

CONSEIL D'ORIENTATION DES RETRAITES

Séance plénière du 22 octobre 2013 à 14h30

« Les outils statistiques et les systèmes d'information sur les retraites »

Document N°2

<i>Document de travail, n'engage pas le Conseil</i>

Le système statistique actuel d'information sur les retraites

Secrétariat général du Conseil d'orientation des retraites

Le système statistique actuel d'information sur les retraites

Ce document décrit les principales sources de données statistiques – données d'enquêtes par sondage et données extraites de fichiers administratifs – qui constituent actuellement le système d'information statistique sur les retraites en France.

Une présentation exhaustive de toutes les données existantes n'est bien sûr pas possible. De très nombreuses sources peuvent en effet aborder des thématiques en lien avec la retraite – même si ce n'est parfois que de manière annexe. On s'attachera donc surtout à présenter les sources principales, c'est-à-dire celles dont sont issues la plupart des informations disponibles pour le public en matière de retraite, qu'il s'agisse de résultats descriptifs sur les retraites ou les retraités, d'études et analyses économiques détaillées, ou de projections à court, moyen ou long terme issues d'outils prospectifs basés sur ces données statistiques¹.

1. Les sources statistiques sur les retraites légalement obligatoires issues de fichiers administratifs

Les revenus des retraités proviennent en grande partie des pensions versées par les régimes de retraite légalement obligatoires. Les données extraites des fichiers administratifs de gestion de ces régimes constituent donc naturellement la principale source d'information sur les retraites.

1.1 Les données statistiques des régimes de retraite

Chaque régime gère son ou ses propres fichiers statistiques. On peut citer, à titre d'exemple de tels fichiers, le **Système National Statistique Prestataires (SNSP)** du régime général, la **base des pensions** du Service des retraites de l'Etat, ou la **base statistique des pensions** de la Mutualité sociale agricole (MSA). Ces trois régimes ont pour particularité d'apparaître dans les programmes statistiques annuels du Conseil national de l'information statistique (CNIS) et, à ce titre, sur le portail internet de la Statistique publique².

D'autres régimes – qu'il n'est pas possible de tous mentionner ici –, sans apparaître dans le champ de la statistique publique, gèrent également leurs propres bases de données et publient les statistiques qu'ils en extraient.

1.2 Les échantillons interrégimes construits par la DREES

La plupart des retraités perçoivent des pensions de plusieurs régimes – régime(s) de base et régime(s) complémentaire(s). La connaissance d'un seul régime à la fois ne fournit donc qu'une information partielle sur les retraites de chaque individu, et des sources de données interrégimes sont nécessaires pour avoir une image plus globale. Les échantillons interrégimes construits par la DREES constituent, en ce domaine, les sources de référence. De nouveaux fichiers interrégimes de gestion ont par ailleurs été créés récemment, notamment

¹ Il convient toutefois de ne pas confondre la disponibilité de sources statistiques et l'existence d'études publiées sur un sujet donné. Le fait qu'aucune analyse n'ait été publiée sur telle ou telle question spécifique ne signifie pas forcément qu'aucune information qui permettrait de l'éclairer ne soit (potentiellement) disponible.

² <http://www.statistique-publique.fr/>

afin de mettre en œuvre les mécanismes d'écrêtement selon le niveau de pension tous régimes (voir **document n°3**) ou dans le cadre des campagnes d'information retraite (voir **document n°7**), mais ils ne font pas, à ce jour, l'objet d'une utilisation statistique.

L'**échantillon interrégimes de retraités (EIR)** permet de reconstituer le montant de la retraite globale (tous régimes confondus, droits directs et droits dérivés) pour un échantillon anonyme d'individus, en rapprochant les données des différents régimes français légalement obligatoires. L'EIR collecte également des informations sur les éléments de calcul du montant de pension : nombre de trimestres ou de points validés, taux et circonstances de liquidation, décote et surcote éventuelle, etc. Il permet ainsi d'observer l'évolution du nombre et de la composition des retraités, de décrire les conditions de départ à la retraite et de reconstituer le montant moyen de pension tous régimes par retraité.

L'EIR est un panel : les générations appartenant à l'échantillon d'une vague sont également sélectionnées lors de la vague suivante, ce qui permet d'observer finement l'évolution des retraites entre deux vagues, en distinguant l'effet de différents facteurs (revalorisations, acquisition de nouveaux droits, arrivée de nouveaux liquidants, disparition des décédés, etc.) et d'analyser la mortalité des retraités en lien avec leurs caractéristiques. La version disponible la plus récente – l'EIR 2008, portant sur les pensions au 31 décembre 2008 – est la sixième vague de cette opération statistique, après celles de 1988, 1992, 1996, 2000, 2004. La prochaine vague – EIR 2012 – est en cours de production par la DREES.

La sélection des individus de l'EIR se fait sur la base de la date de naissance. L'EIR 2004 portait sur les générations âgées de 54 ans ou plus, nées en France ou à l'étranger. Le champ de l'EIR 2008 a été étendu aux générations plus récentes, pour mieux prendre en compte les possibilités de départs anticipés dans certains régimes spéciaux. L'EIR de 2008 rassemble ainsi les observations individuelles pour plus de 250 000 retraités, soit environ 1,6 % de l'ensemble des retraités de droit direct ou dérivé (un retraité sur 60)³. L'EIR est par ailleurs pondéré de manière à être représentatif de l'ensemble des retraités des régimes français. La quasi-totalité des organismes de retraite obligatoire (régimes de base, régimes spéciaux et régimes complémentaires légalement obligatoires) sont interrogés, soit 74 régimes au total pour l'EIR 2008. Les retraites issues de régimes supplémentaires non obligatoires sont en revanche exclues du champ de l'opération, ainsi que les revenus provenant de l'épargne individuelle volontaire.

Parallèlement à l'EIR, l'**échantillon interrégimes de cotisants (EIC)**, construit lui aussi à partir du rapprochement des données des différents régimes, permet de reconstituer l'intégralité de la carrière année après année pour un échantillon anonyme d'individus. L'EIC permet ainsi d'étudier les droits à la retraite acquis dans l'ensemble des régimes par les personnes de différentes générations qui cotisent ou ont cotisé à un régime de retraite au cours de leur carrière. Pour chaque individu, chaque caisse de retraite et chaque année de carrière, les informations renseignées concernent les durées validées (durées cotisées et périodes assimilées), les salaires ou revenus d'activité portés au compte, les points acquis dans les régimes par points, et – lorsque le régime est en mesure de fournir l'information – les conditions d'emploi (quotité de travail, caractéristiques des emplois occupés, etc.)

³ La sélection des générations retenues dans l'échantillon varie selon l'âge à la date de référence : toutes les générations sont retenues pour celles ayant entre 54 et 66 ans, une sur deux entre 66 et 90 ans, une sur trois après 90 ans, et une sur quatre entre 30 et 54 ans. La génération ayant 66 ans est par ailleurs sur-échantillonnée : un retraité de cette génération sur 15 est retenu dans l'échantillon, contre un retraité sur 36 pour les autres générations.

L'EIC de 2009, portant sur les droits acquis jusqu'au 31 décembre 2009, est la troisième vague de cette opération statistique, après celles de 2001 et de 2005. La sélection des individus de l'EIC se fait sur la base de la date de naissance. L'EIC de 2009 portait sur les générations ayant entre 23 et 75 ans au 31 décembre 2009, nées en France ou à l'étranger. L'échantillon couvre une génération sur quatre (personnes nées en 1934, 1938, 1942... jusqu'à 1986), soit près de 23 000 personnes par génération concernée (c'est-à-dire environ 2,7 % du nombre total de personnes de la génération).

L'EIC et l'EIR s'articulent et les personnes échantillonnées dans l'une ou l'autre des bases se recourent pour partie⁴, afin de pouvoir comparer les droits estimés et les pensions effectivement versées. Outre l'intérêt propre de l'EIC (étudier les carrières et l'acquisition des droits), l'EIC peut donc être utilisé conjointement avec l'EIR pour analyser le passage de l'emploi à la retraite.

Afin de compléter les informations fournies par les régimes de retraite, l'EIR et l'EIC sont appariés avec plusieurs panels construits par l'Insee (pour toutes les années de carrière dans le cadre de l'EIC et pour les périodes après 50 ans dans celui de l'EIR) : le **panel DADS**, qui renseignent sur les caractéristiques des emplois occupés par les salariés du privé et des collectivités territoriales et hospitalières ; le **panel Etat**, qui contient des informations équivalentes pour les salariés de l'Etat ; enfin les fichiers de l'**UNEDIC**, qui renseignent sur les allocations de chômage et de préretraite publique.

1.3 Une synthèse de sources pour compléter l'information : l'enquête annuelle auprès des caisses de retraite et le modèle ANCETRE

Du fait de la lourdeur de leur processus de collecte, les échantillons interrégimes de retraités et de cotisants de la DREES ne sont toutefois constitués que tous les quatre ans. D'autres sources sont donc nécessaires, afin de disposer d'une actualisation annuelle et plus récente des principaux résultats concernant les retraites tous régimes.

La DREES recueille ainsi chaque année un certain nombre de résultats agrégés dans le cadre de son **enquête annuelle auprès des caisses de retraite (EACR)**, dont les résultats sont disponibles depuis 2004. Cette enquête porte sur les effectifs de retraités et les pensions moyennes versées par les principaux régimes de retraite en termes d'effectifs⁵ (13 régimes de retraite de base et 5 régimes de retraites complémentaires⁶), ainsi que sur les proportions de retraités concernés par les diverses circonstances de liquidation (décote, surcote, ex-invalidité, etc.) Les informations sont en outre détaillées par sexe et, dans la plupart des cas, par génération.

Les données recueillies permettent de suivre l'évolution annuelle d'un certain nombre d'indicateurs, mais uniquement régime par régime. Pour suivre annuellement, au niveau « tous régimes », les principales grandeurs agrégées (notamment le nombre de retraités et le

⁴ Plus précisément, les personnes nées au cours des dix premiers jours du mois d'octobre des années 1934, 1938, ... jusqu'à 1974, sont retenues à la fois dans le champ de l'EIC de 2009 et (si elles perçoivent une pension de retraite) dans celui de l'EIR de 2008.

⁵ Depuis 2011, des données sont également recueillies auprès de la caisse nationale d'assurance maladie des travailleurs salariés, portant sur les montants de pension et effectifs bénéficiaires de pensions d'invalidité.

⁶ CNAV, MSA salariés et non-salariés (pour ces derniers : régimes de base et complémentaire), RSI (artisans et commerçants, régimes de base et complémentaire), Service des retraites de l'État (fonctionnaires d'Etat civils et militaires), CNRACL, SNCF, RATP, CNIEG, CRPCEN, CAVIMAC, ARRCO, AGIRC, IRCANTEC.

montant moyen de pension tous régimes confondus), la DREES utilise un outil de microsimulation statique, conçu pour synthétiser au mieux les données agrégées par régime et actualisées, renseignées dans l'enquête annuelle auprès des caisses de retraite, et les données individuelles – mais parfois datées – de l'échantillon interrégimes de retraités : le **modèle ANCETRE** (Actualisation annuelle par calage pour l'estimation tous régimes des effectifs de retraités).

Il s'articule autour de deux étapes (voir **document n°11**) : dans un premier temps, il prend en compte l'évolution d'année en année de la structure des polypensions qui est prévisible du fait d'effets de composition démographique et de la montée en charge de certains dispositifs (notamment la création des retraites anticipées pour carrières longues). Il se fonde pour cela sur la dernière version disponible de l'EIR et sur les données démographiques les plus récentes publiées par l'INSEE. Dans un second temps, le modèle repondère les données individuelles pour assurer la cohérence entre les données de l'EIR et les données agrégées actualisées de l'EACR. Cette pondération se fait régime par régime et, dans la plupart des cas, génération par génération.

2. Les autres sources et thématiques connexes

Une limite importante des données issues de fichiers administratifs est que ces données contiennent généralement très peu d'informations autres que celles directement utiles pour calculer les retraites versées par les régimes, et en particulier très peu d'information sur les caractéristiques sociodémographiques des personnes. Pour étudier un certain nombre de problématiques concernant les retraités, d'autres sources statistiques doivent donc être mobilisées. Il s'agit soit d'enquêtes générales, couvrant toute la population française et incluant, à ce titre, des retraités ; soit d'enquêtes spécifiques, étudiant une thématique en particulier ou ciblant une sous-population de retraités ou de personnes âgées.

2.1 Les enquêtes ou thématiques générales

Le premier ensemble recouvre notamment la plupart des grandes enquêtes en population générale de l'Insee, qui peuvent être mobilisées pour des analyses portant sur le sous-champ des retraités contenus dans leur échantillon, chacune sur sa ou ses thématiques spécifiques : ainsi, le **recensement** ou l'**enquête Logement** permettent d'analyser les caractéristiques de logement et la mobilité résidentielle des retraités, les **enquêtes Budget de famille** leur consommation, les **enquêtes Emploi du temps** leurs activités et leur sociabilité, etc.

Parmi les enquêtes plus « ciblées », l'**enquête SHARE** (acronyme en anglais « d'enquête sur la santé, le vieillissement et la retraite en Europe ») occupe une place particulière. Cette enquête est certes ciblée en termes de population étudiée (elle porte sur les personnes de 50 ans et plus), mais elle est transversale en termes de thématiques, car elle aborde un ensemble très vaste de problématiques : emploi, retraite, patrimoine, santé, données psychologiques, famille, entraide, réseaux sociaux, etc. Cette enquête européenne est réalisée tous les deux ans depuis 2004 et se présente sous la forme d'un panel (les personnes échantillonnées étant réinterrogées jusqu'à leur décès). Si cette source permet des comparaisons européennes, l'effectif de personnes interrogées en France est relativement limité (de l'ordre de 3 000 individus pour les trois premières vagues), ce qui empêche souvent des analyses fines sur le seul cas français. Cependant, l'échantillon a été doublé pour la 4^e vague réalisée en 2011. Par ailleurs, la limite liée à la taille de l'échantillon n'est pas

spécifique à l'enquête SHARE : l'échantillon de seniors n'est, par exemple, guère plus important dans les panels généraux (tel que SRCV).

2.2. Les enquêtes ou thématiques spécifiques

Sans prétention d'exhaustivité sur les enquêtes interrogeant ou concernant les retraités ou les retraites, on détaille, dans les sous-parties qui suivent, les principales sources qui peuvent être mobilisées pour des problématiques spécifiques fréquemment suivies.

a) Niveau de vie des retraités

En matière de niveau de vie⁷ et de patrimoine des retraités, les sources de référence sont des opérations statistiques en population générale de l'INSEE : enquête **Revenus Fiscaux et Sociaux** (ERFS), enquête **Patrimoine**, enquête **statistique sur les ressources et conditions de vie** (SRCV).

L'enquête **Revenus Fiscaux et Sociaux** (ERFS) est réalisée chaque année depuis 1996 en complétant les données de l'enquête Emploi de l'INSEE par des informations sur les revenus des ménages enquêtés, de façon à reconstituer leur revenu disponible et leur niveau de vie⁸. Ces données proviennent, selon la nature des revenus, des déclarations fiscales (revenus d'activité et de remplacement, revenus fonciers, impôts...), des organismes de protection sociale (prestations familiales et logement, minima sociaux) ou d'estimations économétriques d'après les données micro et macroéconomiques sur le patrimoine (revenus financiers). Grâce au recours aux données administratives (réputées plus fiables qu'une interrogation directe des ménages sur leurs revenus) et à un échantillon important (celui de l'enquête Emploi), l'ERFS constitue la source de référence sur les niveaux de vie des ménages en France. Le COR se base notamment sur cette source pour comparer le niveau de vie ou la pauvreté des retraités à celui des actifs.

L'enquête **SRCV** est une opération européenne coordonnée par Eurostat et lancée en 2004, qui sert de référence sur les niveaux de vie des ménages en Europe. Elle contient à la fois des données sur les revenus et sur les conditions de vie (permettant notamment de mesurer la pauvreté en termes de conditions de vie). Cette source est peu utilisée pour l'étude des niveaux de vie en France, car son échantillon est plus restreint que celui de l'ERFS. Par ailleurs, jusqu'en 2008, les données sur les revenus provenaient d'une interrogation directe des ménages, ce qui pouvait conduire à une sous-estimation des revenus des retraités. Ce problème ne devrait plus se poser, puisque l'INSEE applique depuis 2008 la même méthode que l'ERFS pour reconstituer les revenus.

En complément de ces sources, certaines enquêtes spécifiques permettent de cibler les retraités ou les personnes âgées à plus bas revenu ou plus bas niveau de vie. L'**enquête sur les allocations du minimum vieillesse**, collectée annuellement par la DREES auprès des principaux régimes prestataires de ces allocations, fournit ainsi des informations, agrégées mais néanmoins ventilées à un niveau fin (par sexe, par âge, par région, etc.) sur les bénéficiaires de minimum vieillesse (aujourd'hui de l'ASPA). L'**enquête auprès des bénéficiaires de minima sociaux** (BMS) de la DREES, dans sa vague réalisée en 2012,

⁷ Pour connaître le niveau de vie des retraités, il faut connaître, outre le montant de leurs retraites, le montant de leurs autres revenus (revenus du capital, etc.) et la structure des ménages dans lesquels ils se trouvent.

⁸ Le niveau de vie est calculé comme le revenu disponible divisé par le nombre d'unités de consommation du ménage.

fournira des informations détaillées sur les revenus et les conditions de vie des personnes bénéficiaires du minimum vieillesse. Les résultats de cette enquête (portant sur les seuls allocataires de minima sociaux) pourront être comparés à ceux de l'enquête SRCV (portant sur des populations plus larges) : en effet le questionnaire de l'enquête et la méthode de mesure des revenus reprend celui de SRCV.

b) La retraite supplémentaire et l'épargne retraite

Les sources statistiques potentielles sur la retraite supplémentaire dans le cadre professionnel et l'épargne retraite à titre individuel sont de trois natures :

- Des données d'enquête auprès des ménages, telles que l'**enquête Patrimoine** de l'INSEE, qui interrogent les ménages de manière précise sur les actifs qu'ils détiennent (dont les produits d'épargne retraite, ainsi que les produits d'épargne généraux tels que l'assurance-vie qui, sans être spécifiquement considérés comme de l'épargne retraite, peuvent être détenus par les ménages en vue de la retraite). Ces enquêtes présentent l'intérêt d'être très riches en données socio-économiques, ce qui permet de mieux caractériser les personnes qui détiennent des produits d'épargne retraite (catégorie socioprofessionnelle, niveau de diplôme, etc.), mais leur limite réside dans le caractère déclaratif des réponses, avec des risques de sous-déclaration ou de mauvaise déclaration induits notamment par la complexité des produits financiers.
- Des données issues d'interrogations exhaustives ou quasi-exhaustives des organismes gestionnaires de produits d'épargne retraite. Ces interrogations sont réalisées et publiées d'une part par les fédérations professionnelles (notamment la fédération française des sociétés d'assurance – FFSA –, le centre technique des institutions de prévoyance – CTIP – et l'association française de gestion financière – AFG), et d'autre part, depuis 2004, par la DREES dans le cadre de son **enquête annuelle sur la retraite supplémentaire facultative**, qui consolide les données des sociétés d'assurance, mutuelles et institutions de prévoyance (IP).
- Enfin, des données d'enquêtes auprès des établissements ou auprès des entreprises, pour ce qui concerne les produits de retraite supplémentaire proposés aux salariés dans le cadre professionnel. Interroger directement les employeurs permet notamment de mieux évaluer le champ des salariés couverts ou potentiellement couverts par ces produits, ainsi que les différences selon les caractéristiques des entreprises (notamment de secteur et de taille). Cependant, à ce jour, seul le PERCO fait l'objet d'un suivi statistique régulier de ce type, dans le cadre de l'**enquête annuelle ACEMO sur la participation, l'intéressement, les plans d'épargne et l'actionnariat des salariés** (dite « PIPA ») de la DARES.

En ce domaine, deux défauts d'information importants subsistent, liés, d'une part, à la couverture incomplète du champ de la retraite supplémentaire (en particulier, les régimes gérés en propre par les entreprises ne font l'objet d'aucune enquête statistique) et, d'autre part, à l'impossibilité de mettre en relation à un niveau individuel, pour les retraités, les montants des retraites supplémentaires facultatives avec ceux des retraites obligatoires de base et complémentaire – faute de disposer d'une source unique et suffisamment précise contenant à la fois l'information sur les retraites légalement obligatoires et celle sur les retraites supplémentaires et l'épargne retraite.

c) Opinions, souhaits et choix concernant la retraite

De nombreux sondages et enquêtes d'opinions sur la retraite sont réalisés par divers organismes publics ou privés, et leurs résultats publiés régulièrement dans les media – leur grand nombre empêche de les lister ici de manière exhaustive. Comme c'est toujours le cas en ce qui concerne l'expression qualitative d'opinions, par nature subjectives, les résultats sont sensibles au champ, à la méthodologie, aux formulations des questions et au contexte du questionnaire, et peuvent de ce fait varier d'une enquête à l'autre. Au sein des organismes du Service statistique public, la DREES réalise annuellement depuis 2000 un **baromètre sur l'opinion des Français à l'égard de la santé, de la protection sociale, de la précarité, de la famille et de la solidarité** qui permet de suivre, entre autres, les évolutions d'un certain nombre d'attentes et d'opinions relatives au système de retraite.

Les données sur les opinions et attentes présentent un intérêt accru lorsqu'elles peuvent être mises en relation avec d'autres données, plus objectives, sur les droits à retraite. Certaines enquêtes prévoient donc, dans le choix de l'échantillon interrogé, la possibilité d'être appariées avec les données des fichiers administratifs des régimes. C'est le cas, par exemple, de **l'enquête sur les Motivations de départ à la retraite** des nouveaux retraités du régime général. Cette enquête, réalisée tous les deux ans par la CNAV, le COR, la DREES et la DSS (première vague sur les nouveaux retraités de 2008-2009 et deuxième vague sur les nouveaux retraités de 2010-2011), permet d'apprécier les motivations de départ à la retraite, mais aussi la connaissance des notions attachées aux droits à la retraite et certaines attentes quant à l'âge de départ à la retraite des personnes parties récemment à la retraite. **L'enquête « Intentions de départ à la retraite »** de 2005 interroge quant à elle les personnes proches de l'âge de la retraite mais pas encore parties à la retraite, âgées de 55 à 59 ans. Pour les deux enquêtes, l'appariement avec les données administratives des régimes permet de confronter les attentes et les représentations subjectives avec des données plus objectives sur les droits à retraite acquis.

d) Retraite et emploi

Deux sujets particuliers sont à la frontière entre les thématiques d'emploi et de retraite : les parcours de fin d'activité – avec la transition, progressive ou directe, de l'emploi vers la retraite – et le cumul emploi-retraite.

Ces problématiques peuvent être éclairées au moyen des données issues des informations administratives des régimes, puisque les informations sur les cotisations et validations de droits fournissent une information sur l'existence d'un emploi. Ainsi, par exemple, les appariements entre l'EIR et l'EIC, ou entre l'EIR et le panel DADS (voir première partie) permettent d'étudier les transitions de l'emploi vers la retraite ou la juxtaposition de la retraite et de l'emploi.

Pour étudier de manière plus détaillée les caractéristiques précises des emplois des personnes en fin d'activité ou en situation de cumul emploi-retraite, **l'enquête Emploi** de l'INSEE peut être mobilisée⁹ En outre, des **modules complémentaires sur le passage de l'emploi à la retraite**, réalisés ponctuellement en complément de cette enquête (les deux derniers ont eu lieu en 2006 et en 2012), apportent un éclairage plus fin sur les circonstances dans lesquelles

⁹ Jusqu'en 2012, l'enquête Emploi ne permettait pas de bien identifier les situations de cumul emploi-retraite, mais le nouveau questionnaire mis en place depuis début 2013 devrait supprimer cette difficulté.

les personnes ont quitté le marché du travail et sur les intentions de départ de celles encore en emploi.

Une problématique spécifique relative aux transitions entre emploi et retraite concerne enfin l'effet sur ces transitions de l'état de santé ainsi que des conditions de santé tout au long de la vie. Pour l'étudier, des enquêtes mêlant questions sur la santé et renseignements d'un calendrier d'activité – c'est-à-dire une chronique, période après période, de toute la carrière jusqu'au passage à la retraite – sont disponibles, telles que l'enquête **SHARE** ou l'enquête **Santé et itinéraire professionnel** (SIP) de 2010. Des bases de données issues d'appariements entre données administratives des caisses de retraite et de maladie, telles que la **base de données Hygie** (couplant des données individuelles de la CNAV et de la CNAM-TS), devraient également permettre d'éclairer ces problématiques.