

Bernard AUBRY
67100 STRASBOURG

Strasbourg, le 3 août 2010

A Monsieur Paul CHAMPSAUR,
Président de l' Autorité de la Statistique Publique
18, boulevard Adolphe Pinard
75675 PARIS cedex 14

Monsieur le Président,

Hier chargé de mission à la Direction régionale de l'Insee-Alsace, je suis aujourd'hui à la retraite. Je n'en ai pas pour autant abandonné les activités statistiques. N'ayant plus directement accès à la richesse des fichiers de l'Insee, je suis comme tout un chacun usager du site de mon ancienne institution.

J'avais pu, avant mon départ de l'Insee, mettre à jour le fichier Saphir (fichier individuel harmonisé des recensements de 1968 à 1999) en intégrant les données issues des quatre premières campagnes du nouveau recensement. Un tel fichier donnait immédiatement accès aux données les plus récentes et leur conférait un surcroît de sens grâce à une mise en perspective historique des séries couvrant plusieurs décennies. La fiabilité du fichier était bonne ; sans doute aurait-elle pu être meilleure encore si l'institution avait pu s'intéresser davantage à cet investissement et traiter elle-même les variables les plus difficiles à harmoniser.

Des séries historiques cohérentes sur la taille des familles m'ont amené à proposer une communication au colloque de la Cudep (Conférence universitaire de démographie et des populations) qui s'est tenu récemment à Strasbourg. A cette occasion, j'ai tenu à m'associer aux doléances des chercheurs qui se plaignent non seulement des conditions d'accès aux données, mais aussi des difficultés de dialogue avec l'Insee. Ainsi Madame Chantal Blayo, Présidente de la Cudep, a regretté que son courrier daté du 21 mars 2008 adressé au Directeur général de l'Insee soit resté à ce jour sans réponse.

Je travaille actuellement sur le thème de la périurbanisation dans le cadre d'une convention que l'APR (Association de prospective rhénane) a signée avec le Ministère de l'Equipement. Je me rends compte concrètement des difficultés que rencontrent les agents des territoires quand ils veulent réaliser des études sur la base des informations disponibles en ligne. Le site de l'Insee est d'un accès encore insuffisamment convivial. Il y manque une présentation synthétique et synoptique des fichiers disponibles. Il manque surtout un mode d'emploi de la statistique locale qui permettrait à chacun de se retrouver dans un cheminement qui mériterait d'être mieux balisé.

Dans les dernières années de ma carrière, j'ai effectué une analyse statistique comparant en détail les exploitations principale et complémentaire du RP1999 dont les résultats parfois contradictoires, diffusés simultanément, créent une gêne chez les utilisateurs. Si les conclusions de cette analyse avaient été prises en considération, cela aurait permis, par exemple, d'éviter la malheureuse erreur qui consiste à affecter à la commune de Mulhouse l'ensemble des salariés de l'établissement Peugeot situé sur la commune de Sausheim. A travers les contacts qui sont encore les miens – chercheurs et acteurs du terrain en particulier au sein de l'APR – il m'apparaît qu'une attention plus grande devrait être portée à l'accessibilité et à la qualité des produits diffusés par l'institution.

En ce temps où l'on ne parle que de la nécessaire modernisation des services publics, je suis d'autant plus sensible à l'amélioration de la statistique que je me suis fortement investi en ce domaine.

Sûr que vous ne sauriez vous-même être insensible à ces questions, je vous prie de croire, Monsieur le Président, à l'expression de ma respectueuse considération.

PJ

Annexe 1 - Réflexions sur l'accès aux données publiques

Annexe 2 – Le site de l'Insee au crible d'une expérience

Copies

Monsieur Jean-Pierre DUPORT, Président du Conseil National de l'Information
Statistique (CNIS)

Monsieur Jean-Philippe COTIS, Directeur général de l'INSEE - (dossier joint)

Madame Chantal BLAYO, Présidente de la Conférence Universitaire de Démographie et
des Populations (CUDEP)

**Annexes à la note du 3 août 2010 destinée au Président de l’Autorité
de la Statistique Publique**

Annexe 1 - Réflexions sur l’accès aux données publiques

Lors du colloque de la Conférence universitaire de démographie et d’étude des populations¹ j’ai présenté une étude sur la taille des familles. La communication s’achevait sur quelques réflexions relatives aux difficultés d’accès aux données publiques. Le texte qui suit reprend les principales idées exposées et tient compte des remarques de la salle.

Nombre de chercheurs se plaignent des difficultés qu’ils rencontrent dans la quête des informations statistiques dont ils ont besoin pour leurs travaux d’investigation. Cette critique n’est pas nouvelle et je l’entendais déjà au début de ma carrière à l’Insee. Elle ne concerne d’ailleurs pas seulement les chercheurs mais également tous ceux qui ont à traiter l’information statistique publique. Avant l’irruption de l’informatique, le traitement et le stockage des données limitaient évidemment les possibilités de diffusion, mais ce n’est plus le cas aujourd’hui avec la généralisation de l’internet.

Le rapport du Conseil d’analyse économique (CAE) intitulé « *Immigration, qualifications et marché du travail* » remis le 27 mai 2009 au ministre de l’Immigration illustre bien le défaut de diffusion. En effet, ce rapport interrompait en 1990 la présentation de séries historiques sur les concentrations d’immigrés et notait, pour le regretter, l’absence de données plus récentes². Et pourtant les séries existaient. Au même moment, la revue *Aspenia*, puis la revue *Commentaire* publiaient des séries historiques

¹ XVème Colloque national de Démographie – Conférence universitaire de démographie et d’étude des populations (CUDEP) – Strasbourg 25-29 mai 2010).

² *La Documentation française*. Paris, 2009 - ISBN : 978-2-11-007726-4. Page 224(renvoi)

de contenu voisin jusqu'en 2005³ ; celles-ci avaient été obtenues par simple adjonction au fichier historique des recensements, Saphir⁴, d'une sélection de données issues des quatre premières campagnes des enquêtes de recensement (2004-2007).

Le besoin d'accéder aisément aux données statistiques publiques ne concerne pas seulement les chercheurs et les acteurs publics, il concerne aussi un public plus large : par exemple des professeurs de collèges et de lycées qui veulent faire comprendre à leurs élèves leur environnement territorial, des associations dont on sait qu'à la faveur de l'allongement de la vie, elles jouent un rôle grandissant dans le domaine de la cohésion sociale. C'est ainsi que personnellement je m'investis dans une approche statistique de la périurbanisation dans le cadre de l'Association de prospective rhénane (APR) qui regroupe des acteurs issus de différents horizons. Les pistes en matière de diffusion que je suggère dans ce dossier dépassent donc les seules remarques faites devant les chercheurs de la CUDEP. Les contraintes de la CNIL ne sont pas un obstacle à une amélioration significative de l'offre statistique actuelle du site de l'Insee.

Accessibilité, clarté, cohérence, comparabilité

L'article 15 du Code européen des bonnes pratiques de la statistique préconise l'accessibilité et la clarté aux données publiques. De son côté l'article 14 porte sur la cohérence et la comparabilité des données.

On se réjouit de la mise en ligne de fichiers individuels, seuls en mesure de satisfaire les besoins des chercheurs, tout en regrettant toutefois qu'ils n'autorisent encore que peu d'exploitations au niveau infra régional⁵. Mais le problème est aussi lié à la comparabilité des résultats dans le temps. S'agissant du recensement, il est certain que, du fait de la rupture engendrée par le caractère novateur de la méthodologie du nouveau recensement, une certaine prudence s'imposait dans la mise à disposition des données séquentielles. Il reste que les expériences que nous avons faites plaident en faveur d'une bonne continuité d'un grand nombre de séries. On n'imagine pas qu'un chercheur n'ait pas le souci de comparer les chiffres les plus récents, déjà disponibles, à ceux du passé, encore peu accessibles. Plutôt que de laisser tout un chacun « bricoler » à sa façon des séries, avec tous les risques inhérents à de telles pratiques, ne vaut-il pas mieux confier cette tâche d'harmonisation aux spécialistes de l'Insee qui sont rompus aux arcanes des méthodes de redressements complexes. Au demeurant certains chercheurs regrettent que les méthodes

³ *Aspenia*, n° 44 (*I giovani stranieri di Francia - Bernard Aubry et Michèle Tribalat*), *Commentaire*, n°126, Été 2009 (*Les jeunes d'origine française – mêmes auteurs*). L'Insee à qui le projet de texte avait été proposé fin 2008 n'avait pas souhaité le publier

⁴ *Saphir : Système d'Analyse de la Population par l'Histoire des Recensements*

⁵ *Il en est ainsi des études sur la périurbanisation dans la mesure où le contenu des données individuelles mises en ligne au niveau communal est pauvre.*

de traitement d'enquêtes soient opaques, alors même que l'article 8 du même Code européen évoque la transparence des procédures statistiques. Plus généralement, au-delà du souhait de la transparence, il n'est pas inutile de rappeler la nécessité d'une pédagogie permettant une meilleure compréhension entre les producteurs et les consommateurs de chiffres.

Quelques pistes concernant l'amélioration de l'accès aux données

+ Engager une coopération entre l'Insee et les chercheurs pour faciliter la **capitalisation** des travaux d'harmonisation. L'Insee pourrait se charger d'une harmonisation minimale (type Saphir) pour les recensements et certaines enquêtes périodiques, telles celles traitant de l'emploi et du logement. Pour le reste, les universitaires pourraient être mis à contribution chacun dans leur domaine, l'Insee apportant son assistance et se chargeant de capitaliser les investissements réalisés. Ainsi par exemple, les séries historiques sur le temps de travail harmonisées par tel chercheur en démographie au prix d'un traitement lourd des différentes bases annuelles mises à sa disposition via le centre Quétélet pourraient être d'une grande utilité pour d'autres chercheurs, des économistes par exemple⁶. Tout investissement universitaire qui améliore les bases de données ou les rend plus accessibles devrait ainsi pouvoir contribuer à l'extension de l'offre statistique.

+ Rendre l'Insee plus réceptif aux remarques, aux critiques faites sur la qualité des résultats⁷, leur fiabilité, mais aussi aux suggestions portant sur le contenu des enquêtes. Les courriers restent trop souvent lettre morte. Les doléances faites à l'Institut de la statistique ne pourraient-elles pas être soumises simultanément au Conseil national de l'information statistique (CNIS) ? Cette procédure aurait un double avantage : sensibiliser les membres du Comité aux difficultés rencontrées par les utilisateurs et inciter l'Insee à donner une réponse aux questions posées.

+ Mettre en place un Comité de chercheurs et d'utilisateurs amenés à consulter les sites des autres instituts nationaux de statistique. Chacun dans son domaine apporterait une pierre à l'édifice puisqu'il serait invité à comparer méthodiquement le contenu, la qualité et les procédures de mise à disposition des données publiques par les services étrangers homologues de l'Insee. Les résultats de ces investigations seraient eux-mêmes organisés et mis en ligne. La publicité ainsi faite pourrait favoriser l'émulation entre les différents services publics en un temps où chacun sait l'importance qu'ont pris les chiffres dans la construction européenne. De son côté, la communauté des chercheurs serait partie

⁶ Je pense notamment à la thèse de Ceren Inan (IEDUB Bordeaux IV)

⁷ Les erreurs de collecte et de traitement sont inévitables. S'il n'est pas souhaitable d'en faire une grande publicité, il n'en est pas moins indispensable d'inventorier les principales anomalies afin d'éviter à ceux qui analysent les résultats de faire de grossières erreurs d'interprétation. Cela vaut en particulier pour les analyses territoriales. Mon expérience en ce domaine à l'Insee m'a appris le peu de cas que l'on fait généralement des suggestions venues de la base.

prenante à l'amélioration de la qualité du patrimoine statistique. Elle se trouverait notamment informée par elle-même de la façon dont les différents offices de statistique traitent les demandes portant sur les données dites sensibles (cryptage, interrogation à distance, etc.), puisqu'il s'agit là d'un problème délicat auquel les responsables des services de diffusion sont confrontés.

+ Les universitaires eux-mêmes de leur côté font-ils ce qu'il faut pour améliorer la situation ? Ne sont-ils pas résignés face à une institution qui leur apparaît souvent comme une forteresse ? Par ailleurs, chacun ne nourrit-il pas le secret espoir d'obtenir un sésame auprès d'un responsable compréhensif⁸ ? Pourtant on rappelle que la statistique est un bien public (« *res nullius* »), que les chiffres n'appartiennent à personne et que les responsables de l'Insee ont la charge de les traiter de façon à les rendre utiles à tous. S'il apparaît que ce n'est pas le cas les citoyens, en l'occurrence les chercheurs, pourraient, et devraient alerter les pouvoirs publics et même l'opinion. Personne n'a à gagner, et surtout pas l'Insee, à un « *affaiblissement*⁹ » de la statistique publique qui conduirait très vite à l'effacement de l'institution. Tôt ou tard la collecte et le traitement des données passeraient dans le domaine marchand et seraient conditionnés aux besoins du moment. Est-ce bien cela que nous voulons ?

⁸ La « forteresse » s'ouvre parfois comme en témoigne l'interview de Rahsaan Maxwell paru dans *Le Monde* du 2 février 2010. Cet universitaire américain a pu travailler au sein de l'Insee et accéder ainsi à des fichiers sensibles (sur l'immigration).

⁹ La formule est empruntée à Claude Thélot (préface de l'ouvrage « *Echantillonnage et méthodes d'enquêtes* », ouvrage collectif publié sous la direction de Pascal Ardilly – Dunod 2004)

**Annexes à la note du 3 août 2010 destinée au Président de l'Autorité
de la Statistique Publique**

**Annexe 2 - Le site de l'Insee au crible d'une
expérience**

Plan

A – La base des « Chiffres clés »

B – Les fichiers détails

C - A propos d'une erreur : la localisation des salariés du site Peugeot-Mulhouse

D – Pour des données communales plus facilement exploitables

Chargé de rassembler pour le compte d'une association les données statistiques existantes sur le thème de la périurbanisation en vue d'analyses devant être effectuées par des géographes et des économistes, j'ai été amené à interroger le site de l'Insee dans sa diversité, quoique d'une façon non exhaustive. L'investigation n'est pas terminée mais elle est assez avancée pour m'autoriser à faire d'ores et déjà quelques remarques.

Les premiers résultats du recensement de 2007 ont été mis en ligne le 1^{er} juillet 2010. On regrette la disparition simultanée des données de 2006. Pouvoir disposer des chiffres des années antérieures serait pourtant précieux, non pas bien entendu pour comparer les chiffres de chaque petite commune prise individuellement, mais pour suivre des tendances lourdes et comparer par exemple l'évolution, dans leur ensemble, des centres-villes et de leurs couronnes.

On souligne d'emblée que le site propose des produits clairs et bien adaptés à certains besoins, mais il s'agit alors plutôt de tableaux ou de cartes utiles pour la consultation plus que pour l'étude proprement dite. En effet les résultats valent pour des territoires fixés à l'avance (commune, aire urbaine, etc) ou à composer, mais il s'agit alors de données figées qui ne se prêtent pas à des calculs rapides d'indicateurs. Si l'on veut effectuer des analyses statistiques, il faudrait pouvoir disposer soit de tableaux communaux

agrégables, soit de fichiers détails. Il en existe en effet, mais ils ne sont guère adaptés à des usages tels que l'étude de la périurbanisation qui implique à la fois de disposer de séries historiques longues pour mettre en évidence la dynamique territoriale, et de pouvoir composer des variantes par adjonction d'une ou de plusieurs communes à des territoires de référence.

On s'intéresse ici plus spécialement à deux ensembles de bases téléchargeables. D'une part les *bases communales*, comportant par construction un enregistrement (une ligne) par commune et d'autre part les bases ou *fichiers détails*, comprenant un enregistrement par individu. Bien entendu les exploitations rendues possibles par les bases détails sont infiniment plus riches que celles offertes par les bases communales. A l'inverse les fichiers détails sont très volumineux et exigent des moyens spécifiques de traitement, à l'inverse des bases communales qui peuvent être utilisées avec les tableurs classiques.

Parmi les bases communales, on distingue les « *Bases de Chiffres clés* » qui renvoient aux tableaux de même nom et les « *Bases de Tableaux détaillés* ». Compte tenu de ce qu'une étude sur la périurbanisation ne peut se concevoir qu'en prenant en compte le mouvement, seule la « *Base des Chiffres clés* » a été utilisée puisqu'elle fournit quelques données comparables pour le recensement de 1999. Par ailleurs on a utilisé la base communale intitulée « *Données harmonisées des recensements de 1968 à 1999* ».

A – Les « Bases de Chiffres clés »

Dans la série *Chiffres clés*¹⁰, le site propose une série de 6 tableaux communaux.

+ L'impression qui domine est celle d'une grande confusion dans la présentation des variables. L'ensemble des codes et leur origine auraient dû faire l'objet d'une synthèse de façon à aider l'utilisateur à faire ses choix. On ne comprend pas pourquoi certaines variables sont proposées au RP2006 et non au RP1999. Plus généralement on regrette de ne pas percevoir de logique évidente dans le choix des variables proposées (voir § D)¹¹.

Par ailleurs, il arrive encore que l'on propose deux chiffres différents à une même date, pour une même variable. Certes les écarts s'expliquent car il s'agit de deux exploitations

¹⁰ Les « *Bases détaillées* » (TD) n'ont pas été utilisées car elles ne donnent apparemment que des chiffres que pour la dernière année (2006 ou 2007)

¹¹ L'Insee diffuse très peu de tableaux élaborés et préfère mettre en ligne quelques fichiers détails thématiques sélectionnant certaines variables, sans qu'on sache quel projet d'analyse a participé à la sélection de ces variables. C'est d'autant moins compréhensible que les recensements classiques étaient d'un abord beaucoup plus simple que les nouvelles enquêtes annuelles. L'Insee devrait avoir à cœur de mettre en ligne des tableaux détaillés tout en laissant la possibilité de confectionner des tableaux plus élaborés à partir des fichiers-détails, mis en ligne avec le maximum de variables, n'obérant pas ici des analyses auxquelles l'Insee n'a pas songé en sélectionnant des variables a priori.

(principale et complémentaire), mais ces différences sont incontestablement gênantes. Il faudrait un mode d'emploi afin que les lecteurs sachent à quoi s'en tenir.

Pourquoi ne pas présenter des séries plus longues, de 1968 à 2007 ?

Dès lors que l'on a pu sélectionner une liste de variables communes aux deux derniers recensements (après redressements dans le cas du RP99), rien ne devrait s'opposer à ce que l'on puisse remonter dans le temps, puisque les recensements antérieurs sont de même facture que celui de 1999.

Voici un exemple. En regroupant des données présentées sur le site, on peut reconstituer une série allant de 1968 à 2006 pour la population active comptée *au lieu de résidence* (actifs occupés et chômeurs) et au *lieu de travail* (emplois). Au vu du tableau, la continuité des séries est tout à fait satisfaisante. Le tableau aurait pu être mis à jour avec les données relatives à 2007.

T.1 - Une série cohérente pour l'Alsace (1968-2006) pour la population active et l'emploi

Catégorie socio-prof.		Données harmonisées 1968-1999					Nouvelle série	
		RP68	RP75	RP82	RP90	RP99	RP99	RP06
Actifs occupés		561008	596310	642472	686415	750115	748099	796239
Expl.agr.	ao1	54164	24895	22108	14308	8723	8515	7406
Art. Com. CE	ao2	41644	33960	36564	35045	37650	37225	36831
Cad. Sup.	ao3	28524	40560	49120	70477	83934	83596	108830
Prof. Int..	ao4	70296	92185	110116	134277	169699	169351	195089
Employés	ao5	129196	148525	162848	175328	200764	200404	214705
ouvriers	ao6	237184	256185	261716	256980	249345	249008	233379
chômeurs		5708	14335	38664	51876	65062	65056	82982
Expl.agr.	ch1	0	0	48	32	157	157	33
Art. Com. CE	ch2	0	5	276	1172	2354	2354	2156
Cad. Sup.	ch3	128	460	728	1648	2714	2714	3303
Prof. Int..	ch4	480	1175	3132	4472	8581	8581	12546
Employés	ch5	1600	4460	16040	21273	22935	22935	29080
ouvriers	ch6	3500	8235	18440	23279	28321	28315	35865
emplois		547836	568995	608816	635819	687767	687767	744441
Expl.agr.	e1	54180	24890	22112	14328	8694	8694	7655
Art. Com. CE	e2	41632	33820	36400	34805	36550	36550	36314
Cad. Sup.	e3	28440	39750	47600	67341	78957	78957	101307
Prof. Int..	e4	69608	89845	106828	127489	158616	158616	185555
Employés	e5	127240	143575	157660	166937	189640	189640	205482

ouvriers	e6	226736	237115	238216	224919	215310		215310	208127
Agriculture	eagri	61384	28810	26600	18512	15050		15050	15205
Industrie	eind	243760	257320	248392	234104	211550		211550	205320
Tertiaire	eter	242692	282865	333824	383203	461167		461167	523916

On donne le total de l'Alsace, mais les données sont disponibles par commune.

Il est très utile de pouvoir disposer de séries cohérentes sur la population active et l'emploi sur une période de près de 40 ans (1968-2007). Nous aurions même pu obtenir une série plus longue (45 ans) si l'année 1962¹² avait été retenue !

Aussi comprend-on mal que l'on n'ait pas chaîné l'ensemble des recensements sur toute la période 1968-2007, alors même que tous les éléments qui auraient permis de faire l'opération sont en ligne, *mais à des endroits différents*. Il apparaît bien quelques différences au niveau du RP99 au titre de la population active, mais celles-ci sont loin d'être rédhibitoires¹³. Pourquoi aurait-on des scrupules à diffuser des séries quand les discontinuités sont faibles (ainsi pour la population active au RP99), alors que l'on met en ligne par ailleurs des séries manifestement incohérentes (cf l'emploi industriel de Mulhouse)?

A propos de la cohérence des définitions territoriales.

Si les variables structurelles ont été harmonisées (il est toujours possible de se référer aux données originales si nécessaire), en revanche rien n'a été fait pour la mise en cohérence des variables géographiques dès lors que celles-ci ont subi de nombreux changements : les fusions de communes, mais surtout les défusions, nombreuses dans certains départements, compliquent le suivi des séries. Ainsi par exemple, quand deux communes qui ont fusionné en 1972 ont « divorcé » en 1978, les deux éléments du couple n'ont pas été identifiés au RP75, de sorte que les séries brutes sont incohérentes¹⁴. Pourtant recréer des données fictives pour les communes momentanément disparues était possible moyennant un investissement informatique assez modeste.

On ne dira jamais assez la gêne occasionnée par ces ruptures, ni le gaspillage en moyens qu'elles créent par le temps perdu à corriger les séries et ce avec plus ou moins de bonheur.

¹² L'année 1962 figure en effet dans le fichier Saphir

¹³ Il semble bien que la rupture dans la série soit la combinaison de deux éléments agissant en sens inverse : la nouvelle série n'inclut pas les actifs de 65 ans et plus, mais elle comprend les chômeurs n'ayant jamais travaillé.

¹⁴ Nous avons proposé une méthodologie relativement simple (au niveau du ménage), malheureusement les quelques moyens informatiques demandés n'ont pu être débloqués.

Il y a d'autres modifications territoriales au niveau communal, mais elles correspondent soit à des échanges de parcelles sans grande conséquence statistique, soit à des cas singuliers (créations de villes nouvelles) devant être traités au cas par cas avec le concours des statisticiens régionaux.

B – Fichiers détails

Il existe six fichiers-détails qui représentent un bon capital de données. Quatre de ces fichiers mentionnent la commune et sont donc précieux pour l'étude de la périurbanisation, tout particulièrement le fichier logement et les fichiers migratoires.

+ Une description des codes peu commode

Associée à chaque fichier-détail, on trouve une liste des variables et des modalités correspondantes. Une présentation synthétique pour l'ensemble des fichiers aurait été appréciée, avec renvoi aux seules listes de modalités nécessaires. En effet, on est tenté de railler la répétition inutile de l'ensemble de modalités aussi évidentes que la liste des nombres de 000 à 120 pour caractériser l'âge (autrement dit, une ligne par âge détaillé, et ce à chaque fois fois !). Au bilan le document perd grandement en lisibilité et l'imprimer est onéreux (l'opération est peu écologique par le gaspillage de papier et d'encre qu'elle occasionne). On aurait pu très probablement réunir en une quinzaine de pages ce qui est proposé en plus de 200 (on dénombre déjà plus de 50 pages pour le seul fichier régional !).

Quelques suggestions pour rendre plus facile l'accès aux données.

- 1) + ***Un principe : chaque fois que l'information existe, il faut pouvoir disposer d'une série historique****. Parmi les avantages des séries historiques : évidemment la mise en évidence des transformations spatiales et structurelles, mais aussi *le repérage immédiat des anomalies*

- 2) + ***Fichiers communaux***** (Excel)
 - . un fichier élémentaire national (une base de 36000 lignes contenant environ 200 variables soigneusement sélectionnées, les mêmes pour chaque RP)
 - . un fichier d'essai pour faciliter la prise en main, et des fichiers par région (afin d'alléger les opérations de téléchargement)

- 3) + ***Fichiers-détails*****
 - . un fichier *unique* national, de référence, de 1962 à 2006 pour quelques variables harmonisées : date RP, région de résidence, sexe, âge révolu, formation, statut d'activité, emploi (par CS regroupée), origine (regroupée), département - pays de naissance, résidence antérieure, de travail, vie en couple, nombre d'enfants, statut du logement...
 - des fichiers régionaux incluant les personnes n'y résidant pas mais qui ont ou qui ont eu un lien avec elle.
 - des fichiers thématiques (sur les familles, les immigrés, le logement, etc).
 - un fichier communal (cf fichier logement incluant quelques données individuelles)

- 4) + ***Harmonisation sur longue période des codes communaux***

- 5) + ***Documentation plus élaborée*** : éviter les répétitions inutiles – harmoniser les libellés –
 - mieux adapter la documentation au contenu de chaque tableau –
 - + **accompagnement de l'utilisateur**, par un mode d'emploi, en forme de « cours » élémentaire de statistique adapté à l'analyse des données figurant sur le site.

- 6) + ***Renvoi vers des listes d'erreurs***, et proposition d'une méthode à suivre pour s'en affranchir (cf note intitulée : « *Sur l'affaiblissement de la statistique – A propos de la qualité du recensement de 1999 – 3 janvier 2008* »).

- 7) + ***Expression graphique*** des résultats adaptée aux domaines. Ainsi, compte tenu du débat actuel autour des générations, on pourrait associer des **pyramides d'âge** aux tableaux, à l'instar de ce que propose Statistique Canada.

**Lors des recensements classiques et avant le développement d'internet, l'Insee publiait de nombreux documents, notamment des volumes thématiques, dans lesquels on trouvait un grand nombre de tableaux détaillés très utiles et immédiatement accessibles. Avec internet le volume n'étant plus un problème, on*

aurait dû pouvoir disposer au moins de tous les tableaux des « fascicules verts » bien connus et très appréciés par les utilisateurs « traditionnels » des recensements.
** Ou plutôt un fichier par recensement, l'essentiel étant que l'utilisateur n'ait aucune difficulté à fusionner les fichiers successifs.

C - A propos d'une erreur : la localisation des salariés du site de Peugeot-Mulhouse

Les anomalies de collecte et de traitement sont inévitables. S'il n'est pas souhaitable d'en faire une grande publicité, il n'en est pas moins indispensable d'inventorier les principales anomalies afin d'éviter à ceux qui analysent les résultats de faire de grossières erreurs d'interprétation. Cela vaut surtout pour les analyses territoriales. Une note décrivant une méthode permettant de repérer et de lister certaines catégories d'erreurs observées au recensement de 1999 avait permis de faire une liste exhaustive des anomalies. Si l'on avait accepté de regarder sérieusement cette note, on aurait sans doute pu éviter de renouveler certaines erreurs. Ainsi, qui peut croire, au vu des fiches communales actuellement en ligne (tableau EMPT8), que la commune de Mulhouse ait vu son emploi *industriel* passer en 7 ans de 6587 à 11795 postes de travail ? De 1999 à 2006, l'augmentation aurait été de 79%. De telles anomalies, si elles ne sont pas accompagnées d'une note d'explication, ternissent l'image de la statistique publique.

Le quidam ne sait pas quelle est la fréquence de ce type d'erreurs : pour être tombé aléatoirement sur un événement malheureux il peut, à tort, être porté à jeter un discrédit plus général sur l'ensemble.

T.2 – fac-similés du tableau EMP T8 de la série Chiffes clés pour la commune de Mulhouse

T. 2a- version antérieure (avant le 1.7.2010)

	2006				1999	
	Nombre	%	dont femmes en %	dont salariés en %	Nombre	%
Ensemble	61 739	100,0	49,1	93,6	53 567	100,0
Agriculture	204	0,3	29,0	80,8	169	0,3
Industrie	11 795	19,1	22,7	97,0	6 587	12,3
Construction	3 165	5,1	10,2	88,0	2 876	5,4
Tertiaire	46 574	75,4	58,5	93,1	43 935	82,0
dont						
- commerce	6 996	11,3	53,9	88,3	7 602	14,2
- services aux entreprises	8 105	13,1	43,0	90,7	7 107	13,3
- services aux particuliers	3 866	6,3	63,2	86,0	3 870	7,2

Sources : Insee, RP1999 et RP2006 exploitations complémentaires lieu de travail.

T2.b - version actuelle (après le 1.7.2010)

EMP T8 - Emplois selon le secteur d'activité en 2007				
	Nombre	%	dont femmes en %	dont salariés en %
Ensemble	61 035	100,0	50,1	93,6
Agriculture	107	0,2	49,8	77,1
Industrie	10 339	16,9	22,2	96,8
Construction	3 298	5,4	11,7	87,5
Commerce, transports, services divers	25 271	41,4	49,9	90,6
Administration publique, enseignement, santé, action sociale	22 019	36,1	69,1	96,6

Source : Insee, RP2007 exploitation complémentaire lieu de travail.

Par rapport au même tableau mis en ligne précédemment, on remarque la disparition des données relatives au RP99. Pourtant les données pour 1999 sont encore proposées, mais elles se trouvent dans le tableau EMP T5. Cette fois le chiffre de 1999 n'est plus de 53567 (exploitation principale), mais de 62780, soit environ 9000 emplois de plus (l'usine Peugeot a donc été intégrée) !

T.3 – fac-similé du tableau EMP T5 de la série Chiffes clés pour la commune de Mulhouse : version actuelle, postérieure au 1.7.2010

EMP T5 - Emploi et activité		
	2007	1999
Nombre d'emplois dans la zone	61 820	62 780
Actifs ayant un emploi résidant dans la zone	39 354	42 034
Indicateur de concentration d'emploi	157,1	149,4
Taux d'activité parmi les 15 ans ou plus en %	56,5	57,4

L'indicateur de concentration d'emploi est égal au nombre d'emplois dans la zone pour 100 actifs ayant un emploi résidant dans la zone.

Sources : Insee, RP1999 et RP2007 exploitations principales lieu de résidence et lieu de travail.

L'origine de l'erreur est simple. Les emplois du site de Peugeot situé à Sausheim, commune limitrophe, ont été localisés à tort à Mulhouse alors qu'ils l'avaient été

correctement en 1999. D'où une incohérence fâcheuse¹⁵. Malheureusement, tout se complique comme le montre le tableau suivant.

T. 4 – Confrontation des données sur l'emploi : commune de Sausheim

Position Sur le site	Tableau	Exploitation	Date d'observation sur le site*	1999	2006	2007
Données locales (CLAP)	RSTAT	Principale	Après 1 ^{er} juillet	4243	-	7469
	N16	-	id	-	-	15034
Recensement	EMP T5	Principale	Avant	4243	6688	
			Après	4243	-	7469
	EMP T8	Complémentaire	Avant	13481	2905	
			après	?		7682

- par rapport au 1^{er} juillet 2010

On reunit dans ce tableau les données mises en ligne **avant et après** le 1^{er} juillet 2010. On distingue d'une part la *position sur le site* de l'Insee (soit dans les statistiques des recensements, soit dans les statistiques locales) et d'autre part *la source proprement dite* (enquêtes de recensement ou fichier Connaissance locale de l'appareil productif (CLAP)). La position dans le site des données RP et CLAP est éloignée et il n'y a pas de lien immédiat et évident (pour un non initié) pour inciter le lecteur à passer d'un thème à l'autre. A souligner qu'il existe plusieurs fichiers CLAP. L'un d'entre eux (RSTAT résumé communal) rapproche quelques données issues de CLAP avec d'autres chiffres communaux émanant d'autres sources (RP, Sirene, revenus fiscaux, etc.).

Comprenne qui pourra !

Le tableau met en évidence des **données contradictoires**...

¹⁵ Les agents de la DR de l'Insee-Alsace connaissent bien cette anomalie. Elle avait été signalée à plusieurs reprises aux responsables nationaux du traitement des recensements. Mais dans le passé les choses étaient plus maîtrisables, dans la mesure où les recensements successifs se faisaient en deux temps. Une première exploitation, exhaustive, appelée aujourd'hui principale correspondait à une codification rapide. Une personne mentionnant Peugeot-Mulhouse comme établissement de travail aura donc automatiquement cette commune pour lieu de travail. En revanche lors de l'exploitation complémentaire, le nom de l'établissement renvoie aux informations contenues dans le fichier Sirene et ce sont ces données qui alimentent le fichier informatique du recensement pour ce qui concerne le lieu de travail et le secteur d'activité.

En 1999, selon les exploitations du RP : EMP T5 donne 4243 emplois, en accord avec tableau CLAP résumé, mais en contradiction avec EMP T8 qui donne 13481 emplois (incluant donc l'automobile, 9500 emplois environ). En 2007, selon CLAP : emplois totaux : 7469 d'après le tableau résumé (en fait le chiffre du RP complémentaire), contre 15034 quand on totalise les 16 postes du tableau NES16.

On se demande pourquoi les tableaux RSTAT (CLAP) ne donnent pas en 1999 les données du complémentaire, à priori plus proches des données administratives (puisque confrontées à Sirene).

... et des évolutions incohérentes :

Augmentation de 4243 à 7469 emplois totaux de 1999 à 2007 d'après tableaux CLAP RSTAT et diminution de 13481 à 7682. Par ailleurs la variation est de 2905 à 7682 pour le tableau EMPT8 et de 6688 à 7469 pour le tableau T5.

Pourquoi une telle erreur est-elle gênante ?

On pourrait penser que les conséquences de telles anomalies ne sont pas trop graves et qu'il suffirait pour l'utilisateur de fusionner les deux communes de Mulhouse et de Sausheim pour avoir une description satisfaisante de la réalité. Cette opération n'aurait de sens que si les deux communes appartenaient l'une et l'autre aux mêmes territoires d'études éventuelles, ce qui n'est jamais le cas. Les conséquences induites sont nombreuses notamment pour les travaux nationaux ou suprarégionaux.

+ On rappelle que cette anomalie ne fait l'objet d'aucune mention susceptible d'alerter l'utilisateur. On peut comprendre qu'une telle erreur puisse être difficile à corriger une fois le processus de diffusion engagé. Pourquoi ne procède-t-on pas à davantage de contrôles de vraisemblance avant la diffusion des résultats.

Et s'il apparaît des erreurs graves, il faut impérativement en dresser une liste et mettre clairement celles-ci à disposition des utilisateurs. C'est là un égard minimal que l'institution se doit de montrer vis-à-vis de ceux qui font confiance à la statistique publique.

+ Il reste que même clairement identifiées, les erreurs ne sont pas faciles à gérer pour le statisticien en herbe qui s'y trouve confronté. Certes celui-ci saura corriger l'anomalie s'il fait un travail monographique. Mais s'il travaille à Brest ou au Havre, loin de Mulhouse pour reprendre l'exemple précité, et s'il compare sa propre commune à celles

d'un ensemble de villes similaires par la taille, il ne verra pas le problème et ses conclusions seront erronées¹⁶.

+ Songeons aussi à l'enseignant d'un collège ou d'un lycée de Mulhouse qui chercherait à intéresser ses élèves au développement économique de la cité. Que dire des conséquences de cette première rencontre avec la statistique publique ?

+ On dispose depuis quelques mois d'une information localisée sur l'emploi (CLAP). Il y aurait peut-être une piste intéressante qui consisterait à intégrer dans les chiffres clés issus des recensements, *à côté des données de l'emploi*, une information complémentaire telle qu'elle est donnée par les sources administratives. Il faudrait alors veiller à accompagner le lecteur en lui montrant que des écarts entre sources ne sont pas alarmants s'ils restent modérés et que, s'ils sont importants, cela signifie probablement que les données doivent être analysées au plus près. On est là dans l'indispensable pédagogie : ***comment faire comprendre l'existence possible d'erreurs et proposer une marche à suivre quand elles se présentent.***

Une réflexion, encore

Comme toute opération très lourde, un recensement, surtout quand il s'opère selon des procédures innovantes, est un échafaudage de procédures complexes. Il suffit que l'une d'elles soit défectueuse, qu'un seul chaînon soit défectueux pour que l'ensemble soit fragilisé. En l'occurrence et sous réserve d'inventaire, on a l'impression que la conception, la préparation et la collecte sur le terrain (du moins si j'en juge d'après quelques témoignages de collègues alsaciens) soient plutôt réussies. En revanche on est enclin à penser que les traitements statistiques ne donnent pas tout ce que l'on aurait été en droit d'attendre. N'avait-on pas insisté sur la promesse que la nouvelle méthode de recensement ne changerait en rien l'appréhension des migrations ? Les agents du territoire rencontrés me disent leur déception et je serais enclin à leur donner raison.

La collecte des données peut connaître des erreurs ici ou là, malgré la qualité du travail des agents recenseurs. ***Mais les erreurs systématiques dans le traitement de ces données s'avèrent beaucoup plus graves que des erreurs localisées et une attention redoublée doit être portée pour les éviter.***

¹⁶ On souhaite bon courage aux chargés d'études de la région mulhousienne, si les élus leurs demandent un rapport clair et concis de la situation et de l'évolution de l'emploi de leur territoire. On fera cependant remarquer que l'inflation de chiffres permet de rédiger un texte triomphant ou au contraire alarmiste, selon les attentes du destinataire de la note !

D – Pour des données communales plus facilement exploitables

Au-delà des chiffres de référence (monographies de territoires), l'utilisateur doit pouvoir accéder à des bases simples qui seules permettent d'effectuer les regroupements spécifiques nécessaires à l'étude des territoires.

Le site propose deux séries de fichiers communaux téléchargeables : six fichiers pour les chiffres clés (CC) et une quarantaine de fichiers détaillés (TD). Cela constitue certes une masse de données imposantes, mais leur accès ne peut se faire qu'au prix de manipulations, avec, au bout du compte, une information peu claire et superfétatoire.

A titre d'exemple on présente sur un même tableau l'ensemble des variables issues des deux fichiers ACT (activité) et EMP (emploi) de la série CC (bases détails) avec indication des effectifs au RP1999 et au RP2006/07. Les libellés sont amputés des quatre premiers caractères et classés par ordre alphabétique et par exploitation (exp : P pour principale et C pour complémentaire).

T. 5 - Les données sur l'activité et l'emploi (bases CC – Alsace)

variable	exp	1999	2006	2007	variable	exp	1999	2006	2007
ACT1524	P		112741	112166	CONSTLT_FSAL	C		4533	5167
ACT1564	C	821031	885902	891965	CONSTLT_NSAL	C		7263	7830
ACT1564	P	822374	887206	892248	CONSTLT_SAL	C		40551	43182
ACT1564_CS1	C	8672	7438	7269	CTSLT_F	C			154378
ACT1564_CS2	C	39579	38987	39387	CTSLT_FNSAL	C			11107
ACT1564_CS3	C	86310	112132	115088	CTSLT_FSAL	C			143271
ACT1564_CS4	C	177932	207634	210706	CTSLT_NSAL	C			32268
ACT1564_CS5	C	223339	243785	245733	CTSLT_SAL	C			285504
ACT1564_CS6	C	277323	269244	266589	EMPLT	C	687767	744441	749008
ACT15P	P	824286	891624	896919	EMPLT	P	685360	745143	749405
ACT2554	P		698777	699800	EMPLT_AGRI	C		15205	13230
ACT5564	P		75688	80283	EMPLT_APESAS	C			214500
ACTOCC	P	751123	801258	804754	EMPLT_COMM	C		105367	
ACTOCC1524	P		91049	89429	EMPLT_CONST	C		47814	51011
ACTOCC1564	C	748099	796239	800259	EMPLT_CS1	C	8694	7655	7438
ACTOCC1564	P	749155	796897	800170	EMPLT_CS2	C	36550	36314	36543
ACTOCC1564_CS1	C	8515	7406	7236	EMPLT_CS3	C	78957	101307	104102
ACTOCC1564_CS2	C	37225	36831	37096	EMPLT_CS4	C	158616	185555	188814
ACTOCC1564_CS3	C	83596	108830	111918	EMPLT_CS5	C	189640	205482	207170
ACTOCC1564_CS4	C	169351	195089	198133	EMPLT_CS6	C	215310	208127	204941
ACTOCC1564_CS5	C	200404	214705	216101	EMPLT_CTS	C			317771
ACTOCC1564_CS6	C	249008	233379	229774	EMPLT_F	C		348344	352745
ACTOCC15P	P	751067	801041	804549	EMPLT_FNSAL	P	19628	21862	22318
ACTOCC15P_ILT1	P	257779	256464	255337	EMPLT_FSAL	P	286960	326250	330285
ACTOCC15P_ILT2	P	398044	450193	454671	EMPLT_INDUS	C		157506	152495
ACTOCC15P_ILT2P	P	493288	544577	549212	EMPLT_NSAL	P	61841	67644	68946
ACTOCC15P_ILT3	P	11594	15274	15647	EMPLT_NSALTP	P	5192	6591	6889
ACTOCC15P_ILT4	P	11636	13892	14318	EMPLT_SAL	C		682727	687008
ACTOCC15P_ILT5	P	72014	65218	64575	EMPLT_SAL	P	623519	677498	680460
ACTOCC15P_TP	P		133369	135812	EMPLT_SALTP	P	108740	118871	120927
variable	exp	r1999b	r2006	r2007	variable	exp	r1999b	r2006	r2007
ACTOCC2554	P		636287	636860	EMPLT_SENTREP	C		86192	

ACTOCC5564	P		69562	73881	EMPLT_SPARTIC	C		52747	
AGRILT_F	C		5308	5053	EMPLT_TERTI	C		523916	
AGRILT_FNSAL	C		2300	2168	ETUD1564	P	127580	122726	122611
AGRILT_FSAL	C		3008	2885	F1524	P		119482	119045
AGRILT_NSAL	C		7729	7113	F1564	P		604347	607936
AGRILT_SAL	C		7476	6117	F2554	P		389615	389328
AINACT1564	P	126321	111095	109289	F5564	P		95250	99563
APESASLT_F	C			142969	FACT1524	P		51157	51008
APESASLT_FNSAL	C			4473	FACT1564	P	367207	414187	418512
APESASLT_FSAL	C			138496	FACT2554	P		329078	331010
APESASLT_NSAL	C			9467	FACT5564	P		33952	36494
APESASLT_SAL	C			205033	FACTOCC1524	P		40478	39945
CHOM1564	P	71120	90308	92078	FACTOCC1564	P		365841	369836
COMMLT_F	C		52385		FACTOCC15P	P		367681	371789
COMMLT_SAL	C		93924		FACTOCC2554	P		294239	296429
CONSTLT_F	C		4811	5475	FACTOCC5564	P		31124	33462
CONSTLT_FNSAL	C		278	307	FCHOM1524	P		10679	11064
FCHOM1564	P	38312	48346	48677	HSAL15P_APPR	P		11035	11352
FCHOM2554	P		34839	34582	HSAL15P_CDD	P		22853	23151
FCHOM5564	P		2828	3031	HSAL15P_CDI	P		337820	336641
FNSAL15P	P		22347	22806	HSAL15P_EMPAID	P		2648	2292
FNSAL15P_AIDFAM	P		951	901	HSAL15P_INTERIM	P		12020	11457
FNSAL15P_EMPLOY	P		9692	9815	HSAL15P_TP	P	14722	19186	19978
FNSAL15P_INDEP	P		11703	12090	HSAL2554	P		304477	302528
FSAL1524	P		39651	39091	HSAL2554_TP	P		9581	9815
FSAL1524_TP	P		11044	11295	HSAL5564	P		31087	32591
FSAL1564	P		343927	347514	HSAL5564_TP	P		2484	2599
FSAL1564_TP	P		106788	108115	INACT1564	P	337262	323198	324967
FSAL15P	P	309362	345334	348983	INDUSLT_F	C		47093	44870
FSAL15P_APPR	P		6534	6923	INDUSLT_FNSAL	C		1192	967
FSAL15P_CDD	P		31529	32641	INDUSLT_FSAL	C		45901	43903
FSAL15P_CDI	P		297451	300403	INDUSLT_NSAL	C		6001	5322
FSAL15P_EMPAID	P		4422	3821	INDUSLT_SAL	C		151505	147173
FSAL15P_INTERIM	P		5398	5195	NSAL15P	P		69331	70673
FSAL15P_TP	P	101451	107376	108740	NSAL15P_TP	P		6807	7094
FSAL2554	P		276420	278357	POP1524	P		239215	238843
FSAL2554_TP	P		84783	85245	POP1564	P	1159636	1210403	1217215
FSAL5564	P		27857	30066	POP15P	P	1398616	1483157	1495065
FSAL5564_TP	P		10961	11575	POP2554	P		779331	778024
H1524	P		119733	119798	POP5564	P		191857	200349
H1564	P		606057	609278	RETR1564	P	83361	89377	93067
H2554	P		389716	388696	SAL15P	P		731710	733876
H5564	P		96607	100785	SAL15P_TP	P		126562	128719
HACT1524	P		61584	61157	SENTREPLT_F	C		37408	
HACT1564	P	455167	473019	473736	SENTREPLT_SAL	C		78422	
HACT2554	P		369699	368790	SPARTICLT_F	C		33141	
HACT5564	P		41736	43789	SPARTICLT_SAL	C		44535	
HACTOCC1524	P		50571	49484	TERTILT_F	C		291132	
HACTOCC1564	P		431056	430334	TERTILT_FNSAL	C		15117	
HACTOCC15P	P		433360	432760	TERTILT_FSAL	C		276015	
HACTOCC2554	P		342047	340432	TERTILT_NSAL	C		40720	
HACTOCC5564	P		38438	40419	TERTILT_SAL	C		483195	
HCHOM1524	P		11013	11673					
HCHOM1564	P	32808	41963	43402					
HCHOM2554	P		27652	28358					
variable	exp	r1999b	r2006	r2007	variable	exp	r1999b	r2006	r2007
HCHOM5564	P		3298	3370					
HNSAL15P	P		46984	47867					
HNSAL15P_AIDFAM	P		406	369					
HNSAL15P_EMPLOY	P		29447	29960					
HNSAL15P_INDEP	P		17132	17538					

HSAL1524	P		49301	48199				
HSAL1524_TP	P		6774	7099				
HSAL1564	P		384866	383319				
HSAL1564_TP	P		18838	19513				
HSAL15P	P	377636	386375	384892				

Attention : les variables du tableau sont présentées par ordre alphabétique, mais dans une composition par blocs

La diversité des variables et des chiffres rend difficile la lecture

On a extrait (tableau T.6) les seules variables concernant l'ensemble des catégories d'actifs (occupés ou non) et des emplois. Les résultats de l'Alsace ont été comparés avec ceux obtenus par l'exploitation du fichier détail régional.

T.6 – Trop de chiffres pour l'Alsace ?

source		Expl.	âge	actifs	Actifs occupés	chômeurs	emplois
Base -détails	Ensemble	P	15-64	885902	796239	-	-
	ensemble	C	15-64	887206	796897	90308	-
	ensemble	P	15+ (?)	-	801258	-	745143
	ensemble	P	15 +	891624	801041	-	744441
Fichiers détails	ensemble		15-64	884662	795088	89573	740216
	ensemble		15 et +	888949	799141	89808	744199

Examinons la colonne des actifs occupés. On dispose de six chiffres (!).

On comprend à la rigueur que le souci de l'exactitude – est-ce bien le terme adéquat ? - apporte quelque chose à certains spécialistes de l'Insee et à certains chercheurs, mais quand il s'agit du grand public et d'acteurs de la vie économique et sociale (les maires et leur entourage par exemple) une telle diversité de chiffres – *soi-disant des chiffres clés* - peut surprendre. De surcroît les chiffres étant peu différents les uns des autres (ils sont tous compris dans une fourchette inférieure à 1%), leur multiplicité n'apporte rien sinon des soucis à l'utilisateur.

Autre remarque. A propos de la catégorie socioprofessionnelle : pour l'emploi la répartition se fait sur l'ensemble, pour les actifs, elle se fait sur les seules personnes de 15 à 64 ans.

Tout cela n'est grave en soi, mais contribue à la confusion. Surtout, **on ne voit pas la logique sous-tendue par ces choix**. Ainsi pourquoi ne dispose-t-on pas des actifs totaux

par sexe, tous âges, en 1999 et en 2006? Une telle information aurait permis de disposer d'une série historique de qualité meilleure (cf T.1)

Par ailleurs on s'étonne que les chiffres du RP 1999 ne soient pas proposés systématiquement en regard de ceux de 2006. Or les chiffres seuls n'ont qu'un intérêt limité. En revanche ils prennent tout leur sens ils sont mis en perspective historique.

Quelques suggestions :

+ Est-il vraiment indispensable de poursuivre la diffusion simultanée de données issues des deux exploitations ? Des principes que l'on croyait invariants comme la décomposition de la population d'une commune en un très grand nombre de catégories, apparaissent aujourd'hui comme inutiles puisqu'ils ont été abandonnés. Dans un ordre d'idées un peu différent, on a franchi un saut « culturel » en faisant accepter que les recensements classiques pouvaient être remplacés par des enquêtes annuelles par sondage et que les résultats informatisés pouvaient être diffusés avec en décimales.

Dès lors que l'on accepte de s'affranchir de la tradition, est-il vraiment illusoire d'imaginer que l'on puisse libérer enfin les utilisateurs des lourdeurs qu'occasionnent la diffusion de chiffres dédoublés ?

+ On a montré que l'on pouvait disposer de séries cohérentes depuis 1968 jusqu'en 2006 pour l'emploi par CS (exploitations complémentaires). Pourquoi ne pas proposer cette série ainsi que quelques autres séries cohérentes ou pouvant être rendues telles (âge quinquennal, taille des familles, composition des logements, etc.) ? Cela permettrait à l'utilisateur d'éviter de faire une gymnastique (parfois acrobatique) qui est de toute façon chronophage pour des utilisateurs souvent pressés par bien d'autres occupations. Cette remarque peut paraître de bon sens, d'autant plus que de présenter des séries cohérentes de 1962 à 2007 sur des variables structurelles élémentaires valoriserait sensiblement les chiffres les plus récents.

+ A ces données censitaires, fondamentales, on adjoindrait quelques données plus récentes, complémentaires, issues par exemple de CLAP (emploi dans une nomenclature en 15 postes) pour les dernières années disponibles.

En conclusion :

Ce que le « client » doit pouvoir comprendre :

- + *Que l'on diffuse des chiffres provisoires, remplacés plus tard par des chiffres semi-définitifs, puis enfin définitifs.*
- + *Qu'il existe des ruptures à des moments donnés (pour cause de changements de nomenclature), pour autant que l'on dispose des moyens de raccordement.*
- + *Que des chiffres d'une même variable à une même date ne soient pas les mêmes, dès lors qu'ils proviennent de sources différentes.*

Ce qu'il a des difficultés à comprendre :

- + *Que l'on diffuse un jour donné, pour une même variable, à une même date, pour un même territoire, des chiffres différents alors même qu'ils proviennent d'une même source, du seul fait qu'ils proviennent d'exploitations différentes, dès lors que l'on dispose de moyens de traitement bien plus performants que dans le passé (cf ci-dessus).*
- + *Que l'on diffuse des chiffres pour une année n , puis pour une année $n+1$, tout en laissant entendre que la variation d'une année à l'autre est entachée de biais ¹⁷.*

¹⁷ Ainsi des données issues des chiffres-clés. S'agissant des communes de moins de 10000 habitants, la variation de la population totale est réputée significative, puisqu'issue d'estimations s'appuyant sur des données administratives. En revanche, les structures sont fixées une fois pour toutes pendant 5 ans, ce sont celles de l'année d'observation. N'est-il pas possible d'extrapoler ces structures (en s'appuyant sur celles de 1999 et de celle de l'année d'observation) pour les estimer pour l'année millésimée ?