

Typology of Information Systems

Essai de typologie des Systèmes d'Informations

Mylène Leitzelman, Henri Dou

CRRM - Marseille Innovation Technopole de Château-Gombert 13451 Marseille cedex 20
mylene@imt950.imt-mrs.fr

Résumé :

A l'heure où il est plus que jamais nécessaire aux entreprises d'exploiter la bonne information au bon moment, il s'agit d'aider celle-ci à choisir le système d'informations adapté à ses besoins et à son organisation et qui l'aidera à prendre ses décisions. L'objet de cet article est de poser un nouvel éclairage sur la grande famille des Systèmes d'Informations. Ces derniers ont été positionnés sur une matrice composée d'un axe du pouvoir sur l'information et d'un axe des besoins en informations. Avec l'émergence des technologies internet une nouvelle typologie se profile qui est encore à définir.

Abstract :

While it is utmost a necessity for a firm to size the good information in due time, the firm must choose an adequate and efficient information system, fitted to its organization and which could extract the right information for decision making. The aim of this article is to enlighten in a new way the family of Information Systems. We have put them on a matrix mixing power on and needs of informations. With new technologies such as Internet, this typology will probably turn into a new one, not defined today.

Mots-clés : système d'informations, data warehouse, architecture distribuée, réseau intelligent

Keywords : information system, data warehouse, distributed network, mediated network

Essai de typologie des Systèmes d'Informations

Typology of Information Systems

Mylène Leitzelman, Henri Dou

Il y a quelque trois décennies, émergeait le concept de système d'informations [MELE 69] avec comme toile de fond la crise socio-économique de la fin des industries dites "mécaniques" de la deuxième révolution industrielle et l'entrée dans l'aire du tertiaire. Tout d'abord le Système d'Informations a été assimilé au système informatique [BOUS 96], depuis, les recherches sur ce sujet reconnaissent qu'un système d'informations est bien plus qu'un parc informatique mis en réseau, mais qu'il est intrinsèquement lié à l'organisation et aux flux économiques et sociaux qui font vivre une organisation.

Les échecs passés du "tout informatique", le foisonnement des types de systèmes proposés sur le marché et la révolution qu'introduit Internet dans le rapport à l'information brouillent un peu plus les entreprises à trouver le bon système d'informations qui s'adapte à leurs besoins. Cet essai de typologie tente de résoudre partiellement ce problème en donnant un nouvel éclairage sur cette grande famille. Dans un premier temps nous donnerons une définition du concept "Système d'Informations", puis nous aborderons la typologie des systèmes en fonction de leur infrastructure et de leurs fonctionnements internes, puis nous terminerons par des exemples de systèmes.

1 - Définition d'un "Système d'Information"

C'est un système qui rend possible la mise en commun d'informations appartenant à des sources séparées.

Système : On entend par système un ensemble d'éléments qui interagissent entre eux en échangeant des informations internes et externes, avec pour support des voies de communication [BOUZE 83].

Globalement, cette définition entend par système, l'outil informatique, cœur technique du système. Aujourd'hui, quand on parle de Système d'informations, on se réfère nécessairement à un support informatique, grâce aux progrès technologiques dans ce domaine. En effet, l'apparition du concept de Système d'informations n'est pas sans lien avec les avancées de la technologie en matière de systèmes intégrateurs d'information (cf. les architectures client/serveur, Intranet, internet, les bases de données distribuées,...).

D'autre part, cette définition intègre aussi les éléments extérieurs à l'outil informatique tels les utilisateurs (au niveau connaissance et ergonomie du système), les fournisseurs et les différentes sources d'informations (formats, textes, images,...), les réseaux et protocoles de communication, etc..., car sans ces éléments, le Système d'informations ne serait pas dynamique.

Information : Dans un Système d'informations, l'information est avant tout digitale, qu'elle soit du texte, du son, des images ou de la vidéo. Elle peut être aussi soit structurée (comme l'information digitale d'une base de données), soit non structurée -*amorphous*- (comme le texte libre, des fichiers de logiciels d'application Word, Excel, ou de messagerie électronique, les images, les schémas,...).

Un système d'informations transforme de l'information brute ou élaborée en informations de décision. Les sources de données dans lesquelles vient puiser le système peuvent indistinctement fournir des informations brutes -tel des textes, des informations événementielles, des données chiffrées-, comme des informations élaborées, tel des enquêtes des études, des statistiques...

En fait, dans un système d'informations, l'information passe d'un premier état, brut ou élaboré, à un état final qui est celui de l'information utile. Ce critère n'est attribué qu'en fonction du jugement et des besoins des commanditaires du système d'informations.

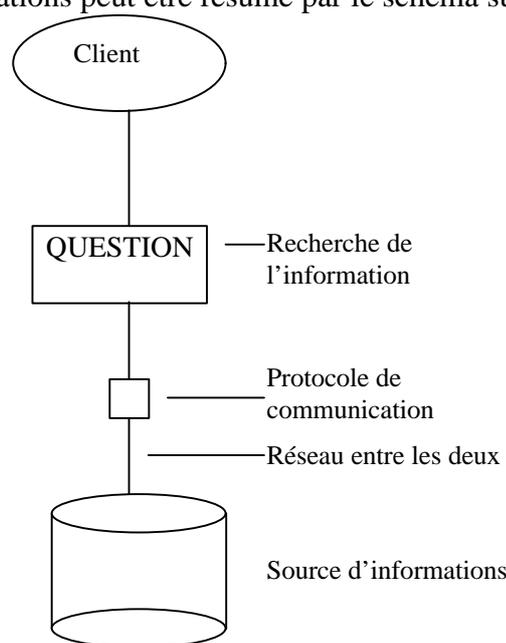
L'importance du besoin de l'utilisateur est primordial et nous verrons plus loin qu'il existe autant de type de Systèmes d'Informations qu'il existe de familles de besoins.

Sources séparées : La source est l'endroit physique et/ou virtuel où se trouvent les informations à l'origine et c'est là que vient puiser le Système d'informations. On entend par source, tout support qui gère et stocke de l'information, comme des logiciels d'application bureautique, des bases de données (objets ou relationnelles), des messageries électroniques,... Ces sources sont souvent multiples, autonomes et hétérogènes, c'est-à-dire qu'elles peuvent être séparées soit physiquement (par l'appartenance à différents logiciels), soit géographiquement (localisées dans différents endroits).

Mais si on s'en tient à cette définition, nous cernons que partiellement le concept de Système d'Informations, car définir ce concept, c'est définir ce qu'il est et aussi ce pourquoi il est. La mise en commun d'informations éparses n'a de sens que si cela vise un but, et ce but est de rendre l'information accessible rapidement et presque sans effort pour l'utilisateur, en agrégeant puis traitant l'information de façon à la rendre directement utilisable [DOU 96].

2 - Essai de typologie

Tout système d'informations peut être résumé par le schéma suivant :



On retrouve les éléments qui permettent de répondre à la définition d'un Système d'Informations, c'est-à-dire, un système qui rassemble en un même point des informations éparses et qui peuvent être interrogées par un utilisateur.

Nous avons donc comme éléments :

- au moins une source d'informations, dans laquelle se trouvent les informations originelles,
- un utilisateur qui pose sa question,
- un module d'interrogation qui traduit la question de l'utilisateur, recherche l'information et restitue les résultats,
- un réseau et un protocole de communication permettant la liaison entre les éléments du système.

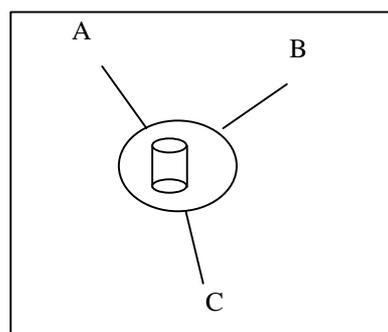
Ce schéma de base se complexifie au fur et à mesure que les sources d'informations deviennent nombreuses et qu'il y a de plus en plus d'utilisateurs connectés au système. Cette complexité se traduit par une multitude de formes que prend un système d'informations pour s'adapter justement à cette complexité. Ainsi le schéma de base vu plus haut se décline en plusieurs types d'infrastructures réseau et de schémas de fonctionnement interne.

Nous avons voulu rassembler dans une matrice toutes nos considérations concernant les typologies possibles de Systèmes d'Informations suivant leur architecture réseau et leurs types de fonctionnement interne. Les types d'infrastructures réseau se déploient sur un axe qui dose le pouvoir sur l'information du pouvoir absolu, au pouvoir partagé entre les partenaires. Alors que les types de fonctionnement interne des Systèmes d'Informations se déploient plutôt sur un axe déterminé par les besoins en recherche d'informations des utilisateurs.

2 -1) Typologie des Infrastructures réseau

Les différents types d'infrastructures réseau peuvent en fait se distinguer en deux groupes suivant un axe de pouvoir plus ou moins partagé. D'un côté de l'axe, nous avons les infrastructures très centralisées, où les sources d'informations sont toutes ramenées en un point avec beaucoup de contraintes de format, de modèle,.. D'un autre côté de l'axe, nous avons les infrastructures réseau dites distribuées, où le « pouvoir » sur l'information est plus ou moins partagé. Un article du CERC a ébauché les types d'architecture réseau qui pouvaient se présenter pour la réalisation de son projet de réseau-santé, *Community Care Network* de Ouest-Virginie [JAGA 94].

- Les infrastructures réseau centralisées



On voit d'après cette architecture client-serveur que la base est centralisée, toutes les informations sont gérées en un seul point. Cela implique que les partenaires A, B et C cèdent totalement tout contrôle et tout pouvoir sur les informations provenant de leurs bases, et que celui qui gère le tout récupère à lui seul toutes les informations. Ce type d'architecture pose un sérieux problème de confidentialité et de sécurité de l'information.

Cette infrastructure implique également que les partenaires doivent transcrire toutes leurs informations dans un même modèle, celui du système central, car dans ce cas un même modèle relie toutes les sources. Si les sources d'informations préexistaient déjà au système, ce travail de transcription peut être considérable et peut entraîner des sur-coûts de développement et de réalisation.

- Les infrastructures réseau distribuées

Le mot « distribué » s'applique essentiellement au pouvoir de chacun sur la gestion et le contrôle de l'information, à croire qu'il est plus ou moins réparti entre les partenaires.

a) L'infrastructure distribuée

A l'extrême bout de l'axe du pouvoir distribué, nous avons une infrastructure réseau quasiment ingérable. Le schéma 1 présente un exemple d'infrastructure distribuée.

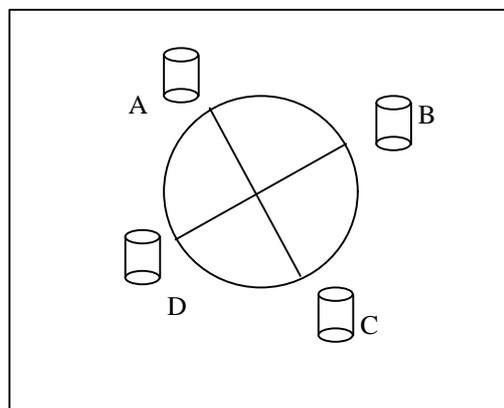


Schéma 1

Cette infrastructure sous-entend une parfaite autonomie des partenaires donc aucune ingérence dans les sources existantes, mais cela va générer aussi un grand nombre de problèmes, à savoir : aucune vision unifiée de l'information, risque de doublons, difficulté d'accès à l'information, problème de confidentialité et temps de réponse plus ou moins long suivant le nombre de partenaires et de sources d'informations. En fait, cette infrastructure préfigure ce qu'on pourrait appeler l'anarchie en terme de réseau.

Face à cette « anarchie », il existe d'autres types d'infrastructures réseau qui proposent des justes milieux au problème de l'autonomie de gestion et de contrôle de l'information.

b) L'infrastructure « fédérée »

Le réseau, dit fédéral, propose une médiation entre un modèle central et l'information distribuée (voir schéma 2).

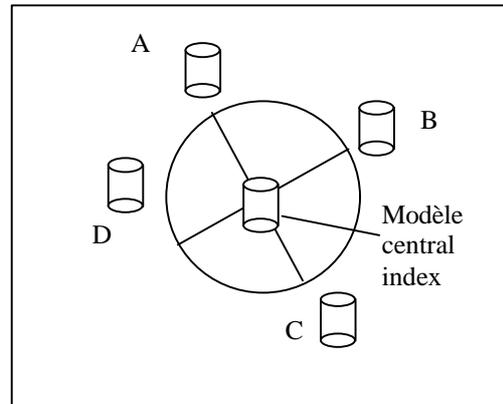


Schéma 2

L'autonomie des partenaires est préservée tout en gardant un contrôle centralisé sur l'accès et l'intégrité de l'information. Pour qu'un tel système marche, les partenaires doivent s'entendre sur les informations qui doivent être partagées pour assurer une sécurité et une certaine confidentialité. L'index de la base centrale représente ce qui est accessible dans chaque source d'informations d'après les concertations des partenaires au niveau de l'information à partager.

Cette infrastructure est un bon compromis entre autonomie et gestion centralisée de l'information car le pouvoir est ainsi distribué, mais elle demande cependant l'adhésion et la concertation des partenaires ainsi que la réalisation d'un index central global dont la programmation peut s'avérer très complexe suivant l'hétérogénéité des sources et le nombre de partenaires impliqués.

c) L'infrastructure médiane

Enfin, l'infrastructure qui laisse le plus de liberté à chacun reste l'infrastructure réseau « intelligente » (ou *mediated network*). Le modèle ainsi que les sources sont distribués (voir schéma 3) et c'est une sorte d'agent intelligent ou *mediator* qui coopère sur le réseau pour fournir un résultat de recherche d'informations attendu.

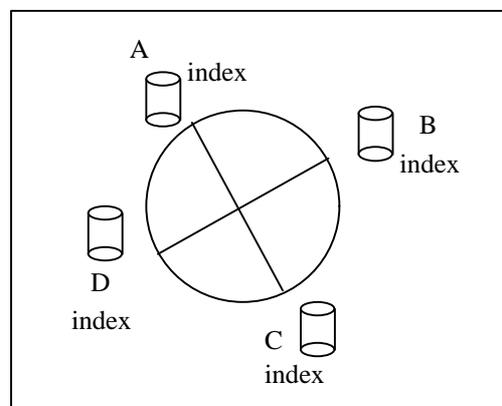


Schéma 3

Dans ce dernier cas, la liberté des partenaires est complète, ceux-ci gèrent à la fois les modèles de leurs bases, leurs informations et leurs index indépendamment des autres

partenaires. De plus cette infrastructure permet de s'adapter aux sources d'informations existantes.

Le pouvoir sur l'information reste ainsi distribué également sur chaque partenaire, ce qu'on pourrait appeler une infrastructure réseau « démocratique ».

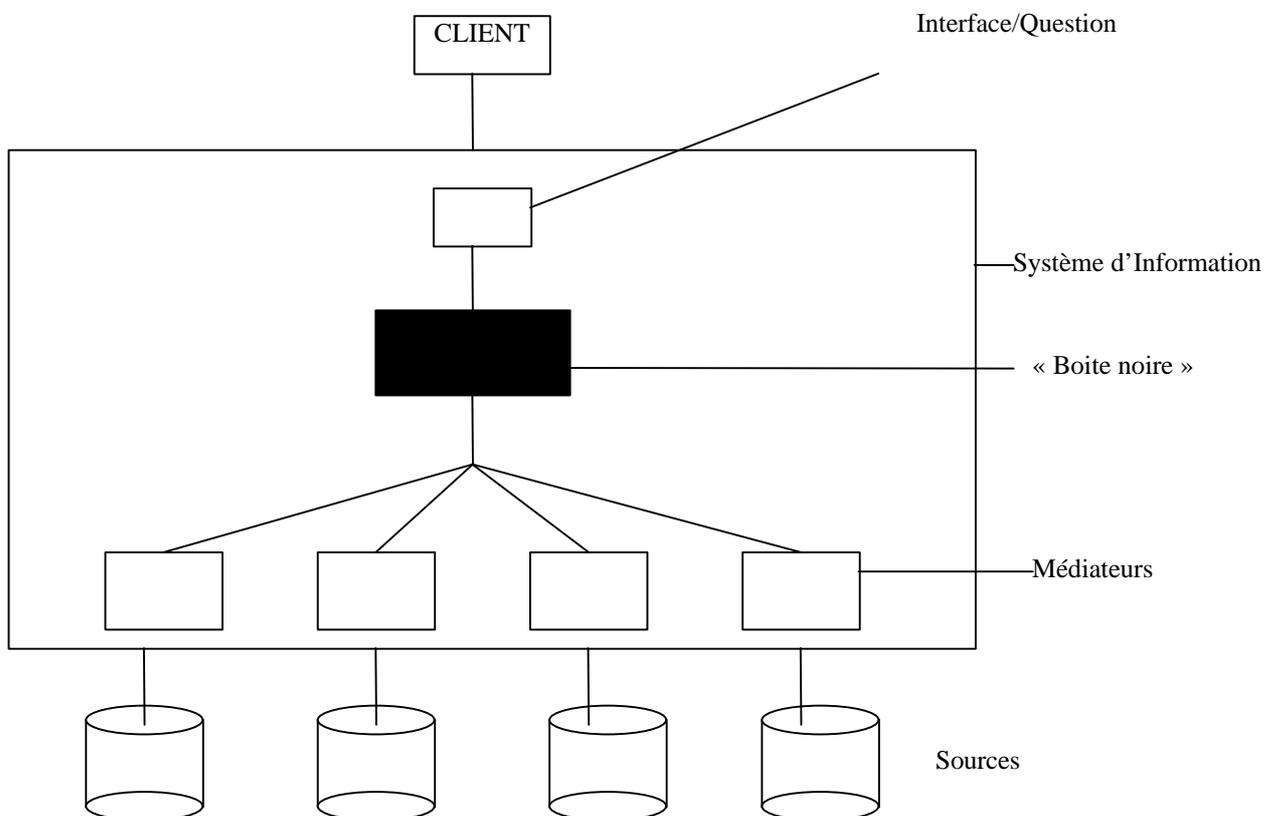
Pour choisir parmi ces différents types d'infrastructure réseaux, plusieurs critères doivent donc être pris en ligne de compte :

- le nombre de partenaires,
- la préexistence de bases d'informations, leur volume, leur format,...
- le coût de l'infrastructure en matériel (réseau câblé, logiciels, hardware,...),
- le pouvoir cédé plus ou moins sur la maîtrise de ses sources d'informations.

2 - 2) Typologie des fonctionnements internes des SI

Nous avons dit précédemment qu'il existait une typologie des Systèmes d'Information d'après la distribution du pouvoir sur l'information, mais ici, nous abordons aussi le fait que selon les besoins en informations des utilisateurs, nous pouvons avoir une organisation interne différente des SI.

Tout d'abord, schématisé, un Système d'Informations peut être représenté comme suit :



Nous pouvons en fait considérer deux grandes familles de Systèmes d'Informations, deux approches distinctes dans la façon de gérer et de se connecter à l'information. Selon Jennifer Widom [WIDO 95], chercheur à l'Université de Stanford, un Système d'informations procède en deux étapes :

- une question est posée, le SI détermine les sources d'informations appropriées qui peuvent répondre à cette question, formule un lot de sous-questions (*sub-queries*) pour chaque source,
- ensuite, les résultats sont obtenus des sources, les informations sont traduites, filtrées et rassemblées pour être données au client (utilisateur).

Ce processus, appelé aussi "approche à la demande" (*on-demand approach* ou *lazy approach*), suit un flux dynamique de la question aux sources d'information, des sources jusqu'à l'interface qui restitue les réponses.

Un autre processus existe dans le SI, plus communément appliqué pour les Data Warehouses :

- l'information qui peut être intéressante est extraite à l'avance de chaque source, elle est traduite, filtrée et fusionnée avec d'autres informations provenant d'autres sources, puis stockée dans un « hangar » ou warehouse, central.
- Quand une question est posée, les résultats sont évalués directement par le warehouse central, sans accéder aux sources originelles de ces informations.

Cette approche, appelée plutôt "approche anticipée" (*in-advance approach* ou *eager approach*) procède différemment de la première en ce sens que le flux dynamique part de la question à une base modèle centrale et revient vers le client (via un interface).

Suivant ces deux configurations, la Boite Noire du schéma de base du SI, représentant en fait le programme de fonctionnement du SI sera, dans le premier cas un adaptateur ou médiateur (le SI s'adapte aux sources d'informations hétérogènes), dans le deuxième cas un intégrateur (le SI intègre dans un même endroit toutes les informations des sources).

Ainsi, pour chaque type de fonctionnement de SI correspond une famille de besoins spécifiques.

Cas de l'utilisation d'un SI intégrateur (type Data Warehouse):

- Les besoins en informations sont connus, du même ordre et souvent répétitifs
- Les réponses sont immédiates (puisque'elles ne proviennent pas des sources)
- L'information est agrégée, filtrée, voire résumée, car ce qui prime c'est son utilisation par des décideurs.
- L'importance est mis sur l'information orientée vers l'analyse en profondeur, le SI fournit une photographie de l'information disponible à l'instant t (sorte de *snapshot*) où la mise à jour n'est pas essentielle. En effet, le SI intègre périodiquement les informations dans la base modèle.
- cette configuration permet de dépasser les liens relationnels et on-line entre les données et fournit un déroulement historique des informations, qui ne dépendent plus alors des sources mères.

Cas de l'utilisation d'un SI adaptateur :

- Les besoins en informations sont souvent imprévisibles, couvrant de larges domaines et s'échelonnant sur de multiples sources hétérogènes,
- Les sources d'informations sont très nombreuses, géographiquement éloignées et de toutes sortes de formats, de modèles,
- Les utilisateurs ont besoin d'informations à la pointe de l'actualité (*state of art*),
- Les questions sont nécessairement traitées de façon plus lentes, car les sources à contacter sont multiples.

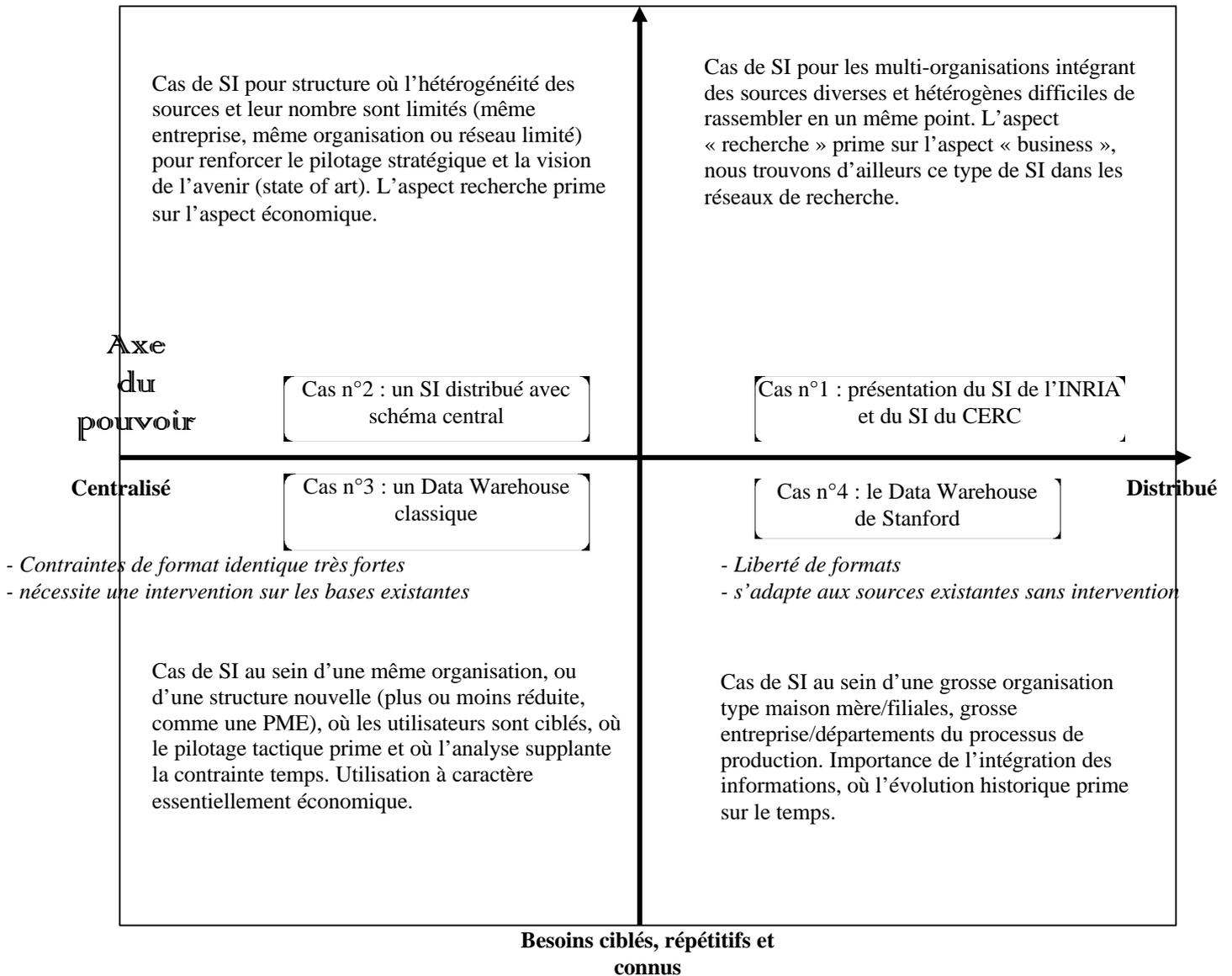
Nous voyons clairement que les avantages de l'un des systèmes représentent les inconvénients de l'autre. Si l'accent est mis sur l'immédiateté, l'historique et la profondeur des informations mais sans souci de mise à jour exacte ni de précision, alors le DW est approprié, ou si il est mis sur l'information à la pointe de l'actualité, couvrant des domaines de façon imprévisible, mais avec certaine lenteur de traitement, alors l'autre système sera choisi.

La matrice de positionnement des Systèmes d'Informations

Nous avons rassemblé ces considérations dans une matrice qui croise ces différentes typologies de Systèmes d'Informations, suivant un axe du **pouvoir sur l'information**, centralisé ou distribué, et un axe **des besoins en informations**, informations ciblées ou informations ouvertes, qui peut être également l'axe du temps, puisque l'une des différences essentielles entre les deux types de fonctionnement de SI repose sur la photographie (*snapshot*) ou l'accès en temps réel à l'information. Pour chaque case de la matrice sont également proposés des exemples concrets de Systèmes d'informations recueillis dans divers ouvrages.

Axe temps

Besoins libres,
imprévisibles



Avant de rentrer plus en détails sur les composants internes au Systèmes d'Information qui sont présentés ci-dessous, quelques définitions s'imposent sur les termes : système de gestion de bases de données, système de gestion de bases de données relationnelles, système de gestion de bases de données distribuées, système de gestion de bases de données hétérogènes [LI 92].

Système de gestion de bases de données (SGBD):

Un système de gestion de bases de données est composé d'une ou plusieurs sources de données rassemblées dans une ou des bases ainsi que d'un programme appelé le Système de Gestion de Bases de Données (SGBD). Le SGBD, d'une part, gère les données en modélisant leur rapport à la structure du système, puis d'autre part gère le rapport utilisateur/données par l'intermédiaire d'un interface graphique.

Système de Gestion de Bases de Données Relationnelles (SGBDR):

Ce système marche comme le précédent système avec ceci en plus que les bases de données sont reliées entre elles par des champs. Le système possède donc un module en plus qui rassemble et gère le modèle des relations entre données et entre bases.

Système de Gestion de Bases de Données Distribuées (SGBDD):

Dans ce cas là, les bases de données sont logiquement reliées entre elles et localement distribuées sur un réseau. Un programme spécial, le Système de Gestion des Bases de Données Distribuées (SGBDD), rend transparent cette distribution à l'utilisateur. Le SGBDD fonctionne en intégrant toutes les données distribuées dans un même modèle avec un même langage.

Système de Gestion de Bases de Données Hétérogènes (SGBDH):

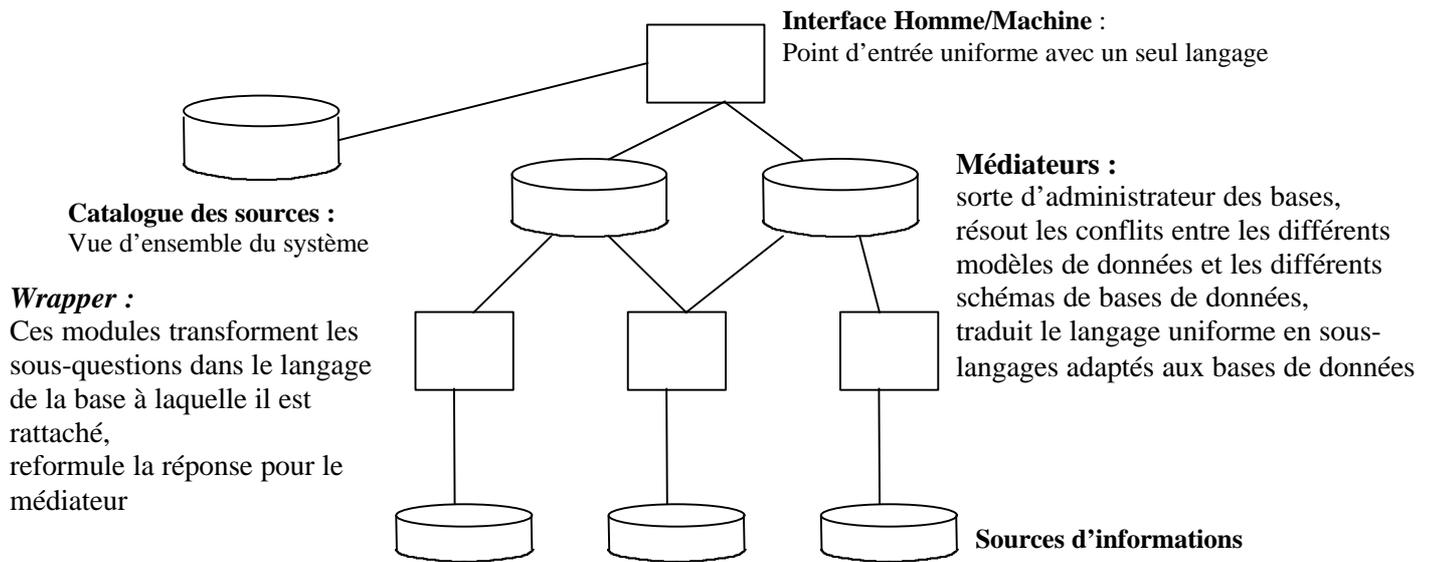
Ce système est en fait un Système de Gestion de Bases de Données Distribuées qui possède des composants hétérogènes entre eux, c'est-à-dire au niveau des bases de données, du langage ou de la structure.

Les cas n°1 et n°2 présentent deux Systèmes d'Information fonctionnant avec deux types distincts de Système de Gestion de Bases de Données Distribuées Hétérogènes (SGBDH). Dans le cas du premier, dont le schéma général est tiré des études de l'INRIA du projet Rodin et dont l'exemple particulier provient du centre de *Concurrent Engineering* de Virginie, le SGBDH s'adapte à n'importe quel source indépendamment, quel que soit les différentes structures des bases, les différents langages, les différents fichiers, etc...

Dans le second cas, le SGBDH procède en intégrant dans un même modèle toutes les modèles gérant les sources d'informations. Il crée en fait un méta-modèle des modèles des bases de données.

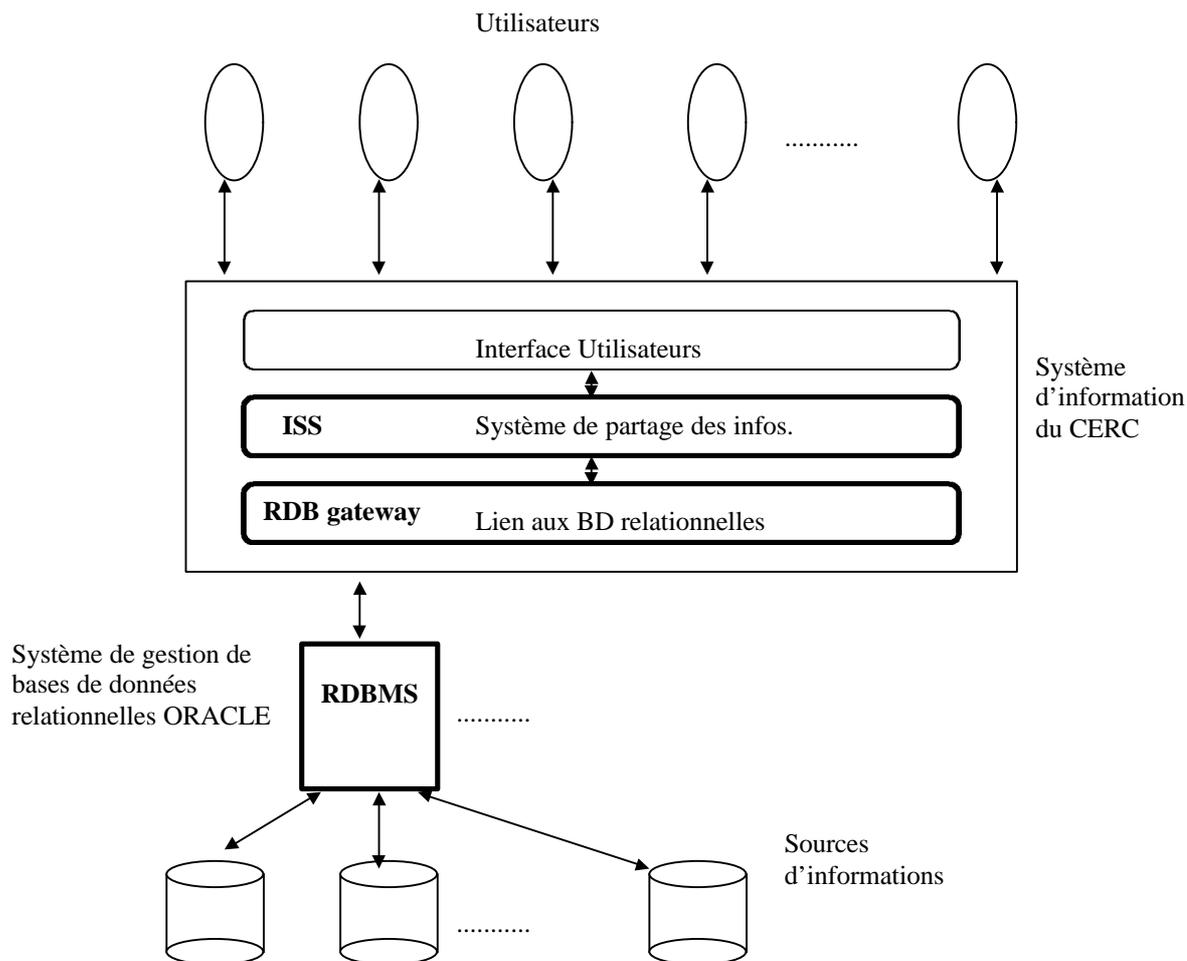
Cas n°1 : Présentation d'un schéma général tiré des études de l'INRIA et d'un cas concret de système d'informations avec Le Système d'Informations Partagées du CERC (Concurrent Engineering Research Center, Université de Ovest Virginie).

L'INRIA a développé dans un programme, le programme DISCO (*Distributed Information Search Component*) [VALD 95], un Système d'Informations permettant d'incorporer dans un même système des sources de données hétérogènes, avec des structures différentes, et rendant possible la gestion de ces informations (utilisation de bases existantes, création de nouvelles sources, transformation des requêtes,...).



La valeur ajoutée du programme DISCO est de fournir aux utilisateurs, aux programmeurs d'applications, aux administrateurs et aux développeurs de sources de données, un outil modulable et ouvert, avec un système de création de médiateurs entre un langage interface unique et des langages propres à chaque source d'informations (module appelé « *wrapper implementor* »).

Dans le même esprit, le schéma ci-dessous présente les grands modules qui forment le Système d'Informations Partagées du CERC ou (Information Sharing System), de l'utilisateur à la donnée source.



Plusieurs utilisateurs peuvent se connecter au système (simultanément ou en différé), l'interface propose une entrée de question et une sortie de réponses adéquates à la question posée. Le Système d'informations traduit la question en autant de sous-questions adéquates aux types de bases de données qui sont connectées au système. Le lien entre le système d'informations et les sources d'informations s'établit, dans le cas du système du CERC, par l'intermédiaire d'un système de management de bases de données relationnelles (RDBMS).

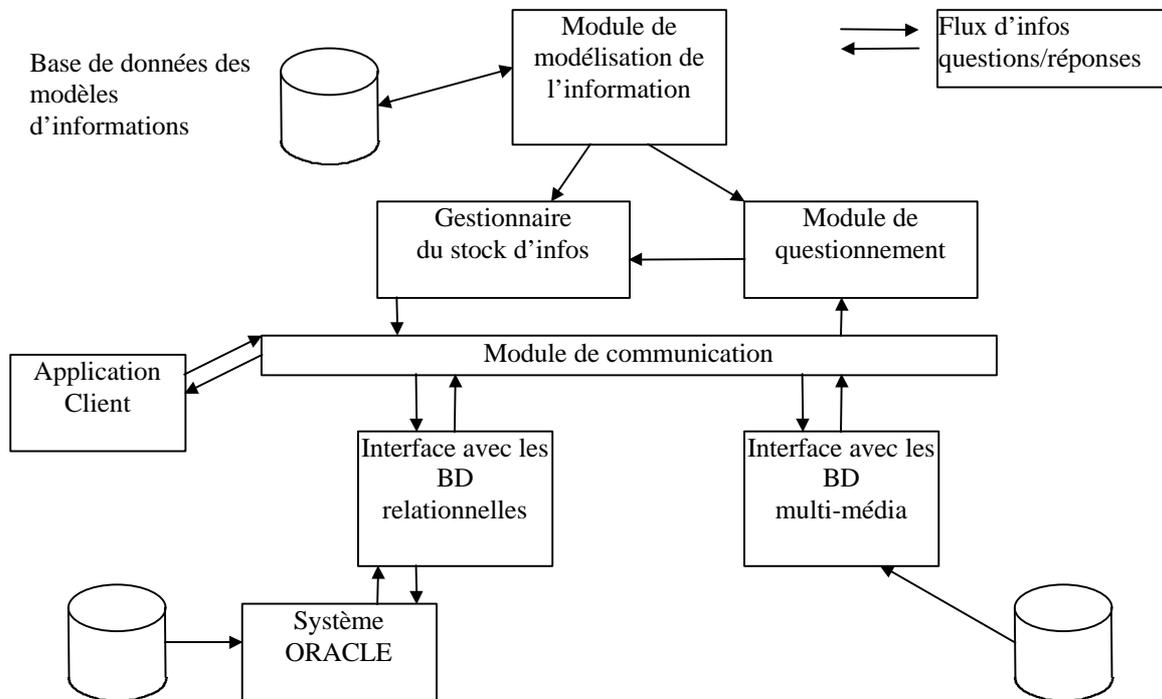
Dans le cas du CERC et contrairement au projet de l'INRIA, le choix s'est porté sur un médiateur spécifique existant qui est le système de management de bases de données relationnelle ORACLE. Ce choix a été conditionné par le fait qu'ORACLE est un des systèmes les plus connus et les plus utilisés dans le monde et avec lequel beaucoup d'applications sont développées. De plus, c'est un système capable de comprendre plusieurs formats et plusieurs structures de données.

Le but de ce Système d'Informations Partagées était, au départ du projet, de fournir au CERC un moyen de diffuser l'information sur les recherches en ingénierie concurrente (*concurrent engineering*). Or comme un certain nombre de sources d'informations préexistaient déjà, il était impératif de réaliser un Système d'Informations ouvert, capable de s'adapter à l'existant.

L'innovation, dans le cas du projet du CERC, a été de mettre au point un Système d'Informations relié à un système de management de bases de données relationnelles ainsi

qu'à des sources d'informations multi-média, tout ceci avec comme point d'entrée un même interface et un unique langage.

Pour comprendre plus en détails le fonctionnement du Système de Partage des Informations (ISS) réalisé au CERC, le schéma ci-dessous présente l'architecture du ISS.

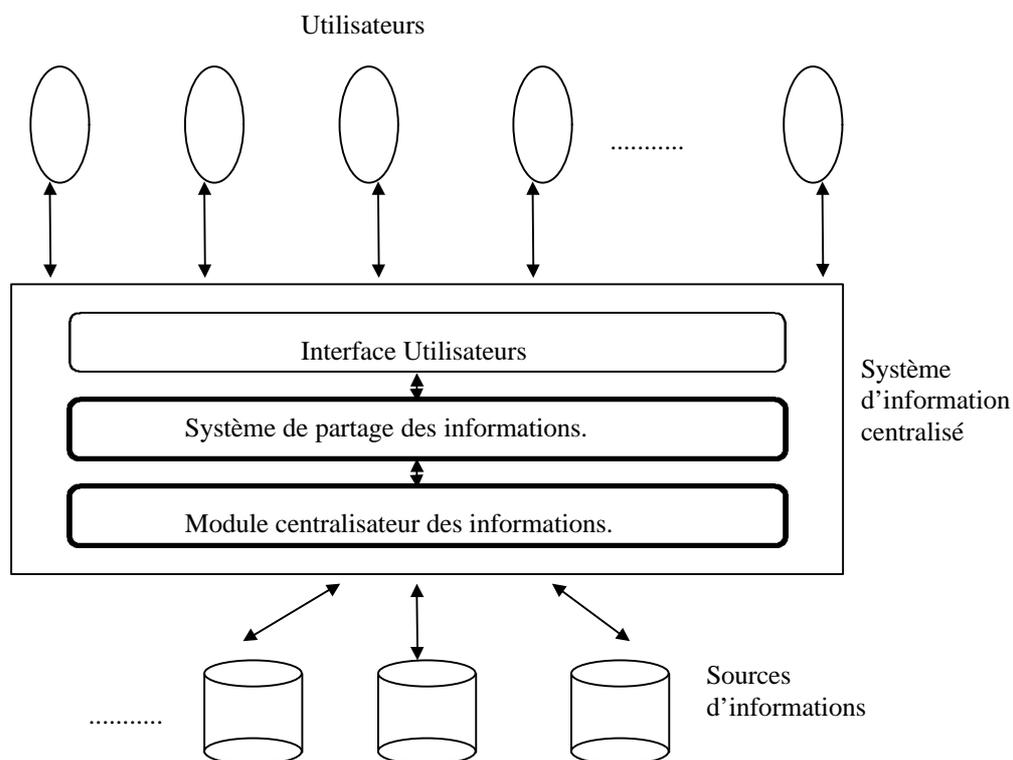


Cas n°2 : Un système d'informations pour sources hétérogènes à schéma unique.

Ce cas-ci relève plutôt de la deuxième approche des Systèmes de Management de Bases de Données Hétérogènes (SMBDH), où un schéma unique remplace le module multi-sources.

En fait, ce genre de Systèmes d'informations possède un modèle intégrateur qui définit un schéma unique pour toutes les sources d'informations. Ce schéma unique est, en quelque sorte, un méta-modèle de tous les modèles des sources. De ce fait, toute transaction passe nécessairement par ce modèle intégrateur, servant ainsi d'intermédiaire entre les sources et l'utilisateur.

Il est clair, aux vues du cas n°1, que ce type de SI est plus contraignant, en terme de flexibilité ainsi qu'en terme de programmation des transcriptions de langages, mais il a l'avantage de faciliter la gestion de la sécurité et de l'intégrité des informations et des flux, grâce au module centralisateur (voir schéma).



Il n'est pas difficile de comprendre que les différents types de Systèmes d'Informations présentés ci-dessus restent quand même peu accessibles aux entreprises, à d'autres organisations, voire au grand public, du fait de leur grande complexité. Le fait est que, seules, les grandes structures de recherche comme l'INRIA ou le CERC peuvent s'investir dans de tels projets, qui sont encore, pour la plupart, à l'état de bêta-test.

Il n'existe, à l'heure actuelle, aucun Système d'Informations qui soit totalement flexible et ouvert à l'hétérogénéité des sources existantes sur le marché et qui questionne ces sources en temps réel et on-line.

Les deux cas suivant présentant deux types de Data Warehouse et, bien que toujours dans le cadre d'un système intégrant des informations éparses, procèdent de façon différente.

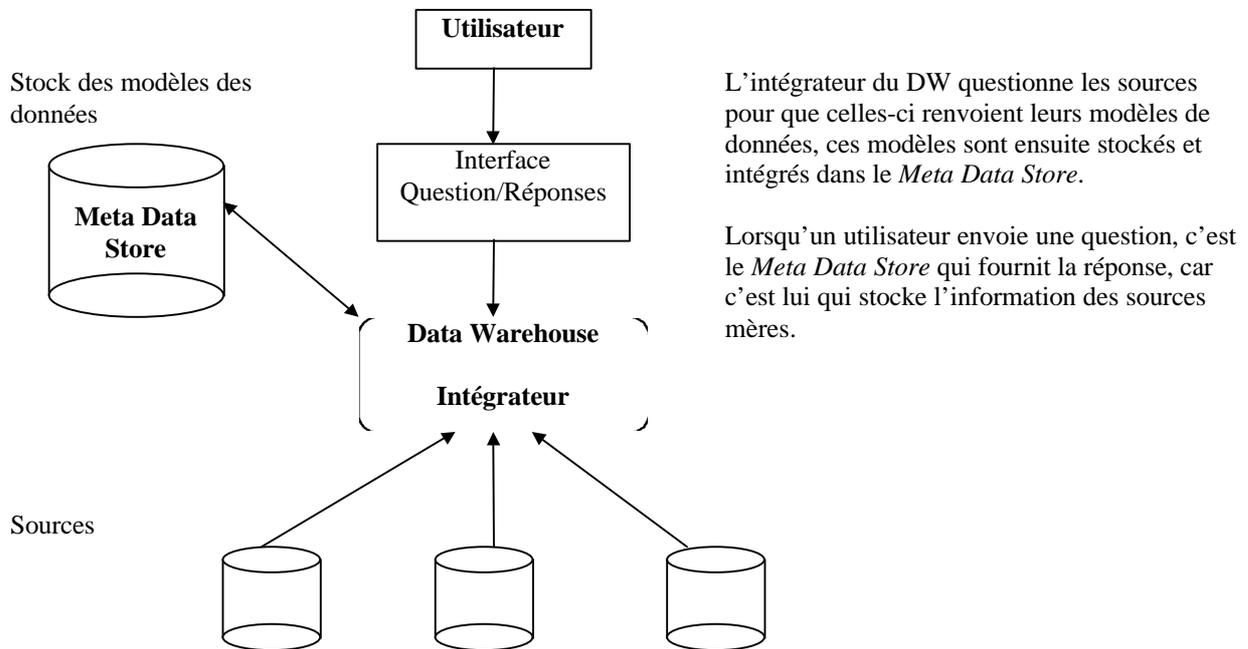
Alors que le challenge relèvé l'INRIA ou le CERC s'approche plutôt d'un challenge technique et informatique, les recherches sur les systèmes d'informations type Data Warehouse s'approchent certes d'un challenge technique mais aussi d'un challenge commercial. En effet, les Data Warehouses ont été développés dans le but d'assister une organisation dans la gestion de ses informations. A l'heure actuelle, plusieurs Data Warehouse sont proposés sur le marché (IBM, Cognos, Seagate,...).

Les Systèmes d'Informations type Data Warehouse:

Les Systèmes d'Informations type Data Warehouse (DW) ont été conçus pour soutenir la vie d'une organisation en terme de gestion des informations dans le temps ainsi qu'en terme de soutien à la prise de décision.

L'architecture d'un Système d'Informations type Data Warehouse se distingue des autres types précédemment vus, dans ce sens où, comme son nom l'indique, le Data Warehouse est

essentiellement composé d'un système de stockage des informations (« *warehouse* », pour traduire « entrepôt » de données).



L'intégrateur du DW questionne les sources pour que celles-ci renvoient leurs modèles de données, ces modèles sont ensuite stockés et intégrés dans le *Meta Data Store*.

Lorsqu'un utilisateur envoie une question, c'est le *Meta Data Store* qui fournit la réponse, car c'est lui qui stocke l'information des sources mères.

Le maître mot du Data Warehouse, c'est de rendre facile la gestion de l'information, et ce sur plusieurs plans.

D'une part, les informations sont classées par sujet génériques (ex : comptabilité, clients, fournisseurs, normes,...) et non pas par fonction (marketing, finance, production,...). Cette structuration implique que les informations sont **filtrées** ; celles qui ne sont pas utilisées par le décideur et son équipe ne sont pas du ressort du DW.

D'autre part, le Data Warehouse a cette particularité d'**intégrer** les informations, c'est-à-dire qu'il rassemble dans un même schéma de structure tous les types de modèle d'informations qu'il rencontre. Par exemple, si dans la source d'informations A, le genre Masculin/Féminin est symbolisé par les codes M/F, dans la source B, les codes sont 0/1, dans la source C, les codes sont X/Y, le Data Warehouse transforme les divers codes et donne un code général (M/F par exemple) [DATA].

L'interrogation est ainsi simplifiée, les utilisateurs utilisent leur propre langage sans avoir la connaissance des modèles, des formats et des langages des différentes sources.

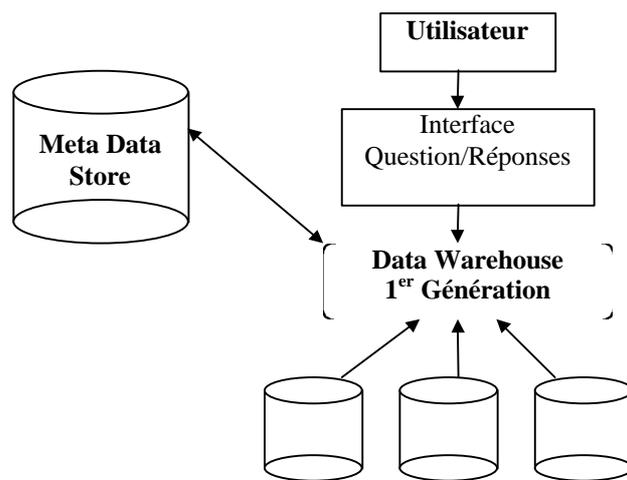
Enfin, la gestion globale des flux d'informations se trouve, dans le Data Warehouse, nettement **simplifiée**, puisqu'il n'existe seulement que deux flux d'informations, l'un qui est la mise en stock des données (photographie des sources à l'instant t), l'autre qui est la visualisation des réponses aux questions posées.

Cette dernière particularité implique, nous l'avons déjà dit, d'une part qu'il n'y a pas de mise à jour d'enregistrements en temps réel comme dans un Système d'Informations type cas n°1

et n°2, d'autre part, l'accès à l'information se fait de façon rapide, car la recherche d'information est optimisée par leur mise en stock.

Il semble, à l'heure actuelle, que le fossé entre les Systèmes d'Informations reliés en temps réel aux sources et les Data Warehouses se rétrécit de plus en plus. En effet, la nouvelle génération de DW évolue vers une forme de Stock d'informations distribuées, où l'accès aux sources en temps réel semble possible [HAMM 95]. Les schémas suivants montrent respectivement le cas d'un DW de la première génération (cas n°3) et le cas du data Warehouse développé à l'université de Stanford (cas n°4).

Cas n°3 : un Data Warehouse de la première génération, type classique



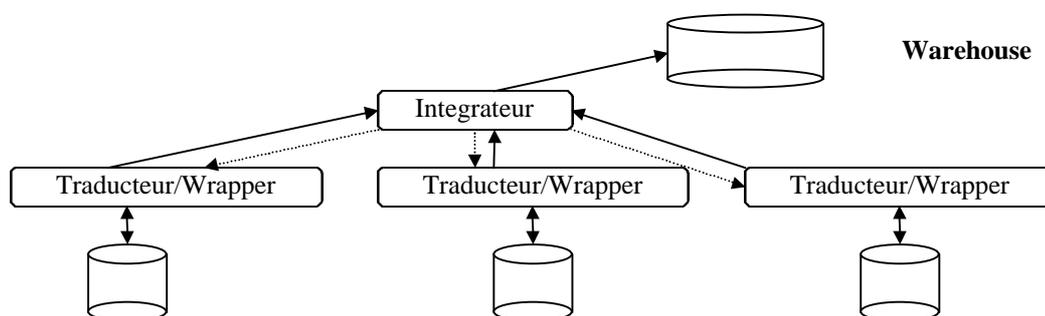
Ce type de DW fonctionne toujours sur le même principe du stock d'informations.

Cependant, il exige de fortes contraintes :

- Les sources sont modifiées dans le but d'intégrer le DW (modification des modèles, des codes, résumés d'infos,...). Cette contrainte sous-entend que l'organisation qui installe ce type de DW, soit possède peu de bases existantes à modifier, soit commence à construire ses sources avec l'installation du DW.
- Aucune accession aux informations du Meta Data Store n'est possible (pas de mise à jour ni de modification possible).

Ce DW est en fait très centralisé et rigide, mais cette rigidité peut répondre à un certain type de besoins, de petites PME, où les utilisateurs sont ciblés et peu nombreux.

Cas n°4 : Le Data Warehouse de l'Université de Stanford



Source : J. Widom, « Research problems in Data Warehousing », Stanford University, 1995

...

L'intégrateur est le module qui gère l'agrégation des données à transmettre dans le Warehouse.

Les traducteurs ou *Wrappers* dépendent du type de source. Ce sont eux qui traduisent les informations de chaque source en information intégrée comprise par le DW. Ainsi, si le DW fonctionne en mode relationnel et qu'une source fonctionne en mode tableur (ligne/colonne), le traducteur permet la lecture du tableur en mode relationnel. De plus ce module est chargé de détecter tout changement dans la source susceptible d'intéresser le DW.

Ces modules sont à rapprocher des Wrappers du cas n°1 de l'INRIA. Ce sont eux qui permettent de faire évoluer les DW vers des Systèmes d'Informations gérant les données en temps réel. Ils rendent notamment possible la mise à jour régulière des informations (pour éviter le phénomène de photographie à l'instant t) et la modification des informations à la source (rajoutant ainsi plus de flexibilité dans la gestion des flux), tout en préservant une rapidité de traitement des questions (car le stock d'informations sert toujours).

Conclusion

Si il existe aujourd'hui une typologie distincte des Systèmes d'Informations, du Data Warehouse au système ouvert on-line, il semble, d'après les recherches actuelles dans ces domaines (Inria, Cerc, Stanford), que ceux-ci vont évoluer vers une optimisation de tous les avantages qu'offrent ces deux types de systèmes. Les chercheurs approfondissent, en effet, la mise au point de Systèmes complexes, mélange de Data Warehouse à composante historique couplé d'un système d'informations opérationnel on-line.

Déjà, les nouvelles générations d'architectures de DW tentent de gommer l'effet centralisateur et complexe des premières générations, en intégrant la distributivité des données dans des "Datamarts" répartis par fonction de l'entreprise [01INF 98].

Aujourd'hui enfin, la fusion des deux types "accès on-line" et Data Warehouse est quasiment accomplie dans le dernier sorti "Data Web", fusionnant les technologies intranet/internet avec le client-serveur et rendant accessible les données décisionnelles via un browser classique.

Ces nouvelles évolutions vont quelque peu changer la donne au niveau de la typologie des Systèmes d'informations. C'est certainement en suivant les évolutions du marché comme les partenariats et alliances stratégiques entre les éditeurs traditionnels d'outils de requête client/serveur (type Cognos, Seagate Software ou Andyne) et les éditeurs d'outils web (type Zanza Software, Infospace ou Data Dynamics)[DATA 98] que l'on pourra trouver de nouvelles orientations et de nouvelles pistes de recherche.

Bibliographie

MELEZE Jacques, « Approche systémique des organisations » 1969

BOUSSAGOL Hélène "Des systèmes informatiques aux systèmes d'informations" SOSI CNRS 1996

BOUZEGHOUB .M "Synthèse des méthodes et des outils d'aide à la conception de systèmes d'Informations", rapport de recherche n°258 INRIA déc 83

DOU Henri , « Veille technologique et compétitivité », p 91, chez Dunod, 1996

JAGANNATHAN V. « Architectural alternatives for Community Care Networks », CERC Technical Report, April 1994

WIDOM Jennifer « Research problems in Data Warehouse », Proc. of 4th Conference on Information and Knowledge Management, nov. 1995

LI X, « A relational database gateway for an information sharing system », CERC, 1992

VALDURIEZ P &, « Scaling heterogeneous databases and the design of Disco », Rapport de recherche n°2704 - Projet Rodin, nov 95

« Data Warehouse Environment », [http ://carbon.cudenver.edu/~jkarimi/is6800/jan30.html](http://carbon.cudenver.edu/~jkarimi/is6800/jan30.html)

HAMMER J & « The Stanford Data Warehousing Project », 1995, [http ://www.db.stanford.edu](http://www.db.stanford.edu)

"Choisir une architecture adaptée à ses besoins", article 01 Informatique n°1487, 27 février 98

"Data web : le décisionnel s'émancipe" Informatiques Magazine - Janvier 1998