

LES OUTILS DE LA QUALITÉ D'UNE ENQUÊTE

Conception de questionnaires et d'outils de saisie

F. Tsai

1. La qualité d'une enquête

La qualité d'une enquête peut être comparée à la qualité d'une chaîne de froid, toute rupture qualitative dans la chaîne détériore la qualité finale du produit. Il convient donc de prévoir les différentes phases, d'organiser leur enchaînement, de concevoir et surveiller chacune des phases de l'enquête sous l'angle qualité, pour l'optimiser lors de la conception, pour minimiser les dégradations lors des traitements. La qualité n'est pas le seul paramètre à prendre en compte lors de la préparation d'une enquête, il y a l'organisation, la formation des différents intervenants, mais la plupart des travaux participent de la recherche de qualité. Si, par exemple, l'accent est mis sur une organisation permettant l'obtention rapide des résultats, au-delà de l'aspect purement médiatique de la chose, il est clair que la qualité du produit "enquête" est, pour beaucoup, fonction de sa fraîcheur, car les chiffres se périment vite. Mais ce souci de rapidité peut impliquer certaines "impasses" au niveau de la qualité, la qualité a un coût, la rapidité aussi. Il faut trouver le bon compromis en fonction des objectifs visés.

D'après les statisticiens suédois, la qualité d'une enquête dépend de trois facteurs :

- la **précision**, fonction de la base de sondage (qualité, fraîcheur), fonction du sondage proprement dit, fonction du taux de réponse et de l'exactitude des réponses (sincérité, bonne volonté, compréhension des questions, accessibilité plus ou moins grande de l'information par recours à la mémoire ou à des documents), fonction de l'enquêteur (formation, compréhension des questions, compréhension et notation des réponses), fonction aussi de la qualité des traitements que l'on fait subir ou non à l'information collectée ;
- les **délais** entre observation et publication ;
- l'**adéquation** au problème posé (conformité à la demande et non pas conformité aux seules idées du statisticien).

Mon propos sera centré sur les **enquêtes par sondage, auprès des ménages**, qui sont réalisées auprès d'un échantillon représentatif de la population étudiée. Mais bien des considérations pourraient s'appliquer à des "recensements" exhaustifs, ou à des enquêtes auprès d'entreprises ou d'autres entités ou bien encore à des enquêtes de type administratif ou à l'exploitation de documents issus de sources administratives.

J'ai distingué les phases suivantes dans le traitement d'une enquête :

- l'échantillonnage, *puisque sondage il y a* ;
- l'élaboration du questionnaire ;
- la collecte de l'information ;
- la saisie et ses différents modes ;
- les traitements manuels et l'assistance informatique ;
- les traitements informatiques.

Je développerai particulièrement tout ce qui implique une interaction entre le statisticien et l'appareil d'enquête, outils et acteurs de la collecte, de la saisie, de la codification, de la vérification et de la correction.

Je ne développerai pas les parties **échantillonnage, et traitements informatiques** qui sont hors de ce champ. Il convient simplement d'indiquer que la qualité générale d'une enquête commence avec la qualité de la base de sondage (qui doit être entretenue sous peine de mal vieillir) et la qualité du sondage qui doit être ajusté au plus près selon les informations recherchées dans l'enquête ; ajusté au plus près également pour minimiser les coûts de collecte.

Pour ce qui concerne les traitements informatiques ou "macro-traitements", ils interviennent normalement après les "micro-traitements" (la balance entre micro et macro traitements est à penser en termes de coûts, de délais, de complémentarité et de qualité globale de l'enquête).

Les traitements informatiques ont lieu dès livraison d'un fichier réputé "propre" par les Directions régionales ou par l'établissement centralisateur, donc après contrôles, apurements, codifications éventuelles. S'il y a encore des contrôles à ce stade, ils viseront plutôt à mesurer une qualité de l'enquête, donneront lieu le cas échéant à corrections automatiques mais se feront en principe sans retour aux dossiers. En principe seulement, car si, à ce stade, le statisticien trouve des anomalies qu'il ne peut expliquer, il y aura retour aux dossiers incriminés pour trouver l'explication ; soit explication correspondant à une réalité non prévue et qui permettra d'expliquer les

tableaux, soit anomalie ou erreur réelle qui n'avait pas donné lieu à contrôle aux stades précédents et qui donnera lieu à correction adaptée à chaque dossier, ou à une correction automatique si cela est jugé préférable. On peut aussi, à ce stade, relever des erreurs systématiques de traitement, erreurs liées le plus souvent à une compréhension erronée de certaines consignes. Selon le taux d'erreur on pourra envisager une correction systématique et apurer manuellement les cas qui ne relèvent pas de ce traitement systématique, ou bien agir uniquement sur les seuls cas aberrants.

Une fois le fichier enfin "propre" au sens du statisticien, interviennent les "macro-traitements", les redressements, les procédures de remplacement des parties absentes ou aberrantes (*Hot Deck par exemple, et là aussi il convient d'être prudent*).

Et enfin, la production des tableaux tant attendus, pour analyse et publication.

Revenons au début de l'enquête, et donc aux phases élaboration du questionnaire, saisie, traitements manuels, contrôles et codification.

2. L'élaboration du questionnaire

La conception du questionnement est une étape vitale dans la recherche de qualité. Il semble évident qu'il faut poser les bonnes **questions** pour obtenir les bonnes réponses sur les sujets étudiés, mais ce n'est pas si simple.

Il faut d'abord que les entités interrogées **comprennent** les questions, que cela rentre dans leur domaine de compréhension, et que l'information demandée leur soit **accessible**.

Dans la plupart des "grandes" enquêtes on interroge une grande diversité de personnes, avec des situations professionnelles et personnelles, des niveaux d'études, des préoccupations, très différentes. Il faut savoir sortir du cadre de référence du statisticien pour que les questions soient compréhensibles par tous sans être hors de portée des uns et ridicules pour les autres.

Mais la liberté n'est pas totale, certaines questions ou enchaînements de questions doivent être relativement **normalisés** pour permettre si possible des comparaisons avec l'enquête précédente, avec d'autres enquêtes, avec des sources extérieures.

L'enchaînement des questions est délicat, il doit suivre une logique compréhensible par la personne interrogée, et avoir un lien logique avec le thème annoncé de l'enquête.

Cette logique doit aussi être accessible à l'enquêteur qui va piloter le questionnement et éviter qu'il ne s'égaré dans un dédale de filtres, prenne de mauvais chemins et pose des questions inutiles tout en ne posant pas les questions utiles.

S'impose également un balisage clair et standardisé des filtres qui justifient que l'on saute telle et telle question sans objet dans le cas d'espèce.

S'il n'y avait que le questionnement, ce serait encore assez simple, mais il y a les **réponses**. Il faut prévoir toutes les réponses possibles et prévoir leur **transcription ou mode de notation**.

Cette complexité, et ce n'est pas fini, fait que doit intervenir une notion de "**charte graphique et logique**" pour les questionnaires, **pour tous les questionnaires livrés au même réseau d'enquêteurs**.

Il faut donc standardiser le graphisme des titres, des sous-titres, des questions et de la numérotation de ces entités. Standardiser les modalités de réponse (*que par exemple le oui soit toujours 1, le non toujours 2, le ne sait pas toujours 9, etc.*) et les exclusions (*les exclusions en italique par exemple*), le graphisme et la logique des systèmes de notation des réponses (*toujours cercler, ou toujours cocher les questions à choix multiple, toujours caler à droite ou toujours caler à gauche les quantités*).

Standardiser le graphisme et la logique des renvois ou filtres qui dépendent des réponses faites, (*toujours renvoyer sur des entités d'un niveau bien défini : question, chapitre, partie ...*) et éviter autant que possible le jeu de piste des renvois soumis à des conditions complexes (*si... et si..., sinon...*), faisant référence à des réponses déjà lointaines et envoyant vers des destinations éloignées trop fines. Plus le filtre envoie loin, plus on doit viser gros (c'est-à-dire des parties importantes).

Le questionnaire est certes l'outil de questionnement, mais il doit aussi être un guide pour l'enquêteur.

Il faut donc prévoir un maximum d'indications pour éviter le recours à des documents annexes, cartes de codes, cahiers d'instructions, pas toujours faciles à mémoriser et peu manipulables lors de l'interview, ceci sans pour autant saturer l'enquêteur, donc aération du questionnaire. Aération qui, à quantité égale d'information contenue, diminue la densité, augmente la lisibilité, ce qui est positif, et l'épaisseur du questionnaire, ce qui est négatif, en le rendant plus pesant pour l'enquêteur et visuellement plus stressant pour l'enquêté. Il faut trouver le juste milieu.

Mais le questionnaire ne sert pas qu'à l'enquêteur, il va être **vérifié**, il faudra parfois visualiser les zones à vérifier en priorité, certains renseignements recueillis en clair devront être codifiés et ceci dans des zones réservées (zones "réservées" dans lesquelles

l'enquêteur n'aura pas la tentation d'entrer - encore un graphisme particulier, *bacs "fermés" par exemple*).

À moins d'opérer une transcription sur un document spécifique, opération coûteuse en temps et douteuse pour la qualité, car source d'erreurs de recopie, ce même questionnaire va être saisi.

La "charte graphique et logique" doit être signifiante pour l'enquêteur et le "vérificateur" mais aussi pour la saisie. Il faut faire figurer sur le questionnaire les indications techniques liées à la logique de saisie (numéros de formats, compteur si format multiple, noms des variables, essayer d'indiquer par un graphisme particulier si tel filtre est programmé, si la machine va "sauter" ou si au contraire il faut saisir toutes les zones même vides, ceci servant de vérification).

Le questionnaire devient le **document de référence**, base de discussion lors de l'élaboration, puis base de travail pour toutes les parties prenantes, porteur de presque toutes les définitions de l'enquête. (*sont "hors questionnaire", les instructions diverses écrites et orales à tous les intervenants, les cartes code éventuelles, les nomenclatures de chiffrage, les contrôles et traitements manuels et informatiques variés que l'on fait subir aux variables collectées, encore que la plupart des contrôles "mécaniques" - modalités, filtres - soit implicitement contenue dans le questionnaire*).

Le problème se pose de manière presque identique pour la réalisation d'un **questionnaire "écran"**, c'est-à-dire d'un questionnaire qui sera collecté "en direct", soit lors d'une interview en face à face avec un ordinateur portable, ou par interview téléphonique avec questionnement "écran" et saisie directe des réponses (*sur un ordinateur de bureau pour les agents qui réalisent ces interviews aux heures de bureau, ou sur un portable pour les enquêteurs chargés de réaliser les interviews, depuis leur domicile aux heures non ouvrables*).

Dans ce cas, le nombre d'intervenants est diminué puisque la saisie est intégrée dans la collecte ainsi qu'une bonne part du travail de vérification manuelle et de contrôle informatique. Le questionnaire "écran" n'est plus un outil destiné au travail de plusieurs acteurs, mais destiné au seul "enquêteur-saisisseur-contrôleur".

Une charte graphique s'impose néanmoins avec force pour bien différencier ce qui est question et doit être lu, ce qui est **modalité à lire** pour faciliter la réponse, ou **modalité à ne pas lire** pour coder la réponse spontanée, ce qui est **consigne ou explicitation destinée au seul enquêteur**, etc.

Dans le cas du questionnaire "écran" intervient de surcroît une notion d'**ergonomie fonctionnelle**.

Même si la charte graphique et logique du questionnaire "papier" participe d'une certaine ergonomie de cet outil, il s'agit malgré tout d'un outil "passif".

Le questionnaire "écran", lui, est "interactif", et l'ergonomie visuelle et logique est particulièrement importante. En l'état actuel, un écran est moins "lisible" qu'un document papier, et ceci milite d'autant plus fortement pour une ergonomie soignée qui permet de s'y retrouver.

Tout d'abord trouver le bon compromis entre aération et permanence de l'affichage, entre aération et défilement incessant des pages.

S'il est agréable d'avoir des écrans peu denses, il est néanmoins utile de garder sous les yeux les réponses aux questions précédentes si elles constituent un ensemble, et que ces réponses soient repérées par un nom de variable évoquant la question.

Cela semble un détail, mais il faut que, au fur et à mesure du défilement des "pages", les questions à lire à l'enquêté se situent toujours au même endroit de l'écran, avec, si possible, un certain type d'affichage ou de couleur, de même pour les modalités de réponse et les explicitations réservées à l'enquêteur, de même pour le système de notation des réponses. Les écrans qui "dansent" provoquent fatigue visuelle et erreurs.

À l'exception des noms de variables, tout ce qui concernait le graphisme lié à la logique et aux attributs de saisie disparaît. Par contre la logique de saisie, enchaînements, filtres et contrôles, bien qu'invisible, est "intégrée" dans le programme qui, en coulisses, fait "fonctionner" le questionnaire. Nous y reviendrons au chapitre "saisie directe".

Si le questionnaire écran contient toutes les informations voulues, il est par contre plus difficilement "le document de référence", base de discussion multi-partites pour la mise au point, ceci malgré les possibilités éventuelles de listage papier, car il ne suffit plus de lire et relire, il faut vérifier toutes les fonctions automatiques. Il ne s'agit plus de maquettage au sens traditionnel, mais de "programmation", et donc d'un produit moins "lisible" par le non informaticien pour ce qui concerne les moteurs de l'interactivité, et plus difficile à vérifier, car il faut vérifier par de multiples essais, tous les chemins possibles, toutes les logiques possibles.

La suppression de la phase "impression" permet de gagner deux bons mois et permet aussi des retouches de dernière minute (*encore qu'il convienne de se méfier de ces retouches qui interviennent après les vérifications de fond*).

Par contre, la préparation du questionnaire écran est un tout, plus complexe, puisque intégrant davantage de fonctions auparavant destinées à différents acteurs intervenant successivement, et dont on pouvait étaler la préparation dans le temps. La pièce est en un acte et on a donc moins le droit à l'erreur.

Le problème se pose différemment pour les questionnaires auto-administrés, soit remis, soit envoyés aux enquêtés. Cas fréquent des "carnets" à renseigner sur une période de temps pour les enquêtes ménages, cas fréquent des questionnaires envoyés aux entreprises. La charte graphique et logique doit ici être simplifiée au maximum pour être clairement lisible et donner une apparence de facilité au remplissage du document. Il doit y avoir donc, juste ce qu'il faut de consignes, des exemples, le minimum indispensable de surcharges "techniques" destinées au vérificateur-chiffreur et à l'opérateur de saisie qui auront à intervenir ensuite.

Bien évidemment, tous ces "questionnaires", papier, écran, auto-administrés, font l'objet d'un ou plusieurs tests auprès de mini échantillons, pour vérifier leur tenue face aux enquêtés, difficultés de compréhension, information difficile à retrouver, réticences à aborder tel ou tel sujet ou à tenir tel carnet de compte, leur passation par les enquêteurs, complexité trop grande, temps d'enquête trop long, etc.

Les tests amènent parfois à des bouleversements profonds dans la structure du questionnement, dans la rédaction des questions, dans le nombre de questions posées, mais donnent toujours lieu à des mises au point utiles. Les questionnaires-tests ainsi recueillis sont parfois utilisés pour tester la suite des traitements, du moins lorsqu'ils sont à un stade d'élaboration proche du stade final.

3. La collecte de l'information

En dehors de quelques enquêtes simples réalisées par téléphone et en dehors des enquêtes réalisées par voie postale - auprès des entreprises le plus souvent, les enquêtes "complexes" auprès des ménages sont réalisées par des enquêteurs pigistes et réparties dans les différentes Directions régionales.

Il s'agit d'abord d'assurer un bon "rendement" de l'enquête en évitant les refus. Refus qui, trop nombreux, risqueraient de biaiser les résultats en empêchant des redressements corrects. Il est clair que si l'on utilise des enquêteurs, solution relativement onéreuse, pour réaliser ces enquêtes, c'est qu'ils ont un meilleur "rendement" qu'une enquête par voie postale ou par téléphone (*il est en effet difficile de contraindre des ménages à répondre, même si l'enquête est obligatoire ; la conviction est sans doute un meilleur garant de qualité et de sincérité dans les réponses que la contrainte*).

L'emploi d'enquêteurs est aussi justifié par le caractère souvent très complexe des enquêtes, complexité moindre qu'il n'y paraît car tout le monde ne répond pas à toutes les questions, mais les questionnaires dans leur globalité sont souvent assez impressionnants, et un enquêteur "interface" est indispensable pour les administrer.

Ces enquêteurs sont **formés**. Formés d'abord à représenter l'organisme enquêteur, par une explication des missions générales de cet organisme, formés ensuite à une certaine déontologie, vis-à-vis des enquêtés (conviction et non menace), vis-à-vis des enquêtes en général (secret professionnel, principe du sondage aléatoire, souci de qualité), vis-à-vis de l'enquête en cours (utilisations, compréhension des questions et des réponses, compréhension des enchaînements, consignes particulières).

Les enquêteurs (terrain ou téléphone) sont (à l'INSEE) **annoncés** par une "lettre-avis" qui a pour but de prévenir le ménage, d'officialiser l'enquête, d'expliquer pourquoi cette enquête est réalisée, mais qui vise surtout à faire la différence entre l'enquêteur et un quelconque démarcheur.

Sont parfois joints, ou remis par l'enquêteur, quelques résultats tirés de l'enquête précédente, ce qui est très "motivant" et apprécié par les enquêtés, qui perçoivent mieux ainsi, l'utilité de leur participation.

4. La saisie

La saisie est un **transfert** d'information d'un support à un autre, transfert avec changement de forme (*de la réponse transcrite sur papier au support magnétique, de la réponse orale directement au support magnétique*).

Ce transfert constitue une **rupture de charge** dans la chaîne de traitement.

Ce transfert implique une **intervention humaine**, et donc, des risques d'erreur.

De quels genres de transferts peut-il s'agir ?

J'évoquerai à peine certains types de saisie peu utilisés pour les enquêtes statistiques.

La saisie des codes-barres par exemple, utilisée à l'INSEE pour lire avec un crayon optique le code SIREN des entreprises et établissements répondants (*le code-barre ayant été imprimé sur le questionnaire envoyé*). Ceci est fiable et évite les erreurs de saisie.

La lecture optique est également utilisée parfois. Soit lecture de cases cochées par la personne interrogée (procédé déjà ancien mais ne concernant que des documents très réduits), soit reconnaissance de caractères. Je ne sais pas si ceci est utilisé dans certaines enquêtes, en tout cas pas à l'INSEE, à ma connaissance, mais il est probable que, les progrès de la scannérisation ajoutés aux progrès logiciels en matière de reconnaissance optique des caractères (*dactylographiés pour l'instant, manuscrits en balbutiements*),

permettront sous peu la "scannérisation-compréhension" de libellés, ceci sans saisie manuelle.

Revenons au présent.

Une saisie de qualité, est une saisie qui ne perd, ni ne crée, ni ne modifie l'information d'origine (*si ce n'est dans sa forme*).

La qualité de la saisie dépend éminemment du mode de transfert et des sécurités qui accompagnent ce transfert.

J'évoquerai :

La saisie de masse, la saisie assistée et la saisie directe

- Chaque type de saisie correspond à un certain "style"
- Les acteurs sont différents ;
- Les rôles sont différents ;
- Les actions automatiques de l'outil de saisie sont différentes ;
- L'organisation de la chaîne de traitement est différente.

4.1 La saisie de masse

La saisie dite "de masse" correspond à la saisie par des "opérateurs ou opératrices de saisie" de documents renseignés. Par "document" il faut entendre questionnaires ou formulaires imprimés sur lesquels certaines zones prédéfinies ont été renseignées ou cochées ou entourées par le répondant ou par un enquêteur. Ceci se pratique par frappe sur un clavier de terminal informatique - généralement relié à un système spécifique - ou sur un micro-ordinateur autonome ou relié à un réseau local à un serveur.

La saisie dite "de masse" correspond à un certain "style" de saisie. L'opérateur ou l'opératrice lit le document, ou plus exactement les renseignements portés dans les zones à saisir et saisit, avec une grande dextérité, en ne regardant l'écran que périodiquement pour vérifier s'il en est bien là ou il croit être, ou lorsque son attention est attirée par un signal sonore d'anomalie.

La saisie de masse souffre d'un handicap majeur, son manque de sécurité. Malgré le professionnalisme des opérateurs et leur faible taux d'erreurs de frappe, il convient d'éviter que des erreurs de frappe (et les touches sont terriblement voisines les unes des autres), ou des erreurs de lecture (et c'est souvent très mal écrit, lorsque ce sont les répondants qui ont renseigné le document) ne détériorent la qualité de l'information saisie. On a donc recours à différents procédés pour éviter ou éliminer, autant que faire se peut, les erreurs de saisie.

Premier procédé, coûteux mais efficace, qui requiert un logiciel adapté, la double saisie ou saisie-vérif. Un deuxième opérateur ressaisit le document et le système signale toute différence par un bip sonore. Ce deuxième opérateur doit alors juger si la divergence a pour origine une erreur lors de la saisie initiale ; il maintient alors sa frappe, ou une erreur lors de sa propre saisie ; dans ce cas, il doit se corriger. Au total aucune divergence ne doit subsister entre première et deuxième saisie.

Ce procédé est souvent mis en œuvre au début de la saisie d'une enquête, des statistiques d'erreurs sont produites, avec indication des zones les plus touchées, et l'attention des opérateurs est attirée sur ces zones.

Lorsque, après rodage, le taux d'erreur est descendu suffisamment bas, le taux étant apprécié selon le caractère sensible ou non des variables, la double saisie peut être abandonnée. Elle peut aussi ne jamais être abandonnée lorsque les critères de qualité l'exigent.

On peut aussi prévoir une double saisie systématique mais réduite à certaines zones sensibles.

Pourquoi ne pas faire de contrôles pour éviter les erreurs, du moins pour les zones stratégiques ?

C'est possible, et il y en a, de type "mécaniques" (*modalités des codes, fourchettes de valeurs admises, cohérence entre questions filtrantes et questions filtrées*).

Mais si l'on prévoit que Sexe = 3 déclenche un message d'erreur en saisie, l'opérateur corrigera sa frappe si l'erreur lui est imputable, par contre il n'est pas habilité à interpréter le document s'il y est réellement noté Sexe = 3. Il devra donc "forcer" cette réponse erronée.

L'erreur devra être détectée ultérieurement par un programme de contrôle, listée et traitée par un gestionnaire qui tentera de déterminer si, d'après le prénom, Sexe = 1 ou 2. Si ceci s'avère impossible pour les Claude ou les Dominique, on laissera Sexe = 3 et ceci sera redressé ultérieurement par des procédés automatiques tel le Hot-Deck.

Ces procédures automatiques sont à utiliser avec précaution, (*on a découvert ainsi dans une enquête sur la santé, des Sexe = 2 (femmes) souffrant de la prostate ou des Sexe = 1 (hommes) enceintes...*).

Pour la cohérence entre questions filtrantes et questions filtrées, la difficulté est la même ; il n'appartient pas à l'opérateur, au cas où l'erreur ne serait pas de son fait, de déterminer si l'erreur se situe au niveau des questions filtrantes ou au niveau du chemin suivi par le répondant ou l'enquêteur. En tout état de cause la saisie de l'incohérence peut être utile pour juger de la qualité de la collecte.

Pour ce qui concerne la saisie de quantités par exemple, on peut prévoir un traitement manuel préalable qui consiste à indiquer le nombre de chiffres significatifs et donc permettre ainsi un contrôle, mais 92 765 au lieu de 29 765 ne sera pas détecté en saisie unique, sauf s'il existe un total qui permette une vérification.

S'il existe un moyen de détecter l'erreur par une cohérence, ce contrôle et sa correction seront reportés à une phase ultérieure. Tout contrôle en saisie qui détecte une erreur ou une anomalie qui peut être le fait soit d'une erreur de frappe soit d'une erreur de notation devra pouvoir être "forcé".

Compte tenu du "style" de la saisie de masse, et du rôle traditionnellement dévolu à l'opérateur, rôle tourné davantage vers la quantité et la rapidité de saisie que vers l'analyse des documents, et même si cela est techniquement possible, on évite de programmer des contrôles tels que l'opérateur face à une anomalie détectée ne sache quelle décision prendre. Il y a, là aussi, un compromis à trouver entre contrôles utiles à la qualité de saisie, pour éventuellement éviter une double saisie, et contrôles ralentisseurs, car ne pouvant généralement pas être corrigés par l'opérateur. Ce compromis est fonction de la nature et de la qualité de remplissage des documents à saisir. Plus le document est renseigné de manière fiable, plus on pourra implanter de contrôles en saisie, puisque les erreurs détectées ne devraient pas provenir du document. Tel est le cas des enquêtes auprès des ménages collectées par des enquêteurs qui veillent à une bonne qualité et une bonne logique de remplissage puis vérifiées en bureau. Tel n'est pas le cas de documents de type administratif ou de documents renseignés directement par les répondants.

Une autre terreur de la saisie de masse est "le décalage", terreur car une longue séquence peut ainsi être complètement pervertie, pour un simple caractère en trop ou en moins.

Le décalage se produit lorsque de longues séries de chiffres sont à saisir (5467895428546, qui peut devenir 546895428546, cas des identifiants), ou lorsque il y a plusieurs chiffres identiques (25000 peut devenir 2500 ou 250000, cas des quantités), ou bien encore lorsqu'il y a une alternance de chiffres analogues (1, 1, 2, 1, 1, 1, 2, 2, 1, 2, 1, 1 ... cas d'une succession de réponses par des oui, non).

Quelle que soit la cause, il faut se donner les moyens d'arrêter les décalages, et de les arrêter vite, pour éviter une reprise trop importante.

Soit on trouve, de-ci, de-là des solutions "naturelles" pour détecter les décalages : clé pour l'identifiant, succession logique de valeurs, variable ou code à modalités particulières, soit on repère dès la conception du questionnaire les zones dangereuses et on implante artificiellement des variables anti-dérapiage. Les cases de contrôle du nombre de chiffres significatifs des quantités en sont un bon exemple. On peut enfin prévoir la saisie d'un caractère particulier " * " par exemple, ou exiger une frappe de type "entrée" à chaque fin de format.

Un autre procédé a été mis en œuvre, la vérification de la saisie par sondage (SVS).

Le principe était le suivant : saisie initiale complète. Reprise en double saisie de certains lots tests tirés aléatoirement par programme. Statistiques de divergences établies pour ces lots entre 1^{re} et 2^e saisie.

Selon les taux de divergences constatés, soit l'ensemble des lots était réputé "bon", soit l'ensemble des lots était réputé "mauvais" et il fallait alors faire la double saisie de tous les lots, à l'exception des lots tests, déjà ressaisis.

Ceci induisait une gestion des lots assez compliquée qui a progressivement conduit à l'abandon de ce procédé.

Par contre, une vérification par sondage de la saisie lors des phases "population légale" et **Colibri** a été fort bien acceptée et a donné d'excellents résultats. Mais la gestion en était centralisée ; ceci explique peut être cela.

La saisie de masse traditionnelle est en train d'évoluer vers des saisies plus sophistiquées, plus lentes mais plus analytiques, plus intégrées, donc au total plus performantes, pour se rapprocher du rôle du vérificateur manuel ou "gestionnaire d'enquêtes" qui de son côté évolue vers la saisie par l'utilisation interactive de l'outil informatique. Nous verrons ceci au chapitre **saisie assistée**

À terme, ces deux "métiers" devraient fusionner.

Ceci ne va pas sans poser de problèmes de tous ordres : statutaires d'abord, les opérateurs de saisie ont un statut particulier avec primes particulières et horaires aménagés, pauses ergonomiques. Lorsque les uns sont amenés à faire le même travail que les autres, ces statuts différents peuvent provoquer frictions et revendications.

Autre problème : celui des qualifications ; en l'état actuel, les opérateurs de saisie sont réputés non qualifiés pour "comprendre" les dossiers. Un programme de formation et de requalification est en cours.

Dernier problème : celui des chasses gardées ; l'introduction de l'informatique dans presque tous les domaines a permis, d'une part la suppression de tâches fastidieuses et d'autre part des gains de productivité importants, chacun veille donc jalousement sur son travail, surtout si celui-ci a, précisément, été valorisé par l'outil informatique, et il est difficile de reprendre des travaux déjà "en main".

4.2 La saisie assistée

Autant la saisie de masse est une saisie peu assistée (les opérateurs ne regardent pas l'écran), et peu contrôlée puisqu'il ne sert à rien de détecter des erreurs que l'opérateur ne sera généralement pas en mesure de corriger, n'ayant pas une connaissance suffisante de l'enquête en général et du dossier particulier, autant la saisie assistée a pour mission première d'utiliser pleinement les possibilités de l'outil informatique pour réaliser une saisie-analyse contrôlée et assistée des dossiers.

Le "style" est ici très différent. Autant le questionnaire tente d'intégrer les contraintes de différents acteurs, autant ce style de saisie tente d'intégrer les fonctions des opérateurs de saisie, et du gestionnaire d'enquête. Le "gestionnaire d'enquête" est réputé avoir une bonne compréhension de l'enquête en général et du dossier qu'il traite en particulier, il a pour mission première d'analyser le document, de le vérifier, de le compléter, de le corriger.

Pourquoi ne pas le faire assister par l'ordinateur dans son travail ?

Il est déjà assisté par la fourniture de listings d'anomalies, qu'il doit corriger, et dont les corrections sont répercutées sur le fichier magnétique dans une phase d'apurement. Les phrases sont cependant dissociées.

Mais on peut pousser davantage l'intégration des tâches en intégrant les phases de saisie, d'analyse, de contrôle, de correction, d'aide à la codification.

La saisie par un "gestionnaire d'enquête", si elle est moins performante en terme de caractères frappés par seconde, si elle est moins fiable au niveau de la frappe, permet par contre une assistance et des contrôles interactifs beaucoup plus poussés et donc une plus grande intégration du travail de traitement du dossier et de son contrôle. On perd en vitesse pure, mais on gagne en délai global de traitement en évitant le découpage du travail en phases distinctes, avec les ruptures de charge induites, en évitant les itérations entre les phases. En évitant aussi, ce qui n'est pas négligeable, la mise au point des différentes phases d'une chaîne dissociée, la formation et le rôdage des différents acteurs, en évitant les redites aux uns et aux autres, en évitant les "couacs" liés aux incompréhensions et aux redites mal reprises d'une phase à l'autre.

On a donc des gains en délais de préparation, des gains en délais de réalisation, des gains en qualité en évitant les pertes en ligne.

Sur quoi peut porter cette intégration ?

Il peut s'agir de consignes purement visuelles, délivrées au fur et à mesure de la saisie, consignes destinées à attirer l'attention sur telle ou telle difficulté, et à vérifier s'il y a conformité du document à ces consignes.

Ceci est en fait peu utilisé puisque le conseil de vérification peut et est en général traduit par un contrôle programmé.

Dans le mode saisie assistée et contrôlée, on peut avoir une densité de contrôles importante : contrôles de structure, de modalités et de cohérence. Il convient à ce stade de bien faire la différence entre erreur et anomalie. Une **erreur** correspond à un cas réputé impossible, faute d'inscription ou de transcription qu'il convient de corriger, si possible ; il y a donc des contrôles qui détecteront des erreurs. Une **anomalie** correspond à un cas peu probable (en général lié à une cohérence), mais admis comme possible. Le contrôle d'anomalie vise à s'assurer qu'il n'y a pas eu erreur de transcription ; s'il y a conformité au document d'origine, l'anomalie sera confirmée. Il convient de bien différencier pour le gestionnaire les erreurs (à corriger), des anomalies (à examiner, et corriger ou confirmer le cas échéant), l'idéal est que le message comporte l'indication : Erreur :, Anomalie :

Il arrive parfois que, après étude des questionnaires, ce qui était réputé "erreur" s'avère correspondre à une réalité que l'on n'avait pas prévue *a priori*.

Le gestionnaire signalera ces cas et le statut du message devra être modifié en message d'anomalie.

Le gestionnaire doit être formé au traitement des messages issus de ces contrôles pour respecter l'information d'origine et ne pas opérer une mise aux normes à tout prix. Problème de déontologie, certes, mais la déontologie n'est pas innée, elle doit être inculquée et elle doit être surveillée.

Il ne faut modifier qu'à coup sûr, pour **corriger une erreur** d'inscription manifeste. Ainsi s'il y a incohérence entre la réponse à une question filtrante et la suite on peut penser à une erreur (*par exemple : Êtes vous locataire ou propriétaire ? Réponse notée : locataire ; la suite des questions renseignées correspond pourtant à la partie Propriétaire, remboursements d'emprunts, nombre de prêts, année d'acquisition, durée des remboursements, etc, tandis que la partie Locataire n'est pas renseignée, on peut raisonnablement penser que le filtre est erroné et le corriger*). Autre cas : on annonce quatre individus dans le ménage et on en trouve cinq renseignés, on peut raisonnablement penser que le quatre était erroné. Si par contre on en annonce cinq et que l'on n'en

trouve que quatre renseignés, il se peut que le cinq soit faux mais il se peut aussi que la description d'un individu ait été oubliée. Dans ce cas il convient d'abord de rechercher dans le document si d'autres données corroborent plutôt le cinq ou plutôt le quatre. Enfin, faute d'autre solution on peut tenter de revenir à la source - enquêteur ou ménage - ce qui implique une réaction rapide par rapport à la collecte (deux ou trois jours, au-delà l'enquêteur a vu trop d'autres ménages pour se souvenir, et le ménage risque d'être réticent face à cette "contre-enquête").

À défaut, on laissera subsister l'incohérence qui sera signifiante quant à la qualité de collecte.

Il est important de quantifier erreurs et anomalies pour mesurer la qualité de collecte et la qualité de saisie, ceci pour chaque contrôle. Ceci implique que les corrections soient, elles aussi, analysées pour déterminer quelle est la part de responsabilité des différents acteurs dans les erreurs et anomalies : enquêté, enquêteur, opérateur de saisie. La mise en œuvre de ces analyses est malaisée mais souvent pratiquée lorsque le traitement est centralisé.

Par ailleurs, un contrôle qui ne ramène rien dans ses filets est soit inutile, soit mal programmé. Un contrôle qui ramène beaucoup dans ses filets a été soit mal programmé, soit révèle un défaut dans la conception de la question et de ses modalités de réponse. Le statisticien devra en être prévenu.

Autre type d'assistance : l'**aide à la codification**. Un ordinateur est très efficace pour enchaîner et afficher les pages d'une nomenclature (en arbre ou non) et réaliser le chiffrement (contrôlé pas à pas) par niveaux successifs, plus efficace que le recours au fascicule, souvent long à parcourir, et plus sûr car il évite le recours à la mémoire, plus rapide après une phase d'apprentissage, mais source d'erreurs, mémoire à laquelle on fera d'autant plus appel que le fascicule est épais et donc consommateur de temps de recherche.

Autre assistance possible : la **codification automatique**, soit "en ligne" soit "en différé".

L'utilisation de telles procédures de codification automatique ne s'improvise pas et implique une préparation de la base qui sera utilisée pour la reconnaissance des libellés et l'attribution du code. Ceci requiert des **experts de** la nomenclature considérée. Cette expertise est plus facile à acquérir et à mobiliser en centralisé.

À partir de la saisie d'un libellé en clair "*carottes*" l'ordinateur cherche et annonce un code, ou propose de choisir entre plusieurs codes possibles (*sans doute faut-il retoucher la base pour éviter ces incertitudes*), ou cherche et ne trouve pas, il faut alors envisager une codification manuelle, et "apprendre" ce nouveau cas au système, si la personne est qualifiée pour ce faire.

Autre difficulté, l'orthographe, soit du répondant, soit de l'enquêteur, soit du gestionnaire, soit de l'opérateur de saisie (*car le problème se pose aussi en saisie de masse*). Faut-il **respecter** ce que dit ou écrit le répondant, ou le traduire dans une forme qui semble correcte ?

Les variantes ne posent pas de difficulté technique. Le code (*pour "carrote"*) devra simplement être "appris" par l'ordinateur en tant que nouveau produit et on lui indiquera le même code que pour le libellé correct ("*carottes*")

Ce qui pose problème est l'**interprétation** faite par les individus selon leur propre compréhension, et qui peut aboutir à des choses très différentes. On tombe ici dans un classique des stages de communication : il y a ce que l'individu a voulu dire, ce qu'il a dit, ce que l'autre a compris, et ceci à chaque stade.

Pour ce qui concerne la saisie des libellés, il faut donc établir une ligne de conduite : soit respect de ce qui est noté, respect coûteux car il faut veiller à la conformité de ce que l'on tape avec l'original, soit bien fixer les degrés de liberté pour "arranger", avec les risques d'interprétation que cela comporte, avec le bénéfice cependant de diminuer les variantes ; l'expérience montre que la conformité est préférable à la remise en forme qui peut aboutir à trop normaliser des cas apparemment absurdes, absurdes parce que non connus ou mal compris par l'intervenant.

Si le taux de réussite de la codification automatique atteint par exemple 80 %, on peut estimer que l'opération est rentable, puisque qu'il ne reste plus à assurer la codification manuelle que de 20 % des libellés.

Mais elle implique la saisie de 100 % des libellés, saisie plus longue que la saisie d'un code.

Les libellés non reconnus, les "rejets", devront être codés manuellement, et "appris" au système, si leur occurrence le justifie, par un **expert**. Il convient donc de peser l'intérêt quantitatif et qualitatif de la mise en œuvre de telles procédures.

Un intérêt annexe est que l'on disposera ainsi d'un catalogue de libellés ; on pourra mieux savoir ce qui "entre" dans tel ou tel code, avec quelle fréquence, quelle pondération, et donc permettre d'affiner la nomenclature, en isolant par exemple tels ou tels libellés dont on avait sous estimé l'importance ou la fréquence parce que "noyés" au sein d'une rubrique, ou en supprimant tel ou tel libellé qui n'est plus jamais rencontré. L'intérêt "marketing" d'un tel catalogue peut aussi entrer en jeu pour décider de pratiquer la saisie des libellés suivie d'une codification automatique, plutôt qu'à une codification manuelle.

4.3 La saisie directe

La saisie directe par un enquêteur des réponses faites par l'enquêté, sans passer par l'intermédiaire du papier, permet de passer en une seule étape des réponses fournies par l'interviewé à un fichier magnétique "relativement" contrôlé.

Ce mode de saisie se rapproche, par sa technologie, de la saisie assistée et contrôlée du "gestionnaire d'enquête", avec un souci d'intégration plus poussé encore en terme d'intervenants, mais moins poussé en termes de contrôles et d'assistance. Donc un "style" très différent de la saisie assistée.

En effet, si pour le "gestionnaire" on peut pousser l'assistance assez loin, il faut par contre bien mesurer ce que l'on peut, et ce que l'on ne peut pas demander à un enquêteur qui opère "en direct", en situation parfois précaire vis-à-vis de l'enquêté, qu'il s'agisse d'une interview en face à face ou d'une interview téléphonique. À la différence du "style" saisie assistée, l'enquêteur ne peut "perdre le contact" avec l'enquêté en étant par trop absorbé par sa machine, comme le conducteur ne peut perdre le contact avec la route en étant trop occupé par ce que lui indique son tableau de bord.

Tout d'abord, nous l'avons évoqué lorsque nous avons traité des questionnaires "écran", le direct impose une ergonomie particulière, une ergonomie "simple" mais rigoureuse, des règles, peu de règles, pas d'exceptions aux règles.

La machine doit lui simplifier la tâche, et non la lui compliquer.

La logique de la saisie directe est, nous l'avons vu plus haut, "invisible". Elle peut théoriquement, à la différence du questionnaire papier, et comme la saisie assistée, devenir très complexe puisque l'enquêteur n'a plus à la gérer. Il faut néanmoins qu'il s'y retrouve et la comprenne.

S'il se retrouve dans un mauvais chemin dû à une erreur de frappe pour une question filtre, il faut d'abord qu'il s'en aperçoive, et qu'il s'en aperçoive rapidement.

Il est probable que les questions de ce "mauvais chemin" n'auront pas de sens si elles avaient lieu d'être "sautées", encore faut-il que l'enquêteur sache revenir, et revenir vite, car cela se passe "en direct", à l'erreur qui a provoqué ce déroutement, pour la corriger et retrouver ainsi la bonne voie.

La ou les questions filtres ne peuvent donc être trop lointaines (en arrière), ni trop nombreuses (pour limiter ses choix de marche arrière), ni trop complexes (*ici aussi éviter les si... et si..., sinon...*).

Cette difficulté milite pour une complexité "compréhensible".

Bien évidemment une explicitation des chemins pris par des messages permet à l'enquêteur de mieux comprendre pourquoi, c'est à dire en fonction de quelles réponses il suit tel chemin. Mais ceci "charge" l'écran..... et complique la programmation.....

La saisie directe permet, et c'est un de ses grands avantages, des contrôles "en direct" qui permettent de vérifier et corriger éventuellement des erreurs de frappe, de compréhension de la question ou de la réponse.

Contrôle égale message, et message compréhensible, par l'enquêteur, mais aussi par l'enquêté auquel l'enquêteur devra lire ou traduire le message. Ceci limite donc la complexité possible des contrôles, vis à vis de l'enquêté mais aussi vis à vis de l'enquêteur, car si correction il doit y avoir, il faut que les variables incriminées soient facilement repérables et donc peu nombreuses, faciles d'accès donc peu lointaines (*donc, ici aussi, éviter les si... et si..., sinon....*).

Enfin le nombre de contrôles ne doit pas être trop élevé, sous peine de transformer l'enquête en interrogatoire contradictoire qui provoquerait un rejet de la part de l'enquêté. Les possibilités théoriquement étendues du "direct" ont donc leurs limites pratiques.

Ces limites font que l'on se bornera à tirer tout le parti possible du "direct", c'est à dire "soulager" l'enquêteur de tout ce qui est "fonctionnement" du questionnement et traiter tout ce qui, en cas d'erreur ou d'anomalie, pour être correctement traité, impliquerait un retour à l'enquêté. Par contre sera reporté sur une phase ultérieure, manuelle ou automatique tout ce qui peut, sans trop de dommage, être traité de manière différée.

Outre les avantages liés à la sécurité accrue (mais pas absolue) apportée par l'enchaînement automatique des questions et la vérification directe auprès de l'enquêté de certaines anomalies non récupérables après coup, la saisie directe offre l'énorme avantage de mettre à disposition des fichiers saisis, directement issus de la phase de collecte, et donc de raccourcir considérablement les délais entre collecte et exploitation, si tant est que l'on n'opère pas ensuite des contrôles et corrections manuelles nombreuses.

L'essentiel étant assuré lors de la saisie, on limitera les micro-traitements (manuels) pour préférer des macro-traitements (automatiques).

5. Les traitements manuels et l'assistance informatique

Avant ou après la saisie "de masse", après la saisie directe, peuvent intervenir différents traitements manuels. Lourds avant et/ou après la saisie de masse, légers ou inexistant après la saisie directe.

Les traitements manuels, encore très pratiqués, correspondent à une conception déjà ancienne du mode de traitement des enquêtes, mais les prestations demandées et leur insertion dans la chaîne de traitement des enquêtes tendent à évoluer.

Autrefois, jusqu'au début des années 80, il y avait obligatoirement traitements manuels avant saisie, pour retranscrire et aménager l'information dans une forme lisible et logique acceptable par la saisie (les bordereaux de chiffrage), pour apporter les compléments nécessaires (compteurs, totaux, cases de contrôle des quantités, codes)

La recopie a disparu avec l'apparition des questionnaires intégrés et pensés en fonction de la saisie.

Il convient de distinguer deux filières pour le traitement des enquêtes :

- la filière **décentralisée pure** ;
- la filière **mixte** (décentralisée puis centralisée) ;

La filière décentralisée pure :

Dans cette filière, la saisie et les traitements manuels sont tous prévus dans les DR de collecte.

Les traitements manuels peuvent être décomposés en deux temps : avant et après saisie.

Les traitements manuels avant saisie sont de plus en plus allégés et ne portent en principe que sur la codification d'une ou deux variables (PCS des individus par exemple). Les vérifications manuelles destinées à faire un retour rapide aux enquêteurs ou enquêtés n'interviennent que si la saisie n'est pas prête ou pas disponible en temps voulu, pour permettre un listage programmé de ces anomalies.

Il peut y avoir traitements manuels avant saisie pour apporter quelques compléments techniques, précisément destinés à la saisie et qu'il n'était pas naturel de demander à l'enquêteur. Mais ceci est rare dans le décentralisé pur.

De plus en plus souvent, une bonne planification des tâches permet d'enchaîner collecte, saisie et listages d'erreurs et anomalies dans des délais compatibles avec le retour vers l'enquêteur et/ou l'enquêté.

On peut estimer que ce qu'il est urgent de contrôler à ce stade, correspond à ce qui est contrôlé ou géré en direct par la saisie directe, c'est-à-dire la structure, les grands filtres, et les variables sensibles liées au thème central de l'enquête.

Les retours vers les enquêteurs ont aussi des vertus pédagogiques pour leur démontrer que leur travail est suivi et qu'ils ne peuvent faire n'importe quoi, que ces retours fassent suite à un contrôle manuel ou automatique.

Des contrôles automatiques plus poussés qui ne justifient pas un retour à l'enquêteur peuvent avoir lieu dans la même phase de contrôle, mais ils sont généralement opérés dans un deuxième temps.

La filière mixte, décentralisée puis centralisée :

Cette filière mixte est plutôt retenue dans le cas des enquêtes complexes et peu répétitives.

Les travaux réalisés dans la partie décentralisée visent alors uniquement à assurer les vérifications pour lesquelles le retour à l'enquêteur ou l'enquêté est indispensable. Soit manuellement si la saisie est opérée par la partie centralisée, soit par des contrôles automatiques si la saisie est opérée au stade décentralisé.

Dans cette organisation, la DR n'opère aucune codification (à l'exception parfois de la PCS), et n'opère pas de contrôles approfondis.

Si la saisie n'a pas été déjà opérée, la partie centralisée va d'abord apporter les compléments nécessaires à la saisie (compteurs, cases de contrôle des quantités, codifications) et opérer une vérification manuelle globale préalable du questionnaire.

La partie centralisée interviendra également pour participer à la définition des contrôles nécessaires, et pour affiner les nomenclatures.

On trouve dans les enquêtes complexes de nombreux libellés notés en clair par l'enquêteur ou le répondant, libellés qui correspondent généralement au thème central de l'enquête, et qu'il faut codifier : produits alimentaires pour l'enquête Consommation Alimentaire, maladies pour l'enquête Santé, produits achetés et dépenses effectuées pour l'enquête Budget de Famille, motifs des déplacements pour l'enquête Transport, etc.

Ces codifications, très spécifiques, basées sur des nomenclatures pas toujours fermées au départ et qui se complètent progressivement, sont généralement traitées de manière centralisée et ne bénéficient pas d'une base toute prête pour une codification automatique, précisément parce que la nomenclature n'est pas figée ou parce qu'elle a évolué depuis la dernière enquête du même type. C'est un atout pour le statisticien que de pouvoir affiner sa nomenclature à partir de l'observation de ce qui a été collecté. Ces codifications sont faites manuellement, à l'exception des produits alimentaires de l'enquête Consommation Alimentaire qui a une fréquence telle (tous les deux ans) que sa nomenclature évolue peu d'une enquête à la suivante. Les autres grandes enquêtes ont

une périodicité de l'ordre de quatre à sept ans qui fait que la nomenclature de l'enquête précédente, même si elle sert de base, est davantage remise en cause.

C'est généralement la présence de ces codifications qui justifie le recours à la partie centralisée, la centralisation permettant une meilleure interactivité avec le statisticien et une meilleure surveillance de l'homogénéité.

Pour ce qui concerne les contrôles, la correction des erreurs et anomalies détectées, lorsqu'elle a lieu, car elle n'est pas toujours possible ou pas toujours justifiée, est basée sur la compréhension approfondie du questionnaire en cause, par recoupements, par mise en cohérence. Ceci requiert beaucoup de soin pour ne pas "lisser" à tout prix les anomalies qui peuvent correspondre à une réalité. Ceci est réalisé de manière itérative jusqu'à obtention d'un fichier "propre", dans lequel ne subsiste plus aucune erreur, mais où peuvent se trouver des anomalies expliquées et confirmées.

Bien que par commodité de langage on parle d'**apurement** dès lors que l'on procède à la correction du fichier magnétique, il ne faut pas confondre l'apurement proprement dit qui consiste à modifier le fichier magnétique et la **correction** qui consiste à rechercher et indiquer quelle correction il convient d'opérer. La correction, hormis les cas de saisie-assistée, intervient généralement sur papier, soit sur le document d'origine (*en veillant à ne pas faire "écraser" l'information d'origine*), soit sur le listing d'anomalies et d'erreurs. L'apurement intervient après la correction.

L'apurement est généralement réalisé par les gestionnaires eux-mêmes, et concerne, soit les corrections d'erreurs, soit l'insertion des codes si la codification intervient après saisie, les zones correspondantes ayant été saisies "à blanc", soit la codification des rejets de la codification automatique.

L'apurement peut être opéré par les opérateurs de saisie, d'après les corrections faites par les gestionnaires, si cet apurement a lieu sur le matériel de saisie.

Les facteurs **délais**, **moyens** (disponibilité des moyens humains et matériels), **complexité** (des différentes phases), **qualité** (globale et donc qualité de chaque phase), interviennent sur les choix d'organisation, sur ce que l'on fait ou ce que l'on ne fait pas à chaque stade, dans quel ordre on le fait, qui le fait, avec quels outils.

À terme, les traitements manuels assistés par informatique ont vocation à disparaître au profit d'un poste de saisie assistée intégrant tous les traitements post-collecte.

Tout ceci correspond à un schéma relativement standard, mais ce schéma peut être morcelé. On peut, par exemple, exploiter une partie d'enquête (la première ou les deux premières vagues d'une enquête à vague) pour commencer à se faire une idée sur ce que donne l'enquête, détecter de macro-errements ...

Quoi qu'il en soit, réaliser une enquête est un processus long et complexe. Même si l'intégration de plus en plus poussée des phases (et la saisie portable représente actuellement le stade le plus intégré), permet de gagner sur les délais entre collecte et publication, ce qui est tout à fait essentiel pour la "fraîcheur" de l'information, il n'empêche que les délais hors-tout de réalisation d'une enquête sont très longs, de l'ordre de trois à quatre ans. Bien évidemment, plus l'enquête est répétitive, plus il y a amortissement et réutilisation des phases de préparation. Mais les enquêtes les plus répétitives sont aussi les plus simples...