

# La place de l'aide à la décision dans le système d'information

Michel Volle

Délégué général du  
« Club des Maîtres d'Ouvrage des Systèmes d'Information »

[michel.volle@polytechnique.org](mailto:michel.volle@polytechnique.org)

**Résumé :** Cet article décrit comment un système d'aide à la décision fonctionne : alimentation, traitements, publication, outils de consultation. Il débute par une étude de cas qui montre les difficultés intellectuelles et « politiques » que comporte sa conception.

**Summary :** This paper presents how a Decision Aid System works: input, data processing, output and publication, available consultation tools, publishing environment. It begins with a case study underlining the intellectual and “political” aspects of its engineering.

**Mots clés :** Aide à la décision, Tableau de bord, Système d'information

\* \*

Les systèmes d'aide à la décision (SIAD) peuvent être classés en trois catégories qui se distinguent par l'échelle de temps comme par l'envergure de la décision :

- le *SIAD opérationnel* évite la surcharge mentale de l'opérateur en lui proposant des solutions permettant de faire face rapidement à des situations complexes (exemple : le SIAD du « quart opération » d'un transporteur aérien qui propose, en cas d'incident, des solutions pour le reroutage ou le remplacement des avions et des équipages). Cet automate, qui relève des systèmes experts, n'équipe que ceux des opérateurs qui peuvent avoir à résoudre des problèmes très difficiles sous la contrainte de l'urgence.
- le *SIAD de gestion* présente aux responsables opérationnels les indicateurs et alarmes quotidiens utiles au pilotage du travail des opérateurs (respect des normes de qualité, charge de travail des ressources). Un SIAD de gestion équipe, ou devrait équiper, chacun des processus de production.
- le *SIAD stratégique* présente aux dirigeants des séries chronologiques mensuelles éclairant l'efficacité et le positionnement de l'entreprise (résultat, part de marché, tendances). Il fournit au comité de direction une évaluation partagée et précoce des indicateurs essentiels.

Nous ne parlons ici que du SIAD stratégique, que nous appellerons SIAD tout court.

\* \*

Le SIAD occupe, dans le système d'information, une place très particulière. Il est pour le SI un point d'aboutissement. Le SI soutient et alimente le SIAD, mais celui-ci ne soutient rien d'autre que l'interface qui permet aux utilisateurs de le consulter. Il a pour fonction de faire rayonner le SI. Sa place, dans l'architecture du SI, est ainsi semblable à celle du coq à la pointe d'un clocher : porté par les fondations et les piliers de l'architecture, il est visible et c'est là tout son rôle. Cela déconcerte parfois l'intégrateur de système qui recherche la synergie entre des éléments qui, comme les deux moitiés d'une voûte, s'appuient mutuellement.

Pour rendre visible un SI, on doit être guidé par un critère de pertinence : cela suppose que l'on ait défini l'action qu'il s'agit d'éclairer et que l'on ait fait passer, à travers la masse des données que brasse le système d'information, le scalpel d'une sélectivité exigeante.

Mais pour le SIAD stratégique il ne suffit pas de sélectionner les données, il faut aussi les transformer. Les bases de données opérationnelles contiennent les dossiers en cours de traitement, mais non les séries chronologiques qui seules sont utiles au stratège. Par ailleurs, même si la qualité des données suffit aux tâches opérationnelles, la mise en perspective chronologique nécessite l'intervention du statisticien qui évalue les données manquantes, redresse les pondérations, corrige les codages.

Pour que l'on puisse définir l'action que l'on veut éclairer, il faut bien sûr que l'entreprise sache ce qu'elle veut être et ce qu'elle veut faire. Or la définition des priorités est entre ses directions – direction financière, direction de la production, direction commerciale, direction de la communication etc. – l'enjeu d'un conflit auquel seul l'arbitrage du directeur général peut mettre un terme.

A ces difficultés techniques et « politiques » s'ajoute enfin une difficulté intellectuelle. Personne ne conteste, par exemple, que la sobriété ne soit l'une des premières qualités d'un tableau de bord. Cependant elle suppose une sélectivité qui semble contraire au réalisme, ce que l'on sélectionne étant découpé dans la masse dont on fait abstraction. Or la hardiesse que réclame la pratique de l'abstraction est peu répandue. Par ailleurs les conventions comptables, souvent étrangères au raisonnement économique, exercent sur le traitement des données une dictature dont il est difficile de s'affranchir.

### **La dimension intellectuelle et « politique » : cas d'un SIAD stratégique**

*Nota Bene* : Nous présentons ici un cas réel et exemplaire. Le nom de l'entreprise a été modifié et des détails ont été omis pour préserver son anonymat.

Les tableaux de bord mis à la disposition des dirigeants d'Elseneur au milieu des années 90 présentaient des défauts manifestes. Les évaluations des données commerciales étaient biaisées (le chiffre d'affaires du dernier mois était sous-estimé d'environ 3 %) ; la présentation des séries chronologiques était fallacieuse : au lieu d'utiliser la correction des variations saisonnières (CVS), qu'ils jugeaient « trop compliquée », les dirigeants réclamaient que l'on présentât les données sous la forme R/R (« réalisé sur réalisé », valeur du mois divisée par la valeur du mois correspondant de l'année précédente) ou R/P (« réalisé sur prévu », valeur du mois divisée par celle prévue lors de la préparation du budget). Ces présentations interdisaient de percevoir l'évolution des indicateurs. Elles comportaient des variantes qui alourdissaient les tableaux de bord sans apporter d'information supplémentaire (R/R et R/P étaient parfois appliqués non à la valeur du mois, mais au cumul depuis le début de l'année, à la somme glissante des douze derniers mois etc.)

Des commentaires paraphrasaient les tableaux de nombres en égrenant des pourcentages ; ils fournissaient peu d'explications (il serait d'ailleurs difficile de commenter de façon pertinente un indicateur fallacieux).

### **Pourquoi l'on préfère un estimateur biaisé**

Le biais de 3 % dans la première évaluation publiée du chiffre d'affaires mensuel d'Elseneur résulte du « principe de prudence » cher aux comptables : lors de la première publication du chiffre d'affaires du mois  $m$ , vers le 15 du mois  $m + 1$ , certaines données ne sont pas encore parvenues. Si une région n'a fourni aucune information, on suppose que son chiffre d'affaires

est nul ; si elle n'a fourni d'information que sur les quantités vendues, on suppose qu'elles ont été vendues avec la plus forte ristourne. L'évaluation définitive, publiée le 15 du mois  $m + 2$ , repose sur une information complète : systématiquement, la deuxième évaluation est supérieure à la première.

Nous avons fait remarquer au responsable de ce calcul que, le biais étant à peu près constant, il était possible le corriger. Il nous a répondu que s'il publiait des données sans biais, une fois sur deux la publication définitive réviserait à la baisse le chiffre d'affaires du mois précédent. Et il a ajouté : « La direction générale admet les révisions à la hausse, mais si je fais une révision à la baisse je me ferai licencier ».

Ainsi le DG ignore qu'un estimateur sans biais est excessif une fois sur deux en moyenne ; il préfère recevoir un estimateur biaisé, au risque de fausser sa perception du marché.

Chaque service de l'entreprise se faisant un devoir de publier un tableau de bord, le PDG en recevait chaque mois quelques dizaines, tous conçus selon les méthodes que nous venons de décrire (donc incompréhensibles et mutuellement incohérents). Il les faisait classer sans les regarder. Lors des réunions de direction, les écarts entre statistiques occasionnaient de pénibles discussions.

Les erreurs de méthode étaient ancrées dans les habitudes. Le contrôle de gestion était trop proche de la comptabilité pour prendre le recul nécessaire au raisonnement économique.

### **Pourquoi il est difficile de construire un tableau de bord**

Un tableau de bord, pour être efficace, ne doit fournir aux dirigeants que les indicateurs qui leur sont utiles : un tableau de bord lourd n'est pas lu. Mais pour sélectionner les indicateurs utiles il faut entrer dans le raisonnement stratégique, compétence rare.

Il faut aussi traiter les données, les corriger, les interpréter pour en extraire l'information : cela suppose des compétences en statistique et en économie qui sont elles aussi rares.

La démarche la plus fréquente consiste à produire des tableaux de bord lourds et à laisser au lecteur le soin de s'y retrouver. L'ambition est de décrire « complètement » la réalité de l'entreprise, but que l'on croit approcher en publiant beaucoup d'indicateurs. Enfin, on suppose que l'« objectivité » consiste à transcrire les données comptables en s'abstenant de toute rectification, alors qu'elles comportent des défauts évidents.

### ***Le projet***

Il fallait créer du nouveau en évitant l'appellation « tableau de bord », déjà accaparée. La direction de la stratégie proposa au DG d'établir un « bloc-notes tendanciel » qui ferait apparaître l'évolution des principaux indicateurs.

Le « bloc-notes » devait se restreindre aux données essentielles. Pour faire apparaître les tendances, chaque série serait corrigée des variations saisonnières et une extrapolation serait ajustée sur ses dernières valeurs. Les données seraient accompagnées de commentaires signalant les retournements de tendance et expliquant les points aberrants (grèves, nouveaux produits, événements géopolitiques etc.)

La cellule élémentaire du bloc-notes comporterait ainsi :

- le graphique d'une série CVS (ou de quelques séries qu'il est utile de comparer), prolongé par une extrapolation de tendance,
- une fenêtre contenant les nombres essentiels (niveau atteint, taux de croissance mensuel, taux tendanciel),
- une fenêtre contenant un bref commentaire.

Une page du bloc-notes contiendrait une ou deux cellules élémentaires.

Les données ainsi produites seraient différentes des données « officielles » diffusées par le contrôle de gestion car elles seraient corrigées des biais d'observation. Il conviendrait donc de limiter leur diffusion aux membres du comité de direction (quinze personnes).

Un « club du bloc-notes » rassemblerait ses rédacteurs et les personnes qui leur fournissent les données. Celles-ci auraient ainsi le privilège de voir les séries clés de l'entreprise et de participer à leur interprétation ; elles fourniraient par la même occasion les commentaires nécessaires : la création de ce club visait à désamorcer la réticence des « propriétaires » de l'information en les intéressant à la réussite du bloc-notes.

### ***La mise en place***

Le DG fut convaincu par une maquette et demanda à la direction de la stratégie de mettre le bloc-notes en place ; le travail a été réalisé par une petite équipe.

Le « club du bloc-notes » a été lancé immédiatement. Les participants sont assidus. C'est un élément clé du dispositif : les réunions intéressent les participants qui fournissent des commentaires indispensables. Si l'entreprise n'avait pas jusqu'alors de bonnes méthodes en matière d'indicateurs et de tableau de bord, elle dispose d'experts qui connaissent les données et en donnent un commentaire éclairant. Les réticences à fournir l'information sont effacées.

Le bloc-notes, publié vers le 25 du mois  $m + 1$ , comporte une partie fixe (séries régulièrement suivies) et une partie variable qui éclaire un problème jugé particulièrement intéressant. Il se peut qu'un indicateur temporaire devienne permanent s'il répond à une priorité durable.

Une difficulté s'est présentée lors de l'élaboration du premier « bloc-notes » : pour construire une CVS, il faut disposer de données sur trois ans au moins. Or Elseneur venait d'absorber une autre entreprise. Il fallut reconstituer sur une durée suffisante l'entité fictive composée par la réunion d'Elseneur et de cette entreprise.

Les premières éditions du bloc-notes ont eu du succès auprès du comité de direction. Quelques exemples d'actualité ont ensuite permis de prouver l'efficacité du concept et de la CVS : ainsi le bloc-notes a permis à l'entreprise de percevoir de façon précoce les effets de la crise financière asiatique de 1998.

### ***Le fonctionnement***

La constitution du stock des séries CVS a été progressive, et le domaine cible a évolué pour adapter le bloc-notes aux priorités stratégiques. En outre la partie variable a été enrichie d'un éditorial, petite étude très appréciée sur le sujet important du mois.

Dans sa première définition, le bloc-notes devait fournir les rubriques permettant un calcul de rentabilité et un suivi des parts de marché. Mais la stratégie de l'entreprise s'articule désormais selon quatre dimensions entre lesquelles le PDG s'efforce de maintenir l'équilibre :

- 1) rapport qualité / prix du produit ;
- 2) avantages concurrentiels ;
- 3) gestion sociale et cognitive ;

#### 4) fonction de production et création du cash-flow.

Ces quatre dimensions ont déterminé la présentation du bloc-notes. La gestion sociale et cognitive est le point le plus délicat : si le suivi social est réalisé par un sondage hebdomadaire, on ne possède pas d'indicateurs sur l'évolution du capital intellectuel.

La philosophie que reflète le bloc-notes a ainsi toutes les chances d'être pérenne, quels que soient les PDG futurs de l'entreprise.

Le bloc-notes est purement descriptif même si les commentaires apportent des éléments d'explication. Or on peut avoir besoin de modéliser des schémas de causalité, fussent-ils hypothétiques, pour les tester sur les données d'observation. Les tentatives d'application de l'économétrie ont cependant été décevantes. Le raisonnement repose plutôt sur la comparaison visuelle de l'allure des séries : l'économétrie se fait ainsi « à l'œil » et non par le calcul.

La plupart des dirigeants n'ont plus fait de mathématiques après le lycée. Il est impossible de leur présenter un graphique en échelle semi-logarithmique et *a fortiori* ils ne sont pas prêts à entendre une expertise économétrique.

Le DG a jugé dangereux de mettre en circulation un document qui révèle les forces et faiblesses de l'entreprise : sa diffusion est restée limitée au comité de direction, ce qui lui confère d'ailleurs un grand prestige. Cependant il a été mis sur l'Intranet de l'entreprise, sous un contrôle d'accès rigoureux. Certains spécialistes des divers domaines ont ainsi accès chacun aux indicateurs concernant son domaine. Les membres du comité de direction trouvent sur l'Intranet, outre le contenu de la dernière édition du bloc-notes, la série complète des éditoriaux et l'historique des commentaires. On y trouve également les définitions et précisions sur les périmètres couverts, ce qui facilite l'interprétation des données et limite les malentendus.

#### ***Les apports du bloc-notes***

Le bloc-notes est présenté au comité de direction par le directeur de la stratégie. Il se fait « briefer » avant la réunion par le responsable du bloc-notes, auquel il rapporte les questions posées lors de la réunion.

Le comité de direction examine attentivement le bloc-notes : sa discussion occupe trois quarts d'heure dans une réunion de quatre heures. Les pertes de temps naguère occasionnées par les divergences entre statistiques ne se produisent plus.

Les calculs réalisés par l'équipe du bloc-notes n'ayant jamais été mis en défaut, un climat de confiance s'est créé. Les chiffres ne sont pas discutés. En cas d'évolution soudaine, la première question porte sur la qualité des chiffres, puis très vite le comité de direction passe à l'interprétation du phénomène.

Le bloc-notes a permis de graver dans la mémoire des dirigeants le profil de certaines séries ainsi que certains ordres de grandeur essentiels. Les commentaires sécurisent le chiffrage en montrant que l'information a été analysée et contribuent à l'efficacité de la lecture. Même si certains dirigeants prétendent avoir tout compris avant qu'on ne le leur explique, aucun d'entre eux ne réclame la suppression des commentaires.

L'extrapolation des tendances nourrit le commentaire : *la tendance est l'information principale fournie par le bloc-notes*. Elle a d'autant plus de valeur qu'elle a été validée par les experts du domaine. La succession des blocs-notes permet de signaler ses inflexions et retournements. Certains dirigeants disent voir la tendance sur le graphique sans qu'on la leur indique, mais ils ne pourraient sans doute pas détecter ses retournements.

Si la sélectivité du bloc-notes est une ascèse, elle est appréciée. Il est accompagné d'annexes fournissant des ventilations par marché. Elles ne sont là que pour un éventuel complément d'information.

Le bloc-notes est une aide pour la décision, mais une aide indirecte : il ne dicte pas à lui seul la décision, qui suppose une réflexion et une concertation approfondies. Comme il permet aux dirigeants de partager une même connaissance des ordres de grandeur et des tendances, il leur fournit un cadre conceptuel commun qui facilite la préparation de la décision. Par ailleurs les indicateurs révèlent parfois un problème de façon précoce et facilitent sa perception collective au sein du comité de direction, étape nécessaire (même si elle n'est pas suffisante) de la solution.

### ***Difficultés « politiques »***

La pérennité du bloc-notes semble assurée. Le rattachement à la direction de la stratégie devrait être durable : il faut qu'il soit établi par une direction généraliste. S'il était établi par le contrôle de gestion, qui est rattaché à la direction financière, il accorderait trop de place aux données financières.

Cependant le contrôle de gestion n'a jamais accepté que le tableau de bord soit établi par un autre service. C'est le principal risque politique. Lorsque le bloc-notes a été lancé, le contrôleur de gestion a été invité aux réunions du club mais n'a pas accepté d'y participer. Son successeur participe aux réunions ; il s'intéresse aux indicateurs, pose des questions, mais ne contribue pas à leur interprétation. Lorsqu'il participe à une réunion le climat du club n'est d'ailleurs pas le même : le contrôle de gestion est craint parce qu'il détient un pouvoir sur les budgets. En sa présence les participants se sentent moins libres de parler.

Le conflit s'est durci lorsque le bloc-notes a publié un indicateur mensuel du coût unitaire des produits. Le contrôle de gestion produisait des comptes d'exploitation par produit (CEP) trimestriels. L'équipe du bloc-notes a conçu un modèle économétrique qui permet d'estimer des coûts mensuels en s'appuyant sur les indicateurs et en se calant sur le CEP. L'écart constaté après coup avec la mesure que fournit le CEP est instructif. Ce travail a été contesté par le contrôle de gestion, et le conflit a dû remonter à l'arbitrage du DG.

Le CEP a d'ailleurs des défauts : fondé sur une comptabilité de trésorerie et non sur une comptabilité au fait générateur, il porte la trace des aléas des dates de facturation et de paiement. Cependant il était politiquement impossible pour le bloc-notes de s'écarter des coûts publiés par le CEP, du moins au niveau agrégé de la nomenclature des produits. Par contre aux niveaux détaillés les erreurs du CEP sont par trop apparentes et l'équipe du bloc-notes a dû s'affranchir de la contrainte du calage.

Le contrôle de gestion a finalement supprimé le CEP pour construire le modèle ACP (Analyse des Coûts de Production) alimenté par le progiciel SAP qui équipe Elsevier. Comme SAP ne fournit pas de coûts unitaires mensuels, ACP les évalue en multipliant les volumes par des coûts standards annuels : cela revient à supposer que les coûts unitaires sont constants durant l'année, hypothèse très grossière. ACP fonctionne mal et ne satisfait personne.

L'équipe du bloc-notes met au point une méthode qui s'appuiera sur ACP et l'améliorera en répartissant au mieux la « poubelle » (quelques centaines de millions d'euros par mois) où s'accumulent les écarts entre les coûts constatés et l'estimation fondée sur les coûts standards. Le bloc-notes publiera ainsi des coûts détaillés différents de ceux que fournit ACP et on peut s'attendre à un nouvel affrontement.

Enfin le contrôle de gestion a entrepris de mettre sur l'Intranet de l'entreprise, dans la foulée de SAP, un « tableau de bord du DG ». Il entend ainsi publier 150 indicateurs mensuels. Ce

projet coûteux, qui emploie une armée de consultants, tourne le dos à la méthode artisanale employée pour le bloc-notes. Il ferait retomber l'entreprise dans l'ornière du tableau de bord lourd produit automatiquement, peu commenté et éventuellement mal vérifié.

La cohabitation de deux tableaux de bord, si elle se produit un jour, suscitera la confusion dans l'esprit des dirigeants. Il semble toutefois peu probable que le contrôle de gestion parvienne à construire un tableau de bord convenable. Pour le moment, le futur « tableau de bord du DG » ne publie que cinq indicateurs, diffusés à l'aide d'un outil HTML qui prend 30 secondes pour afficher des images de courbes figées. Il est choquant que ce projet ait pu consommer une telle part des ressources budgétaires, mais il est vrai que le contrôle de gestion central peut accéder à ces ressources-là plus aisément que les autres services...

### **La dimension technique du SIAD stratégique : sélectivité et traitements**

Le SIAD est un outil d'observation qui vise, à partir de données de gestion et/ou de statistiques, à donner aux managers d'une entreprise les moyens d'identifier des alertes de gestion, de suivre l'évolution de l'activité et de disposer d'outils d'investigation de sujets ou phénomènes particuliers. Il ne fournit pas les explications ni les commentaires qui relèvent d'une phase de travail postérieure à l'observation.

Le SIAD tire parti de l'ensemble des données produites ou acquises par l'entreprise, ensemble dont il fournit une présentation synthétique. Cela suppose (1) que le SIAD soit alimenté potentiellement par toutes les applications de l'entreprise, (2) qu'il résolve les problèmes de comparabilité et de redressement des données qui se posent lorsque l'on utilise des sources diverses.

Le SIAD vise à présenter des informations utiles. Ceci implique qu'il soit construit selon des critères de sélectivité en choisissant, parmi les statistiques qu'il est possible de produire, celles qui peuvent servir à telle ou telle catégorie d'utilisateurs. Sa construction suppose donc une analyse des besoins, elle-même fondée sur une segmentation des utilisateurs en sous-populations homogènes en ce qui concerne les missions à remplir et les besoins correspondants.

Le SIAD vise à fournir aux utilisateurs un outil de consultation commode, d'une ergonomie aisée, de façon à minimiser les tâches de recherche de l'information et de présentation des résultats.

Produire des statistiques en adressant au coup par coup des requêtes à une application opérationnelle est coûteux en traitement. Le SIAD protège donc les bases de données opérationnelles en s'intercalant comme un tampon entre elles et les utilisateurs et en préparant la plupart des statistiques dont ces derniers ont besoin.

Les outils fournis par le SIAD pour remplir ces divers objectifs sont :

- des tableaux de bord comportant des alertes ;
- des tableaux préformatés contenant l'essentiel de la statistique d'activité et d'environnement ;
- des tableaux et graphiques restituant des résultats synthétiques en utilisant la technologie « hypercubes » ;
- la publication d'analyses sophistiquées à la demande (analyse de corrélation, simulation etc.) utilisant les outils de « datamining ».

*La technologie « hypercubes »* permet à l'utilisateur de construire, à partir de tableaux multidimensionnels intermédiaires, les séries chronologiques ou tableaux croisés dont il a

besoin. Le contenu des hypercubes doit être défini *a priori*, à partir de l'analyse des besoins, de sorte qu'ils satisfassent au mieux les besoins (cf. annexe).

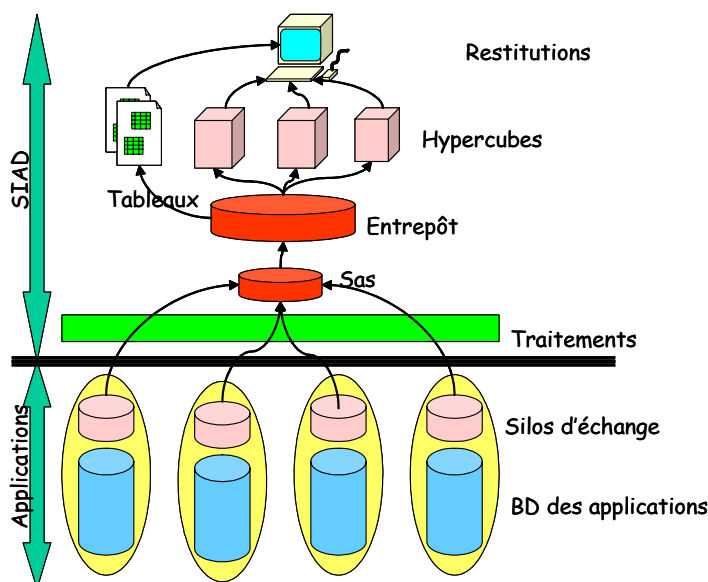
Etant sélectif, le SIAD ne peut pas répondre à *toutes* les questions imaginables, mais seulement à la plupart des questions. Il peut arriver qu'un utilisateur recherche une information que le SIAD ne fournit pas. Il faut pourtant que l'on puisse lui répondre. Ce sera la tâche d'une équipe d'analystes habilités à interroger la base de données intermédiaire. Toutefois le délai de réponse sera plus long (quelques heures ou quelques jours) que celui de la consultation des hypercubes (quelques secondes).

L'évolution des hypercubes, comme celle du dispositif dans son ensemble, est assurée par une cellule d'administration centrale en relation avec les analystes.

### Architecture du SIAD

Un SIAD peut être présenté selon trois couches :

- L'alimentation par les applications opérationnelles ;
- Le stockage historisé, l'agrégation et la constitution des hypercubes ;
- La publication sous forme d'alarmes, tableaux préformatés, tableaux croisés et graphiques.



Seuls les hypercubes sont consultables par les utilisateurs, qui peuvent ainsi construire une grande diversité de tableaux croisés.

La base de données du SIAD est exploitée par l'équipe d'analystes : (1) pour produire les hypercubes, (2) pour répondre à des demandes complexes formulées par les utilisateurs.

La constitution de la base de données du SIAD à partir des applications qui l'alimentent nécessite :

- une procédure permettant d'extraire périodiquement de chaque application les données nécessaires. Ces données sont rangées dans une base de données appelée « silo d'échange » dont la responsabilité appartient à l'application source,
- un traitement réalisé par le SIAD pour vérifier et redresser les données avant de les intégrer à la base de données du SIAD.

La structuration de l'entrepôt de données fait l'objet d'une modélisation formelle qui précise :



- les « axes », variables qualitatives dont le croisement définit les tableaux que le SIAD pourra produire (exemple : mois, région, secteur d'établissement, etc.) ;
- les « attributs », variables qui seront ventilées dans les cases des tableaux (exemple : nombre de personnes, montants en francs, nombre d'entretiens, etc.).

Les hypercubes sont produits, à partir de la base de données du SIAD, par agrégation de données individuelles (nous utilisons ici le mot « individu » au sens qu'il a en statistique pour désigner une unité particulière au sein d'une « population » qui peut être un ensemble d'établissements, de clients, de fournisseurs, de francs de dépense etc.) Ainsi, si le SIAD contient les données détaillées sur les clients, les hypercubes permettront de construire des tableaux croisés décrivant la population des clients.

### ***Base de données du SIAD***

La base de données du SIAD est détaillée ; lorsque l'application source contient des enregistrements individuels, la base de données du SIAD est elle-même individuelle. Cependant, alors que la base de données opérationnelle est « vivante », qu'elle subit des modifications par mises à jour continues, le SIAD est une base « morte » enregistrant des situations passées dont elle doit permettre de reconstituer l'enchaînement chronologique.

Considérons par exemple la base de données des clients. A chaque client correspondent un identifiant fixe et des variables qui évoluent dans le temps. Seule une *sélection* parmi ces variables intéresse le SIAD ; celui-ci est donc à la fois exhaustif (en ce qui concerne la population étudiée) et sélectif (en ce qui concerne les variables observées).

Il est *a priori* possible d'utiliser deux méthodes différentes pour constituer la base de données du SIAD : prendre une suite périodique de « photographies » instantanées de la base vivante, ou considérer les *événements* qui modifient cette base (on appelle « événement » une modification datée d'un enregistrement individuel, affectant l'une des variables sélectionnées pour le SIAD).

1 - Supposons le SIAD alimenté par une « photographie » périodique de la base vivante (par exemple on copie dans le silo d'échange, chaque vendredi à 20h00, des enregistrements fournissant, pour chaque client, les valeurs des variables sélectionnées). La dimension historique du SIAD sera alors obtenue en considérant la succession de ces enregistrements hebdomadaires. Le volume de la base de données du SIAD croîtra progressivement, par empilement de fichiers hebdomadaires, ce qui peut poser à terme un problème de volumétrie.

2 - Utiliser les « événements » pour nourrir la base du SIAD, et non des « photographies » périodiques, apporte deux améliorations :

- le volume de la base de données est plus faible, puisque les enregistrements qui n'ont pas été modifiés ne sont pas recopiés,
- le fait que les événements soient datés permet de construire, lors des exploitations, un découpage chronologique quelconque (par semaine, mois etc.), alors que le calcul exact de données mensuelles à partir de données hebdomadaires n'est pas possible.

Soulignons, pour écarter un risque de malentendu, que l'on entend par « événement » la modification d'une variable interne au dossier considéré, qu'elle soit calculée ou obtenue par observation. Ainsi le SIAD n'a pas à reproduire les traitements réalisés au sein de l'application source, mais à recueillir leurs résultats.

Si l'on choisit d'utiliser les « événements », la base de données du SIAD comporte :

- une photo initiale de la base source, constituée par la liste exhaustive des enregistrements individuels identifiés dans la base opérationnelle et, pour chaque enregistrement, la liste sélective des variables observées,
- des enregistrements individuels datés correspondant à chaque événement, de telle sorte que le traitement de la base permette de reconstruire l'évolution historique de chaque individu.

Pour simplifier, et par abus de langage, on appelle souvent « événement » tout court chacun de ces enregistrements.

NB : pour limiter la volumétrie on distingue, parmi les variables observées, celles dont on souhaite suivre l'historique et celles dont on souhaite seulement connaître l'état actuel. Les événements concernant les variables dont on souhaite suivre l'historique sont conservés en mémoire ; pour les autres variables, seul l'événement le plus récent est conservé et il « écrase » les événements antérieurs. Cette distinction se fait à l'intérieur de la base du SIAD : elle ne concerne pas le silo d'échange qui doit contenir temporairement tous les événements.

### ***Silo d'échange***

A partir de ce qui précède, il est aisé de concevoir le silo d'échange que chaque application doit constituer. Au démarrage du SIAD, l'application construit la base de données initiale, qui indique pour chaque individu la valeur des variables sélectionnées par le SIAD. Cette base de données servira ensuite à amorcer la base de données du SIAD.

Le « silo d'échange » est une base de données qui stocke selon la technique que l'on a retenue soit les événements, soit les photographies, au fur et à mesure de leur occurrence. Si l'on a choisi de recueillir les événements il faut prévoir, à l'intérieur de l'application, un mécanisme qui détectera les événements et enverra les enregistrements correspondants vers le silo d'échange.

Le moteur d'alimentation du SIAD consulte périodiquement le silo, recopie ses éléments vers une base temporaire nommée « sas » (ils seront ensuite traités pour alimenter la base du SIAD), puis le purge.

Le « silo d'échange » est de taille modeste, son volume se limitant à celui de la dernière photographie ou à celui des « événements » survenus entre deux consultations par le SIAD. Il peut contenir des indications techniques visant à garantir la qualité de l'alimentation du SIAD ; il faut en effet s'assurer périodiquement (et ce n'est pas facile) :

- que l'image figurant dans le SIAD ne diverge pas de la réalité par suite d'un cumul d'erreurs dans la collecte des événements ;
- que les opérations visant à « nettoyer » l'application (« purge » d'enregistrements désuets, corrections des codages et identifiants) ne suscitent pas des erreurs en provoquant des événements artificiels ;
- que les modifications des classifications et nomenclatures utilisées dans l'application sont correctement répercutées dans le SIAD.

### ***Commande adressée par le SIAD à une application***

On voit maintenant ce que doit contenir la commande adressée par le SIAD à une application :

- définition des « individus » qui seront observés : *a priori*, tous les « individus » gérés par l'application intéressent le SIAD (clients et commandes, fournisseurs et offres) ;
- liste des variables qui seront observées par le SIAD sur chacun de ces individus ;

- indications techniques visant à garantir la qualité de l'alimentation du SIAD.

Les responsables de l'application devront, à partir de cette commande, faire développer l'outil qui alimentera le silo conformément à la méthode décrite ci-dessus. Il n'est pas indispensable que le SIAD indique dès la passation de sa commande la liste exacte des variables qu'il voudra observer : une fois l'alimentation du silo d'échange en place, cette liste peut être modifiée car elle constitue un paramètre de ce mécanisme.

### ***Traitements réalisés par le SIAD***

Les données brutes issues d'une application opérationnelle ne se prêtent jamais telles quelles à une exploitation statistique comme celle que réalise le SIAD : il faut corriger les erreurs, estimer les données manquantes etc. Entre le silo d'échange et la base de données du SIAD s'intercale donc un traitement des données. Ce traitement peut poser problème lors de l'utilisation : lorsque l'on remplace une donnée manquante par une estimation, cela peut donner une information utilisable au niveau France entière, mais fausser les proportions au niveau d'une commune ou d'une région. Il faudra donc lors de l'utilisation des données disposer de contrôles garantissant leur représentativité.

Le traitement comporte deux étapes : la première apporte des corrections purement techniques visant à garantir la valeur statistique des données. L'autre apporte des transcodages visant à assurer la compatibilité des données avec les définitions réglementaires et comptables.

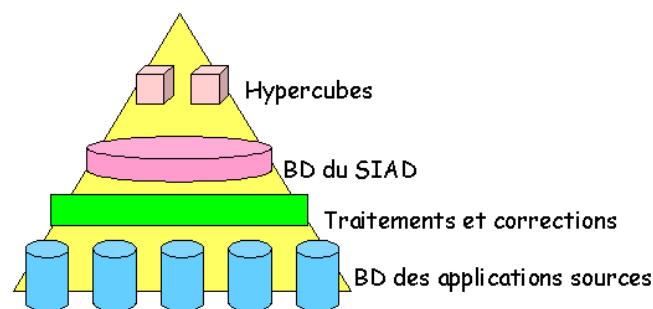
### **Apports et limites du SIAD**

Le SIAD a pour but de fournir des observations qui alimenteront, après recoupement avec d'autres sources (économiques, démographiques, marketing etc.) la compréhension du marché et permettront de suivre l'activité, d'analyser son efficacité, d'optimiser les moyens, le tout en vue de faciliter l'orientation de l'action.

Le SIAD n'a, par contre, pas vocation à fournir des indicateurs pour un pilotage opérationnel au jour le jour ou pour le suivi de dossiers individuels : il faut que chaque processus de production fournisse les indicateurs et alarmes qui permettent aux responsables opérationnels de piloter les travaux au plus près de leur réalisation. Cependant le SIAD peut conforter la gestion : un responsable peut trouver, dans les hypercubes produits par le SIAD, telle série chronologique qu'il confrontera avec des données de gestion pour évaluer l'efficacité du travail de son unité.

### ***Apport du SIAD stratégique à l'analyse***

Certains représentent l'architecture du SIAD (que l'on nomme souvent « datawarehouse ») par une pyramide. Sa base est constituée des applications qui l'alimentent, son sommet par les hypercubes et autres outils d'observations synthétiques :

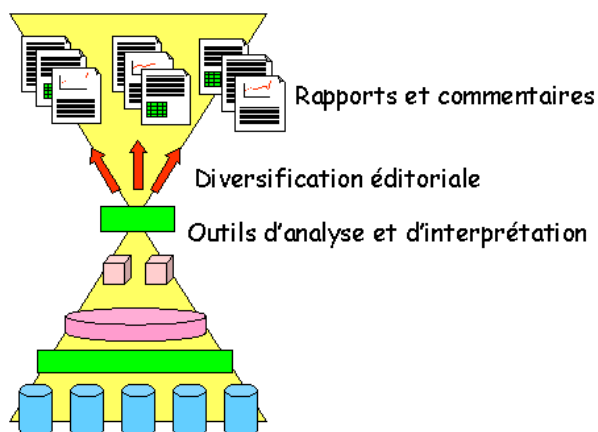


Mais le SIAD est bâti à partir des données d'observation, et il faut distinguer l'observation de l'explication : il faut, pour comprendre ce qui se passe, associer l'observation à la connaissance des théories concernant l'objet observé (un microscope permet de voir des

bactéries, mais ne les explique pas). C'est en complétant le SIAD par des outils d'analyse des données et d'économétrie, et en le confrontant à des modèles explicatifs, que l'on pourra comprendre ce qui se passe sur un marché. Or si le SIAD alimente ces outils, il ne les comporte pas.

Il est donc utile de compléter le schéma ci-dessus en représentant les opérations éditoriales qui s'appuient sur le SIAD ; utilisant les données d'observation synthétiques et les confrontant à des modèles explicatifs, elles produisent des résultats interprétés et commentés destinés à diverses populations d'utilisateurs (responsables régionaux, responsables de ligne de produit etc.)

La représentation n'a plus alors la forme d'une pyramide mais celle d'un diabolos :



S'il est souvent nécessaire pour ces interprétations d'utiliser les méthodes de l'analyse des données ou de l'économétrie il est recommandé de rien laisser paraître de ces démarches techniques dans la publication, qui ne doit en accueillir que les résultats : il ne convient pas en effet de laisser les échafaudages en place après la construction d'un immeuble.

Le SIAD est un puissant outil d'observation ; il appelle un dépassement vers l'explication et le commentaire. Ce dépassement implique, pour servir les diverses populations d'utilisateurs concernées, la diversification éditoriale que représente la partie haute du diabolos.

### Annexe : à propos des hypercubes

Un « hypercube » est, comme son nom l'indique, un tableau à plusieurs dimensions représentant la répartition d'une population selon deux variables ou plus. L'œil humain ne pouvant lire que les tableaux à deux dimensions, l'hypercube sera, s'il possède plus de deux dimensions (ce qui est le cas général), un être purement informatique mais invisible pour l'être humain. Son utilité réside dans la multiplicité des tableaux à deux dimensions ou des séries chronologiques que l'on peut obtenir, à partir d'un hypercube à n dimensions, par sommation sur n-2 ou n-1 indices.

*Exemple* : supposons que la population soit celle des clients (personnes physiques) et que les critères soient la région, le mois, le métier, la tranche d'âge et le sexe. L'hypercube est alors un tableau à cinq dimensions. Sa case courante contient le nombre de clients tel mois dans telle région, ayant tel métier, tel sexe, et appartenant à telle tranche d'âge. Si l'on utilise les lettres I, J, K, L et M pour désigner les indices qui repèrent les modalités de chacune de ces variables, le nombre figurant dans la case courante de l'hypercube est noté  $x_{ijklm}$ .

Dans le jargon des datawarehouses on dit « axe » au lieu de « variable » et « segment » pour désigner l'ensemble des individus possédant la même modalité d'une variable. Pour répartir une population selon les modalités d'une variable, il faut qu'il s'agisse d'une variable qualitative (comme « mois » ou « région ») ; les variables quantitatives (comme « revenu » ou

« âge ») doivent, pour être représentées par un tableau, être rendues qualitatives en définissant des classes (« classe d'âge », « tranche de revenu »). On distingue les variables qualitatives pures (« sexe », « région ») et les variables qualitatives « ordinales », qui comportent un ordre naturel (« tranche de revenu », « classe d'âge »).

A partir de l'hypercube on peut, par sommation sur certains indices, obtenir des tableaux qui croisent les variables deux à deux ; si l'on utilise une notation comme

$$\sum_m x_{ijklm} = x_{ijkl}$$

la case courante du tableau qui croise les variables I et J contiendra le nombre :

$$\sum_{klm} x_{ijklm} = x_{ij..}$$

Ces tableaux peuvent être eux-mêmes redéfinis en regroupant les modalités d'une variable selon une classification plus agrégée : en regroupant des mois on peut obtenir des années, en regroupant des départements on peut obtenir des régions, etc.

Les logiciels usuels comportent des outils commodes permettant à l'utilisateur :

- de sélectionner les variables qu'il veut croiser et regrouper les modalités que l'on souhaite agréger ;
- de visualiser les données (courbes, histogrammes, « fromages » etc.) ;
- de réaliser sur le tableau de nombres tous les calculs jugés opportuns ;

L'expérience montre que les utilisateurs maîtrisent rapidement ces outils.

### **Références**

Isabelle Boydens, *Informatique, normes et temps*, Bruylant Bruxelles 1999

Michel Volle, *Le métier de statisticien*, Economica 1984

Michel Volle, « Fonctionnement d'un Système Informatique d'Aide à la Décision (SIAD) », *Extraction et gestion des connaissances (EGC 2004)*, CEPAD 2004

Michel Volle, site [www.volle.com](http://www.volle.com)