

16 juin 2016

Bernard AUBRY
Statisticien

Chiffres pour tous

Combien de personnes sont âgées de 65 ans ?

L'utilisateur s'emmêle dans les chiffres... l'Insee aussi

L'Insee dispose d'un fantastique réservoir d'informations chiffrées mais il ne sait pas suffisamment les mettre en valeur. Le site de l'institution, la « vitrine » de l'Insee, manque manifestement de clarté et l'accessibilité aux données laisse à désirer, surtout si on le compare à celui de ses homologues étrangers. De plus cette imperfection est susceptible d'engendrer des erreurs comme on le verra. D'où vient que la fonction de diffusion, qui devrait pouvoir bénéficier de la volonté affichée de l'Etat d'entrer résolument dans l'ère du numérique, semble n'avoir pas trouvé d'écho dans ce service statistique où, pourtant, la matière s'y prête bien?

L'Insee fête cette année ses 70 ans. Plutôt que de communiquer dans l'autosatisfaction, l'institution devrait s'engager dans une large ouverture à la critique constructive. Dans un pays qui a bénéficié des Lumières et qui aime à s'enorgueillir de cet héritage, accepter la critique ne devrait pas être une épreuve insurmontable (hors d'atteinte ne convient pas parce qu'on ne cherche pas les épreuves, on cherche à les éviter !). Ce pourrait être aussi une chance pour réussir une cure de jouvence.

PLAN

I – Avoir 65 ans

- 1 – Les différences observées, selon l'origine des chiffres
- 2 – Les données préétablies
- 3 – Le traitement des fichiers détails
 - a – Le fichier « *Données harmonisées des recensement de population* »
 - b – Le fichier détail régional¹ : focus sur la campagne 2011
 - c - Confrontation des variables AGED et ANAI
 - d – Quelles sont les conséquences de telles divergences?
- 4– Illustration des écarts observés
 - a- L'effet sur la pyramide : les 55 – 75 ans
 - b- Séries temporelles 2004-2011: confrontation ANAI-AGED
 - c- Saphir : une erreur dans la mise à jour ?

II- Que faire ?

- 1- Quelques pistes pour améliorer la diffusion du nouveau recensement
- 2- Un nouveau « *Projet de service public* » ?
 - a- La langue de bois ne sied pas à l'Insee
 - b- Réanimer le sens du service public
 - c- Innovation et territoires

III– Annexes - documentation et compléments

- 1- Les sources
- 2- Sur les valeurs de la variable CAMP incohérentes

Introduction

Ce n'est pas être anormalement curieux que de chercher à savoir combien on dénombre en France de personnes âgées de 65 ans (1)². Avec un peu de chance, on obtient un chiffre sur le site de l'Insee, en cliquant sur un lien bien précis, par exemple celui qui nous relie à la pyramide des âges. Mais il s'agit alors d'une information isolée et non d'un tableau qui permettrait un minimum d'analyse statistique dans le temps. Quant au mot-clé « *recensement* », il ne conduit nullement à l'objectif recherché, puisqu'il renvoie à des statistiques par classes d'âge et non à un âge détaillé. Pourtant la distribution de la population

¹ pour en savoir plus sur les fichiers détail, on pourra consulter le document suivant : <http://www.apr-strasbourg.org/detail-document-686-.html>

² Les chiffres entre parenthèses renvoient aux annexes

par âge détaillé était jusqu'à présent l'un des tableaux figurant en tête des fascicules de diffusion des résultats des recensements.

La diffusion des statistiques publiques ne passe plus aujourd'hui par des documents papier, mais par le canal d'internet. Le site de l'Insee, dans sa version actuelle est réputé d'une grande confusion. Le site ne propose pas, comme ses homologues étrangers, une arborescence claire des thématiques qui permet de trouver son chemin. Sauf à bien connaître le monde de la statistique, des enquêtes, des sources, des thématiques, des principaux indicateurs, etc., on est porté à lâcher prise très vite, tant l'offre de données est vaste... et surtout très mal ordonnée. La complexité du nouveau système de recensement et du mode de diffusion des résultats n'a fait qu'aggraver la situation. Il n'est pas question ici de remettre en cause le nouveau système de collecte puisqu'il présente bien des avantages par rapport à la méthode traditionnelle³. Il offre des opportunités pour une meilleure connaissance de la société. Mais la complexité inhérente à la diffusion des résultats n'a pas vraiment été anticipée. Il manque une vision d'ensemble. Pourtant, avec le nouveau recensement, mais aussi avec des informations localisées d'origine le plus souvent administrative et la capacité de mobiliser les données fournies par les recensements réalisés de façon analogue ou presque de 1962 à 1999, l'Insee possède un immense capital qu'il ne valorise pas comme il le devrait.

Ce document se focalise sur un objectif bien précis ; il s'agit de savoir combien de personnes ont un âge donné, en l'occurrence 65 ans, et comment évolue dans le temps cette population. L'intérêt d'une telle information n'est pas anecdotique. Les à-coups dans la pyramide des âges sont cause de difficultés pour tous les acteurs qui ont en charge les services à la population (en l'occurrence les personnes âgées). Plus généralement, dès lors que l'on veut calculer des indicateurs démographiques significatifs (taux de fécondité, de mortalité...) et de santé (prévalence des maladies), il faut pouvoir disposer de dénominateurs solides. On verra comment, si l'on n'y prend garde, les risques sont grands d'engendrer des erreurs grossières.

I - Avoir 65 ans

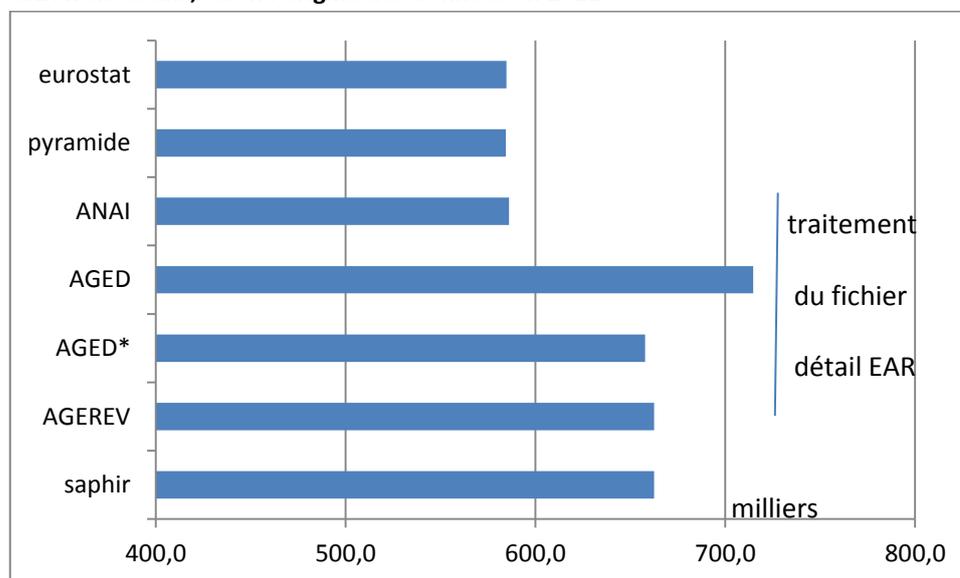
1 – Les différences observées, selon l'origine des chiffres

Le graphique G1 compare sept résultats obtenus soit par consultation du site d'Eurostat, soit par traitement des fichiers détail mis en ligne (fichier annuel pour l'EAR 2011), soit en ayant recours au fichier historique des recensements, intitulé *Données harmonisées des*

³ La revue Economie et statistique a consacré en mai 2016 un dossier au recensement rénové : http://www.insee.fr/fr/publications-et-services/sommaire.asp?ref_id=ECO483

recensements de la population 1968-2111, soit par l'ouverture de liens dont le chemin d'accès est plus ou moins facile à trouver. Les écarts observés sur la population âgée de 65 ans sont considérables, puisqu'ils oscillent entre 585000 et 715000 personnes. Les liens correspondants figurent en annexe.

G.1 Avoir 65 ans, selon l'origine des données en 2011



Conventions d'écriture pour distinguer l'origine des données (liens en annexe (1) :

-Eurostat

-Pyramide

-en majuscules les noms des variables du fichier détail, à savoir

- ANAI : année de naissance
- AGED et AGED-1: Age détaillé (par différence de millésime)
- AGEREV : Age révolu le jour de référence (le 3^{ème} jeudi de janvier)
- Saphir: fichier détail «Données harmonisées des recensements de la population »

On verra plus loin que tous ces écarts s'expliquent, mais deux questions fondamentales se posent: « Comment le quidam peut-t-il s'y retrouver? » et : « Que devrait faire l'institution pour crédibiliser les résultats et les rendre ainsi plus utiles à la compréhension de la société? »

2 – Les données préétablies

La consultation du site de l'Insee sur le mot-clé « recensement » est très décevante. Ni l'année de naissance, ni l'âge détaillé ne figurent dans la liste des modalités disponibles. Au mieux, même au plan national, on ne peut avoir que des chiffres pour une tranche d'âge très large, à savoir les 60 ans et plus. En revanche on dispose d'estimations par âge quinquennal, par sexe et département qui s'appuient sur les résultats des recensements. Avec les mots clé

« *Pyramide des âges 2011* », on dispose, en sus du graphique, d'un tableau par sexe et âge détaillé (disponible également pour l'année 2016).

Mais l'information la plus facile à utiliser, car elle répond exactement à la question, est celle mise en ligne sur le site d'Eurostat. Il s'agit d'estimations annuelles au 1^{er} janvier proposées par sexe, et région allant de 1975 à 2015. Les chiffres des dernières années sont affichés comme provisoires. L'intérêt de cette base est de permettre des comparaisons au niveau des régions de l'Union européenne (et même au-delà). Dans une région transfrontalière comme le Rhin supérieur, disposer immédiatement de données comparables est évidemment un avantage appréciable.

Les deux séries fournissent des résultats assez voisins. Il s'agit de la population au 1^{er} janvier. Ont 0 an les enfants nés l'année précédente, en l'occurrence l'année 2010 (pour la date de référence du 1.1.2011). L'inconvénient, majeur, de ces données, c'est qu'il est impossible de les décomposer selon les variables du recensement (nomenclatures territoriales – aire urbaine, commune... - et structurelles – diplôme, activité économique...). D'où la nécessité, pour en savoir davantage, d'avoir recours aux fichiers détail.

3 – Les données issues d'un traitement des fichiers détail

a) Le fichier «*Données harmonisées des recensements de la population*»

Le fichier intitulé « *Données harmonisées des recensements de la population* »(1) donne immédiatement la série des effectifs par sexe, *âge révolu* et région pour les recensements de 1968, 1975, 1982, 1990, 1999, mais on verra plus loin que la mise à jour pour les années 2006 et 2011, actuellement la dernière année disponible, n'a pas été faite correctement. C'est un fichier volumineux et, pour obtenir des résultats, il faut disposer d'un logiciel capable de traiter des masses de données importantes. L'utilisateur peut décomposer les résultats sur quelques critères, malheureusement trop peu nombreux.

b) Le fichier détail régional : focus sur la campagne 2011

Les fichiers détail mis en ligne annuellement sur le site de l'Insee proposent trois variables adéquates au problème posé et offrent ainsi la possibilité d'examiner de façon approfondie la question de la mesure des effectifs de la population âgée de 65 ans.

Rappelons que la diffusion des résultats se fait par millésime. EAR 2011 caractérise les résultats qui correspondent à l'agrégation des cinq campagnes constitutives, à savoir celles qui ont été effectuées au début des années suivantes : 2009, 2010, 2011, 2012, 2013. Il n'existe pas de code spécifique pour repérer dans le fichier les individus observés à telle ou telle campagne. Mais par le jeu du millésime, de l'âge, de l'année de naissance il est facile d'attribuer à chaque personne présente dans le fichier l'année de la collecte le concernant. On crée alors un code spécifique à la campagne (CAMP).

Savoir combien de personnes ont un âge donné n'est pas aussi simple qu'il y paraît si l'on ne dispose pas des dates précises (jour, mois et année) de naissance et d'observation. Déjà dans le passé, les recensements traditionnels induisaient parfois en erreur ceux qui ne prenaient pas le temps de réfléchir puisque les tableaux de résultats distinguaient deux notions différentes. L'une concernait l'âge révolu à la date de référence du recensement (un jour de la première décade de mars), l'autre l'âge obtenu par différence entre le millésime du recensement et l'année de naissance.

Par exemple : un enfant né le 1^{er} février de l'année 1999 aura 0 an à la date de référence (collecte du recensement en mars, un mois plus tard) dans les deux cas de figure. En revanche, s'il est né le 1^{er} décembre 1998, il aura ou 0 an ou 1 an selon la définition de l'âge retenu. Ainsi, tout dénombrement de la population au 1^{er} janvier de l'année du recensement fait coïncider les deux définitions de l'âge.

Avec le nouveau recensement, les choses sont devenues beaucoup plus délicates pour toute personne travaillant à partir des fichiers détail mis en ligne sur le site de l'Insee. Les données sont datées au jour de la collecte, le troisième jeudi du mois de janvier. Pour simplifier l'écriture, on supposera qu'il s'agit du 20 janvier.

Que dit le fichier détail, lorsqu'on totalise les effectifs pour le pays tout entier?

T1 - Ages dans les générations 1945 et 1946 : observation au titre de la campagne 2011

	Naissance en 1945 (ANAI)		Naissance en 1946 (ANAI)	
	Entre le 1.1 et le 20.1.1945	Entre le 20.1 et le 31.12.1945	Entre le 1.1 et le 20.1.1946	Entre le 20.1 et le 31.12.1946
	e0	e1	e2	e3
Effectifs (milliers)	7,7	108,5	8.2	145,9
Age d'après ANAI	65	65	64	64
Age d'après AGEREV	66	65	65	64
Age d'après AGED	66	66	65	65

Avoir 65 ans n'est pas en soi un concept clair. Il y a différentes façons d'avoir 65 ans. En effet, si on est né en 1945, avoir 65 ans ce peut être avoir eu son anniversaire dans le courant de l'année 2010 (2010 – 1945 = 65 ans), ce qui correspond ici à la variable ANAI. L'effectif correspondant est la somme e0 + e1.

Avoir 65 ans, c'est aussi être né entre le 20 janvier 1945 et le 19 janvier 1946. Il s'agit là de l'âge révolu (AGEREV) à la date de la collecte. Il s'agit alors de la somme e1 + e2.

Mais avoir 65 ans, c'est aussi être né en 1946 si l'on effectue une totalisation sur la modalité 65 de la variable AGED, à savoir e2 + e3.

En fait ANAI et AGED correspondent à des années entières. Dans ce contexte, AGEREV, qui chevauche deux générations, semble peu adapté à l'analyse statistique : il est dépendant de la date de référence (depuis 2004 : le troisième jeudi de janvier).

La collecte 2011 au titre de l'EAR 2011 conduit à trois résultats :

Selon ANAI : 116 200 (e0 + e1)

Selon AGEREV : 116 700 (e1 + e2)

Selon AGED : 154 100 (e2 + e3)

L'écart qui sépare AGED des deux autres est la conséquence du début brutal du baby-boom qui a produit ses effets surtout à partir de 1946.

c) Une confrontation des variables ANAI (année de naissance) et AGED (âge détaillé)

Quand on traite non plus une seule campagne (CAMP 2011), mais une série de 5 campagnes (EAR 2011), la totalisation effectuée sur la variable ANAI n'a plus rien à voir avec celle effectuée sur la variable AGED. Ainsi:

- Dans le cas de AGED, les éléments constitutifs appartiennent à **5 générations différentes** dont on sait qu'elles peuvent connaître des ruptures d'effectifs.
- Dans le cas d'ANAI, les éléments constitutifs appartiennent à la **même génération** c'est-à-dire que l'on additionne des effectifs par nature très proches les uns des autres. Ils ne diffèrent que par la mortalité, les mouvements migratoires et du fait de redressements statistiques d'ampleur mineure.

T.2 – 65 ans : comment se construisent les codes AGED et ANAI⁴ (EAR 2011)

CAMP	ANAI	AGED	AGED-1
0	1,3	1,3	1,1
2009	120,9	117,2	114,1
2010	118,5	118,5	115,3
2011	116,1	154,1	116,1
2012	114,5	161,6	152,7
2013	114,8	161,9	158,4
Total	586,2	714,7	657,8

En grisé : les chiffres qui font les grandes différences.

⁴ La variable AGEREV ne figure pas dans le tableau car elle est à cheval sur deux années de naissance

On constate en effet que la variable ANAI, qui se réfère à la même année de naissance, garde une valeur à peu près constante au cours des cinq collectes (écart de 6000 entre les extrêmes). La valeur centrale, pour 2001 (116,2), est proche de la moyenne des 5 collectes (117,0⁵).

En revanche la sommation sur la variable AGED se fait sur des effectifs très différents. Ainsi, en 2011, 65 ans correspond à la génération 1946, en 2012 à la génération 1947 et en 2013 à la génération 1948, dont les effectifs sont supérieurs de plus d'un tiers à la génération 1945. Mais même si l'on décalait AGED d'un an (AGED-1), l'écart resterait important.

Alors que dans les recensements classiques, âge et année de naissance concordent, avec le nouveau système, il n'y a concordance que pour les deux collectes de 2009 et 2010 si l'on retient AGED et pour les trois premières collectes (2009, 2010, 2011).

d) – Quelles sont les conséquences de telles divergences?

Les problèmes de cohérence n'ont pas la même importance selon la position de l'âge dans la pyramide. Il y a deux domaines critiques : l'un touche les plus jeunes, l'autre les sexagénaires. D'une façon générale, il y a risque de problème là où les effectifs des générations successives varient fortement.

T. 3 - Comparaison des résultats pour quelques âges / années de naissance

Age*	Génération	Eurostat	Pyramide	ANAI	AGED	AGED -1	AGEREV	Saphir	naissances
-3	2013			6,6					
-2	2012			161,3					
-1	2011			320,5		38,1			
0	2010	805,5	804,9	488,7	38,1	790,9	788,9	788,9	802,2
1	2009	799,7	799,1	643,1	790,9	791,0	790,6	790,6	793,4
2	2008	803,6	803,1	801,8	791,0	801,2	801,5	801,5	796,0
3	2007	801,0	800,4	797,5	801,2	803,1	802,3	802,3	786,0
4	2006	818,2	817,7	814,8	803,1	805,5	805,9	805,9	796,9
5	2005	803,2	802,7	802,3	805,5	803,8	804,0	804,0	774,4
63	1947	813,8	813,2	815,0	821,6	765,3	770,3	770,3	870,5
64	1946	773,8	773,2	775,8	765,3	714,7	717,1	717,1	843,9
65	1945	584,8	584,4	586,2	714,7	657,8	662,5	662,5	645,9
66	1944	570,8	570,4	573	657,8	603,5	606,0	606,0	629,9

*Dans ce tableau, âge et génération sont ici liés par la relation : ANAI+AGE=2010. Par ailleurs, AGEREV et Saphir sont identiques par construction.

⁵ Sans la collecte 0 (voir annexe 2)

Le tableau T.3 compare sept séries de chiffres pour les très jeunes et ceux qui ont autour de 65 ans auxquelles on en adjoint une huitième, celle du nombre de naissances enregistrées correspondant aux générations retenues.

Chez les plus jeunes, on voit d'emblée un problème lié à la méthode de diffusion retenue. La génération 0 an, celle née au cours de l'année 2010, est formée pour l'essentiel des enfants nés en France et encore vivants à la fin d'année. *Il y a une bonne cohérence avec les chiffres d'Eurostat et ceux de la pyramide diffusée par l'Insee qui donnent tous deux une répartition par âge au 1^{er} janvier 2011. C'est normal puisqu'il s'agit de données provenant toutes les deux d'estimations effectuées par la même institution, peut-être même par le même service⁶. Des redressements appropriés sont effectués pour corriger les effets « pervers » engendrés par la méthode retenue pour la réalisation du nouveau recensement.*

Mais, chez les plus jeunes, La série ANAI, composée des personnes recensées normalement en 2009, 2010, 2011, 2012 et 2013, est tronquée. ANAI=2013 ne peut se rencontrer que pour les enfants nés entre le 1 et le 20 janvier de la dernière collecte (2013). On leur donne donc un âge « négatif » (- 3 ans) puisque la référence 0 an correspond à l'année 2010. Les enfants nés en 2012 ne figurent que dans les collectes 2012 et 2013 (âge -2), etc. Pour les premiers âges, si l'on doit travailler à partir des fichiers détail, il vaut donc mieux retenir AGEREV et AGED-1, afin de se rapprocher des données au premier janvier que l'on trouve dans les pyramides des âges publiées par l'Insee et Eurostat.

Les divergences au titre de la classe 65 ans évoluent de 585 000 (Eurostat/pyramide/ANAI) à 658 000 (AGED-1) et même à 715 000 personnes (AGED). On rappelle que ANAI rassemble cinq collectes de la même génération, ici de faibles effectifs, que AGED-1 est formé de deux collectes aux effectifs forts et de trois autres nettement plus faibles, que AGED comprend une collecte forte de plus. *Pour les générations nées à la sortie de la guerre, il vaudra mieux retenir ANAI pour se rapprocher des données au 1^{er} janvier publiées par l'Insee et Eurostat.*

Ces écarts à différents endroits de la pyramide dont la solution n'est pas la même pose donc un problème pour construire une pyramide, en retenant une même variable à partir des fichiers détail. On ne peut guère imaginer commencer sur une variable dans le bas de la pyramide puis changer de pied au fil des âges !

4 – Les principaux écarts illustrés

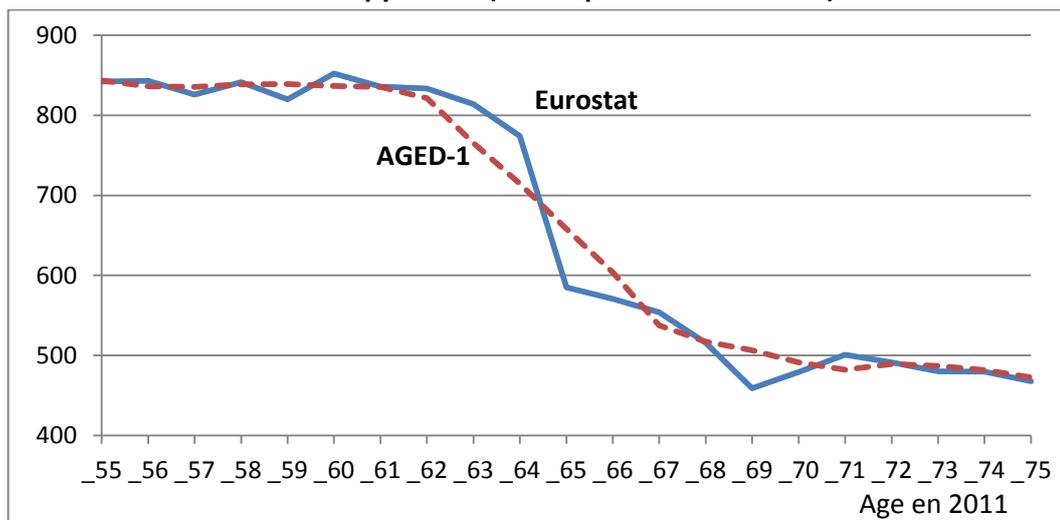
Pour mettre en évidence certaines contradictions entre les effectifs, on présente trois graphiques. Le graphique G.2 représente une classe d'âge centrée sur la génération 1945 (de 55 à 75 ans en 2010) et compare les données d'Eurostat avec la variable AGED-1. Les deux

⁶ Des divergences sont parfois la conséquence d'une discontinuité dans le champ géographique (Mayotte en particulier).

autres décrivent des évolutions temporelles. G3 porte sur la série des 11 collectes disponibles (2004 à 2014) et G4 sur une période plus longue (1990-2011) pour 4 âges (63 à 66 ans). On rappelle que les séries Eurostat/pyramide qui sont manifestement les plus crédibles, ne peuvent se décliner sur les variables du recensement, sinon par région et sexe. Quant à composer des tableaux avec les autres variables, notamment avec la série ANAI, on peut le faire à condition de bien réfléchir à la signification des indicateurs obtenus.

a) L'effet sur la pyramide: les 55-75 ans

G.2 - l'effet sur la pyramide (extrait pour les 55 à 75 ans) en 2011



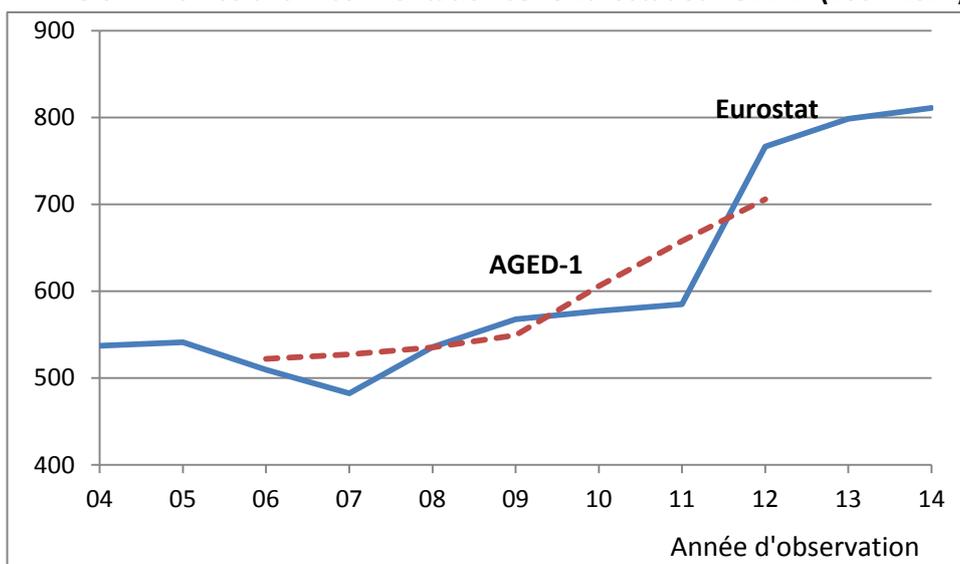
Le graphique confirme que la série Eurostat rend bien compte de l'effet du baby-boom en 1946. En revanche, la série AGED-1 donne à peu près les mêmes résultats aux accidents de pyramide près. La pyramide se trouve considérablement lissée.

Ce lissage risque de produire un écrasement des évolutions d'indicateurs qui peut avoir des conséquences fâcheuses pour les études et les rapports qui sont rédigés à partir de l'exploitation des fichiers détail (par exemple quand il s'agit de taux de chômage ou d'emploi dans la population en âge d'entrer ou de sortir de l'activité).

b) – Séries temporelles 2004-2011 – confrontation ANAI - AGED-1

Bien évidemment, les divergences qui apparaissent à un moment donné ont des répercussions sur les indicateurs de variation d'une année à l'autre. Le graphique G.3 compare les effectifs des 65 ans selon deux sources : le fichier Eurostat et la variable AGED-1 du fichier détail.

G.3 – Avoir 65 ans – confrontation série Eurostat et AGED-1 (2004-2014)



Le graphique appelle deux remarques. D'une part le mode de diffusion de l'EAR prive l'utilisateur de 4 années sur onze (2004, 2005, 2013 et 2014). D'autre part, alors que la courbe Eurostat décrit les changements annuels, la courbe lissée AGED-1 ne dit pas grand-chose. C'est d'ailleurs pour cela que l'Insee recommande de ne pas publier de variations annuelles de l'EAR (en toute rigueur on ne devrait pouvoir comparer que les variations de l'année $n+5$ par rapport à n , ce qui est particulièrement frustrant !⁷).

Les données des EAR, en principe comparables, sont elles-mêmes « écrasées » puisqu'elles regroupent 5 ans ce qui, au bout du compte, appauvrit considérablement l'intérêt des résultats obtenus. .

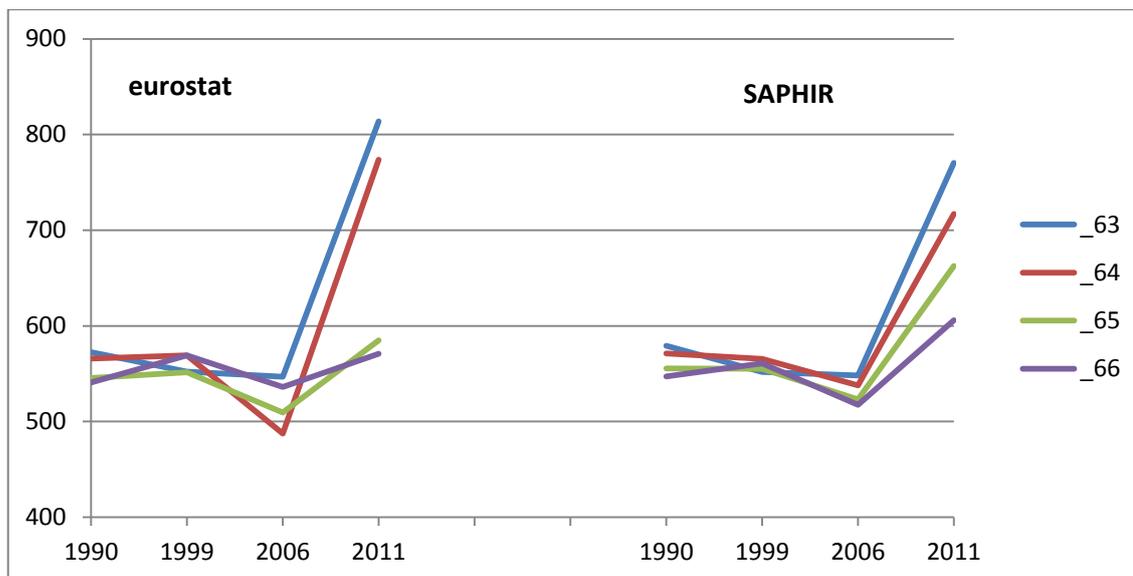
Pourquoi l'Insee n'introduit-il pas la variable CAMP dans les fichiers détail? Le traitement des données serait alors beaucoup plus facile à effectuer. De toute évidence des séries annuelles sur 11 ans auraient manifestement un sens, pour autant, évidemment, que l'on soit prudent dans le choix des variables géographiques et structurelles.

d) – Saphir : une erreur dans la mise à jour sur les âges 63 à 66 ans ?

Les écarts observés sur les graphiques précédents sont certes regrettables, mais ils ont une explication rationnelle, sinon cohérente. Il n'en est pas de même lorsque l'on compare la série Saphir avec la série Eurostat.

⁷ Pourquoi 5 ans ? Parce que c'est le temps d'une rotation. Si l'on avait choisi une autre option (étalement de la collecte sur sept ans, ou sur trois ans), on aurait une autre description de la réalité.

G.4 – Evolution des effectifs par âge détaillé : Eurostat met clairement en évidence la rupture de 1946, l’EAR non.



En réalité il s’agit du fichier soi-disant harmonisé, incorrectement mis à jour sur le site⁸.

A gauche, on observe une série composée des données d’Eurostat (disponible depuis 1990, donc aussi en 2011). A droite, il s’agit de la série Saphir (recensements classiques de 1990 à 1999, mis à jour en 2006 et 2011). L’image que l’on a des variations d’une année à l’autre n’est manifestement pas la même.

Pour comprendre ce qui se passe, on se réfère au tableau T.4 qui donne les chiffres correspondant aux graphiques avec en plus, pour 2011, un chiffre correspondant à ANAI ou à AGED-1.

T.4 - Les chiffres du graphique: la mise à jour de Saphir n’aurait pas dû se faire sur la variable AGED-1⁹, mais sur ANAI

Age	EUROSTAT				ANAI	SAPHIR				AGED-1
	1990	1999	2006	2011		1990	1999	2006	2011	
_63	573	552	547	814	814	579	552	548	770	765
_64	566	569	488	774	774	571	565	538	717	715
_65	545	552	510	585	585	556	555	523	663	658
_66	541	569	536	571	571	547	561	517	606	604

⁸ En fait, sur le site, l’Insee a d’abord fait parfois référence à Saphir, puis a omis de le faire. Au fond, il vaut peut-être mieux qu’il en soit ainsi, puisque l’esprit de Saphir (Système d’Analyse de la Population par l’Historique des Recensements) était précisément de mettre à disposition des utilisateurs un fichier assurant une cohérence satisfaisante des effectifs observés lors des recensements successifs...

⁹ En fait la mise à jour a été faite avec la variable AGEREV, très voisine de AGED, avec décalage d’un an

Comprendre le tableau :

- En 1990 et 1999: cohérence entre Saphir et Eurostat

- En 2006 et 2011 :

* la variable ANAI est cohérente avec Eurostat, donc pas de problème

*la variable AGED-1 est cohérente avec Saphir...

... mais ANAI et AGED-1 ne sont pas cohérentes entre elles comme on l'a montré. Par conséquent la mise à jour de Saphir par la variable AGED-1 n'est pas cohérente avec celle d'Eurostat qui fait référence à ANAI.

Si donc l'Insee lui-même ne domine pas les chiffres qu'il diffuse, qu'en sera-t-il des autres utilisateurs, qui n'ont ni la même expérience du traitement statistique, ni l'accès à toutes les séries archivées ?

II - Que faire?

1 – Quelques pistes pour améliorer la diffusion du nouveau recensement

On imagine que les responsables de l'Insee se sont déjà interrogés sur la nécessaire amélioration de la diffusion des résultats du nouveau recensement. Cependant, bien que l'on ne dispose pas de tous les éléments nécessaires à l'analyse, on se propose d'exprimer certaines réflexions alimentées à la fois par l'examen du système de diffusion actuel et par l'expérience du fonctionnement interne de l'institution. On s'appuiera aussi sur quelques textes, notamment le *Code européen des bonnes pratiques statistiques*(2) qui énonce quinze principes, dont cinq au moins ont à voir avec la diffusion des résultats des recensements.

Examinons les enjeux de la diffusion des résultats du recensement, d'une façon générale.

Si l'on se place au plan national, mais aussi régional, la réponse à la question « *Que faire ?* » semble s'imposer. Il faut publier des résultats annuels relatifs à chaque collecte, quitte à apporter quelques redressements supplémentaires (mineurs) aux fichiers. Ces résultats porteraient sur les principales variables et donneraient la possibilité de signaler sans tarder les principaux changements en cours dans la société. Il faudrait envisager l'abandon des variables AGEREV et AGED au profit d'une seule, ANAI (année de naissance), une variable liée sans ambiguïté à une variable AGE (ainsi en 2010¹⁰, ANAI = 1945 et AGE=65 auraient la même

¹⁰ Ce qui implique nécessairement un décalage: soit fixer au 31.12 la référence, soit diminuer systématiquement d'une unité la variable AGE.

signification), à condition toutefois de trouver une méthode d'estimation pour les trois premiers âges. Les conséquences sur la qualité des données seraient minimes, mais l'avantage en matière de **lisibilité** serait considérable¹¹. La base serait parfaitement cohérente avec celle d'Eurostat.

A propos de la crédibilité des chiffres publiés. Comment peut-il se faire que les chiffres traités par l'Insee et mis en ligne sur le site européen ne se retrouvent pas sur le site de l'Insee? On dispose ainsi, depuis le 25 février 2016 d'un tableau, par sexe, âge détaillé et anciennes régions (26 régions) au 1^{er} janvier 2015 sur le site d'Eurostat, chiffres qui ne semblent pas destinés à figurer sur le site de l'Insee¹². C'est aussi semble-t-il le cas pour d'autres domaines, les flux migratoires par exemple. Les chiffres mis en ligne sur Eurostat ne le sont pas sur le site de l'Insee.

Cependant la mise en ligne de données annuelles pose **un problème majeur** : que faire en matière de diffusion des résultats pour les niveaux infrarégionaux ? En effet, chaque année seulement une commune sur 5 de moins de 10000 habitants fait l'objet d'une collecte.

Deux options entre autres semblent se présenter :

- a) Envisager une double diffusion, à l'instar de ce qui se faisait dans le passé (d'abord des résultats portant sur un logement sur quatre, puis un autre portant sur l'exhaustif¹³) : une première diffusion au plan national, éventuellement régional, en juin de l'année n+1 de résultats de la dernière collecte et une autre deux ans plus tard, par agrégation des 5 dernières collectes, au niveau communal.
- b) Effectuer une diffusion annuelle et unique, moyennant des redressements appropriés, faisant appel à des techniques élaborées prenant en compte des informations exogènes.

De toute façon, l'agrégation de cinq recensements présente des inconvénients importants puisqu'ils « écrasent » les changements rapides intervenus sur les territoires. C'est vrai au niveau national et régional, mais ça l'est surtout à l'échelle des communes où, en l'espace de quelques mois, la fermeture/l'ouverture d'un établissement peut perturber notablement le marché du travail local. Compte tenu de l'existence d'autres fichiers communaux (fichiers d'établissements, système SIRENE, fichiers des demandeurs d'emploi notamment), on pourrait imaginer, *et ce serait une autre conception de la diffusion des données locales*, qu'à toute diffusion de résultats communaux issus des recensements, soient associés, par un lien *ad hoc* et de façon claire, des informations complémentaires, éventuellement qualitatives, qui aideraient les chargés d'études, en particulier ceux qui ont pour tâche d'éclairer les décideurs locaux, dans leur travail de présentation d'une image de leur territoire. On ne doit pas oublier

¹¹ Une correction serait d'autant plus facile à effectuer qu'elle n'aurait comme effet que de modifier très marginalement le code pondération (IPONDI)

¹³ On acceptait alors de trouver des différences pour une même variable. Il est fréquent que les statisticiens diffusent des données provisoires, puis quelques années plus tard des données définitives.

que le système actuel fournit non seulement des données anciennes mais aussi des données qui ne correspondent pas à la réalité du millésime auxquelles elles sont référencées.

L'Insee prône l'innovation (voir plus loin). L'institution doit montrer qu'elle ne la combat pas, que la notion de « veille » n'est pas un mot creux ; elle doit relever le défi du numérique pour faire mieux encore que ses homologues étrangers en matière de diffusion des données grâce à des sources d'une exceptionnelle richesse.

2 – Un nouveau « *Projet de service public* » ?

Les défauts de la diffusion ne peuvent être dissociés de certaines faiblesses managériales qui entravent le fonctionnement efficace de l'institution.

a – La langue de bois ne sied pas à l'Insee

Faut-il souligner que faire des propositions n'a de sens que dans le contexte d'un minimum d'ouverture de l'institution (4). Pour avoir une chance d'aboutir, il faut que les acteurs qui prennent des décisions soient perméables aux suggestions qui ne sont pas nécessairement issues du cénacle. L'expérience montre amplement qu'il existe à l'Insee comme dans bien d'autres "corps constitués" une surdité à ce qui vient de l'extérieur ou des rangs inférieurs. Plusieurs des idées qui sont exprimées ici l'ont déjà été d'une façon ou d'une autre à travers des documents internes pendant la période de mon activité à l'Insee et, plus tard, dans le cadre des réflexions qu'une participation aux travaux de l'APR (5) a rendu possibles. L'Insee donne, hélas plutôt dans l'autosatisfaction comme en témoigne ce texte extrait de la mise en ligne de l'annonce des 70 ans de l'institution:

« L'innovation constitue une pièce essentielle d'une démarche plus générale : la qualité. Des normes de qualité - y compris dans les règlements européens depuis 1997 - s'imposent, notamment sur les délais de publication et la « méta-information ». En 2006, un plan qualité est mis en place à l'Insee, il étend et prolonge le processus commencé en 1999 avec la démarche « Statistique publique et entreprises ». La transparence de l'activité de l'Institut, indispensable à sa crédibilité, est mise en avant, aussi bien à propos de la gouvernance de la statistique publique que des processus statistiques ou des résultats eux-mêmes. La mise en œuvre de la Lolf (loi organique relative aux lois de finances) modernise le management et encourage l'efficacité du dispositif statistique. Enfin, l'Insee s'oriente vers un suivi plus systématique de la satisfaction des utilisateurs ».

Ce texte est un petit bijou de langue de bois pour qui connaît le fonctionnement interne de l'Insee. Au moins pour ce qui concerne la statistique démographique et sociale, il n'est pas certain que la diffusion des données de l'Insee satisfasse les attentes des utilisateurs familiers des sites des homologues étrangers. Une institution sérieuse n'a rien à gagner à manier la langue de bois et moins encore un institut statistique: les chiffres, pour être crus, doivent être portés par un langage rigoureux. Force est de reconnaître que le CNIS, censé mettre en contact

l'Insee avec les utilisateurs, n'est pas parvenu à réveiller l'Insee¹⁴. Au-delà du manque d'esprit critique de l'institution sur ses véritables performances, il faut ajouter un manque de curiosité pour les pratiques de nos voisins européens, pourtant accessibles lors de visites de leurs sites¹⁵. L'Insee a été capable d'imposer un système de recensement de population original, alliant plutôt judicieusement les contraintes administratives et juridiques. A notre avis le succès technique est gâché par un système de diffusion bâclé.

b – Réanimer le sens du service public

Le fichier historique des recensements mis en ligne par l'Insee (appelé « *Données harmonisées des recensements de la population pour la période 1962-2011* ») ressemble étrangement à une extraction du fichier Saphir pour la période 1968-1999 qui aurait été mise à jour avec les données des EAR 2006 et 2011. Mais la relation malsaine qu'a entretenue la hiérarchie parisienne avec le fichier Saphir mis au point par l'Insee-Alsace n'a pas été de toute évidence bonne conseillère, puisque la base actuellement mise en ligne est erronée¹⁶...

Pourtant la création du fichier historique et la première phase de sa mise au point n'auraient pu se faire sans un regard bienveillant de la part de la direction générale. En effet celle-ci avait engagé, en 1989, un *Projet de service public* (PSP) qui offrait la possibilité aux agents qui le souhaitaient, de participer activement au service public de la statistique. Tous les agents, et pas seulement le sommet de la hiérarchie, étaient appelés à s'impliquer davantage, à faire des propositions¹⁷. L'institution devait « *mobiliser toutes les compétences pour une meilleure efficacité* ».

Las! Les réflexes propres à l'administration ont vite repris le dessus et il en est advenu du *Projet de service public* ce qu'il advient à bien des réformes : il a été vite oublié. Des principes sans valeur intrinsèque (par exemple l'obligation de mobilité, pour que l'agent surtout ne s'attache pas trop à son travail (*sic*) ont bien vite supplanté le souci de rigueur (une valeur que les statisticiens se plaisent à exciper) et la mission d'informer au mieux les citoyens et les décideurs.

L'investissement Saphir n'a pas pu pleinement s'accomplir. Les exploitations du fichier ont juste été tolérées et les propositions d'extension à d'autres enquêtes ignorées. La décision de mise sur le site d'un fichier Saphir est venue tardivement (décision de la DG d'avril 2008, soit près de 20 ans après le début des opérations, fin 1989). Il s'agit d'un fichier appauvri et privé de ses potentialités. Saphir est devenu un fichier individuel alors qu'il s'agissait d'un fichier hiérarchisé avec des niveaux Familles et Ménages permettant des analyses rétrospectives

¹⁴ Il existe au sein du CNIS (Conseil National de l'Information Statistique) une Commission Nationale d'Evaluation du Recensement de la Population (CNERP) : <http://www.cnis.fr/cms/Accueil/activites/Organisation/Cnerp>

¹⁵ Un récent article du Figaro (23 mai 2016) compare les sites Destatis et Insee (3).

d'une grande richesse. La direction régionale d'Alsace n'a guère été récompensée de son initiative puisque la mise à jour du fichier a été, par le fait du prince, transférée à une autre direction régionale. On a vu les erreurs auxquelles ont conduit ces décisions bureaucratiques qui croient pouvoir faire de la bonne statistique sans prendre en compte l'engagement professionnel des agents. L'objet premier de Saphir - réunir en un même fichier l'essentiel de l'information collectée lors des recensements et de ce fait de traiter des séries longues sans avoir à se soucier de la cohérence des chiffres - a été perdu de vue. Il est pour le moins cocasse de voir les erreurs commises par l'Insee échapper au fichier d'Eurostat, où d'aucuns pensent que la bureaucratie sévit encore plus intensément qu'à l'Insee. C'est cocasse, mais désolant. L'institution doit réviser ses méthodes de management. Cette légèreté dans la gestion des ressources humaines entretient, bien inutilement, la défiance à l'égard de la statistique publique. Il ne faut pas s'étonner si, comme l'a écrit Claude Thélot, on assiste à un certain discrédit de la statistique¹⁸.

3 - Innovation et territoires

A l'époque où les premiers travaux de fusion des fichiers des recensements ont débuté (fin de l'année 1989), l'Insee-Alsace se singularisait et la revue régionale, *Chiffres pour l'Alsace*, partageait avec celle de la Corse le premier rang en matière de taux de pénétration (nombre d'abonnements payants rapporté à la population régionale). La région Alsace, singulière en soi, avait coutume de prendre des initiatives. Elle fut ainsi la première à engager des relations suivies avec ses homologues d'outre-frontière (Office statistique du Bade-Wurtemberg et de la Rhénanie-Palatinat, Offices fédéral et cantonaux suisses)¹⁹. Plus tard elle réalisera des enquêtes auprès des travailleurs frontaliers, une enquête sur le mode de vie des Alsaciens, un inventaire des équipements sportifs, un suivi détaillé de l'emploi d'après les bordereaux Urssaf, une enquête sur les liaisons industrielles, etc. Cela marchait à merveille jusqu'à ce que la DG intervienne et décide elle-même de la spécialisation de chaque région. Paris distribua des « pôles de compétence » dans les régions, sans guère se soucier de la présence ou non de compétences locales. De sorte que la DR d'Alsace, qui n'était alors plus en cour, se vit gratifier d'un demi-pôle (partagé avec l'Insee Haute-Normandie), sur un domaine pour lequel elle n'avait pas montré d'aptitude particulière et qui n'avait aucun intérêt conceptuel, à savoir la

¹⁸ Sur le thème *Statistique et Société*, Claude Thélot, inspecteur général de l'Insee, puis conseiller-maître à la Cour des comptes, écrit notamment : «... Satisfaction car la victoire de la statistique – en tout cas comme moyen de mesurer un certain nombre de phénomènes même si cette mesure est partielle – est assez forte, assez grande, assez notable. Et en même temps crainte que cette victoire même n'entraîne un certain discrédit de la statistique. Nous pensons en effet qu'elle n'a pas gagné en crédibilité et cela sur deux plans. Dans une partie de l'opinion éclairée, le fait que la statistique dise quelque chose « d'exact » n'est plus aussi cru qu'autrefois et qu'en particulier la justesse de l'information chiffrée qui est ainsi obtenue, qui est ainsi diffusée, est mise en cause pour différentes raisons... ». *Raisons explicitées plus loin. Il y aurait matière à réflexion pour les agents de l'Insee lors de leurs réunions annuelles ! Ce texte est extrait de l'ouvrage « Echantillonnage et Méthodes d'enquêtes » sous la direction de Pascal Ardilly—Dunod, 2004, page 2.*

réponse au téléphone, travail consistant à dispatcher auprès des spécialistes les questions posées par les clients de l'institution. On retrouve bien ici une caricature de l'administration française, une variante du « mal français » que tout le monde connaît et dénonce mais qui n'a jamais pu trouver de solution.

En 2008, l'ensemble des chefs de service d'études et de diffusion de la DG se sont retrouvés pour leur séminaire annuel dans la charmante bourgade de Klingenthal. Une telle localisation aurait pu être l'occasion de favoriser des échanges avec d'autres statisticiens.

D'une part avec leurs collègues d'Outre-Rhin, d'autant plus que les agents de l'Insee ne sont guère encouragés à regarder ce qui se passe chez leurs partenaires européens. Il y a quarante ans, nos collègues d'Outre-Rhin étaient en admiration devant le système statistique français. Aujourd'hui nous aurions beaucoup à apprendre de leur expérience(3), non seulement en matière de mise à disposition au public des données statistiques, mais aussi en matière de méthodes de production statistique.

D'autre part avec des statisticiens régionaux pour aborder les questions spécifiques à la région (le transfrontalier, l'expérience Saphir...).

Des échanges pour faire le point autour des « *bonnes pratiques en matière de diffusion* » et, pourquoi pas ? une réflexion sur l'évolution du management dans l'esprit du « *Projet de service public* » n'auraient pas été de trop. Mais rien de tout cela ne se produisit. La lecture du compte rendu de la réunion de la DG est instructive. Elle donne l'image d'un groupe qui « *tournerait en rond dans l'Hexagone* », un groupe qui aurait pris le parti du repli sur soi.

L'Insee aurait bien besoin d'un nouveau « Projet de service public » en forme de mise à jour du précédent, un PSP qui insisterait davantage sur la nécessité d'une pédagogie appropriée, notamment à destination des médias, d'une plus grande disponibilité à la critique tant interne qu'externe et d'une incitation plus forte à s'ouvrir à l'international.

III – Annexes - documentation et compléments

1- Les sources

*Eurostat

Par région de la classification européenne nuts 2 (disponible pour les 28 pays de l'UE, mais aussi pour quelques autres pays voisins comme l'Albanie, la Turquie, etc.
<http://ec.europa.eu/eurostat/fr/data/database>

Arborescence : Exploration des données/base de données par thèmes/Statistiques générales et régionales/St. régionales par classification NUTS/St. Démographiques régionales/Population au 1^{er} janvier NUTS2

*Pyramide : On accède à toutes les pyramides d'âge de 1901 à 2011 avec le lien
http://www.insee.fr/fr/themes/detail.asp?ref_id=ir-sd2010&page=irweb/sd2010/dd/sd2010_population.htm

Curieusement pour la pyramide des âges 2011, les chiffres sont issus d'une « estimation provisoire ». Pour combien de temps encore ? En fait la série a l'air d'être abandonnée. En faisant :

https://www.google.fr/search?q=pyramide+des+ages+2011&ie=utf-8&oe=utf-8&client=firefox-b&gfe_rd=cr&ei=KdZPV4XII-t8wfbm4yACg

on obtient une autre pyramide avec des chiffres estimés en 2013.

*Fichiers détail annuels des EAR

http://www.insee.fr/fr/themes/detail.asp?req_id=99&ref_id=fd-rp2011&page=fichiers_detail/RP2011/telechargement.htm#RP2011_INDREG

Les fichiers détail ne peuvent être exploités que par des utilisateurs disposant de logiciels de traitement statistiques, mais ces logiciels sont de plus en plus répandus et les ordinateurs de plus en plus puissants.

*Saphir- Fichier historique des recensements Ce fichier ressemble étrangement au fichier Saphir (Système d'analyse de la population par l'historique des recensements).

<https://www.google.fr/#q=donn%C3%A9es+harmonis%C3%A9es+des+recensements+de+la+population>

D'autres liens permettent en principe d'accéder à des données d'âge :
<http://www.bdm.insee.fr/bdm2/choixCriteres?codeGroupe=63>, une arborescence mais peu de choix Ce qui est bizarre c'est que le tableau s'arrête en 2011 sur très longue période. Ou
<http://www.bdm.insee.fr/bdm2/choixTheme?code=44>.

A noter également :

Estimation annuelle par région et département de 1975 à 2015 par **âge quinquennal** :

http://www.insee.fr/fr/themes/detail.asp?ref_id=estim-pop®_id=99

D'une façon générale, on rappelle que les recherches sur le site de l'Insee sont d'autant plus pénibles à effectuer qu'elles s'inscrivent dans un système de diffusion très mal organisé.

2- Sur l'Insee et le Projet de service public

*Les 70 ans de l'Insee (juin 2016) : document d'information

<http://www.insee.fr/fr/insee-statistique-publique/default.asp?page=connaître/histoire-insee.htm>

*Sur le Projet de service public de l'Insee (PSP 2011)

<http://www.apr-strasbourg.org/detail-document-702-.html>

*Le service public en recherche. Extrait d'un ouvrage de la Documentation française, un article de François Turlot, inspecteur général de l'Insee, qui a inspiré le PSP.

<http://www.apr-strasbourg.org/membre/2054/aubry-bernard/categorie/15-mthodes.html>

*Code des bonnes pratiques de la statistique européenne :

[http://ec.europa.eu/eurostat/documents/3859598/5921941/KS-32-11-955-FR.PDF/f1dfd441-391c-4c15-94f2-b34a43697f55 -](http://ec.europa.eu/eurostat/documents/3859598/5921941/KS-32-11-955-FR.PDF/f1dfd441-391c-4c15-94f2-b34a43697f55-)

Ce document énonce et commente 15 principes. Plusieurs d'entre eux pourraient servir à structurer une critique plus approfondie sur la diffusion des données sociales (principes de **cohérence, de comparabilité, d'accessibilité, de clarté**).

3- Sur l'ouverture à l'international

Les agents de la fonction publique ne sont guère incités par leur hiérarchie à regarder ce qui se passe au-delà de la frontière. Ce critère compte peu pour la carrière.

Pourtant la comparaison des sites nationaux est toujours instructive. On invite les lecteurs à ouvrir les liens « *Insee coopération internationale* » et « *Destatis internationale Kooperation* ». Ils constateront deux façons de présenter les actions de coopération. Bien noter toutefois que c'est le fond, plus que la forme qu'il faut comparer, ce qui exigerait une analyse spécifique.

Pour l'Insee :

<http://www.insee.fr/fr/insee-statistique-publique/default.asp?page=insee-activites/methodes-statistiques.htm>

Pour Destatis :

<https://www.destatis.de/DE/UeberUns/UnsereAufgaben/InternationaleKooperation/StatistikKooperation/KooperationStatistik.html>

on accède notamment à une brochure d'une présentation attrayante:

https://www.destatis.de/DE/UeberUns/UnsereAufgaben/InternationaleKooperation/StatistikKooperation/Publikationen/I2016.pdf?_blob=publicationFile

Dans Le Figaro du 23 mai 2016, Jean-Pierre Robin compare les sites Destatis et Insee. Selon lui, la confrontation ne se fait pas à l'avantage de la France. Il y voit la marque des "*rapports sociaux opaques et confus en France*" et conclut qu'"*un peuple a les statistiques qu'il mérite*".

Par ailleurs, l'Insee devrait se préoccuper de l'usage qui est fait des chiffres qu'il communique à fin de diffusion sur des sites étrangers. En l'occurrence le site de l'IPUMS (base de données internationale, université du Minnesota) comprenait d'importantes anomalies dans sa section France (les données pour la région PACA étaient omises pour 4 recensements et celles de Corse pour 3 recensements)

<http://www.apr-strasbourg.org/membre/2054/AUBRY-Bernard/document/604.html>

4 - Sur la communication de l'Insee

L'institution reste encore peu ouverte aux critiques. La tentation est de ne pas répondre au courrier ou de « *botter en touche* ».

Ainsi pour ce qui concerne un dossier resté sans réponse qui portait sur la diffusion des données publiques, adressé le 3 août 2010 au Directeur général de l'Insee, avec copie au Président de l'Autorité de la statistique publique et du Président du CNIS.

<http://www.insee.fr/fr/themes/donnees->

[locales.asp?typgeo=COM&typesearch=territoire&search=Mulhouse+%2868224%29&ref_id=emp2012](http://www.insee.fr/fr/themes/donnees-locales.asp?typgeo=COM&typesearch=territoire&search=Mulhouse+%2868224%29&ref_id=emp2012) Le document signalait notamment une anomalie concernant la commune de Mulhouse.

Qu'il soit difficile de corriger cette erreur peut se comprendre, que l'on n'informe pas l'utilisateur est bien plus grave puisque de telles anomalies contribuent à jeter le discrédit sur la statistique publique en général.

Quand le journal Le Monde consacrait (1 août 2012) un article aux doutes de beaucoup de maires sur la qualité des chiffres de population légale de leur commune, la réponse de l'Insee fut peu convaincante. Elle laissait croire que l'incertitude au plan national du nouveau recensement était de 15000 personnes. Il s'agissait ainsi de montrer que le plan de sondage du nouveau recensement est remarquable. Mais le lecteur aurait besoin de disposer d'une indication sur les incertitudes de collecte (omissions, doubles comptes, erreurs diverses), notamment dans les quartiers réputés difficiles.

5 - Sur l'Association de Prospective Rhénane (APR)

Le site de l'association <http://www.apr-strasbourg.org/> propose un certain nombre de documents sur le thème de l'innovation et de la créativité, tout particulièrement s'agissant des territoires.

On trouvera également sur mon blog personnel des documents permettant de comprendre Saphir et sa démarche, en particulier

<http://www.apr-strasbourg.org/membre/2054/aubry-bernard/document/681-saphir-prsentation-partie-1.html>

Pour en savoir plus sur les fichiers détail, on pourra notamment consulter de document : « *La diffusion des recensement à l'heure de l'open data* » : <http://www.apr-strasbourg.org/detail-document-686-.html>

2- A propos des valeurs de CAMP incohérentes

On a vu (T.2) que quelques milliers d'individus ne pouvaient pas être classés dans l'une cinq collectes.

Le repérage de l'année de la collecte repose sur la relation :

$$\text{millésime de l'EAR} - (\text{ANAI} + \text{AGED})$$

Cette formule correspond à l'année centrale de la période de 5 ans : ainsi pour l'EAR2011, la valeur centrale correspond à la somme ANAI=1945 + AGED=66.

Mais on a vu que pour une même année de naissance, on rencontre cinq âges différents: 64, 65, 66, 67 et 68 ans. Entre l'âge réel et l'âge « central », on a donc un écart qui, en principe prend les valeurs -2, -1, 0, 1 ou 2.

De fait, pour les millésimes 2006 à 2008 de l'EAR, on observe parfaitement cette propriété. En revanche à partir de l'EAR 2009, on observe des âges sortant de l'intervalle. Ce sont toujours alors des écarts supérieurs à +2. Les écarts peuvent même atteindre 6 ans.

Les tableaux qui suivent correspondent à une exploitation spécifique des trois principaux fichiers détail (régional, cantonville et logement). Ils ne concernent que la population des ménages. Il faut donc considérer ces résultats que comme des ordres de grandeur.

T.5 – Population née en 1945, effectif selon l'âge (milliers)

Age	Ecart d'âge (par rapport à 66 ans)	Effectif
61	5	0,4
62	4	0,5
63	3	0,4
64	2	116,3
65	1	114,1
66	0	111,5
67	-1	110,2
68	-2	110,5

A l'EAR 2011, les personnes nées en 1945 devraient avoir entre 64 et 68 ans. En fait on en trouve quelques centaines qui ont un âge supérieur à l'âge attendu (AGED).

T.6 – Population née en 1945, les écarts anormaux selon le millésime de l'EAR (milliers)

Ecart d'âge	2009	2010	2011	2012
3	0,8	0,7	0,4	0,7
4		0,6	0,5	0,3
5			0,4	0,3
6				0,2

En 2011, on retrouve les chiffres du tableau 5

T.7 – Evolution des effectifs par campagne pour quatre villes *ayant franchi le seuil des 10000 habitants entre 1999 et 2006

Camp/EAR	2006	2007	2008	2009	2010	2011	2012
0				49,6	33,2	22,9	12,1
2004							
2005							
2006	53,8	54,5	54,9				
2007							
2008							
2009				5,0	10,6	10,6	10,8
2010					10,6	10,7	10,8
2011						11,1	11,1
2012							11,5
Total	54,1	54,5	54,9	54,5	54,5	55,3	56,3

**Liste des villes: Tournon-sur-Rhône (07324) – Auriol (13007) – Sassenage (38474) – Biscarosse (40046) – Pornichet (44132). Toutes ces communes avaient en 1999 une population inférieure à 10000. Elles avaient dépassé le seuil en 2006, première année de diffusion du nouveau recensement.*

Avant 2009, on avait coïncidence entre la collecte et le millésime (sauf pour quelques centaines de personnes en 2006). A partir de 2009, une partie de la population correspond à une campagne (ainsi environ 10% de la population totale la première année, sauf en 2009 où la nouvelle collecte tournante a démarré sur un faible espace). Puis ce sont deux campagnes, etc. En 2013, l'ensemble de la population appartiendra à une campagne identifiable.