

NATIONS UNIES

COMMISSION ECONOMIQUE
POUR L'EUROPE

ОБЪЕДИНЕННЫЕ НАЦИИ

ЭКОНОМИЧЕСКАЯ КОМИССИЯ
ДЛЯ ЕВРОПЫ

UNITED NATIONS

ECONOMIC COMMISSION
FOR EUROPE

SEMINAIRE

СЕМИНАР

SEMINAR

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE POUR L'EUROPE



CES/SEM.35/SV/R.1 */
19 juillet 1996

CONFÉRENCE DES STATISTICIENS EUROPÉENS

FRANCAIS

Original: FRANCAIS et ANGLAIS

Séminaire sur les Statistiques Officielles - Passé et Futur
(Lisbonne, Portugal, 25-27 septembre 1996)

SESSION V: NOTRE LEGAT POUR LES GENERATIONS A VENIR

**DOCUMENTATION ELECTRONIQUE DES OUTILS STATISTIQUES:
UNE EXPERIENCE FRANCO-ROUMAINE**

Rapport présenté par l'Institut national de la statistique et
des études économiques (INSEE) (France) et
la Commission Nationale de statistiques (Roumanie) **/

1. La préoccupation de l'Institut National de la Statistique et des Etudes Economiques (**INSEE**) pour la documentation de ses opérations statistiques est née il y a une quinzaine d'années, dans un double souci de transmission du savoir-faire à l'intérieur de l'Institut et de recherche de cohérence dans le traitement des données. M. Edmond MALINVAUD, alors Directeur Général de l'**INSEE**, décida d'initialiser les travaux sur ce sujet, en confiant la responsabilité à M. Claude POULAIN, dont les recherches conduisirent à la mise au point d'un instrument qui reçut le nom de **DDS (Dictionnaire de Données Statistiques)**.

*/ Les services des Nations Unies à Genève ne peuvent garantir la distribution de ce document avant la date du Séminaire du fait qu'il a été soumis au-delà requis de 10 semaines.

**/ Préparé par Dominique Crosnier, (INSEE) France et Liana Marina, CNS, Roumanie.

GE.96-

2. Cet instrument est à considérer sous un double aspect: celui d'une méthode d'organisation de la documentation d'opérations statistiques et celui d'outils informatiques en assurant la mise en oeuvre concrète. Dans un premier temps, cet instrument a été dédié aux enquêtes statistiques, type d'opération essentiel pour l'INSEE. Dans ce contexte, le point de vue des

producteurs était privilégié, puisqu'il s'agissait de formaliser en premier lieu la documentation utilisable dans la gestion de ce type d'opération. Un modèle conceptuel a été conçu à cette intention (**modèle Enquête**) et un outil informatique lui a été associé (Enquête). Néanmoins, les diffuseurs firent assez vite valoir leurs propres besoins, ce qui a conduit à la définition d'un autre modèle conceptuel destiné à représenter une documentation simplifiée des opérations statistiques depuis leurs sources jusqu'à leurs modes de diffusion, un autre outil informatique lui étant associé : ISIS (Inventaire du Système d'Information Statistique). Enfin, une première tentative de généralisation de l'outil Enquête fut mise en chantier (Texo). Il permet, dans certaines limites, d'adapter le "modèle Enquête" à d'autres situations. En pratique, il a servi à la construction de bases de recherche documentaire.

3. Plus récemment, depuis 1992, dans le cadre d'un programme de coopération entre l' **INSEE** et la Commission Nationale pour la Statistique de Roumanie (**CNS**), cette dernière a exprimé son intérêt à la fois pour cette idée de documentation des opérations statistiques et pour l'instrument développé par l' **INSEE**. Une action de coopération entre les deux Instituts a donc été entreprise à ce sujet. Elle a été financée par l'Union européenne au travers du programme **PHARE**.

4. Dans un premier temps, la **CNS** a pu s'initier aux outils mis en place par l' **INSEE**, à l'aide de l'implantation à Bucarest de l'outil E **NQUÊTE** dans une version micro-informatique. Ensuite, elle a procédé à certains développements à partir des outils existants (1994). Aujourd'hui, elle a décidé de poursuivre la collaboration avec l' **INSEE** pour la mise en oeuvre des évolutions de cet instrument.

5. Ce document se propose de procéder à un bilan de l'ensemble de cette opération, en traitant les trois points suivants:

1. Une présentation de l'instrument **DDS**. On ne traitera que du "modèle Enquête" et de l'outil Enquête qui lui est associé. En effet, il est d'une part inutile d'envisager les différents modèles et outils existants pour exposer les principes du **DDS**, et d'autre part seul l'outil E **NQUÊTE** est disponible à la fois à l' **INSEE** et à la **CNS**.
2. Les réalisations faites à l' **INSEE** et à la **CNS** à l'aide de cet outil.
3. Les évolutions à prévoir à la lumière de ces expériences concrètes.

I. LE DDS SOUS LA FORME DU "MODELE ENQUETE"

6. Sous la forme du "modèle Enquête", le **DDS** est d'abord une méthode d'organisation de la documentation d'enquêtes statistiques. Il a aussi une traduction opératoire sous forme d'un outil informatique.

Le DDS méthode de documentation

7. C'est une méthode, dans la mesure où il s'appuie sur un modèle de conceptualisation des métadonnées (Cf. **Annexe 1**), à l'aide duquel on organise la documentation d'une enquête statistique de la manière suivante:

- Définition des " objets " appelés " **type de fiche** " qui permettent la description et le classement des différents éléments concourant à la mise en oeuvre d'une opération statistique dans ses phases de conception, de réalisation, d'exploitation et de diffusion.
- Pose des liens que ces différents éléments entretiennent entre eux.

Les objectifs de cette méthode sont multiples :

- Constituer et utiliser la documentation dans le processus d'élaboration de l'enquête statistique. Il s'agit en particulier d'homogénéiser la diffusion de l'information entre les différents intervenants.
- Stocker la mémoire d'une enquête sur un "lieu informatique unique".
- Utiliser cette mémoire pour informer, pour concevoir une nouvelle opération, pour archiver, ou pour diffuser les métadonnées.

Les types de fiche du modèle Enquête

8. Le "modèle Enquête" définit dix " **types de fiche**" principaux:

- | | |
|-----------------|----------------|
| . variable | . incident |
| . questionnaire | . programme |
| . question | . nomenclature |
| . concept | . tableau |
| . spécification | . fichier |

9. Au centre du modèle, se trouve le type de fiche " **variable statistique**". Tous les autres, qu'ils soient en amont ou en aval, entretiennent une relation étroite avec lui.

En amont:

- Le **questionnaire**, document de référence, et les **questions** qui sont le plus souvent à l'origine des variables.
- Les **concepts** qui fondent et explicitent la définition d'une variable.
- Les notes de **spécification** qui précisent les procédures de collecte et/ou les algorithmes de traitement de l'information.

En aval :

- Les **incidents** où l'on pourra stocker tous les commentaires liés à l'ensemble du traitement, qu'ils se soient produits au cours de la collecte ou à un autre moment.
- Les textes des **programmes** informatiques de contrôle, de redressement, de génération des variables calculées.
- La composition des **tableaux** utilisés dans la diffusion des résultats.
- La description des **fichiers** informatiques construits tout au long de l'opération.
- Les **nomenclatures** utilisées pour classer les unités observées par l'opération.

Les champs des types de fiche

10. Chaque type de fiche est décrit par un ensemble de propriétés. Certaines sont communes à tous les types de fiche, d'autres sont spécifiques à certains d'entre eux. Cet ensemble est constitutif du type de fiche. L'**Annexe 2** donne la liste de ces propriétés appelées "champs".

Les liens

11. A priori, des liens réciproques et de plusieurs types peuvent exister entre tous les types de fiche. Dans la forme applicative du modèle, on a choisi de simplifier cet état théorique, en ne retenant qu'un nombre limité de possibilités de liens (Cf. **Annexe 1**) et en les hiérarchisant de telle façon qu'on puisse toujours établir un lien entre deux types de fiche quelconques, même s'il est indirect et transite par un troisième type de fiche. Il existera donc des liens "amont" et des liens "aval", des liens "directs" et des liens " indirects " entre les différents types de fiche.

Le mode de stockage de la documentation

12. A partir de cette structuration de l'information, tout document textuel pourra être enregistré dans une **fiche**, qui sera l'élément unitaire de la base. Une fiche est constituée de trois parties: Une première partie décrit, par la valeur de ses propriétés, la fiche à laquelle correspond le document textuel. La deuxième partie d'une fiche est constituée des documents textuels eux-mêmes: contenu d'une note, libellé d'une question, définition d'un concept, source ou référence d'un programme informatique, tableau de modalités d'une variable, relation d'un incident, source d'une demande de tableau, ensemble des postes d'une nomenclature, descriptif d'un fichier. Enfin dans la troisième partie, on décrit les liens qu'une fiche déterminée entretient avec d'autres fiches de la base. L' **Annexe 3** donne un exemple d'une fiche du type variable.

Le DDS outil informatique

Les aspects techniques

13. L'outil Enquête existe en trois versions :
- Une version sous MVS dans laquelle l'éditeur de texte est celui d'Ispf, le SGBD est Oracle. Elle a été développée avec la version 5 d'Oracle, puis maintenue dans les versions 6 et 7. Elle utilise le logiciel de composition de documents Dcf et intègre une interface avec SAS. Cette version n'est implantée qu'à l' **INSEE**.
 - Une version micro monoposte dans laquelle l'éditeur de texte est WORD 5.5, le SGBD Oracle. Le développement a été fait avec la version 5 d'Oracle. Cette version ne contient pas l'interface avec SAS. Cette version est implantée à l' **INSEE** et à la **CNS**.
 - Une version micro réseau local sous machine UNIX, dans laquelle l'éditeur de texte est Word 2.0, le SGBD Oracle. Elle a été développée avec la version 6 d'ORACLE, elle ne contient pas non plus l'interface avec SAS. Ce développement a été réalisé en collaboration entre la **CNS** et l' **INSEE**.

L'organisation

14. Il existe trois types d'intervenants:
- **L'administrateur "Oracle" de la base**
Il est le seul habilité à ouvrir une base , à enregistrer les gestionnaires, à effectuer des sauvegardes de l'ensemble d'une base (contenu et éléments Oracle).
 - **Les gestionnaires**
Ils sont déclarés, par un nom symbolique, auprès de l'administrateur de la base et peuvent effectuer toutes les opérations de gestion sur la base, y compris les sauvegardes au sens **DDS**, c'est à dire seulement le contenu d'une base.
 - **Les consultants**
Ils ne sont pas spécifiquement déclarés auprès de l'administrateur d'une base. Il y ont accès pour consulter ou récupérer l'information qui s'y trouve. En aucun cas ils ne peuvent modifier le contenu de la base. Toute personne possédant un identifiant informatique est un consultant potentiel.

Les principales fonctionnalités

15. Selon les cas, elles sont disponibles pour tous, ou seulement pour les gestionnaires.

Opération	Intervenants
Gestion	Gestionnaires
Interrogation de la base	Gestionnaires et consultants
Edition - récupération	Gestionnaires et consultants
Livre des codes	Gestionnaires
Interface avec SAS	Gestionnaires

Gestion

Chargement de fiches peut s'effectuer en batch ou en mode interactif, aussi bien pour l'introduction de fiches que pour leur modification. Un bilan de cohérence permet de vérifier si cette introduction a été faite conformément aux règles de gestion de l'outil. L'introduction des textes (ou leur modification), peut se faire directement dans la base Oracle, ou en passant par l'éditeur de texte (Ispf, Word 5.5 ou Word 2.0 selon les cas). Cette option est particulièrement utile quand on récupère le texte sous forme de fichier informatique.

16. Les principales autres fonctions de gestion sont les suivantes:

- Suppression de fiches;
- Ajout de liens, modification de l'ordre des liens;
- Changement du nom d'une fiche;
- Introduction ou/et correction de mots-clés;
- Archivages (de fiches, historique).

17. L'archivage de fiches peut revêtir un caractère assez important. En effet, on peut décider, à un moment de l'élaboration de la documentation d'une enquête statistique, qu'il n'est plus utile de conserver un document textuel dans sa relation complète au reste de la documentation (valeur des champs, liens), mais que l'on a cependant intérêt à en garder la trace. Par exemple, il peut s'agir d'une version intermédiaire d'une note de spécification. On a alors la possibilité de le conserver dans la base en le versant en archive où il sera stocké sans ses liens ni ses valeurs de champs autres que son titre. L'archivage historique, quant à lui, peut être réalisé sur support informatique en déchargeant le contenu d'une base triée par type de fiche.

Interrogation de la base

18. Elle peut se faire de plusieurs manières :

- sélection sur les noms des fiches (quand ils sont connus);
- sélection sur les types de fiche;
- sélection sur les différents champs des types de fiche. Selon les cas, on aura des choix fermés (liste de modalités d'un critère) ou des choix ouverts (sélection sur une chaîne de caractères), par exemple sur le nom d'un auteur ou sur des plages de valeurs (dates);
- sélection sur des chaînes de caractères contenues dans le document textuel;
- sélection à partir des liens entre les fiches;

- sélection sur des mots-clés.

L'édition - récupération

19. Il existe deux états standards principaux: l'édition de fiches ou l'édition de texte. Dans le cas de l'édition de fiches, on obtient le contenu (la valeur des différents critères) correspondant à la partie descriptive de la fiche, ainsi que le texte. Cette édition n'existe que sur papier. L'édition du seul texte lié à une fiche peut se faire dans un fichier ou sous une forme papier. On peut récupérer des fiches, en partie ou en totalité, soit pour faire des mises à jour, soit pour les transférer dans une autre base **DDS**.

Livre des codes

20. Une sortie standard permet d'éditer dans un fichier, puis sur papier, la liste des variables d'une enquête avec:

- les noms symboliques des variables;
- la partie texte de la fiche variable, comprenant les titres développés des variables;
- la liste des modalités associées et des commentaires pouvant inclure tout type de référence, voire des tables de passage.

21. Si, en plus, l'enquête est exploitée à l'aide du logiciel SAS, on peut, dans la version sous MVS, y adjoindre les effectifs et les fréquences correspondants pour chacune des variables. On peut également regrouper les variables par thème.

L'interface avec SAS

22. En dehors de la possibilité précédente d'introduction automatique des distributions dans le livre des codes, et toujours dans le cas de la version sous MVS, l'interface avec SAS permet d'effectuer certains contrôles, dont deux sont particulièrement utiles. Exhaustivité - Ce contrôle permet de vérifier que toutes les variables d'une table SAS sont aussi décrites dans le **DDS** et que toutes les variables déclarées dans le **DDS** existent bien dans une table SAS.

23. Type - Ce contrôle permet de vérifier que les valeurs prises par une variable appartiennent bien à une liste de valeurs permises. En particulier on pourra éditer, à l'aide d'une PROC FREQ classique, les valeurs trouvées pour une variable dans une table SAS, alors qu'elles ne sont pas prévues dans le **DDS**.

II. LES REALISATIONS

Les réalisations françaises

24. L'outil ENQUÊTE étant dédié en priorité aux enquêtes statistiques et plus particulièrement aux enquêtes auprès des ménages, c'est dans ce domaine qu'il y a eu le plus de réalisations. Comme à ce jour, il n'existe à l' **INSEE**

aucune obligation de constitution d'une base de documentation, et qu'en conséquence ce travail s'effectue sous la forme du volontariat, toutes les opérations statistiques de l' **INSEE** n'en sont pas munies. Néanmoins, la plupart des enquêtes "régulières" auprès des ménages ont constitué, à un moment ou à un autre, une base **DDS**, l'exception notable étant le Recensement de la population. En revanche, l'utilisation de ce genre de méthode reste embryonnaire dans le domaine de la statistique d'entreprise et dans celui des enquêtes régionales. Par ailleurs, quelques tentatives encore timides sont en cours de réalisation pour des opérations dont la source statistique est différente d'une enquête "classique". C'est le cas, par exemple, de l'indice des prix à la consommation.

25. Il convient toutefois de noter qu'à chaque fois qu'une opération a été menée sous la forme d'un partenariat, et que l' **INSEE** s'était engagé à en faire la documentation, alors les unités en charge de ces opérations à l'**INSEE** ont utilisé des bases **DDS**, afin de mettre à la disposition des différents partenaires une documentation homogène et cohérente (Enquêtes Transport, Santé, Mobilité et Insertion Sociale, Loisirs).

26. En dehors des réalisations effectuées à l'aide de l'outil Enquête, deux autres bases méritent d'être citées: la base documentaire de Comptabilité Nationale, faite avec avec l'outil Texo, et la base ISIS.

27. Pour préciser les idées, on citera les principales bases existantes et l'essentiel de leur contenu.

Les principales bases existantes

28. Dans le domaine des enquêtes auprès des ménages:

- La série des enquêtes **Emploi** 1980 et 1990
- La série des années 80 contient 739 fiches dont 569 variables et 28 fichiers.
- La série des années 90 contient 660 fiches dont 498 variables, 9 concepts et 13 nomenclatures.
- Les enquêtes **Formation** et **Qualification Professionnelle**
- La base de l'enquête 1985 contient 1783 fiches dont 549 variables et 248 programmes.
- La base de l'enquête 1992 contient 2024 fiches dont 1180 variables, 124 programmes et 447 tableaux.
- L'enquête **Logement**
- La base de l'enquête 1988 contient 929 fiches dont 559 variables, 119 tableaux et 6 nomenclatures.

- La base de l'enquête 1992 contient 1024 fiches dont 828 variables et 33 spécifications
- L'enquête **Actifs Financiers**
- Elle contient 3819 fiches dont 2846 variables et 43 fichiers.
- L'enquête **Santé** 1991
- Elle contient 1516 fiches dont 1033 variables, 111 spécifications et 18 fichiers.
- L'enquête **Transports** 1993
- Elle contient 1388 fiches dont 993 variables et 12 nomenclatures.
- L'enquête **Mobilité et Insertion Sociale** de 1995 a donné lieu à la création de deux bases distinctes selon les populations concernées.
- La première contient 1332 fiches dont 821 variables, 37 spécifications, 46 programmes et 6 nomenclatures.
- La seconde contient 1139 fiches dont 656 variables, 108 programmes et 7 nomenclatures.

29. Dans le domaine de la statistique d'entreprise:

- L'enquête annuelle **Commerce**
- Elle contient 3056 fiches dont 726 variables.

Avec l'outil ISIS

- La base **ISIS** (Inventaire du Système d'Information Statistique)
- Elle contient 13541 fiches qui permettent de recenser 215 enquêtes et 9886 questions, 1033 publications et la définition de 565 concepts.

Avec l'outil Texo

- La base documentaire de **Comptabilité Nationale**
- Elle contient 3472 fiches qui permettent de stocker 107 publications et 700 résumés de notes.

Pour les bases qui utilisent l'outil Enquête, on remarquera qu'elles sont, dans une grande proportion, composées des définitions des variables issues des enquêtes, reliées bien sûr aux questions et questionnaires correspondants. Pour le reste, la documentation composant les bases est fonction des besoins de chacun. Certaines d'entre elles contiennent beaucoup

de spécifications, d'autres peu. En revanche, les textes des programmes informatiques de traitement de l'information ne figurent pas toujours, mais surtout un déficit important est à constater en matière de concepts.

L'utilisation des bases DDS

30. Quand il existe des bases **DDS** correspondant à des versions successives d'une même enquête, alors les informations contenues dans les premières ont été utilisées pour la version la plus récente, au stade de sa conception. Pour des enquêtes dont il n'y a eu jusqu'à présent qu'une seule édition documentée, alors deux fonctions principales ont été utilisées. D'une part l'archivage et d'autre part l'édition du livre des codes à l'aide de l'interface SAS (Cf **Annexe 4**). Pour les enquêtes faites en partenariat, c'est toute la documentation qui a été mise à la disposition des différents partenaires.

Les réalisations roumaines

31. En Roumanie, la première application concrète de documentation d'une opération statistique a été faite sur l'enquête annuelle **AMIGO** (enquête Emploi roumaine), pour les années 1994 et 1995. Cette base contient 238 fiches dont 90 variables, 57 spécifications, 3 nomenclatures et 22 concepts. Elle a donné lieu à plusieurs utilisations: Toute la documentation de l'enquête annuelle a pu être utilisée pour la gestion de l'enquête emploi trimestrielle qui existe depuis 1996, les deux opérations ayant beaucoup de caractéristiques en commun. La base de l'enquête **AMIGO** elle-même a permis de confectionner un guide d'instructions pour les enquêteurs, d'éditer le livre des codes (Cf **Annexe 5**), ainsi que les spécifications pour les programmeurs (conditions de contrôle entre les variables). En outre, elle a permis de rationaliser l'enquête annuelle 1996. Enfin, la base étant particulièrement riche en matière de concepts, elle a permis de répondre à une demande extérieure en la matière, accompagnant les données issues de l'enquête. Cette demande a été exprimée par l'Institut de Prévision, l'Institut pour la qualité de la vie, l'Institut de Recherches économiques, les Ministères du Travail et des Finances.

32. L'outil Enquête a également été employé pour documenter des indicateurs macroéconomiques. Pour s'adapter à la situation du "modèle Enquête", on les a traité comme des variables. De cette façon, on a pu documenter leurs définitions. Néanmoins, le modèle a dû être "forcé", dans la mesure où il ne répond pas complètement au problème à traiter.

33. Ainsi, l'outil Enquête autorise un lien et un seul entre une variable et une nomenclature. Or, dans le cas présent, il était nécessaire d'attacher plusieurs nomenclatures à un même indicateur. Pour pallier cet inconvénient, les nomenclatures ont été introduites en commentaire des textes des variables, ce qui n'est évidemment pas satisfaisant, car il n'est a priori pas "naturel" d'aller rechercher des nomenclatures dans les commentaires du texte d'une variable.

Les difficultés rencontrées

34. Deux grandes difficultés ont été rencontrés dans la pratique de la confection de bases de documentation. La première est parfaitement illustrée par l'exemple précédent. Le caractère contraignant des modèles conceptuels qui sous-tendent les outils existants rendent leur utilisation délicate pour des opérations dont les caractéristiques diffèrent sensiblement de celles attachées aux trois modèles. Même en restant dans le domaine des enquêtes, c'est déjà vrai pour celles qui sont faites auprès des entreprises. A fortiori, le phénomène devient encore plus patent dès que l'on s'intéresse à des opérations dont la source n'est pas une enquête classique : indices, indicateurs, comptes, fichiers d'origine administrative, etc... "Forcer" l'un quelconque des trois modèles pour adapter l'organisation de la documentation d'une opération peut conduire à des incohérences dans le contenu des bases, comme de traiter des nomenclatures en commentaire du texte d'une variable. En passant d'une base à une autre, il n'existe plus de garantie d'avoir un traitement cohérent de l'information qui assure une recherche efficace.

35. La seconde grande difficulté est d'une autre nature. Elle est liée à la charge de travail qu'impose la mise en place d'une documentation dans un contexte de pénurie relative en moyens humains. Ce poids qui tient aux méthodes de travail et aux caractéristiques techniques des outils actuels peut décourager les concepteurs d'une opération statistique de la documenter, alors que la demande en documentation technique devient de plus en plus importante.

36. Ce double constat impose une réflexion quant aux modifications à apporter à cette procédure pour qu'elle puisse davantage satisfaire les besoins auxquels elle doit répondre. Dans ce document nous n'aborderons que la partie relative aux outils, celle relative aux méthodes de travail étant beaucoup plus spécifique à chaque organisation concernée par cette question.

III. LES PERSPECTIVES

37. L'expérience accumulée à l' **INSEE** depuis plusieurs années ainsi que celle, plus récente, de la **CNS** conduisent à une double conclusion.

38. D'une part, la méthode d'organisation de la documentation telle qu'elle est conçue dans le **DDS** est bien une réponse satisfaisante aux problèmes rencontrés dans ce domaine. D'autre part, l'aspect opérationnel du **DDS** (les outils informatiques) laisse apparaître un certain nombre d'insuffisances, dont l'origine est à la fois fonctionnelle et liée au fait que ces outils, développés il y a maintenant six ans, ont besoin de s'adapter aux progrès de la technologie informatique. Ainsi, ils doivent évoluer dans deux directions. L'une correspond à une amélioration de l'existant. L'autre correspond à la prise en compte de besoins non encore satisfaits, qui s'expriment avec de plus en plus de vivacité. L'un de ces besoins est lié à la nécessité de généraliser la possibilité de documentation des opérations statistiques; l'autre à l'évolution récente des techniques de diffusion. Face à cette situation, la décision a été prise, en commun accord entre l'**INSEE** et la **CNS**, de procéder à une rénovation des outils. Elle est en cours de développement et devrait aboutir à la fin de l'année 1996.

L'amélioration de l'existant

39. Deux points sont à considérer:

- les contraintes propres aux outils;
- l'homogénéité entre les outils;

Les contraintes propres aux outils

40. Les principales critiques, exprimées par les utilisateurs, sur les outils existants concernent les éléments suivants: le traitement des documents textuels, la circulation dans les bases, la gestion des liens, l'ergonomie. On ne traitera pas ici tous ces points en détail, et on se contentera d'en donner les caractéristiques principales:

Le traitement des documents textuels

41. Dans la configuration actuelle, les documents textuels introduits dans une base **DDS** le sont avec une triple contrainte :

- Lignes de 72 caractères maximum.
- Perte de l'enrichissement initial des documents. Toute récupération de texte, pour impression à partir d'une base **DDS**, exige que l'on refasse cet enrichissement pour obtenir une édition soignée.
- Aucun autre type de document que du texte ne peut être stocké dans une base **DDS**.

42. L'amélioration consiste donc à permettre que les textes puissent être stockés, dans une base quelconque, sans contrainte sur la longueur des lignes, que l'enrichissement soit conservé et qu'on puisse introduire d'autres types de documents (graphiques, organigrammes...).

La circulation dans les bases

43. Dans la configuration actuelle, la sélection d'un sous-ensemble de fiches peut s'avérer assez complexe, en termes de manipulation des outils informatiques. Ceci peut être d'autant plus décourageant qu'on les utilise de façon irrégulière, alors que ce mode d'interrogation d'une base **DDS** est bien évidemment fondamental. La rénovation doit donc permettre, à l'aide des interfaces graphiques et des moyens de circulation offerts par la technologie sous **WINDOWS**, que le mode d'interrogation d'une base soit simplifié et que la récupération des textes soit plus aisée.

La gestion des liens

44. L'un des intérêts majeurs d'une base documentaire construite avec un **SGBD** est de pouvoir introduire des liens entre les fiches. On a ainsi toute chance de pouvoir spécifier précisément la description de chaque fiche et de pouvoir circuler dans l'ensemble de la documentation par l'intermédiaire de ces liens. Dans la configuration actuelle des outils, l'introduction des liens nécessite la saisie manuelle et la mémorisation des noms de fiches, ce

qui peut décourager les utilisateurs de le faire et conduire à appauvrir les bases. L'amélioration consiste à rendre plus facile le mode d'introduction des liens, voire à l'automatiser quand cela est possible.

L'ergonomie

45. De façon générale, l'ergonomie des outils est liée à la technologie de l'époque à laquelle ils ont été développés. L'amélioration vise à recourir aux possibilités offertes par la technologie récente en matière de convivialité des outils informatiques. Aide en ligne, intégration de l'outil au poste de travail du statisticien, communication plus adaptée avec d'autres outils de gestion des opérations.

L'homogénéité entre les outils et la communication entre les bases

46. A l'heure actuelle, on ne dispose pas d'un outil unique, quel que soit le modèle conceptuel considéré. Cette situation a les conséquences suivantes:

- Les fonctionnalités ne sont pas parfaitement homogènes entre les différents outils.
- D'un outil à l'autre, la communication ne s'effectue pas de façon simple. Par exemple, on ne peut pas recopier directement des informations qui figurent dans une base de type Enquête pour alimenter ISIS et réciproquement, alors que cela constituerait à la fois un gain de temps et de cohérence. On peut, par exemple, concevoir qu'ISIS contienne une information riche en matière de définition de concepts, et qu'elle soit facilement récupérable dans toute base Enquête particulière.
- Les différentes versions de l'outil Enquête ne sont pas cohérentes entre elles. En dehors de versions différentes du traitement de texte utilisé, l'interface SAS, à partir duquel on peut en particulier éditer les distributions des variables dans le dictionnaire des codes, n'existe que dans la version sous MVS.

47. Le travail d'amélioration consiste donc à développer **un outil unique**, aux fonctionnalités homogènes, quel que soit le modèle conceptuel utilisé.

Extension des capacités de l'outil

Adaptabilité du modèle

48. L'outil ENQUÊTE est fondé sur un modèle conceptuel lié explicitement à un type d'opération statistique dont la source est une enquête et même plus particulièrement du type enquête auprès des ménages. De la même manière, ISIS et TEXO sont sous-tendus par des modèles conceptuels adaptés à leurs objets de départ. Tous les cas de figure pouvant intéresser les statisticiens ne correspondent pas nécessairement à ces schémas fixes. Une opération statistique dont la source est d'origine administrative, une opération statistique en continu comme l'indice des prix à la consommation, des données de synthèse comme celles de la Comptabilité Nationale, le traitement de

séries chronologiques comme celles des Comptes Trimestriels... nécessitent un modèle conceptuel différent de ceux existants dans les outils actuels.

49. La documentation des séries chronologiques peut illustrer ce propos. En effet, quand on travaille sur des séries infra-annuelles, un élément essentiel de leur traitement réside dans les corrections statistiques qui sont apportées aux données brutes. Il s'agit, par exemple, des corrections pour variations saisonnières et des corrections pour effets de jours ouvrables. Dans ces cas, aucun des types de fiche des modèles existants n'est réellement adapté à la prise en compte de l'information utile à la compréhension des méthodes utilisées dans ces traitements. Il faut donc imaginer un nouveau type de fiche, relié au type de fiche variable dont l'intérêt demeure. On pourrait l'intituler "traitement" et il serait caractérisé par des "champs" qui fourniraient les informations suivantes:

Champs	Exemples de valeurs des champs
Nature du traitement	CVS, CVS-CJO, CJO-CVS...
Fréquence du traitement CVS	mois, trimestre, année...
Fréquence du traitement CJO	"
Composition du traitement CVS	additif, multiplicatif
Composition du traitement CJO	"
Type d'ajustement retenu pour les CJO	jours ouvrables, jours non ouvrables...
Date de début du calcul des CVS	
Date de fin du calcul des CVS	
Date de début du calcul des CJO	
Date de fin du calcul des CJO	
Date de début d'application des coefficients correcteurs	
Date de fin d'application des coefficients correcteurs	
Logiciel utilisé pour le traitement des CVS	
Logiciel utilisé pour le traitement des CJO	
Définition des points aberrants	

50. Les textes de ce type de fiche permettraient de fournir le détail des éléments d'explication méthodologique complémentaires: définition des points aberrants, choix du type de traitement des CJO, choix de l'ordre d'application des corrections...Ce qui est vrai de cet exemple l'est de beaucoup d'autres. Il est donc indispensable, si l'on souhaite pouvoir répondre aux différentes demandes, que l'outil unique soit adaptable aux différentes situations.

51. Dans ce contexte, l'organisation des bases devrait être modifiée. Le rôle de l'administrateur des bases se trouve sensiblement renforcé par rapport à la situation actuelle. En effet, il sera de sa responsabilité de mettre en œuvre, après discussion avec les responsables de l'opération statistique, la traduction technique des spécificités du modèle choisi, dans le cadre initial de la méthode d'organisation de la documentation. Il sera aussi de son rôle de faire en sorte que ces spécificités soient cohérentes d'une base à une autre. Il s'agit ainsi de préserver la possibilité

d'échanges et de récupération d'informations d'une base à une autre, des informations de même type se retrouvant selon un descriptif identique. Si l'on cherche des notes d'instruction à des enquêteurs, par exemple, cela doit apparaître, quelle que soit la base, dans un même type de fiche avec des valeurs de champ identiques. Le rôle des gestionnaires d'une base n'est pas modifié. Ils ont les mêmes attributions qu'antérieurement.

52. La possibilité d'extension à différents types d'opérations statistiques peut engendrer un besoin de confidentialité relative de certaines bases qui peuvent être "privées". En conséquence, les consultants sont partagés en deux catégories: privilégié ou ordinaire. Dans les deux cas, ils ne peuvent procéder qu'à des opérations de consultation et/ou de récupération dans une base.

- un consultant privilégié est toujours déclaré auprès de l'administrateur et il a accès à l'information contenue dans toutes les fiches.
- un consultant ordinaire ne peut accéder qu'aux fiches "publiques". Il ne peut le faire dans une base réputée "privée" qu'après avoir été déclaré auprès de l'administrateur de la base.

Portabilité de l'outil

53. L'adaptabilité du modèle n'est pas la seule voie que doit emprunter une rénovation de l'outil pour satisfaire au besoin d'extension de son utilisation. Il y a aussi sa "portabilité". Les développements actuels ont été faits à partir du même SGBD, choisi à la fois par l' **INSEE** et la **CNS**, à savoir Oracle. Ceci limite les possibilités de diffusion de l'instrument **DDS**. En conséquence, l'outil informatique doit être muni de la capacité à fonctionner avec d'autres SGBD (Paradox, Sybase, Informix...), dans les mêmes conditions qu'avec Oracle.

Diffusion des métadonnées

54. La demande, qui est longtemps restée latente, de métadonnées par les utilisateurs des données, devient de plus en plus forte et elle est certainement appelée à connaître un développement important et rapide. Elle concerne au premier chef les diffuseurs. Ils jugent comme idéal de disposer d'une aide en ligne facilement consultable à partir des données elles-mêmes. En conséquence l'outil doit être muni d'une nouvelle fonctionnalité, permettant de fournir la génération automatique de documents hypertexte à partir des informations contenues dans une base **DDS**, les liens hypertexte reproduisant les liens définis dans la base.

Les bénéfices attendus de l'extension des capacités de l'outil

55. De la mise à la disposition des statisticiens d'un outil rénové, il est attendu une meilleure rationalisation de l'ensemble de la production, **quelle que soit la nature de l'opération statistique et quel qu'en soit le lieu d'implantation.**

- Les concepteurs disposeront d'un instrument facilitant la rationalisation des méthodes d'organisation d'une opération statistique (cas d'opérations nouvelles), ainsi que leur transmission, dans les cas où la conservation de cette mémoire est une composante essentielle des conditions de la production (opérations en continu, opérations ponctuelles lourdes et répétées).
- Les gestionnaires disposeront d'une aide facilement accessible aux spécificités de déroulement d'une opération statistique (transmission à des gestionnaires ou à différents intervenants d'une chaîne de production, des éléments de documentation nécessaires à leur activité, soit par accès direct aux bases, soit par l'intermédiaire de fichiers d'aide). Ce besoin est particulièrement important dans les cas où l'opération statistique fait intervenir des méthodes de gestion déconcentrées.
- Des concepteurs ou gestionnaires d'autres opérations statistiques pourront trouver l'information méthodologique sur des opérations dont ils ne sont pas maîtres d'oeuvre mais qui les concernent directement.
- Les utilisateurs des données produites par les statisticiens pourront trouver l'information méthodologique sur les opérations qui sont à l'origine des données qu'ils utilisent.
- Les diffuseurs pourront documenter leurs produits de diffusion quand la demande leur en est faite.
- Les archivistes trouveront les éléments permettant de documenter ce qu'ils souhaitent conserver.

Les moyens mis en oeuvre

56. Ils sont de deux ordres: techniques et humains

Les techniques utilisées

57. La mise en oeuvre d'une évolution de cette nature a tout avantage à recourir à deux techniques complémentaires: **la programmation orientée objet et l'architecture client serveur:**

- Compte tenu des objectifs de la rénovation en matière d'amélioration de l'ergonomie de la gestion d'une base, d'adaptabilité du modèle et de portabilité, il apparaît que les techniques de **la programmation objet** sont tout à fait adaptées au développement d'une telle application. Elle permettent une modélisation des problèmes proche du domaine réel décrit et offrent de grands avantages en termes de réutilisabilité (maintenance facilitée, extensibilité ouverte).

- Par ailleurs, **l'architecture Client/Serveur** permet de répondre aux questions relatives à la souplesse d'implantation des bases, à la communication entre les bases et avec d'autres outils de gestion, ce qui permet l'intégration de l'application au poste de travail du statisticien.

Dans ce contexte, l'outil de développement retenu a été DELPHI (de Borland), qui concilie les avantages conceptuels de la programmation objet avec la souplesse du client/serveur concernant l'implantation des données.

Les moyens humains

58. Cette entreprise de rénovation de l'outil **DDS** se réalise sur la base d'une collaboration étroite entre les deux Instituts (**INSEE** et **CNS**). Conjointement, et à la suite du succès de leurs premières expériences, ils ont décidé de mettre à profit leurs compétences, à la fois fonctionnelles et techniques, pour travailler en commun à l'élaboration de ce nouveau produit, la répartition des tâches s'étant faite en fonction des compétences de chacun.

CONCLUSION

59. La rénovation des outils informatiques de l'instrument **DDS** est une condition nécessaire au développement de la constitution de bases de documentation des opérations statistiques. Pour autant, elle n'est pas une condition suffisante. En effet, la volonté de documenter les opérations statistiques ainsi que la mise en place des méthodes de travail nécessaires à cet objet sont des conditions indispensables pour assurer ce développement. Or, elles sont du domaine de l'organisation des Institutions concernées et non plus de celui de la technique pure.

60. On peut cependant penser que cette volonté va aller en s'amplifiant dans un avenir proche, tant le besoin de connaître les méthodologies d'élaboration des données statistiques s'accroît. Dès lors, un instrument rénové devrait fournir aux producteurs de ces données un outil de travail satisfaisant.