

Sirus, le répertoire d'entreprises au service du statisticien



Ali Hachid* et Marie Leclair**

Tout statisticien intéressé par la collecte et l'analyse de données économiques a besoin de s'appuyer sur un répertoire d'entreprises, par exemple pour constituer une base de sondage ou confronter ses données d'enquête ou administratives à des valeurs de référence. En France, depuis plus de cinquante ans, la statistique publique s'est appuyée sur le répertoire Sirene géré par l'Insee pour produire de l'information économique. Conçu pour répondre à des besoins administratifs, Sirene ne permet pas de satisfaire pleinement les attentes du statisticien d'entreprise d'autant plus qu'il n'est pas souhaitable d'en alourdir la charge de gestion.

À partir de 2008, l'Insee a choisi de développer un répertoire statistique d'entreprises : Sirus, répertoire des unités statistiques, est adossé à Sirene et répond aux exigences du statisticien : il prend en compte la notion d'entreprise au sens économique, il intègre de nouvelles unités comme les groupes de sociétés, il met en œuvre des traitements spécifiques comme la cessation statistique. L'ambition est de doter le système statistique public d'un outil partagé, servant de référence à l'ensemble des productions. L'enjeu est également de renforcer la cohérence d'ensemble des statistiques diffusées dans le domaine économique.

 *Any statistician interested in collecting and analysing economic data needs to rely on a business register, for example to build a sampling frame or to compare survey or administrative data with reference values. In France, for over fifty years, official statistics have relied on the SIRENE register managed by INSEE to produce economic information. Designed to meet administrative needs, SIRENE does not fully meet the expectations of business statisticians, especially as it is not desirable to increase its management burden.*

Since 2008, INSEE has chosen to develop a new register: SIRUS, the statistical business register, is linked to SIRENE and meets the requirements of the statistician: it takes into account the notion of enterprise in the economic sense, it integrates new units such as groups of companies, and it implements specific processing such as statistical cessation. The ambition is to endow the official statistical system with a shared tool, used as a reference for all productions. The challenge is also to strengthen the overall consistency of business statistics.

* Responsable des répertoires statistiques Sirus et Citrus, division Infrastructures et répertoires statistiques, Insee, ali.hachid@insee.fr

** Cheffe du département Répertoires, infrastructures et statistiques structurelles, Insee, marie.leclair@insee.fr

Une question centrale pour le statisticien est de savoir si la statistique qu'il produit représente bien la population qu'il souhaite couvrir. Pour cela, il a besoin de données de référence, qui décrivent cette population couverte et auxquelles il confronte ses propres données. Les répertoires jouent ce rôle en recensant l'ensemble des unités statistiques d'intérêt.

Pour la production des statistiques d'entreprise, le répertoire inter-administratif **Sirene** a longtemps été la référence en France (<https://www.insee.fr/fr/information/1972043>). À partir de 2008, un répertoire complémentaire dédié spécifiquement aux usages des statisticiens publics est entré dans le paysage : **Sirus** (Système d'immatriculation au répertoire des unités statistiques) a, de ce fait, un champ et des unités d'observation différents de Sirene. Référentiel pour la statistique d'entreprise, il centralise de nombreuses informations provenant de différents processus statistiques. Il offre une large palette de services, aux responsables d'enquêtes comme aux systèmes d'information utilisateurs.

► Répertorier les entreprises pour produire des statistiques représentatives

Recueillir une information *via* un questionnaire adapté, redresser la non-réponse ou ventiler ses statistiques selon des critères variés ne sera pas pertinent si la population sur laquelle ces statistiques sont calculées est trop particulière pour qu'on en tire une quelconque généralité. Pour éviter cet écueil, le responsable d'enquête tire aléatoirement un échantillon dans une base de sondage. Celle-ci doit comprendre l'ensemble des unités d'intérêt pour l'enquête : ce sera, par exemple, l'ensemble des unités ayant une activité industrielle pour **l'enquête annuelle de production** (EAP) ou encore l'ensemble des entreprises créées un semestre donné pour le **système d'information sur les nouvelles entreprises** (Sine).



Si chaque enquête mobilise une base de sondage(...) il faut, pour la constituer, s'appuyer sur un référentiel recensant exhaustivement les unités statistiques : un répertoire.



Si chaque enquête mobilise une base de sondage qui est fonction de son champ d'intérêt, il faut, pour la constituer, s'appuyer sur un référentiel recensant exhaustivement les unités statistiques : un répertoire.

Pour répondre au besoin de constitution de bases de sondage, le répertoire doit toutefois aller plus loin que le seul recensement des unités statistiques ; il est important de cumuler d'autres informations pour affiner le champ des bases de sondage à élaborer. Pour l'enquête annuelle de production citée précédemment, il est nécessaire par exemple de savoir quelles entreprises ont une activité industrielle et ce même si ce n'est pas leur activité principale ; pour le système d'information sur les nouvelles entreprises, il est fondamental de connaître la date de création de l'entreprise.

Les informations ainsi centralisées dans le répertoire peuvent être nombreuses : variables permettant de définir les entreprises productives marchandes (i.e. le champ de la statistique d'entreprises), mais également le statut juridique de l'entreprise, le secteur d'activité,

la catégorie d'entreprise au sens de la loi de modernisation de l'économie¹, le secteur institutionnel, le statut marchand, productif, exploitant ou employeur, le chiffre d'affaires, les effectifs salariés, etc.

Ce recensement des unités statistiques est indispensable aux statistiques produites à partir d'enquêtes, mais pas seulement : cela est tout autant nécessaire pour produire une statistique sur les entreprises à partir de données administratives. Le champ des données administratives ne couvre que les unités sujettes à l'obligation administrative qui a généré la source, et il faut pouvoir le confronter à la population que l'on souhaite représenter : il s'agit alors de corriger, si nécessaire, les biais de couverture de la source administrative (Rivière, 2018). Par exemple, toutes les entreprises n'établissent pas de liasses fiscales² ; c'est le cas des **microentreprises**. De ce fait, toute statistique d'entreprise s'appuyant sur ces liasses fiscales doit être redressée pour prendre en compte le chiffre d'affaires de ces microentreprises.

► Comment connaître exhaustivement la population des entreprises ?

Dans le cas français, le statisticien a l'opportunité de pouvoir s'appuyer sur le répertoire inter-administratif **Sirene** (Système informatique pour le répertoire des entreprises et des établissements) qui est géré par l'Insee depuis 1973 : chaque entreprise doit, selon le code du commerce, s'y immatriculer, signaler toute modification de sa situation (un changement d'adresse par exemple ou la création d'une nouvelle implantation). Une attention particulière est notamment apportée dans sa gestion pour éviter un doublon d'immatriculation. Depuis 1997, l'utilisation du numéro **Siren** est obligatoire pour toutes les déclarations administratives et les formalités des entreprises, ce qui garantit la couverture du répertoire mais également sa mise à jour et la justesse des informations. Le champ de Sirene est même plus large que celui de la statistique d'entreprise c'est-à-dire l'ensemble des entreprises ayant une **production marchande** de biens ou de services : depuis 1983, les entreprises publiques mais aussi les administrations doivent s'immatriculer au répertoire Sirene. Les associations, lorsqu'elles sont employeuses, qu'elles demandent des aides ou paient des impôts, sont également enregistrées.

Ce répertoire Sirene est donc une ressource essentielle pour construire un répertoire à des fins statistiques. La plupart des pays peuvent s'adosser à un répertoire administratif, mais celui-ci est rarement géré par l'institut national statistique comme c'est le cas en France. Cette particularité permet de bien prendre en compte les besoins statistiques : la possibilité de mener des opérations d'amélioration de la qualité sur les adresses, de s'assurer de la bonne codification de l'activité, etc. En l'absence d'un tel répertoire comme base de départ, il faut entreprendre un travail conséquent pour initialiser un répertoire statistique d'entreprises : l'appariement et la synthèse de différentes sources administratives permet par confrontation d'essayer de dépasser les limites de couverture

¹ La loi LME -article 51- publiée le 4 août 2008 et le décret 2008-1354 du 18 décembre 2008 « relatif aux critères permettant de déterminer la catégorie d'appartenance d'une entreprise pour les besoins de l'analyse statistique et économique ». Voir les références juridiques en fin d'article.

² Une liasse fiscale est un ensemble de déclarations fiscales remises par les professionnels (commerçants, indépendants, professions libérales, au régime réel normal) ou les sociétés soumises à l'impôt sur les sociétés.

de chaque source ; on peut également recenser les entreprises soit méthodiquement, comme un recensement de population, soit en moissonnant par exemple les sites *web* ou les annuaires privés (Unece, 2015).

► Vers la création d'un répertoire statistique des entreprises

Grâce à sa fraîcheur et à sa couverture large, Sirene a longtemps été exploité par les statisticiens qui ont directement adossé leurs processus statistiques à ce répertoire.

Toutefois, utiliser le même répertoire pour des usages administratifs et statistiques pose plusieurs difficultés. La gestion d'un répertoire administratif n'est pas la même, en effet, que celle d'un répertoire statistique. Par exemple, il n'est pas possible de changer la valeur de la variable pour un usage statistique ; toute modification de Sirene peut en effet avoir des incidences légales pour l'entreprise et il importe d'être conforme à la déclaration de l'entreprise. Cependant, le statisticien pourra exclure une entreprise de sa base de sondage (notamment en cas de cessation d'activité) pour éviter à l'avenir toute réinterrogation inutile. Mais ceci serait bien sûr inenvisageable dans le cadre d'un répertoire administratif, compte tenu des conséquences pour l'entreprise en cas de cessation erronée.

De même, exiger de Sirene l'intégration de nouvelles variables à finalité statistique accroît sa charge de gestion et ajoute un flou sur le statut de ses variables statistiques et de leur usage. Prenons l'exemple du code d'activité principale exercée de l'entreprise (**APE**) : cette variable figure dans le répertoire Sirene mais sa finalité est avant tout statistique³. Pour autant, sa présence dans le répertoire Sirene en fait une information, publique, facilement mobilisable par tous, pour obtenir par exemple des aides ou des subventions, ouvrir des droits ou des obligations. De ce fait, cet usage « administratif » et ses conséquences pratiques amènent beaucoup d'entreprises à contester la codification par l'Insee de leur activité principale. Cette information demeure dans le répertoire inter-administratif Sirene, ainsi que la **catégorie d'entreprise**, autre variable statistique ; mais toutes deux montrent l'équilibre difficile à établir au quotidien lorsqu'on diffuse des variables statistiques au niveau individuel.

“ L'Insee a voulu dissocier les usages statistiques et administratifs des répertoires. ”

Du fait de ces difficultés, l'Insee a voulu dissocier les usages statistiques et administratifs des répertoires. Un répertoire statistique, Sirius, a été ainsi créé à partir de 2008 en sus du répertoire Sirene. La séparation se parachèvera avec le passage à Sirene 4 prévu en 2023, dont l'usage sera purement administratif (Alviset, 2020). Toutes les variables ayant une finalité statistique seront répertoriées dans Sirius, qui deviendra le point d'entrée pour tous les besoins statistiques ; *a contrario*, Sirius

ne sert qu'aux besoins statistiques. C'est ainsi un outil interne au service statistique public qui n'a pas vocation à être diffusé à l'extérieur (**encadré 1**).

³ « L'attribution par [...] l'Insee, à des fins statistiques, d'un code caractérisant l'activité principale exercée (APE) en référence à la nomenclature d'activités ne saurait suffire à créer des droits ou des obligations en faveur ou à charge des unités concernées » (décret n° 2007-1888). Voir références juridiques en fin d'article.

► Un champ de la statistique d'entreprise différent du répertoire Sirene

Deux autres raisons ont justifié la séparation entre répertoire administratif et répertoire statistique. La première est la volonté de mieux définir et contrôler le champ de la statistique d'entreprise. Du fait de son usage administratif, Sirene est très adhérent aux besoins d'identification d'unités concernées par le recouvrement de l'impôt ou des cotisations sociales.

► Encadré 1. Sirius, un large spectre de services pour les statisticiens du service statistique public

Le répertoire Sirius n'est accessible qu'au sein du service statistique public. Les différents besoins des statisticiens vis-à-vis d'un répertoire ont préalablement été recensés : Sirius offre ainsi une palette de services personnalisés, tenant compte du mode d'accès souhaité et du champ d'intérêt du client (par exemple les entreprises agricoles, les filiales de groupe, etc.). Ces différents services sont :

- **l'extraction des caractéristiques d'une liste d'unités** : à partir de leur identifiant, la liste d'unités est enrichie d'informations (chiffre d'affaires, activité économique principale, statut d'activité, catégorie juridique, etc.) en valeur courante ou à une date de valeur précise ;
- **la constitution de référentiel et de base de sondage** : pour les clients « *premium** » ayant des besoins récurrents, le champ d'intérêt du client est codé directement dans le répertoire. Par exemple, il permet de fournir la liste complète des unités agricoles actives au moins un jour pendant une année donnée, accompagnées de leurs caractéristiques en fin d'année ;
- **un service de recherche** dans le répertoire notamment en vue d'identifier des entreprises ou des établissements à partir de leur dénomination ou de leur adresse d'implantation ;
- **un service d'abonnement personnalisé** permettant à chaque application cliente de recevoir hebdomadairement l'ensemble des mises à jour au répertoire impactant les unités de leur champ d'intérêt ;
- **des web services inter-applicatifs** permettant à des applications ou des processus clients de dialoguer de manière interactive avec le répertoire Sirius, par exemple pour récupérer à la volée des données sur une entreprise à afficher sur une interface métier ;
- **un service de mise à jour du répertoire** permettant à chaque processus statistique de faire remonter une information sur une unité, comme une cessation observée dans le cadre de la collecte d'une enquête auprès d'une entreprise. Le répertoire Sirius joue alors le rôle de rediffuseur de l'information à l'ensemble du réseau des statisticiens d'entreprise ;
- **une interface web** dédiée permettant aux statisticiens d'explorer le répertoire Sirius dans toutes ses dimensions : navigation dans la structure des groupes de société, consultation de l'historique des événements affectant une entreprise, etc. ;
- **des bases de données archivées** et mises à disposition des statisticiens 4 fois par an, offrant un instantané du répertoire, pour un usage en *self* à partir d'un logiciel de traitement statistique.

* Il s'agit des producteurs des enquêtes structurelles (Esane), du recensement de la population, de l'indice de chiffre d'affaires, du SSM Agriculture, etc.

Un certain nombre d'unités légales sont immatriculées pour remplir des obligations administratives sans pour autant avoir une réelle consistance économique.

Un certain nombre d'unités légales (voir ci-après) sont immatriculées pour remplir des obligations administratives sans pour autant avoir une réelle consistance économique (on peut par exemple citer les sociétés civiles de moyens pour la gestion de ressources partagées dans les professions libérales).

De ce fait, le champ utile de Sirius est plus restreint que celui de Sirene. On en exclut par exemple les associés gérants, les loueurs personnes physiques, les sociétés civiles de moyens précédemment citées, la majorité des sociétés civiles immobilières, les associations non employeuses, etc. Alors que Sirene comptabilisait en juillet 2020 plus de 14 millions d'unités légales actives, Sirius en dénombre environ 8,4 millions (*figure 1*).

Le champ de Sirius est cependant un peu plus large que celui de la statistique d'entreprise : Sirius recense ainsi les unités employeuses, y compris quand elles sont non marchandes, y compris publiques ou associatives, etc. Ainsi parmi les 8,4 millions d'unités légales actives dans Sirius, un peu plus d'un million correspondent à des unités hors champ de la statistique d'entreprise, unités légales non marchandes et non productives.

► Une unité légale n'est pas toujours une unité statistique

La seconde raison à l'origine de la création de Sirius est la définition de l'unité d'intérêt du répertoire : ce que recense Sirene n'est pas toujours ce qui intéresse le statisticien. En effet, Sirene immatricule des **unités légales** et leurs **établissements**. Par exemple, une entreprise

de vente de chaussures constituera une unité légale avec plusieurs établissements (ses différents magasins) qui seront tous enregistrés dans Sirene.

Ce que recense Sirene n'est pas toujours ce qui intéresse le statisticien.

Or, cette unité légale, définie d'un point de vue juridique, n'est pas forcément la plus pertinente pour analyser la vie des entreprises, leurs décisions, leurs

stratégies : avec le développement des groupes de sociétés, les unités légales détenues par d'autres peuvent perdre tout ou partie de leur autonomie de décision. Dans notre exemple précédent, imaginons que les magasins soient possédés par un groupe qui dispose également d'une centrale d'achat et d'une entité assurant la publicité de la marque : il sera plus intéressant pour comprendre les décisions économiques d'analyser ces unités légales (magasins, centrale d'achat et entreprise de publicité) toutes ensemble. Pour définir le contour de ce groupe de sociétés, on se fonde sur les liaisons financières entre les unités légales. Le groupe est une nouvelle unité statistique, mais il n'a pas de véritable existence juridique et on ne le trouvera pas de ce fait dans Sirene.

► Documenter le lien entre « entreprises au sens économique » et unités légales

À cette notion de groupe (définie par les liaisons financières), les économistes et statisticiens ont ajouté la notion d'entreprise, au sens économique, définie comme « *la plus petite combinaison d'unités légales qui constitue une unité organisationnelle de production de biens et de services, jouissant d'une certaine autonomie de décision, notamment pour l'affectation de ses ressources courantes* »⁴. En reprenant toujours notre exemple, imaginons que le groupe de vente de chaussures possède également, pour des raisons financières, des salles de sport mais gérées de manière complètement autonome des magasins de chaussures. Il sera plus intéressant d'un point de vue économique de considérer au sein du groupe d'une part l'entreprise en charge des ventes de chaussures (comprenant les magasins, la centrale d'achat et l'entreprise de publicité) et d'autre part une entreprise comprenant le réseau de salles de sport (**encadré 2**). Cette entreprise, au sens économique, est une convention : elle n'a pas de réalité juridique, mais du point de vue de l'économiste cette convention

► Figure 1 - Le répertoire Sirius en chiffres standardisés pour les élèves français

14 MILLIONS D'ENTREPRISES*

DONT :

4,5 MILLIONS CESSÉES ADMINISTRATIVEMENT,
1,7 MILLION CESSÉES STATISTIQUEMENT.



(situation au 19 juillet 2022)

10 000
UNITÉS MISES À JOUR
QUOTIDIENNEMENT



50
STATISTIENS
CONSULTENT
QUOTIDIENNEMENT
L'APPLICATION



130 000
NOTIFICATIONS
PAR SEMAINE
AUX APPLICATIONS
STATISTIQUES
PARTENAIRES



* depuis l'initialisation de la base en septembre 2011, dont 8 millions d'entreprise actives représentant 8,5 millions d'unités légales.

⁴ Règlement européen 696/93 du 15 mars 1993, voir les références juridiques en fin d'article.

est la plus apte à décrire de manière pertinente les décisions de l'entreprise. D'un point de vue pratique, le contour de ces entreprises au sens économique est construit par les statisticiens à partir des groupes de société, en mobilisant des méthodes dites de profilage, manuelles ou automatiques (Haag, 2019).

En France, les statistiques structurelles d'entreprise sont construites depuis l'exercice 2017 à partir de l'entreprise au sens économique, et non plus des unités légales. Cette évolution a eu pour conséquence une réévaluation de certains agrégats économiques, comme la production car les flux monétaires sans réelle consistance économique entre unités légales du même groupe sont annulés (Chanteloup et Haag, 2019) : ainsi la prise en compte des entreprises conduit à une « baisse » d'environ 7 % du chiffre d'affaires total (les ventes au sein de l'entreprise ne sont plus comptabilisées avec le passage à la notion d'entreprise). De plus, la répartition sectorielle est modifiée en faveur de l'industrie, car les entreprises industrielles rassemblent en leur sein des unités légales industrielles, mais aussi

du commerce et des services, activités « support » de l'activité principale industrielle de l'entreprise.

Par rapport au répertoire inter-administratif Sirene qui recense les unités légales, Sirius répertorie ainsi ces différentes unités statistiques et identifie les liens entre elles : les établissements d'une unité légale, les unités légales, les entreprises au sens économique, leur appartenance à un groupe.

Sirius répertorie ainsi ces différentes unités statistiques et identifie les liens entre elles.

Afin d'en faciliter la maniabilité et la maintenance, seules les informations couramment utilisées par le statisticien sont répertoriées dans Sirius, soit les principales caractéristiques des unités (date de création, activité principale, effectifs salariés, etc.) ainsi que les liens entre ces unités. Les informations plus détaillées doivent être recherchées ailleurs : un répertoire de groupes⁵ ainsi que des données sur les restructurations⁶ complètent Sirius.

► Au-delà de la constitution de bases de sondage

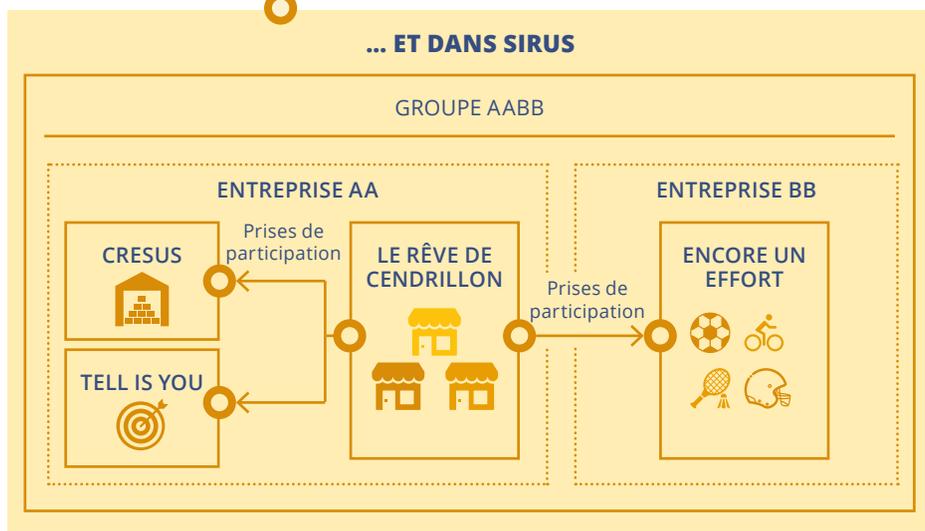
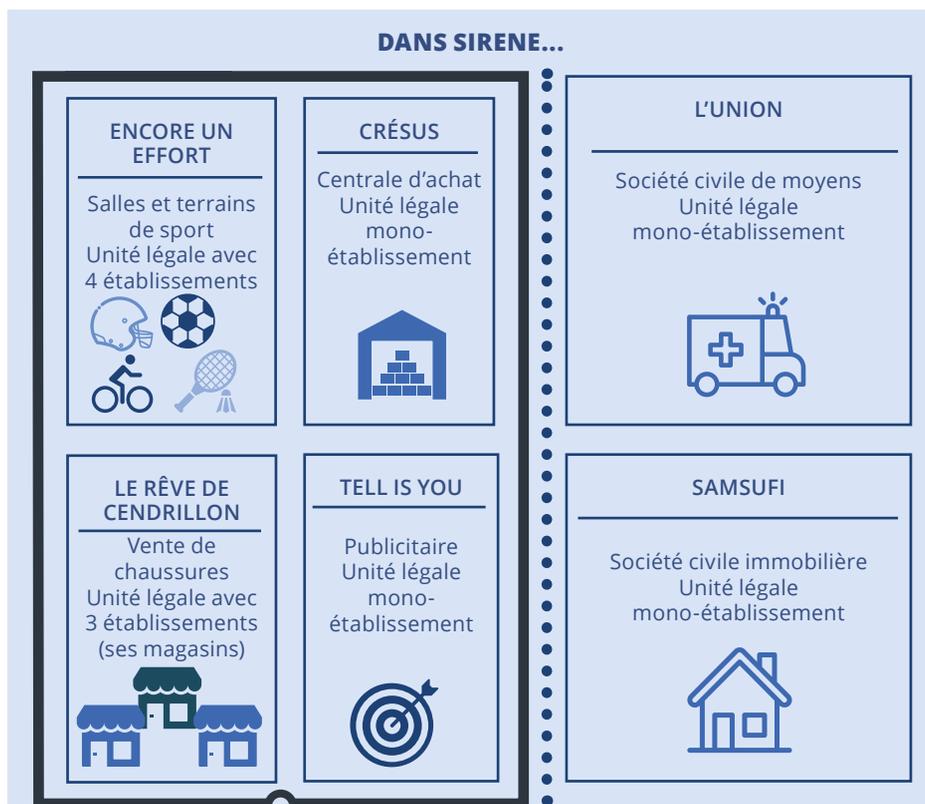
Le recensement exhaustif des unités statistiques fait aujourd'hui de Sirius le répertoire de référence pour constituer des bases de sondage sur les entreprises. Mais la centralisation de toutes ces informations sur les entreprises permet d'autres usages de Sirius : on y calcule par exemple la notion de cessation statistique des entreprises.

Qu'une entreprise ne soit plus présente dans les sources administratives mobilisées par la mise à jour de Sirius est un indice qui laisse penser que l'entreprise a cessé son activité, de même si les données de cette entreprise n'ont pas été actualisées depuis un certain temps. Cette information est cruciale pour la précision des enquêtes qui utilisent Sirius : une telle entreprise, cessée statistiquement, ne répondra pas à une enquête. Elle ne devrait même pas donner lieu au redressement de sa non-réponse (par exemple à une extrapolation de son chiffre d'affaires à partir des entreprises répondantes). Identifier les cessations statistiques améliore la qualité des résultats d'enquête, et permet évidemment d'épargner des coûts de collecte inutiles.

⁵ Le répertoire Lifi (liaisons financières).

⁶ Le répertoire Citrus (Coordination des informations et des traitements sur les restructurations d'unités statistiques).

► **Encadré 2. Sirene et Sirius : des unités en moins... et des unités en plus**





Identifier les cessations statistiques améliore la qualité des résultats d'enquête, et permet évidemment d'épargner des coûts de collecte inutiles.



La mise à jour des informations, en continu, dans Sirius en fait également la source privilégiée pour étudier la démographie d'entreprises : le nombre de créations, mais aussi les stocks et les cessations d'entreprises. Sirius est utilisé pour la production de ces statistiques depuis 2022⁷ (Insee, 2022). Ainsi, la démographie d'entreprise est désormais calculée en concept entreprise et bénéficie des traitements de cessation statistique. De plus cela permet d'avoir un champ cohérent avec les autres productions de la statistique d'entreprise.

En France, les données d'enquêtes et les données administratives se fondent presque toujours sur le numéro Siren ; il est ainsi aisé de les apparier pour en combiner les informations. Cependant, il existe encore de rares cas où cette information est indisponible (un exemple parmi d'autres est l'enquête sur le transport routier de marchandises, qui échantillonne des véhicules), voire erronée. Dans ce cas, il faut réussir à retrouver l'identité de l'entreprise (on dit aussi « sireniser » les informations), en se fondant sur les caractéristiques connues de l'entreprise dans la source mobilisée, comme la dénomination de l'entreprise ou son adresse d'implantation. Sirius, qui regroupe en son sein de nombreuses informations sur chaque entreprise, est l'outil idéal pour les identifier et restituer un ensemble large de données les caractérisant.

Enfin, un nouvel usage a été confié à Sirius : la mesure de la charge statistique⁸ des entreprises. Sirius compile pour chaque entreprise concernée, les différentes enquêtes de la statistique publique qui l'ont interrogée ainsi que le temps qu'elle a passé à répondre ; la question est en effet obligatoire à la fin de chaque enquête, depuis la commission Warsmann sur l'allègement des démarches administratives pour les entreprises (Assemblée nationale, 2011). La charge statistique de réponse aux enquêtes pour les entreprises constitue une information précieuse si on veut coordonner négativement les échantillons, c'est-à-dire choisir, quand on tire ces échantillons, avec une probabilité plus élevée les entreprises pour lesquelles la charge est la plus faible. Elle permet de répartir la charge d'enquêtes entre entreprises de manière plus équitable (Insee, 2011).

► **Sirus, au centre de nombreux processus statistiques** —

Pour construire Sirius, on s'adosse au répertoire Sirene, dont on complète les informations par de nombreux autres processus statistiques, qui eux-mêmes se fondent sur Sirius (pour tirer des échantillons, identifier des unités, définir leur champ, etc.). Sirius est ainsi au cœur d'un système complexe.

⁷ Ces séries statistiques étaient déjà produites auparavant, mais à partir de Sirene.

⁸ La charge statistique correspond au temps total que consacrent les entreprises à répondre aux enquêtes du service statistique public.



Sirus exploite les informations provenant de différentes sources administratives pour enrichir la connaissance de ses unités.



En effet Sirus exploite les informations provenant de différentes sources administratives pour enrichir la connaissance de ses unités : les données sur la TVA, les liasses fiscales, les données sur l'emploi, les données de l'URSSAF Caisse Nationale (UCN)⁹ permettant de repérer les microentrepreneurs, les données des Douanes pour repérer les entreprises importatrices ou exportatrices (*figure 2*).

La centralisation de toutes ces informations permet en retour de pouvoir décrire finement les unités qui composent le répertoire et de s'assurer de leur complétude ; elle permet également de fournir un référentiel commun à l'ensemble des statistiques d'entreprise, ce qui est un enjeu pour que celles-ci, bien que produites à partir de processus différents, demeurent comparables. Se fondant toutes sur Sirus, la définition de l'activité, de la catégorie d'entreprise, de ce qui est dans ou hors du champ devrait être identique d'une enquête à l'autre.

► **Historisation des informations, cohérence interne : deux facteurs de complexité**

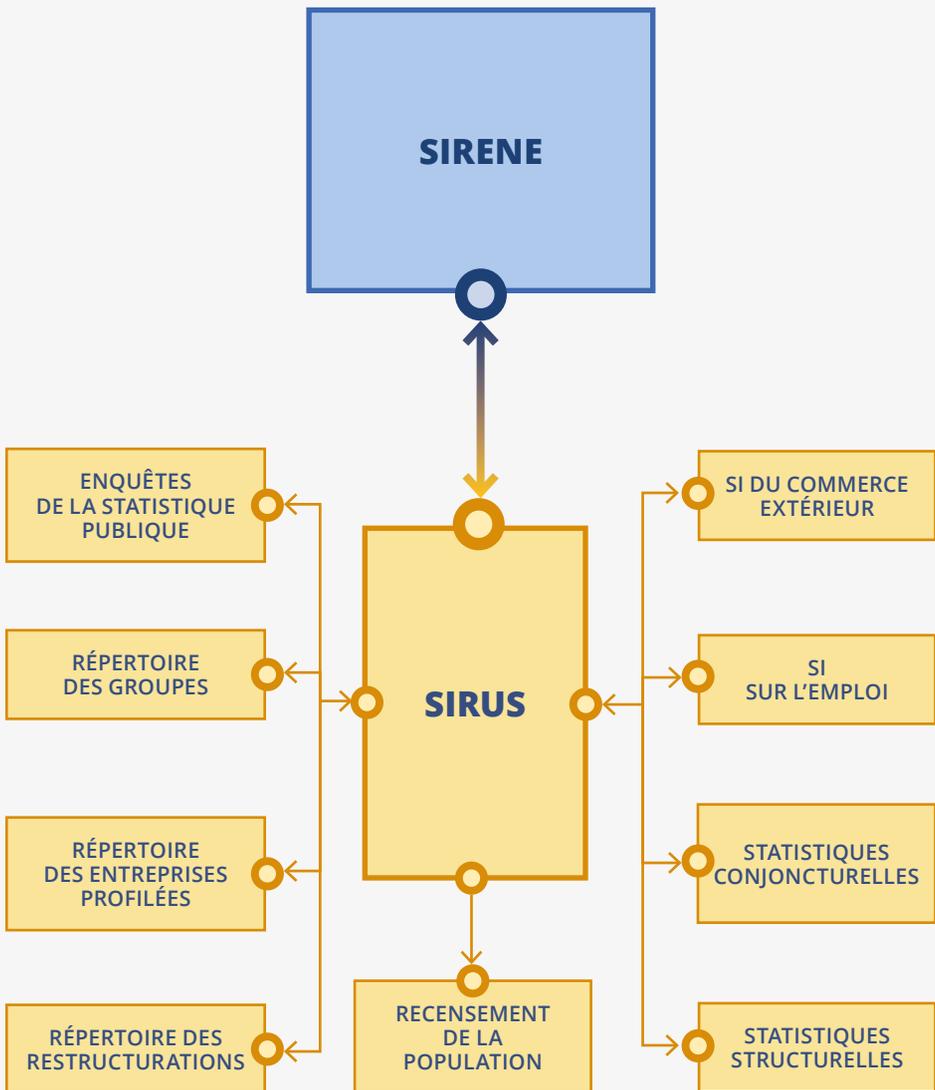
Mais derrière cette ambition, il y a bien sûr des difficultés : Sirus est mis à jour en continu par différentes sources d'information. Ses évolutions reflètent tout à la fois une réalité économique (par exemple, en décembre 2021, l'ensemble des unités actives n'est pas le même qu'un an auparavant) mais aussi la disponibilité de l'information, parfois avec retard (la cessation d'une entreprise en décembre 2021 n'est connu qu'en mars 2022). Sirus historise ainsi l'information pour prendre en compte au cours du temps les changements qui interviennent réellement sur les unités statistiques mais aussi les mises à jour des sources : les modifications ont ainsi une date de traitement (ou une date de prise en compte dans le répertoire) et une date d'effet (ou de validité dans la réalité économique). Cependant, compte tenu du nombre de sources impliquées pour la mise à jour de Sirus (de l'ordre d'une dizaine), toutes les variables et tous les changements ne peuvent pas être historisés, de manière à conférer à Sirus une certaine maniabilité.

Une autre difficulté tient également à la cohérence interne du répertoire : certaines variables, comme la catégorie d'entreprise ou encore le champ de la statistique d'entreprise, sont calculées par Sirus sur la base des différentes informations que lui apporte chaque source. Les interdépendances entre variables du répertoire Sirus peuvent être telles qu'il n'est pas toujours aisé de répercuter une correction sur l'ensemble du processus. Dès lors, une correction peut créer un dilemme entre l'incohérence potentielle qu'elle peut créer au sein du répertoire et le souci d'exactitude et de fraîcheur (avoir la dernière information d'une source de données). Et il arrive que l'on soit contraint en dernier recours de restaurer la base à partir d'une version antérieure à la prise en compte d'une information erronée.

⁹ Issues de l'obligation de déclarations du chiffre d'affaires aux Urssaf pour les microentrepreneurs.

Enfin, toutes ces sources d'information peuvent modifier des unités pour une même caractéristique (une enquête structurelle annuelle peut modifier le code d'activité principale exercée alors que l'entreprise aura demandé au répertoire Sirene de modifier ce code pour une raison administrative). Dès lors, il convient d'arbitrer entre ces différentes mises à jour pour savoir à laquelle Sirius donne la préséance : faut-il choisir l'information la plus récente ? Ou considérer que certaines sources sont prioritaires sur certains champs ?

► **Figure 2 - Sirius, au cœur des systèmes d'information statistique**



Pour gérer ces conflits, un ensemble de règles ont été établies visant pour chaque variable à déterminer la priorité de chaque source.

Pour gérer ces conflits, un ensemble de règles ont été établies visant pour chaque variable à déterminer la priorité de chaque source en fonction de la date d'effet proposée de la mise à jour (Sirene ne peut pas, par exemple, mettre à jour l'activité d'une entreprise pour l'année N si cette entreprise a été enquêtée la même année par une enquête annuelle structurelle).

Les enjeux sont importants car pour être utilisé, Sirius doit apporter à ses utilisateurs une information la plus fraîche possible, la plus riche mais aussi cohérente, facilement mobilisable et traçable (l'utilisateur voudra pouvoir retrouver la base de sondage tel qu'il l'aura défini à un instant T à partir de Sirius).

► Des répertoires nationaux aux répertoires supranationaux

Bien que les *business registers* (répertoires d'entreprise) ne soient pas une « statistique » fournie par les instituts nationaux statistiques à Eurostat, ils font l'objet de beaucoup d'attentions dans les travaux européens et internationaux, compte tenu de leur rôle central dans la production de toutes les autres statistiques d'entreprise : manuels et groupes de travail visent à harmoniser et à améliorer la qualité de ces répertoires (*Unece, 2015 ; Eurostat, 2021*).

Eurostat ou la commission statistique des Nations Unies incitent ainsi les instituts statistiques à réaliser des autoévaluations de la maturité de leur *business register* : la couverture de ces répertoires, en termes de **secteurs institutionnels**, mais aussi d'unités statistiques est un élément de cette évaluation, ainsi que le nombre de sources prises en compte, la fréquence de leur mise à jour, leur cohérence interne et externe, le détail géographique, les supports légaux les fondant, leur interopérabilité¹⁰ (*Collet et Iskrenova, 2021 ; Al Kafri et Hermans, 2021*).

Si ces répertoires sont un enjeu au niveau européen et international, c'est aussi par ce que la pertinence de leur dimension nationale est remise en cause par la mondialisation.

Si ces répertoires sont un enjeu au niveau européen et international, c'est aussi par ce que la pertinence de leur dimension nationale est remise en cause par la mondialisation (*Sturm, 2021*). Par définition, les répertoires nationaux d'entreprise couvrent le territoire national : on trouve donc dans le répertoire français les unités légales françaises et ce que l'on appelle la « trace française » des groupes et des entreprises au sens économique. En effet, certaines entreprises sont transnationales et une partie des unités légales sur le territoire français peut appartenir à une entreprise dont le centre de décision se trouve

à l'étranger. Dans ce cas, la statistique française, et les répertoires, ne suivent que la trace de l'entreprise qui se trouve sur le territoire français mais ne capte ainsi qu'une partie des choix économiques qu'opère l'entreprise.

¹⁰ Voir également l'article de Pascal Rivière dans ce même numéro.

Parce qu'il peut être plus pertinent pour comprendre et analyser les décisions de ces entreprises de se situer à un niveau supranational, un répertoire des groupes européens (*European Group Register*, EGR) est construit par les différents instituts nationaux statistiques européens sous l'égide d'Eurostat ; des tentatives similaires sont faites au niveau mondial par l'OCDE (*Analytical Database on Individual Multinationals and Affiliates*, Adima) et la commission statistique des Nations Unies (*global enterprise group register*), prenant en compte le poids considérable des entreprises multinationales dans l'économie mondiale (Snyder, Di Matteo, Pilgrim et Ostolaza, 2021).

► Quelles évolutions à l'avenir pour Sirius ?

Le répertoire européen des groupes se construit en se fondant sur les répertoires nationaux et permet de définir les enveloppes des groupes européens. À terme, une interopérabilité sera exigée entre ces répertoires nationaux, dont Sirius, et le répertoire européen des groupes, c'est-à-dire que Sirius devra pouvoir marquer les groupes européens avec les identifiants du répertoire européen et assurer la cohérence des données des deux répertoires sur le champ des entreprises multinationales.

Ces évolutions sont prévues par le nouveau règlement européen sur les statistiques d'entreprise *Structural Business Statistics* (Colin, 2019) qui a également relevé le niveau d'exigence sur les *business registers* nationaux en termes de variables à couvrir (de manière obligatoire ou recommandée, comme le marquage des *Special Purpose Entity*¹¹) et a réaffirmé le rôle central de ces répertoires, véritable colonne vertébrale de la statistique d'entreprise, pour garantir la cohérence des données dans le temps, entre les États membres et entre les différents domaines (statistiques de court terme, statistiques structurelles d'entreprises, démographie d'entreprises, statistiques du commerce selon les caractéristiques des entreprises, etc.).

Poussé par ces règlements, Sirius devrait donc s'enrichir de nouvelles variables dans les années à venir : le numéro de TVA intracommunautaire, l'identifiant LEI (Legal Entity Identifier), l'identifiant des entreprises et des groupes du répertoire européen (EGR) afin de faciliter les appariements avec d'autres sources.

Un autre enjeu important pour Sirius est la refonte du répertoire inter-administratif Sirene auquel il est adossé : à partir de janvier 2023, l'écosystème des formalités d'entreprise va être profondément modifié avec l'arrivée d'un guichet unique d'entreprises qui va remplacer tous les centres de formalité d'entreprise et amener une dématérialisation complète de ces formalités. Le répertoire Sirene va être profondément refondu pour prendre en compte ce nouveau contexte (Alviset, 2020). L'enjeu pour Sirius est de continuer à produire les mêmes services alors que sa principale source sera grandement modifiée.

¹¹ Les SPE sont des entités légales sans réelle consistance économique, créées pour donner à leur propriétaire des avantages spécifiques fournis par la juridiction d'accueil.

► Bibliographie

- ALKAFRI, Saleh et HERMANS, Hank, 2021. The Maturity Model for Statistical Business Registers. In : *27th Meeting of the Wiesbaden Group on Business Registers*. [en ligne]. 20-24th September 2021. Inegi, Mexico, Session n°7. [Consulté le 2 septembre 2022]. Disponible à l'adresse : https://www.inegi.org.mx/eventos/2021/wiesbaden/doc/7_5_PAP_HERMANS_AL_KAFRI.pdf.
- ALVISET, Christophe, 2020. La troisième refonte du répertoire Sirene : trop ambitieuse ou pas assez. In : *Courrier des statistiques*. [en ligne]. 29 juin 2020. Insee, N° N4, pp. 101-121. [Consulté le 2 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4497083/courstat-4-7.pdf>.
- ASSEMBLÉE NATIONALE, 2011. *Rapport n°3787 fait sur la proposition de loi n°3706 de M. Jean-Luc Warsmann relative à la simplification du droit et à l'allègement des démarches administratives*. [en ligne]. 5 octobre 2011. Tome I de M. Étienne Blanc. [Consulté le 2 septembre 2022]. Disponible à l'adresse : <https://www.assemblee-nationale.fr/13/rapports/r3787-tl.asp>.
- CHANTELOUP, Guillaume et HAAG, Olivier, 2019. De la définition juridique à la définition économique de l'entreprise : méthode et mode d'emploi. In : *Les entreprises en France. Édition 2019*. [en ligne]. 3 décembre 2019. Insee Références, pp. 45-58. [Consulté le 2 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4255789>.
- COLIN, Christel, 2019. FRIBS : un nouveau cadre commun pour les statistiques d'entreprises européennes. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 110-124. [Consulté le 2 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4254231/courstat-3-9.pdf>.
- COLLET, Isabelle et ISKRENOVA, Iliyana, 2021. Data-quality indicators for European statistical business registers. In : *27th Meeting of the Wiesbaden Group on Business Registers*. [en ligne]. 20-24 septembre 2021. Inegi, Mexico, Session n°7. [Consulté le 2 septembre 2022]. Disponible à l'adresse : https://www.inegi.org.mx/eventos/2021/wiesbaden/doc/7_2_PAP_ISKRENOVA_COLLET.pdf.
- EUROSTAT, 2021. *European business statistics methodological manual for statistical business registers*. [en ligne]. 16 février 2021. Manuals and Guidelines. [Consulté le 2 septembre 2022]. 2021 edition. Disponible à l'adresse : <https://ec.europa.eu/eurostat/fr/web/products-manuals-and-guidelines/-/ks-gq-20-006>.
- HAAG, Olivier, 2019. Le profilage à l'Insee : une identification plus pertinente des acteurs économiques. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. Insee. N° N2, pp. 86-102. [Consulté le 2 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/version-html/4168409/courstat-2-9.pdf>.
- INSEE, 2022. *Rebond des créations d'entreprises en juin 2022*. [en ligne]. 19 juillet 2022. Informations Rapides n°185. [Consulté le 2 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/6473986>.
- RIVIÈRE, Pascal, 2018. Utiliser les déclarations administratives à des fins statistiques. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 14-24. [Consulté le 2 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/3647013/courstat-1-5.pdf>.
- SCHUHL, Pierrette, 2018. Le Legal Entity Identifier - Contexte international et rôle de l'Insee. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 58-68. [Consulté le 14 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/information/3646983?sommaire=3647035>.

- SNYDER, Nancy, DI MATTEO, Ilaria, PILGRIM, Graham et OSTOLAZA, Rodolfo, 2021. Future Improvements and Collaboration on the GGR and ADIMA. In : *27th Meeting of the Wiesbaden Group on Business Registers*. [en ligne]. 20-24 septembre 2021. Inegi, Mexico, Session n°6. [Consulté le 2 septembre 2022]. Disponible à l'adresse : https://www.inegi.org.mx/eventos/2021/wiesbaden/doc/6_4_PAP_SNYDER_OSTOLAZA.pdf.
- STURM, Roland, 2021. Units unchanged or units unchained: How to react to globalisation from a statistical units' perspective ? In : *27th Meeting of the Wiesbaden Group on Business Registers*. [en ligne]. 20-24 septembre 2021. Inegi, Mexico, Session n°6. [Consulté le 2 septembre 2022]. Disponible à l'adresse : https://www.inegi.org.mx/eventos/2021/wiesbaden/doc/5_1_PAP_STURM.pdf.
- UNECE, 2015. Data sources for SBR. In : *Guidelines on statistical business register*. [en ligne]. Juillet 2015. Chapitre n°6. [Consulté le 2 septembre 2022]. Disponible à l'adresse : https://unece.org/DAM/stats/publications/2015/ECE_CES_39_WEB.pdf.

► Fondements juridiques

- Règlement (CEE) n° 696/93 du Conseil, du 15 mars 1993, relatif aux unités statistiques d'observation et d'analyse du système productif dans la Communauté. In : *Journal officiel des Communautés européennes*. [en ligne]. Mis à jour le 31 décembre 2007. [Consulté le 17 octobre 2022]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:31993R0696>.
- Loi n° 2008-776 du 4 août 2008 de modernisation de l'économie (1). In : *site de Légifrance*. [en ligne]. Mise à jour le 26 août 2021. [Consulté le 17 octobre 2022]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000019283050/>.
- Décret n° 2007-1888 du 26 décembre 2007, portant approbation des nomenclatures d'activités et de produits françaises. In : *site de Légifrance*. [en ligne]. Mis à jour le 31 décembre 2007. [Consulté le 17 octobre 2022]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000017765090/>.
- Décret n° 2008-1354 du 18 décembre 2008 relatif aux critères permettant de déterminer la catégorie d'appartenance d'une entreprise pour les besoins de l'analyse statistique et économique. In : *site de Légifrance*. [en ligne]. [Consulté le 17 octobre 2022]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000019961059/>.