Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Conforme à l'original produit,
Début du texte, page suivante.

BULLETIN OFFICIEL DES ARMEES

Edition Chronologique n°2 du 9 janvier 2009

PARTIE PERMANENTE
Administration Centrale

Texte n°2

DIRECTIVE N° 6/DEF/DGSIC
relative aux composants de logiciels décisionnels.

Du *5 décembre 2008*

DIRECTION GÉNÉRALE DES SYSTÈMES D'INFORMATION ET DE COMMUNICATION.

DIRECTIVE N° 6/DEF/DGSIC relative aux composants de logiciels décisionnels.

Du 5 décembre 2008

NOR D E F M 0 8 5 2 8 4 6 X

Références :

Ordonnance n° 2005-1516 du 8 décembre 2005 (JO n° 286 du 9 décembre 2005, texte n° 9 ; BOC, p. 8645. ; BOEM 120-0.3.1).
Décret n° 2006-497 du 2 mai 2006 (n.i. BO ; JO n° 103 du 3 mai 2006, texte n° 9 ; JO/135/2006. ; BOEM 160.1).

Pièce(s) Jointe(s) :

Deux annexes.

Classement dans l'édition méthodique : BOEM 160.11

Référence de publication : BOC N°2 du 9 janvier 2009, texte 2.

---

1. PRÉSENTATION GÉNÉRALE ET GUIDE D'USAGE.

1.1. **Présentation.**

La présente directive définit les règles et les critères de décision pour l'acquisition, la réalisation et la mise en œuvre d'une plateforme d'informatique décisionnelle (*Business Intelligence*) intégrant un logiciel d'intégration de données (ETL), une solution de stockage de données et une solution de restitution de données aux utilisateurs au ministère de la défense

Elle a pour but de fournir une base commune de règles techniques et fonctionnelles à respecter au sein des entités du ministère de la défense.

1.2. **Niveaux de préconisation.**

Les règles présentées dans ce document ont différents niveaux de préconisation repris du référentiel général d'interopérabilité (RGI) et inspirés de la RFC 2119 :

- OBLIGATOIRE : ce niveau de préconisation signifie que la règle édictée indique une exigence absolue de la directive ;

- RECOMMANDÉ : ce niveau de préconisation signifie qu'il peut exister des raisons valables, dans des circonstances particulières, pour ignorer la règle édictée, mais les conséquences doivent être comprises et pesées soigneusement avant de choisir une voie différente ;

- DÉCONSEILLÉ : ce niveau de préconisation signifie que la règle édictée indique une prohibition qu'il est toutefois possible, dans des circonstances particulières, de ne pas suivre, mais les conséquences doivent être comprises et le cas soigneusement pesé ;

- INTERDIT : ce niveau de préconisation signifie que la règle édictée indique une prohibition absolue de la directive.

1.3. **Modalités d'application.**

La présente directive s'applique à tous les composants, projets, programmes, opérations incluant des logiciels existants développés pour une utilisation générale ou des logiciels développés spécifiquement.

Ces règles définissent la cible et sont applicables à tout nouveau projet ou toute évolution majeure. L'ensemble des systèmes d'information du ministère de la défense devront être en conformité avec la présente directive au premier janvier 2010.

1.4. **Gestion des exceptions pour les projets.**

Elles sont instruites par la CMTSIC* et font l'objet d'une approbation par le directeur général de la direction générale des systèmes d'information et de communication (DGSIC).

2. CADRE DOCUMENTAIRE.

2.1. **Documents applicables.**

Recommandations nationales du cadre commun d'interopérabilité (CCI).

Référentiel général d'interopérabilité (RGI) et référentiel général de sécurité (RGS).

Politique de référencement intersectoriel de sécurité (PRIS).

Directive sur les logiciels (DGSIC001).

Directive n° 5/DEF/DGSIC du 7 avril 2008 portant sur les systèmes de gestion de base de données relationnelle.

2.2. **Normes et standards applicables.**

[RFC 2119] - Mots-clés pour niveaux d'obligation.

[ISO*/CEI 9075 1992*] - SQL*92 ou SQL2.

[ISO/CEI 10646 2003*] - jeu universel de caractères codés sur plusieurs octets (UNICODE version 5.0.0).

[REC-XML-19980210*] - spécifications XML* 1.0.

[REC-XML11-20060816] - spécifications XML 1.1 (seconde édition).

[JSR-054*] - Java JDBC* 3.0 specification.

[JSR-114*] - Java JDBC Rowset implementations.

[JSR-168] - Portlet APIS.

3. DOMAINE COUVERT ET EMPLOI.

L'informatique décisionnelle (en anglais : DSS pour décision support system ou encore BI pour business intelligence) désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données d'une entreprise en vue d'offrir une aide à la décision et de permettre aux responsables de la stratégie d'une entreprise d'avoir une vue d'ensemble de l'activité traitée.

Ce type d'application utilise en règle générale un entrepôt de données (ou *datawarehouse*) pour stocker des données transverses provenant de plusieurs sources hétérogènes et fait appel à des traitements lourds de type traitement par lots (batch) pour la collecte de ces informations.

Les outils de simulation, de statistique et d'aide à la décision orientés vers la recherche opérationnelle sont hors du champ d'application de la présente directive.

L'informatique décisionnelle s'insère donc dans l'architecture plus large d'un système d'informations. En effet, les données « métier » sont stockées dans une ou plusieurs bases de données relationnelles ou non relationnelles. Ces données sont extraites, nettoyées, transformées et chargées dans un entrepôt de données par un outil de type ETL.

Un entrepôt de données prend la forme d'un entrepôt de données et/ou d'un magasin de données (datamarts). En règle générale, un entrepôt de données globalise toutes les données applicatives de l'entreprise. Le magasin de données est généralement alimenté par l'entrepôt de données et forme un sous-ensemble d'informations concernant un métier particulier de l'entreprise.

Les outils d'aide à la décision (appelés également outils décisionnels, de restitution ou de business intelligence) permettent ensuite de présenter et d'analyser les données issues de l'entrepôt de données ou des magasins de données afin d'en dégager des informations qualitatives nouvelles qui seront la base de décisions tactiques ou stratégiques. Ces outils peuvent couvrir plusieurs fonctionnalités telles que le requêtage, le reporting, l'analyse multidimensionnelle (technologies OLAP et ses déclinaisons), le pilotage (tableaux de bord) et l'exploration de données (datamining).

Les trois composantes majeures du système décisionnel sont couvertes :

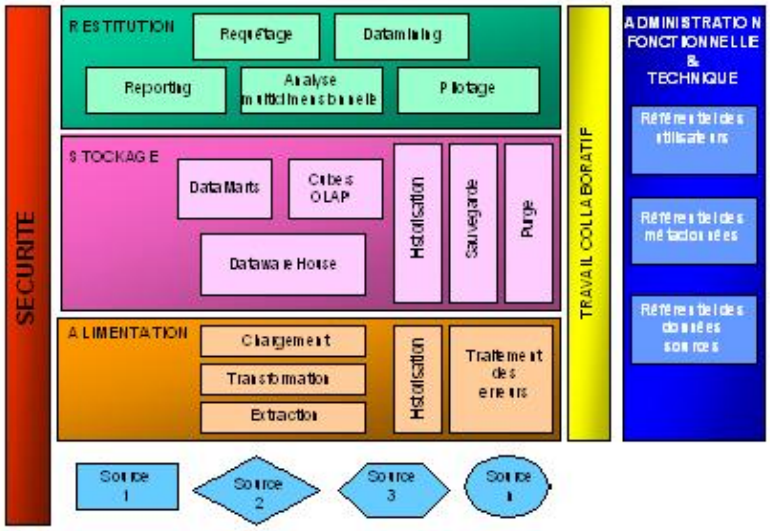
- alimentation : cette composante extrait, transforme et nettoie les données des systèmes sources pour les charger dans un entrepôt de données ;

- stockage : cette composante comprend un entrepôt de données central, un ou plusieurs espaces de stockage multidimensionnel (OLAP) et une ou plusieurs bases de données spécialisées appelées aussi magasins de données ou d*atamarts ;*

- restitution : c'est l'ensemble des outils permettant la présentation et l'analyse de l'information issue du système décisionnel vers l'utilisateur final.

À ces trois principales composantes s'ajoutent les fonctions transverses de sécurité, de travail collaboratif, et d'administration fonctionnelle et technique.

L'architecture fonctionnelle d'un système décisionnel peut être modélisée de la façon suivante :



3.1. **Le logiciel d'intégration de données.**

Les logiciels ETL (extraction, transformation, loading) sont chargés d'extraire les données de différentes sources, de les transformer et de les charger dans un système décisionnel.

3.1.1. ***Besoins satisfaits par l'ETL.***

Les services offerts par les ETL ont pour but de répondre aux besoins suivants :

- réconciliation et enrichissement des données : croisement des données élémentaires provenant de systèmes sources distincts ;

- transformation des données : application de règles plus ou moins complexes de traitement et de calcul des données ;

- qualité des données : assurer, garantir la cohérence et la pertinence des données de l'application décisionnelle face à des sources multiples et hétérogènes. Identifier les rejets fonctionnels et techniques et les règles pour leur recyclage ;

- traçabilité et historisation : identification de l'origine de la donnée et historisation des données sources ;

- documentation des processus d'alimentation : documentation technique pour l'exploitation et la maintenance des traitements. Documentation fonctionnelle du processus d'alimentation.

L'ETL permet de gérer tous les mouvements de données en batch et en temps réel dans l'entreprise . Cette solution peut être utilisée sur des projets décisionnels, d'intégration/d'échange de données inter-applicatives, ou de migration de données.

3.1.2. ***Les services fournis par le logiciel d'intégration des données.***

L'outil ETL doit proposer les services suivants :

- Un service de traitement : il accède nativement aux bases de données sources, cibles et aux référentiels de l'ETL ;

- référentiel de métadonnées : les métadonnées décrivent les règles et les processus de transformation des données du système décisionnel. La gestion des métadonnées permet notamment de mesurer l'impact d'une modification sur le système décisionnel ;

- interface de conception et de modélisation des flux de traitements : interface graphique intuitive qui permet de définir les mouvements de données intégrant l'accès, la transformation et le chargement de celles-ci. Elle permet de définir des processus de mapping, de transformation (simples ou complexes) et de contrôle de la pertinence des données pour alimenter l'entrepôt de données sans avoir recours à la programmation. Elle utilise les mécanismes de copier-coller et glisser-déplacer. L'outil devra proposer un ensemble de fonctions de transformation pré-définies ;

- grande polyvalence d'accès aux données : nativement un ETL doit accéder aux SGBDR, aux fichiers plats, aux fichiers XML, aux fichiers MS XLS, aux fichiers OpenDocument Calc, aux données ERP (avec des connecteurs souvent spécifiques) et aux données non structurées ;

- service d'analyse de la qualité des données : il dispose en standard de services spécialisés dans la qualité des données.

3.2. **Le stockage des données.**

Le stockage au sein des SGBDR est traitée par la directive portant sur les systèmes de gestion de base de données relationnelle.

La fonction de stockage a pour but la mémorisation des données issues de l'alimentation, des référentiels et des données de restitution. Elle présente une structure utilisable comme support par la fonction de restitution.

3.2.1. ***L'architecture.***

Dans une architecture décisionnelle, le SGBDR joue un rôle prépondérant au niveau de l'alimentation (operational data store) et du stockage des données élémentaires (entrepôt de données).

L'operational data store (ODS) est une base facultative de stockage des données temporaires. À ce niveau, les données peuvent faire l'objet de contrôle qualité (contrôle de la cohérence et de la pertinence des données par rapport à des règles fonctionnelles et techniques) et peuvent subir des traitements de réconciliation (enrichissement des données, réconciliation par rapport à des référentiels de domaines fonctionnels...) et de transformation (calculs, tris...).

L'entrepôt de données (datawarehouse) correspond à un espace de stockage pour les données issues des systèmes opérationnels. Il est le support du processus d'aide à la décision en fournissant des données :

- orientées sujet (les données sont organisées autour des sujets majeurs et des métiers afin de rendre possible des analyses transverses aux structures fonctionnelles et organisationnelles) ;

- partagées (la sémantique des données et leurs valeurs effectives sont partagées) ;

- intégrées (les données sont mises en forme et unifiées) ;

- historisées (l'évolution des données est suivie dans le temps) ;

- non volatiles (les données sont permanentes et ne peuvent pas être supprimées).

Le magasin de données est un sous-ensemble de l'entrepôt de données contenant les données d'un métier particulier. Il a vocation à présenter les données aux utilisateurs avec les meilleures performances possibles.

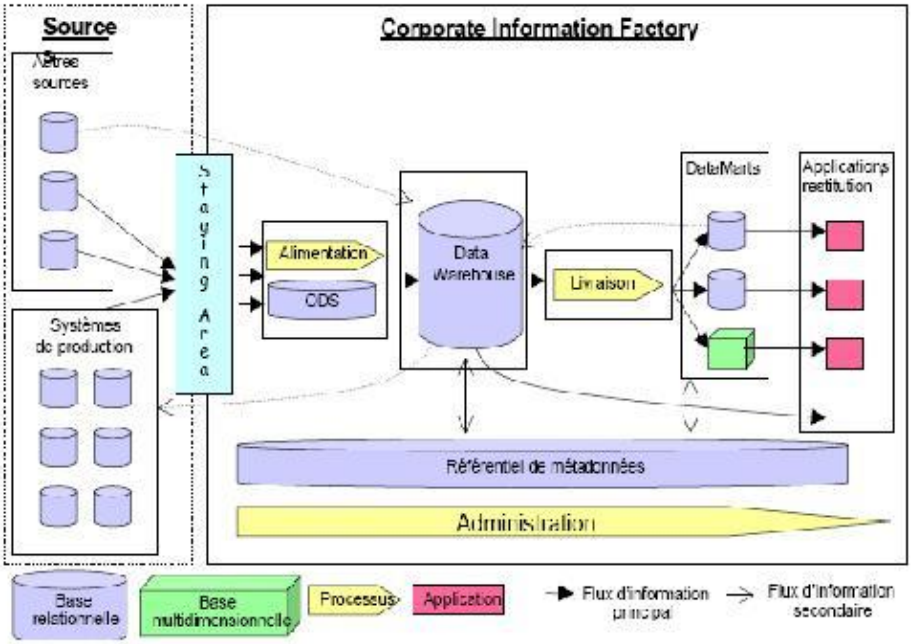
Il n'y a pas de règle absolue en matière de choix d'architecture. Cependant, dans la pratique, les solutions les plus souvent choisies sont :

- un modèle relationnel à l'identique des systèmes opérationnels pour l'ODS ;

- un modèle relationnel pour l'entrepôt de données ;

- un modèle multidimensionnel pour les magasins de données .

Le modèle du magasin de données peut être également relationnel. Il est lié à son utilisation et donc au type d'outil de restitution.



Architecture générale d'un système décisionnel.

*3.2.2. **La modélisation.***

La pérennité, l'évolutivité et les performances du système décisionnel reposent essentiellement sur sa modélisation.

Le modèle de données des systèmes opérationnels est optimisé pour des ajouts/modifications/suppressions portant sur des données élémentaires. Il est conçu sur la base d'un modèle relationnel normalisé.

Les systèmes décisionnels traitent des volumes de données importants tant en chargement qu'en consultation. Pour cette raison, la modélisation des données est spécifique.

Elle requiert un savoir faire très particulier mais surtout de l'expérience et la capitalisation sur les projets. Cette expertise à forte valeur ajoutée est à introduire dès la phase de conception des projets. La modélisation du système décisionnel ne doit pas être considérée comme un livrable technique mais bien comme un livrable fonctionnel qui requiert un travail conjoint de la maîtrise d'ouvrage et de la maîtrise d'œuvre étayé éventuellement par une expertise sur le sujet.

La modélisation dimensionnelle (modèle en étoile ou en flocon) est une méthode de conception logique qui vise à présenter les données sous une forme standardisée intuitive et permet des accès performants. Cette dernière est fréquemment mise en œuvre dans les datamarts.

*3.2.3. **Les technologies on line analytical processing.***

Le terme OLAP (on line analytical processing) est utilisé dans le monde décisionnel pour désigner l'ensemble des technologies permettant l'analyse multidimensionnelle.

Sous la terminologie OLAP, on distingue, sans exhaustivité :

- l'implémentation d'une modélisation « OLAP » dans une base de données SGBDR classique : R-OLAP ;

- l'implémentation d'une modélisation « OLAP » dans une base de données spécifique : M-OLAP ;

- le H-OLAP est une solution hybride exploitant R-OLAP et M-OLAP.

Dans les bases de données MOLAP, les différents agrégats sont pré-calculés et stockés physiquement. La restitution des informations est immédiate. Le stockage des agrégats peut engendrer des volumétries considérables.

L'évolution des besoins vers des analyses « temps réel » et portant sur des volumes de données de plus en plus importants renforce l'utilisation de la technologie OLAP.

L'OLAP constitue un choix recommandé dans une logique de magasin de données et pour répondre à des besoins tels que la simulation, le datamining, la prévision et l'élaboration budgétaire. Le choix d'une des trois technologies dépend des besoins fonctionnels.

3.3. **La restitution.**

On appelle restitution, toute présentation de l'information issue du système décisionnel (entrepôt de données, magasin de données) vers l'utilisateur final. Il en existe plusieurs types afin de répondre aux divers besoins des utilisateurs.

Le requêtage permet à l'utilisateur d'accéder à l'information avec ou sans mise en forme prédéfinie dans un langage natif de la base de données (SQL, MDX). L'outil de requêtage doit permettre de masquer la complexité de la requête à l'utilisateur en lui proposant une interface conviviale.

Le reporting consiste à mettre en page l'information sous forme de tableaux ou de graphiques et à les diffuser aux utilisateurs. C'est le mode de restitution le plus répandu. Différents types de reporting existent en fonction du profil de l'utilisateur : reporting de masse (mise en forme prédéfinie, informations pré-calculées), reporting paramétré (variables saisies par l'utilisateur), reporting ad-hoc (pour les utilisateurs qui maîtrisent leurs données, les règles de fonctionnement des tables et des indicateurs).

L'analyse multidimensionnelle consiste à présenter l'information selon plusieurs axes d'analyse (vision appelée cube). La navigation dans l'information s'effectue en partant du niveau le plus agrégé vers les niveaux les plus détaillés (concept de drill down) ou de façon transversale (concept de drill through - changement d'axe d'analyse pour visualiser les données sous un angle différent). Cette analyse des données s'appuie sur les technologies OLAP.

Le pilotage, plutôt destiné aux dirigeants, permet de donner une vision synthétique des données sous forme de tableau de bord avec des indicateurs clairement interprétables.

Enfin, le datamining, forme la plus aboutie de l'analyse, consiste à utiliser des algorithmes mathématiques pour détecter des corrélations entre les données et dégager des tendances. Les données sont analysées par le système sans que l'utilisateur pose des hypothèses ou indique dans quelle direction les corrélations doivent être recherchées.

Avec le développement des technologies Web, ces fonctionnalités peuvent être fédérées au sein d'un portail décisionnel qui permet notamment de gérer les accès des utilisateurs, de classer les rapports par sujet, de publier les documents créés.

3.4. **Les fonctions transverses.**

3.4.1. ***Sécurité.***

Compte tenu du caractère sensible de l'information mise à disposition sous forme de rapports, de tableaux de bord ou d'analyses, la solution décisionnelle devra disposer d'une gestion de la sécurité élaborée, permettant de sécuriser l'accès à l'information suivant des critères multiples :

- limitation de l'accès aux fonctions de base de la solution décisionnelle (accès, visualisation, planification, création des documents analytiques) en fonction du profil utilisateur ou des groupes d'appartenance ;

- limitation de l'accès aux documents analytiques ou dossiers les contenant en fonction du profil utilisateur ou du/des groupes d'appartenance ;

- possibilité de filtrer le résultat des requêtes en fonction du profil utilisateur ou du/des groupes d'appartenance.

3.4.2. ***Travail collaboratif.***

Le système doit offrir des fonctionnalités de travail collaboratif, comme par exemple un circuit de validation des données lors de leur intégration, de la saisie de commentaires ou d'un document analytique*.

3.4.3. ***Administration.***

L'administration des composants de la plateforme décisionnelle est centralisée et s'effectue à l'aide d'un client léger. À partir d'une console unique, l'administrateur pourra gérer :

- la gestion des droits sur les modules de la solution ;

- l'administration des utilisateurs et des groupes ;

- la configuration du mode d'authentification des utilisateurs ;

- la gestion des différents services et leur paramétrage ;

- l'audit de la plateforme qui permet de gérer de façon plus appropriée le système décisionnel et les comptes utilisateurs individuels en donnant plus d'informations sur les actions menées par les utilisateurs et sur les documents analytiques auxquels ils accèdent.

3.4.4. ***Interopérabilité et interfaçage de l'informatique décisionnelle.***

Il est conseillé de considérer l'informatique décisionnelle comme un composant d'une architecture de services. Elle doit donc avoir des interfaces standardisées pour pouvoir coopérer avec les autres services.

En outre, la solution décisionnelle doit proposer en standard, un environnement de développement complémentaire pour l'ensemble des composants de son offre, permettant de personnaliser et d'étendre les possibilités de la plateforme logicielle.

4. LES RÈGLES.

La directive est déclinée sous 2 angles : technique (RT) et fonctionnelle (RF). Les règles sont numérotées séquentiellement par catégorie.

4.1. **Règles techniques.**

4.1.1. ***Généralités.***

- RT 01 : il est obligatoire de pouvoir activer des modules indépendants (open source ou issus d'autres éditeurs) au cours de la vie d'un projet décisionnel ;

- RT 02 : il est obligatoire d'être en mesure d'utiliser les navigateurs les plus courants du marché (Microsoft Internet Explorer, Mozilla Firefox, etc.) ;

- RT 03 : il est obligatoire de disposer d'une interface graphique supportant plusieurs langues dont la langue française (ETL et outil de restitution).

4.1.2. ***Le logiciel d'intégration de données.***

- RT 04 : il est recommandé que l'ETL puisse être utilisé aussi bien de façon autonome ou intégrée dans une infrastructure plus vaste intégrant SOA (SOAP, XMLRPC) et ESB ;

- RT 05 : il est recommandé de pouvoir afficher les implications d'une modification des données physiques sur les processus ETL (graphiques de dépendance) ;

- RT 06 : il est obligatoire que le processus ETL soit en mesure d'extraire ou de générer divers formats standards de fichiers (fichiers textes, fichiers CSV, fichiers Microsoft Excel, fichiers Open Office, fichiers XML) ;

- RT 07 : il est recommandé que le processus ETL généré soit compatible au minimum avec les systèmes Windows, Unix, Linux ;

- RT 08 : il est recommandé que le processus ETL généré utilise un compilateur ou un interpréteur ou une machine virtuelle standard (exemple : Java, C,...) ;

- RT 09 : il est obligatoire que l'ETL puisse s'interfacer avec les protocoles HTTP, IMAP, POP3, FTP et SMTP ;

- RT 10 : il est obligatoire que l'ETL puisse appeler des programmes externes ou des commandes système du système d'exploitation ;

- RT 11 : il est obligatoire que l'ETL permette le stockage des erreurs ou alertes dans une base de données au standard ANSI2 ;

- RT 12 : il est obligatoire pour l'ETL de disposer d'un composant graphique permettant de générer les correspondances entre les colonnes en entrées, les colonnes en jointure et les différentes colonnes en sortie (mapping). Les rejets (jointures non réalisées) doivent être gérés de façon native ;

- RT 13 : il est obligatoire pour l'ETL de disposer d'un composant gérant les tris et l'agrégation des données ;

- RT 14 : il est obligatoire pour l'ETL d'être en mesure de s'interfacer avec les bases de données relationnelles du marché (Oracle, IBM DB2, MS SQL Server, Sybase, Teradata, Informix, et équivalent ...) et logiciels libres (Open Source) (comme MySQL, PostgreSQL, Ingres, et équivalent...) en privilégiant les connexions natives ;

- RT 15 : il est recommandé de pouvoir visualiser le code généré par l'ETL ;

- RT 16 : il est obligatoire que l'ETL puisse contrôler la qualité des données (conformité technique et /ou fonctionnelle) pendant leur traitement ;

- RT 17 : il est obligatoire de pouvoir informer les utilisateurs de l'outil des comptes rendus d'exécution au minimum par courriel ;

- RT 18 : il est obligatoire que l'ETL dispose d'une console permettant la surveillance des exécutions des différents processus ;

- RT 19 : il est obligatoire que l'ETL dispose d'un outil de navigation pour les méta-données ;

- RT 20 : il est obligatoire que l'ETL permette la parallélisation des traitements ;

- RT 21 : il est obligatoire que l'ETL mette à disposition des journaux spécifiques indiquant notamment l'heure de début/fin de traitement et le nombre de lignes lues/écrites/en erreur ;

- RT 22 : il est obligatoire que l'ETL dispose d'un mode pas à pas ou débogage ;

- RT 23: il est obligatoire que l'ETL permette la livraison automatisée d'une ou plusieurs interfaces d'un environnement (de développement, de recette, de production,...) à un autre ;

- RT 24 : il est recommandé que l'ETL puisse échanger avec l'outil MEGA ;

- RT 25 : il est obligatoire que des contraintes de précédence puissent permettre de mettre en œuvre la logique d'enchaînement de l'exécution des tâches (succès, échec, etc.) ;

- RT 26 : il est recommandé que l'ETL puisse marquer une source (par exemple dans le cadre d'une reprise de données différentielles) ;

- RT 27 : il est recommandé que l'ETL puisse intégrer les fonctions de Change Data Capture.

4.1.3. ***Stockage.***

- RT 28: il est obligatoire d'être en mesure de proposer une interface JDBC ou/et ODBC pour accéder aux bases de données ;

- Ensemble des règles définies au sein de la directive n° 5/DEF/DGSIC du 7 avril 2008 portant sur les systèmes de gestion de base de données relationnelle .

4.1.4. ***Restitutions.***

- RT 29 : il est recommandé de pouvoir visualiser le code généré par l'outil de restitution ;

- RT 30 : il est recommandé de pouvoir afficher les implications d'une modification des données physiques sur les documents analytiques (analyse d'impacts) ;

- RT 31 : il est recommandé que l'interface à disposition de l'utilisateur final soit bâtie sur des technologies de portail standardisées (exemple: portail Java supportant la spécification JSR 168 Portlet) ;

- RT 32 : il est obligatoire que l'outil de restitution supporte les standards XML, SQL ANSI2 et MDX ;

- RT 33 : il est recommandé que l'outil de restitution supporte le standard XMLA (OLAP) ;

- RT 34 : il est recommandé que l'outil de restitution utilise un langage de programmation standard (JAVA,C++,...) pour les développements complémentaires ;

- RT 35 : il est recommandé d'être en mesure de personnaliser le comportement des différents documents analytiques par script et/ou procédures spécifiques ;

- RT 36 : il est recommandé de pouvoir visualiser le code généré par l'outil de restitution ;

- RT 37 : il est recommandé de pouvoir commenter les documents analytiques (activité collaborative) et générer un document de présentation ;

- RT 38 : il est recommandé de pouvoir créer les documents analytiques à partir des données physiques des bases de données et d'une couche sémantique ;

- RT 39 : il est obligatoire de pouvoir positionner librement les objets dans un document analytique ;

- RT 40 : il est recommandé de disposer d'outils de mesure et d'analyse des requêtes soumises ;

- RT 41 : il est obligatoire que les documents analytiques puissent être générés en plusieurs formats (HTML, PDF, MS Office, XML, Open Document, fichier texte) ;

- RT 42 : il est recommandé que chaque document analytique puisse avoir différentes versions au minimum au niveau du contenu ;

- RT 43 : il est recommandé que la génération de documents analytiques soit possible à la volée ou avec planification ; le résultat pourra être mémorisé et/ou avec gestion des versions dans un système documentaire adhérent aux standards JSR 170 ;

- RT 44 : il est recommandé de pouvoir visualiser le diagramme de navigation entre les documents analytiques ;

- RT 45 : il est recommandé de supporter le multifenêtrage (restitution des documents analytiques dans une fenêtre spécifique, plusieurs documents analytiques dans des fenêtres spécifiques) ;

- RT 46 : il est recommandéde pouvoir personnaliser la mise en page notamment par application de feuilles de style ;

- RT 47 : il est obligatoire de disposer d'un procédé permettant la limitation du nombre de lignes ou du temps unité centrale de traitement (CPU: central processing unit) utilisé par une requête lancée par l'utilisateur ;

- RT 48 : il est recommandé de pouvoir effectuer une sauvegarde du résultat après la réalisation de la requête.

*4.1.5. **Administration fonctionnelle et technique des outils.***

- RT 49 : il est recommandé d'être en mesure de stocker dans une base de données au format ANSI2 des statistiques permettant l'utilisation d'un outil tiers de surveillance.

- RT 50 : il est obligatoire de disposer d'outils d'analyse d'activité de la plateforme décisionnelle.

- RT 51 : il est obligatoire de disposer d'un outil d'administration en mode client léger.

- RT 52 : il est obligatoire de disposer d'un outil d'administration permettant de gérer les droits d'accès aux données et traitements.

- RT 53 : il est obligatoire de disposer d'un ordonnanceur de tâches.

4.1.6. ***Sécurité.***

- RT 54 : il est obligatoire de supporter un système d'authentification unique (SSO : single sign-on) avec possibilité de recensement des usagers sur annuaire lightweight directory access protocol (LDAP) ;

- RT 55 : il est recommandé que les solutions décisionnelles (ETL et outils de restitution) supportent le protocole SSL (HTTPS) ;

- RT 56 : il est déconseillé de forcer à utiliser des modules d'extension (plug-ins) spécifiques dans le navigateur ;

- RT 57 : il est obligatoire de disposer d'un moyen de gestion et de supervision des utilisateurs ;

- RT 58 : il est obligatoire que tout utilisateur du système décisionnel ou de l'un de ses composants soit enregistré dans un environnement centralisé, compatible LDAPv3 ;

- RT 59 : il est obligatoire que les utilisateurs (y compris les administrateurs) puissent changer eux même leur mot de passe sur le système ;

- RT 60 : il est recommandé de pouvoir modifier le/les port(s) d'écoute initialisé(s) par défaut lors de l'installation de la plateforme décisionnelle ;

- RT 61 : il est recommandé de pouvoir paramétrer le/les port(s) d'écoute initialisé(s) par défaut lors de l'installation de la plateforme décisionnelle ;

- RT 62 : il est recommandé de pouvoir tracer les actions des utilisateurs connectés ;

- RT 63 : il est recommandé de pouvoir journaliser les actions des utilisateurs.

4.2. **Règles fonctionnelles.**

4.2.1. ***Règles relatives aux fonctionnalités on line anatycal processing.***

- RF 01 : il est recommandé que l'outil dispose d'un assistant pour la création des cubes physiques, logiques ou virtuels(dont la définition des axes et des dimensions) à partir des données de l'entrepôt ;

- RF 02 : il est recommandé de pouvoir disposer de plusieurs hiérarchies pour un même axe (exemple pour l'axe temps : hiérarchie année - mois - semaine et hiérarchie année - trimestre - mois) ;

- RF 03 : il est recommandé de disposer d'un outil d'optimisation des cubes (volumétrie, temps de traitement, temps d'accès) ;

- RF 04 : il est obligatoire de disposer d'une interface de navigation dans les cubes.

4.2.2. ***Règles relatives aux fonctionnalités de restitution.***

- RF 05 : il est recommandé de disposer d'un outil possédant des capacités d'exploration des données :

  - descendant (drill down) ;

  - ascendant (drill up) ;

  - sur les lignes ou les colonnes ;

- zoom à droite et zoom à gauche ;

- d'un tableau ;

- sur les éléments d'un graphique (libellé, histogramme, etc.) ;

- RF 06 : il est recommandé de permettre l'affichage de divers types de graphiques :

- histogramme 2D ;

- histogramme 3D ;

- camembert 2D ;

- camembert 3D ;

- courbe de points ;

- nuage de points.

- RF 07 : il est recommandé que la définition des documents analytiques puisse être stockée et échangée dans un format intelligible et structuré (XML par exemple) ;

- RF 08 : il est obligatoire de disposer d'une interface graphique de construction des documents analytiques ;

- RF 09 : il est obligatoire de disposer d'un aperçu avant impression des documents analytiques ;

- RF 10 : il est recommandé de pouvoir personnaliser les graphiques (annotation des axes, titre, légende, zones de texte).

4.2.3. ***Formation, support.***

- RF 11 : il est obligatoire de disposer des moyens de formation adaptés à la version de la suite décisionnelle utilisée ;

- RF 12 : il est recommandé d'être en mesure de souscrire à un support technique en rapport avec la criticité de l'application ;

- RF 13 : il est obligatoire de disposer des documentations utilisateurs et techniques en français.

Pour le ministre de la défense et par délégation :

L'ingénieur général des télécommunications,
directeur général des systèmes d'information et de communication,

Henri SERRES.

ANNEXE I.

GLOSSAIRE ET ACRONYMES.

Agrégat : un agrégat est une consolidation d'indicateurs pré-calculés dans le but d'offrir de meilleures performances aux restitutions et requêtes.

Application source : une application dont la fonction est de saisir des transactions ou d'autres mesures d'un processus en principe propre à l'organisation. L'application source peut être extérieure à l'organisation, mais saisir des informations nécessaires à l'entrepôt de données.

Attribut : une colonne (un champ) dans une table de dimension.

Authentification/identification : l'authentification a pour but de vérifier l'identité dont une entité se réclame. Généralement l'authentification est précédée d'une identification qui permet à cette entité de se faire reconnaître du système par un élément dont on l'a doté. En résumé, s'identifier c'est communiquer son identité, s'authentifier c'est apporter la preuve de son identité.

Axe d'analyse : voir dimension.

Base de données multidimensionnelle : base de données dans laquelle les données sont présentées sous forme de cubes, par opposition aux tables d'une base de données relationnelle.

Bean : dans le monde Java, un Bean est un composant logiciel réutilisable qui, pour obtenir ce « label », doit respecter certaines conventions sur le nommage des méthodes, la construction et le comportement.

CGAT : cadre général d'architecture technique.

CMTSIC : commission ministérielle technique des SIC. Commission ministérielle spécialisée instaurée par l'arrêté DEFD0600690A du 6 juin 2006.

Cube : nom d'une structure dimensionnelle sur une base de données OLAP (online analytical processing) aussi appelée multidimensionnelle.

Datamining : l'exploration de données, datamining (forage de données) ou encore Extraction de Connaissances (ECD en français, KDD en Anglais), a pour objet l'extraction d'un savoir ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques, et l'utilisation industrielle ou opérationnelle de ce savoir.

Datamart (magasin de données): sous ensemble du datawarehouse contenant les données du datawarehouse pour un métier particulier de l'entreprise (département, direction, service, gamme de produit).

Datawarehouse (entrepôt de données) : base de données partagée, intégrée, historisée, non volatile et organisée pour le support d'un processus d'aide à la décision.

Dimension (table de dimension) : une entité indépendante dans un modèle multidimensionnel, qui sert de point d'entrée vers des mesures additives situées dans la table de faits et comme mécanisme utilisé pour effectuer sur ces mesures des coupes en tranches et en dés.

Document analytique : document issu d'une analyse ou d'une requête.

Entrepôt de données ou « datawarehouse » en anglais : base de données partagée, intégrée, historisée, non volatile et organisée pour le support d'un processus d'aide à la décision.

ESB : l'entreprise service bus est une technologie informatique intergicielle permettant à des applications hétérogènes d'interagir au travers de services standards qu'elles mettent à disposition.

ETL (extract, transform and loading) : outil permettant d'extraire les informations d'une base de données ou de fichiers, de les modifier et d'alimenter une autre base de données.

Fait (table de faits) : le fait concerne l'événement déclencheur exprimé sous la forme d'une valeur généralement numérique et additive et stockée dans une table de faits. Dans un modèle en étoile, la table de faits contient des indicateurs ou mesures, déduits du fait. Ils sont pré-calculés selon des axes de dimension. La table de fait est reliée aux tables de dimensions.

Forage transversal (drill throught) : requêtes visant des données désignées de manière semblable dans plusieurs tables de faits et les réunissant dans un même document analytique. S'effectue généralement en 2 passes.

Forage vers le bas (drill down) : on désigne par forage un mode de navigation dans les données qui consiste, à partir d'un niveau agrégé, de la détailler (par transitions successives) vers des niveaux de granularité plus fins.

Forage vers le haut (drill up) : l'élimination d'un intitulé de ligne ou son remplacement dans un document analytique pour obtenir un total de ligne d'un jeu de réponse. On emploiera également le terme « d'agrégation dynamique ».

Granularité : niveau de détail des faits. Plus la granularité est fine, plus la volumétrie est importante.

GS : groupe de standardisation. Instance de gouvernance technique placée sous l'égide de la CMTSIC*.

HTTP : hyperText Transfer Protocol, protocole utilisé par un butineur (navigateur Web) pour accéder à un serveur Web.

Indicateur : élément de mesure de l'activité (ex : nombre de salariés à l'effectif, masse salariale, heures de formation, ...). Les indicateurs sont qualifiés par des champs qui sont le plus souvent numériques et cumulatifs et des clés pour relier les faits à chaque dimension.

IETF : internet engineering task force : groupe informel, international, ouvert à tout individu, qui participe à l'élaboration de standards pour l'Internet. Produit les RFC*.

Intranet : utilisation des technologies de l'Internet à des fins internes à une entreprise. L'Intranet permet de bénéficier de l'économie d'échelle acquise par les logiciels sur l'Internet et d'outils de développement orientés objet. On peut réaliser maintenant sur l'Intranet la totalité des applications métiers et services communs. L'Intranet nécessite une administration soigneuse des droits d'accès.

Interopérabilité technique : l'interopérabilité des services correspond à la possibilité de fonctionner indifféremment sur des réseaux différents. En informatique, l'interopérabilité signifie l'aptitude de deux ou plusieurs systèmes (logiciels ou matériels) à fonctionner ensemble en utilisant des standards communs.

ISO : international organization standardization, organisation non gouvernementale qui constitue un réseau d'instituts nationaux de normalisation de 157 pays, selon le principe d'un membre par pays.

ISO/CEI 9075 1992 : révision majeure du langage normalisé et standardisé SQL. Elle est connue sous le nom de SQL2 ou SQL-92.

ISO/CEI 10646 2003 : normalise le jeu universel de caractères codés sur plusieurs octets. Elle s'applique à la représentation, à la transmission, à l'échange, au traitement, au stockage, à la saisie et à la présentation des langues du monde sous forme écrite et de symboles complémentaires.

IP : (internet protocol) protocole de communication correspondant globalement au niveau 3 du modèle OSI. Il est utilisé par le réseau Internet et sur les réseaux intranet et extranet.

JDBC : java database connectivity. Ensemble de classes Java qui permet aux applications développées selon cette technologie d'accéder par le biais d'une interface commune à une base de données relationnelle pour laquelle il existe un pilote JDBC.

JSR-054 : java spécification request n° 54, éditée sous l'égide de la société Sun, décrivant la norme JDBC 3.0 qui introduit, entre autres, la notion d'ensemble de connexions partagées (pool de connexions).

JSR-114 : cette JSR complète la JSR-054 en encapsulant l'accès aux données dans un seul Bean nommé rowset.

JSR-168 : standardisation du java community process définissant la gestion du cycle de vie de composant web (portlets) ajoutés à un portail d'entreprise. Les composants supportant ce standard sont portables d'un portail à un autre. La JSR 168 est la spécification des composants web (portlets) définissant le contrat entre les conteneurs de composants web (portlets) et les composants web (portlets).

JSR 170 : standardisation du java community process définissant un ensemble de règles en rapport avec le stockage et la gestion de contenus électroniques. La spécification définit la création de la structure des contenus, la gestion de leurs versions, des recherches avancées (plein texte ou sur les meta-données), des verrous ainsi que l'accès sécurisé aux contenus.

Magasin de données : sous ensemble de l'entrepôt de données contenant les données de l'entrepôt de données pour un métier particulier de l'entreprise (département, direction, service, gamme de produit).

Modèle en étoile : technique d'organisation des données utilisée en décisionnel. L'implantation classique consiste à considérer un modèle en étoile avec au centre la table des faits et les dimensions autour. Les branches de l'étoile sont des relations de 1 à plusieurs. Chaque modèle porte un potentiel de restitutions relatives à un thème.

Modèle en flocon : modèle normalisé dans lequel une dimension plate à une seule table est décomposée en une structure arborescente comportant éventuellement plusieurs niveaux imbriqués. Bien que les flocons de neige puissent être considérés comme un enrichissement du modèle dimensionnel, ils diminuent le plus souvent la facilité de compréhension et les performances de la navigation. Les économies de places sont généralement insignifiantes par rapport au volume total de l'entrepôt de données. Cependant, les tables de dimension normalisées en flocons de neige peuvent exister dans la zone de préparation pour faciliter la maintenance des dimensions.

Modélisation multidimensionnelle : méthodologie de modélisation des données favorisant la performance et la facilité de réalisation de requêtes, à partir d'un ensemble de mesures de base. Dans l'environnement d'un SGBD relationnel, on construit une table de faits comportant un enregistrement pour chacune des mesures. Cette table de faits est ensuite entourée d'un jeu de tables de dimension qui décrit avec précision ce qui est connu au moment de la prise de mesure représenté par chaque enregistrement. La structure caractéristique d'un schéma dimensionnel conduit à utiliser l'appellation schéma en étoile. Grâce à leur symétrie inhérente, les modèles dimensionnels s'avèrent être compréhensibles, prévisibles, extensibles et très adaptés aux requêtes ad hoc. Ce modèle sera le fondement logique de l'architecture OLAP.

ODS (operational data store) : base de données dans laquelle les données brutes sont collectées et traitées en vue d'être intégrées dans le datawarehouse.

OLAP (Online Analytical Processing) : mode de stockage des données spécifiquement optimisé pour l'analyse. Les données sont stockées dans un cube à N dimensions où toutes les intersections sont pré-calculées. Ce mode de stockage apporte un temps de réponse instantané quelle que soit la question initiale.

MOLAP (multidimensional OLAP) : implémentation de la technologie OLAP permettant la navigation multidimensionnelle à partir d'un système de gestion de base de données multidimensionnel.

ROLAP (relational OLAP) : implémentation de la technologie OLAP permettant la navigation multidimensionnelle à partir d'un système de gestion de base de données relationnel.

HOLAP (Hybrid OLAP) : combinaison des technologies ROLAP et MOLAP.

OMG : object management group est une association américaine dont l'objectif est la standardisation et la promotion du modèle objet sous toute ses formes.

POP3 : post office protocol version 3, c'est le protocole qui permet de récupérer les messages sur le serveur de messagerie.

Plugin : (ou *plug-in)* est employé pour désigner un programme qui interagit avec un logiciel principal, appelé programme hôte, pour lui apporter de nouvelles fonctionnalités.

REC-XML-19980210 : recommandation du W3C* concernant le langage de balisage extensible XML en version 1.0.

REC-XML11-20060816 : recommandation du W3C concernant le langage de balisage extensible XML en version 1.1.

RFC : request for comment : série de documents et normes concernant l'Internet. Peu de RFC sont des standards mais tous les standards de l'Internet sont enregistrés en tant que RFC.

SGBDR : système de gestion de base de données relationnelle.
Une base de données construite selon le modèle relationnel stocke les informations dans des structures tabulaires à deux dimensions formées de colonnes et d'enregistrements. Chaque colonne représente un domaine de valeurs et les valeurs des colonnes d'un même enregistrement entretiennent une relation. Des relations sont également définies entre les tables. Elles permettent d'obtenir rapidement une vue logique des informations.

SMTP : simple mail transfer protocol, c'est un protocole de communication utilisé pour transférer les messages vers le serveur de messagerie (MTA*), entre serveurs de messagerie (MTA*).

SOA : service oriented architecture (architecture orientée services).
Modèle d'interaction applicative mettant en œuvre des connexions entre divers composants logiciels. Un service désigne une action exécutée par un composant « fournisseur » à l'attention d'un composant « consommateur », fondé éventuellement sur un autre système.

SOAP : simple object access protocol. Protocole de remote procedure call orienté objet bâti sur XML permettant de formaliser les échanges de services web.

SQL : structured query language ou langage structuré de requêtes.
Le SQL est un pseudo langage informatique de type requête standard et normalisé, destiné à interroger ou à manipuler une base de données relationnelle. SQL a été adopté comme recommandation par l'institut de normalisation américaine (ANSI) en 1986, puis comme norme internationale par l'ISO en 1987 sous le nom de ISO/CEI 9075 - Technologie de l'information - Langage de base de données - SQL.

SQL92 ou SQL2 : version du langage SQL synthétisant le mieux le modèle relationnel. A ce jour, c'est la version la mieux implémentée par les éditeurs de SGBDR.

SSO : single sign on, solution logicielle permettant de s'identifier une seule et unique fois. Un serveur d'authentification joue alors le rôle d'intermédiaire entre toutes les applications du système d'information. Dans de nombreux cas le point d'entrée est le portail d'entreprise.

Table de faits : dans un schéma en étoile (modèle multidimensionnel), c'est la table centrale comportant les mesures numériques de la performance. Elle se caractérise par une clé composite dont chacun des éléments est

une clé étrangère pointant sur une table de dimension.

Thème (ou domaine) : domaine de préoccupation des utilisateurs. Un thème est porté par un ou plusieurs modèle(s) en étoile qui porte un potentiel de restitutions. Ce qui définit précisément un sujet est la contribution à des enjeux identifiés, des cibles utilisateurs distinctes mais partageant les mêmes préoccupations.

XML : eXtended markup language est un langage informatique de balisage générique. Le world wide web consortium (W3C), promoteur de standards favorisant l'échange d'informations sur l'Internet, recommande la syntaxe XML pour exprimer des langages de balisages spécifiques.

XML/A : XML pour l'analyse (XMLA) Protocole utilisé pour accéder aux bases OLAP en mode web services.

XMLRPC : est un protocole RPC (remote procédure call), une spécification et un ensemble de codes qui permettent à des processus s'exécutant dans des environnements différents de faire des appels de méthodes à travers un réseau.

W3C : world wide web consortium.
Le W3C est un consortium fondé en octobre 1994 pour promouvoir la compatibilité des technologies liées à Internet. Le W3C émet des recommandations qui ont valeur de standards auprès des industriels.

Web Service : un service web est un programme informatique permettant la communication et l'échange de données entre applications et systèmes hétérogènes dans des environnements distribués. Il existe plusieurs protocoles comme W3C SOAP, XMLRPC... pour pouvoir réaliser des web services.

Zone de préparation des données : préparation des données : ensemble de processus servant à nettoyer, transformer, combiner, administrer, archiver et préparer des données sources en vue de leur utilisation dans l'entrepôt de données. La zone de préparation de données comporte tout ce qui se trouve entre l'application source et la zone de présentation des données. Aucune requête ne doit viser la zone de préparation de données, qui n'est nullement organisée pour assurer la sécurité, les indexations ou agrégats requis pour l'obtention de bonnes performances et encore moins l'intégration de multiples sources de données.

Zone de présentation des données : lieu où les données sont organisées, stockées et offertes aux requêtes des utilisateurs, aux outils d'accès aux données et à d'autres applications analytiques. Toutes les requêtes doivent s'effectuer vers la zone de présentation des données.

ANNEXE II.
RÉFÉRENCES.

[CCI] Recommandations nationales du cadre commun d'interopérabilité des systèmes d'information publics. Circulaires du premier ministre du 21 janvier 2002 et du 4 décembre 2002 (n.i. BO).

[CGAT] Recommandations du CGAT n° P04 F22 : « Document de référence architecture messagerie V1.2 du 7 avril 2006 ».

[GTEP] : groupe de travail spécialisé dans l'examen des projets SIAG et SIOC. Les modalités de présentation des projets de SIAG au GTEP de la CSIAG sont fixées par l'instruction n° 713/DEF/SGA du 24 juin 2004 (n.i.BO).

[ISO*/CEI 9075 1992*] - SQL*-92 ou SQL2.

[ISO/CEI 10646 2003*] - jeu universel de caractères codés sur plusieurs octets (UNICODE version 5.0.0).

[PRIS] Politique de référencement intersectoriel de sécurité v2.0 du 1er juin 2005 (ADAE et DCSSI).

[REC-XML-19980210*] - spécifications XML* 1.0.

[REC-XML11-20060816] - spécifications XML 1.1 (seconde édition).

[JSR-054*] - Java JDBC* 3.0 specification.

[JSR-114*] - Java JDBC Rowset implémentations.

[JSR-168] - Standardisation du java community process définissant la gestion du cycle de vie de composant web (portlets) ajoutés à un portail d'entreprise.

[RFC 2119] - Key words for use in RFCs to indicate requirement levels (best current practice 03/1997).

[RGI] : référentiel général d'interopérabilité défini par ordonnance n° 2005-1516 du 8 décembre 2005 relative aux échanges électroniques entre les usagers et les autorités administratives, et entre les autorités administratives.

[RGS] : référentiel général de sécurité défini par ordonnance n° 2005-1516 du 8 décembre 2005 relative aux échanges électroniques entre les usagers et les autorités administratives, et entre les autorités administratives.