. De la même manière, toute demande de sortie (tableaux, graphiques, fichiers…) du système déclenche une procédure spécifique de vérification de non-rupture de la confidentialité.

Enfin, les chercheurs d'un même projet peuvent collaborer et partager leurs travaux depuis leur environnement de travail Windows.

5. L'intégration des technologies du bigdata

Nous travaillons, pour le traitement des données massives, à l'intégration de technologies Hadoop (bigdata) dans l'environnement CASD. Un premier pilote, dans le cadre du projet commun Mines Telecom et Genes, sera disponible au courant du troisième trimestre 2014. L'idée est de fournir des outils performants (R ou Python connectés à un cluster Hadoop) pour le traitement de données massives. Deux datascientists ont été recrutés au deuxième trimestre 2014 pour mener à bien ce projet.

6. Les données couvertes aujourd'hui par la plateforme CASD

Si historiquement, la plateforme ne couvrait que les données de l'Insee, elle comprend aujourd'hui un champ beaucoup plus vaste. Il s'agit dans tous les cas de **données individuelles très détaillées (données brutes) où seuls les noms, prénoms et numéro de sécurité sociale (NIR) ont été enlevés**. Il s'agit donc de données indirectement nominatives. Elles constituent un patrimoine riche structuré, documenté.

Dans le public, le CASD met aujourd'hui à disposition **des données des ministères de la justice, de l'éducation, de l'agriculture, des finances pour les données fiscales**... Pour ces dernières, il a été nécessaire de modifier la loi (loi ESR de 2013) et qu'un décret soit publié en 2014 pour qu'elles puissent être mises à disposition des chercheurs. Le décret précise explicitement que l'accès ne peut s'effectuer qu'au moyen du centre d'accès sécurisé aux données (CASD) du Genes.

Dans le privé, plusieurs sociétés se sont adressées au CASD parce qu'elles désiraient accroitre la sécurité de leurs données en accès externe dans le cadre de collaborations avec des chercheurs, des start up ou des consultants. C'est le cas par exemple de la banque postale, de Generali, de la banque publique d'investissement (BPI), de la société MAPP, d'ERDF, de RTE... Dans ce dernier cas, il s'agissait de mettre en place un datalab sécurisé pour travailler sur des technologies *big data* dans un environnement distant et sécurisé. Pour ces sociétés,

confier leurs données au CASD a été un moyen de répondre à leur besoin de sécurité et de permettre ainsi la réalisation de projets collaboratifs innovants impliquant de nombreux acteurs externes. Ces exemples s'inscrivent parfaitement dans la démarche actuelle d'« open innovation » mêlant à la fois le savoir- faire métier, la recherche, et les capacités d'innovation des start up.

La technologie du CASD et son infrastructure d'hébergement et de diffusion sécurisée des données est valorisée à l'étranger pour répondre à des problématiques similaires. Elle est ainsi pressentie pour être adoptée au niveau européen (Commission européenne / Eurostat), à la suite d'un premier pilote qui a rempli pleinement ses objectifs.

7. Le CASD et la sécurité des données

La technologie du CASD a été brevetée en France et à l'étranger et fait l'objet régulièrement d'un audit de sécurité par des sociétés externes spécialistes de la sécurité informatique.

Le CASD prépare son homologation aux standards de référence nationale et internationale en matière de sécurité et de qualité, pour renforcer la confiance de ses partenaires :

- **ISO 27001**, pour sa gestion de la sécurité ;
- la qualification au **Référentiel général de sécurité** (RGS) et le **Certificat de Sécurité de Premier Niveau** (CSPN) délivrés par l'Anssi ;
- l'agrément d'**hébergeur de données de santé**, délivré par le Ministre en charge de la santé.

Concernant la certification ISO 27001, la procédure est déjà bien avancée. Une personne dédiée, certifiée auditeur de la norme, travaille avec les équipes informatiques du CASD pour la réalisation des documents de certification. L'objectif est d'avoir un audit à blanc pour la fin de l'année 2014 et une certification au premier trimestre 2015. La partie analyse et quantification des risques est en cours.

B. Résultats de l'enquête sur l'accès aux données bancaires et financières pour les chercheurs

Gunther Capelle-Blancard, Professeur à l'Université Paris 1 Panthéon-Sorbonne

Dans le cadre de la commission « Système financier et financement de l'économie » du Conseil National de l'Information Statistique (Cnis), un questionnaire a été adressé aux chercheurs sur la visibilité et l'accessibilité des données bancaires et financières. Le questionnaire a été envoyé par email *via* la liste du Groupement de Recherche « Monnaie Banque Finance » du CNRS (GdRe n°335).

Entre mars et avril 2013, 55 chercheurs ont répondu anonymement à ce questionnaire, tous universitaires, à plus de 95 % en poste en France. Ces chercheurs sont pour la plupart spécialisés en économie bancaire, et dans une moindre mesure en finance de marché. Ce document synthétise les réponses.

Les sources de données sur le système financier

La principale source de données, tant pour la recherche que pour l'enseignement, est la Banque de France. Les organismes internationaux (BCE, BRI, FMI, OCDE) sont aussi régulièrement cités, tout comme l'Insee. En revanche, peu d'enseignant-chercheurs (moins de 10 %) déclarent utiliser d'éventuelles données de l'AMF ou des organismes professionnels.

L'utilisation de données bancaires individuelles

Parmi les répondants :

- 9 sur 10 considèrent « indispensable » l'utilisation de données bancaires individuelles (i.e. non-agrégées) pour la recherche ;
- 7 sur 10 estiment toutefois que l'accès à ces données est « difficile » ou « très difficile », particulièrement pour les données confidentielles (9 sur 10) ;
- 9 sur 10 évaluent l'accès aux données bancaires confidentielles plus restrictif en France que dans les autres grands pays de l'OCDE ;
- 4 sur 10 considèrent que la qualité des données bancaires s'est améliorée et 2 sur 10 que l'accès est plus facile (les autres ne voient pas d'évolutions notables).

Parmi les obstacles à une plus grande utilisation des données individuelles, arrivent en tête :

- l’accès aux données confidentielles (très problématique dans 88 % des cas) ;
- le coût (très problématique dans 69 % des cas).

Parmi les répondants, une moitié a déjà utilisé des données bancaires individuelles dans le cadre de ses travaux de recherche. Parmi eux :

- 18 % ont déjà utilisé des données individuelles de la Banque de France ; 15 % déclarent qu’ils ignoraient que la Banque de France mettait à disposition de telles données ; 27 % déclarent qu’ils ignoraient la procédure à suivre ; 18 % déclarent que les données individuelles de la Banque de France leur auraient été utiles, mais qu’ils ont préféré utiliser d’autres bases de données plus accessibles ; 12 % déclarent qu’ils en ont fait la demande, mais que celle-ci n’a jamais abouti ;
- 34 % ont déjà utilisé des données individuelles d’autres Banques Centrales ;
- 69 % ont utilisé la base de données (privée) Bankscope.

Parmi ceux qui n’ont jamais utilisé de données bancaires individuelles dans le cadre de leurs travaux de recherche :

- 39 % n’en ont jamais eu besoin ; 57 % n’y ont pas eu accès ;
- 87 % ignoraient que la Banque de France permettait aux chercheurs d’accéder (sous conditions) à des données bancaires individuelles.

Les données sur les marchés de capitaux et les autres intermédiaires financiers

La majorité des répondants estiment que ces données sur les marchés de capitaux ou les autres intermédiaires financiers (OPCVM, fonds de *private equity*, *hedge funds*, fonds souverains, fonds ISR…) sont accessibles, aussi bien en France qu’à l’étranger, même si ces données ne sont pas en accès libre.

Parmi les obstacles à une plus grande utilisation des données concernant les marchés de capitaux, arrivent en tête :

- Le coût (très problématique dans 43 % des cas)
- L’éparpillement des transactions entre une multitude de plateformes et les difficultés d’accès aux données confidentielles (très problématique dans 23 % des cas chacun)

Les données utilisées sont à 43 % des données agrégées (fréquence mensuelle ou inférieure), à 30 % des données journalières sur les marchés réglementés, à 11 % des données infra-journalières sur les marchés réglementés et à 13 % des données sur les marchés de gré à gré.

C. Bref historique des recommandations du Cnis sur l'accès aux données bancaires

Gilles Pouzin, membre du bureau du Cnis

Depuis plus de dix ans, la connaissance du secteur bancaire et financier, pourtant réputé parmi les plus stratégiques et sensible pour la stabilité et la croissance économique, se heurte à de nombreuses difficultés de mesure et d'accès aux données.

Ayant identifié ces lacunes dès le début des années 2000, le Conseil national de l'information statistique (Cnis) s'est attaché à étudier les moyens de les combler en confiant le soin à des groupes de travail constitués d'experts de formuler des recommandations en ce sens. La succession de recommandations formulées dans trois rapports précédant celui-ci, démontre la volonté de persévérer dans la recherche de solutions pour améliorer la transparence de l'activité financière.

Le 11 juillet 2005, le Cnis mandatait un groupe de travail, présidé par Edouard Salustro, sur les « statistiques structurelles fondées sur les groupes d'entreprises et leurs sous-groupes », qui devait notamment « *évaluer les apports du profilage de groupes pour former des unités statistiques ''entreprises'' au sens du règlement européen 696/931, c'est-à-dire des unités qui sont la combinaison de sociétés contrôlées par le même groupe d'entreprises et ayant une autonomie décisionnelle, souvent appelées sous-groupes ou branches opérationnelles par les groupes eux-mêmes ; prendre en compte leur dépendance financière* ».

Si on comprend l'importance d'appréhender le contour de ces branches opérationnelles et de leur dépendance financière pour les groupes influents dans tous les secteurs économiques, cette mission était d'autant plus essentielle concernant les groupes financiers eux-mêmes. Malheureusement, ce groupe de travail s'est heurté à des difficultés ne lui permettant pas d'avancer concrètement sur ce dernier point.

En janvier 2008, ce rapport précise que ce groupe de travail « *n'a pas pu prendre en compte dans son analyse de la situation actuelle et dans ses propositions les cas particuliers d'une part des fonds d'investissement, d'autre part du secteur financier ; s'agissant de ce dernier, l'analyse doit prendre en compte tant les groupes de ce secteur (et particulièrement les sociétés non-financières de ces groupes) que les sociétés de ce secteur mais relevant de groupes hors secteur financier. Le GT demande que cette analyse soit prochainement conduite, notamment avec les services compétents. Un complément à l'actuel rapport devra alors être produit* ».

Le 27 mars 2008, alors que la crise bancaire se propage (quelques jours après la faillite de la banque *Bear Sterns*), le bureau du Cnis mandate un nouveau groupe de travail pour explorer la connaissance statistique des activités bancaires, présidé par Gilles de Margerie.

Deux problématiques abordées par le mandat de ce groupe de travail s'inscrivent parfaitement dans la mission confiée au présent rapport. On y retrouve :

- la même préoccupation d'améliorer la connaissance statistique du secteur bancaire : « *Ainsi devra être examiné (…) ce que signifie la notion de branches opérationnelles au sein des groupes financiers, prenant notamment en compte les pôles banques de détail/de financement/etc., ainsi que les classifications statutaires, banques/assurances/etc.* » ;

- les mêmes difficultés d'accès aux données bancaires pour améliorer sa connaissance statistique : « *Évaluer le coût et les problèmes méthodologiques qu'impliquerait la constitution d'une base de données des groupes du secteur financier. Aujourd'hui, la collecte des comptes des sociétés financières n'est pas centralisée. Elle est organisée, principalement pour des besoins propres, par les autorités administratives chargées du contrôle de ces sociétés (CB, ACAM) et selon des normes de restitution et des plans comptables différents. Consolider ces données pour former les comptes des groupes financiers peut donc être source de problèmes. Une solution alternative pourrait être de constituer une base de données directement à partir des comptes consolidés publiés par les groupes financiers.* »

Vu l'ampleur de la tâche, il faudra deux ans au groupe de travail présidé par Gilles de Margerie pour explorer les besoins de mesures statistiques propres au secteur bancaire et financier.

En février 2010, le groupe de travail constitué autour de Gilles de Margerie, par Sanvi Avouyi-Dovi, Jean-Marie Fournier et Denis Marionnet, remet son rapport consacré aux « Statistiques sur les groupes financiers ».

Ce rapport préconise notamment de « *mobiliser les informations contenues dans les différentes bases de données existantes en développant les échanges d'informations entre les différentes institutions* » en espérant que « *la mise en place de l'Autorité de Contrôle prudentiel en 2010, qui sera investie du pouvoir de délivrer les agréments et de contrôler les entreprises des secteurs de la banque et de l'assurance devrait avoir un effet très structurant* ».

Recommandation n°2 du rapport du Cnis de février 2010 : « *Le GT recommande qu'une convention soit établie entre la Banque de France, l'Insee et tout autre organisme compétent en vue de préciser la nature et le périmètre des échanges d'informations nécessaires à l'amélioration du suivi des groupes financiers. Elle devrait préciser les règles et modalités d'échange d'informations dans ce domaine entre les différentes institutions* ».

En 2011, dans le sillage des recommandations de ce groupe de travail, et à force que le sujet de l'amélioration de la connaissance des activités bancaires soit débattu au Cnis, un « groupe technique » est mis en place, c'est-à-dire un groupe de travail « interne » entre professionnels de la statistique sans représentants des utilisateurs.

En juin 2012, ce « groupe technique » du Cnis piloté par trois représentants de la Banque de France (Alain Duchâteau, Émilie Crète et Daniel Gabrielli) rend son rapport intitulé « Lignes de métiers bancaires » dont les recommandations portent davantage sur la connaissance des activités bancaires au niveau agrégé que sur une amélioration de l'accès aux données bancaires individuelles permettant d'étudier leurs activités.

Contrairement aux rapports précédents de 2008 et 2010 recommandant de rechercher un meilleur accès aux données bancaires pour mieux appréhender l'activité des entreprises et groupes de ce secteur, le rapport du Cnis confié à la Banque de France en 2012 recommande, au contraire, « *de veiller au respect du secret statistique et notamment à ce que les situations individuelles ne puissent malencontreusement transparaître au sein d'une publication trop fine ou granulaire* ».

***

Alors que la Cnis recherche depuis dix ans les moyens d'améliorer l'accès aux données bancaires pour développer la connaissance statistique des activités bancaires, le groupe de travail actuel s'inscrit dans la continuité de cet historique.

Les membres du groupe du travail et les auteurs de ce rapport espèrent que leurs recommandations pourront cette fois aboutir à des avancées concrètes.

D. Charte du G8 pour l'ouverture des données publiques (extraits)

Une transcription en français de la Charte est disponible sur le portail de la Modernisation de l'action publique : http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/charte-g8-ouverture-donnees-publiques-fr.pdf

L'accès aux données publiques permet aux individus et aux organisations d'acquérir de nouvelles connaissances, de mettre au point des innovations qui amélioreront la qualité de vie de chacun, et de contribuer à une meilleure diffusion de l'information à l'échelle d'un pays et entre les États.

S'ils recueillent une multitude de données, les gouvernements et les entreprises ne les partagent pas toujours selon des modalités permettant au public de les découvrir, de les utiliser et de les comprendre facilement. C'est là une occasion manquée.

Les citoyens exigent aujourd'hui de pouvoir accéder à l'information et aux services par voie électronique au moment et selon les modalités qui répondent à leurs attentes. C'est de plus en plus souvent le cas pour les données publiques elles-mêmes. Nous arrivons à un tournant qui marque le début d'une nouvelle ère au cours de laquelle il sera possible de se servir des données ouvertes pour acquérir des connaissances, générer des idées et produire des services afin de créer un monde meilleur pour tous.

L'accès aux données ouvertes permet d'accroître la transparence de l'action des gouvernements et des entreprises (…).

Des données publiques librement accessibles et gratuitement réutilisables peuvent être à la source de services et des produits innovants susceptibles d'aider les individus à faire face plus facilement à la vie moderne. Utilisées de la sorte, ces données amènent le secteur privé à innover, contribuant à la création de nouveaux marchés, de nouvelles entreprises et de nouveaux emplois.

Nous, membres du G8, convenons que les données ouvertes constituent une ressource inexploitée qui présente un potentiel immense pour favoriser l'essor de sociétés plus fortes, mieux interconnectées, qui répondent mieux aux besoins de nos citoyens et contribue nt à l'innovation et à la prospérité.

Nous convenons par conséquent de nous conformer à un ensemble de **principes** qui régiront l'accès aux données mises à disposition par les gouvernements du G8 ainsi que leur diffusion et leur réutilisation. Ces principes sont les suivants :

- Données ouvertes par défaut
- De qualité et en quantité

- Accessibles et réutilisables par tous
- Ouvrir les données pour améliorer la gouvernance
- Ouvrir les données pour encourager l'innovation

(…)

Principe n°1. Données ouvertes par défaut

Nous reconnaissons que l'accès libre aux données publiques et leur réutilisation gratuite sont d'une importance majeure pour la société et pour l'économie.

Nous convenons de faire progresser nos gouvernements vers l'ouverture par défaut de nos données publiques.

Nous reconnaissons que l'expression « données publiques » doit être entendue dans son sens le plus large possible. Elle peut recouvrir les données produites par des organismes gouvernementaux nationaux, fédéraux, locaux ou internationaux, ou par le reste du secteur public.

Nous reconnaissons qu'il convient de respecter les lois nationales et internationales qui ont trait à la propriété intellectuelle et aux données personnelles ou sensibles.

Nous établirons un principe d'ouverture par défaut des données publiques, conformément à la présente Charte, tout en reconnaissant qu'il existe des motifs légitimes justifiant que certaines données ne puissent pas être diffusées.

Principe n°2. De qualité et en quantité

Nous reconnaissons que les gouvernements et le secteur public détiennent de grandes quantités de données susceptibles de présenter un intérêt pour les citoyens.

Nous reconnais sons également que l'ouverture de données de haute qualité peut nécessiter du temps, et qu'il importe de travailler ensemble et de consulter des utilisateurs de données ouvertes, à l'échelle nationale et au-delà, afin de déterminer quelles données il convient de diffuser en priorité et d'améliorer.

Nous :

- diffuserons des données ouvertes de grande qualité qui soient à jour, complètes et exactes. Dans la mesure du possible, les données seront disponibles sous leur forme initiale non modifiée, et présenteront le meilleur degré de granularité possible ;

- veillerons à ce que l'information contenue dans les données soit rédigée en langage simple et clair, de manière à être comprise par tous, étant entendu que la présente Charte ne prévoit pas d'obligation de traduction ;

- nous assurerons que les données fassent l'objet d'une description complète afin que leurs usagers disposent de suffisamment d'information pour comprendre leurs forces et leurs faiblesses, leurs limites sur le plan de l'analyse, les exigences en matière de sécurité et les modalités pour les traiter ;

- diffuserons les données dès que possible, permettrons aux utilisateurs de fournir un retour d'information, puis les réviserons afin de garantir qu'elles soient conformes aux normes les plus élevées de qualité de données.

Principe n°3. Accessibles et réutilisables par tous

Nous convenons d'ouvrir nos données publiques d'une manière qui aide chacun à y accéder et à les réutiliser.

Nous reconnaissons que les données publiques ouvertes devraient être accessibles et réutilisables gratuitement pour en favoriser l'usage le plus répandu.

Nous convenons que lorsque des données ouvertes sont diffusées, elles devraient l'être sans faire l'objet d'obstacles d'ordre bureaucratique ou administratif, comme des exigences d'enregistrement, qui peuvent dissuader les usagers d'y accéder.

Nous :

- diffuserons les données en formats ouverts dans toute la mesure du possible, ce qui permettra de s'assurer que les données soient accessibles au plus grand nombre d'utilisateurs possible et pour le plus grand nombre de fins possible ;

- diffuserons le plus de données possible, et nous promouvrons les avantages de l'ouverture et encouragerons un accès libre et gratuit aux données publiques dans les cas où cela n'est pas encore réalisable pour moment. Dans de nombreux cas, cela impliquera de fournir des données dans des formats multiples, afin qu'elles puissent être à la fois traitables par des ordinateurs et compréhensibles par des humains.

Principe n°4. Ouvrir les données pour améliorer la gouvernance

Nous reconnaissons que l'ouverture des données publiques renforce nos institutions démocratiques et favorise de meilleures politiques publiques pour répondre aux besoins de nos citoyens, dans nos pays comme à travers le monde.

Nous reconnaissons également que l'ouverture des données suscite de plus en plus d'intérêt de la part d'autres organisations et initiatives multilatérales.

- (…)

Principe n°5. Ouvrir les données pour encourager l’innovation

Conscients de l’importance que revêt la diversité afin de stimuler la créativité et l’innovation, nous convenons que plus les personnes et les organisations qui utilisent nos données seront nombreuses, plus les avantages sociaux et économiques de l’ouverture des données seront substantiels. Cela s’applique aux usages à fins commerciales tout autant qu’aux usages à fins non commerciales.

Nous :

- nous emploierons à mieux faire connaître l’ouverture des donnés publiques et à encourager les citoyens, tels que les développeurs d’applications et les organisations de la société civile œuvrant dans le domaine de la promotion de l’accès à l’information, à réaliser le potentiel de la réutilisation des données ouvertes ;

- donnerons à une future génération d’innovateurs en matière de données les moyens dont ils ont besoin en leur fournissant des données en formats lisibles à la machine.

E. Les bases de données commerciales

Gunther Capelle-Blancard, Professeur à l'Université Paris 1 Panthéon-Sorbonne

Il existe une offre commerciale très large de données sur les marchés financiers, à des coûts toutefois élevés. La plupart des données produites à usage commerciale ont avant tout comme débouché le secteur financier lui-même, gros consommateur de données, particulièrement de données de marché. Les deux principaux fournisseurs de données financières sont Thomson-Reuters et Bloomberg. S'y ajoutent, pour ce qui est de l'information boursière, de nombreuses entreprises de services financiers, tels que McGraw Hill Financial (qui publie notamment les indices Dow Jones). Ces bases de données sont très complètes, mais peu accessibles du fait de leur coût, au minimum plusieurs dizaines de milliers d'euros, même si quelques Universités ou Grandes Ecoles en France réussissent à financer sur contrat des abonnements (souvent à tarif négocié).

Dans ce complément, nous examinons en particulier le cas de deux bases de données : l'une sur les banques (Bankscope), l'autre sur les marchés (Eurofidai).

1. Bankscope

Pour ce qui est des données sur les banques, la base la plus utilisée est la base de données Bankscope, commercialisée par Bureau van Dijk, qui propose une couverture géographique mondiale et un historique très long, pouvant aller jusqu'à 25 ans (avec une fréquence annuelle, voire trimestrielle sur la période récente). Cette base de données fournit des informations générales (statut juridique, domiciliation du siège social, classement mondial, rating…) ainsi que des données détaillées sur les bilans, les comptes résultats, et la structure actionnariale de plusieurs milliers de banques. Bankscope fournit aussi des informations (quoique partielles) sur les éléments hors-bilan, sur la nature des fonds propres (réserves, titres hybrides, etc.) et sur les fonds propres réglementaires (*core tier one* ou *risk weighted asset*).

D'après sa plaquette de présentation, Bankscope a plus de 10 000 clients dans le monde, dont un tiers d'institutions financières ; elle est utilisée par plus de 90 % des banques. Son coût pour la recherche est toutefois élevé, avec un abonnement de l'ordre de 12 000 euros par an qui ne couvre pas toutes les années (jusqu'à 16 000 euros pour l'accès aux données trimestrielles). Bankscope est de loin la principale base de données utilisée par les chercheurs en économie bancaire – par défaut, il faut bien le reconnaitre, en raison de la grande difficulté

d'accéder à d'autres sources de données, notamment les données dont disposent les autorités de supervision.[53]

Plus récente, la base SNL Financial a une couverture moins importante (particulièrement sur l'Europe, où les données ne sont disponibles qu'à partir de 2005). En revanche, le travail d'harmonisation des données est réputé meilleur. Les abonnements pour ces bases de données sont d'une dizaine de milliers d'euros (tout dépend de la durée, de la couverture, des options offertes, etc.).

Répétons-le, ces bases de données ne diffusent que des données non-confidentielles. Leur valeur ajoutée tient à leur interface et, éventuellement, à leur travail d'harmonisation. Ces bases permettent en effet d'avoir un accès aisé à des données simultanément pour un grand nombre de banques.

Outre la question du coût, se pose la question de la qualité de l'information collectée, notamment par Bankscope. Cette base de données n'est en effet que partiellement renseignée, avec un nombre important de données manquantes, même pour certaines grandes banques et dans des pays où la publication d'information comptable et financière est obligatoire, comme la France. Il apparaît aussi que cette base de données souffre de nombreux biais.

Dans le cadre d'un réseau de recherche organisé entre 1999 et 2001 par la Banque Centrale Européenne avec l'ensemble des Banques centrales nationales de la zone euro, des économistes ont comparé les résultats obtenus avec les données banques centrales et les données Bankscope.[54] Ils montrent qu'en appliquant rigoureusement la même méthodologie à des échantillons de données différents, on obtient des résultats très différents : « *On montre aussi que la base de données Bankscope ne rend pas compte de façon satisfaisante de l'hétérogénéité des comportements bancaires. En effet, sur plusieurs points, les données issues de Bankscope suggèrent des réponses très différentes de celles obtenues à partir des bases de données plus complètes disponibles dans les banques centrales nationales.* » Ce travail met clairement en lumière non seulement la supériorité des données collectées par les Banques centrales, qui sont plus complètes, plus précises et plus détaillées, mais aussi l'effet important que cela peut avoir dans les recommandations, et donc la conduite des mesures de politique économique.

Ce travail n'est pas le seul du genre. Un document de travail de la Banque des Règlements Internationaux a également montré que Bankscope souffrait d'un biais de sélection important.[55] Plus récemment encore, deux chercheurs en thèse de l'École d'Économie de Paris, en poste à l'ACPR et à la Banque de France, ont publié un document de travail

[53] On recense ainsi plusieurs milliers d'articles académiques sur Google Scholar qui utilisent Bankscope.
[54] Michael Ehrmann, Leonardo Gambacorta, Jorge Martínez-Pagés, Patrick Sevestre et Andreas Worm, Financial systems and the role of banks in monetary policy transmission in the euro area, Notes d'Études et de Recherche # 93, Nov. 2002.
[55] Kaushik Bhattacharya, How good is the BankScope database? A cross-validation exercise with correction factors for market concentration measures, BIS Working Papers No 133, Sept. 2003.

appréciant la qualité de la base de données Bankscope.[56] Leur étude fait apparaître de nombreuses difficultés :

- L'absence d'explication quant à la façon dont les données sont collectées et éventuellement redressées, et plus généralement quant aux spécificités du jeu de données et aux points de vigilance.

- Une couverture dans le temps relativement limitée et variable selon les pays et les banques de telle sorte que certaines périodes ne peuvent faire l'objet d'études du fait de données absentes ou de très mauvaise qualité (du début des années 1980 au milieu des années 1990).

- Un très grand nombre de données non renseignées. L'interruption des données peut être le résultat d'une faillite, d'une opération de fusion-acquisition, ou autre sans que l'on sache pourquoi. Certaines données sont parfois seulement manquantes entre deux dates, là encore sans que l'on sache pourquoi. Comment alors traiter ces données ? Faut-il les supprimer ? Les considérer comme des valeurs nulles ? Cela entraîne nécessairement des biais, bien connus des statisticiens, qui peuvent avoir une influence significative sur les résultats. Par ailleurs, cela limite considérablement les possibilités d'agrégation (notamment au niveau national).

- La question de la consolidation : Bankscope ne renseigne pas du tout la nature d'un établissement : s'agit-il d'une filiale, d'une maison-mère, dispose-t-on de comptes sous-consolidés ? Cela pose d'énormes problèmes, car : i) on procède vite à un double comptage (en comptant les actifs de la filiale et de la maison mère, cette dernière incluant les premiers par consolidation) ; ii) il est là aussi très délicat de procéder à des agrégations. Plus généralement, les relations intra-groupes constituent aujourd'hui un enjeu de recherche important et Bankscope ne permet nullement de traiter ces questions.

- Les définitions des variables : Bankscope ne suit pas rigoureusement les évolutions comptables et réglementaires. Par exemple, Bankscope ne fournit pas les fonds propres règlementaires conformes aux recommandations de Bâle 3.

[56] Thibaut Duprey et Mathias Lé, Bankscope Dataset: Getting Started (January 15, 2015). Disponible sur SSRN: http://dx.doi.org/10.2139/ssrn.2191449.

2. Eurofidai

En France, il existe une base de données très riche, avec un long historique, sur les actions, les indices boursiers et les fonds de placement collectifs. Cette base de données a été créé à l'initiative du CNRS, *via* l'institut de recherche Eurofidai.

Ces données ne sont pas soumises au secret statistique. Il s'agit essentiellement de données non-confidentielles, mais souvent éparses. La plupart du temps, ces bases de données sont également particulièrement volumineuses (plusieurs millions d'informations). Eurofidai fait un précieux travail de collecte de l'information, d'harmonisation et de formatage. Les données proposées par Eurofidai peuvent se répartir en deux catégories : les données journalières (voir tableau 2), et les données très haute-fréquence (projet BEDOFIH). Pour ces dernières, qui sont souvent très lourdes à gérer, Eurofidai propose également des services personnalisés (mise à disposition de serveurs sécurisé par exemple).

Eurofidai commercialise ses données *via* la plateforme IODS. Cette initiative portée notamment par l'Insead, la Caisse des dépôts et consignations, et le CNRS et qui propose aux établissements de recherche qui souscrivent à un abonnement, une plateforme d'accès à plusieurs bases données bancaires et financières. L'abonnement de base (5 000 € HT par an) comprend outre la base de données journalières Eurofidai, la base Lipper (Thomson Reuters) sur les fonds de placement collectifs, la base Altares (réseau Dun & Bradstreet) qui recense des données financières sur 2 500 000 entreprises françaises, ainsi, entres autres, que la base macro-financière de Observatoire de l'Épargne Européenne.[57]

L'essentiel du financement d'Eurofidai provient du CNRS et, pour ce qui est des données haute-fréquence, du programme « Equipements d'excellence » (Equipex) dans le cadre des « Investissements d'avenir ». Dans la mesure où il s'agit de financements publics, il convient d'envisager que ces données soient désormais librement accessibles et gratuitement.

[57] En supplément, il est aussi possible d'accéder à d'autres bases, comme Factset sur l'actionnariat des entreprises et les prévisions des analystes, Compustat pour les données financières d'entreprise au niveau mondial, Barchart pour les marchés au comptant et à terme de commodités, etc.

Tableau 1. Les données de marché Eurofidai

Produit	Caractéristiques	Disponibilité
Actions	• Couverture des zones Europe, Asie, Pacifique et Moyen-Orient : 61 pays, 42 places électroniques • Affichage de toutes les places européennes et asiatiques où un titre est côté et non uniquement la place principale • Ligne principale : code EUROFIDAI unique et permanent pour les différentes bases (actions, options, opérations sur titres) • Variables additionnelles calculées par EUROFIDAI	1977-2013 pour la France 1980-2013 pour l' Europe 1986-2012 pour l' Asie
Indices de référence EUROFIDAI (« benchmark »)	• Calculés par EUROFIDAI à partir de ses propres données de qualité sur les actions • En US dollars et en euro • Indices par pays et pour l'Europe (par exemple, un indice pour les 27 pays européens est proposé (1991-2013 en dollars et 1999-2013 en euro)) • Indices par secteur • Indices factoriels et portefeuilles de référence (basés sur la taille, le book-to-market et le momentum) • Variables additionnelles et indices sur mesure calculés sur simple demande • Couverture complète des marchés européens et asiatiques	1977-2013 pour la France 1988-2013 pour l'Europe 1992-2012 pour l'Asie
Autres indices	• Indices calculés par d'autres fournisseurs de données	1980-2014
Cours de change	• Environ 14 000 cours de change • Au comptant • A terme sur demande	1975-2014 au comptant 2003-2011 à terme
Fonds mutuels	• Fonds émis en Europe et en Asie • Marchés de gré à gré (OTC) et marchés organisés • Classés par pays ou par monnaie d'origine • Couverture complète des fonds en euros, 1980-2012 • Couverture complète des fonds dans les autres devises pour les 10 dernières années • Informations très précises sur les caractéristiques des fonds sur une base quotidienne : informations générales (manager du fonds, nom, site web, brochure...), frais, "benchmark"...	1980-2013 marchés de gré à gré 1985-2013 marchés organisés
Opérations sur titres	• Donnés détaillées (split, dividendes, radiation...)	1977-2014
Codes : table de correspondance	• Table de correspondance reliant les différents codes des émetteurs et des instruments	-

Source : https://www.eurofidai.org/accueil_bases_boursieres.html

F. Cas spécifiques d'accès aux données

1. Les données bancaires individuelles consolidées aux États-Unis : éléments sur les « Call reports »

Renaud Lacroix, Direction Générale des Statistiques, Banque de France

Sous l'impulsion des institutions de supervision américaines, la *Federal Deposit Insurance Corporation* (FDIC), l'*Office of the Comptroller of the Currency* (OCC) et le *Federal Reserve System*, réunies au sein du FFIEC[58], le système américain de collecte des informations bancaires consolidées a été intégralement refondu en 2005. Ce projet de modernisation plus connu sous le nom de *Call Report Modernization Project* visait d'une part à optimiser la transmission des données bancaires et d'autre part à créer un entrepôt de données publiques. L'utilisation de la technologie XBRL a permis de mettre en place un système de collecte intégré facilitant ainsi la transmission et le contrôle des données. La base de données publique est perçue comme une avancée majeure, notamment par les chercheurs.

1. Le reporting bancaire consolidé US, les "Call reports"

Le reporting *Consolidated Report of Condition and Income* plus communément appelé « call report » est remis trimestriellement, sous un format relativement stabilisé depuis 1984, à la Federal Deposit Insurance Corporation (FDIC) par l'ensemble des institutions financières américaines[59], conformément à la section 1817 du *Federal Deposit Insurance Act*. Chaque trimestre, plus de 8 000 institutions transmettent, au plus tard 30 jours calendaires après la date d'arrêté de fin de trimestre, le détail de leur compte de résultats, de leur bilan et hors-bilan à un niveau consolidé[60] :

- des informations détaillées sur les revenus et charges du compte de résultat, sur les variations trimestrielles des capitaux propres, les radiations et recouvrements de prêts ainsi que les variations des provisions pour pertes sur prêts, et les revenus des filiales étrangères ;

- l'actif et le passif détaillés par instruments (ventilation fine des expositions par type d'obligations, par type de crédit,…) ou par contrepartie (transactions interbancaires

[58] Federal Financial Institutions Examination Council
[59] Les institutions financières concernées sont les *national banks*, les *state member banks*, les *insured state non member banks*, les *savings associations* et les *credit unions*, ainsi que les *thrifts* (soit l'ensemble des institutions financières).
[60] https://www.fdic.gov/regulations/resources/call/crinst/callinst2014_dec.html

par type de banques). Des informations précises sur les dépôts des banques, par maturité, par montant doivent être renseignées ;

- des informations détaillées du hors-bilan : lignes de crédit, garanties, positions par types de produits dérivés, et notamment les positions sur produits dérivés négociés de gré à gré ventilés par contrepartie ;
- des informations spécifiques au système américain de couverture contre les faillites des établissements de dépôts telles que l'assurance-dépôt FDIC;
- des informations portant sur les ratios de capital et les pondérations de risques de chaque type d'actifs;
- des informations sur la titrisation, par type de crédits titrisés ;
- des informations sur les services fiduciaires.

Pour cette collecte, plus de 2 500 champs de données doivent être renseignés.

Ce reporting est défini à partir de normes comptables nationales, qui diffèrent parfois des normes comptables IFRS et par conséquent de celles de l'Union Européenne qui a adopté les normes IFRS en 2002 (Règlement CE n° 1606/2002) pour la plupart des banques (seules les banques actives sur un marché régulé y sont contraintes).

2. Le Call Report Modernization Project

Les institutions en charge de la supervision des institutions financières américaines, la FDIC, l'OCC (*Office of the Comptroller of the Currency*) et le *Federal Reserve System*, réunies au sein de la FFIEC, ont initié et abouti en 2005 à une refonte intégrale du système de collecte des « call reports ». La FFIEC a ainsi défini des exigences en termes de définitions des données et de contrôles qualité, regroupées dans une taxonomie XBRL, transmise aux fournisseurs privés de logiciels.

Les logiciels utilisés par les banques intègrent l'ensemble des définitions, la hiérarchie des informations à renseigner mais également des contrôles qualité entre les tables renseignées d'une part et sur une table entre plusieurs échéances d'autre part. Les logiciels fournissent ainsi une vérification automatique des données, respectant ainsi, avant l'envoi des données au régulateur, les exigences en termes de qualité. Le non-respect de règles de validation doit par ailleurs être dûment justifié. Les données sont envoyées *via* un mode de transmission sécurisé dans un entrepôt de données, la *Centralized Data Repository* (CDR). La bonne réception est notifiée par un mail auprès de la banque. Les données sont accessibles au public *via* la CDR quelques heures après leur transmission par la banque. Si tant est que la banque ait à faire des modifications, celles-ci peuvent être effectuées en ligne.

La quasi-totalité des données collectées est mise à la disposition du public, pour chaque institution, sur le site web du Central Data Repository (CDR) du FFIEC[61]. Ce site met à disposition du public les reportings trimestriels individuels pour les remises postérieures au T1 2001 sous format pdf, xbrl, et sdf. Des reportings regroupant les banques par taille ou par type sont également proposés. En outre, le FDIC met à disposition un outil, le *Uniform Bank Performance Report* (UBPR) rassemblant, pour chaque banque, les principaux indicateurs issus de ces reportings sur plusieurs échéances, permettant ainsi à chacun d'évaluer la performance et le risque afférent à une banque[62].

Sauf demande dument justifiée de la banque, la plupart des informations collectées sont mises en ligne et accessibles *via* le site internet dédié, à l'exception de certaines informations :

- les informations relatives à l'assurance-dépôt auprès de la FDIC sont jugées confidentielles ;
- l'estimation des expositions douteuses ;
- les encours de prêts à l'habitat et de crédits à la consommation ainsi que les expositions aux entreprises présentant un profil de risques élevé (le caractère risqué du profil est défini par la FDIC), ainsi que les montants titrisés de ces encours ;
- l'exposition à une seule contrepartie la plus élevée, et la somme des 20 expositions les plus élevées ;
- la répartition des différents crédits aux ménages par tranche de probabilité de défaut ;
- les réserves de garantie en lien avec les prêts hypothécaires.

[61] https://cdr.ffiec.gov/public/
[62] Un exemple pour Bank of America : https://cdr.ffiec.gov/Public/Reports/UbprReport.aspx?rptCycleIds=80%2c79%2c78%2c76%2c75&rptid=283&idrssd=480228&peerGroupType=&supplemental=

2. Accès aux données FIBEN pour fluidifier le marché secondaire des prêts aux PME – Extrait de la note du conseil d'analyse économique, n°18

Jacques Cailloux, Nomura International, correspondant du CAE
Augustin Landier, École d'économie de Toulouse (TSE), Université de Toulouse 1, IDEI
Guillaume Plantin, Sciences Po Paris, membre du CAE

La Banque de France produit des données financières sur les entreprises françaises dont le chiffre d'affaires est supérieur à 750 000 €. Regroupées dans le fichier FIBEN (fichier bancaire des entreprises), ces données informent sur le passif de l'entreprise, l'historique récent des incidents de paie- ment, et attribuent une note à l'entreprise qui vise à mesurer sa capacité à honorer ses engagements financiers à un horizon de trois ans. Cette note dépend des données de bilan, mais aussi d'une appréciation qualitative issue des échanges entre la Banque de France et l'entreprise.

Actuellement, ces données sont accessibles à l'ensemble des établissements de crédit, des sociétés d'assurance-crédit et d'assurance-caution. Les informations FIBEN sont cou- vertes par le secret professionnel et ne peuvent donc pas être communiquées à des tiers non autorisés. Cette confidentialité des données évite à la Banque de France, qui les produit, d'être considérée comme une agence de notation et régulée comme telle. Fluidifier et standardiser l'information concernant le risque financier des entreprises, sans donner d'avantage informationnel à certaines catégories d'acteurs, aurait cependant une valeur sociale. Cela pourrait faciliter l'émergence d'un marché secondaire des prêts et favoriser l'innovation sur ce marché. Un accès plus large aux informations FIBEN pourrait par exemple permettre des ventes de prêts individuels des plus grosses PME, l'escompte de créances commerciales, etc. Cela ouvrirait l'analyse des produits titrisés à des acteurs divers – pas uniquement des établissements de crédit mais aussi des acheteurs potentiels ou des consultants offrant des prestations d'analyse. Plus généralement, cela élargirait l'analyse économique du marché du crédit aux PME. Les personnes susceptibles de participer à la formation des prix sur le marché du crédit ne sont pas toutes salariées des établissements de crédit ou des assureurs.

Produite par la Banque de France, cette information pourrait soit être vendue à tous les acteurs économiques, soit être mise dans le domaine public au titre de l'open data. Le projet de loi relatif à la simplification de la vie des entreprises avait envisagé l'ouverture de FIBEN aux assureurs au titre de leur activité de crédit, aux plateformes de *crowdfunding* et aux sociétés de gestion pour l'activité des fonds de prêt. Le Sénat, qui l'a examiné en novembre 2014, a cependant décidé de supprimer l'article 30 qui contenait cette disposition, en invoquant le « coût », « la faisabilité » de FIBEN, et le risque que la vie des entreprises ne s'en trouve au final « complexifiée ». Il est regrettable de voir ce sujet enterré sans motifs substantiels, dans la mesure où le fichier FIBEN existe déjà, et qu'il a déjà fait la preuve de son utilité pour les acteurs bancaires qui en ont l'usage. Il y aurait tout à gagner à ouvrir ces données, et ce le plus largement possible, afin de faciliter de nouvelles relations de financement en dehors du secteur bancaire.

3. Les données sur les crédits aux entreprises et les incidents de paiements

Catherine Refait-Alexandre, Professeur à l'Université de Franche Comté

Régis Blazy, Professeur à l'Université de Strasbourg

Les données sur le crédit aux entreprises sont indispensables à l'analyse du financement bancaire des entreprises, qui constitue la source principale du financement de l'économie, en particulier en France.

1. Les statistiques agrégées

Au niveau agrégé, des résultats d'enquête sont accessibles gratuitement. Ainsi, la BCE réalise trimestriellement une enquête sur le financement bancaire au sein de la zone euro. (https://www.ecb.europa.eu/stats/money/surveys/lend/html/index.en.html). L'enquête donne des informations globales sur la demande de crédit des ménages et des entreprises, et sur l'évolution des conditions d'octroi de prêt (encours globaux, demande de garanties et de clauses de sauvegardes, échéance). Le questionnaire est disponible mais pas les résultats bruts de l'enquête.

La BCE publie des résultats d'enquêtes semestrielles sur l'accès au financement des PME. (https://www.ecb.europa.eu/stats/money/surveys/sme/html/index.en.html). Une analyse riche est faite de l'évolution du financement, en terme d'encours et de tarification et des causes de cette évolution : modification du risque des PME, modification de la volonté des banques de leur octroyer des crédits ou encore modification de la demande de financement.

La Banque de France (notamment l'Observatoire des Entreprises, Direction des entreprises) fournit également des études sur le financement bancaire des entreprises, sous forme de bulletin (https://www.banque-france.fr/publications/bulletins-de-la-banque-de-france/les-bulletins-de-la-banque-de-france.html) ou de cahier statistique (https://www.banque-france.fr/publications/bulletins-de-la-banque-de-france/dernieres-statistiques-mensuelles-parues-dans-le-bulletin-de-la-banque-de-france.html).

Ces informations sont riches, mais elles constituent un travail abouti, un « produit fini » alors que le chercheur a besoin d'un « produit brut » pour réaliser son propre travail d'analyse. Aux États-Unis, les chercheurs ont souvent accès aux résultats bruts des enquêtes ainsi réalisées.

Ainsi, les résultats de l'étude nationale sur le financement des petites entreprises (*National Survey of Small Business Finances*) réalisée par la Réserve fédérale américaine sont librement accessibles sur son site (http://www.federalreserve.gov/pubs/oss/oss3/nssbftoc.htm). La méthodologie, les questions posées et les réponses sont données. Les questions de l'enquête

concernent les caractéristiques des entreprises (localisation, âge, activité, concentration de la détention du capital, etc.) ses documents comptables et des éléments concernant ses principaux propriétaires (nom, sexe, degré d'éducation, âge, race, ethnie etc.) Les questions portent également sur les services financiers utilisés par les entreprises (dette à court terme, crédit-bail, utilisation de cartes de crédit, épargne, etc.) et le nom des institutions qui leurs fournissent ces services. Par exemple, sont renseignés : le nombre de prêts refusés, le nombre de prêts pour achat de véhicule, les raisons pour lesquelles l'entreprise utilise tel ou tel service financier.

2. Les états financiers des entreprises

Les documents comptables publiés par les entreprises permettent de connaître l'endettement financier global, les dettes vis-à-vis des établissements de crédit, les concours bancaires courants et les taux d'intérêt versés (globalement). Souvent est également accessible le montant d'emprunt remboursé (en annexes). Ces informations ne permettent pas toujours une analyse fine du financement bancaire, mais elles en donnent une première idée.

Sans être exhaustif, on peut mentionner cinq sources qui fournissent ces informations comptables de manière centralisée.

- La base de données Altares concerne exclusivement les entreprises françaises. Elle est vendue par IODS dans son abonnement de base. Son prix est de 5 000 euros HT par an. Le principal problème est l'absence de données pour de nombreuses entreprises : soit parce que les comptes ne sont pas disponibles, soit parce que les comptes ne sont pas tous renseignés.

- La base de données Diane est assez similaire, avec un coût à peu près équivalent.

- La Banque de France dispose de la base de données FIBEN, qui est alimentée par les liasses fiscales des entreprises et qui comprend les comptes de plus de 2 millions d'entreprises (elle remplace depuis 2009 la Centrale des Bilans qui collectait des documents comptables des entreprises adhérentes, sur la base du volontariat).

  La base FIBEN sert également à nourrir les bases « Dirigeants et associés » et « Dossier Risques » qui sont commercialisées par le Service des Produits Entreprises et Banques de la Banque de France.

- L'Insee a constitué depuis de nombreuses années une base exhaustive de bilans, à partir des liasses fiscales et de données internes. Il s'agit du fichier Suse (Système unifié de statistiques d'entreprises). L'Insee présente le fichier Suse comme « un ensemble cohérent de données individuelles et statistiques sur les entreprises par l'exploitation conjointe de deux sources d'information : l'une fiscale, qui regroupe les déclarations des entreprises auprès de la Direction générale des Impôts et l'autre statistique, l'enquête annuelle d'entreprise ». La base Suse peut être rapprochée

d'autres bases également gérées par l'Insee (bases de liaisons financières, enquêtes créations…).

- Il existe d'autres bases de données donnant donc accès aux comptes des entreprises tout pays confondus. Les deux principales sont Facset et Thomson One Banker. L'abonnement à Facset coûte 9 000 euros HT par an, et celui à Thomson One Banker coûte plus de 15 000 euros HT par an. Outre leur coût élevé, ces bases de données ne sont pas toujours satisfaisantes, car elles se caractérisent également par de nombreuses données manquantes. Ceci est vrai même pour un historique assez récent (deux ou trois ans en arrière uniquement), et dès que l'on souhaite des lignes de comptes un peu détaillées.

3. Les données sur les crédits aux entreprises octroyés par les banques

Certaines bases de données procurent quelques informations sur les crédits octroyés aux entreprises ou sur leur niveau de risque. Ainsi, Deal Scan, l'une des bases de données incluses dans Thomson One Banker, procure des informations précises sur le marché des prêts (montant, échéance, taux d'intérêt, clauses de sauvegarde etc.) Malheureusement, ces informations ne concernent que les prêts syndiqués, et non les prêts bancaires indivis octroyés par une banque unique. La base de données Altares donne des indications sur le nombre de banques en relation commerciale avec une entreprise, ainsi que le nom de ces banques. Malheureusement, la fiabilité de ces informations n'est pas très bonne. Altares fournit aussi une mesure du risque de défaillance (le score « Conan-Holder ») des entreprises. Mais cette donnée n'est pas toujours renseignée.

Une autre possibilité est la réalisation d'une enquête afin de collecter des informations précises et fines relatives au financement bancaire des entreprises. Les obstacles sont nombreux, au-delà du coût (matériel et financier). Si l'enquête est réalisée auprès des entreprises, se pose la question de l'objectivité des réponses, et la fiabilité des résultats se heurte au taux très bas de réponse (10% est généralement considéré comme un bon taux de retour). Les enquêtes réalisées auprès des banques sont sans doute plus fiables. Mais elles exigent la collaboration d'un groupe bancaire, prêt à divulguer ses fichiers clients. Se pose alors le problème de la représentativité de l'échantillon, puisque les fichiers proviennent du coup d'un seul groupe bancaire.

4. Les données sur les incidents de paiement et le risque de faillite des entreprises

Une première source d'information émane de l'Insee qui collecte systématiquement, depuis plusieurs décennies, les jugements d'ouverture et de clôture de procédures collectives. Ces jugements font l'objet d'une publication préalable au BODACC (Bulletin Officiel des Annonces Civiles et Commerciales). Le fichier « défaillance » ainsi constitué permet de faire des croisements avec d'autres bases gérées par l'Insee. Ces données sont accessibles aux utilisateurs des bases de l'Insee, dans le cadre de la procédure du Comité du secret statistique.

L'Insee publie régulièrement des statistiques agrégées de suivi des défaillances à partir de cette base.

Des données plus complètes existent au sein des greffes des tribunaux de commerce – qui gèrent le *Registre de commerce et des Sociétés* et vendent les informations collectées par les tribunaux de commerce, *via* Infogreffe.fr. Ces données ne concernent que les entreprises faisant l'objet d'une procédure judiciaire (en aval d'un défaut de paiement : liquidation et redressement judiciaires, sauvegarde) et sont théoriquement publiques (à l'exception des procédures confidentielles de prévention telles que les conciliations ou des mandats *ad hoc*). Toute collecte systématique nécessite cependant une autorisation préalable. L'état complet d'endettement d'une société soumise à une procédure collective[63] est vendu 46,86 euros. Des informations relatives à une éventuelle procédure collectives peuvent également être achetées (1,56 euros par société). Le dossier complet est vendu 68,63 euros par entreprise.

En pratique, la collecte de ces données se fait manuellement. Contrairement au Royaume-Uni (*Companies House*) ou aux États-Unis (base *Pacer*©), les greffes français ne scannent pas l'entièreté des dossiers de faillite, ni ne sont dotés d'outils leur permettant une collecte automatisée du contenu détaillé des dossiers de défaillance ; les greffes disposent néanmoins d'outils informatiques d'identification des entreprises et des procédures. Les données stockées par les greffes renseignent sur l'ensemble du passif des entreprises défaillantes (salariés, État, fournisseurs). En revanche, il est impossible d'accéder aux dossiers de conciliation – contrairement, là encore, à ce qui se passe au Royaume-Uni (*cf.* procédures équivalentes). C'est évidemment regrettable dans la mesure où l'absence d'information individuelle sur les dossiers de conciliation empêche de juger des performances de cette même procédure introduite en 2005 et reprenant l'ancien règlement amiable de 1984.

En amont de la défaillance (*i.e.* pour les simples défauts ou retards de paiement), la centrale des risques de la Banque de France[64] constitue une autre source d'informations sur les incidents de paiements des entreprises françaises. Elle recense l'ensemble des crédits octroyés par les banques à leurs clients (au-delà de 25 000 euros). Dans le cadre d'échanges de données avec les Centrales des Risques d'autres membres de la zone euro, la Banque de France dispose également de cette information pour des sociétés non françaises. Il y a par ailleurs un projet de centrale des risques pour les pays de la zone euro en cours de discussion au sein de l'Eurosystème. Ce projet vise à disposer de données de crédits, principalement aux entreprises, harmonisées selon une logique prêt par prêt (et non emprunteur par emprunteur comme c'est le cas actuellement en France). Le calendrier de mise en œuvre du projet prévoit actuellement un démarrage en 2018. Toutefois, le lancement effectif du projet est soumis à l'approbation du Conseil des Gouverneurs.

[63] La procédure collective est une procédure de redressement ou de liquidation judiciaire organisant le règlement du paiement des créances d'une entreprise en cessation de paiement. Une entreprise en difficulté peut être concernée par : i) une procédure de sauvegarde, à l'initiative du chef d'entreprise, si l'entreprise n'est pas en état de cessation des paiements ; ii) un redressement judiciaire ou une liquidation judiciaire, si l'entreprise est en état de cessation des paiements. Une procédure collective interrompt le cours des poursuites, et fait l'objet de mesures de publicité. Définition : Infogreffe.fr

[64] https://www.banque-france.fr/fileadmin/user_upload/banque_de_france/Mission/Services_rendus/note115.pdf

Les informations gérées par la Centrale des Risques sont inclues dans le Fichier Bancaire des entreprises (FIBEN), ainsi que les informations issues des liasses fiscales évoquées plus haut, l'historique des incidents de paiement et enfin des informations générales sur le secteur d'activité, l'ouverture éventuelle d'une procédure collective ou encore le nombre de banques avec lesquelles l'entreprise est en relation (http://guide.fiben.fr, et par exemple : http://guide.fiben.fr/index.php?id=4878)

Grâce à ces informations, la Banque de France élabore un indicateur de risque de défaut pour environ 200 000 entreprises. Ces « scores » correspondent aux probabilités de défaillance d'une entreprise à un horizon de trois ans ; ils sont calculés à partir de séries statistiques portant sur des ratios propres à chaque secteur d'activité. Ces scores sont accessibles moyennant un abonnement.

4. Les données sur les crédits aux particuliers et le surendettement

Gunther Capelle-Blancard, Professeur à l'Université Paris 1 Panthéon-Sorbonne

Les données sur le crédit aux ménages sont une source précieuse d'informations, notamment pour étudier les problèmes de surendettement et d'instabilité financière.

1. Les dossiers des ménages surendettés

La Banque de France détient un fichier exclusif sur les ménages surendettés. Ces données sont particulièrement sensibles puisqu'elles relèvent de la vie privée. Le Fichier National des Incidents de Remboursement des Crédits aux Particuliers (FICP) a été créé en 1989. Il recense les incidents de paiement caractérisés liés aux crédits accordés à des personnes physiques pour des besoins non professionnels ainsi que les situations de surendettement. Au 31 décembre 2014, 2 677 712 personnes y était inscrites, pour 3 311 693 incidents de paiements. Ce fichier est dédié aux établissements de crédit afin de leur permettre de mieux apprécier les risques découlant de l'octroi de crédits aux particuliers. Il n'est pas accessible autrement.

2. Fichier positif des crédits bancaires

Il n'existe pas en France de registre national des crédits, dit « fichier positif »[65], qui recense l'ensemble des crédits aux particuliers. L'idée d'un tel fichier serait de fournir aux prêteurs une information détaillée de l'endettement de leur client et ainsi prévenir les situations de surendettement.

Les « fichiers positifs » existent dans de nombreux pays de l'Union européenne (outre la France, le Danemark, la Finlande et la Grèce n'y ont pas recours) ainsi qu'aux États-Unis. Dans la très grande majorité des cas, ces registres sont gérés par des institutions privées qui commercialisent ces informations, à l'exception de la Belgique où ce fichier est géré publiquement.

En France, la création d'un fichier positif fait l'objet de débats depuis de nombreuses années. En 2010, un comité chargé de préfigurer sa création a été mis en place dans le cadre de la loi « Lagarde » (art. 49, loi n°2010-737 du 1er juillet 2010) sur la réforme du crédit à la consommation. La Cnil a alors néanmoins émis un avis négatif.

Le projet de fichier positif a ensuite été repris dans le cadre de la loi de modernisation de l'économie (loi « Hamon »), mais cette disposition a été rejeté par le Conseil constitutionnel

[65] par opposition au fichier sur les incidents de paiement, dit « fichier négatif ».

jugeant qu'il « porte une atteinte au droit au respect de la vie privée qui ne peut être regardée comme proportionnée au but poursuivi ».

Le Conseil constitutionnel reconnaît l'intérêt d'un tel dispositif pour prévenir le surendettement, mais juge en fait son ampleur disproportionné. Un tel dispositif reviendrait en effet à ficher près de 12 millions de particuliers, pour une durée de plusieurs années.

L'idée ne semble toutefois pas complètement abandonnée puisque durant l'été 2014, le ministre des Finances, Michel Sapin, a mandaté le président du Comité consultatif du secteur financier (CCSF), Emmanuel Constans, pour qu'il examine s'il existe d'un point de vue juridique des marges de manœuvre permettant la mise en place d'un registre des crédits aux particuliers en France.

5. Les données de marché

Carine Romey, AMF

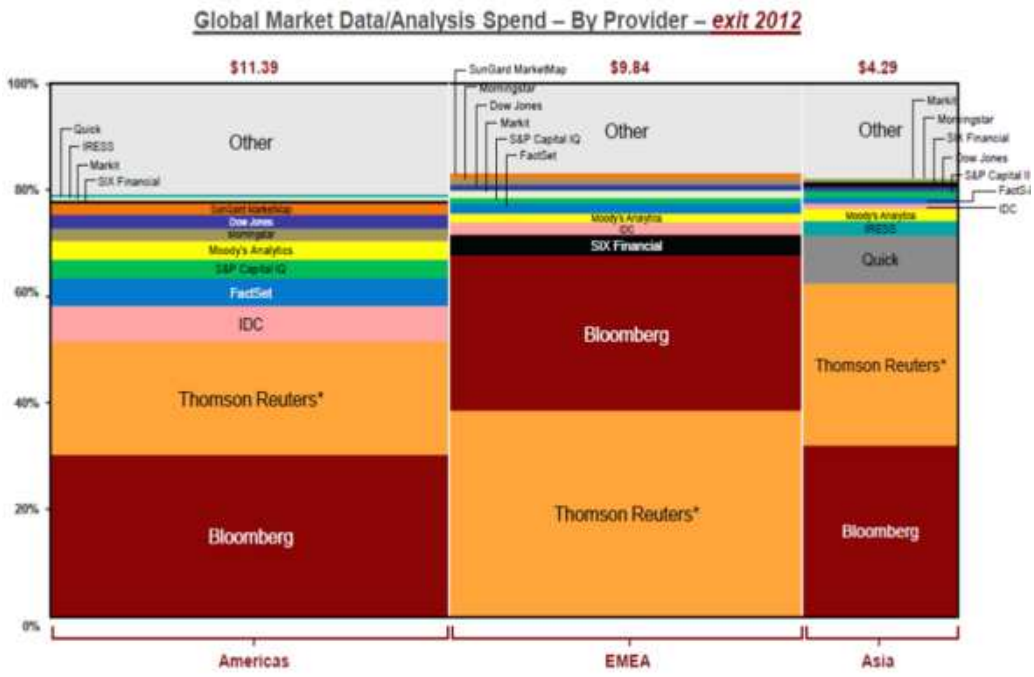
À la différence des données bancaires principalement issues de *reporting* règlementaires, les données de marché sont spécifiques car protéiformes tant au niveau de leur typologie que par le nombre d'« acteurs-producteurs » de ces données (les bourses, les agences de notations, les brokers, les autorités de supervision, les diffuseurs privés d'informations, etc.). En outre, la problématique de l'accessibilité des données de marché ne peut être dissociée de celle de « l'industrie des données de marché ».

Le terme « données de marché » (*market data*) désigne l'ensemble des données financières, mais aussi les applications associées. Ce terme générique recouvre généralement plusieurs dimensions :

- des classes d'actifs variées (action, titres de dette, titre du marché monétaire, taux de change, matières premières, produits dérivés, etc.) ;
- des types de données différentes sur l'ensemble des classes d'actifs (données de prix (cours coté ou OTC, fourchettes bid/ask, courbe de taux, etc.), données de référence (référentiels émetteur, rating, codification, lieu de cotation, etc.), données sur les opérations sur titres (émission de titres, détachement de dividende, date du prochain coupon, etc.) ;
- des sources multiples : données propriétaires, données issues des bourses, données générées par des entreprises privées spécialisées, données réglementaires issues de reporting, données publiques diffusées sur le web ;
- des modes de diffusion spécifiques : les professionnels de la finance utilisent essentiellement des données « en temps réel », ce qui suppose une infrastructure parfois très lourde, alors que les besoins en termes de recherche n'ont besoin que de données historiques.

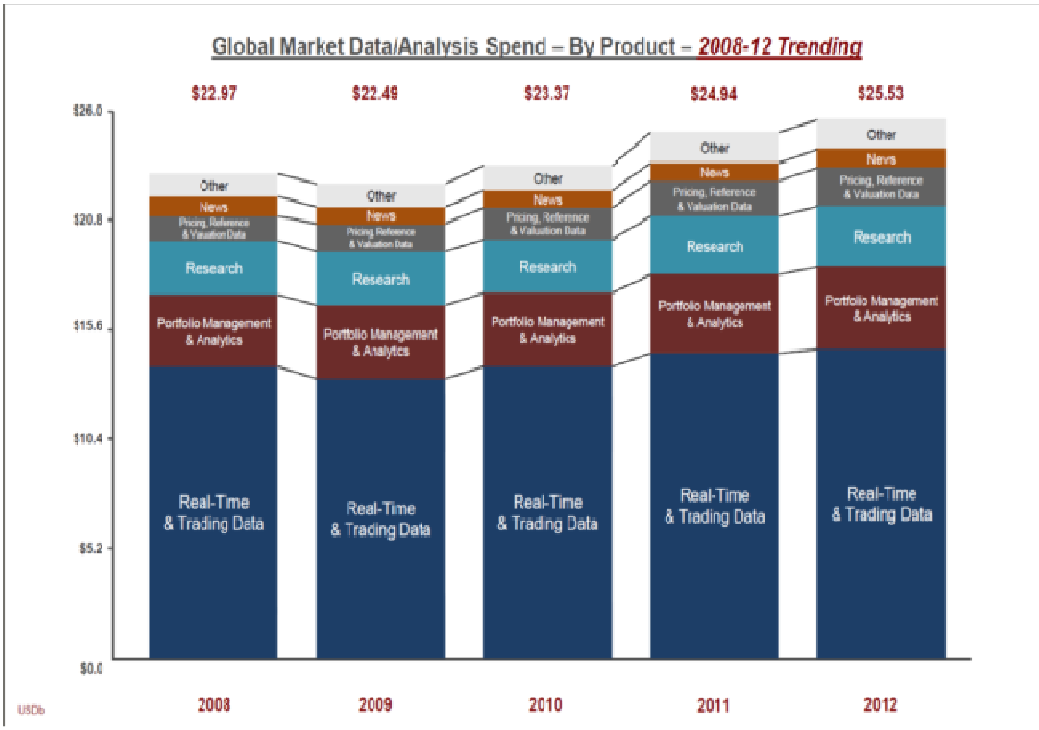
Au niveau mondial, l'industrie des « données de marché » représente près de 25 milliards de dollars, un marché quasi monopolistique, Thomson Reuters et Bloomberg se partageant en Europe près de 70% du marché (*cf.* figure 1, page suivante). La moitié des besoins se concentrent sur des données de transaction, notamment en temps réels, suivis par les outils analytiques offerts par les diffuseurs. Viennent ensuite, des besoins relatifs à l'analyse financière, la valorisation des actifs illiquides, les référentiels, les news…

Figure 1 : Part de marché des *data vendors* fin 2012 par zones géographiques



Source : Burton Taylor

Figure 2 : Évolution des dépenses par types de produits depuis 2008



Source : Burton Taylor

Le contexte actuel se caractérise au niveau européen par un élargissement important des missions des régulateurs (seconde directive MIF, EMIR, AIFM, vente à découvert, UCITV) qui, en renforçant les exigences en matières de transparence et de reporting[66] renforcent dans le même temps le besoin d'accès à de nouvelles sources d'informations financières.

Les acteurs financiers sont confrontés à des problématiques spécifiques dans l'usage de ces données qu'elle achète notamment celle des droits de licence et de la propriété des données.

Les acheteurs de données, ne sont pas propriétaires des données qu'ils achètent. En effet, le fournisseur de données donne un droit d'usage limité à certaines utilisations. Ces problèmes de coût et d'accès mettent fondamentalement en exergue un problème de droit entourant la propriété des données : il y a 15-20 ans les fournisseurs vendaient des plateformes techniques, alors qu'ils vendent aujourd'hui une propriété intellectuelle. Or on peut s'interroger sur leur droit de propriété intellectuelle sur les données qu'ils fournissent (données de marché, données agrégées, etc.). De plus, les fournisseurs ne se portent pas responsables des données qu'ils fournissent quand bien même ils s'en déclarent propriétaires

Dans ce contexte, **une réflexion est en cours au niveau européen sur l'accès aux données de marchés sur une base « commercialement raisonnable »**. Au travers de la nouvelle directive MIF, la commission souhaite corriger la situation de monopole dont profiteraient les bourses pour facturer les données de marchés à de prix excessifs. La directive révisée prévoit que les données de marché soient librement commercialisables par des entreprises privées indépendantes. Cependant l'accès à ces données doit avoir lieu dans des conditions « commercialement raisonnables ». Une des mesures consisterait à organiser plus de transparence et de favoriser la publication centralisée des tarifs. À ce stade, ces mesures s'appliqueront non seulement aux plateformes mais aussi aux APA (qui assurent la publication des transactions OTC), aux *consolided tapes* et aux internalisateurs systématiques.

[66] Conformité réglementaire, analyse des risques, multiplication des plates-formes de *trading* électroniques, analyse en temps réel toujours plus fine, valorisation de produits de gré à gré...

6. Les données sur les fonds de placement

Raphaëlle Bellando, Professeur à l'Université d'Orléans

La composition des portefeuilles des fonds de placement collectifs constitue une source d'information essentielle pour la recherche et, plus largement, pour la compréhension des mécanismes de formation des prix des actifs financiers. Une mise à disposition des données françaises pourrait faciliter une recherche pour l'instant largement limitée au cas des États-Unis.

L'étude du comportement des gérants de fonds de placements et de leurs choix de portefeuille est fondamentale à plusieurs égards :

- Une partie importante de l'épargne des ménages transite directement ou indirectement (par le canal de l'assurance-vie en France) par ces supports, et il est donc nécessaire de s'assurer que les actions des gérants de fonds vont dans le sens des intérêts des détenteurs de parts, et surtout, qu'elles ne perturbent pas le processus de formation des prix d'actifs (actions, taux d'intérêt).

- La littérature montre qu'il existe un problème d'agence : les sociétés de gestion ont intérêt à voir grandir leur actif sous gestion, et pour attirer les investisseurs, leurs gérants peuvent agir dans un sens défavorable à ces derniers, et d'une façon qui perturbe la formation des prix. Ainsi l'importance actuelle de la gestion déléguée ne serait pas automatiquement gage d'efficience des marchés et pèserait in fine sur la qualité de l'allocation des fonds sur les marchés financiers.

Les données de compositions des fonds permettent de répondre à des questions précises dérivant de ce problème d'agence[67]. Les styles de gestion affichés sont-ils respectés ? Y-a-t-il des phénomènes de Window Dressing ? Les gérants de fonds prennent-ils des risques spécifiques dans le cadre de la compétition qu'ils se mènent pour attirer les investisseurs ? Les gérants agissent-ils de manière indépendante ?

Cette dernière question, celle du mimétisme[68] des gérants de fonds est assez emblématique de la littérature académique et illustre bien l'intérêt de disposer de données de portefeuilles des fonds. Les premiers travaux empiriques sur le mimétisme des fonds de placement datent des années 1990 et visaient essentiellement à voir si de tels comportements pouvaient affecter l'efficience des marchés, ou renforcer la volatilité des marchés en période de crise. Le dernier

[67] L'existence de biais comportementaux des gérants de fonds est aussi parfois traitée, elle ne dérive pas du problème d'agence, mais aurait des effets sur les prix d'actifs. Dans cet ordre d'idée, un papier récent a montré que les gérants de fonds avaient une tendance à privilégier des titres d'entreprises dirigées par des hommes.

[68] On parle de mimétisme si un investisseur fonde ses décisions d'achat ou de vente sur celles des autres, délaissant potentiellement ses propres informations.

rapport sur la stabilité financière globale du Fonds Monétaire International[69] remet cette question au goût du jour. Il montre que les décisions des fonds ont un impact sur les prix au moins pour les actifs les moins liquides. Il montre surtout une montée importante du niveau de mimétisme entre fonds américains, ainsi l'indicateur de mimétisme pour les fonds de détail aurait été multiplié par cinq entre 2006 et 2014 (par quatre pour les fonds à clientèle institutionnelle). Il s'inquiète enfin des possibilités de *run* et de vente en urgence de parts des fonds, susceptibles d'alimenter une crise systémique.

Les pratiques et règles en matière d'obligation de publication des détentions des fonds sont très diverses dans le monde. En 2013[70], seule l'Inde obligeait ses fonds à publier mensuellement la composition de l'actif des fonds, tandis que la plupart des fonds au Danemark et en Norvège, et les trois-quarts des fonds américains fournissent ces données à un rythme mensuel aux institutions qui le demandent.

En France, les sociétés de gestions sont tenues de communiquer aux investisseurs qui en font la demande un document détaillant la composition de leur actif lors de l'élaboration de leur rapport d'activités semestriels.

Dans le même temps, les fonds de droit français sont tenus de déclarer la composition de leur portefeuille au titre par titre.[71] La Banque de France collecte ainsi pour l'Euro-système depuis 1998 des données extrêmement détaillées à une fréquence mensuelle pour les fonds monétaires depuis 1998, pour les autres fonds les données ont été collectées à un rythme trimestriel entre 1998 et 2011, et à fréquence mensuelle depuis. Or ces données relèvent du règlement européen n°2533/98. Autrement dit, elles peuvent être transmises pour les besoins de la recherche à la condition que le déclarant ne puisse être identifié. À titre d'exemple, les données concernant les fonds actions ont été mises à disposition en 2005 après anonymisation des fonds et des titres. En un sens, cette mise à disposition représentait une opportunité exceptionnelle, en termes de coût, de qualité et d'exhaustivité. Les travaux américains, qui dominent de façon écrasante la littérature, reposent sur l'acquisition d'une base de données très onéreuse : la base CDA/Spectrum commercialisée aujourd'hui par Thomson Reuter. Le cas américain semble d'ailleurs montrer l'intérêt des données de *reporting* obligatoires relativement aux données collectées pour des bases de données commerciales, sur la base du volontariat.[72] Enfin, des travaux européens ont été réalisés sur des échantillons de taille très réduite : Walter et Weber (2006) et Frey, Herbst & Walter (2014) utilisent des données semestrielles (collectées à la main) de fonds allemands orientés sur leur marché national.

Pour autant la question de l'anonymisation des données est extrêmement problématique. Anonymiser les fonds ne permet pas la prise en compte du style de gestion et de l'orientation géographique ou sectorielle des fonds. L'anonymisation des titres est encore plus délicate.

[69] Ce point fait l'objet du chapitre 3 du Global Financial Stability Report, Avril 2015. http://www.imf.org/external/pubs/ft/gfsr/2015/01/pdf/c3.pdf

[70] Global Fund Investor Report, Morningstar Fund Research, 15 mai 2013.

[71] Les collectes concernent aussi les autres composantes de l'actif ou du passif, et diverses informations sur les parts et leurs détenteurs.

[72] Schwarz et Potter (2014) « Revisiting Mutual Fund Disclosure » http://ssrn.com/abstract=2093688 montrent par exemple que les portefeuilles déclarés à la SEC font apparaitre plus de *Window Dressing* que lorsqu'ils sont issus des bases de données commerciales.

Elle rend impossible la prise en compte des caractéristiques des titres (secteur, dynamique temporelle) pourtant essentielle pour répondre aux questions évoquées ci-dessus.

À la fin des années 2000, une discussion s'était engagée au sein du service responsable de ces données, sur leur mise à disposition en ligne pour les chercheurs. Il était question d'anonymiser les fonds, mais pas les titres, ce qui aurait permis de croiser le fichier avec des informations complémentaires sur les titres (secteur, dynamique de prix...). Faute de suivi, ce projet n'a pas abouti.

Un débat existe dans la littérature sur le fait d'exiger des fonds une publicité quant à la composition de leur portefeuille. Le fait qu'on puisse copier les stratégies des fonds performants actifs pourrait leur enlever le bénéfice de la rente qu'ils tirent de leur acquisition d'information. De plus, les investisseurs pourraient jouer contre les fonds : en vendant, par exemple, des titres en portefeuille ils feraient baisser leur cours et pèseraient sur la performance du fonds. Une publication permettrait à l'opposé de résoudre en partie les problèmes d'agence en forçant les fonds à mieux s'aligner sur leur stratégie annoncée, ou en révélant des comportements inefficients (prises de risque, mimétisme, *window dressing*).[73]

Puisqu'il ne s'agit pas ici d'exiger la publicité intégrale de ces données mais d'en favoriser l'accès aux chercheurs, les « effets négatifs » consistant à copier ou jouer contre un fonds ne peuvent apparaître dans ce cas.[74] À l'opposé, un effet « discipline de marché » pourrait jouer si des études portant sur d'éventuelles inefficiences de la gestion d'actif étaient publiées, sans aucune référence aux noms des entités étudiées. On peut même imaginer que ces travaux soient utilisés par le régulateur (l'AMF) pour améliorer sa surveillance des fonds.

Ainsi, il conviendrait probablement d'autoriser un accès sans restriction aux données dans le cas de travaux susceptibles d'éclairer le régulateur. Cela pourrait se faire par un contrat de confidentialité engageant l'acquéreur à ne pas transmettre les données et à ne pas publier d'informations nominatives sur les fonds. Ou par une intégration temporaire dans l'organisme de régulation.

[73] Ces effets sont détaillés par exemple dans Ge & Zheng, The Frequency of Mutual Fund Portfolio Disclosure, http://ssrn.com/abstract=557186

[74] D'autant que l'argument suppose une divulgation immédiate des informations, ce qui ne serait pas nécessaire dans le cadre de travaux de recherche.

7. Epargne retraite supplémentaire en France : quelles données, quelle transparence ?

Jérôme Glachant, Professeur à l'Université Paris 1 Panthéon-Sorbonne

La générosité déclinante des systèmes de retraite par répartition implique un recours grandissant des individus à des produits d'épargne retraite privés, qu'ils soient souscrits à titre privé ou professionnel. Les fonds placés sur ces supports ont en commun d'être cumulés durant la vie active puis décumulés lors de la retraite, avec des contraintes importantes sur le calendrier des flux (produit « tunnel » sans déblocage anticipé, sortie en rentes, …) et ces produits sont assortis de conditions fiscales avantageuses à l'entrée ou à la sortie.

En France, le développement de l'épargne retraite a été favorisé par la loi de réformes des retraites du 21 août 2003. Ces produits définissent les régimes de retraite dits supplémentaires et font l'objet d'une enquête annuelle réalisée par la DREES : l'enquête retraite supplémentaire facultative.[75]

Les organismes gestionnaires de l'épargne retraite sont sous la supervision de l'Autorité de contrôle prudentiel et de résolution (ACPR) et/ou de l'Autorité des marchés financiers (AMF). Ces deux organismes disposent ainsi d'informations microéconomiques sur la gestion financière des dispositifs, notamment en matière de performance et de niveau des frais de transaction.

Pour préparer sa retraite, l'individu doit effectuer des choix d'allocation parmi les différents supports, et une information complète et transparente sur les dispositifs serait extrêmement utile. Pourtant, cette transparence est faible et l'objet de cette note est d'examiner comment elle pourrait être améliorée. Dans cette perspective, l'exemple américain est intéressant, car il illustre la manière dont les autorités publiques peuvent se structurer en plateforme d'accès aux données.

La première section décrit les dispositifs d'épargne retraite en France. La deuxième section souligne les difficultés européennes en matière de transparence sur les performances et présente le cas de la société américaine BrightScope. La troisième section analyse les produits français en matière d'information et formule quelques recommandations.

1. L'épargne retraite supplémentaire

Les produits disponibles

[75] http://www.drees.sante.gouv.fr/l-enquete-retraite-supplementaire-facultative

Les produits concernés peuvent être des produits facultatifs souscrits à titre individuel (principalement le Plan d'Epargne Retraite Populaire (PERP), les contrats « Madelin » destinés aux travailleurs non-salariés, les plans Préfon et Complément retraite mutualiste (COREM) destinés aux fonctionnaires) ou bien des produits dits d'entreprise, souscrits ou proposés collectivement (principalement le Plan d'Epargne Retraite Collectif (PERCO), les contrats collectifs « Article83 » ou « Article39 »).

D'un point de vue réglementaire, ces produits se répartissent en trois catégories avec des conséquences sur la gouvernance, le cadre prudentiel et la transparence :

1. Les produits d'assurance retraite relèvent du Code de l'assurance et sont souscrits par un contrat de groupe auprès d'un organisme d'assurance : principalement les PERP, contrats « Madelin », contrats « Article 83 », contrats « article 39 ».

2. Les produits d'épargne salariale, régis par le Code du travail, sont gérés par des sociétés de gestion sous forme principalement de Fonds Communs de Placement en Entreprise (FCPE) : les PERCO.

3. Les produits spécifiques réservés à des catégories particulières et fonctionnant par point : COREM, Préfon, COREL,…

Ces produits se distinguent également par la nature des garanties offertes aux souscripteurs :

- Les produits à cotisation définies sont les plus nombreux (PERP, PERCO, contrats « Madelin », « Article 83 »). Le niveau de la prestation (rente servie) n'est pas fixé par le contrat et l'épargne retraite est ainsi exposée aux risques financiers avec une individualisation du compte.

- Les produits à prestations définies sont des retraites d'entreprise de type « Article 39 ».

- Les produits à points (COREM, Préfon, COREL,…) organisent une mutualisation parmi les souscripteurs avec un équilibre comptable obtenu par une valorisation des droits par point.

L'enquête Retraite supplémentaire facultative de la DREES est une enquête exhaustive effectuée auprès de tous les intermédiaires proposant ces produits (sociétés de gestion en épargne salariale, sociétés d'assurance, mutuelles et institutions de prévoyance). Pour chaque produit, l'enquête collecte des données sur le nombre d'adhérents et de cotisants, leurs caractéristiques, le montant des cotisations et des prestations ainsi que les encours. Selon la dernière enquête[76] de la DREES, le montant total des encours (provisions mathématiques) s'élevait à 185 Mds d'euros fin 2013 pour 3 millions d'adhérents.

Choisir parmi les produits

[76] http://www.drees.sante.gouv.fr/IMG/pdf/retraite-edition-2015.pdf

L'enquête de la DREES n'a pas pour objet de fournir des informations concernant la performance financière de ces produits. Cette performance est d'autant plus difficile à évaluer qu'elle concerne des produits d'épargne à long terme, en fait des plans d'épargne à « effet tunnel » dont le taux de rendement interne pour chaque adhérent n'est connu qu'*ex post*, c'est-à-dire une fois l'ensemble des flux entrants (cotisations) et sortants (prestations) effectués. Cela n'empêche évidemment pas le calcul d'une performance comptable mesurée « au fil de de l'eau », et qui peut orienter les ménages dans leur décision de souscription, de contribution voire de liquidation pour chaque produit, qu'il soit à cotisations définies ou par point.

Pour se constituer un complément de retraite, l'individu a en effet la possibilité d'arbitrer parmi les différents supports. Cet arbitrage individuel s'effectue à plusieurs niveaux :

- Entre les produits d'épargne retraite « supplémentaires » et les autres supports d'épargne, qu'ils soient immobiliers ou financiers (contrats d'assurance-vie, Plan d'épargne en action, livrets, détention directe…).

- Parmi les produits d'épargne retraite « supplémentaires ». Un même individu peut ainsi disposer d'un PERP, d'un contrat « Madelin », s'il est non salarié, ou d'un contrat COREM ou Préfon s'il est fonctionnaire. S'il est salarié du privé, il peut, à titre individuel, abonder le PERCO proposé par son entreprise ou effectuer des versements libres sur le contrat « Article 83 ».

Pour que ces choix d'allocation soient pertinents, il est crucial que des informations comparables sur la performance des différents supports d'épargne retraite, qu'ils soient proposés dans l'entreprise ou souscrits individuellement, soient disponibles. Même si on doute de la capacité de l'individu à réaliser de la planification financière en vue de sa retraite, ces informations peuvent être traitées et mise en forme par des nouveaux intermédiaires pour fournir du conseil indépendant. Dans cette perspective, la France et l'Europe sont en retard sur les États-Unis

2. Transparence du marché de l'épargne retraite : l'Europe et les États-Unis

L'opacité européenne

Dans son dernier rapport[77], l'organisation non gouvernementale « *Better finance for all* » souligne l'opacité du marché européen des produits d'épargne retraite en ce qui concerne la mesure de la performance et des coûts d'intermédiation. Pour « *Better finance for all* », les performances de ce marché ne sont connues ni des clients, ni des régulateurs, que ce soient des performances moyennes par types de support ou individuelles par produits et

[77] Better Finance for all (2014), "Pension Savings : The Real Return", 2014 Edition, http://www.betterfinance.eu/fileadmin/user_upload/documents/Research_Reports/en/Pensions_Report_2014_FINAL_-_EN_FOR_WEB.pdf

gestionnaires. L'OCDE publie régulièrement[78] des taux de rendement moyen des plans d'épargne retraite, qui ne peuvent servir de guide opérationnel aux choix des individus.

Cette opacité nuit au développement et à la performance de ce marché. Ainsi, la Commission européenne fournit chaque année un tableau de bord des différents marchés[79] du point de vue des consommateurs, et parmi 52 marchés de biens et services, celui des produits d'investissement et d'épargne retraite obtient le score le plus faible en termes de confiance dans le marché, de comparabilité des offres, et de satisfaction générale.

Cette insatisfaction des consommateurs n'est pas exclusivement due au manque de transparence du marché, elle est aussi inhérente à des classes de produits « complexes ». Il n'empêche qu'une information transparente et facile d'accès permettrait l'émergence d'une nouvelle classe d'intermédiaires, qui pourraient apporter aux épargnants des conseils à bas coût. Dans cette perspective, il est pertinent de se tourner vers les États-Unis, où le marché de l'épargne retraite est considérablement plus développé[80].

États-Unis : le cas BrightScope

Aux États-Unis, BrightScope[81], une entreprise créée en 2008, fournit gratuitement au public des informations, des conseils et une échelle de notation sur plus de 50 000 plans d'épargne retraite d'entreprise de type 401(k), plans à cotisations définies avec abondement joint employé-employeur, auxquels peuvent se comparer nos PERCO et « article 83 » . BrightScope collecte les informations de base auprès du *U.S. Department of Labor*, régulateur auprès duquel les employeurs proposant des plans 401(k) sont tenus annuellement de remplir le formulaire 5500 et de déposer un audit annuel pour les plans de plus de 100 participants. Pour chaque plan, le formulaire 5500 contient des informations sur la gouvernance du plan, ces dépenses, les options d'investissement proposées et les bilans comptables associés. Ces informations rendent transparents les coûts de gestion et la nature de la relation entre l'initiateur du plan (l'entreprise) et les gestionnaires des fonds (*mutual funds* de type OPCVM). Les informations obtenues auprès du *Department of Labor* sont complétées par des données en provenance d'autres organismes publics comme la *Securities and Exchange Commission* (SEC) ou la *Financial Industry and Regulatory Authority* (FINRA). En rendant disponibles ces informations et en les standardisant, BrightScope permet un point d'accès à des données diverses concernant l'épargne retraite en entreprise.

Cette activité a été rendue possible par l'administration Obama, qui a favorisé à partir de 2009 le principe du *Open Government*, favorisant l'accès du public aux données collectées par les

[78] Voir par exemple : http://www.oecd.org/pensions/PensionMarketsInFocus2013.pdf
[79] http://ec.europa.eu/consumers/consumer_evidence/consumer_scoreboards/10_edition/index_en.htm
[80] Towers Watson réalise chaque année une étude sur le marché mondial de l'épargne retraite : fin 2014, son volume est de 36199 Mds USD, dont 22000 Mds USD pour les Etats-Unis (contre seulement 171 Mds USD pour la France). http://www.towerswatson.com/en/Insights/IC-Types/Survey-Research-Results/2015/02/Global-Pensions-Asset-Study-2015
[81] https://www.brightscope.com/ A propos de l'expérience BrightScope, on peut consulter l'article écrit par ces deux fondateurs : http://beyondtransparency.org/chapters/part-2/from-entrepreneurs-to-civic-entrepreneurs/

organisations publiques de régulation et de contrôle. En 2007, Ryan Alfred et Mike Alfred, les fondateurs de BrightScope, étaient partis du constat que la transparence des données publiques ne suffisait pas dès lors que des obstacles techniques empêchaient l'accès *effectif* et à moindre coût. Le principe de l'*Open Governement* ou du *Government as a platform* nécessite que l'accès aux données soit facilité d'un point de vue technique et réglementaire. C'est précisément ce qu'a facilité l'administration à partir de 2009.

Retour en France

L'exemple de BrightScope et des plans d'épargne retraite est riche d'enseignement pour la situation en France, et peut-être en Europe. Comme nous l'avons déjà indiqué, le résident français actif a à disposition une large gamme de produits dédiés d'épargne retraite, qu'ils soient individuels ou collectifs et parmi lesquels il peut *choisir*. Il ne dispose malheureusement pas d'une plateforme comparable à BrightScope. La société Profideo édite annuellement des guides comparatifs[82] sur la retraite collective (Article 83, Article 39), les PERP et les contrats « Madelin ». Sans mettre en cause la qualité de ces sites, on ne peut que constater l'écart avec BrightScope ; les données disponibles ne semblent pas provenir d'organismes publics mais sont collectées directement auprès des gestionnaires. Quels sont les obstacles qui, en France, empêchent l'émergence d'intermédiaires de données tel BrightScope dans le domaine de l'épargne retraite ?

Le premier obstacle est le cadre réglementaire de chaque produit d'épargne retraite et les obligations de transparence des gestionnaires vis-à-vis de leurs clients, du public et des organismes régulateurs. Ce cadre est hétérogène selon les produits concernés, même lorsqu'ils sont proches sur leurs caractéristiques (abondement volontaire de l'individu, cotisations définies)

Il en est ainsi des contrats « Madelin » et des PERP, qui sont tous deux des contrats d'assurance de groupe souscrits par une association loi 1901 auprès d'un organisme assureur. Les individus adhèrent ensuite volontairement et individuellement à cette association pour bénéficier du contrat. L'association souscriptrice défend les intérêts des adhérents et supervise la gestion de l'organisme assureur, qui a des obligations de transparence vis-à-vis de l'association. Ces obligations sont parfaitement définies pour le PERP (voir le décret n° 2011-1635 du 23 novembre 2011 et l'arrêté associé), mais moins pour le contrat « Madelin ». Pour le PERP, l'article L144-2 du Code des Assurances prévoit que le rapport annuel de gestion soit transmis à l'ACPR, qui tient également un registre des associations dites Groupement d'Epargne Retraite Populaire (GERP) souscriptrices de PERP. L'ACPR centralise ainsi l'intégralité de l'information concernant les PERP : caractéristiques des plans, résultats financiers, coûts de gestion associés. Certains GERP et/ou organismes d'assurance rendent publiquement disponibles les rapports de gestion[83], qui sont de toutes façons fournis aux

[82] Ces informations sont en partie disponibles sur le site http://www.lesdossiers.com/. On trouve également des éléments comparatifs en consultant le site http://www.francetransactions.com/perp/comparateur-perp.html.
[83] Voir par exemple le rapport de gestion de Batiretraite de SMABTP Vie : http://www.smabtp.fr/SGM/upload/docs/application/pdf/201307/rapport_annuel_batiretraite_perp_2012_.pdf

adhérents des plans. Sur son site, l'ACPR fournit la liste des GERP mais ne propose pas les rapports annuels de gestion. Pour les contrats « Madelin », les obligations de transparence sont peu codifiées dans la loi en comparaison des PERP, et les associations souscriptrices moins reconnues. On constate ainsi que pour deux produits « individuels » qui sont de proches substituts, les obligations légales sont différemment explicitées, ce qui freine la transparence.

Les contrats « article 83 » (art. 83 du CGI) sont des contrats de groupe régis par le Code des Assurances, qui relèvent de la retraite professionnelle supplémentaire (art. L143). Le contrat est ainsi souscrit de manière facultative par l'employeur mais est à adhésion obligatoire pour tous les salariés avec un abondement obligatoire de l'employeur et libre de l'employé. Comme pour le PERP et le « Madelin », ces plans sont à cotisations définies et la réglementation prévoit des possibilités de transfert d'un contrat sur l'autre, ce qui suppose une uniformisation réglementaire dont on est encore éloigné. L'article L143-2 du Code des assurances prévoit l'existence d'un comité de surveillance du contrat Article 83 « chargé de veiller à la bonne exécution du contrat et à la représentation des intérêts des adhérents » (voir Article R-143-2 du Code des Assurances). Les obligations de transparence vis-à-vis du souscripteur, des adhérents et bénéficiaires et de l'autorité de contrôle sont beaucoup plus réduites que pour les PERP. Le rapport annuel et les comptes de gestion sont « remis *à leur demande* aux souscripteurs (…) et sont tenus à la disposition des adhérents et bénéficiaires » (Art. L143-6 du Code des Assurances). Il en est de même pour le rapport tri-annuel indiquant la « politique de placement et les risques techniques et financiers ». Le dépôt de ces rapports auprès de l'ACPR n'est pas obligatoire. Très peu d'informations est donc disponible sur la performance financière de ces produits.

Le Plan d'Epargne Retraite Collectif (PERCO) est un dispositif collectif, ouvert à l'ensemble des salariés d'une entreprise avec une adhésion facultative et des versements volontaires, qui peuvent être complétés par l'entreprise. À la différence des contrats de groupe souscrits auprès d'organisme d'assurance, le PERCO est régi par le Code du travail dans la catégorie de l'épargne salariale et donne accès à des fonds communs de placement, en général des FCPE, au nombre minimal de 3. La gestion administrative est confiée à un organisme de tenue de compte et de conservation (TCCP) et la gestion financière à une ou des sociétés de gestion qui gère(nt) l'actif des FCP. Chaque FCPE est doté d'un comité de surveillance avec des représentants des salariés. D'un point de vue réglementaire, le PERCO relève du champ de compétence de l'AMF, car il constitue une enveloppe pour accéder à des OPCVM, enregistrés auprès de cet organisme. En particulier, l'AMF rend disponible dans sa base GECO le Document d'informations clés pour l'investisseur (DICI) pour chaque fonds agréé et distribués en France. Ce prospectus inclut des informations sur la performance financière et le niveau des frais de gestion. Cette information est précieuse pour chaque fonds, mais ne permet pas d'identifier les relations entre les fonds figurant dans un même PERCO, ainsi que les niveaux des commissions associées aux PERCO.

Nous n'évoquons pas ici les plans spécifiques de type COREM, Préfon,… organisés par l'article L-441 du Code des Assurances qui fonctionne par point sur le principe de la

capitalisation provisionnée. La comparaison avec les autres dispositifs est moins aisée car ces plans ne sont pas à cotisations définies.

3. Quelles recommandations ?

La réglementation et les pratiques ne permettent pas la mise à disposition publique des informations relatives à la gestion des produits d'épargne retraite supplémentaire. Des données microéconomiques sont collectées par l'ACPR et l'AMF, mais elles restent éparses et ne couvrent pas l'intégralité du marché. Cette situation ne favorise pas un fonctionnement efficace du marché de l'épargne retraite qui a besoin d'informations pour alimenter le choix des individus. Ce choix complexe requiert l'émergence de nouveaux intermédiaires qui, à l'instar de BrightScope aux États-Unis, pourraient s'alimenter des données publiques en contribuant aussi à leur standardisation. Pour aller dans cette direction, deux obstacles doivent être levés.

Le premier obstacle est relatif à l'hétérogénéité de la réglementation s'appliquant à chacun des produits. Les exigences de transparence et d'information vis-à-vis des adhérents et du régulateur sont ainsi élevées sur le PERP, et plus réduites pour les « Madelin », « Article 28 » ou PERCO. Ces produits à cotisations définies étant forts substituts, il est nécessaire d'aller vers une uniformisation des règles de transparence en s'alignant sur le fonctionnement des PERP/GERP avec une centralisation de l'information vers l'ACPR (assurance retraite) et l'AMF ou le Ministère du Travail (épargne salariale). La création d'un prospectus uniformisé annuel pour ces produits d'épargne retraite, à l'image de celui existant pour les OPCVM, serait très utile. En particulier, il est essentiel que les épargnants aient accès aux données relatives à la performance et au montant des frais.

Le second obstacle est lié à la pratique des organismes publics qui ont fait beaucoup d'effort pour délivrer des informations générales[84], mais qui fournissent moins d'informations microéconomiques. La pratique pourrait là aussi s'inspirer des OPCVM puisque l'AMF fournit avec sa base GECO l'intégralité des prospectus OPCVM. L'AMF et l'ACPR pourraient ainsi unir leurs efforts pour délivrer les informations microéconomiques sur les produits d'épargne retraite.

[84] Ainsi, l'ACPR, l'AMF et la Banque de France se sont unis pour créer le portail d'information Assurance-Banque-Epargne Info Service (http://www.abe-infoservice.fr/accueil.html).

8. Les données nécessaires au calcul du risque systémique

Chistophe Hurlin, Professeur à l'Université d'Orléans

La crise financière de 2008 a souligné l'importance du risque systémique qui nait de l'interdépendance entre les situations et les décisions individuelles; de sorte que la prise de risque d'une institution financière est capable d'affecter l'ensemble du système financier. La mesure de ce type de risque, et en particulier la contribution qu'y apporte une institution donnée, est un problème complexe. Il intéresse vivement les régulateurs puisque que la mesure du risque systémique est un préalable à la maîtrise du phénomène de contagion et à la mise en place des régulations macro-prudentielles. On distingue trois grandes sources de risque systémique : (i) la prise de risque systémique, (ii) les phénomènes de contagion et (iii) les phénomènes d'amplification. L'analyse de la prise de risque systémique consiste à analyser les raisons réglementaires ou économiques qui poussent les institutions financières à prendre des positions très risquées et souvent corrélées. Les phénomènes de contagion correspondent plus à une vision « passive » du risque systémique et recouvrent l'ensemble des phénomènes tendant à expliquer la transmission ex post de pertes financières d'une institution financière à l'autre. Enfin, les mécanismes d'amplification expliquent comment de petits chocs sur le système financier peuvent engendrer in fine des pertes de grandes ampleurs et des faillites en cascade (l'exemple typique étant celui du mécanisme des ventes forcées ou *fire sales*).

Dans la littérature académique et la pratique réglementaire[85], deux principales approches, correspondant chacune à un type de données particulières, peuvent être distinguées.

L'approche la plus fréquemment utilisée dans la littérature académique consiste à proposer des mesures globales, visant à englober potentiellement ces différentes sources de risque systémique. Ces mesures sont généralement fondées sur des données de marché (prix d'actif, primes de CDS, etc.) et parfois, des données comptables, disponibles publiquement (permettant par exemple de calculer le levier). L'exemple emblématique de ces mesures globales est la SRISK.[86] Ces mesures globales peuvent être calculées en temps réel et permettent un suivi régulier de l'évolution du risque systémique. Mais l'utilisation de données de marché suppose que l'hypothèse d'efficience soit vérifiée et les interdépendances mesurées dans les prix d'actifs (notamment des actifs bancaires) reflètent effectivement les interdépendances existant entre les portefeuilles d'activité des institutions financières.[87]

[85] Benoit, S., Colliard, J.E., Hurlin, C. et C. Pérignon, (2015). Where the Risks Lie: A Survey on Systemic Risk, Document de travail.

[86] Voir Acharya, V. V., R. Engle, et M. Richardson, (2012). Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks, *American Economic Review*, 102(3), 59-64 ; Brownlees, T. C., et R. F. Engle, (2015), SRISK: A conditional capital shortfall index for systemic risk measurement, Document de travail.

[87] Voir de Bandt O, J.C Heam, C. Labonne et S. Tavolaro, 2013, Measuring the systematic risk in a post crisis word, Débats économiques et financiers n°6, ACPR. Tavolero, S. et F. Visnovsky, (2014), What the information content of the SRISK mesures as supervisory tool, Débats économiques et financiers n°10, ACPR.

La deuxième approche consiste plus directement à identifier une source de risque systémique et à en déduire une mesure spécifique. La mesure des effets associés à l'une ou l'autre des trois sources systémiques évoquées plus haut requiert l'identification et la quantification des « liens », définis au sens large, pouvant exister entre les institutions financières. Ces liens peuvent être représentés sous la forme de matrices d'exposition croisées, expositions mesurées tant du côté de l'actif que du côté du passif des bilans bancaires. Ils peuvent être aussi déduits de la composition du portefeuille d'actifs de différentes banques à partir d'une mesure statistique de similarité.

Dans l'état actuel des règles de confidentialité, la mise à disposition de ces données nécessaires à cette deuxième approche et qui sont en France recueillies par l'ACPR, n'est généralement pas possible[88] (voir règles de confidentialité de l'ACPR). Même si dans l'idéal, l'accès à ce type de données devrait s'effectuer au niveau européen voire mondial, on voit bien l'intérêt pour l'organisme régulateur d'accentuer et de développer les coopérations scientifiques avec des chercheurs extérieurs dans le cadre de ces travaux. Ceci illustre parfaitement la recommandation n°12.

[88] Toutefois, dans un souci de transparence des marchés, certaines données agrégées sur les portefeuilles d'actifs sont rendues publiques après un certain délai, comme par exemple celles concernant les expositions des banques européennes utilisées dans le cadre des stress tests menés par l'European Banking Authority (EBA).

9. Le reporting pays par pays des établissements financiers : de nouvelles données disponibles, mais difficiles à comparer et analyser

Lucie Watrinet et Thomas Dauphin, CCFD Terre solidaire

En 2013, un pas majeur a été fait vers la transparence des établissements financiers (établissements de crédit, compagnies financières et compagnies financières holding mixtes, et entreprises d'investissement), au niveau français et européen : les établissements financiers européens sont en effet tenus, depuis 2014, de rendre publiques des informations concernant leurs activités et les impôts qu'ils payent dans chacun des pays où ils sont présents.

Cette nouvelle obligation de transparence, appelée « reporting pays par pays » est une recommandation clé des organisations de la société civile qui luttent pour plus de justice fiscale. En effet, la publication obligatoire de ces informations permet de donner une idée plus précise de l'usage qui est fait par les établissements financiers des paradis fiscaux et notamment de vérifier si elles y ont des activités réelles ou si elles les utilisent pour délocaliser artificiellement des bénéfices, échapper à l'impôt ou pour créer des véhicules de gestion de certains actifs risqués et contourner ainsi leurs obligations réglementaires.

Deux initiatives parallèles ont contribué à l'instauration du reporting pays par pays pour les établissements financiers : la directive européenne du 26 juin 2013 sur les exigences de fonds propres[89] (CRD IV) et la loi française du 26 juillet 2013 de séparation et de régulation des activités bancaires[90]. Cette dernière est l'aboutissement d'un projet de loi déposé le 19 décembre 2012 adopté au terme d'une Commission Mixte Paritaire au Parlement et incluant un éventail de mesures pour réduire les risques systémiques et l'aléa moral des banques : séparation des activités de détail et de spéculation, renforcement du pouvoir des autorités de contrôle, encadrement des rémunérations, etc.

Il est indéniable que la position proactive de la France en la matière a contribué à influer positivement le cours des négociations à Bruxelles. Les standards de transparence requis par les deux législations sont d'ailleurs identiques, à savoir :

1) la dénomination des implantations et la nature de leur activité ;
2) le chiffre d'affaires ;
3) les effectifs en équivalent temps plein ;
4) le résultat avant impôt ;
5) le montant des impôts sur les bénéfices ;
6) les subventions publiques reçues.

[89] Article 89 de la directive 2013/36/UE du 26 juin 2013
[90] Article 7 de la loi bancaire n° 2013-672 du 26 juillet 2013

La loi bancaire dispose que ces données sont mises à disposition publiquement et gratuitement, **en deux temps :**

- les catégories 1), 2) et 3) dès 2014 pour l'exercice 2013 ;
- toutes les catégories à partir de 2015 pour l'exercice 2014.

C'est sur la base de la première série de données que la Plateforme Paradis Fiscaux et Judiciaires (www.stopparadisfiscaux.fr/) a réalisé en novembre 2014 un rapport[91] et émis des recommandations.

Le décret de la loi bancaire[92] publié un mois et demi après notre rapport a répondu à deux de nos préoccupations : désormais les banques sont obligées de publier ces nouvelles informations **sous forme de deux tableaux distincts** (le premier avec la liste des filiales et activités, le second avec les chiffres pays par pays) et **dans leur rapport annuel**, ce qui facilite la recherche de ces données.

Néanmoins, les limites suivantes subsistent :

- **Les formats utilisés par les banques pour publier ces données rendent leur exploitation difficile.** Les données ayant été publiées en format pdf, un travail de saisie manuelle est nécessaire pour permettre les comparaisons. Outre la charge de travail que ce retraitement des données représente, il comporte surtout un risque important d'erreur, raison pour laquelle la société civile a toujours plaidé en faveur d'un format unique accessible gratuitement en ligne et sous forme exploitable.

- **La typologie utilisée pour qualifier les activités des banques n'est pas harmonisée,** ce qui donne des résultats très disparates et impossibles à comparer : la Société Générale distingue ainsi 8 activités pour toutes ses filiales (dont certaines extrêmement vagues comme « société financière ») là où BPCE en distingue 86, par exemple. Une typologie précise et uniformisée des activités des banques devrait permettre de mieux cerner les objectifs poursuivis par les banques quand elles s'implantent dans les paradis fiscaux. Cependant, ni la loi ni le décret d'application ne suggérant une telle typologie, les banques ont toute latitude pour qualifier comme elles l'entendent l'activité de leurs filiales.

- **La liste complète des banques soumises à cette obligation est difficile à établir.** Si la loi bancaire indique que sont pour le moment concernés par cette obligation de reporting « *les établissements de crédit, compagnies financières et compagnies financières holding mixtes, et entreprises d'investissement* », il n'existe pas de liste claire des établissements bancaires soumis à cette obligation de publication et le décret

[91] « Que font les plus grandes banques françaises dans les paradis fiscaux ? », 13 novembre 2014
[92] Décret n°2014-1657 du 29 décembre 2014 - art. 1

d'application ne fait aucune précision à ce sujet. Il faut donc chercher ces publications dans les rapports annuels de nombreux établissements bancaires établis en France.

- **Des variations sur leur périmètre de consolidation.** Les informations exigées par la loi bancaire de 2013 ne sont obligatoires que pour les entités consolidées. Sur ce point, le décret renvoie à l'article L233-16 du Code du commerce selon lequel entrent dans le périmètre de consolidation les sociétés que la maison-mère détient exclusivement (*via* la détention d'au moins la moitié des droits de vote) ou conjointement, ou sur lesquelles elle exerce une influence notable (présumée en cas de fraction au moins égale au cinquième des votes). Néanmoins, les normes comptables internationales (*International Financial Reporting Standard*) prévoient de rares exceptions à l'inclusion d'entités dans le périmètre de consolidation même quand celles-ci relèvent de la définition d'une entité consolidable. La non-significativité d'une entité est l'une de ces exceptions, et elle est à l'appréciation des établissements financiers eux-mêmes lors de l'établissement de leur rapport financier. Le risque de subjectivité est donc non-négligeable : l'utilisation de ce critère de non significativité leur permet en effet de décider qu'en dessous de certains seuils (taille de bilan, chiffre d'affaires, effectifs) les entités sont « non significatives » et donc non consolidées. Par exemple, BNP Paribas a relevé ses seuils de consolidation en 2011[93], ce qui explique que BNP soit passée de 1409 entités (dont 360 dans les paradis fiscaux) en 2011 à 870 entités (dont 214 dans les paradis fiscaux) en 2012. Ce « jeu » sur le périmètre de consolidation peut encourager la prolifération d'entités de petite taille et donc permettre aux banques de se soustraire à la publication d'informations sur certaines de leurs activités.

Notons que les pays du G20 ont adopté à Brisbane le 14 novembre 2014 une première série de mesures proposées par l'OCDE visant à lutter contre l'érosion des bases fiscales et le transfert des bénéfices, parmi lesquelles un modèle de reporting pays par pays (Action 13). Ce reporting n'est certes pas, a priori, public mais il va plus loin que celui auquel sont soumises les banques européennes : en effet doivent être aussi fournies, en plus des informations sur les bénéfices avant impôts, chiffres d'affaires et nombre d'employés, des données concernant les bénéfices non distribués, les actifs corporels hors trésorerie et équivalents de trésorerie, les impôts sur les bénéfices dus et les impôts sur les bénéfices acquittés.

[93] Communiqué de presse de BNP Paribas, 21 avril 2011.

10. L'initiative LEI (*Legal entity identifier*) : enjeux et mise en œuvre

Philippe Roussel, Cnis

La crise bancaire et financière de 2008 a montré une insuffisance de données fines, granulaires, permettant de mieux identifier les risques. À l'initiative des régulateurs, le G20 soutient en novembre 2011 (sommet de Cannes) la création du LEI (*Legal entity identifier*) – dispositif d'identifiant unique des intervenants sur les marchés financiers –, afin de faciliter la gestion et le contrôle des risques, tant par le secteur privé que par les autorités publiques. En juin 2012, le sommet du G20 de Los Cabos valide le dispositif. Le règlement européen EMIR en impose désormais l'usage pour les produits dérivés. À terme, compte tenu des diverses communications légales ou réglementaires de données (« reportings ») à réaliser auprès des régulateurs, son utilisation devrait se généraliser aux autres produits financiers.

Il s'agit donc d'une initiative des régulateurs conduite dès l'origine en étroite coordination avec le secteur privé. Une meilleure information sur les parties prenantes à des transactions intéresse les régulateurs et superviseurs, pour assurer une supervision micro- et macro-prudentielle. Cette initiative a reçu un soutien du secteur privé (banques), mais aussi des opérateurs des marchés financiers, qui y voient une source d'économies substantielle. Les institutions financières disposent de nombreux référentiels et parvenir à réconcilier leurs données internes représente un coût énorme. Disposer à terme d'un répertoire mondial des entités financières devrait constituer une source potentielle d'économies.

Le GLEIS (*Global LEI System*) est le dispositif mondial d'identification unique des intervenants sur les marchés financiers mis en place par le *Financial Stability Board* (FSB) à la demande du G20. Il est supervisé par un comité de surveillance règlementaire (ROC) composé de régulateurs financiers à travers le monde. En France, la Banque de France, le Trésor et l'AMF participent au ROC. Le GLEIS est composé d'une unité centrale (GLEIF, *Global LEI Foundation*) qui coordonne et supervise les actions des unités opérationnelles locales (LOU). Ces dernières reçoivent les demandes et attribuent les LEI aux entités demandeuses, puis diffusent librement les informations concernant les entités auxquelles elles ont attribué un LEI.

La GLEIF diffuse également centralement les informations concernant les entités auxquelles un LEI a été attribué par un LOU. Les informations diffusées concernent l'identification des entités (LEI, raison sociale, adresse, état d'activité, identifiant au registre légal, raison sociale du siège social, adresse du siège social, identifiant au registre légal du siège social, …). La liste est définie par le standard ISO17442. L'objectif de cette diffusion est de permettre aux différents intervenants d'identifier sans ambigüité les participants à des transactions financières.

L'Insee a été désigné par le ministre de l'économie pour assurer le service français d'attribution des LEI sur le périmètre des entités françaises (entreprises et fonds de droit français) : c'est l'unité opérationnelle locale française (cf. https://lei-france.insee.fr/).

L'Insee attribue donc des LEI à vie (comme pour le Siren) aux unités qui en font la demande, en s'appuyant sur le référentiel Sirene et le répertoire des fonds d'investissements de l'AMF. Le LEI, attribué puis certifié, est rendu public et peut être utilisé dans le cadre des « reportings » réglementaires. Au fur et à mesure de la généralisation de ces derniers, le nombre d'entités identifiées devrait s'accroître. Actuellement, ce sont essentiellement les acteurs intervenant sur les marchés dérivés qui sont enregistrés. Au 1er avril 2015, l'Insee a attribué 22 000 LEI, dont environ 40 % à des fonds d'investissement.

Outre l'accès en données ouvertes (« *open data* ») à un répertoire mondial des entités juridiques (360 000 unités au 1er avril 2015), le LEI permettra de produire des statistiques sur les déclarants et les contreparties, de cartographier les groupes et les liaisons entre unités et, à terme, de mieux analyser le fonctionnement du marché financier et de mieux anticiper les risques, *via* les obligations réglementaires de « reporting ».

Il est d'ores et déjà possible d'accéder sur internet aux 360 000 données sur les entreprises ou fonds d'investissement. Ces données constituent une première base de travail. Elles permettent ainsi de constater, au premier avril 2015, que 9 921 fonds ont été identifiés dans les îles Caïmans, et 344 aux Bahamas. Cette nouvelle source d'information gratuite constituera un outil tout particulièrement utile, puisqu'il permettra de répondre au besoin traditionnel des statisticiens de disposer du bon identifiant pour la bonne contrepartie. L'ambition affirmée par les pouvoirs politiques de constituer une cartographie des groupes et des liaisons entre les entités, même si elle s'avère relativement ambitieuse, est déjà en gestation. Les gains statistiques existent et militent pour une intervention de plus en plus forte du monde statistique pour soutenir le LEI.

G. Annexes

Composition du groupe de travail

Président :	Gunther Capelle-Blancard, Professeur à l'Université Paris 1 Panthéon-Sorbonne et Président de la commission « Système financier et financement de l'économie » du Cnis
Rapporteurs :	Raphaëlle Bellando, Professeure à l'Université d'Orléans
	Renaud Lacroix☆, Direction Générale des Statistiques, Banque de France
Membres :	Michel Boutillier, Professeur à l'Université Paris Ouest Nanterre La Défense
	Clément Guillo, Insee
	Denis Marionnet, Autorité de Contrôle Prudentiel et de Résolution
	Gilles Pouzin, Journaliste
	Carine Romey, Autorité des Marchés Financiers
	Catherine Refait-Alexandre, Professeure à l'Université de Franche-Comté
	Amine Tarazi, Professeur à l'Université de Limoges
	Laurent Weill, Professeur à l'Université de Strasbourg
Coordination :	Philippe Roussel, Cnis

☆ *Renaud Lacroix a souhaité ne pas être cité comme co-rapporteur dès lors qu'il n'approuvait pas certaines des recommandations du groupe de travail.*

Auditions

- Pierre Bocquet, Fédération Bancaire Française
- Antoine Frachot, Directeur général du Genes
- Kamel Gadouche, Directeur du CASD
- Henri Verdier, Directeur d'Etalab

Remerciements

Ce travail est le fruit de longues réflexions menées depuis plusieurs années auprès d'un grand nombre d'acteurs, producteurs ou utilisateurs de données. Je tiens ainsi à remercier Guillaume Arnoult (Université Paris 1), Agnès Bénassy-Quéré (Université Paris 1), Christian de Boissieu (Université Paris 1), Christian Chavagneux (Alternatives économiques), Jézabel Couppey-Soubeyran (Université Paris 1), Jérôme Creel (OFCE), Anne-Laure Delatte (CNRS), Mathilde Dupré (CCFD-Terre solidaire), Thibaut Duprey (Ecole d'Economie de Paris), Mathias Lé (École d'Économie de Paris), Jean Merckaert (Revue Projet), Jérôme Glachant (Université Paris 1), Olena Havrylchyk (Université Lille 1), Thierry Philiponnat (2° Investing), Laurence Scialom (Université Paris Ouest), Lucie Watrinet (CCFD-Terre solidaire), Gabriel Zucman (London School of Economics). Je remercie également vivement les membres du groupe de travail pour leur collaboration et leur investissement personnel, ainsi que les services de la Cnil pour leurs remarques expertes et avisées. Je remercie enfin chaleureusement l'équipe permanente du Cnis, Dominique Allain, Pierre Audibert et Philippe Roussel, pour leur efficacité, ainsi que sa présidente, Yannick Moreau, pour son soutien et sa confiance.

Gunther Capelle-Blancard