

SIMES - 961620  
Système d'Information Multimédia  
Pour l'Environnement Subsaharien

# Spécification de l'Architecture du Système d'Information SIMES

---

*Deliverable number : D1.1*

*Nature:P*

*Contractual Date of Delivery: 14 March 1998*

Task WP3.1 : Specification of the common software bases

*Nom du rédacteur : Jean-Claude Derniame*

*Institut : Loria*

*E-mail : [Jean-Claude.derniame@inria.fr](mailto:Jean-Claude.derniame@inria.fr)*

## **Abstract:**

Cette première version présente une spécification informelle du support du système d'information SIMES (Système d'Information Multimédia pour l'Environnement Subsaharien). Le système d'information SIMES peut être appréhendé au travers de deux visions : une vision assez analytique, centrée sur les composants, qui fait du système SIMES une boîte à outils, et une vision où l'aspect fédérateur de composants logiciels est mis en avant. Le bus logiciel qui va supporter les spécifications est présenté.

This document presents an informal specification of the Information System SIMES (Multimedia System Information for Subsaharian Environment). SIMES can be seen in two different views : an analytic one focused on its components making SIMES be a tools box, and a global one where the federative aspect of the software components is enhanced. Here, we present the Software bus.

## **Keyword List:**

Système d'Informations, Fédération des composants, Méta-modélisation, Bus Corba, Internet/Web.

Information System, Components integration, Meta-modelling, Corba bus, Internet/Web



# Table des matières

<b>1. SYSTÈME D'INFORMATION SIMES .....</b>	<b>5</b>
1.1. INTRODUCTION.....	5
1.1.1. <i>Présentation du projet SIMES</i> .....	5
1.1.2. <i>Notre vision du système</i> .....	5
1.2. SIMES BOÎTE À OUTILS GÉNÉRIQUES.....	5
1.2.1. <i>Les composants sources d'informations</i> .....	6
1.2.2. <i>Les composants outils</i> .....	7
1.3. VISION FÉDÉRATION DE COMPOSANTS.....	7
1.4. ANALYSE .....	8
1.4.1. <i>La représentation et la gestion des connaissances dans le système</i> .....	8
1.4.2. <i>Le fonctionnement des composants logiciels</i> .....	9
1.5. LES MÉTA-INFORMATIONS.....	13
1.5.1. <i>La définition des méta-informations</i> .....	13
1.5.2. <i>La description des méta-informations</i> .....	13
1.5.3. <i>Le stockage des méta-informations</i> .....	13
1.5.4. <i>L'acquisition des méta-informations</i> .....	13
1.6. CONCLUSION ET PERSPECTIVES. ....	14
<b>2. LE BUS LOGICIEL SIMES .....</b>	<b>15</b>
2.1. INTRODUCTION.....	15
2.2. BACKGROUND .....	15
2.2.1. <i>Présentation de PCIS-2</i> .....	15
2.2.2. <i>Collaborations Internationales du LORIA</i> .....	15
2.2.3. <i>Architecture à la CORBA</i> .....	15
2.3. PROPOSITION SIMES .....	16
2.3.1. <i>PCIS 2 versus SIMES</i> .....	16
2.4. DISCUSSION D'IMPLÉMENTATIONS.....	17
2.4.1. <i>Collecte centralisée</i> .....	18
2.4.2. <i>Collecte distribuée</i> .....	19
2.5. PERSPECTIVES .....	20
<b>REFERENCES .....</b>	<b>21</b>

## Liste des figures

Figure 1-1 Composants du système d'information SIMES	10
Figure 2-1 "The three tiers architecture"	17
Figure 2-2 Architecture de l'infrastructure SIMES	18
Figure 2-3 Une première vue géographique de l'architecture	19
Figure 2-4 Une deuxième vue géographique de l'architecture	19



# 1. Système d'Information SIMES

## 1.1. Introduction

### 1.1.1. Présentation du projet SIMES

Le projet SIMES WISE-DEV "Système d'Information Multimedia pour l'Environnement Subsaarien" est un projet financé conjointement par l'Union Européenne et la Banque Mondiale. Il se situe dans le cadre du programme Européen INCO-DC dont l'objectif est de développer un partenariat avec les pays en développement.

SIMES WISE-DEV vise à associer des partenaires africains et européens pour l'application et l'adaptation des nouvelles technologies de l'information à la compréhension et à la maîtrise de l'environnement subsaharien. Ses objectifs scientifiques et technologiques devront favoriser les compétences particulières en informatique environnementale.

Les thèmes abordés dans l'ensemble du projet sont les suivants :

- Acquisition et traitement d'information
- Bases de données
- Modèles statistiques
- Traitement d'images
- Intelligence artificielle
- Systèmes d'information
- Génie logiciel

### 1.1.2. Notre vision du système

Le système d'information SIMES, en ce qui concerne notre contribution, peut être appréhendé au travers de deux visions : une vision assez analytique, centrée sur les composants, qui fait du système SIMES une boîte à outils, et une vision où l'aspect fédérateur est mis en avant.

Ces deux visions sont présentées respectivement dans la section 1.2 et la section 1.3. L'architecture de l'infrastructure qui doit supporter ces propositions est présentée dans le deuxième chapitre .

## 1.2. SIMES boîte à outils génériques

Dans le document destiné à l'Union Européenne, il est stipulé que les méthodes et outils conçus pour le système SIMES sont destinés à être réutilisés pour la création d'un réseau de Systèmes d'Information sur l'Environnement (S.I.E.) Subsaarien. Le système SIMES apparaît de ce fait d'une part comme une boîte à outils intégrée sur le Web, d'autre part comme une architecture support d'intégration des outils.

Plus précisément, le système est vu comme un ensemble de composants dont certains (notamment les outils) sont conçus de façon générique et de manière à être greffés/intégrés/ pluggués facilement et rapidement dans un système ouvert et évolutif. Ces composants peuvent être grossièrement répartis en deux catégories, les composants sources d'information et les composants outils.

### 1.2.1. Les composants sources d'informations

Ce sont des composants qui contiennent et fournissent de l'information au système SIMES. Ils peuvent être répartis en deux catégories :

- les sources d'informations qui sont gérées par des organisations partenaires de la mise en place d'un système de type SIMES et dont l'autonomie doit, de ce fait, être respectée. Rentrent dans ce cadre, les bases de données avec des données hydrologiques, des images satellites, des références documentaires, des photographies de paysages ou encore des données issues d'enquêtes. Il pourra également s'agir de sites Web, de S.I.G1, d'un répertoire de fichiers ou même d'un document ou fichier non archivé dans une base de données (exemples : article d'un chercheur, rapport d'étude).
- les sources d'informations propres au système SIMES/WISE-DEV qui sont en fait des bases de données où sont capitalisées des informations extraites des sources d'informations de la catégorie précédente.

Les opérations pilotes prévues dans le cadre du projet SIMES apporteront deux sources d'informations gérées par des partenaires du projet SIMES. A savoir, l'observatoire de la pêche dans le Delta Central du Niger et l'observatoire de la vallée du fleuve Sénégal.

#### 1.2.1.1. L'observatoire de la pêche dans le Delta Central du Niger

Ce système est opérationnel depuis fin 1996. Il a été conçu à la suite des recommandations finales des recherches menées par le Grand Programme (ORSTOM) de recherche sur le Delta Central du Niger. Il s'agit d'un système de collecte et de restitution rapide d'informations pertinentes sur les ressources en poissons et son exploitation, intégrant le comportement des acteurs et les modes de production.

- Ce système basé à l'IER à Mopti (Mali) consiste en :
  - ❖ une série d'enquêtes répétées sur trois zones échantillon du Delta par trois enquêteurs de l'IER et de l'OPM. Ces enquêtes concernent respectivement l'état d'occupation de 72 villages et campements (fréquence bimestrielle), l'activité et la mobilité des ménages (fréquence bimestrielle), les débarquements des pirogues de pêche sur une quinzaine de sites et les mises en place de dispositifs fixes (environ 70 barrages).

une base de données installée sur un PC et développée avec le logiciel Fox-Pro. Cette base de données relationnelle contient actuellement 12 tables décrites dans le rapport d'études "Mise en place d'un système de suivi de la pêche dans le Delta Central du Niger : Concepts et méthodes" [4].

des procédures systématiques, mais non encore automatisées pour la constitution de synthèses statistiques, graphiques, textes, l'ensemble aboutissant à la production d'un bulletin (papier) périodique bisannuel destiné principalement aux décideurs régionaux et nationaux.

un site Web qui contient des informations générales sur les zones de suivi (hydrographie, informations de cadrage initial, images de barrages de pêche, etc...) ainsi qu'une description du contenu des bulletins. Ce site donne aussi accès aux différents bulletins semestriels édités jusqu'ici, ainsi qu'à leurs notes de conjoncture respectives. Des graphiques (présentés sous forme de séries temporelles) sur les effectifs de ménages et les techniques de pêche dans les différentes localités enquêtées sont également proposés. Le

contenu et la présentation de ce site sont encore en cours d'évolution. L'URL du site est : <http://orleans.orstom.fr/ext/pechedcn>.

des données bioécologiques (descriptif morphologique du paysage, caractéristiques biométriques et état physiologique des poissons) collectées sur trois zones d'échantillonnage sont archivées dans des fichiers Excel.

La base de données et le site Web sont donc tous les deux des sources d'informations et de données potentielles pour SIMES.

### **1.2.1.2. L'observatoire de la vallée du fleuve Sénégal**

Cet observatoire est en cours de mise en place. Il a pour objectif d'étudier les différentes problématiques de gestion environnementale présentes dans la vallée du fleuve Sénégal. En collaboration avec des organismes travaillant sur cette zone, il est prévu une collecte de données assez variées : données hydrologiques, données météorologiques, données pédologiques et géomorphologiques, données sur les aménagements et les infrastructures, données agricoles et sylvo-pastorales. Il est également prévu d'utiliser les images fournies par METEOSTAT pour des applications telles que l'estimation de la biomasse, la croissance végétale, les rendements agricoles, ou encore le suivi des feux de brousse.

### **1.2.2. Les composants outils**

On distingue plusieurs types d'outils, comme les présente la figure 1-1:

- ❖ des outils d'acquisition de l'information : extraction et intégration de l'information.

des outils de gestion de la connaissance : représentation, gestion et exploitation de la connaissance au sein du système afin de faciliter la recherche d'information, l'extraction de la connaissance.

des outils de traitement de l'information : traitement statistique, traitement d'images, modélisation, etc...

des outils de diffusion de l'information : gestion de l'interface utilisateur, des profils utilisateurs, des notifications.

des outils d'administration et de coordination du système.

En conclusion, dans cette première vision, l'accent est mis sur les caractéristiques des composants logiciels : indépendance et autonomie pour ce qui est des sources d'informations gérées par des partenaires, généricité, interopérabilité et portabilité pour les outils.

### **1.3. Vision fédération de composants**

Dans cette seconde approche du système SIMES, on se focalise beaucoup plus sur l'aspect fédération de sources d'informations ainsi que sur celui de l'administration et la coordination des différents composants.

Dans l'environnement comme dans beaucoup d'autres domaines où les études et les recherches sont fréquentes, l'existant en terme de systèmes archivant, traitant ou produisant de l'information est important. Pour cette raison, le système SIMES ne se veut pas être un système "*pensé et conçu du début jusqu'à la fin*". Il s'agit plutôt d'un système d'information pouvant s'appuyer sur des systèmes existants et capable d'évoluer par l'intégration (ou le retrait) de nouveaux composants. Par conséquent, un système instancié du système SIMES peut être vu comme une fédération de composants existants ou à venir.

En général, l'existant en termes de sources d'information est divers (bases de données, sites Web, SIG, etc.) et réparti sur des sites géographiquement distincts. On est donc en présence de systèmes de logiques et de structures différentes (par exemple : bases de données et site Web). Il ne s'agit pas comme c'est parfois le cas dans les SGBD répartis, de donner une vision unique de plusieurs bases de données en construisant un schéma de

données global. Dans le cas du système SIMES, il s'agira d'extraire en partie ou en totalité des informations contenues dans les systèmes locaux et de les intégrer dans un référentiel de connaissances tenant compte des spécificités des problématiques environnementales. De plus, de la valeur ajoutée pourra être apportée à ces informations en y appliquant des traitements (statistiques, traitements d'images, etc.).

Cette fédération fait apparaître deux niveaux de systèmes :

- des systèmes ayant une existence propre en dehors de celle du système SIMES : ce sont en général les sources d'informations gérées par des partenaires que nous appellerons par la suite *observatoires fédérés* ou *systèmes fédérés* dont on ne doit pas contraindre le fonctionnement.

Un système fédérateur : l'observatoire ou *système fédéral* qui représente en fait le système SIMES et qui comprend tous les composants : ceux qui contiennent de l'information, ceux qui permettent de l'acquérir, de l'archiver, de la traiter et de la diffuser.

Chaque système fédéré est une composante du système fédérateur SIMES en ce sens qu'il alimente ce dernier en informations. Mais ces systèmes fédérés doivent être relativement découplés au système fédérateur pour que leur autonomie et leur indépendance ainsi que l'évolutivité du dispositif entier soient garanties. Les difficultés dans la conception du système SIMES vont résider dans :

- la gestion de la cohérence entre l'observatoire fédéral et les observatoires fédérés notamment la capacité à respecter l'autonomie des observatoires fédérés, et à concevoir un système fédérateur qui survive à l'évolution des systèmes fédérés ainsi que la gestion des mises à jour.
- la représentation et la gestion des connaissances
- l'évaluation de la pertinence des informations
- l'administration et la coordination des composants logiciels du traitement de l'information (statistiques, traitement d'images, modélisation, etc...).

Au total, ces deux visions (boîte à outils et fédération de composants) ne sont pas incompatibles mais se complètent et permettent d'apprécier toutes les particularités de ce système.

## **1.4. Analyse**

Le système SIMES fait apparaître deux niveaux d'analyse liés :

### **1.4.1. La représentation et la gestion des connaissances dans le système**

Il s'agit de définir un système de représentation et de gestion des connaissances dans SIMES. Ce sont ces connaissances qui permettront d'établir des liens/relations entre les différentes informations stockées. Ce système constituant le noyau à travers lequel se fera l'extraction de connaissances des informations capitalisées, il se doit donc d'être souple et évolutif (les problématiques environnementales s'enrichissant ou évoluant fréquemment). Sa conception est compliquée par l'exigence d'une absence de contraintes sur les systèmes sources d'information.

Elle nécessite qu'on élucide les questions relatives aux types d'informations extraites et à la définition des méta-informations associées (voir section 1.5), à la manière d'extraire l'information (extracteurs), à la définition d'une ontologie spécifique permettant de classifier l'information, aux systèmes d'indexation.

Pour la gestion de l'information dans le système, il apparaît nécessaire de définir :

- des protocoles d'extraction et d'intégration des informations et des méta-informations

un système de classification et d'indexation des connaissances. On peut penser que les recherches pourront explorer le domaine d'ontologies (par exemple : utilisation d'un référentiel basé sur une ontologie des méta-informations du système) ou encore l'analyse statistique de co-occurrences de mots dans des documents textuels.

des outils de gestion de cette métaconnaissance qui pourront être des agents (utilisant par exemple des techniques de l'IA, du traitement naturel ou encore du "datamining" pour effectuer des inférences sur les méta-informations).

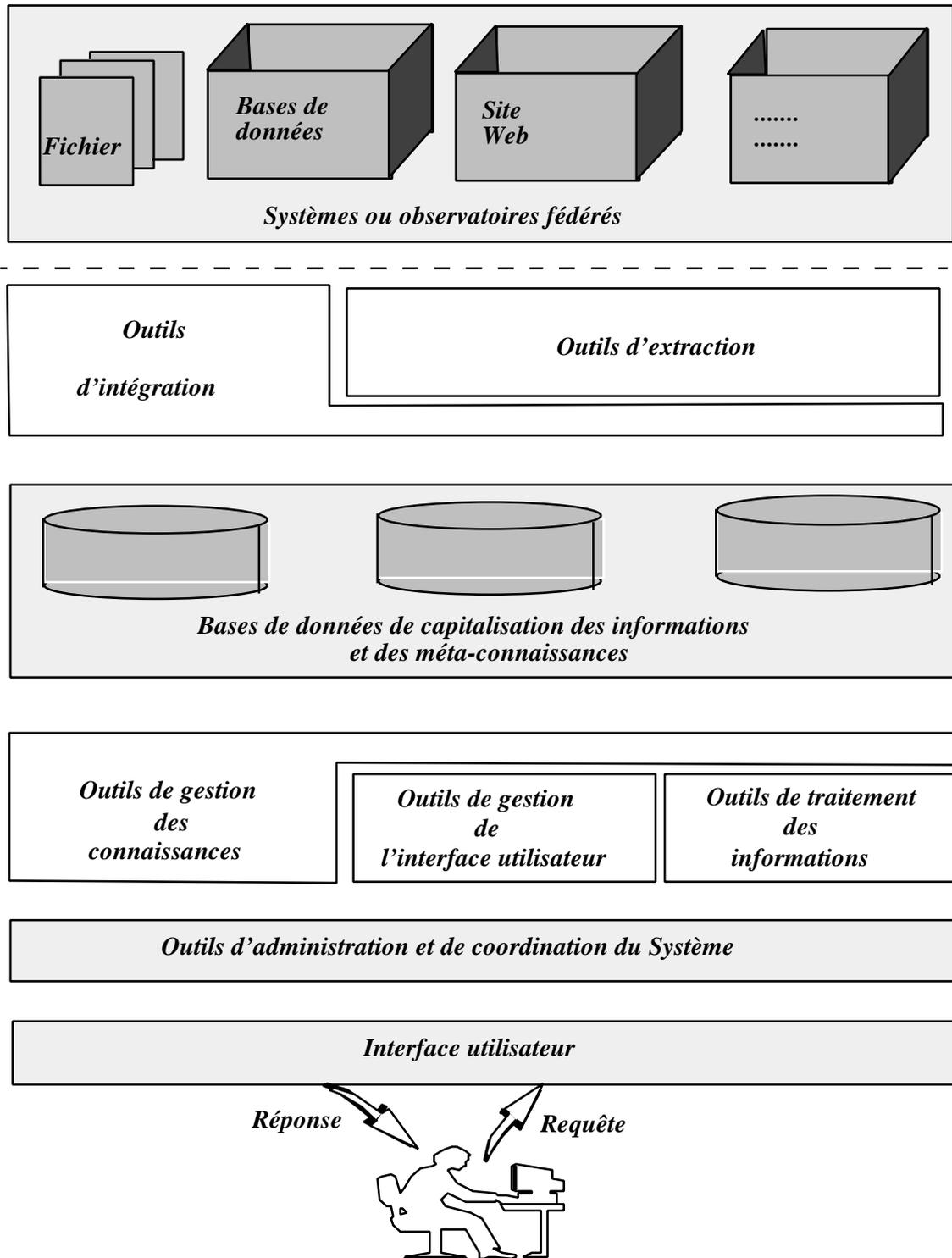
#### **1.4.2. Le fonctionnement des composants logiciels**

A ce niveau, il s'agit de spécifier les composants logiciels et de décrire comment ces composants coopèrent entre eux. La métaphore du bus logiciel qui est présentée au chapitre 2 est une solution architecturale intéressante pour ce niveau.

Nous allons brièvement présenter, en nous servant du schéma en couches du système illustré par la figure 1-1, quelques outils qui permettent d'élaborer les réponses à fournir aux utilisateurs.

##### **1.4.2.1. Outils d'extraction de l'information**

Figure 1.1 : Composants du système d'information SIMES



Il s'agit principalement d'outils (extracteurs) qui vont recueillir de l'information de manière périodique ou qui sauront aller la retrouver lors de l'élaboration de la réponse à la requête d'un utilisateur. On peut imaginer qu'il y aura autant de types d'extracteurs que de types de systèmes que l'on souhaite fédérer.

#### 1.4.2.1.1. Le concept d'extracteur

L'acquisition suppose un moyen d'extraire l'information des sources et donc la définition du concept d'extracteur. Comme son nom l'indique, l'extracteur est un composant logiciel qui sera utilisé pour :

- établir une passerelle entre un composant observatoire fédéré et les composants du module d'administration et de coordination du système. Ainsi, chaque système observatoire fédéré disposera d'un extracteur. A chaque système fédéré sera associé un extracteur, instancié d'un des extracteurs types et spécialisé en utilisant les informations spécifiques au système fédéré.

récupérer des données ou documents du composant observatoire fédéré (BD, sites Web, etc.) et éventuellement des méta-informations qui faciliteront leur classification ou indexation. L'information extraite est ensuite confiée à un outil d'intégration.

Comment cet extracteur pourra-t-il fonctionner?

Un extracteur peut être vu comme un agent "proxy" de la source d'information sur laquelle il travaille. Il est créé (par instanciation) et initialisé à chaque ajout d'un nouveau composant de type observatoire fédéré.

#### **Création et initialisation**

L'instanciation d'un composant extracteur se fait par la déclaration d'un nouvel observatoire fédéré au module d'administration. C'est au cours de cette inscription que les informations d'ordre général sur l'observatoire fédéré et sur son contenu seront indiquées (manuellement) afin de permettre l'initialisation de l'extracteur. Une mise à jour ultérieure de cette initialisation doit être possible, notamment en cas d'évolution ou de mise à jour de l'observatoire. Comme exemples d'informations générales permettant de spécialiser un extracteur, on peut citer :

- Sur l'extracteur : son système cible, ses modes de fonctionnement, son(ses) protocole(s) d'extraction.
- Sur le système source d'informations : nom du système, objectif, propriétaire, adresse ou localisation, type, date de création, taille, etc...
- Sur le serveur qui héberge le système source d'informations : nom du serveur, adresse ou localisation, capacité de traitement (nombre de requêtes simultanées maximum), etc...

0 Sur le contenu du système fédéral (source d'informations) : thème général, sommaire, types de données, description des tables cibles avec fréquence ou mode d'extraction, liste ou arborescence de répertoires, etc...

L'extraction

La définition de plusieurs protocoles d'extraction est nécessaire, car non seulement les sources d'informations sont hétérogènes (DB, Site Web, etc.) mais les types de données

extraites au sein d'un même système peuvent être différents (tables entières, sélection d'une table, donnée, document HTML, image, etc.). L'extracteur peut fonctionner selon plusieurs modes. Voici quelques exemples pris dans le cas d'une BD interrogeable par SQL :

1. Une fois pour toutes :
  - exécuter une requête du genre "describe table" qui liste les définitions des colonnes de la table.
  - demander à un outil de gestion des connaissances d'interpréter son résultat "ontologiquement" et de mettre à jour le système de représentation des connaissances
2. De façon régulière. Il s'agira d'aller extraire des informations en respectant une certaine fréquence. Par exemple, un extracteur pourrait, tous les mois :
  - exécuter une requête "select from" sur la table T de la BD Y.
  - demander à un outil d'intégration de "ranger" le résultat dans une BD.
  - avertir le système du rafraîchissement des données à moins qu'on ne soit dans un système où c'est le module d'administration qui déclenche l'extraction en fonction d'une fréquence de mise à jour.
3. En ligne. Après la traduction par un agent de la requête d'un utilisateur en SQL, un extracteur pourrait :
  - faire exécuter cette requête SQL sur la base de données fédérée (on suppose que les droits d'accès de l'utilisateur à la BD seront au préalable vérifiés par l'administration ou par l'extracteur).
  - transmettre le résultat au module de coordination afin qu'il soit renvoyé à l'utilisateur.

#### **1.4.2.2. Outils d'intégration de l'information**

Ce sont des outils (intégrateurs) qui, disposant de l'information et des méta-informations correspondantes, sauront la "ranger" dans le système. Plus précisément, il s'agira de stocker cette information dans une base de données de capitalisation, et de mettre à jour le système de représentation de connaissances afin que cette information soit facilement retrouvée et judicieusement exploitée.

Les protocoles d'intégration seront utilisés aussi bien pour intégrer les informations qui résultent d'une extraction que pour les documents (fichiers) qui vont être intégrés (en mode "push" d'un système fédéré) dans le système SIMES.

Cette intégration touche donc aux questions concernant l'archivage, le classement et le référencement des informations extraites dans le système SIMES :

- Les informations seront-elles archivées dans leur forme originelle ou seront-elles au préalable codées ou tout au moins normalisées ?

Quels systèmes de classement sont adéquats (classement selon une ontologie spécifique ?) et comment pourront-ils évoluer ?

Comment se fera l'indexation des bases de capitalisation d'informations ?

#### **1.4.2.3. Outils de gestion des connaissances**

Ils s'occupent de la gestion et de l'exploitation du système de représentation ou du référentiel des connaissances. Leur matière première est essentiellement la méta-information qui permet de représenter et de mettre à jour l'information sur l'information. Il pourra s'agir d'outil (ou d'agents) de construction d'index, de classification d'une information, de recherche d'informations ou encore d'extraction de connaissances.

#### **1.4.2.4. Outils de traitement des informations**

Il s'agit d'outils qui effectueront un traitement apportant une valeur ajoutée aux informations capitalisées dans le système ou extraites "en ligne" d'un système fédéré. Ce traitement pourra être effectué soit systématiquement et le résultat stocké, soit ponctuellement si une requête de l'utilisateur le requiert. Ces outils pourront, par exemple, effectuer des traitements statistiques, du traitement d'image ou des simulations. Il se serviront du système de représentation de connaissances pour mieux utiliser les données qu'ils traiteront.

#### **1.4.2.5. Outils de gestion de l'interface**

Dans cette catégorie, on retrouve les outils permettant de gérer les profils des utilisateurs et le service de notifications. Mais on y range aussi les outils qui traiteront l'information sur l'utilisation du système notamment les journaux des accès ou des requêtes (riches en informations permettant d'améliorer l'offre de d'informations et de services faite à l'utilisateur).

#### **1.4.2.6. Outils d'administration et de coordination du système**

Le module d'administration et de coordination du système a un fonctionnement qui peut s'apparenter en termes d'objectifs à celui d'un chef d'orchestre. Il connaît l'existence et l'état de tous les composants du système et dispose d'un module permettant de déclarer et d'intégrer un nouveau composant. Le module de coordination - constitué des premiers outils après l'interface utilisateur à s'occuper réellement de la requête de l'utilisateur - se charge d'orchestrer le traitement de la requête (traduction de la requête en un ensemble d'ordres permettant sa résolution).

### **1.5. Les méta-informations**

#### **1.5.1. La définition des méta-informations**

Compte-tenu du domaine (environnement), il faut définir une ontologie/norme des méta-informations ou méta-données qui devront être manipulées. Faudra-t-il partir de normes existantes ou construire une norme spécifique?

#### **1.5.2. La description des méta-informations**

Se fera-t-elle via un langage "ad hoc" ou par l'utilisation de standards comme SGML? SGML semble être un bon candidat comme moyen fédérateur de l'organisation de la documentation.

#### **1.5.3. Le stockage des méta-informations**

Les méta-informations seront-elles stockées dans la même BD que les informations elles-mêmes, ou séparément ?

#### **1.5.4. L'acquisition des méta-informations**

On peut choisir de contraindre le créateur d'un document à spécifier ces méta-informations. Outre le fait qu'on ne respecte pas l'exigence forte d'absence de contraintes sur les systèmes sources d'informations, cela pose le problème dans le cas des documents existants et ne disposant pas de ces méta-informations.

On peut choisir de n'imposer aucune contrainte sur les sources d'informations :

- soit en spécifiant "manuellement" ces spécifications au moment de l'intégration de chaque document/information dans le système... fastidieux!
- soit en développant quelque chose de semi-automatique ou de complètement automatique. Cette automatisation est à étudier selon les cas. Par exemple : BD relationnelle (avec interrogation possible en SQL), Site Web (utilisation des méta-tags que le créateur aura bien voulu spécifier ou contraindre les créateurs de documents à utiliser des tags spécifiques au système de représentation défini), S.I.G., etc.

### **1.6. Conclusion et perspectives.**

Dans ce chapitre, nous avons présenté une approche du système SIMES à deux visions. La première vision met l'accent sur les caractéristiques des composants logiciels : indépendance et autonomie pour ce qui est des sources d'informations gérées par des partenaires, généricité, interopérabilité et portabilité pour les outils. La deuxième vision qui se focalise beaucoup plus sur la fédération des sources d'informations ainsi que sur celui de l'administration et la coordination des différents composants du système. Des problèmes à résoudre dans le deuxième étant : l'extraction des d'informations, la gestions des incompatibilités des types de ces informations et la définition des méta-informations qui les sont associées. Cette vision est approfondie dans le chapitre suivant qui montre des orientations techniques possibles.

Plusieur points font actuellemnt l'objet de notre réflexion:

Caractèrisation de la notion d'observateur: un observateur est un outil chargé d'élaborer des résultats à partir d'un ensemble de documents jugés pertinents, ce que nous appellerons un *contexte*. Une observation peut être un traitement statistique, un un traitement d'iomage, un; traitement de document à l'aide des outils présents sur la plateforme.

Un contexte (ensemble de documents, sous-ensemble du repository ) est construit par un extracteur: celui-ci il construit un sous-ensmble de ce contexte, qui est encore un contexte. C'est un outil de recherche classique ou un outilde selection ( en cochant par exmple, ou en entourant une zone géographique). Un "montreur" permet de visualiser un contexte en s'appuyant sur une "métaphore".Des exemple sont:

- la métaphore (classique) du bureau: glisser, empiler, mettre à la corbeille,etc..
- une métaphore "géographique": sélectionner un ezone, l'agrandir, la restreindre, s edéplacer sur une carte, etc... et sélectionner au fur et à mesure kles documengt pertinents deu contexte.
- une métaphore de "panier au marché" ou caddie, qui permet de broser divers contextes et de remplir au fur et à mesure le contexte à construire, etc...

L'interface de l'utilisateur sera construit en fonction de ces concepts pour permettre de choisir un contexte, en définir des "vue" grâce aux diveres métaphores disponibles, appliquer à ces des extracteurs pour construire un nouveau contxtte et appliquer des observateurs.

Enfin, la navigation devra aussi permettre de nommer un contexte et d'y revenir, tout comme elle devra permettre de revenir en arrière, défaire, refaire, etc..

## **2. Le Bus logiciel SIMES**

### **2.1. Introduction**

L'infrastructure devra permettre le stockage de données apr essence hétérogènes (textes simples, graphiques, tableaux, enquêtes de terrain, cartes géographiques,etc..., la mise en oeuvre d'outils très divers ( statistiques, extractions de documents, observateurs, traitements graphiques, traitement d'iamges). Elle sera distribuée et accessible par Internet (WWW, ftp, mail , telnet).Tous les choix concernant l'infrastructure ne sont pas encore fixés. Il sont en cours de raffinement et d'intégration des différentes couches. Ce chapitre concerne les couches basses de l'architecture du bus logiciel SIMES. Deux grandes voies sont possibles: s'appuyer sur des standards reconnus d'accès à des bases de données comme ODBC, ou s'appuyer sur une solution "propriétaire" comme une base de données relationnelle + un interface web intégrée, comme le propose Oracle 8. Ce chapitre expore la première hypothèse.

### **2.2. Background**

Une source d'inspiration pour cette couche est celle du projet PCIS-2.

#### **2.2.1. Présentation de PCIS-2**

PCIS-2 (Process Components Infrastructure architecture Specification) est un projet franco- américain (DGA/DRET- US Navy) de définition d'une architecture d'intégration et d'interopérabilité d'outils logiciels hétérogènes distribués sur Intranet (1997- 1999). PCIS "2" parcequ'il fait suite au projet PCIS (Portable Common Interface Set) de l'OTAN ayant abouti à la spécification d'une plate-forme orientée objet de support aux environnements de développement logiciel. La plate-forme doit au maximum utiliser des standards (Java, Corba, HTML), s'appuyer sur des produits commerciaux et être opérationnelle sur Solaris2.6 et NT 4.0. La plate-forme est sensible aux procédés. Le modèle est celui de LCPS2 étendu par des facilités de fragmentation et de réutilisation. Les objets sont stockés dans un "repository" conforme ODMG, complété par un service de gestion des versions et configurations. Ils sont rendus accessibles par Corba. Les outils sont encapsulés (Java) pour être invocables depuis tout site et la plate-forme fournit tous les adaptateurs de types nécessaires ("traders"). L'ensemble est complété par un service d'administration et de support à l'interface utilisateur.

#### **2.2.2. Collaborations Internationales du LORIA**

Loria et Arizona State University ont défini les spécifications, Sema réalise le démonstrateur qui sera utilisé en 1999 pour de futurs développements de l'US Navy et de la DGA (prochaine famille de sous-marins français et système américain de surveillance de l'océan). Un premier prototype a été présenté à la conférence Software Engineering de Salt Lake City (avril 98) voir aussi [1], [3] et [2]. Jean Claude Derniame est responsable de l'équipe d'architecture et intervient avec Pierre Tiako pour l'aspect "procédés", tous deux intervenant dans le projet SIMES.

#### **2.2.3. Architecture à la CORBA**

L'architecture logique de l'infrastructure d'une telle plate-forme peut être schématisée en trois couches comme le montre la figure 2-1. Une approche en trois couches est

caractérisée par une couche application, la partie client de l'interface utilisateur et un ensemble de services rendant les applications disponibles aux utilisateurs.

L'architecture PCIS-2 est prévue pour être distribuée sur un réseau du type Internet et le bus de communications est supposé être CORBA. Mais il peut aussi inclure d'autres mécanismes de distribution tels que Java Remote Method Invocation (RMI). La partie framework de PCIS-2 comprend un gestionnaire de configuration, un reposoir d'artefacts et des services pour la coopération et les procédés de développement.

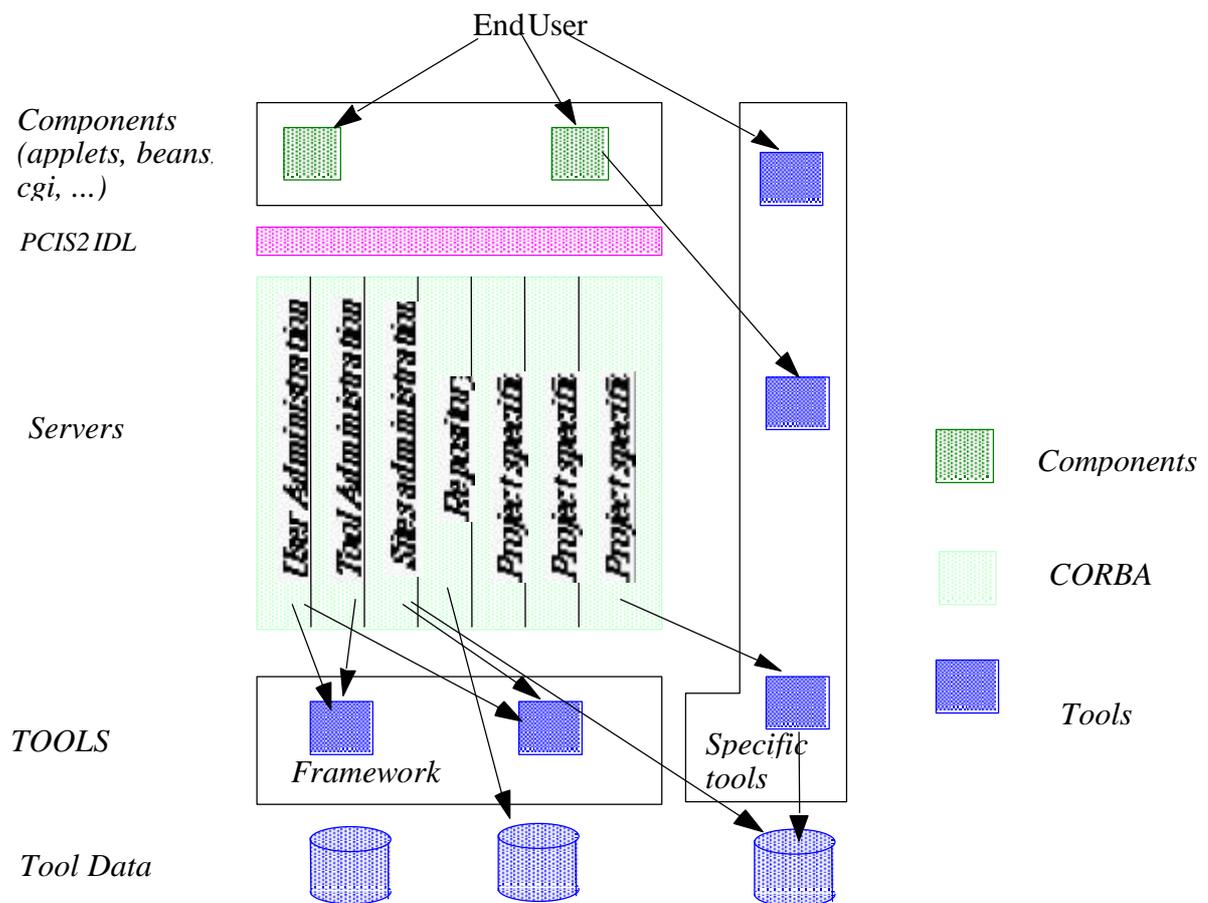
### **2.3. Proposition SIMES**

Nous retiendrons une vue plus concrète de l'architecture pour SIMES, basée, elle aussi sur une approche en trois couches figure 2-1 :

- une première couche composée de deux parties :
    - les données de tout type qui sont rendues accessibles aux différents outils (provenant par exemple des observatoires comme décrit au ,
    - des outils du commerce (COTS) ou des gouvernements (GOTS), et de leurs données. Il comprend les outils du "framework" lui-même, entre autres les différents SGBD nécessaires, les outils de traitement des standards: HTML, Java, etc... et les outils spécifiques au domaine, tels que les outils de cartographie, les grapheurs, etc...
  - une deuxième couche contenant les services offerts par le système; accès, recherche, gestion des usagers, statistiques, etc... et les interfaces en IDL visibles des clients
- la dernière couche, la partie client des applications est composée des applets, CGI, RMI et autres artefacts d'accès.

#### **2.3.1. PCIS 2 versus SIMES**

Les objectifs du projet PCIS-2 partagent avec ceux de SIMES ceux d'intégrer un environnement logiciel basé sur le Web formé de produits, d'outils et de standards déjà existants (intégration a posteriori). Les choix de technologie objet, Java/C++, Corba, technologie Web, HTML et SGML sont cohérents avec ces objectifs.



**Figure 2.1** : L'architecture en trois niveaux

L'adoption du standard de l'OMG pour l'accès aux objets est fondamental. L'usage de IDL (Interface Description Language de Corba) pour décrire la syntaxe des interfaces des services de l'infrastructure et du langage naturel (narratif) pour leur sémantique permettra une spécification de qualité et leur interopérabilité.

L'usage de CORBA dans PCIS se justifie pleinement par la nature et la taille des applications envisagées dans le développement d'applications de systèmes de combats. Cependant la situation est différente dans SIMES. *Il n'est sans doute pas indispensable d'envisager l'obligation pour chaque site de collecte et de service de données d'être équipé d'un logiciel du type IIOP.*

Pour les développements à réaliser, et en particulier pour les services, l'adoption de Java et de C++ quand c'est nécessaire pour des raisons de compatibilité ascendante, sont un gage de portabilité efficace des réalisations : ce choix semble important aussi pour SIMES.

#### **2.4. Discussion d'implémentations**

La figure 2-2 montre une implémentation possible de cette architecture.

Une première observation s'impose sur cette architecture fonctionnelle : on voit sur ce schéma deux installations de l'architecture SIMES, reliées par un bus qui est un Intranet

(ou simplement Internet). Chacune d'elle regroupe des sites connectables entre eux de façon plus ou moins permanente, plus précisément leurs ressources sont connues sur les autres sites et des accès à ces ressources peuvent être activés à tout moment. Sur ces sites se trouvent répartis divers outils (de traitement d'images, de traitement de textes, des outils statistiques, etc...) et des bases de données correspondantes. L'ensemble se présente à l'extérieur comme une machinerie unique avec ses fonctionnalités traditionnelles de stockage et recherche d'informations, de traitement et d'interface utilisateur. Celle-ci est définie pour être uniforme sur l'ensemble de SIMES. Réalisée en Java, elle s'appuie sur l'interface Web habituelle. S'y ajoute une unité d'administration dans laquelle sont déclarés les outils mis à disposition (avec leur fonctionnalité ainsi que leur localisation sur le bus), les utilisateurs et leurs droits (clients, gestionnaires de certaines données ou administrateur de site). Ces deux installations de SIMES peuvent éventuellement communiquer entre elles par tout autre moyen que l'Internet : liaisons spécialisées, envois de disquettes et/ou de CD-Rom, etc...; ce qui est représenté par l'interconnexion entre deux SIMES.

### 2.4.1. Collecte centralisée

Une première représentation géographique de cette architecture est donnée dans la figure 2- 3.

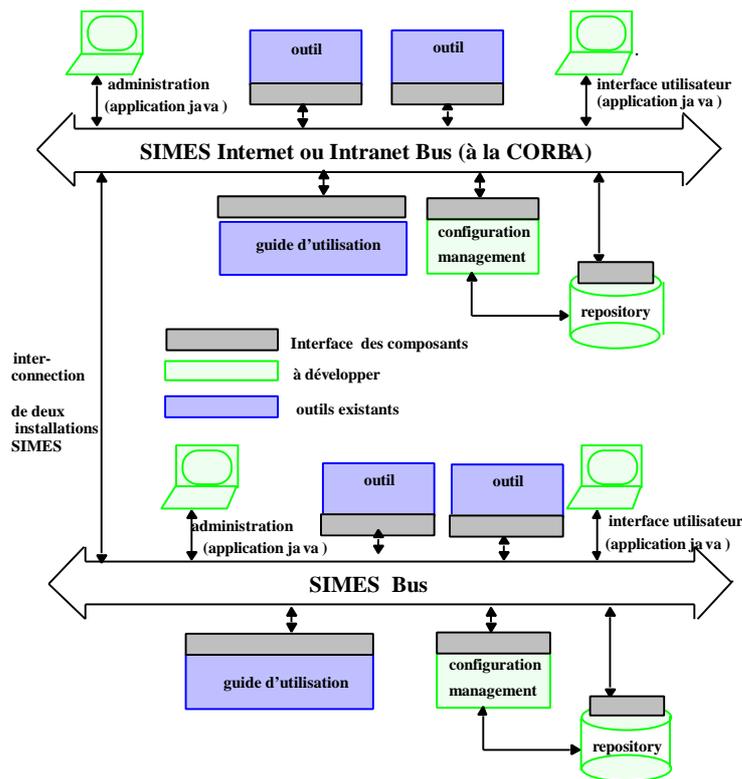
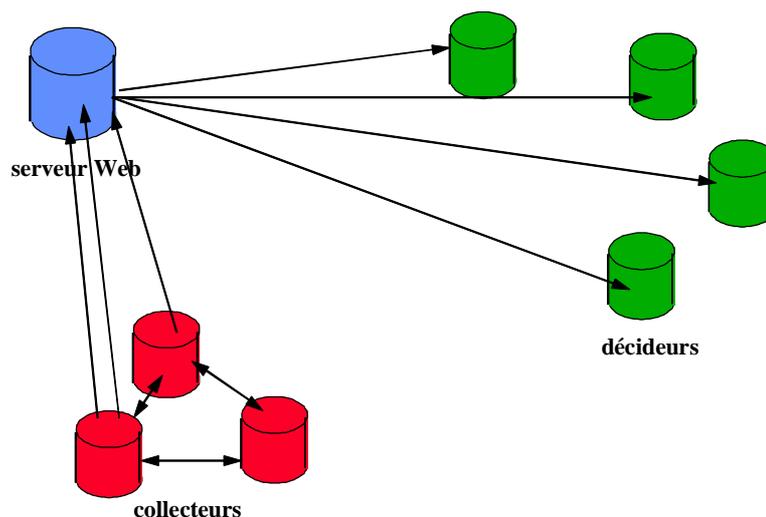


Figure 2.2: Architecture de l'infrastructure SIMES

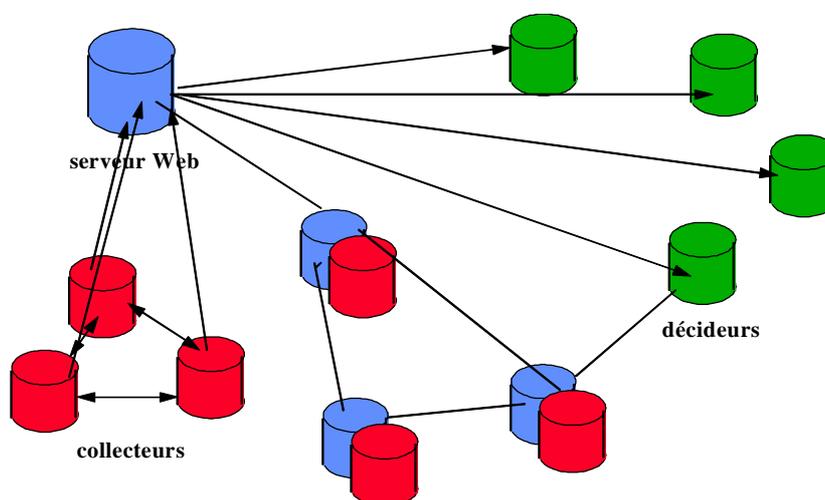


**Figure 2.3 :** Une première vue géographique de l'architecture

Il faut bien voir que celle-ci est limitative : ici, il est supposé que toutes les informations sont converties pour être présentées en HTML sur un serveur Web unique qui rassemble toutes les informations des sites de collecte. Ces informations sont interrogées depuis les sites de consultation appelés "décideurs". Cette solution est bien sûr supportée par l'architecture SIMES mais ce n'est pas sa vocation principale.

#### 2.4.2. Collecte distribuée

Dans la figure 2-4 on peut voir la solution générale apportée par SIMES : chaque site de collecte, seul ou en association avec un site similaire, possède son propre serveur Web d'information et c'est l'ensemble du réseau de serveurs qui est interrogé par les sites "décideurs".



**Figure 2.4 :** Une deuxième vue géographique de l'architecture

- Du point de vue opérationnel, les outils spécifiques (extracteurs, intégrateurs, outils de traitement d'informations, etc...), les outils généraux (traitement de textes, d'images, de statistiques), mais aussi les outils graphiques, les outils de mise en forme (courbes /ex), les outils d'assistance, etc... qui constituent la plate-forme SIMES sont

autant d'agents intelligents au sens des systèmes multi-agents classiques qui opèrent sur cette plate-forme de distribution.

- Ces outils (existants pour la plupart dans le commerce ou en logiciels libres) sont encapsulés ("wrapped") pour dialoguer avec le bus. Ainsi, un extracteur peut demander au bus l'ouverture et l'accès à un fichier de données situées sur un site distant et travailler sur elles comme si elles étaient situées localement.

### **2.5. Perspectives**

Dans la version actuelle de SIMES nous ferons l'hypothèse que les données sont dans des formats prédéfinis et pour lesquels la plate-forme dispose des traducteurs de format nécessaires pour les rendre lisibles aux outils : le problème de l'adaptation automatique des formats aux outils est considéré pour l'instant comme étant en dehors du scope du projet. Ce problème est actuellement traité à Nancy par Olivier Perrin dans un autre cadre mais ses résultats pourront être intégrés.

## REFERENCES

- [1] J.-C. Derniame and T. Lindquist. Toward distributed and composable process components. In *International Workshop on Research Directions in Process Technologies*, Nancy-France, July 1997. Available on <http://www.elet.polimi.it/people/fuggetta/rdpt97.html>.
- [2] J.-C. Derniame and P.F. Tiako. Toward process components mobility in federated process sensitive engineering environments. Technical Report 98-R-045, LORIA - INPL, Nancy - France, 1998. Submitted to 13th IEEE International Conference on Automated Software Engineering.
- [3] J.C. Derniame, T. Lindquist, and P.F. Tiako. Towards distributed and composable process components in pcis ii project. Second Technical Research and Development Project on Software Tools, US Navy - French MOD, San Diego, June 1997.
- [4] Morand and Kodio. Mise en place d'un système de suivi de la pêche dans le delta central du niger : Concepts et méthodes. Rapport d'études, Niger, 1996.