

Fonctionnement d'un Système Informatique d'Aide à la Décision (SIAD)

Michel Volle*

*Volle sarl
21, rue Weber
F-75116 PARIS
michel@volle.com
<http://www.volle.com>

Résumé. Cet article présente le fonctionnement d'un SIAD : alimentation, traitement des données qui l'alimentent, production des résultats, outils de consultation mis à la disposition des utilisateurs, exploitation éditoriale.

1. But d'un SIAD

Le SIAD est un outil d'observation et de description 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 inévitablement 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 toutes 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 chacune 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 :

- le tableau de bord comportant des alertes ;

Fonctionnement d'un Système Informatique d'Aide à la Décision (SIAD)

- des tableaux préformatés contenant l'essentiel de la statistique d'activité et d'environnement ;
- des tableaux et graphiques restituant les résultats d'interrogations en utilisant la technologie " hypercubes " ;
- la restitution d'analyses sophistiquées (analyse de corrélation, simulation etc.) utilisant les outils de " datamining ".

Remarques :

1) La technologie " hypercubes "

Cette technologie permet à l'utilisateur, par la production de tableaux multidimensionnels intermédiaires, de construire par sélection les séries chronologiques ou les tableaux croisés dont il a besoin. Le contenu de ces hypercubes doit être défini *a priori*, à partir de l'analyse des besoins, de sorte qu'ils satisfassent au mieux les besoins des utilisateurs (cf. annexe)

2) Le " datamining "

Etant sélectif, le SIAD ne peut pas répondre à *toutes* les questions imaginables mais seulement à la plupart des questions. Il peut donc 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 en région et à la DG, habilités à utiliser des requêtes et à interroger la base de données intermédiaire pour répondre à l'utilisateur. Toutefois le délai de réponse sera plus long (quelques heures ou quelques jours) que celui de la consultation des hypercubes (quelques secondes).

3) Administration du dispositif

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

2. Architecture du SIAD

Un SIAD peut être présenté selon trois couches :

- L'alimentation par les applications opérationnelles constitue la première couche ;
- La deuxième couche est constituée par le stockage historisé, l'agrégation et la constitution des cubes ;
- La restitution sous forme d'alerteurs, de tableaux préformatés, de tableaux croisés et de graphiques constitue la troisième couche.

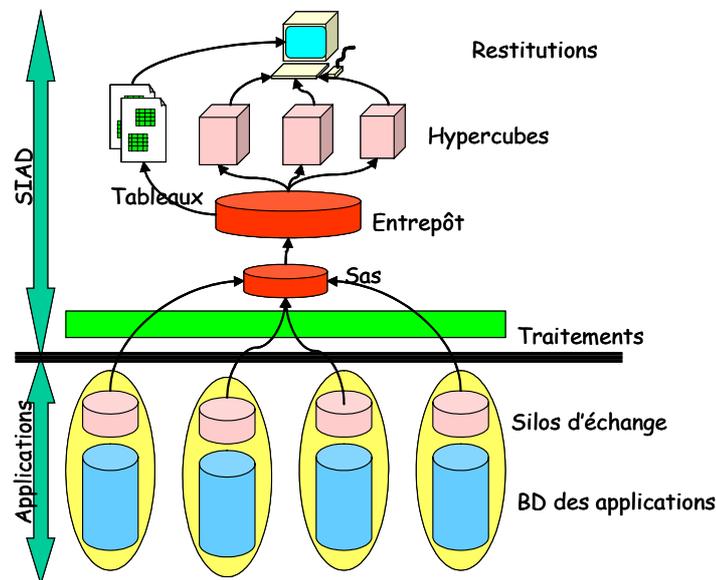


FIG. 1 – Les couches du SIAD

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

La base de données du SIAD est exploitée par une équipe d'analystes : (1) pour produire les hypercubes, (2) pour répondre à la demande à des requêtes complexes envoyées 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 par l'application 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. Ainsi, si le SIAD contient les données détaillées sur les clients, les hypercubes permettront de construire des tableaux croisés décrivant la population des clients. Nous utilisons ici le mot " individu " au sens qu'il a en statistique pour désigner une unité particulière au sein d'une population ; il ne s'agit pas nécessairement d'un être humain (la

Fonctionnement d'un Système Informatique d'Aide à la Décision (SIAD)

" population " considérée peut-être un ensemble d'établissements, de clients, de fournisseurs, de francs de dépense etc.).

2.1. Base de données du SIAD

La base de données du SIAD est nécessairement détaillée, et 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 ", c'est-à-dire subit des modifications par mises à jour continues, le SIAD est une base " morte " qui enregistre les situations passées dont elle doit permettre de reconstituer l'enchaînement chronologique.

Considérons la base de données des clients. A chaque client correspondent un identifiant fixe et des variables qui évoluent dans le temps. Seule une *sélection* de ces variables intéresse le SIAD ; celui-ci sera donc à la fois exhaustif (en ce qui concerne les individus composant 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.

Option 1 : Suite de photographies

Supposons que le SIAD soit alimenté par une copie 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 est alors obtenue en considérant la succession de ces enregistrements hebdomadaires. Le volume de la base de données du SIAD croît progressivement, par empilement de fichiers hebdomadaires, ce qui peut poser à terme un problème de volumétrie.

Option 2 : Evénements

On appelle " événement " toute modification datée d'un enregistrement individuel affectant l'une des variables sélectionnées pour le SIAD. Entre deux événements, l'enregistrement reste le même. 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 réduit puisque les enregistrements qui n'ont pas été modifiés ne sont pas copiés,
- le fait que les événements soient datés permet de construire lors des exploitations un découpage chronologique quelconque (par semaine, par mois etc.), alors que par exemple 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 au moins de l'enregistrement individuel considéré (donc d'une variable interne au dossier considéré, qu'elle soit calculée ou obtenue par observation), et non un événement au sens courant du terme, concernant l'être réel représenté par le dossier, et qui peut entraîner ou non une modification des variables observées. Ainsi le SIAD n'a pas à reproduire les traitements réalisés au sein de l'application source, dont il recueille les résultats.

Si l'on choisit la deuxième option, la base de données du SIAD comporte :

- une photo initiale de la base source, constituée par la liste exhaustive des enregistrements individuels identifiés et, pour chaque enregistrement, la liste sélective des variables observées,

- des enregistrements individuels datés correspondant à chaque événement, de sorte que le traitement de la base permette de reconstruire l'évolution historique de chaque individu. Pour simplifier, et par abus de langage, nous appellerons " 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. Observons que 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.

2.2. Silo d'échange

A partir de ce qui précède, il est aisé de concevoir la nature du silo d'échange que chaque application doit constituer. Au démarrage du SIAD, l'application construit la base de données initiale, indiquant pour chaque enregistrement individuel la valeur des variables observées par le SIAD. Cette base de données sera ensuite traitée pour amorcer la base de données du SIAD.

Un " silo " est constitué par une base de données qui stocke soit les événements, soit les photographies, au fur et à mesure de leur occurrence. Il faut donc prévoir, à l'intérieur de l'application, un mécanisme qui détecte les événements et envoie les enregistrements correspondants vers le silo qui les stocke.

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

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

2.3. 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) ;

Fonctionnement d'un Système Informatique d'Aide à la Décision (SIAD)

- 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 réaliser le développement permettant d'alimenter le silo conformément à la méthode décrite ci-dessus. Observons qu'il n'est pas rigoureusement indispensable que le SIAD indique dès la passation de sa commande la liste exacte des variables qu'il voudra observer : en effet, une fois le mécanisme d'alimentation du silo d'échange mis en place, cette liste peut être modifiée aisément (elle constitue un paramètre de ce mécanisme).

2.4. 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 une opération complexe de traitement des données. L'existence de ce traitement peut elle-même poser problème par la suite : 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 faut donc lors de l'utilisation des données disposer de contrôles ou d'alarmes 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.

3. Les apports du SIAD

3.1. Apport du SIAD à la gestion

Le SIAD a pour but de fournir des données observées alimentant, après recoupement avec d'autres sources (économiques, démographiques, marketing etc.) la compréhension du marché et permettant de réaliser le suivi de l'activité, l'analyse de son impact, l'optimisation des moyens, de façon à faciliter l'orientation de l'action. Le SIAD a donc vocation à fournir les indicateurs de pilotage permettant à un responsable opérationnel d'évaluer la qualité et la productivité du travail fourni par des établissements ou des équipes, indicateurs qui impliquent un recoupement avec des données que le SIAD ne comporte pas (volume et qualité des ressources employées, délais de traitement des affaires, etc.).

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 donc que chaque application soit munie des outils permettant aux responsables opérationnels de piloter par domaine les travaux au plus près de leur réalisation. Cependant le SIAD peut contribuer à l'alimentation de ces outils : un responsable peut trouver, dans les hypercubes produits par le

SIAD, telle série chronologique qu'il recoupera avec des données de gestion pour évaluer l'efficacité du travail de son unité.

3.2. Apport du SIAD à l'analyse

Certains représentent l'architecture du SIAD (on dit aussi " datawarehouse ") par une pyramide. Sa large base est constituée des diverses applications qui l'alimentent, le sommet par les hypercubes et autres outils d'observations synthétiques :

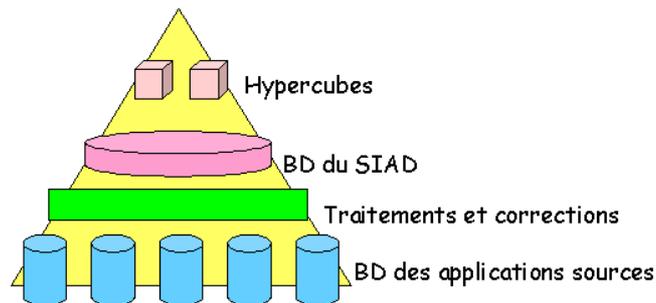


FIG. 21 – La « pyramide », représentation répandue mais partielle

Un SIAD est bâti à partir des données d'observation, et il faut distinguer l'observation de l'explication : un microscope permet de voir les bactéries, mais ne les explique pas ; il faut, pour comprendre ce qui se passe, associer l'observation à la connaissance des théories concernant l'objet observé. C'est en complétant le SIAD par des outils d'analyse des données et d'économétrie, et en le confrontant aux modèles explicatifs, que l'on pourra l'utiliser pour comprendre ce qui se passe sur un marché. Le SIAD alimente ces outils mais ne les comporte pas.

Il est donc utile de représenter les opérations éditoriales s'appuyant sur le SIAD ; utilisant les données d'observation synthétiques, elles permettent de produire des résultats interprétés et commentés destinés à diverses populations d'utilisateurs (responsables régionaux, responsables de ligne de produit etc.).

S'il est souvent nécessaire pour l'interprétation 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 recueillir que les résultats (il ne convient pas en effet de laisser les échafaudages en place après la construction d'un immeuble). La représentation n'a plus alors la forme d'une pyramide mais celle d'un diabolos :

Fonctionnement d'un Système Informatique d'Aide à la Décision (SIAD)

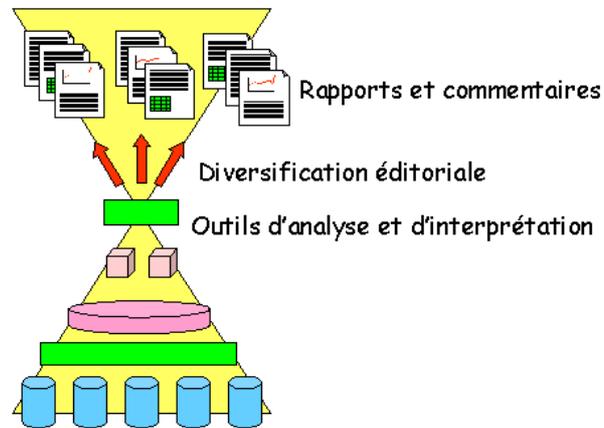


FIG. 3 – Le « diabololo », représentation complète du SIAD

Le SIAD est un puissant outil d'observation ; il appelle donc un dépassement nécessaire vers l'explication et le commentaire. Ce dépassement implique, pour pouvoir servir les diverses populations d'utilisateurs concernées, une diversification éditoriale.

Références

- [Boydens, 1999] I. Boydens, *Informatique, normes et temps*, Bruylant Bruxelles 1999.
[Hirji, 2001] K. K. Hirji, "Exploring Data Mining Implementation", in *Communications of the ACM*, juillet 2001
[Volle, 1984] M. Volle, *Le métier de statisticien*, 2^{ème} édition, Economica 1984.
[Volle, 1997] M Volle, *Analyse des données*, 4^{ème} édition, Economica 1997.
[Volle, 2002] M Volle, *Histoire d'un tableau de bord*, 20 novembre 2002
(<http://www.volle.com/travaux/tdb.htm>)

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 peut lire que des tableaux à deux dimensions ; l'hypercube est donc illisible, invisible, s'il possède plus de deux dimensions, ce qui est le cas général ; son utilité réside dans la multiplicité des tableaux ou séries 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), 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. La case courante de l'hypercube contient le nombre de clients tel mois dans telle région, qui avaient tel métier, tel sexe, et qui appartenaient à telle tranche d'âge. Si l'on utilise les lettres I, J, K, L et M pour désigner les indices servant à

repérer les modalités de ces variables, le nombre qui figure dans la case courante de l'hypercube est noté x_{ijklm} .

Dans le langage des datawarehouses, on dit "axe" au lieu de "variable" et "segment" pour désigner l'ensemble des individus possédant une même modalité de la 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 pouvoir 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 des indices, obtenir les tableaux croisant les variables deux à deux; si l'on utilise une notation du type :

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

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, si l'on regroupe les modalités d'une variable selon une classification plus agrégée. Ainsi en regroupant des mois on peut obtenir des années, en regroupant des départements on peut obtenir des régions, etc.

Les logiciels usuels de datawarehouse comportent :

- des outils commodes pour sélectionner les variables que l'on veut croiser, et regrouper les modalités que l'on souhaite agréger ;
- des outils de représentation graphique (courbes, histogrammes, "fromages" etc.) facilitant la visualisation des données ;
- des fonctionnalités de tableur permettant de réaliser sur le tableau de nombre tous les calculs jugés opportuns ;

En pratique l'expérience montre que les utilisateurs ont tôt fait d'apprendre à se servir de ces outils.