HAL
open science

Archivage de données SAP

Michel Bolognesi

To cite this version:

Michel Bolognesi. Archivage de données SAP. Base de données [cs.DB]. 2010. dumas-00530235

HAL Id: dumas-00530235

https://dumas.ccsd.cnrs.fr/dumas-00530235

Submitted on 28 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

CENTRE REGIONAL ASSOCIE DE LYON
Centre d'enseignement du pays de GEX

MEMOIRE

présenté en vue d'obtenir

le DIPLOME D'INGENIEUR CNAM

SPECIALITE : INFORMATIQUE

OPTION : INFORMATIQUE, SYSTEMES D'INFORMATION

par

Michel BOLOGNESI

Archivage de données SAP

Soutenu le 17 septembre 2010

JURY

PRESIDENT : **Christophe Picouleau**

MEMBRES : **Bertrand David**
Claude Genier
Jacques Savioz
Jean-Luc Philippona

Remerciements

Difficile d'être exhaustif après ces années de recherches, mais je tiens à remercier toutes les personnes suivantes qui m'ont aidé à mener à bien ce travail de mémoire :

- Monsieur Claude Genier, professeur CNAM, pour son soutien, sa disponibilité et son encouragement pour ce cette formation d'ingénieur.
- Monsieur Jacques SAVIOZ, coordinateur SAP, Romande Energie SA, pour la coordination et l'organisation des séances avec les différents intervenants chez Romande Energie.
- Madame Laure BOULIANNE, juriste, Romande Energie SA, pour tous les renseignements concernant le code des obligations suisse.
- Monsieur Pascal MARMY, responsable système, Romande Energie SA, pour les conseils techniques concernant le système d'archivage.
- Monsieur George CORONADO, responsable de l'application JetForm, société Cisel SA, pour la collaboration dans la création des fichiers pdf.
- Monsieur Jean-Luc PHILIPPONA, responsable étude et développement, neo technologies SA.
- Madame Régine PRALONG, société Affaires classées, pour la relecture et la correction du présent rapport.
- Monsieur Jean-Pierre WERMEILLE, consultant logistique SAP, neo technologies SA, pour la relecture et la traduction en anglais.
- Monsieur Martial GUERRY, analyste programmeur, neo technologies SA, pour les différentes conseils concernant l'archivage.
- Monsieur Laurent MARGELISCH, consultant SAP, neo technologies SA, pour ses explications concernant les modules de finance.
- Madame Carmelina BOLOGNESI, mon épouse, pour son soutien et sa patience.

REMERCIEMENTS **3**

INTRODUCTION **9**

I PREMIERE PARTIE : PERIMETRE DU PROJET **11**

1 NAISSANCE DU PROJET **11**

2 ENVIRONNEMENT **11**

2.1 PRESENTATION DES SOCIETES 11
2.2 GENERALITE SUR LE PGI (PROGICIEL DE GESTION INTEGRE) 12
2.3 PGI SAP EN GENERAL 13
2.4 IMPLEMENTATION DE SAP A ROMANDE ENERGIE 16

3 OBJECTIFS DU PROJET **18**

II LES CONCEPTS DE L'ARCHIVAGE ELECTRONIQUE **21**

1 DEFINITION **21**

1.1 TERMINOLOGIE 22

2 BESOIN DE CONSERVATION **22**

3 ENVIRONNEMENT JURIDIQUE **23**

3.1 DROIT SUISSE DES SOCIETES 23
3.2 SPECIFICITE CONCERNANT LES FACTURES ET PIECES JUSTIFICATIVES AU REGARD DE LA TVA . 26
3.3 PROCEDURE CIVILE 27
3.4 TEMPS DE CONSERVATION 27

4 CONTRAINTES TECHNIQUES DU SYSTEME D'ARCHIVAGE. **28**

4.1 FORMAT DES DOCUMENTS ELECTRONIQUES 28
4.2 CONTRAINTES SECURITAIRE 31
4.3 LA SIGNATURE ELECTRONIQUE 34
4.4 CYCLE DE VIE DE L'INFORMATION ARCHIVEE 36
4.5 SUPPORT PHYSIQUE. 38
4.6 SYSTEME D'ARCHIVAGE ELECTRONIQUE (SAE) 44

III PRINCIPE DE L'ARCHIVAGE SAP **51**

1 DEFINITION **51**

2 MOTIVATIONS **51**

3 PRINCIPE **52**

4 OBJET D'ARCHIVAGE **53**

4.1 PROGRAMMES ASSOCIES AU PROCESSUS ARCHIVE 54

4.2 STRUCTURE D'UN OBJET D'ARCHIVAGE 56
4.3 TUPLES DE TABLES NON ARCHIVES 57
4.4 DEPENDANCE ENTRE OBJETS ARCHIVE 57
4.5 PROGRAMME DE LECTURES DES FICHIERS ARCHIVES 58
4.6 TEMPS DE RESIDENCE DANS LA BASE DE DONNEES 58
4.7 TABLES INFORMATION 59
4.8 FONCTIONS SPECIFIQUES (USER-EXIT) 60
4.9 REMARQUES 60

5 PROCESSUS 60

5.1 PARAMETRAGE PROCESSUS D'ARCHIVAGE (CUSTOMIZING) 61
5.2 PROCESSUS PRINCIPAL 65
5.3 PROCESSUS ANNEXES 67

6 ARCHIVELINK 68

6.1 PRINCIPE 69
6.2 OBJET METIER 69
6.3 TABLES DE LIENS 70
6.4 PROTOCOLE HTTP. 74

7 BILAN DE CONFORMITE 75

IV REALISATION DU PROJET 79

1 PREPARATION DU PROJET 80

1.1 ANALYSE DE LA CROISSANCE DU VOLUME DE LA BASE DE DONNEES 80
1.2 ANALYSE SELON LA LOI PARETO 81
1.3 CIBLER LES OBJETS D'ARCHIVAGE 83
1.4 TEMPS DE RETENTION DES DONNEES 93
1.5 VISUALISATION DES ARCHIVES 95
1.6 EQUIPE PROJET 95

2 *BUSINESS BLUEPRINT* (CONCEPTION GENERALE) 96

2.1 ARCHITECTURE TECHNIQUE 97
2.2 VISUALISATION DES ARCHIVES 98
2.3 AUTORISATIONS 104

3 REALISATION 104

3.1 ARCHIVAGE DES DONNEES 104
3.2 ARCHIVAGE DES DOCUMENTS 107

4 PREPARATION FINALE 111

4.1 TEST ARCHIVAGE DES DONNEES 111
4.2 TESTS ARCHIVAGE DES DOCUMENTS 112

5 *GO LIVE* ET SUPPORT 113

5.1 RESULTAT APRES LE PREMIER CYCLE D'ARCHIVAGE DES DONNEES 113

5.2 RESULTAT APRES L'ARCHIVAGE DES DOCUMENTS 114

V CONCLUSION 117

1 BILAN DU PROJET 117

2 CONSIDERATIONS PERSONNELLES 119

3 AUTRES PROJETS 120

3.1 PAYNET 120

3.2 ARCHIVAGE DE DONNEES SAP CHEZ ALCAN ALUMINIUM VALAIS SA 121

3.3 ARCHIVAGE DES DOCUMENTS ESB BIENNE 122

3.4 GESTION DES DOCUMENTS AUX SERVICES INDUSTRIELS DE LAUSANNE 122

VI ANNEXES 125

1 ILM OU GESTION DU CYCLE DE VIE DE L'INFORMATION. 125

1.1 DEFINITION 125

1.2 OBJECTIF 125

1.3 DEMARCHE 125

1.4 CYCLE DE VIE DE L'INFORMATION 126

2 NOM DU FICHIER LOGIQUE 127

2.1 CREATION D'UN CHEMIN LOGIQUE 127

2.2 DEFINITION DU NOM DE FICHIER LOGIQUE 127

2.3 MOTS RESERVES 128

3 HTTP CONTENT SERVER DE SAP 129

4 CODE SOURCE ZP_ISU_LOAD_FACTURE_PDF 131

5 EXEMPLE DE PROGRAMME DE LECTURE DES ARCHIVES 136

VII BIBLIOGRAPHIE 139

1 LIENS INTERNET 139

2 SUPPORT DE COURS SAP 141

VIII LISTE DES FIGURES 143

IX LISTE DES TABLEAUX 145

Introduction

SAP (SAP Systems Analysis and Program Development) est un progiciel intégré de gestion d'entreprise, composé de plusieurs modules liés étroitement entre eux. Il permet de satisfaire les exigences, même les plus pointues, d'un vaste éventail de sociétés, de nature, de milieu et de métier très différents. Ce progiciel s'appuie solidement sur une base de données, qui, au fil du temps, augmente considérablement de volume. Cet accroissement implique une perte de performances progressive et exponentielle, une augmentation des coûts liés à la gestion du volume de données et un temps inacceptable d'indisponibilité en cas de problème.

L'objectif premier de ce projet d' « Archivage de données SAP » défini il y a cinq ans était d'étudier et de mettre en œuvre une solution visant à maîtriser les problèmes liés à la croissance du volume de la base de données mentionnés précédemment. Cela signifiait améliorer les temps de réponse des transactions, réduire les coûts d'exploitation et diminuer le temps d'indisponibilité du système en cas de problème. Cependant, au cours de ces années, le client, pour lequel ce projet était prévu, a renvoyé la date de mise en œuvre à maintes reprises. Néanmoins, l'apprentissage de l'archivage électronique, l'étude de la mise en œuvre du projet et une réalisation partielle a été réalisée.

De manière subjacente à l'objectif premier du projet, d'autres aspects indissociables à l'archivage ont dû être étudiés tels que la gestion de contenu, le système de stockage et les bases juridiques concernant la conservation des documents. Ces derniers, en particulier la gestion de contenu, ont fait l'objet de plusieurs réalisations non prévues dans les besoins initiaux.

La rédaction de ce mémoire va s'articuler entre plusieurs parties. La première se concentre sur le périmètre du projet : ses acteurs, l'environnement SAP installé et les perspectives. La deuxième vise à démystifier le concept d'archivage électronique. Ensuite, la troisième se focalise sur le fonctionnement des outils d'archivage SAP. La quatrième décrit la partie réalisé du projet. Enfin, la dernière partie sera dédiée à une brève projection des projets annexes et des considérations personnelles.

I Première partie : périmètre du projet

Cette première partie vise à exprimer brièvement la raison de la naissance de ce projet d'archivage, le postulat de base et les évolutions de ce dernier, influencé par l'environnement et par l'évolution des besoins exprimés par les acteurs. Tout cela aboutira à la définition du périmètre du projet.

1 Naissance du projet

Le projet est né de manière atypique, c'est-à-dire que le besoin n'a pas été exprimé par le client, mais par l'éditeur du progiciel SAP. Ce dernier préconise qu'après un certain temps d'utilisation du progiciel, il faut prévoir et mettre en œuvre de l'archivage. Neo technologies SA, mon employeur, a proposé à Romande Energie SA, son client, d'étudier et mettre en place un système et des processus d'archivage.

Le projet n'a pas encore abouti car ce besoin n'est pas vital pour le client et ce dernier a repoussé sa réalisation à maintes reprises car il avait d'autres priorités.

2 Environnement

2.1 Présentation des sociétés

2.1.1 neo technologies SA

neo technologies SA est une société de services informatiques et de consulting qui offre des prestations de conception et réalisation, d'exploitation et support de systèmes d'information. Reconnue sur le marché comme centre de compétences « Utilities », elle réalise des projets de développement et d'intégration de nouvelles technologies dans ce secteur. Outre le marché de l'énergie qu'elle maîtrise depuis de nombreuses années, neo technologies œuvre dans le secteur public et offre aux collectivités un service et des solutions de gestion s'appuyant sur la plateforme SAP.

2.1.2 Romande Energie SA

Romande Energie SA est la société opérationnelle du Groupe Romande Energie coté en bourse. Ce dernier est le cinquième fournisseur en Suisse de production, distribution et vente d'électricité.

L'Etat de Vaud et plus de deux cents communes vaudoises (dont 130 liées par une convention) sont, ensemble, ses plus gros actionnaires. Romande Energie détient des participations importantes dans Alpiq et les Forces Motrices Hongrin-Léman.

2.1.3 SAP

SAP est une société allemande fondée en 1972. Grâce à ses nombreuses années d'expérience et son incessante innovation, SAP est devenu un véritable leader mondial dans le logiciel de gestion d'entreprise[1] et inter-entreprise.

2.2 Généralité sur le PGI (progiciel de gestion intégré)

Le progiciel de gestion intégré se caractérise par une conception unique pouvant couvrir l'ensemble des besoins du système d'information[2] de l'entreprise. Dans ce dessein, il est subdivisé en modules qui couvrent chacun un des domaines ou périmètres de gestion de l'organisation.

En cas d'impact d'un module, l'information est mise à jour en temps réel dans l'ensemble des autres modules associés. Par exemple, la saisie d'une vente génère automatiquement une écriture comptable en partie double dans le journal des ventes, avec calcul automatique de la TVA collectée. Le grand livre et le compte de résultat sont automatiquement impactés.

Grâce à ce lien étroit entre modules, le PGI facilite la recherche et l'analyse d'informations lors d'audits.

Chaque entreprise a ses propres flux et modes de fonctionnement. Dans ce but, un PGI est paramétrable à souhait et peut être adapté pour le développement de programmes spécifiques. Cependant, les développements sont, en général, réduits au minimum, car ils sont très coûteux à la réalisation et surtout à la maintenance, sans compter les risques de dysfonctionnements mal maîtrisés sur les autres modules.

---

[1] Soit nommé par l'acronyme anglais ERP « Enterprise Ressource Planning », soit, dans le monde francophone, par la dénomination PGI (Progiciel de gestion intégré). Cependant, la terminologie anglo-saxonne prime.

[2] Un système d'information (noté SI) représente l'ensemble des éléments participant à la gestion, au stockage, au traitement, au transport et à la diffusion de l'information au sein d'une organisation. (*Wikipédia*)

2.3 PGI SAP en général

Comme déjà évoqué, SAP, depuis sa fondation, a progressé et innové aussi bien fonctionnellement que techniquement, pour satisfaire une clientèle très diverse, de milieux et domaines très différents. On pourrait presque affirmer que SAP a créé son propre monde très riche et varié, dont je vais me limiter à mentionner les principales solutions en m'appuyant sur la Figure 1 : intégration des solutions mySAP [3].

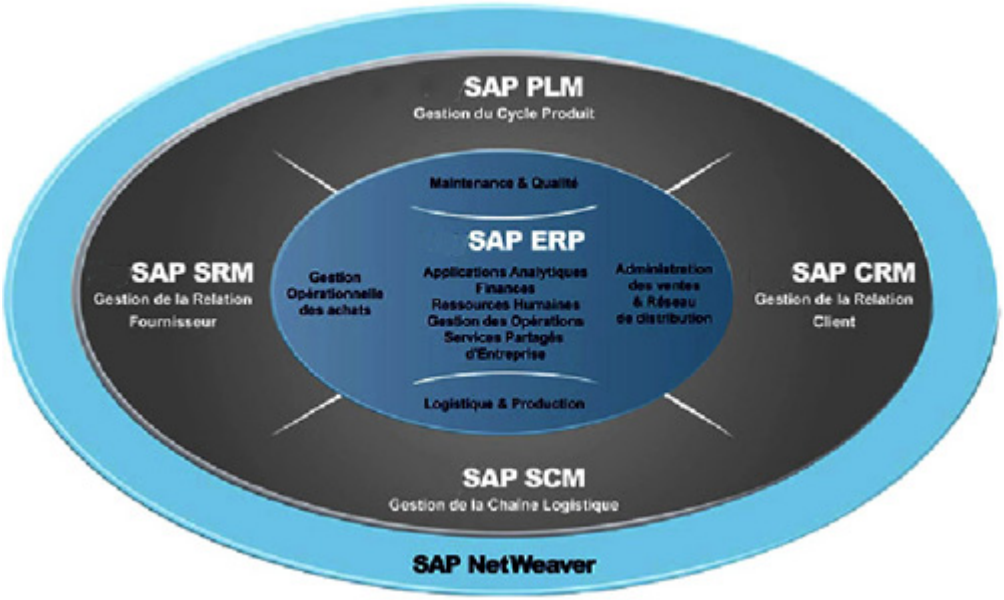


Figure 1 : intégration des solutions mySAP

2.3.1 SAP ERP *(Entreprise Ressource Planning)*

SAP ERP est une suite de modules de gestion relative à la finance, aux services d'entreprise, aux applications analytiques, aux fonctions de self service et à la gestion des ressources humaines et des opérations. En quelques mots, c'est le vrai PGI de l'offre SAP. Le projet « Archivage de données SAP » se déroule dans cet environnement.

2.3.2 SAP PLM *(Product Lifecycle Management)*

SAP PLM est une application spécifique pour la gestion des cycles de vie des produits. PLM est un environnement qui assure un accès rapide et fiable aux informations les plus récentes pour tous les acteurs du processus de développement du produit, de fabrication et de maintenance.

[3] Régulièrement, l'éditeur allemand change et modifie la dénomination de ses produits. Cependant, le contenu et les fonctionnalités deviennent beaucoup plus stables.

2.3.3 SAP CRM *(Customer Relationship Management)*

SAP CRM est une application centrale sur laquelle les différents canaux de distribution traditionnels viennent se greffer pour la gestion de la clientèle. Elle fournit les outils nécessaires pour le support et le service de la clientèle, ainsi que pour le marketing et la vente.

2.3.4 SAP SRM *(Supplier Relationship Management)*

SAP SRM est une application pour la gestion des relations fournisseur. Elle met à disposition un système de collaboration pour toutes les fonctions qui touchent à la gestion des achats et à la relation fournisseur.

2.3.5 SAP SCM *(Supply Chain Management)*

SAP SCM est une application pour la gestion globale de la chaîne logistique allant du fournisseur du fournisseur au client du client.

2.3.6 Bus d'intégration SAP NetWeaver

Depuis 2004, SAP a mis à disposition une plate-forme applicative et d'intégration nommée NetWeaver. Elle permet la communication entre la suite logicielle mySAP Business Suite, représentées par la figure 1, et toutes autres tierces applications. Ce socle technologique est compatible avec les normes internet telles que le protocole HTTP ou le langage XML, ainsi que les services Web. NetWeaver se caractérise également par son ouverture et son interopérabilité avec Microsoft .NET et les environnements J2EE (Java 2 Platform Enterprise Edition) tels que IBM WebSphere.

2.3.7 Architecture SAP

SAP ERP est un progiciel client/serveur n-tiers subdivisé en trois couches : présentation, application et base de données selon la figure 2.

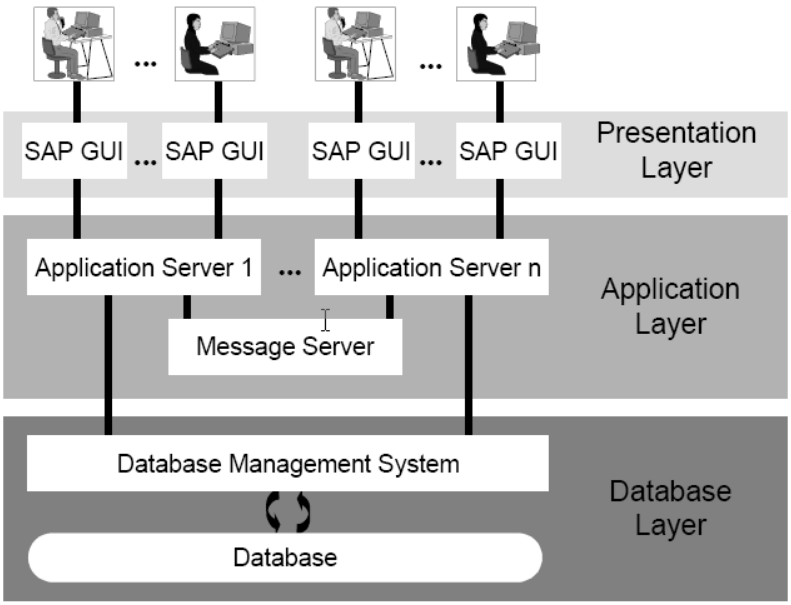


Figure 2 : modèle en trois couches.

2.3.7.1 Couche représentation (Presentation layer)

SAPgui (*graphical user interface*) est la couche représentation installé sur le poste des utilisateurs pour visualiser ou modifier les textes, les champs de la base de données, les étapes d'un flux de travaux, etc..

2.3.7.2 Couche applicative (Application Layer)

La couche applicative, comme son nom l'indique, est un ensemble de programmes, d'applications exécutées par un ou plusieurs serveurs. Le lien entre ces serveurs est fait par un dispatcher. Ce dernier reçoit les requêtes de la couche présentation et l'achemine vers le serveur disponible. Le serveur interagit avec la couche présentation, le gestionnaire de base de données et avec toutes autres connections externes tel que NetWeaver mentionné précédemment.

2.3.7.3 Couche base de données (DataBase Layer).

La couche base de données est un système central contenant toutes les données. Dans ce dessein, un gestionnaire de base de données (Database Management System) est trait d'union entre la couche applicative et une base de données. Dans le cadre de ce projet, ORACLE version 10 est la base de données utilisée.

2.3.7.4 Paysage système

Chaque solution SAP est installée dans un paysage système distinct défini par trois environnements représentés par la figure 3: développement, qualité et production. Un

environnement est une instance de la couche applicative et de base de données installée sur un ou plusieurs serveurs appartenant au même domaine réseau.

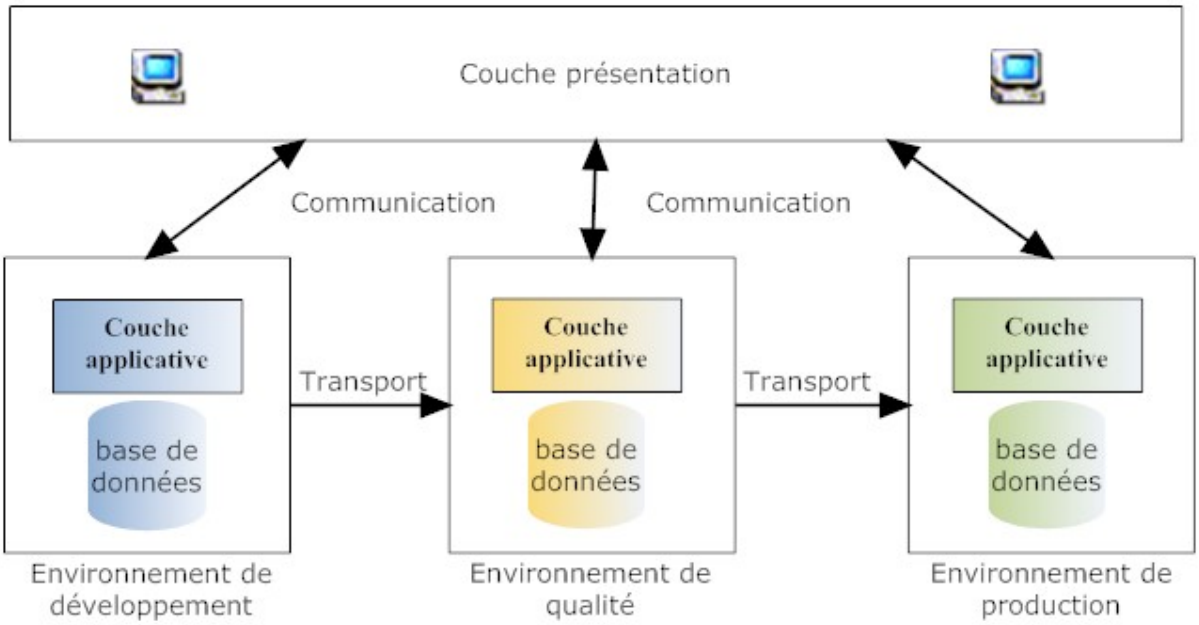


Figure 3 : paysage système

L'environnement de développement est utilisé pour le codage des programmes spécifiques, pour le paramétrage et pour les tests sommaires. La vérification de la pertinence de ces derniers est accomplie par des tests approfondis dans l'environnement de qualité. Après approbation, les nouvelles fonctionnalités sont utilisées dans l'environnement de production. Le passage du code des programmes spécifiques et du paramétrage se fait par l'appel de fonctions distantes entre les environnements appelés transport.

2.4 Implémentation de SAP à Romande Energie

Depuis 2002, Romande Energie a implémenté SAP pour couvrir la quasi-totalité des besoins de son système d'information, principalement par les modules suivants :

- **IS-U** (*Industrie system utilities*) : c'est un module spécialisé pour toutes les entreprises du domaine de la vente et de la distribution d'électricité, gaz, eau, multimédia et chauffage à distance. Ce dernier comprend plusieurs parties :
    - La gestion des données de base : commerciales (les clients, les contrats, etc.) et techniques (les appareils, les installations, etc.).
    - La gestion du processus de facturation et de comptabilisation
    - La gestion de la maintenance par des contrôles périodiques des appareils du réseau, la délivrance de certificats de conformités des installations, etc.
    - La gestion des échanges des courbes de charges de consommation.

- **MM** (*Material Management*) : c'est le module logistique de SAP. Il gère les achats d'articles et les stocks d'articles (les mouvements de stocks : entrées et sorties, transferts de stocks). MM permet en particulier le calcul des besoins, la gestion des achats et des réapprovisionnements, mais aussi les demandes d'achat et les contrats. MM gère aussi les emplacements magasin (WM Warehouse Management) et contrôle les factures d'achat.
- **SD (***Sales and Distribution***) :** c'est le module d'administration des ventes. Il gère les appels d'offres, les offres, les contrats, les commandes clients, les expéditions et livraisons, les remises, la facturation et le système d'information commercial.
- **PM** (Plant Maintenance) : c'est le module de maintenance des appareils installés sur le réseau de distribution électrique. PM permet de décrire le référentiel technologique des postes techniques et des équipements. Il traite également les ordres de maintenance jusqu'à l'achèvement de l'ensemble des opérations d'entretien des équipements et des appareils.
- **FI** *(Financial) :* c'est le module comptable contenant toutes les écritures comptables de ventes, d'achats et d'immobilisations, qui se centralisent dans la comptabilité générale. FI se subdivise en plusieurs sous modules :
    - FI-AR (Accounts Receivable) : comptabilité client
    - FI-AP (Accounts Payable) : comptabilité fournisseur
    - FI-AM (Assets Management) : comptabilise les immobilisations
    - FI-GL (General Ledger) : comptabilité générale
- **CO** *(Costing)* : c'est le module de contrôle de gestion de l'entreprise. Il permet l'analyse des coûts par centre de profits, par catégorie de produits/services/projets. Il permet également l'édition du compte de résultat, de reportage d'activités et d'analyses financières.
- **HR** *(Human Resources)* : c'est le module de la gestion du personnel. HR permet de gérer les recrutements, les paies des employés, les compétences des employés, de suivre les temps de travail et les évolutions de carrière, de gérer les demandes et les frais de déplacement.

3 Objectifs du projet

Après la présentation des acteurs et de l'environnement, ce travail se consacre aux objectifs de ce projet d'archivage de données SAP tels qu'ils m'ont été attribués :

1. **Découvrir** le milieu de l'archivage électronique, son langage, ses pratiques et son application de manière générale.
2. **Décrire** les outils propres au système SAP dans le domaine de l'archivage de données, ainsi qu'un système d'archivage pour le stockage des fichiers d'archive.
3. **Cibler** les objets d'archivage potentiels par une analyse des besoins techniques et fonctionnels tels que la volumétrie des données, les problèmes de performances, la suppression d'entités organisationnelles obsolètes.
4. **Définir** les besoins en termes de rétention de données et les contraintes légales de données.
5. **Présenter** une structure de projet complète avec les intervenants ainsi que les responsabilités dans un projet d'archivage SAP.
6. **Réaliser**, dans un premier temps, une maquette dans le système de qualité pour valider les concepts théoriques étudiés. Puis, dans un deuxième temps, réaliser la mise en œuvre de l'archivage dans le système productif en orchestrant tous les intervenants de la structure du projet présenté.
7. **Planifier** l'exécution périodique des lots de programmes d'archivage et l'archivage à long terme.

II Les concepts de l'archivage électronique

Cette partie vise à donner un aperçu théorique et générique de l'archivage électronique, sans prendre en considération les spécificités de l'éditeur allemand. Il s'agit en quelque sorte de l'abécédaire de l'archivage électronique, largement inspiré par l'ouvrage « Dématérialisation et archivage électronique » (voir VII Bibliographie référence 1) et la compilation de plusieurs ressources trouvées sur la toile (voir VII1Liens internet).

Pour appréhender le sujet, le premier chapitre va définir l'archivage électronique. Il sera suivi par les besoins de conservations qui convergent principalement aux contraintes légales. Par conséquence le troisième chapitre est consacré aux lois et aux codes des obligations qui définissent dans une certaine mesure les caractéristiques d'un système d'archivage. Le dernier chapitre relève le défi de décrire comment ces caractéristiques peuvent être mises en œuvre par les outils et technologies actuelles.

1 Définition

L'archivage électronique consiste à archiver des données sous forme électronique sur un support électronique dédié à cet effet. Les informations sont donc dématérialisées, c'est-à-dire converties en données numériques (pour des informations issues de supports analogiques tels que papier, film, microfilm, etc.) ou simplement transposées (pour des informations issues de formats électroniques).

A noter que par abus de langage, les termes de sauvegarde et d'archivage sont parfois confondus. Cependant, ces notions sont bien différentes :

La sauvegarde : elle a comme finalité de permettre une copie des données d'origine dite « copie de sécurité » afin d'éviter de les perdre en cas de dysfonctionnement du dispositif sur lequel elles sont enregistrées. La durée de conservation est déconnectée de la valeur du contenu et relative à la périodicité de la sauvegarde.

L'archivage [4]: il est défini comme l'ensemble des actions visant à identifier, recueillir, classer et conserver des informations, en vue de consultation ultérieure, sur un support adapté et sécurisé, pour la durée nécessaire à la satisfaction des obligations légales ou des besoins d'information. Contrairement à la sauvegarde, les données archivées sont considérées comme figées (non modifiables) et aisément consultables.

[4] Selon norme NF Z42 013

1.1 Terminologie

Avant de continuer, il est opportun de définir certains termes utilisés dans le contexte de l'archivage électronique: donnée, information, connaissance, document numérique et dossier.

- Donnée : est la représentation d'une idée ou d'un fait sous une forme conventionnelle, codée et structurée pour que son traitement puisse être automatisé, généralement à l'aide d'outils informatiques.
- Information : est un ensemble de données liées entre elles et porteuses de sens.
- Connaissance : est une information que son utilisateur, dans un contexte donné, entend explorer pour une action particulière.
- Document numérique : est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs.
- Dossier : est constitué par l'ensemble de document ou /et d'informations qui tracent et documentent une action.

2 Besoin de conservation

L'archivage électronique est une réponse au besoin d'une entité de conserver des documents. La conservation est liée principalement à trois facteurs majeurs : l'obligation légale et d'audit, le contentieux et le besoin opérationnel de l'entité. Néanmoins, ce sont principalement les exigences légales et d'audit qui ont influencé les bonnes pratiques de mise en œuvre d'un système d'archivage.

- L'obligation légale et d'audit : est encadrée par la réglementation et les codes d'obligations cantonales, fédérales ou internationales. Ces derniers sont requis en cas d'audit fiscal ou financier, ou d'audit interne.
- Le contentieux : est la capacité de fournir, lors d'un litige, des documents probants, authentiques et dans le temps imparti. Cet ensemble est le plus difficile à cerner car la matière contentieuse est, par nature, imprévisible. L'indentification de ces archives relève d'une évaluation des risques de non disponibilité de l'information en cas de recherche. La valeur de preuve étant décidée par le juge, il s'agit d'anticiper au mieux la survenance et l'évolution des litiges.

- Le besoin opérationnel de l'entité : constitue l'ensemble des informations qui sont utiles à l'entité pour son fonctionnement soit par le passé, soit pour le futur. Parmi ces informations, il y a les comptes rendus de séances stratégiques, les procédés, les projets de brevets abandonnés, les études, etc.

3 Environnement juridique

Les sociétés suisses ont l'obligation de conserver les documents au regard du droit des sociétés et du droit fiscal. Les entités soumises à surveillance, tels les intermédiaires financiers ou les banques, ont des exigences supplémentaires relatives à cette obligation de conservation. Ce chapitre vise à donner un bref aperçu de la législation fédérale concernant l'archivage électronique. Le but est de mettre en évidence cette problématique. Il va de soi que des particularités demeurent selon les domaines (notamment les professions médicales) et que des prescriptions cantonales entrent également en ligne de compte.

De manière générale, les lois et règlements servent de base à la détermination des règles de conservation, qui spécifient aussi bien la durée que le support des documents. Malheureusement, en Suisse, peu de textes fixent des délais relatifs à la conservation des documents pour les entreprises privées. Le Code des Obligations (CO) et l'Ordonnance concernant la tenue et la conservation des livres de comptes (Olico) donnent des indications uniquement pour les documents comptables. Les lois sur les archives publiques ou d'autres textes légaux, comme la LPD (Loi Fédérale sur la protection des données), apportent quelques compléments. Il est également souvent admis de prendre modèle sur certaines pratiques étrangères.

3.1 Droit suisse des sociétés

En droit suisse, toute entité inscrite au Registre du commerce a l'obligation de conserver les documents commerciaux pendant une période de 10 ans ou plus selon les cas. Cette obligation s'applique en particulier aux sociétés anonymes et aux sociétés à responsabilité limitée.

Les documents qui entrent dans le champ d'application de cette obligation générale de conservation et d'archivage sont :

- les documents comptables de la société, à savoir les divers états financiers que la société est tenue de conserver en vertu des dispositions de droit comptable (comme le bilan, les comptes de résultats, etc.),
- les justificatifs y afférents (factures, reçus, relevés de comptes, etc.),

- les procès-verbaux (assemblées générales, réunions du conseil d'administration) ,
- les contrats et
- la correspondance commerciale (entrante et sortante).

3.1.1 Correspondance

Concernant la correspondance en particulier, les entreprises devraient au moins conserver celle qui présente un intérêt du point de vue comptable. D'une manière générale, les documents purement internes, tels notes, projets, correspondance interne et documents similaires, n'ont pas à être conservés et archivés.

Il n'est néanmoins pas toujours évident de tracer une limite claire entre documents internes et externes. A titre d'exemple, un contrat de travail sera qualifié de document externe dans ce contexte.

3.1.2 Forme d'archivage

Le droit suisse des sociétés contient des dispositions concernant la forme de l'archivage : seuls les comptes de résultats et de bilans doivent être conservés sous forme écrite et dûment signés. Par contre, les documents sociaux, les documents comptables, la correspondance et les autres documents peuvent être conservés sous forme écrite ou sur un support électronique ou comparable, à condition que l'authenticité soit garantie. Les données sur supports électroniques doivent toutefois être accessibles à tout moment.

3.1.3 Disposition sur la conservation électronique

L'ordonnance fédérale sur la tenue et la conservation des livres de comptes (Olico art. 957) donne certaines directives. Ces dernières sont considérées de la bonne pratique pour la conservation des documents de manière générale. Voici le contenu des principaux testes :

Intégrité (authenticité et infalsifiabilité) (art.3)

« Le mode de tenue, de saisie et de conservation doit garantir que les livres, les pièces comptables et la correspondance ne puissent être modifiés sans que la modification soit apparente ».(art. 3)

Disponibilité (art. 6)

« Jusqu'à la fin du délai de conservation, toute personne autorisée doit pouvoir, en tout temps et dans un délai raisonnable, consulter et vérifier les livres, les pièces comptables et la correspondance. »

Organisation (art.7)

« Les informations archivées sont séparées des informations actuelles, ou sont signalées de telle manière que la distinction soit possible. La responsabilité des données archivées doit être clairement réglée et consignée dans un document.

L'accès aux données archivées doit être possible dans un délai raisonnable. »

Archives (art.8)

« Les informations sont inventoriées systématiquement et protégées contre les accès non autorisés. Les consultations et les accès sont enregistrés. Ces enregistrements sont soumis à la même obligation de conservation que les supports de données. »

Supports d'information autorisés (art.9)

1. *« Sont autorisés pour la conservation de documents:*
   a. *les supports d'information non modifiables, notamment le papier, les supports d'images et les supports de données non modifiables.*
2. *Les supports d'information modifiables si:*
   a. *des procédés techniques (p. ex. signature électronique) sont utilisés, qui garantissent l'intégrité des informations enregistrées;*
   b. *le moment où les informations ont été enregistrées peut être prouvé sans possibilité de falsification (p. ex. grâce à un système d'horodatage);*
   c. *les autres prescriptions relatives à l'utilisation du procédé en question qui existent au moment de l'enregistrement sont respectées, et les procédures et les modes d'utilisation de ces supports sont consignés et les informations nécessaires (protocoles, journal de bord des connexions [log files]) sont également conservées.*

   Les supports d'information sont réputés modifiables lorsqu'ils peuvent être modifiés ou effacés sans que l'opération soit détectable sur le support de données lui-même (p. ex. bandes magnétiques, disquettes magnétiques ou optico-magnétiques, disques durs ou disques amovibles, disques à l'état solide [solid-state]). »

Contrôle et migration des données (art. 10)

Les supports informatiques doivent faire l'objet d'un contrôle de leur intégrité et de leur lisibilité. Les données peuvent être modifiées et transférées sur un autre support et/ou format dans le cas d'une migration, mais les éléments suivants doivent être garantis :

- Les informations restent complètes et exactes.
- La disponibilité et la lisibilité continuent de satisfaire aux exigences légales.
- La migration doit faire l'objet d'un procès-verbal.

3.1.4 Lieu de conservation

Concernant le lieu d'archivage des documents, il n'y a aucune disposition en droit suisse qui impose que ces documents doivent être conservés en Suisse (mise à part certaines obligations concernant la législation en matière de blanchissement d'argent[5]). Il n'y a donc, aucun obstacle à ce que des documents soient archivés à l'étranger, pour autant que les personnes habilitées puissent y accéder dans un délai raisonnable.

3.2 Spécificité concernant les factures et pièces justificatives au regard de la TVA

La conservation des factures et des autres pièces justificatives est une des obligations fondamentales des assujettis. Elle permet à l'administration de procéder aux vérifications de la comptabilité et des décomptes au regard de la TVA (art. 43 OLTVA).

Les données et les informations qui sont transmises ou conservées électroniquement ou de façon analogue et qui sont pertinentes pour la déduction de l'impôt préalable, pour la perception ou pour le recouvrement de l'impôt ont la même force probante que celles qui sont lisibles sans moyen auxiliaire, pour autant que les conditions suivantes soient remplies:

- Preuve de l'origine
- Preuve de l'intégrité
- Incontestabilité de l'envoi et de la réception

Les assujettis, doivent garantir que ces données et ces informations puissent être lisibles en tout temps pendant la durée du délai légal de conservation. A cet effet, l'assujetti doit mettre à disposition les moyens nécessaires. Il doit, sur demande, permettre à l'Administration fédérale des contributions d'accéder directement aux données conservées électroniquement. En ce qui concerne la tenue et la conservation par un moyen électronique ou de façon analogue des livres comptables, des pièces comptables et de la correspondance commerciale, les prescriptions de l'art. 957 du code des obligations (CO) sont applicables par analogie.

---

[5] Des exigences spécifiques résultant de la législation et de la réglementation bancaire et de la loi fédérale sur le blanchiment d'argent, prévoit que les intermédiaires financiers, auxquels appartiennent les banques, doivent établir et conserver les documents concernant les transactions et les clarifications obtenues en vertu de la législation anti-blanchiment. A ce sujet le serveur sur lequel ces documents conservés sous une forme électronique doit être situé en Suisse. Si tel n'est pas le cas, il faut que l'intermédiaire financier dispose en Suisse soit d'une copie papier des documents pertinents, soit d'une copie électronique.

3.3 Procédure civile[6]

Lorsqu'une procédure civile advient entre deux parties adversaires, des titres sont présentés pour prouver les faits. La partie qui invoque un titre doit en prouver l'authenticité si la partie adverse la conteste sur la base de motifs suffisants (art. 175).

Ces titres sont des documents, tels que des écrits, des dessins, des plans, etc. (art. 174). Cela n'importe pas qu'ils soient sur un support traditionnel ou digital, ou qu'ils soient issus d'un support électronique dès l'origine ou d'un document traditionnel scanné. Cela résulte également du droit à la preuve (art. 150). On peut même faire valoir une copie classique ou de produits de l'électronique à valeur probante lorsque l'archivage de cette dernière est conforme aux règles du doit commercial (Olico).

3.4 Temps de conservation

Le temps de conservation légal des documents commerciaux est fixé en général à dix ans. Cependant, il convient d'éviter de confondre délai de conservation et délai de prescription. En effet, si certaines actions judiciaires dans lesquelles un document est requis sont toujours en cours avant le terme de la durée de prescription, ce dernier ne doit en aucun cas être détruit.

C'est la raison pour laquelle plusieurs experts en archives[7] préconisent un délai de rétention bien supérieur. Le tableau ci-dessous en donne un aperçu :

[6] La procédure civile est un ensemble de moyens mis à la disposition des particuliera par l'Etat pour leur permettre d'obtenir la consécration de leurs droits privés, qu'ils soient contestés ou subordonnées au prononcé d'une décision.

[7] Voir http://www.archives.ch/

Type de document	Délai légal	Délai suggéré
DOCUMENTS COMPTABLES Journal Livre d'inventaire, Grand livre, Bilan annexe, Comptes de résultat, Journaux auxiliaires, Balances, etc.	10 ans après la clôture de l'exercice	11 ans après la clôture de l'exercice
FACTURES ET PRÊTS Factures, Contrats d'assurance, Contrats de prêts, etc.	10 ans après la clôture de l'exercice.	11 ans après la clôture de l'exercice
DOCUMENTS FISCAUX Impôts, TVA, etc.	10 ans après la clôture de l'exercice.	11 ans après la clôture de l'exercice
DOCUMENTS BANCAIRES Relevés de comptes, Bordereaux de remises, etc.	10 ans après la clôture de l'exercice.	11 ans après la clôture de l'exercice
CONTRATS COMMERCIAUX Marchés, Plans, Cahiers des charges, etc.	Aucun délai n'est fixé par la loi	30 ans
INVESTISSEMENTS Contrats d'acquisition, Contrats de cessions, Contrats de participations, etc.	Aucun délai n'est fixé par la loi	30 ans
SALAIRES Livre de paie, Bulletin de paie, etc.	10 ans	80 ans
SAUVEGARDES INFORMATIQUES	Aucun délai n'est fixé par la loi	10 ans

Tableau i: Temps de conservation

4 Contraintes techniques du système d'archivage.

Ce chapitre relève le défi de décrire comment les lois et les obligations vues précédemment peuvent être mises en œuvre par les outils et technologies actuelles, en abordant notamment les notions suivantes : le format de fichier, la sécurité, la signature électronique, le cycle de vie de l'information, le support physique et le système d'archivage.

4.1 Format des documents électroniques

Come défini précédemment, un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une

lisibilité partagée entre son concepteur et ses lecteurs. Ces règles sont régies par des formats de fichiers.

La lisibilité des formats de fichiers est liée au matériel physique, au système d'exploitation et au logiciel. Il est possible de distinguer les types de format que l'on peut rencontrer du plus « ouvert » au plus « fermé » :

- Standard : lorsqu'ils sont produits par un organisme de normalisation et que leur utilisation est libre.
- Ouvert : quand les spécifications sont libres.
- Propriétaire : lorsqu'ils sont définis par une entreprise privée et que leur utilisation est soumise à des droits.
- Fermé : quand les spécifications sont tenues secrètes par leur propriétaire.

Les formats fichier peuvent, pour différentes raisons, disparaître au cours du cycle d'archivage du document. Pour cette raison, il est nécessaire de pouvoir accéder aux spécifications dudit format et ainsi pouvoir écrire au besoin un programme destiné à l'interpréter. Ce sont les raisons pour lesquelles, dans le cadre d'archivage, il est fortement recommandé d'utiliser des formats standards ou ouverts tel que XML ou PDD/A.

4.1.1 XML (*eXtensible Makup Language*)

XML est un langage informatique de balisage générique. Il sert essentiellement à stocker/transférer des données au format texte Unicode[8] structurées en champs arborescents. Ces principaux atouts sont :

- Format standard ouvert, gratuit et libre de droits. Il est défini sous la responsabilité du World Wide Web Consortium (W3C).
- Facilement lisibles et compréhensibles car il est au format texte unicode.
- XML est un métalangage extensible à souhait et permet de créer ses propres balises[9].

[8] Unicode est une norme informatique, développée par le Consortium Unicode, qui vise à donner à tout caractère de n'importe quel système d'écriture un nom et un identifiant numérique, et ce de manière unifiée, quelle que soit la plate-forme informatique ou le logiciel (voir http://fr.wikipedia.org/wiki/Unicode)

[9] Marque qui permet d'identifier un élément d'un fichier, sa fonction, son emplacement (http://www.linternaute.com/dictionnaire/fr/definition/balise/)

- Sa structure arborescente permet de modéliser la majorité des problèmes informatiques.
- Il est universel et portable : le format texte tient compte des différents jeux de caractères est compréhensible par tous les systèmes d'exploitation.
- Il est déployable : il peut être facilement distribué par n'importe quels protocoles à même de transporter du texte, comme HTTP[10].
- Il a un niveau élevé d'intégrabilité: un document XML est utilisable par toute application pourvue d'un *parser*[11].

Grâce à tous ses atouts, XML facilite la migration des documents qui peuvent ou doivent être migré lors du cycle de vie de conservation.

4.1.2 PDF/A

Avant de parler de PDF/A (« A » comme archive), il convient de préciser ce qu'est le format PDF. Le PDF (*Portable Document Format*) a été initialement créé par Adobe Systems, pour gérer ses échanges de documents en interne. Son usage s'est ensuite largement répandu principalement grâce au fait qu'il permettait de reproduire fidèlement des textes et images complexes à l'écran comme sur l'imprimante, quelle que soit l'application ou la plate-forme utilisée pour le créer.

Le PDF est un langage de représentation de documents, riche en fonctionnalités qui n'a cessé de s'étoffer au travers de ses versions successives. Les documents PDFs peuvent être simples ou complexes, structurés ou non structurés. Ils peuvent contenir du texte, des images, des graphiques et même des contenus multimédia (documents sonores ou images animées). Les annotations, les métadonnées, les liens hypertextes, les signets ainsi que la signature électronique sont également supportés par ce format de fichier.

Depuis juillet 2008, les spécifications et publications sont sous la tutelle de l'organisme international de normalisation ISO sous la référence ISO 32000-1.

PDF/A est une famille de normes ISO (ISO 19005) qui définit des sous-ensembles du format PDF à des fins d'archivage à long terme de documents électroniques. Cette norme définit les exigences pour la création de documents adaptés à l'archivage, en se basant sur le

[10] *hypertext markup language* : langage informatique permettant de désigner la structure d'un document à l'aide de balises et d'établir des liens hypertextuels entre ses différents constituants et d'autres textes. (http://fr.encarta.msn.com/dictionary_2016011710/HTML.html)

[11] Logiciel permettant d'analyser un code XML et d'en déterminer sa structure syntaxique afin d'effectuer divers traitements, comme par exemple la visualisation.

format de fichier PDF. Elle spécifie en détails quels sont les types de contenu autorisés et ceux qui ne le sont pas. Ces spécifications sont destinées à assurer la lisibilité dans le temps des fichiers, indépendamment des applications et des systèmes d'exploitation dans lesquels ils ont été produits. PDF/A repose sur la version 1.4 de PDF. Pour assurer la portabilité dans le temps, certaines fonctionnalités contenues dans le PDF 1.4 sont exclues de la norme, comme le mode transparent, l'inclusion de séquences sonores, d'images animées, d'images fixes comprimées avec JPEG 2000, etc. En revanche, le PDF/A impose certaines fonctionnalités comme l'incorporation des polices de caractères nécessaires à l'affichage des textes, l'ajout des métadonnées au format XMP[12], etc.

Les principaux avantages de ce format PDF/A sont :

- Spécifications disponibles : La publication de la norme garantit la possibilité pour toute entreprise ou toute administration de pouvoir accéder librement aux spécifications du format et donc de pouvoir créer librement des logiciels de lectures ou de production de fichiers conformes à PDF/A.
- Metadonnées : Des métadonnées, comme le titre, l'auteur, la date de création du document, sa date de modification, le sujet, les mots-clefs, etc. peuvent être enregistrées directement dans le fichier PDF/A. Ainsi, les documents peuvent d'une part être classés automatiquement sans intervention manuelle et, d'autre part, ces dernières peuvent garantir l'authenticité du document.
- Recherche texte avancée : PDF/A peut enregistrer les documents en mode "texte", ce qui permet une recherche avancée sur un mot, des expressions, des chiffres. Cela n'est pas le cas pour des fichiers de type image tels que JPEG ou TIFF (utilisés également pour l'archivage) sans l'aide d'un ROC[13].

4.2 Contraintes sécuritaire

Les documents électroniques définis par le format de fichiers décrits doivent être sécurisés en respectant les contraintes découlant de l'ordonnance fédérale sur la tenue et la conservation des livres de comptes (Olico art. 957) évoquée précédemment.

[12] Extensible Metadata Platform ou XMP est un format de métadonnées basé sur XML. XMP permet d'enregistrer sous forme d'un document XML des informations relatives à un fichier : titre, auteur, historique des modifications, etc. Bien qu'ouvert à tout type de données pouvant intégrer un document XML, XMP prédéfinit la façon de stocker un certain nombre d'informations les plus courantes.

[13] La reconnaissance optique de caractères (ROC) : désigne les procédés informatiques pour la traduction d'images de textes imprimés ou dactylographiés en fichiers de texte.

Ci-après seront mentionnées les contraintes d'intégrité, de traçabilité, de confidentialité, d'accessibilité, d'indentification et de traitement des virus.

4.2.1 Intégrité

Techniquement, l'intégrité se concrétise par la création d'une empreinte électronique du document. L'empreinte est le résultat d'un calcul effectué à l'aide d'un algorithme approprié appelé fonction de hachage[14]. Cette fonction permet de garantir les propriétés suivantes :

- Si l'on change ne serait-ce qu'une virgule dans le document, son empreinte change.
- La probabilité que deux documents aient la même empreinte est quasi nulle.
- Il est impossible de reconstituer le document à partir de sa seule empreinte.

Il est intéressant de noter qu'il y a une différence entre intégrité technique et intégrité au sens juridique. La première, comme vu précédemment, est mise en cause dès l'instant où un seul bit constituant le document est modifié. La seconde consiste à conserver le sens de l'information du document sans s'attacher nécessairement à la forme. Par exemple le fait de changer un accent dans un texte ne va pas changer fondamentalement le sens, alors que cela suffira à lui faire perdre son intégrité technique.

Néanmoins, même si le contrôle d'intégrité ne repose pas uniquement sur l'intégrité technique, il en est largement tributaire car il est évident que ce type de contrôle peut être facilement automatisé, contrairement à un contrôle sur le sens d'un contenu. A ce sujet, il est possible de distinguer trois types de contrôles : contrôle ponctuel, contrôle régulier par sondage et contrôle continu.

- <u>Contrôle ponctuel</u> : ce contrôle est fait uniquement au moment où l'objet est interrogé. Le calcul d'empreinte est effectué pour comparer le résultat de l'empreinte faite au moment où l'objet est entré dans le système d'archivage. Le principal inconvénient réside dans le risque de se trouver dans une situation paradoxale où le système d'archivage sera juste capable d'avertir l'utilisateur que l'objet réclamé a été changé pendant sa conservation, sans pouvoir lui restituer l'objet sans modification.

[14] Voir chapitre « La signature électronique » pour plus de détails

- Contrôle régulier par sondage : afin d'éviter la situation décrite précédemment, ce type de contrôle est opéré de façon systématique, régulière et automatique sur des objets choisis. Le choix est fait de manière aléatoire ou sur la base de catégorie d'objets, par exemple en fonction de leurs criticités.
- Contrôle continu : ce type de contrôle est une extension du précédent, c'est-à-dire que les vérifications sont faites sur l'ensemble des objets et non pas uniquement sur un échantillon.

4.2.2 Traçabilité

La traçabilité doit être présente à tous les niveaux du système d'archivage afin d'enregistrer les « traces » de l'ensemble des opérations effectuées tant par les utilisateurs du système que par ses administrateurs. Elle doit également garantir l'intégrité des objets quant au sens de leur contenu, en conservant la trace et les modalités des différentes migrations éventuellement effectuées.

4.2.3 Confidentialité

Le système d'archivage doit permettre de vérifier les accès aux objets uniquement par les personnes autorisées en assurant un complément à la confidentialité. Cela se traduit par la mise en place d'autorisations, comptes utilisateur, mots de passe et parfois de systèmes de chiffrement.

4.2.4 Accessibilité

L'accessibilité du document est un élément crucial dans tout système d'archivage car elle est déterminante pour retrouver l'information. Cette notion revêt en fait deux composantes : l'organisation des données et les performances techniques.

- L'organisation des données : est composée d'une structure cohérente et efficace permettant de trouver rapidement et facilement les documents. Cette structure repose sur un plan de classement et la constitution d'index.
- Les performances : sont directement liées au type d'infrastructure physique (serveurs, disques, etc.) et réseau.

4.2.5 Identification, authentification

Afin de respecter la confidentialité, il est nécessaire qu'un utilisateur voulant se connecter possède les droits afférents, c'est-à-dire qu'il est bien habilité à accéder à

l'information. Le principe est en général d'identifier la personne grâce à un système d'identifiant (*login*) assorti d'un mot de passe. Malheureusement, rien n'indique de façon certaine qu'il s'agit bien de l'utilisateur autorisé. En effet, même si l'identifiant et le mot de passe sont corrects, il se peut très bien que l'utilisateur ait dérobé ces derniers à son propriétaire.

Afin de pallier cet inconvénient, il est possible de faire recours à un système d' « authentification forte ». Un tel système fait usage de cartes à puces ou d'authentifications biométriques comme les empreintes digitales.

4.2.6 Traitement des virus

Dire que les virus informatiques, pris au sens large, sont en constante évolution et qu'il faut s'en protéger serait une véritable lapalissade. Toutefois, il est important de rappeler que tout document doit être conservé sain. Malgré tout, pour pallier toute éventualité qu'un virus ait tout même pu échapper aux contrôles, il se doit de conserver la définition virale.

4.3 La signature électronique

Certains documents électroniques concernent les contrats. A ce sujet, il est important de relever qu'en droit suisse, la forme des contrats est régie sur le principe de liberté de la forme[15]. Dès lors, le contrat conclu oralement ou encore par voie électronique est tout à fait valable et protégé par la loi, à partir du moment où l'accord nécessaire à sa formation est effectivement donné[16].

Cependant, dans la perspective d'un litige, l'enjeu sera de convaincre le juge. De ce fait, l'exigence d'une certaine forme devient importante pour lever toutes ambigüités du « contrat oral ». En pratique, ces derniers sont sous forme écrite et signés à la main par celui qui s'oblige[17]. Le code des obligations, sous certaines conditions exposées ci-après, prévoit que même la signature électronique est assimilée à la signature manuscrite[18].

---

[15] « La validité des contrats n'est subordonnée à l'observation d'une forme particulière qu'en vertu d'une prescription spéciale de la loi » (art. 11 CO)

[16] « Le contrat est parfait lorsque les parties ont, réciproquement et de manière concordante, manifesté leur volonté. Cette volonté peut être expresse ou tacite » (art. 1 CO)

[17] Art. 14 al. 1 CO

[18] « La signature électronique qualifiée, basée sur un certificat qualifié émanant d'un fournisseur de service de certification reconnu au sens de la loi du 19 décembre 2003 sur la signature électronique est assimilée

Sans entrer dans le détail de la cryptographie, le principe de signature électronique nécessite la délivrance d'un bi-clé constitué d'une clé publique et d'une clé privée. Cette dernière doit absolument rester secrète et à la seule connaissance de son détenteur. A 1'inverse, la clé publique peut être divulguée, en général assortie d'autres renseignements, le tout étant contenu dans un certificat électronique.

4.3.1 Certificat électronique

Il s'agit d'un document sous forme électronique attestant du lien entre les données de vérification de signature électronique telles que les clés publiques et un signataire. Il joue véritablement le rôle de pièce d'identité électronique tel qu'un passeport dans le monde physique . Ce dernier contient par exemple :

- le nom du propriétaire,
- la clé publique,
- la date d'expiration du certificat,
- le numéro d'identification (unique) du certificat,
- le nom de l'autorité de certification qui a délivré le certificat (ou PSCE),
- etc.

4.3.1.1 L'autorité de certification

Comme pour un passeport, le certificat électronique doit être émis et validé par une autorité compétente et autorisée à délivrer ce dernier[19]. Il existe trois grandes classes de certificats qui se différencient au niveau du contrôle de l'information contenue dans le certificat :

- **Certificat de classe I** : ne garantit pas l'identité du titulaire du certificat mais seulement l'existence du courriel de celui-ci.
- **Certificat de classe II** : les informations concernant le titulaire et son entreprise sont contrôlées par l'autorité de certification sur la base de pièces justificatives qui sont transmises en général par voie postale. On parle de contrôle sur pièces.

à la signature manuscrite. Les dispositions légales ou conventionnelles contraires sont réservées » (art. 14 al. 2bis CO). Pour plus de détails voir http://www.bakom.admin.ch/themen/internet/00467/index.html?lang=fr .

[19] http://www.seco.admin.ch/sas/00229/00251/index.html?lang=fr : Dans ce lien, il y un répertoire dans lequel figurent les entreprises suisses qui délivrent et gèrent des certificats électroniques qualifiés et qui sont reconnues.

- **Certificat de classe III** : par rapport à la classe II, un contrôle supplémentaire de l'identité du titulaire est effectué physiquement par un agent de l'autorité de certification. On parle de contrôle de face à face.

4.3.2 Principe de fonctionnement du certificat électronique.

Le principe de fonctionnement du certificat électronique se base principalement sur deux éléments : la signature du document et la vérification de la signature.

4.3.2.1 Signature du document.

La signature du document commence par l'empreinte numérique de ce dernier et le chiffrement de cette empreinte au moyen de la clé privé. Le résultat de ce chiffrement correspond à la signature numérique du document. Ensuite, le document est transmis au(x) destinataire(s) accompagné de sa signature et du certificat électronique correspondant.

4.3.2.2 Vérification de signature

La clé publique transmise dans le certificat permet de déchiffrer la signature numérique et donc de retrouver l'empreinte originale du document. Cette dernière est comparée à une empreinte que l'on prend du document reçu. Si les deux sont identiques, le document transmis est authentique et émane bien du possesseur de la clé publique, comme défini dans le certificat électronique. En effet, seule la clé publique contenue dans le certificat est capable de déchiffrer la signature obtenue avec la clé privée correspondante du bi-clé, attestant du même coup de son origine.

Reste cependant à vérifier la validité dudit certificat auprès de l'autorité compétente.

4.4 Cycle de vie de l'information archivée

Le temps de conservation légal des documents commerciaux, comme vu précédemment, est fixé à dix ans. Implicitement, cette obligation impose un cycle de vie de l'information archivé. Ce cycle est linéaire qui suit les phases de vie de l'information [20]: capture dans le système, maintenance pendant la durée requise et destruction ou transfert.

[20] Voir annexe : ILM ou gestion du cycle de vie de l'information

4.4.1 Identification et capture dans le système

Toute information archivable doit être identifiée et capturée dans le système d'archivage selon un processus. Ce dernier peut être par exemple :

1. Enregistrement dans le système d'archivage avec un identifiant unique.
2. Rattachement au référentiel de conservation et au plan de classement.
3. Vérification de la qualité du format et du support avec refus du fichier ou déclenchement d'une procédure de conversion.
4. Capture des métadonnées existantes relatives aux données et au contexte de production : date, nom des documents, auteurs, destinataires, etc.
5. Vérification de l'unicité des données, c'est-à-dire vérification qu'elles ne sont pas déjà été archivées.
6. Ajout, si besoin, de métadonnées complémentaires.
7. Allocation d'une adresse de stockage.

4.4.2 Maintenance

La maintenance, c'est-à-dire la conservation des informations dans le système d'archivage pendant toute la durée requise, comporte trois volets : la maintenance des formats, la maintenance des métadonnées et la maintenance des droits d'accès.

4.4.2.1 Maintenance des formats

La pérennisation des données suppose une maîtrise des formats et des supports avec, en cas d'obsolescence, des migrations systématiques. Pour ce faire, des contrôles réguliers de la qualité des données et de leur lisibilité doivent être entrepris tout au long de la conservation. En cas de déficience, des actions correctives doivent être appliquées sans omettre de tracer toute modification pour respecter les exigences légales.

4.4.2.2 Maintenance des métadonnées

Au fil du temps, les métadonnées relatives aux plans de classement ainsi qu'au référentiel peuvent changer. Il est donc nécessaire de les modifier et de les faire évoluer pour garantir la recherche, et par conséquence l'accès aisé, aux utilisateurs. Comme pour la maintenance des formats, les changements des métadonnées doivent être tracés.

4.4.2.3 Maintenance des droits d'accès

Enfin, la maintenance du système inclut une gestion rigoureuse des habilitations ou droits d'accès, afin que les départs et changements de poste des collaborateurs soient mis à jour dans le système d'archivage. On évite ainsi que des fichiers archivés soient inaccessibles du fait que la ou les personnes autorisées ont quitté l'entreprise alors que les données doivent toujours être conservées.

4.4.3 Destruction

Les documents et données périmés doivent être détruits à échéance de leur durée de conservation. Si ces derniers présentent un intérêt pour la mémoire historique, ils peuvent soit changer de statut tout en restant dans le système d'archivage, soit être transférés vers un centre d'archives historiques.

4.4.3.1 Mise en application

Toute action de destruction devrait nécessiter l'intervention de trois acteurs : le responsable de l'archivage qui établit la liste des données à détruire, le responsable des contenus qui valide la destruction en vérifiant qu'il n'y a pas d'affaire en cours nécessitant un prolongement de la conservation, et le responsable du stockage qui effectue la destruction.

4.5 Support physique.

Aujourd'hui, un grand nombre de supports physiques sont disponibles pour l'archivage. Cependant, il est généralement admis qu'il y a deux familles : les supports magnétiques (bande et disque magnétiques) et les supports optiques (CD, DVD et disques magnéto-optique).

La grande différence des supports optiques par rapport aux à ceux magnétiques réside dans leur fiabilité, et surtout dans leur pérennité. Elle est bien plus élevée et n'est pas soumise aux phénomènes physiques naturels dus, par exemple, au caractère magnétique de la bande pouvant à long terme provoquer un phénomène de « collage». Même si certains constructeurs de disques optiques n'hésitent pas à annoncer des durées de garantie très longues pour leurs supports, il faut modérer une telle information. En effet, au bout de nombreuses années, il y a de fortes chances que les lecteurs n'existent plus dans ce format et qu'en conséquence, on se trouve avec un disque illisible. Par ailleurs, tant pour les CD que pour les DVD, l'information est stockée d'une façon permanente par modification du substrat, à l'origine de la notion de WORM (*Write Once Read Many*). C'est pourquoi, les supports optiques avaient encore

récemment une large tendance à être privilégiés pour régler les problèmes d'archivage. Néanmoins, l'inconvénient majeur vient de son prix de revient, nettement plus élevé que celui de la bande magnétique. La solution qui aujourd'hui semble se répandre de plus en plus est liée à de nouvelles technologies basées sur le disque magnétique.

4.5.1 Le WORM

Avant d'expliciter les différentes technologies basées sur le disque magnétique, il est important de développer la notion de WORM (*Write Once Read Many*). Cette dernière, comme mentionné précédemment, était originalement attribuée essentiellement aux disques optiques. La norme ISO 18509 élargit la notion de WORM à une méthode d'enregistrement dont la propriété intrinsèque est d'être non effaçable, non réinscriptible et non modifiable. Elle distingue trois types de WORM :

- Type A: transformation permanente du support, principe des disques optiques avec modification du substrat ;
- Type B: utilisation d'un microcode WORM inclue dans le support au moment de sa fabrication, reconnu par le lecteur ou le contrôleur et protégé de l'effacement et de la réécriture dans des conditions normales d'utilisation, principe des disques magnéto-optiques ou des bandes équivalentes ;
- Type C: génération d'un micro-code enregistré avec l'information et destiné à traiter cet enregistrement comme un enregistrement de type WORM par le logiciel de gestion du support, le protégeant du même coup de l'effacement et de la réécriture dans des conditions normales d'utilisation, principe des disques magnétiques. Dans certains cas, la protection de type WORM peut être limitée à une durée de conservation associée aux données à protéger.

Le type C permet ainsi aux technologies à base essentiellement de disques magnétiques de pouvoir se positionner naturellement pour résoudre des problématiques d'archivage.

4.5.1.1 Le disque dur magnétique.

Pourquoi cet acharnement de la part des constructeurs à utiliser le disque dur magnétique ? Il y a au moins quatre raisons : la rapidité d'accès et de transfert des données, la fiabilité du support, la migration technologique et la possibilité de supprimer un fichier.

- Rapidité d'accès et de transfert des données : les supports optiques sont en effet généralement peu performants. Toutefois, concernant l'archivage, c'est-à-dire des documents qui ne seront probablement consultés que de manière exceptionnelle, le critère de rapidité n'est pas forcément prédominant.
- Fiabilité du support : la durée de vie des supports optiques étant relativement courte (5 à 20 ans), il faut prévoir un cycle de duplication régulière, permettant d'assurer la pérennité voulue. Les disques magnétiques intégrés à des baies de type RAID[21], peuvent être remplacés à chaud, les données se trouvant dupliquées et automatiquement régénérées.
- Migration technologique : les normes et technologies inhérentes aux disques, lecteurs, armoires... évoluent rapidement. En cas de migration, il est plus simple de vider le contenu d'une armoire ancienne dans une nouvelle, plutôt que relire tous les supports optiques ou toutes les bandes magnétiques.
- Possibilité de supprimer un fichier : cela peut paraître paradoxal de vouloir supprimer une archive, mais la loi impose désormais que les données ne soient pas conservées au-delà d'une durée normale d'utilisation. Or, sur un support optique, les possibilités de suppression sont pour le moins limitées.

4.5.2 Le Worm Logique

Il existe en fait trois approches différentes destinées à assurer cette conservation sécurisée : le worm logiciel, les CAS et le stockage en grille.

4.5.2.1 Worm logiciel

La première de ces approches, sans doute la plus « simple », consiste à utiliser des baies de stockage traditionnelles sur disque et à y ajouter une couche logiciel en amont destinée à gérer cette notion de WORM logique et ainsi à bloquer toute tentative de réécriture, de modification ou de suppression.

En résumé, le principe repose sur une horloge logicielle indépendante de la baie qui écrit sa mesure systématiquement sur les disques WORM toutes les 10 secondes. Toutes les fonctions liées à la manipulation des fichiers worm sont validées par l'horloge. Si jamais l'on

[21] RAID (*Redundant Array of Independent Disks*) : est une technologie permettant de stocker des données sur de multiples disques durs afin d'améliorer, en fonction du type de RAID choisi, la tolérance aux pannes et/ou les performances de l'ensemble. (voir http://fr.wikipedia.org/wiki/RAID_(informatique))

remplace le contrôleur par un contrôleur d'une nouvelle armoire avec une nouvelle horloge et avec une autre mesure de temps, le logiciel s'en aperçoit et se recale. Les disques sont de type RAID. Lors des manipulations, le contrôleur effectue un calcul de contrôle et, en cas d'erreur, le segment est réécrit. Avec le RAID, les données sont cryptées et réparties, ce qui rend impossible la lecture ou la modification directe.

L'avantage de ce dispositif réside dans sa simplicité :

- n'importe quel volume peut devenir un WORM,
- l'application n'a pas besoin d'une API[22] pour accéder aux archives, il suffit d'y accéder comme un NAS[23].

4.5.2.2 Le CAS (Content Addressed Storage)

Une autre approche, est l'adressage de fichiers fixes appelé CAS (*Content Addressed Storage*). Ce dernier crée une empreinte numérique pour chaque objet entrant. Cette empreinte devient ensuite le seul moyen pour consulter l'objet ainsi stock

Il s'agit en fait d'une logique apparentée à celle des consignes. En effet, vous déposez un objet et en retour vous obtenez un ticket (adresse de contenu). Ultérieurement, seul ce ticket vous permettra de retrouver l'objet déposé. La façon dont est conservé ce dernier est totalement prise en charge par le système.

Par exemple, habituellement, pour stocker le fichier "resultats.xls", le système va inscrire les données sur le disque et enregistrer sur une table d'allocation, le nom du fichier, l'adresse physique correspondante (ou les adresses si le fichier est réparti sur plusieurs segments), ainsi que quelques données supplémentaires comme la taille ou la date d'écriture. Ce système comporte deux inconvénients dans une utilisation d'archivage :

- Il peut arriver qu'un fichier strictement identique puisse être enregistré plusieurs fois ou dans un autre répertoire.
- Il est impossible de distinguer deux exemplaires d'un fichier avec le même nom, la même taille, mais avec un chiffre modifié, par exemple.

---

[22] API (Interface pour la programmation d'applications) : Ensemble de bibliothèques permettant une programmation plus aisée car les fonctions deviennent indépendantes du matériel. (voir http://www.futura-sciences.com/fr/definition/t/informatique-3/d/api_465/)

[23] NAS (*Network Attached Storage*) : est une unité autonome, reliée à un réseau dont la principale fonction est le stockage de données en un gros volume centralisé pour des clients-réseau hétérogènes. (voir http://fr.wikipedia.org/wiki/Stockage_en_r%C3%A9seau_NAS)

Dans le cas de l'archivage, il est important de ne pas multiplier les versions identiques d'un même fichier. Au contraire, il est intéressant de distinguer une version mise à jour. Pour pallier ce problème, toute modification du contenu déclenche automatiquement la création d'une nouvelle adresse de contenu. A l'inverse, si un même objet se présente, l'adresse de contenu étant déjà existante, il n'y aura pas de doublement en matière de stockage. Les accès aux ressources sont rapides. La gestion et l'administration d'une telle solution sont considérablement simplifiées, même pour de gros volumes. Par ailleurs, le système est en mesure de garantir l'intégrité du contenu à long terme et la confidentialité des informations peut être obtenue par un chiffrement systématique des données ainsi conservées.

4.5.2.3 Le stockage en grille

La troisième approche répond à une logique d'organisation en cellules, faisant appel au concept de stockage en grilles ou « grid ». De telles solutions intégrées d'archivage sont plutôt indépendantes des applications, mais elles utilisent néanmoins une méthode d'accès aux enregistrements normalisée.

Cette approche est constituée de plusieurs cellules interconnectées via un réseau Ethernet[24]. Chacune d'entre elle participe à la solution et une demande d'archivage est ainsi répartie sur l'ensemble des cellules. Une cellule possède un processeur, un espace de stockage pour l'indexation des « contenus » et les « méta données » ainsi qu'un espace de stockage pour les données des contenus.

L'un des principaux avantages de cette approche est de permettre la réalisation facile de systèmes de stockage complexes à partir d'éléments standards. Une telle technologie permet en effet la réalisation de solutions très performantes, indépendantes du nombre d'enregistrements gérés, évitant du même coup tous les phénomènes liés aux baisses de performances que l'on peut observer au cours des montées en charge. Un autre avantage et non des moindres réside dans sa capacité à effectuer des migrations très progressives, cellule par cellule. Enfin, l'exploitation de tels systèmes est simplifiée et surtout allégée à l'extrême. Ainsi, même si en apparence le coût d'acquisition peut sembler plus onéreux qu'un système traditionnel, une comparaison plus complète tenant compte des coûts d'exploitation sur trois ans ne laisse plus aucun doute. L'intérêt est encore plus évident si l'on prend en considération la notion de migration, évoquée précédemment.

[24] Ethernet est un protocole de réseau local à commutation de paquets.

4.5.2.4 Principaux acteurs du marché

Ce sous-chapitre vise à donner quelques exemples de grands acteurs du marché utilisant les approches décrites précédemment :

- **IBM** : Ce constructeur bien connu, disposant toujours de juke-boxes de disques optiques à son catalogue pour l'environnement iServer (AS/400), propose également depuis quelques années une solution d'archivage WORM à base de disques magnétiques, le Totalstorage DR (pour Data Retention). Ce système repose sur le module SSAM (System Storage Archive Manager), une extension du logiciel TSM (Tivoli Storage Manager), connu pour la gestion du stockage. Outre la garantie de protection WORM des données sur disque et leur cryptage, le Totalstorage DR offre des fonctions sophistiquées et auditables de conservation des informations répondant aux exigences multiples et évolutives des législations et des outils de gestion de contenu. Pour les conservations de longue durée, et dans le but d'assurer un véritable ILM (*Information Lifecycle Management*)[25], le Totalstorage DR prend à son compte les évolutions technologiques et assure, au moment opportun, la migration des données vers une unité de conservation WORM plus récente.
- **Network Appliance** : De son côté, NetApp à développé un solution sogiciel défend nommée SnapLock permettant de faire du Worm logique.
- **EMC** : propose au travers des systèmes Centera une solution avec une technologie basée sur l'approche du CAS (Content Aeeess System) vue précédemment.
- **Hitachi**: En tant que fournisseur de baies de stockage, Hitachi a développé un gestionnaire d'archivage avec son logiciel Data Retention Utility. Ce logiciel utilitaire offre ainsi la fonction Worm sur les systèmes de stockage sur disque Hitachi dans les environnements de type systèmes ouverts ou de type mainframe.

  Hitachi offre également une véritable solution dédiée à l'archivage: Archive Content Platform. Cette solution est basé sur l'approche de stockage à grille.
- **Hewlett Packard (HP)** : HP propose une solution intégrée d'archivage RISS (*Reference Information Storage System*) indépendante des applications avec

[25] Voir annexe ILM ou la gestion du cycle de vie de l'information (voir VI1).

une méthode d'accès aux informations normalisées. Cette solution utilise une approche du stockage en grille.

- **SUN/Storagetek**: Fournisseur de solutions d'archivage, ce constructeur a une approche de la gestion du stockage sous la forme de « classes de service » indépendante des technologies utilisées. Cette approche permet la mise en œuvre d'une véritable politique ILM (*Information Lifecycle Management*) intégrée au processus métier et garantit une gestion très efficace des fonctions d'intégrité et de traçabilité indispensables dans des contextes d'archivage à valeur probante.

4.6 Système d'archivage électronique (SAE)

En conclusion de ce chapitre consacré aux concepts de l'archivage électronique, je vais brièvement exposer le modèle de référence OAIS[26] (Open Archival Information System) concernant le système d'archivage électronique (nommé par la suite SAE), puis terminer par l'architecture technique du SAE.

4.6.1 Fonctionnalité du modèle OAIS

Sans entrer dans les détails du modèle OAIS, il est intéressant de relever qu'il s'agit d'un modèle abstrait. Il définit une terminologie et des concepts. Il identifie les acteurs, décrit les fonctions et les flux d'informations, et propose un modèle d'information adapté à la problématique de l'archivage numérique sur le long terme[27]. Cependant ce modèle n'est pas un recueil de spécifications techniques destinées à être directement implémentées.

Parmi tous les concepts exprimés par OAIS, il y a cinq fonctions relevantes (voir la Figure 4): la fonction de versement, la fonction de stockage, la fonction de gestion des

[26] OAIS (*Open Archival Information System*) : est le résultat d'un travail du CCSDS (*Consultative Committee for Space Data Systems*), organisme international de normalisation des agences spatiales, réalisé à la demande d'ISO. Ce travail, auquel ont été associés des représentants de bibliothèques et d'archives institutionnelles, a abouti en 2002 à la rédaction d'un document qui spécifie de manière très générale l'architecture logique et les fonctionnalités d'un système d'archivage. OAIS ne préjuge pas de la nature des objets à archiver, mais le modèle est particulièrement bien adapté aux objets numériques. Le modèle de référence OAIS est aujourd'hui une norme internationale ISO (ISO 14721).

[27] Le long terme est défini comme suffisamment long pour être soumis à l'impact des évolutions technologique, c'est-à-dire, en ce qui concerne le document numérique, entre 3 à 10 ans (même si cela est très tributaire du format).

données descriptive, la fonction de consultation et de communication et la fonction d'administration.

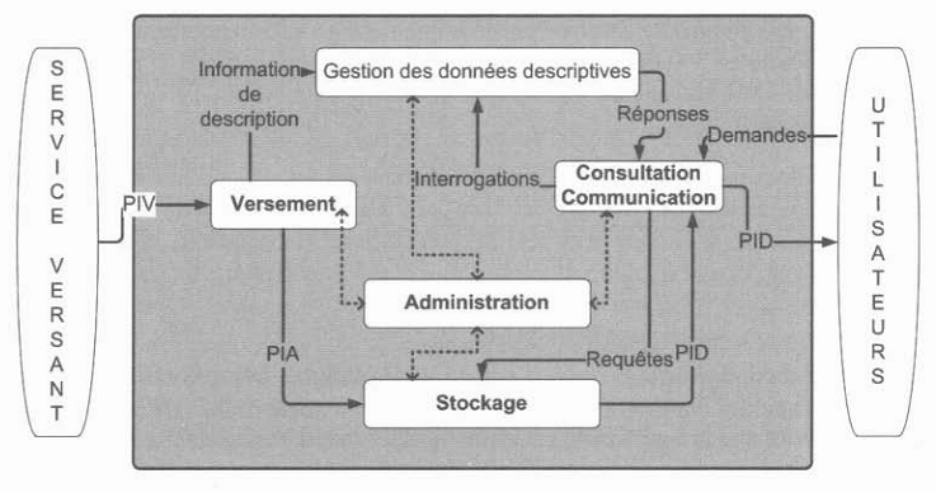


Figure 4 : fonctions du système d'archivage électronique

4.6.1.1 Fonction versement:

La fonction de versement permet le traitement des paquets d'informations en provenance des services versants dans leur ensemble. Cette fonction inclut ainsi tous les mécanismes de préparation, transmission, contrôle, rejet, complément d'information ainsi que tous les traitements de ces informations pour une intégration dans le dispositif de stockage des contenus et celui de gestion des données descriptives.

4.6.1.2 Fonction stockage

La fonction stockage gère l'ensemble des services liés à la conservation des paquets d'informations archivés à partir du moment où ils sont mis à sa disposition par la fonction de versement jusqu'à leur destruction/élimination s'il y a lieu, tout en garantissant leur intégrité. Cette fonction prend entre autres en compte les aspects de choix de supports et de gestion de l'ensemble des migrations.

4.6.1.3 Fonction gestion des données descriptives

La fonction gestion des données descriptives assure la conservation, la mise à disposition et la mise à jour des informations descriptives associées aux contenus d'informations, conservés par la fonction stockage. Ces informations doivent servir aux utilisateurs comme point d'entrée au SAE et permettre de retrouver les données qu'ils recherchent en assurant le lien avec leur identification de localisation dans le système de stockage.

4.6.1.4 Fonction de consultation et de communication

La fonction de consultation et de communication prévoit l'ensemble des mécanismes permettant d'accéder, de consulter et de livrer les informations disponibles dans le SAE, qu'il s'agisse des données descriptives ou du contenu lui-même. Elle comprend la mise à disposition d'une interface de consultation, d'un système de recherche effectuée à partir des données descriptives, d'un principe de visualisation du résultat, la sélection de contenus à communiquer et la livraison effective de ces contenus sous forme de paquets d'informations diffusés.

4.6.1.5 Fonction d'administration

La fonction d'administration permet d'assurer l'exploitation de l'ensemble du système d'archivage électronique et sa pérennisation, ainsi que la gestion des utilisateurs du SAE au sens de leurs droits d'accès. En fonction de ses propres besoins, l'administration pourra également élaborer différentes statistiques sur les fonctions de versement, la consultation, ou encore sur celles directement liées à l'exploitation. Ces statistiques permettront par exemple d'anticiper toute augmentation des capacités de stockage ou de planifier les migrations de supports et de format.

4.6.2 Architecture technique du SAE

L'architecture technique représentée par la Figure 5 doit être considéré comme un exemple de « cas d'école ».

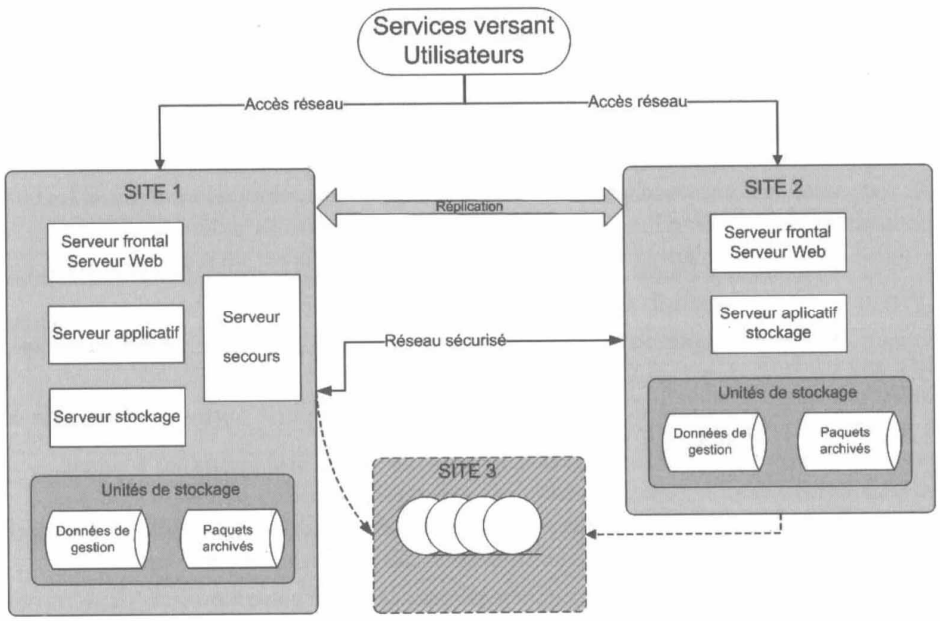


Figure 5 : Exemple d'architecture technique du SAE

Cette architecture se compose de trois sites idéalement éloignés géographiquement, pour des raisons de sécurité. En fait, grâce à cette distance, les paquets d'informations sont stockés au moins trois fois.

4.6.3 Site 1

Le premier site présente le détail des différents composants nécessaires à la bonne marche du SAE d'un point de vue essentiellement fonctionnel. Dans le présent « cas d'école », chaque composante est séparée par un serveur physique différent. L'intérêt d'avoir plusieurs serveurs est double :

- Premièrement, l'ensemble sera plus performant car d'une part chaque serveur possède son propre processeur et, d'autre part, tous les serveurs peuvent être dimensionnés et optimisés en fonction du service qu'ils doivent fournir.
- Deuxièmement, en cas de dysfonctionnement, seuls les services du serveur concerné seront hors service.

Cependant, dans la pratique, il est tout à fait envisageable de regrouper complètement ou partiellement l'ensemble de ces derniers.

4.6.3.1 Gestion du stockage

La gestion du stockage a deux fonctionnalités distinctes: le serveur de stockage et l'unité de stockage.

- **Le serveur de stockage** : gère les données de gestion et il doit évoluer et être mis à jour au cours du temps, en fonction des différentes modifications qui peuvent intervenir sur les données correspondantes.
- **L'unité de stockage** : conserve le contenu de l'information archivée. De ce point de vue, cette fonctionnalité est plus statique et l'évolution se résume à un changement éventuel du support, voire du format au cours du temps.

4.6.3.2 Serveur frontal ou serveur Web

Le serveur frontal est une interface des communications entre les applications ainsi qu'un point d'entrée uniformisé pour les services du SAE.

Par exemple, dans les architectures web, on peut utiliser un serveur frontal HTTP (comme Apache) pour traiter les requêtes générales et renvoyer certaines demandes de service vers un serveur d'application (comme Tomcat).

4.6.3.3 Serveur applicatif

Le serveur applicatif est dédié à fournir tous les services applicatifs comme le classement, les fonctions de recherche, etc. du SAE.

4.6.3.4 Serveur de secours

Le serveur de secours ne constitue pas une obligation, mais simplement un moyen de gagner en efficacité dans le cas d'un dysfonctionnement de l'un des serveurs. En effet, si l'on dispose d'un serveur préparamétré, le temps d'indisponibilité des services sera évidemment moins important.

4.6.4 Site 2

Le deuxième site permet d'une part de répartir la charge, et, d'autre part, de faire fonction de site de secours.

Contrairement au premier site, toutes les fonctions de stockage peuvent être regroupées dans un même ensemble, ce qui représente un mode de fonctionnement minimaliste dans la mesure où il suppose que l'on travaille sur un seul et même type de support.

4.6.5 Site 3

Le troisième site garantit une troisième copie des données. Il est mis à jour de manière asynchrone de manière à pouvoir corriger des paquets de donnée en erreur lors d'un dysfonctionnement des deux autres sites.

III Principe de l'archivage SAP

Cette partie vise à exposer les outils fournis par SAP pour la réalisation du projet d'archivage des données SAP. Dans un premier temps, elle va donner une définition, expliciter les motivations et le principe d'un tel archivage. Elle sera ensuite suivie par la présentation de l'objet d'archivage, le processus associé et l'ArchiveLink. Pour terminer, un comparatif entre les directives du code des obligations en terme de conservation et les outils seront présentés.

1 Définition

L'archivage consiste à déplacer des données d'un système SAP vers un serveur de fichiers ou système de gestion de contenu, et à en assurer la relecture totale ou partielle avec SAP (voir Figure 6 : définition).

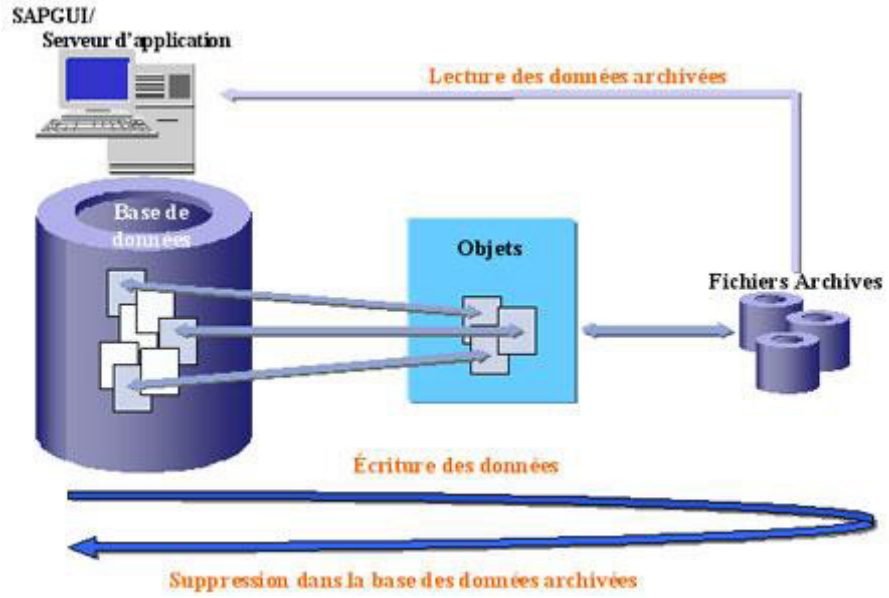


Figure 6 : définition

2 Motivations

Les motivations qui justifient un archivage des données sont multiples :

- **Temps de réponses du système** : au fil du temps, la taille de la base de données augmente et, en particulier, toutes les tables représentant le coeur de l'activité de l'entreprise. Cela implique une perte progressive et souvent exponentielle des temps de traitement et de réponse du système, causant du désagrément aux utilisateurs.

- **Technique** : l'archivage permet de réduire le temps de sauvegarde, de restauration et de la durée des traitements, ainsi que la difficulté voire impossibilité à réorganiser une base de données d'une taille importante.
- **Economique** : cela diminue les coûts de stockage car les données archivées peuvent être stockées sur un support moins onéreux.

3 Principe

Le principe de l'archivage se base sur trois étapes principales : extraction, écriture et suppression.

- Le but de l'extraction est de lire les tables qui modélisent un objet dans la base de données normalisée[28], pour reconstituer ce dernier. Par exemple, la facture (voir Figure 7) est modélisée par les tables telles que VBAK et VBAP.
- Les données extraites lors de l'étape précédente sont écrites dans des fichiers archives. Ils sont ensuite entreposés dans un système de fichiers ou un gestionnaire de contenu.
- Enfin, les données extraites sont supprimées des tables.

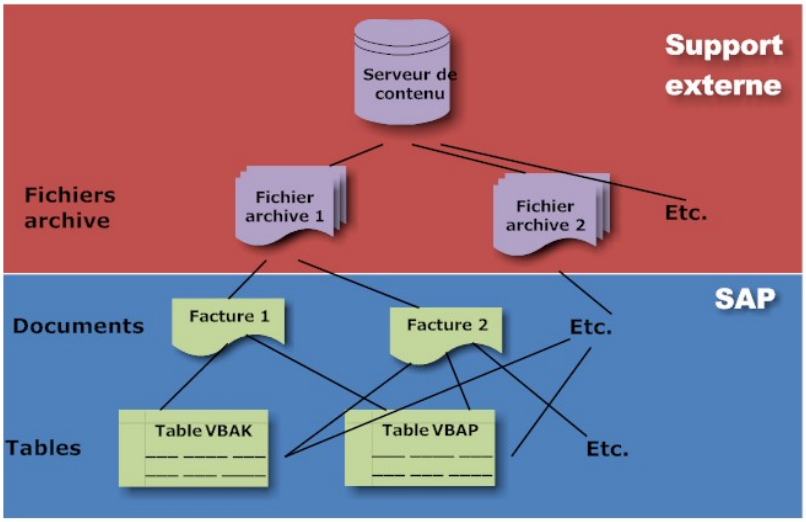


Figure 7 : Principe d'archivage

La consultation des fichiers archives se fait généralement par les mêmes écrans de transaction que pour les données dans la base. Le but est de faciliter et de rendre la consultation la plus aisée possible. Par contre, le temps d'affichage des fichiers archives est

[28] La normalisation vise à éliminer les inconvénients dus à la présence de redondances dans les schémas relationnels de données.

généralement plus long, aucune donnée n'est modifiable et certains champs peuvent être non renseignés[29].

4 Objet d'archivage

Le chef d'orchestre d'archivage des données est l'objet d'archivage[30]. Ce dernier est une sorte de couteau suisse fournissant tous les outils nécessaires pour la création de fichiers archive, de tout ou d'une partie d'un document, de façon cohérente et dans un état stable.

Par exemple (voir Figure 8) une facture de décompte (objet métier) est composée d'une en-tête et des postes. Ces deux parties sont modélisées par plusieurs tables sur lesquelles les objets d'archivage ISU_PRDOCH et ISU_PRDOCL se basent[31].

[29] Les causes de non affichage sont en général dues à une différence de version, c'est-à-dire que si un nouveau champ est ajouté après que l'archivage ait été fait, il ne pourra en aucun cas être complété dans les archives.

[30] Attention : l'objet d'archive n'a pas les mêmes caractéristiques qu'un objet en programmation OO (Orienté Objet), il ne s'implémente pas, ne se dérive pas, etc.

[31] Note : la représentation de la facture par deux objets archives évite la redondance dans les fichiers et permet donc une certaine forme de normalisation. En plus, vu que la lecture des fichiers archives est assez lente, il pourrait être envisageable d'archiver uniquement les postes (en général c'est le plus volumineux) tout en gardant les en-têtes dans la base de données, ce qui améliorerait les temps de consultation.

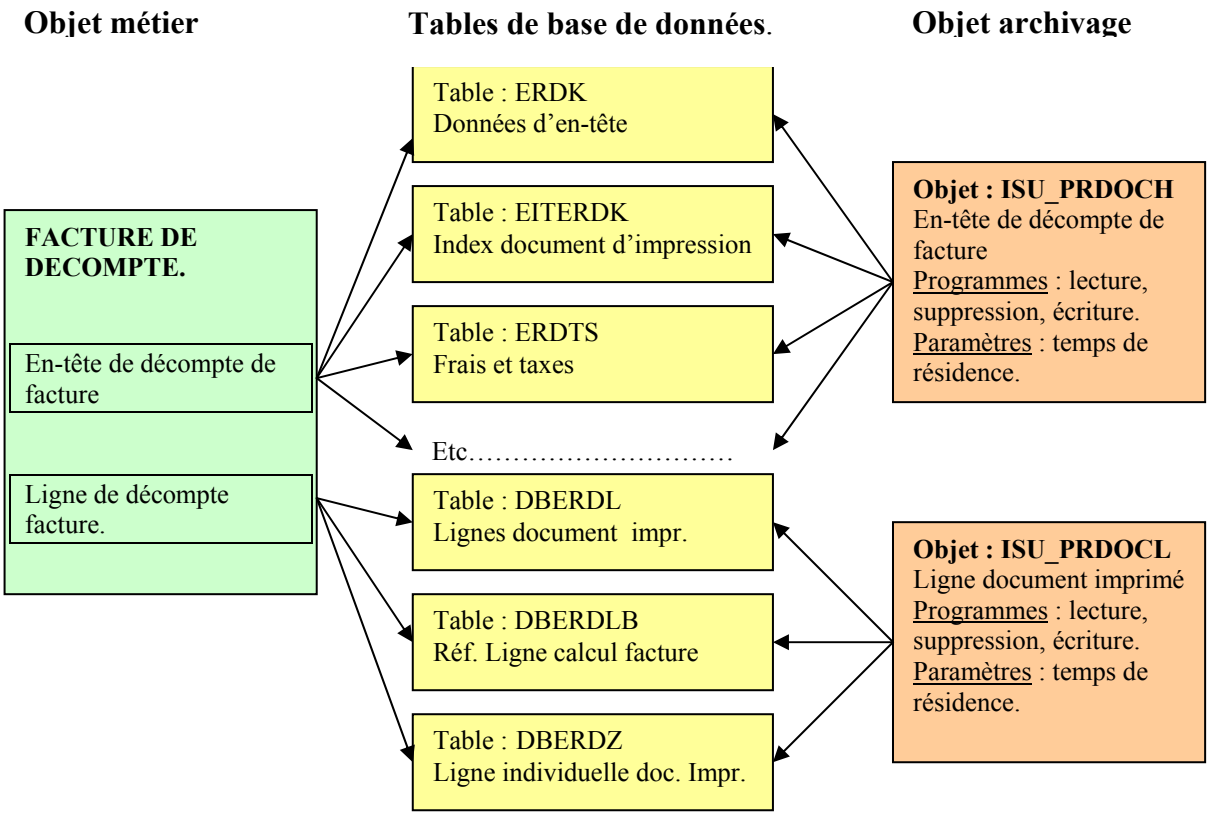


Figure 8 : Exemple facture de décompte

La suite de ce chapitre vise à décrire toutes le « lames des ce couteau » permettant de mener à bien l'archivage des données.

4.1 Programmes associés au processus archive

Le processus d'archivage est une suite d'étapes matérialisées par des programmes et des paramètres associés (voir Figure 9).

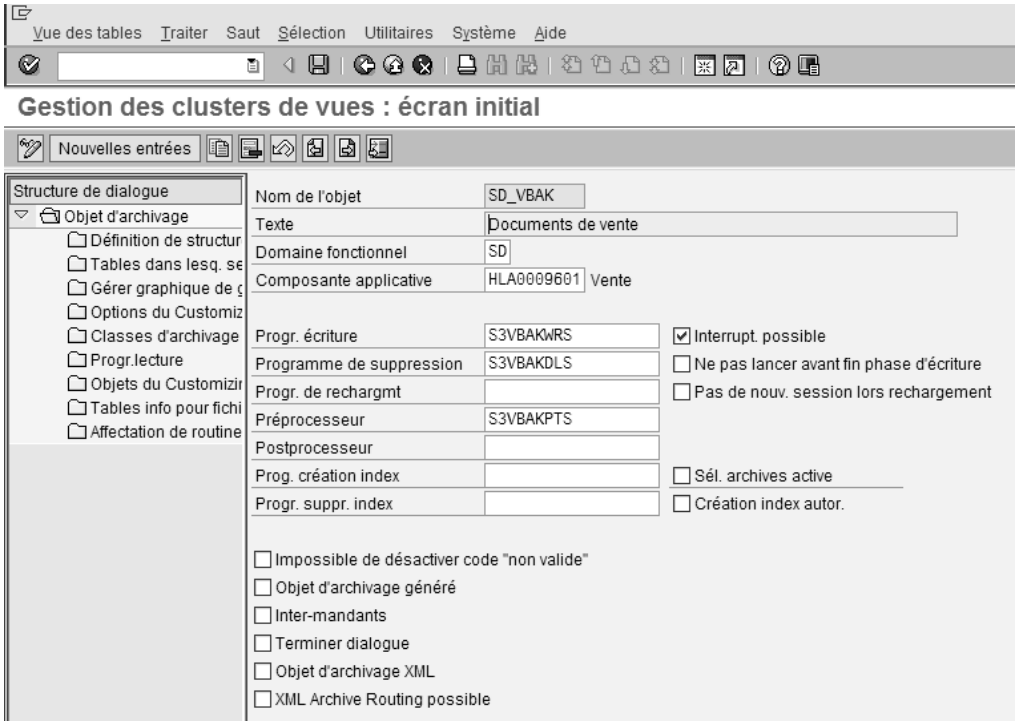


Figure 9 : Copie d'écran d'un objet d'archivage (transaction AOBJ)

- **Programme d'écriture** lit les tables et crée des fichiers d'archive.
- **Programme de suppression** permet de supprimer les données de la base de données, mais, dans certains cas, seul un témoin de suppression[32] est placé dans les données.
- **Programme de rechargement** comme son nom l'indique, il permet de recharger les données à partir d'un fichier d'archive dans la base de données. Cependant, l'utilisation est réservée quasiment uniquement dans le cas où des données ont été archivées par mégarde.
- **Préprocesseur :** certains objets exigent un programme de prétraitement qui prépare les données pour l'archivage. Ce programme marque les données à archiver, mais il ne supprime aucune information de la base de données.
- **Postprocesseur** est effectué après que le programme de suppression ait eu lieu. En général, il est utilisé pour supprimer de la base de données les données marquées d'un témoin par le programme de suppression.

[32] Le témoin de suppression est un statut indiquant que l'objet n'est plus fonctionnellement utilisé mais résident dans la base de données. C'est le cas des articles qui ne sont plus fabriqués par le fournisseur, et sont donc supprimés, mais qui ont déjà été facturés. L'ERP dans ce cas empêche la suppression physique pour garantir la cohérence de ces pièces.

- **Programme de création d'index** permet d'enregistrer des index dans des tables de la base de données SAP pour indexer les fichiers archives. Par cet astuce, l'accès aux archive est plus rapide et aisé.
- **Programme de suppression des index** permet de supprimer les index créés par le programme de création d'index.

4.1.1.1 Paramètres associés aux programmes

Il y a un certain nombre de paramètres liés aux programmes précédents :

- **Lancer les jobs de suppression à la fin de la phase d'écriture :** cela impose que tous les jobs de suppression ne soient lancés qu'après la génération de tous les fichiers du processus d'archivage.
- **Création de fichiers lors du rechargement** : lors du rechargement de fichiers archives, de nouveaux fichiers archives sont créés si les objets de données ne sont pas tous rechargés dans la base de données.
- **Témoin d'archivage inter-mandants** indique que les données à archiver doivent être archivées pour tous les mandants.[33]
- **Objet d'archivage XML** : quelques rares programmes d'archivages utilisent un connecteur au format XML au lieu de celui standard SAP (l'Archive Development Kit : ADK).

4.2 Structure d'un objet d'archivage

Un objet d'archivage, comme déjà mentionné, doit garantir la cohérence des données archivées. Dans ce dessein, il possède une structure issue de la définition des tables utilisées par l'objet métier représenté par la Figure 10.

Définition table 1	Définition table 2	Définition table 3	Etc.

Figure 10 : structure objet archivage

Les programmes d'archivage utilisent cette structure pour écrire et lire les fichiers archives.

[33] Le mandant est une unité indépendante d'un point de vue commercial, organisationnel et technique dans un système SAP. Les mandants possèdent leurs propres fiches, ensembles de tables, droits d'accès et un certain nombre de paramétrage spécifique.

4.2.1 Classes d'archivage

Pour rendre plus facile la définition de la structure, il est possible de lui associer des classes d'archivage. Ces dernières sont une définition de structure qui peut être réutilisée par plusieurs objets d'archivage.

4.3 Tuples de tables non archivés

Certains objets d'archivage peuvent supprimer des données d'une liste de table définie sans archiver pour autant ces données (voir Figure 11).

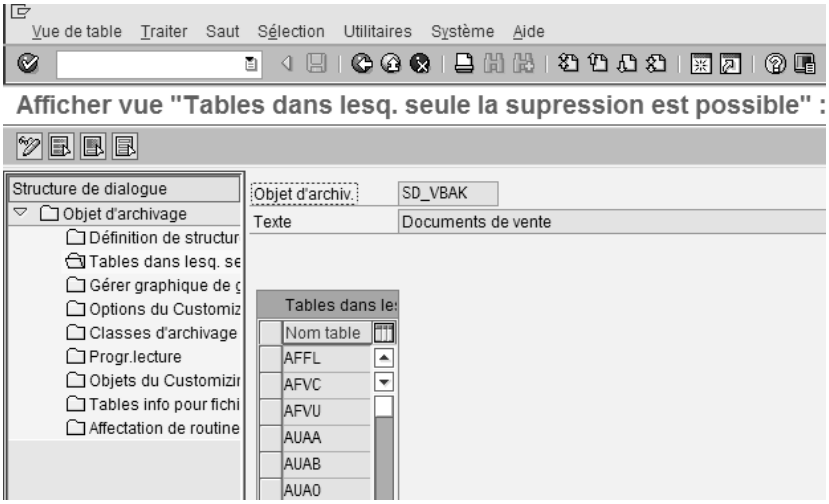


Figure 11 : Liste des tables supprimées uniquement

Ces tables contiennent généralement des données qui peuvent être reconstruites à tout moment à l'aide des données du système comme par exemple des tables d'index. En clair, ce sont généralement des données qui ne doivent pas nécessairement être conservées.

4.4 Dépendance entre objets archive

Dans certains cas, un objet d'archivage ne peut pas à lui seul garantir la consistance de la base de données mentionnée précédemment. SAP a prévu une dépendance entre objets, c'est-à-dire des contrôles garantissant une cohérence.

Par exemple pour archiver un document de vente (voir Figure 12) il faut qu' au préalable les documents d'achats, les factures, les demandes d'achats, etc. soient archivés.

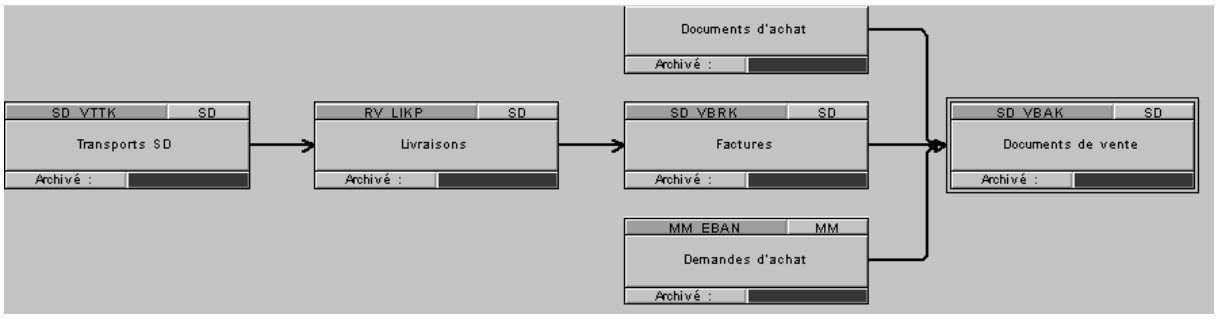


Figure 12 : exemple de dépendance entre objets

4.5 Programme de lectures des fichiers archives

A chaque objet d'archivage, il est possible de lui associer plusieurs programmes (voir Figure 13) conçus pour la lecture de fichiers archives. Ces programmes sont utilisés par les transactions de visualisation des données ou pour des fins statistiques.

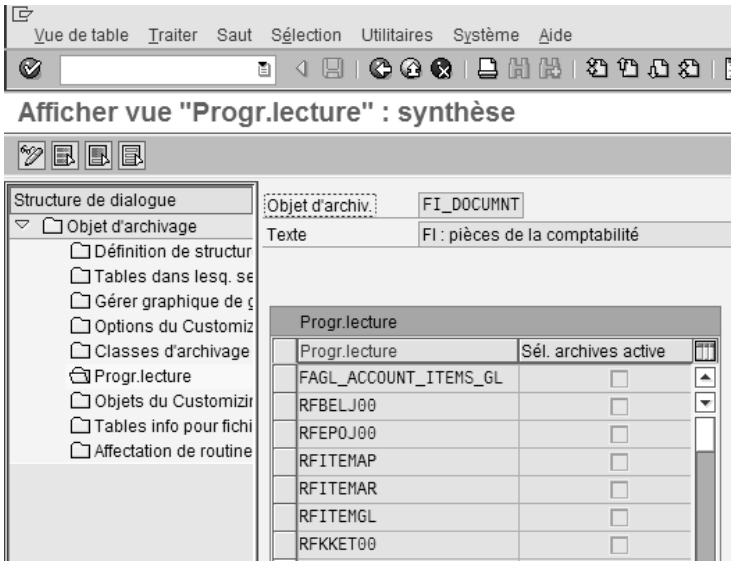


Figure 13 : liste des programmes pour la lecture des fichiers archives

4.6 Temps de résidence dans la base de données

Chaque objet archivage a des besoins de résidence dans la base de données différents en fonction des contraintes métiers. Pour ce faire, il est possible de lui associer un ou plusieurs codes transactions (voir Figure 14) qui permettent de définir ce temps de résidence par plusieurs critères.

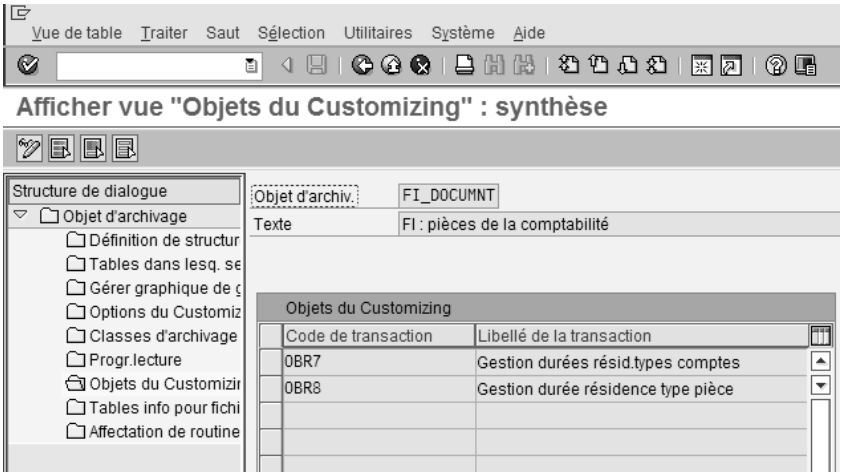


Figure 14 : liste des transactions permettant de définir le temps de résidence

Par exemple, pour une pièce comptable, il est possible de définir le temps de résidence en fonction de la société et du type de pièce (voir Figure 15).

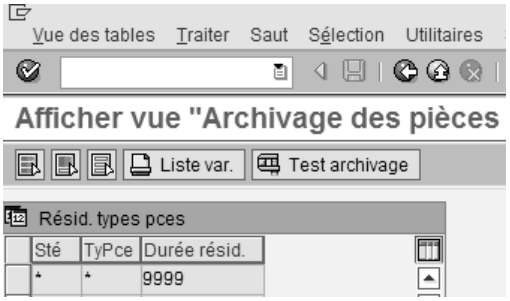


Figure 15 : exemple de définition temps de résidence

4.7 Tables information

La lecture des fichiers archives est faite de manière séquentielle, bloc par bloc. Pour optimiser la lecture de ces derniers, il est possible d'associer à l'objet d'archivage des tables d'index et d'informations (voir Figure 16).

Figure 16 : Liste des tables index

4.8 Fonctions spécifiques (user-exit)

La plupart des programmes des objets d'archivage sont déjà fournis par SAP, mais l'entreprise a parfois des particularités non prévues par l'éditeur. Les fonctions spécifiques ou « user-exit », pour utiliser le jargon SAP, sont appelées pendant l'exécution du programme standard pour modifier le comportement sans que le code de l'éditeur ne soit modifié.

4.9 Remarques

A la fin de ce chapitre consacré à l'objet d'archivage, il est important de faire remarquer que tous ces points de paramétrage sont interdépendants. Autrement dit, chaque élément doit interagir ave les autres afin de pouvoir obtenir les résultats escomptés.

Par exemple si une table d'information est créée et associée à l'objet d'archivage sans que le programme de création et suppression d'index ainsi que de lecture des archives soit créé ou conçu pour pendre en compte cette dernière, il n'y aura pas l'effet escompté.

5 Processus

Le processus d'archivage est une suite d'actions séquentielles matérialisées par des programmes exposés précédemment. La gestion de ce dernier, ainsi que d'autres fonctionnalités associées à l'archivage, sont orchestrées par la transaction SARA (voir Figure 17).

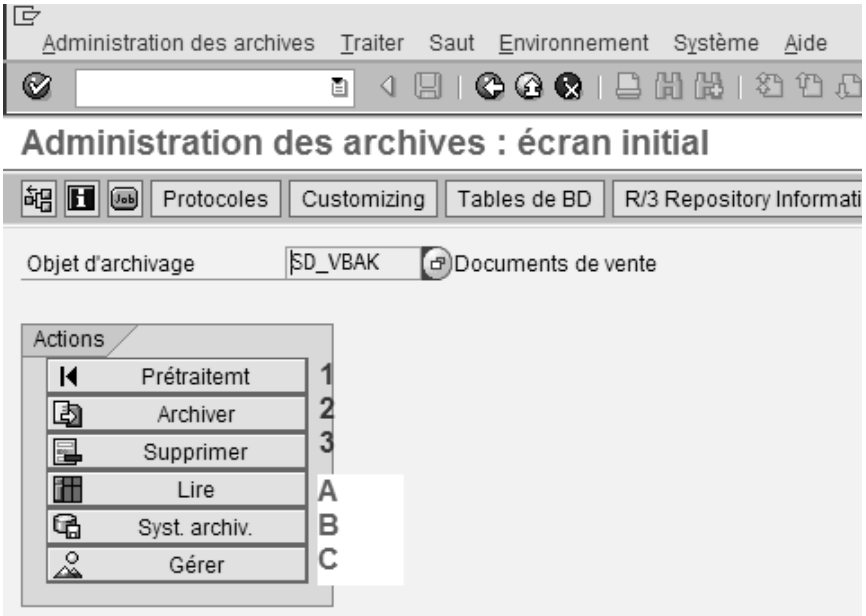


Figure 17 : Copie d'écran de la transaction SARA

La transaction SARA affiche uniquement les actions paramétrées (voir chapitre 4.1) et elles doivent être exécutées dans l'ordre du haut vers le bas (numérotées de 1 à 3 dans la figure 21). Certaines actions (notées par la lettre de A à C dans la figure 21) sont optionnelles ou informatives. En résumé, des paramètres influencent le processus et ce dernier peut être subdivisé en deux parties : processus principal et optionnel.

5.1 Paramétrage processus d'archivage (Customizing)

Par la transaction SARA, il est possible d'accéder aux points de paramétrage qui permettent d'influencer le processus d'archivage (voir Figure 18).

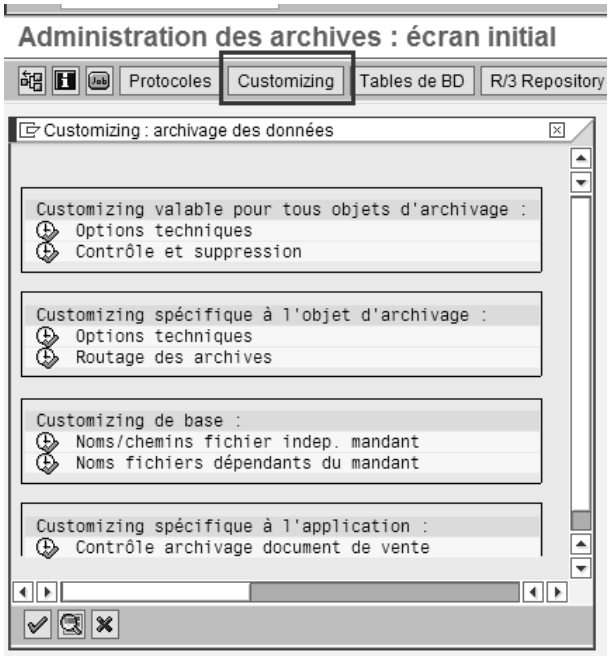


Figure 18 : paramètres processus archivage

La suite de ce sous-chapitre vise à donner une brève description de ces principaux paramètres.

5.1.1 Paramétrage valable pour tous objets d'archivage

Il y a deux ensembles de paramétrage (*customizing*) concernant l'ensemble des objets : options techniques et contrôle et suppression.

5.1.1.1 Options techniques

Les principales options techniques (voir Figure 19) indépendantes d'un objet d'archivage sont :

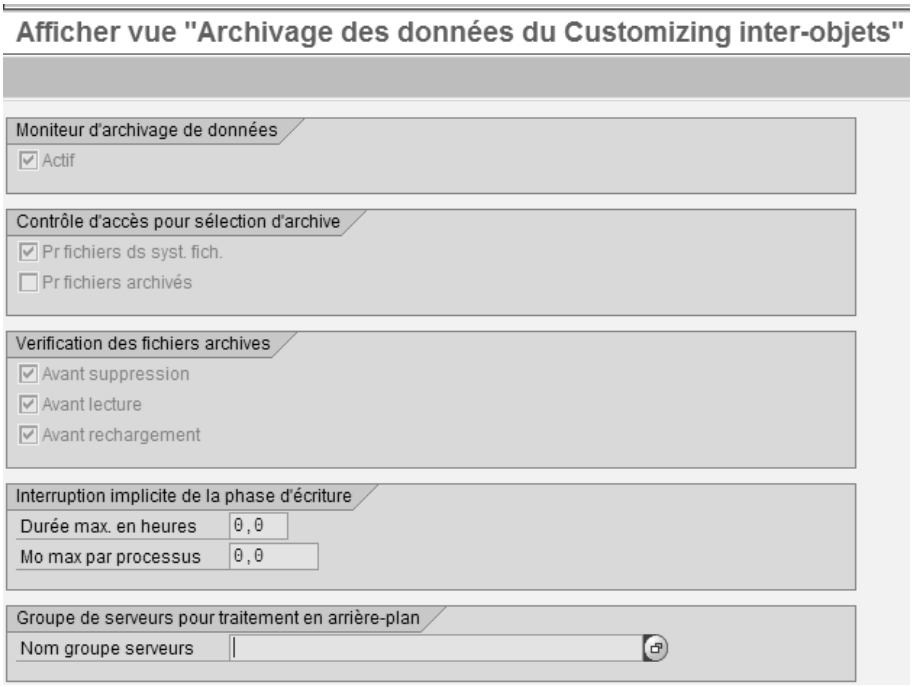


Figure 19 : Options techniques pour tous les objets archivage

- **Contrôle d'accès pour sélection d'archivage** : il permet de définir le contrôle d'accès soit délégué au système d'archivage soit par un contrôle sur le fichier lui-même ou les deux.
- **Interruption implicite de la phase d'écriture** : si une valeur différente de 0 est saisie le processus d'écriture est interrompu après le temps saisi.
- **Groupe de serveurs pour traitement en arrière-plan** : pour une instance de SAP il est possible de lui associer plusieurs serveurs. Un choisissant un serveur (ou groupe) l'ensemble du processus sera exécuté sur ce dernier. Ce paramètre est particulièrement utile dans le cas où un serveur est dédié à l'archivage sinon il vaut mieux laisser gérer le choix par le système.

5.1.1.2 Contrôle et suppression

Certains objets d'archivage peuvent partager un ensemble de paramètres. Il est fort probable que cette fonctionnalité sera développée davantage par SAP car, à l'heure actuelle, seul un objet peut utiliser ce point de paramétrage.

5.1.2 Paramétrage spécifique à l'objet archivage

Il y a deux ensembles de paramètres spécifiques à chaque objet d'archivage : options techniques et routage des archives.

5.1.2.1 Options techniques

Voici un descriptif des différentes options techniques :

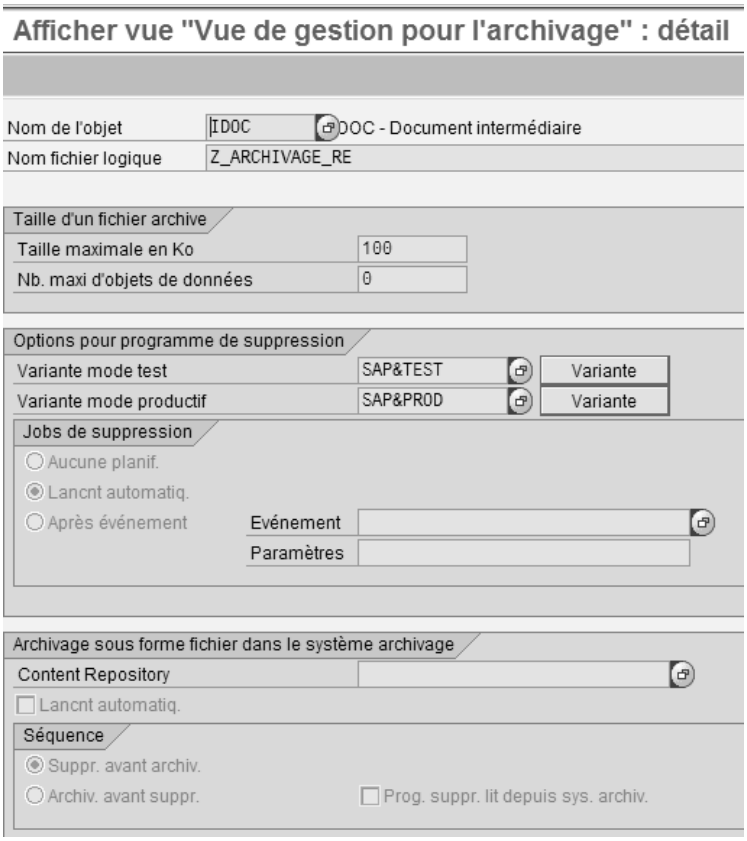


Figure 20 : Options techniques spécifiques objets archivage

- **Nom fichier logique** : le nom du fichier logique est une définition qui permet au système de définir le nom du fichier archive (pour plus d'informations, se référer à l'Annexe 2).
- **Taille d'un fichier archive** : elle permet de définir la taille maximum d'un fichier archivage, soit en limitant le kilo octet, soit en limitant le nombre maximal d'objets.
- **Options pour programme de suppression** : chaque objet d'archivage a son propre programme de suppression avec des paramètres spécifiques. Cependant, tous les objets ont au moins un paramètre test et un paramètre productif.

- Variante[34] mode test : la variante de test est utilisée par le processus pour lancer le programme de suppression en mode test, c'est-à-dire que le programme simula la suppression.
- Variante mode productif : la variante productive est utilisée par le processus lors de la suppression des données.

- **Jobs de suppression** : pendant le processus d'archivage, il est possible d'influencer le lancement du programme de suppression, soit manuellement (aucune planification), soit automatiquement (après le programme d'écriture), soit après un événement.[35]
- **Archivage sous forme dans le système archivage** : les fichiers archives peuvent êtres stockés dans un système d'archivage (voir III6 Archivelink) au lieu d'être placés dans un répertoire (voir Annexe 2).

  Si le code de lancement automatique est choisi, les fichiers archives seront transférés dans un système d'archivage après leur traitement. En plus de ce choix, on pourra opter pour trois comportements différents :
  - Suppression avant archivage : le programme de suppression est exécuté avant que le fichier archive soit stocké dans le système d'archivage.

    N.B. : cette option permet de gagner en performance, mais elle comporte des risques en cas de problèmes.
  - Archivage avant suppression : contrairement au précédent paramètre, les fichiers archives sont en premier entreposés dans le système d'archives et ensuite seulement les données sont supprimées de la base de données.

    NB : Cette option apporte un bon compromis de sécurité/performance.
  - Programme de suppression lit depuis le système d'archivage : cette option indique que la donnée est supprimée après la lecture de cette dernière dans le fichier archive.

[34] Une variante est le moyen mis à disposition par SAP pour mémoriser les valeurs des paramètres de l'écran de sélection. Grâce aux variantes, les utilisateurs peuvent lancer les programmes sans devoir ressaisir toujours les mêmes paramètres.

[35] SAP encapsule un moteur de Workflow, il est donc possible de créer ou d'utiliser des déclencheurs tels que la fin d'un programme, une action de l'utilisateur, etc. pour lancer le programme de suppression.

NB : cette option est de loin la plus sécurisante, mais elle nécessite un traitement bien plus long.

5.1.2.2 Routage des archives

En raison des problèmes de conformité légale, les entreprises, en particulier les multinationales, sont de plus en plus contraintes de stocker les données archivées séparément afin de répondre aux exigences de conservation des différents pays.

Par cette fonction, il est donc possible de créer des règles et des conditions (voir Figure 21) permettant de router les fichiers archives vers des systèmes d'archivages et/ou des noms de fichiers logiques différents. Grâce à ce routage, ces entreprises peuvent jouir d'une plus grande souplesse.

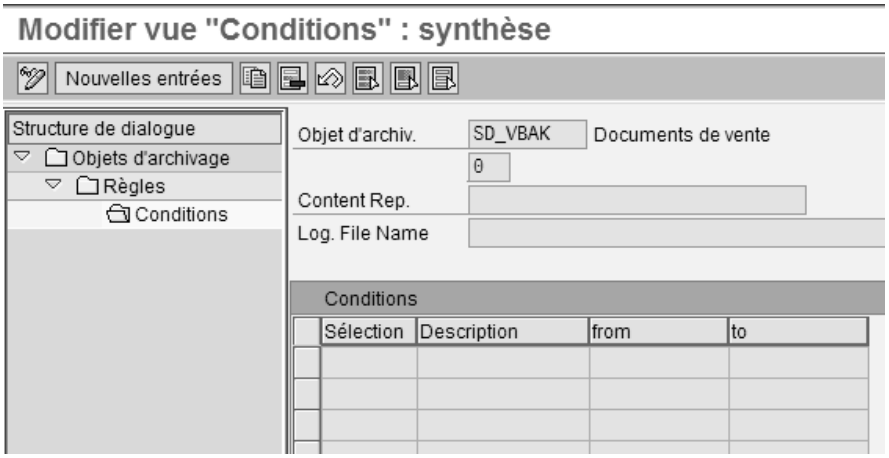


Figure 21 Options de routage

5.2 Processus principal

Le processus principal est composé de trois étapes obligatoires et de deux étapes facultatives (dans le sens qu'elles ne doivent pas être systématiquement présentes dans tous les objets). Les étapes obligatoires sont : écriture, stockage et suppression. Les étapes facultatives sont : prétraitement et post traitement.

5.2.1 Prétraitement (ou préprocesseur)

Certains objets d'archivage requièrent un prétraitement pour vérifier et préparer les données avant l'étape de l'écriture. L'utilisateur choisit un certain nombre de filtres par rapport aux critères à sa disposition. Un certain nombre de contrôles sont effectués pour garantir la cohérence des données filtrées. Enfin, les données jugées cohérentes selon les filtres sont marquées avec le statut « à archiver ».

Il faut noter que normalement cette étape est englobée dans l'étape « écriture », mais pour des raisons de performances, elle a été scindée en deux.

5.2.2 Ecriture et stockage

Le programme d'écriture est, généralement, le point de départ de l'archivage. Le but principal de ce dernier est de créer des fichiers d'archive selon la structure de l'objet et les stocké.

Le programme s'articule autour trois parties : filtre, contrôle et écriture.

- Filtre : par une interface graphique, l'utilisateur peut choisir un certain nombre de paramètres et de filtres.
- Contrôle : les données sont sélectionnées selon les filtres et contrôlées pour garantir la cohérence. Dans le cas d'objets avec prétraitement, cette partie se limite à sélectionner uniquement les données avec statut « à archiver ».
- Ecriture : toutes les données cohérentes sont écrites dans un fichier d'archive[36] stocké.

5.2.3 Suppression

Cette étape, comme son nom l'indique, supprime physiquement de la base de données les tuples qui ont été enregistrés dans un fichier. Cependant, pour certains objets, la suppression physique est remplacée par un témoin de suppression.

Il existe différentes sortes de suppressions avec des caractéristiques différentes :

- Uniquement suppression : les données sont supprimées sans aucun contrôle. Cette option est assez risquée, mais plus rapide et adaptée pour des objets de type techniques ou des objets avec post traitement.
- Vérification avant suppression : cette option fait un contrôle de lecture du fichier archive et supprime ensuite les données, à la condition que le test soit probant.
- Vérification après suppression : le contrôle de lecture du fichier archive est fait uniquement après la suppression.

[36] Le fichier archive est un fichier compressé environ 3 fois avec une structure propriétaire. Ce dernier est lisible uniquement par SAP. Pour faciliter la manipulation et gagner en rapidité de lecture, un paramètre permet de définir une taille maximale. Cela implique que pendant l'étape d'écriture plusieurs fichiers peuvent être créés.

5.2.4 Post traitement (ou postprocesseur)

Comme pour le prétraitement, cette étape a été créée pour scinder en deux la suppression pour des raisons de performances. Dans ce cas, l'étape de suppression met un témoin de suppression sur les données qui seront physiquement effacées pendant le post traitement.

5.3 Processus annexes

En annexes aux étapes d'archivage, il est possible de lancer des routines de rechargement, de lecture et de gestion des archives (voir Figure 22).

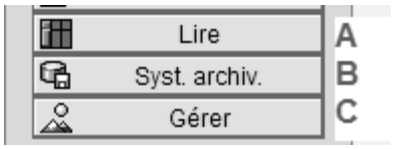


Figure 22: étapes processus annexes

5.3.1 Rechargement (bouton Syst. Archiv.)

Le rechargement permet de recharger les données supprimées par le processus d'archivage en partant des fichiers archives. Cette opération est à utiliser avec beaucoup de précautions car elle peut engendrer des incohérences fonctionnelles et/ou des conflits d'unicité.

Par exemple si une commande de vente est archivée depuis plusieurs années et pendant ce temps tous les codes du canal de distribution ont été modifiés suite à une réorganisation, après le rechargement, le canal de distribution risque de n'être plus pertinent selon la nouvelle organisation. Il pourrait également arriver que le numéro de commande, clé unique de la table VBAK, soit utilisé par une nouvelle, ce qui causerait un conflit d'unicité.

Pour toutes ces raisons, le rechargement est très rarement utilisé, sauf pour corriger un archivage erroné..

5.3.2 Visualisation (bouton lecture)

La visualisation, dans ce contexte, est une fonctionnalité fournie pour vérifier que l'écriture des fichiers archive s'est déroulé correctement.

5.3.3 Gestion des archives

Cette fonctionnalité dédiée aux opérateurs d'archives permet de visualiser les processus d'archives avec les fichiers associés (voir Figure 23)

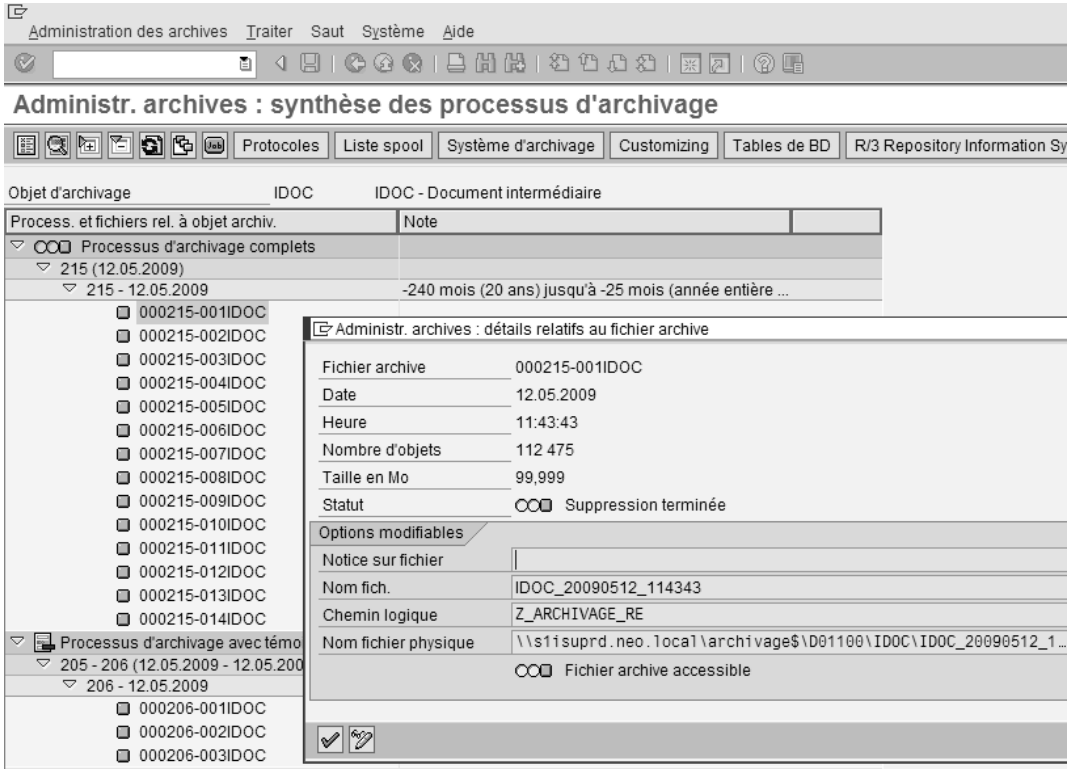


Figure 23: gestion des archives

6 ArchiveLink

ArchiveLink est une interface et un service de communication pour lier tous documents entrants (*« Incoming Document »*) ou sortants (*« Outgoing Documents »*) entre SAP et une autre application. Les flèches de la Figure 24 représentent ce lien fourni par ArchiveLink et le serveur encerclé au milieu, de la Figure 24, symbolise une application.

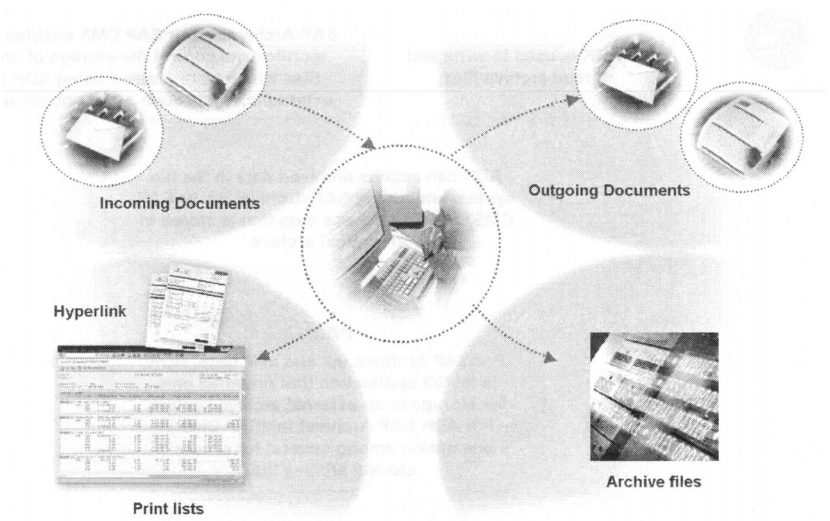


Figure 24 : documents entrants et sortant entre SAP et d'autres applications.

6.1 Principe

Archivelink peut être comparé, dans une certaine mesure, à un lien hypertexte. Comme tous les internautes le savent, ce lien représenté par un texte bleu souligné permet d'ouvrir une page html qui peut être stockée dans un autre site internent. ArchiveLink est un module transversal permettant de faire un lien unique entre un objet métier et des fichiers annexes. Pour ce faire, SAP met à disposition un paramétrage spécifique qui s'articule autour de trois éléments principaux : objet métier, table de lien et protocole HTTP.

6.2 Objet métier

Un objet métier ou « *Business Objet* » s'inspire de la théorie de la programmation objet avec des attributs, des méthodes, des événements, etc. Dans ce contexte, l'objet est une implémentation d'une classe représentant un élément réel comme un article, une commande, un bon de livraison, une facture, etc. Il faut remarquer que le module SAP d'archivage est, lui aussi, un objet métier, à ne pas confondre avec les objets archives. Les objets archive servent uniquement pour l'archivage et sont, dans une certaine mesure, un sous-ensemble des objets métier. Par exemple l'objet métier facture décompte est archivé avec les objets archive ISU_PRDOCH et ISU_PRDOCL (voir Figure 8).

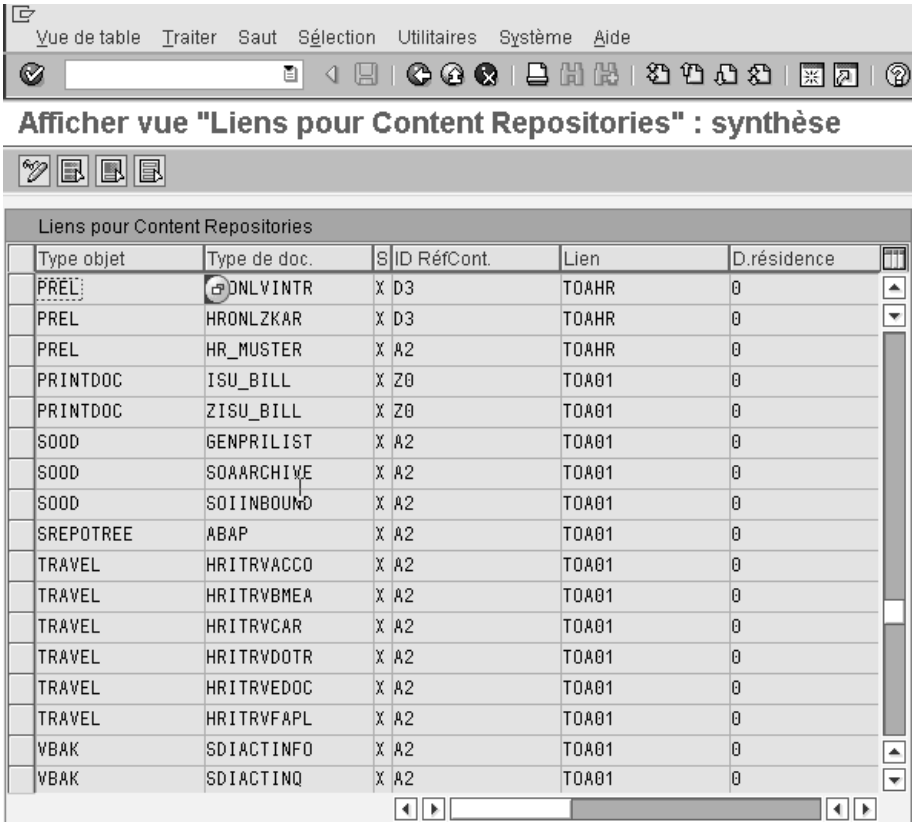
Pour illustrer le principe, un article (objet métier) lui est associé à plusieurs fichiers annexes (grâce aux tables de liens) comme les spécifications, les caractéristiques qualités, etc.

6.3 Tables de liens

Comme son nom l'indique, ces tables de liens de la base de données SAP permettent de créer un lien unique entre l'objet métier et le document stocké sur le serveur de contenu. Ces tables sont utilisées par les programmes de création, lecture et suppression des fichiers annexes. Elles sont transparentes pour l'utilisateur, sauf certaines représentées par des vues de paramétrage utilisées pour la mise en œuvre d'ArchiveLink. Le propos n'est pas de faire une cartographie exhaustive de toutes ces dernières, mais simplement de comprendre comment ce lien est défini par les vues « Liens pour Content Repositories », «Content Repositories » et comment ces liens sont enregistrés par la table « TOA01 ».

6.3.1 Vue « Liens pour Content Repositories »

La vue de paramétrage « Liens pour Content Repositories » permet de définir comment le lien doit s'opérer (voir Figure 25).



Type objet	Type de doc.	S	ID RéfCont.	Lien	D.résidence
PREL	ONLVINTR	X	D3	TOAHR	0
PREL	HRONLZKAR	X	D3	TOAHR	0
PREL	HR_MUSTER	X	A2	TOAHR	0
PRINTDOC	ISU_BILL	X	Z0	TOA01	0
PRINTDOC	ZISU_BILL	X	Z0	TOA01	0
SOOD	GENPRILIST	X	A2	TOA01	0
SOOD	SOAARCHIVE	X	A2	TOA01	0
SOOD	SOIINBOUND	X	A2	TOA01	0
SREPOTREE	ABAP	X	A2	TOA01	0
TRAVEL	HRITRVACCO	X	A2	TOA01	0
TRAVEL	HRITRVBMEA	X	A2	TOA01	0
TRAVEL	HRITRVCAR	X	A2	TOA01	0
TRAVEL	HRITRVDOTR	X	A2	TOA01	0
TRAVEL	HRITRVEDOC	X	A2	TOA01	0
TRAVEL	HRITRVFAPL	X	A2	TOA01	0
VBAK	SDIACTINFO	X	A2	TOA01	0
VBAK	SDIACTINQ	X	A2	TOA01	0

Figure 25 : copie d'écran de la vue « lien pour Content Repositories ».

Les champs clef de cette table sont :

- **Type objet** : ce code définit le nom de la classe de l'objet métier

- **Type de document** : ce code définit le format de fichier. L'architecte de la base de données a ajouté un niveau supplémentaire à la définition des formats de fichier pour permettre de mieux le décrire selon le contexte (voir Figure 26)

Vue de table Traiter Saut Sélection Utilitaires Système Aide

Afficher vue "Tous types de documents" : synthèse

Tous types de documents

Type de do...	Désignation complète	Format de document	Statut
VIOMVANS	Lettre bail	PDF	
VIOMWFRKAU	Changem.locataire : validation caution	PDF	
VIONKABR	Décompte des charges	PDF	
VIOOWNER01	Décompte du propriétaire	PDF	
VIOPEBS	Caution personnelle	PDF	
VIORACC	Conditions bail dans devise origine	PDF	
VIORCCOLL	Dépôt garantie pour bail	PDF	
VIOSLST	Décompte CA	PDF	
VIOSOST	Factures pour baux	PDF	
VIOTNNTCHG	Correspondance pour changement locataire	PDF	
VIOVMANP	Révision du loyer selon loyer référence	PDF	
VIOWBNL	Evaluat.logements Pays-Bas	PDF	
VIOWDRANG	Annulation d'une offre de bail	PDF	
VIOWOMV	Bail d'habitation	PDF	
VIOWOMVANS	Bail habitation - Lettre	PDF	
ZISU_BILL	Copie de facture ISU	PDF	

Figure 26 : exemple d'entrée de la table type de document.

- **S** : c'est un champ de type booléen. Si X est activé, cela signifie qu'une seule classe d'objet métier et qu'un type de document peuvent être actifs.
- **ID Réf.Cont.** : ce code est le nom du content serveur.
- **Lien :** le nom de la table de lien

Champ non clef :

- **D. résidence** : il indique le nombre de jours au bout desquels un document archivé peut être supprimé dans les tables de liens.

En résumé, cette vue définit le serveur de contenu et la table de lien pour une « classe d'objet métier » et un « format de fichier ».

6.3.2 Content Repositories

La vue de paramétrage « Content Repositories » de la Figure 27 définit le moyen de communiquer avec le serveur de contenu qui est distant.

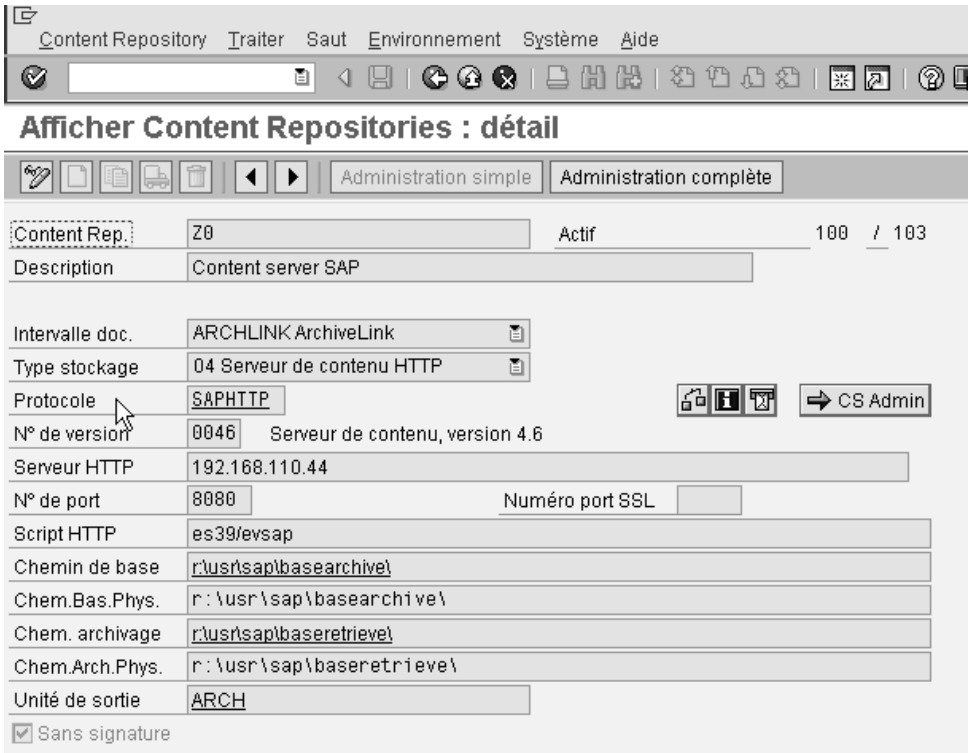


Figure 27 : Vue de paramétrage du Content Repositories.

Le seul champ clef de cette table est le code « Content Rep ». Ce dernier est simplement un champ alphanumérique à deux positions. Les autres champs significatifs sont :

- **Intervalle document** : il correspond à la tranche de numéro interne utilisée pour la génération de noms de fichiers et de numéros d'ordres univoques pour les ordres asynchrones.
- **Type de stockage** : ce code permet de définir le type de communication : http, RFC, etc.
- **Protocole** : par ce champ, SAP permet d'utiliser plusieurs protocoles, mais il reste souhaitable d'utiliser uniquement le saphttp, qui se base sur un protocole standard.
- **N° de version** : c'est la version du protocole saphttp.
- **Serveur http :** c'est l'adresse IP du serveur de contenu.
- **N° de port** : c'est le numéro de port du serveur de contenu.
- **Numéro port SSL** : c'est le numéro de port SSL, si utilisé (?), du serveur de contenu
- **Script http** : cette routine est fournie par le serveur de contenu pour établir la communication.
- **Chemin de base** : ce champ définit l'emplacement du disque où le fichier à stocker est placé temporairement avant qu'il soit entreposé sur le serveur de contenu par une commande http.

- **Chemin d'archivage** : lors de la consultation d'un fichier stocké sur le serveur de contenu, il est copié temporairement dans cet emplacement du disque pendant le temps de lecture de SAP.
- **Unité de sortie** : derrière ce code se cache l'imprimante utilisée pour le stockage des listings sortants de SAP vers le serveur de contenu.
- **Sans signature** : si cette coche est validée, les échanges http entre SAP et le serveur de contenu se font avec une signature de cryptage.
-

Ces boutons servent à vérifier si la communication avec le serveur de contenu se fait correctement.

6.3.3 Table TOA01

Pour des raisons de performances lors d'un très grand nombre d'entrées, SAP a prévu, dans son architecture, la possibilité de créer autant de tables de lien que le besoin se fait sentir. Cependant, la structure est et doit être identique à l'exemple de la Figure 28.

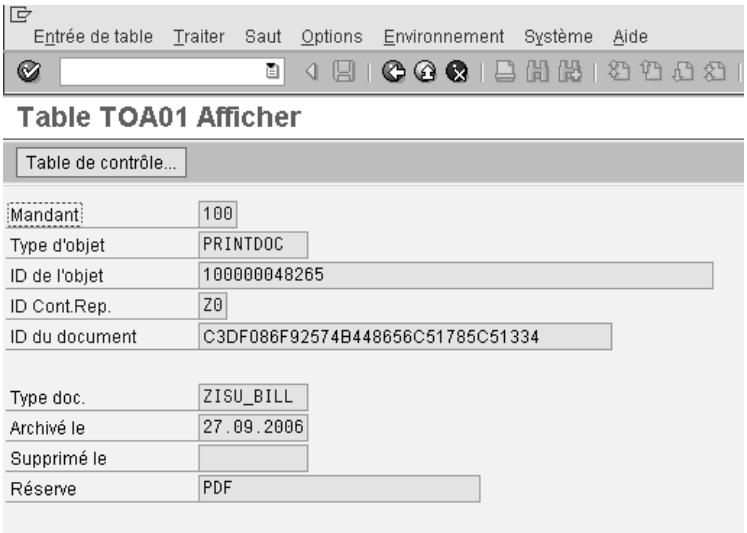


Figure 28 : exemple d'un tuple de la table d'une table de lien.

Les cinq premiers champs de la table des liens permettent de rendre ce lien unique. Ces champs sont les suivants :

- **Mandant** : le mandant est une séparation logique des bases de données à l'intérieur d'une installation du système SAP.
- **Type d'objet** : ce code définit le nom de la classe de l'objet métier. Dans l'exemple de la copie d'écran, il correspond à un numéro de facture.
- **ID de l'objet** : c'est le numéro unique de l'objet métier. Dans cet exemple, c'est le numéro de facture.

- **ID Cont. Rep.** : ce code est le nom du serveur de contenu.
- **ID du document** : ce code est créé par un algorithme permettant de le rendre unique du côté SAP et du côté serveur de contenu. Ce dernier est donné toujours en paramètre pour tous les échanges entre ces deux systèmes.

Les autres champs donnent quelques informations supplémentaires :

- **Type doc.** : ce code définit le format de fichier.
- **Archivé le** : cela indique la date à laquelle le fichier a été stocké sur le serveur de contenu.
- **Supprimé le** : cela indique la date de suppression du fichier.
- **Réserve** : ce champ a été ajouté pas SAP pour des besoins futurs.

6.4 Protocole http.

Archivelink utilise le protocole http version 1.1 développé par le consortium World Wide Web (www.w3c.org). Cependant, SAP utilise une certaine syntaxe pour communiquer avec les contents externes. Cette dernière est exposée sommairement dans ce chapitre.

6.4.1 Syntaxe

http://servername:port/script?command&parameters

- **servername** : c'est le nom du serveur par lequel la requête est adressée.
- **Port** : c'est le numéro de port du serveur (servername) par lequel la requête est adressée.
- **Script** : c'est le nom du programme pour accéder au serveur interpellé. Ce dernier peut être une DLL,[37] un script CGI[38] ou ASP[39].
- **Command** : c'est une commande pour définir l'accès au serveur distant par des opérations de lecture, création, changement et suppression.

---

[37] DLL (Dynamic Link Library) est une fonction qui fait partie d'une bibliothèque dynamique. Cette dernière est appelée lors de l'exécution d'un programme, mais n'en faisant pas partie. L'avantage principal d'une bibliothèque dynamique est la réduction de la taille d'un exécutable, puisque certaines parties du logiciel se situent dans le système, et l'introduction de validations identiques pour chaque programme, puisque tous sont liés à une seule et même bibliothèque.

[38] CGI (Common Gateway Interface) est une interface normalisée utilisée par les serveurs HTTP. Ce dernier, au lieu d'envoyer le contenu d'un fichier (page HTML, image...), exécute un programme puis retourne le contenu généré, comme s'il s'agissait d'un contenu de fichier.

[39] ASP (Active Server Pages) est un langage de programmation développé par Microsoft permettant l'exécution de script sur le serveur.

- **Parameters** : pour chaque commande, il faut ou il peut y avoir des paramètres spécifiques.

6.4.2 Exemple de commande http

L'exemple suivant est une commande http qui demande au serveur de contenu de renvoyer un document :

http://192.168.110.143:1090/ContentServer/ContentServer.dll?get&pVersion=0046&contRep=Z0&docId=F31C5490BDF4E44B959AE10424CC96D0

- **192.168.110.143** : adresse IP du serveur de contenu (voir 6.3.2).
- **1090** : numéro du port de serveur de contenu (voir 6.3.2).
- **ContentServer/ContentServer.dll** : chemin et nom du programme pour l'accès au serveur (voir 6.3.2),
- **Get** : est la commande de lecture
- **Version=0046** : paramètre qui indique la version de protocole http utilisée par ArchiveLink.
- **Rep=Z0** : nom du « Content Repositorie » (voir 6.3.2)
- **docId=F31C5490BDF4E44B959AE10424CC96D0** : numéro unique du document dont la correspondance se trouve dans la table des liens (Figure 29) .

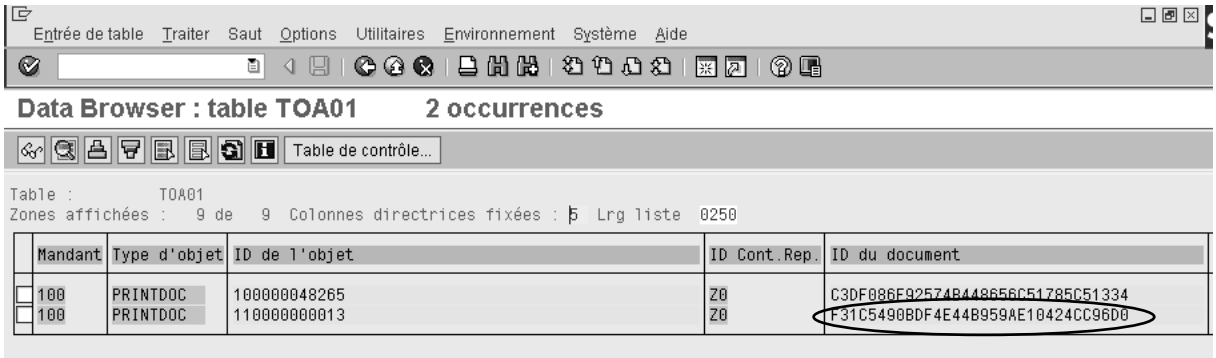


Figure 29: Table des liens

7 Bilan de conformité

Après ce survol des fonctionnalités fournies par l'éditeur allemand concernant l'archivage des données, ce chapitre effectue un bilan de conformité en fonction des exigences légales vues dans la section précédente, car, théoriquement, l'archivage de données est un sous ensemble de l'archivage électronique.

Disposition sur la conservation électronique (voir II3.1.3 Disposition sur la conservation électronique)	**Degré de couverture** ☺ 😐 ☹	**Conformité des outils d'archivage des données SAP**
Intégrité (authenticité et infalsifiabilité)	😐	Les fichiers archives sont dans un format propriétaire. Ils sont illisibles par aucun d'autre moyen que par l'environnement SAP et en plus aucun outil standard SAP permet de les modifier. Dans ce sens il serait presque possible de considérer que l'intégrité est respectée.
Disponibilité	😐	Les fichiers archives sont visualisés par les mêmes écrans que ceux de la transaction qui a généré la donnée. L'accès est donc rapide et assez facile. Cependant, il faut noter que si l'environnement ayant généré les fichiers archives n'est plus disponible, il sera très difficile, voire impossible, de les relire.
Organisation	☺	Le système d'autorisation standard de l'environnement SAP s'applique également pour l'accès aux fichiers archives. Des problèmes d'autorisation dans la lecture des fichiers archives ne peuvent donc pas surgir durant la conservation.
Séparation entre les informations archivées et celles actuelles	☺	Lors de la consultation, l'utilisateur doit choisir une transaction spécifique pour consulter les archives et une mention « Archive » apparaît sur l'écran de visualisation.
Contrôle et migration des données	☹	L'éditeur garantit qu'avec son environnement, les données seront toujours lisibles même après des changements successifs de versions. Cela peut être acceptable pour la conservation des données à court terme, cependant ce n'est pas satisfaisant pour la conservation à moyen et long terme.
Temps de conservation	☹	Le processus d'archivage ne prend pas en compte la suppression automatique des archives à la fin du délai de conservation. L'opérateur doit gérer cela de manière manuelle et la suppression est faite directement dans le système d'archivage. <u>N.B.</u> : il ne faut pas confondre temps de résidence dans la base de données avec temps de conservation des archives. Le premier est un paramètre de contrôle utilisé par le processus d'archivage pour limiter l'archivage avant une certaine date, alors que le second est le temps de résidence des fichiers archives dans le système d'archivage.

Conclusion :

Je pense que les outils pour l'archivage des données SAP ne sont pas adaptés pour l'archivage électronique pour deux raisons principales :

- Format propriétaire : comme évoqué dans le chapitre II4.1 si pour une raison quelconque les progiciels SAP devaient disparaître il sera très difficile, voire impossible, de relire les archives.
- Une contrainte supplémentaire s'ajoute au format de fichier propriétaire, c'est la quasi impossibilité de relire des fichiers sans le système qui en a été l'origine. En clair, même si théoriquement tout système SAP est en mesure de lire des fichiers archives, il reste très difficile de le faire sans les tables index stockées dans la base de données d'origine.

Pour ma part, je crois que de manière générale il est plus judicieux de générer les documents soumis à des contraintes légales dans un format tel que pdf/A et de les conserver dans un système d'archivage.

En conclusion, l'archivage des données SAP est un excellent moyen pour purger les tables et ainsi préserver les performances, faciliter l'administration et réduire le temps d'indisponibilité en cas de problème de la base de données.

IV Réalisation du projet

Le projet d'archivage de données SAP est avant tout un projet d'intérêt technique. Si chaque administrateur de base de données connaît les avantages de garder la base la moins volumineuse possible, les utilisateurs du progiciel souhaitent, au contraire, que toutes leurs données demeurent en ligne pour les consulter le plus aisément possible, la taille de stockage n'étant guère leur problème. Cependant, ces deux visions divergentes doivent coexister et il faut trouver le meilleur compromis tout en respectant les exigences légales, organisationnelles et techniques.

SAP a mis au point une méthodologie appelée ASAP (pour *Accelerated* SAP) pour aider les chefs de projet à mener à bien les projets d'intégration du progiciel allemand. Le présent projet peut également être géré et suivi à l'aide de cet outil.

ASAP s'appuie sur une feuille de route appelée *roadmaps*, jalonnée de cinq étapes (voir Figure 30) : préparation de projet, *business blueprint*[40] (ou conception générale), réalisation, préparation finale et *go live* et support.

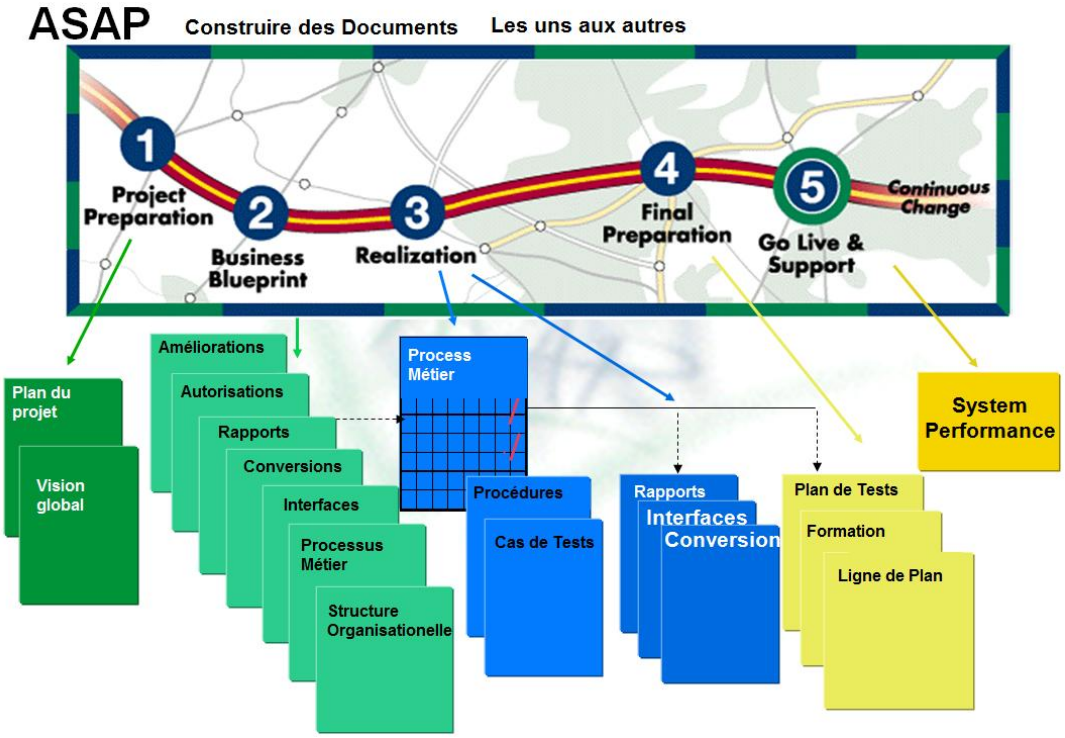


Figure 30: feuille de route ASAP

[40] « blueprint » en référence au tirage de plans transparents originaux transformant dans des machines à tirage de papier à l'ammoniaque des papiers jaunes désazotés en plans bleus.

1 Préparation du projet

La préparation de ce projet a débuté par une brève formation effectuée par SAP sur les outils d'archivage. Ces informations théoriques ont ensuite été complétées par un témoignage expliquant la mise en œuvre d'un tel projet dans l'entreprise Migros SA. Enfin, la participation à plusieurs séminaires consacrés à l'archivage électronique, l'étude approfondie des ressources citées en annexe et de nombreux tests dans le système ont permis de mieux comprendre son fonctionnement.

Après ce travail préliminaire qui était nécessaire, le « vrai » projet a commencé par une analyse de la croissance du volume de la base de données, le choix des objets d'archivage par une analyse selon la loi Pareto des tables les plus volumineuses, la proposition du temps de résidence en ligne et la définition de l'équipe de projet.

1.1 Analyse de la croissance du volume de la base de données

Cette analyse de la croissance du volume de la base de données est un indicateur permettant d'évaluer grossièrement le niveau d'urgence de la mise en œuvre le projet d'archivage de données. Grâce à la pente de la courbe de croissance, il est possible d'estimer quand il faut prévoir l'achat d'espace disque supplémentaire ou le remplacement des serveurs pour garantir le même niveau de temps de réponse. L'autre alternative, objectif de ce projet, est d'archiver avant qu'il ne faille entreprendre des modifications physiques.

1.1.1 Courbe de croissance

Le tableau ci-dessous représente l'évolution mensuelle du volume de la base de données durant les 13 derniers mois (voir **Tableau ii**).

	avr.09	mai.09	juin.09	juil.09	août.09	sept.09	oct.09	nov.09	déc.09	janv.10	févr.10	mars.10	avr.10
Volume net Go	643	669	677	683	698	509	525	544	562	577	592	602	615
Croissance Go		26	8	6	15	-189	16	19	18	15	15	10	13
Croissance moyenne (à partir d'octobre 2009) en Go												**15**	

Tableau ii: croissance de la base de données

La Figure 31 illustre avec un graphique les valeurs du tableau précédent. En septembre, il y a une baisse très importante du volume suite à des actions de purge (décrites plus loin) et à sa réorganisation.[41]

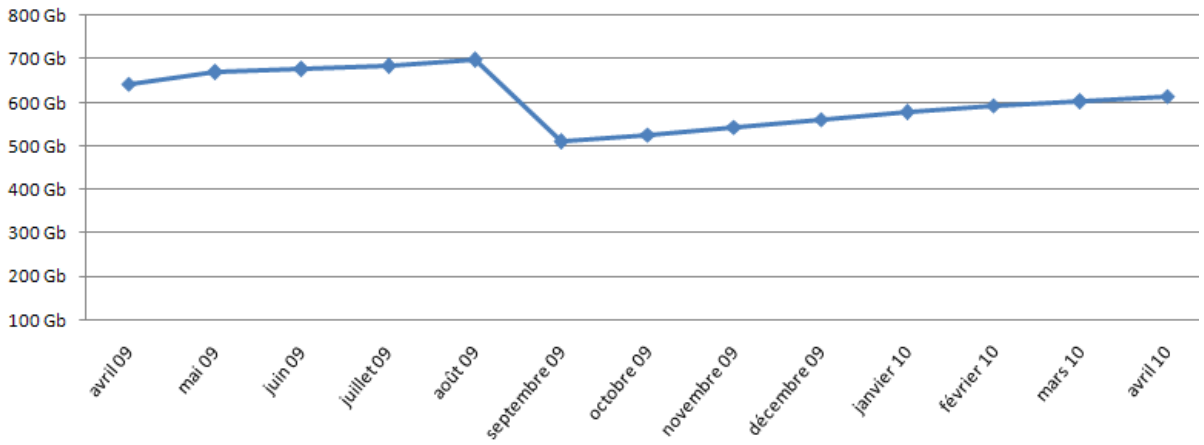


Figure 31 : croissance de la base de données

1.2 Analyse selon la loi Pareto

A l'heure actuelle, la base de données SAP comporte 72'419 tables différentes. Il est utopique de toutes les archiver, cela engendrerait beaucoup d'efforts inutiles. Par contre, il serait plus judicieux de classer les plus volumineuses (voir Tableau iii) et d'axer l'archivage sur ces dernières pour que l'impact sur le volume de la base de données soit significatif. Cette démarche s'inspire sur la loi de Pareto[42] : environ 80 % du volume est le produit de 20 % des tables.

[41] La réorganisation de la base de données est une opération qui vise à modifier la structure physique de cette dernière. Cette démarche permet de récupérer les espaces libres entre les segments au sein d'un data file, et donc de réduire la taille de ces fichiers.

[42] La loi de Pareto, aussi appelée loi des 80/20, est une loi empirique inspirée des observations de Vilfredo Pareto, économiste et sociologue italien : environ 80 % des effets sont le produit de 20 % des causes. Cette « loi », bien qu'empirique, a été formalisée en mathématiques par la distribution de Pareto (voir http://fr.wikipedia.org/wiki/Loi_de_Pareto).

Table	Description de la table	Go
DBERCHZ1	Lignes individuelles de document	49,55
DFKKOPK	Postes pour la pièce compte courant	18,00
DFKKMOP	Postes pour pièce compte courant type	13,92
DFKKOP	Postes pour la pièce compte courant	12,82
CDCLS	Comptabilité : index secondaire pr comptes généraux	12,41
DBERCHZ3	Lignes individuelles de docmt (dées de montants et de prix)	10,80
COEP	Objet CO: Postes individuels liés à la période	9,01
DBERDL	Lignes document d'impression	8,25
DBERDLB	Référence d'une ligne doct imprimé à ligne doct CalcFacturat	7,55
EDI40	Document de relevé	7,25
DBERCHZ2	Lignes individuelles de document (données d'appareil)	7,22
COSP	Objet CO : totaux coûts - écritures externes	5,99
CDHDR	En-tête de document de modification	4,52
GLPCA	EC-PCA : postes individuels au réel	4,46
KNVP	Base de données clients : rôles partenaire	3,95
DFKKOPBW	Extraction de postes non soldés	3,07
KNVV	Fiche client : données ventes	3,04
DFKKKO	Données d'en-tête pour pièce compte courant	2,91
CE41000_ACCT	SIE	2,76
ERDK	Facture / données d'en-tête	2,68
DFKKCIBW	FICA : extraction postes comparés	2,45
EABL	Document de relevé	2,44
REPOSRC	Codes source de programmes	2,43
PTDW_TIMES_DB	Extraction des données de temps (durées du travail)	2,39
D010TAB	Table pour utilisation table <-> programme	2,36
BALDAT	Protocole d'application : données d'un protocole	2,35
SWELOG	Table pour entrées de la trace des événements	2,20
DFKKZP	Lot de paiements : données de paiement	2,13
EDIDS	Enregistrement de statut (IDOC)	1,91
WBCROSSGT	Index pour types globaux - cas d'emploi Workbench	1,89
DBTABLOG	Enr. de prot. pr modifications tables	1,75
DFKKZR	Formulaire de paiement	1,69
ERCH	Données de facture	1,66
CE41000	SIE	1,65
MSEG	Segment de document pour article	1,60
ETTIFN	Val. fixes installation (normales)	1,46
DOKCLU	Formulaire de paiement	1,46
GVD_SEGSTAT	Oracle monitoring: Copy of GV$SEGSTAT	1,32
BSIS	Comptabilité : index secondaire pr comptes généraux	1,19
SRRELROLES	Service relation objet : rôles	1,19
SOFFCONT1	SOFF : table pour contenus de documents (import/export)	1,19
RFBLG	Codes source de programmes	1,19
BDCP	Pointeur de modification	1,15
PPOIX	Index ligne résultat du calcul de paie -> ligne d'écriture	1,13
FKKMAZE	Historique de relance, lignes pièce	1,13
ZISU_LOG_FACT	Facture imprimée	1,13
D010INC	Table des utilisations des includes ABAP	1,12
COEJ	Objet CO: Postes individuels sur l'année	1,07
DPAYP	Programme de paiement - données pour le poste payé	1,07

Tableau iii: classement des 50 tables les plus volumineuses

1.3 Cibler les objets d'archivage

A partir de ce classement, il est possible de déterminer les objets d'archivage grâce à la transaction DB15. Cette dernière renvoie la liste des objets d'archivage par rapport une table et vice-versa (voir Figure 32).

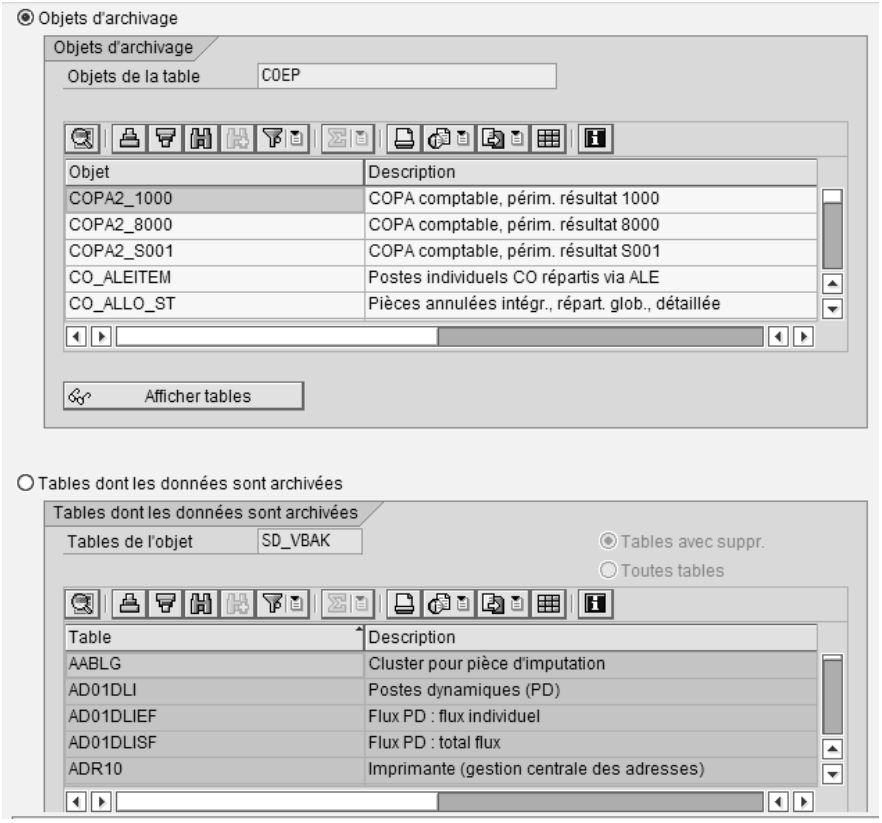


Figure 32: Transaction DB15

Toutefois, pour les tables n'ayant pas d'objet, il faut évaluer la possibilité d'en créer un sur mesure. A noter que le développement est en général conséquent et, surtout, qu'il doit être maintenu tout au long de la conservation des données. Compte tenu de ce fait, les bonnes pratiques suggèrent en général d'éviter un tel développement.

Pour revenir à l'exemple de la table COEP de la Figure 32, un grand nombre d'objets d'archivage mis à disposition par SAP permet d'archiver cette dernière, mais il faut cibler quel est l'objet le plus pertinent, c'est-à-dire celui qui va avoir un meilleur impact par rapport l'environnement. Dans ce but, SAP a édité un guide[43] répertoriant les tables les plus courantes, avec des conseils sur le meilleur moyen de les purger et les archiver. Si toutefois des tables ne sont pas répertoriées, la solution est de tester tous les objets d'archivages pour déterminer le plus pertinent.

[43] Data Management Guide for SAP Business Suite (https://websmp203.sap-ag.de/ilm)

Après avoir trouvé la meilleure méthode, il faut rechercher la dépendance (voir III4.4) grâce à la fonction graphique de réseau proposée depuis la transaction SARA (voir Figure 33).

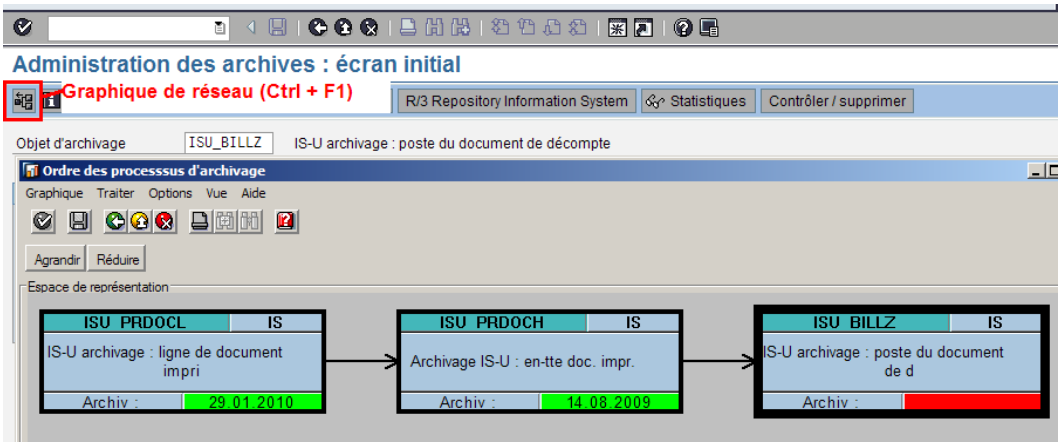


Figure 33: graphique de réseau

Pour terminer, cette analyse par l'approche ascendante,[44] il est opportun de situer les objets dans leur contexte fonctionnel, c'est-à-dire de déterminer quel est le domaine d'utilisation des transactions qui produit les tuples des tables successibles à l'archivage. Cette recherche permet de définir les interlocuteurs susceptibles de participer au projet d'archivage.

1.3.1 Correspondance tables / objets

Ci-dessous (voir Tableau iv) est résumée la liste des tables avec leur objet d'archivage correspondant. Pour certaines, dont plusieurs objets permettent de les archives, quelques explications complémentaires sont données.

Table	Description de la table	Objet	Description objet
PPOIX	Index ligne résultat du calcul de paie -> ligne d'écriture	PA_PIDX	HR : fichiers index pour enregistrement dans FI/CO
MSEG *)	Segment de document pour article	MM_MATBEL	Gestion des articles : documents article
DBERDL	Lignes document d'impression	ISU_PRDOCL	IS-U archivage : ligne de document imprimé
DBERDLB	Référence d'une ligne doct imprimé à ligne doct CalcFacturat	ISU_PRDOCL	IS-U archivage : ligne de document imprimé
ERDK	Facture / données d'en-tête	ISU_PRDOCH	Archivage IS-U : en-tête doc. impr.
ETTIFN	Val. fixes installation (normales)	ISU_FACTS	Valeurs fixes pour l'installation
EABL	Document de relevé	ISU_EABL	IS-U archivage : résultats de relevé
DBERCHZ1	Lignes individuelles de document	ISU_BILLZ	IS-U archivage : poste du document de décompte
DBERCHZ2	Lignes individuelles de document (données d'appareil)	ISU_BILLZ	IS-U archivage : poste du document de décompte

[44] Approche ascendante (ou *bottom up*) c'est-à-dire commencer l'analyse par le détail pour aller vers le général.

DBERCHZ3	Lignes individuelles de docmt (dées de montants et de prix)	ISU_BILLZ	IS-U archivage : poste du document de décompte
ERCH	Données de facture	ISU_BILL	Archivage IS-U: E/T document calcul de facturation
DFKKMOP *)	Postes pour pièce compte courant type	ISU_BBP	IS-U archivage : plan de paiement
EDI40	Document de relevé	IDOC	IDOC - Document intermédiaire
EDIDS	Enregistrement de statut (IDOC)	IDOC	IDOC - Document intermédiaire
DFKKZP	Lot de paiements : données de paiement	FI_MKKPAYL	FI-CA : lot de paiement
FKKMAZE	Historique de relance, lignes pièce	FI_MKKDUNN	FI-CA : historique de relance
DFKKOP	Postes pour la pièce compte courant	FI_MKKDOC	FI-CA : pièce
DFKKOPK	Postes pour la pièce compte courant	FI_MKKDOC	FI-CA : pièce
DFKKKO	Données d'en-tête pour pièce compte courant	FI_MKKDOC	FI-CA : pièce
BSIS	Comptabilité : index secondaire pr comptes généraux	FI_DOCUMNT	FI : pièces de la comptabilité
RFBLG	Codes source de programmes	FI_DOCUMNT	FI : pièces de la comptabilité
KNVP	Base de données clients : rôles partenaire	FI_ACCRECV	Données de base clients
KNVV	Fiche client : données ventes	FI_ACCRECV	Données de base clients
GLPCA *)	EC-PCA : postes individuels au réel	EC_PCA_ITM	Facture centre de profit : postes réels/de budget
CE41000	SIE	COPAC_1000	COPA analytique - objet résultat, périmètre résult
CE41000_ACCT	SIE	COPAC_1000	COPA analytique - objet résultat, périmètre résult
COEJ *)	Objet CO: Postes individuels sur l'année	COPA1_1000	COPA analytique, périm. résultat 1000
COEP *)	Objet CO: Postes individuels liés à la période	COPA1_1000	COPA analytique, périm. résultat 1000
COSP *)	Objet CO : totaux coûts - écritures externes	COPA1_1000	COPA analytique, périm. résultat 1000
CDCLS *)	Comptabilité : index secondaire pr comptes généraux	CHANGEDOCU	Documents de modification
CDHDR *)	En-tête de document de modification	CHANGEDOCU	Documents de modification
BALDAT	Protocole d'application : données d'un protocole	BC_SBAL	Objet d'archivage pour journal des applications
DBTABLOG	Enr. de prot. pr modifications tables	BC_DBLOGS	Archivage des modif. de la table du Customizing
BDCP	Pointeur de modification	Aucun objet	Cette table peut être purgée par la transaction BD22
D010INC	Table des utilisations des includes ABAP	Aucun objet	Dictionnaire données: aucune action à entreprendre
D010TAB	Table pour utilisation table <-> programme	Aucun objet	Dictionnaire données: aucune action à entreprendre
DFKKCIBW	FICA : extraction postes comparés	Aucun objet	-
DFKKOPBW	Extraction de postes non soldés	Aucun objet	-
DFKKZR	Formulaire de paiement	Aucun objet	-
DOKCLU	Formulaire de paiement	Aucun objet	-
DPAYP	Programme de paiement - données pour le poste payé	Aucun objet	Cette table peut être purgée par le programme RFKPYD00.
GVD_SEGSTAT	Oracle monitoring: Copy of GV$SEGSTAT	Aucun objet	-
REPOSRC	Codes source de programmes	Aucun objet	-
SOFFCONT1	SOFF : table pour contenus de documents (import/export)	Aucun objet	Cette table peut être purgée par le programme RSSORE00
SWELOG	Table pour entrées de la trace des événements	Aucun objet	-
WBCROSSGT	Index pour types globaux - cas d'emploi Workbench	Aucun objet	-
ZISU_LOG_FACT	Facture imprimée	Aucun objet	Cette table n'a pas été développée par SAP
SRRELROLES *)	Service relation objet : rôles	Aucun objet	Cette table peut être purgée par le programme RSRLDREL

Tableau iv: correspondance table / objet d'archivage

*) Concerne les tables ayant plusieurs objets.

1.3.1.1 Tables CDHDR et CDCLS : changement de document

Le document « Data Management Guide for SAP Business Suite » a été une précieuse aide pour déterminer que l'objet pertinent pour ces tables est CHANGEDOCU.

1.3.1.2 Tables COEJ, COEP et COSP

Ces tables du module SAP de comptabilité sont archivées par plusieurs objets. SAP a mis à disposition trois programmes RARCCOA1, RARCCOA2 et RARCCOA3 pour déterminer quel est le plus pertinent. L'analyse s'effectue en deux étapes : il faut d'abord exécuter le programme RARCCOA1 puis ensuite lancer RARCCOA2 ou/et RARCCOA3 pour mettre en forme l'analyse du premier.

Après l'exécution des programmes RARCCOA1 et RARCCOA2 (voir Figure 34), l'objet COPA2_1000 influence plus de 50% du volume total.

```
Système              P01

Analyse le           26.05.2010
       Extrait de données généré par programme       RARCCOA1
       Extrait de données généré le                  25.05.2010

Tables analysées :
                     COEP
                     COEJ
                     COSP
```

RANK	ObjetArch.	Tpe	PéAn	Ex.	ALL	COEP	COEJ	COSP	COSS	COST
* To	COPA2_1000				29 616 354	17 497 709	64 630	12 054 015	0	0
* To	COPA2_8000				8 623 571	5 909 012	0	2 714 559	0	0
* To	PS_PROJECT				3 640 135	2 947 844	297 656	394 635	0	0
* To	CO_COSTCTR				3 426 340	2 415 569	878 579	132 192	0	0
* To	CO_ORDER				1 216 863	953 651	208 727	54 485	0	0
* To	PM_ORDER				26 563	11 321	0	15 242	0	0
* To	CO_ITEM				223	223	0	0	0	0
* To					34	0	0	34	0	0
** T					46 550 083	29 735 329	1 449 592	15 365 162	0	0

Figure 34: résultat d'analyse par le programme RARCCOA2

1.3.1.3 Table GLPCA

Grâce à la variante ARCHIVE de la transaction TAANA, il est déterminé plus simplement quel objet choisir (voir Figure 35).

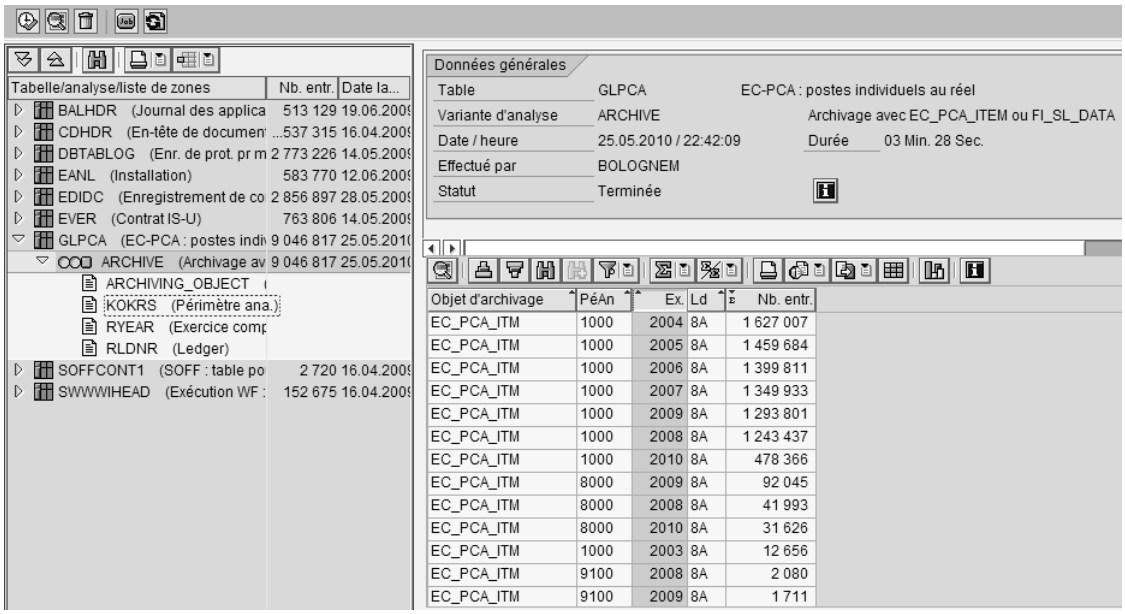


Figure 35: analyse de la table GLPCA par la transaction TAANA

*1.3.1.4 Tables D010**

Toutes les tables avec le préfixe D010, comme D010INC et D010TAB, contiennent les informations du dictionnaire des données utilisées par le fonctionnement des programmes SAP. La croissance de ces dernières demeure normale et rien ne doit être entrepris pour les archiver ou les purger.

1.3.1.5 Table SOFFCONT1

Il est possible de joindre des fichiers liés aux objets tels que des factures, articles, commandes, etc. Ces derniers sont stockés principalement dans la table SOFFCONT1, mais SAP recommande fortement d'utiliser un lien de stockage externe (c'est le principe de l'archivage des documents voir chapitre IV3.2). Cette table peut être purgée par le programme RSSORE00.

1.3.2 Domaine d'utilisation et dépendance

Après avoir trouvé les objets d'archivage correspondant aux tables volumineuses, il est important de s'intéresser aux dépendances de ces derniers (voir III4.4) et de situer l'origine fonctionnelle des données. Dans ce but, les objets sont regroupés en 6 catégories : contrôle de gestion, finance, industrie, logistique, ressources humaines et technique.

1.3.2.1 Contrôle de gestion

Par définition, le module de contrôle de gestion apporte des informations à des fins de planification, de rapport d'activité, d'analyse financière, etc.

EC_PCA_ITM - Facture centre de profit : postes réels/de budget

L'objet EC_PCA_ITM est utilisé pour archiver les lignes de calcul des centres de profit qui permettent de calculer un résultat analytique. Ils sont alimentés avec les produits dès leur comptabilisation dans la comptabilité générale, et indirectement avec les charges par l'intermédiaire des centres de coûts, ordres internes[45] ou éléments d'OTP[46], qui leur sont rattachés.

COPAC_1000 (COPA analytique - objet résultat, périmètre résult) et COPA1_1000 (COPA analytique, périm. résultat 1000)

Les objets COPAC_1000 et COPA1_1000 sont utilisés pour archiver les données des résultats du périmètre de résultat de la société 1000 (Objets de résultats CO-PA). Un périmètre analytique est une entité organisationnelle dans laquelle sont suivis les charges et les produits. Un périmètre analytique peut rassembler une ou plusieurs sociétés et un périmètre de résultat est affecté à un périmètre analytique.

1.3.2.2 Finance FI

Toutes les pièces comptables enregistrées dans les modules de comptabilité générale, comptabilité débiteurs, fournisseurs et immobilisations génèrent un enregistrement comptable dans le module FI.

FI_DOCUMNT - FI : pièces de la comptabilité

L'objet FI_DOCUMNT sert à archiver l'ensemble de ces pièces, aussi bien pour l'en-tête (code de la société, date comptable, etc.) que pour les postes (comptes généraux, TVA, objet analytique, etc.) de cette dernière.

1.3.2.3 Finance module FI-CA

Le module FI-CA (*Contract Accounts Receivable and Payable*) est un grand livre auxiliaire débiteur. Ce dernier a été conçu pour traiter et suivre un grand volume de documents débiteurs tels que factures, relances, payements, plans d'épurement, etc.

---

[45] Un ordre interne est utilisé pour assurer le suivi d'une partie des coûts et parfois des produits de l'entreprise.

[46] Un élément d'OTP (Organisation Technique Projet) permet d'imputer des coûts dans la gestion de projet.

Bien que FI-CA offre beaucoup de fonctionnalités pour améliorer le processus de gestion des débiteurs, il ne remplace pas le module de comptabilité FI-GL[47]. En fait, les écritures des débiteurs collectées par FI-CA sont agrégées et rapportées en comptabilité générale par traitement périodique.

FI_MKKDOC - FI-CA : pièce

L'objet FI_MKKDOC permet d'archiver toutes les pièces du module FI-CA de la même manière que FI_DOCUMNT fait pour les pièces de comptabilité générale.

FI_MKKDUNN

L'objet FI_MKKDUNN permet d'archiver les relances si au préalable les pièces, par l'objet FI_MKKDOC, l'ont été (voir Figure 36).

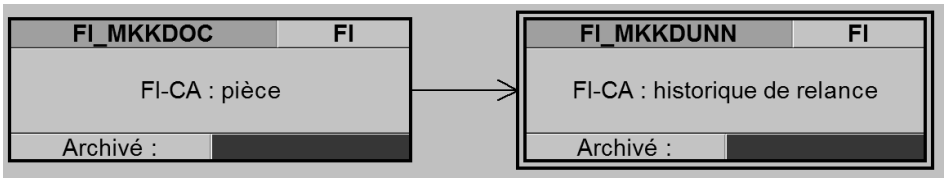


Figure 36: Dépendance objet FI_MKKDUNN

FI_MKKPAYL - FI-CA : lot de paiement

Les fichiers du trafic de payement provenant de la banque ou de la poste sont dans un premier temps stockés dans les tables telles que la DFKKZP. Ensuite, un complexe algorithme de rapprochement génère les pièces des débiteurs FI-CA.

L'objet FI_MKKPAYL archive toutes les données intermédiaires des lots de payements.

FI_ACCRECV - Données de base clients

L'objet FI_ACCRECV permet d'archiver les données de base des clients. Il est cependant dépendant de beaucoup d'autres objets (voir Figure 37).

[47] FI-GL (*General Ledger*) : est le module de comptabilité générale de SAP

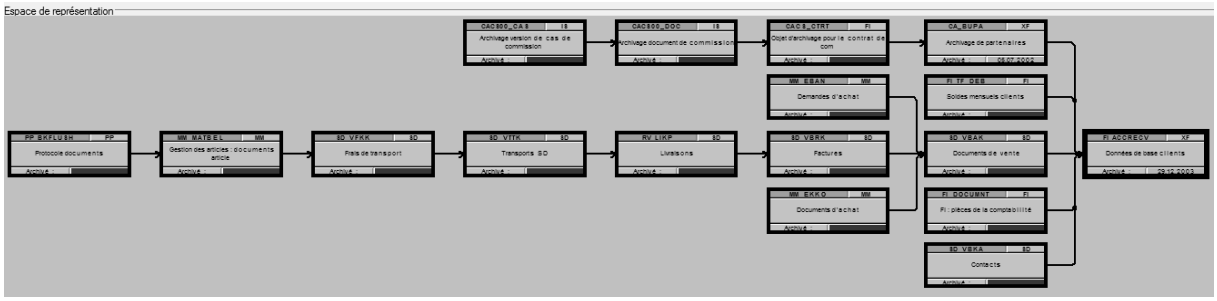


Figure 37: dépendance objet FI_ACCRECV

Compte tenu de cette forte dépendance, l'effort pour gagner quelques Ko est disproportionné par rapport au temps de réalisation. Pour cette raison et en restant cohérent au principe « petite effort pour un grand résultat », cet objet n'est pas pris en compte dans la liste des objets à archiver.

1.3.2.4 Industrie (IS-U)

Le cœur de l'activité de Romande Energie SA est la vente et la distribution d'électricité. Il n'est pas surprenant de retrouver une forte croissance des tables concernant ces activités.

ISU_PRDOCL - ligne de document imprimé

L'objet ISU_PRDOCL permet d'archiver les factures énergies appelées dans le jargon de ce module « lignes de documents imprimés ».

ISU_PRDOCH - En-tête document imprimé

L'objet ISU_PRDOCH permet d'archiver les en-têtes des factures si, au préalable, les postes l'ont été par l'ISU_PRDOCL (voir Figure 38).

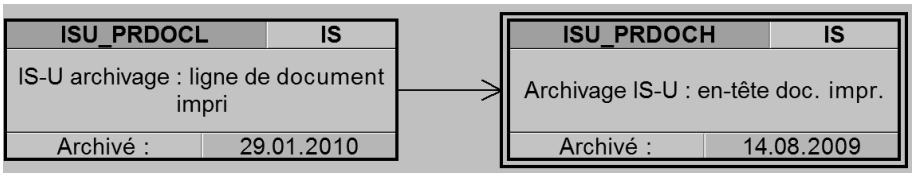


Figure 38: dépendance objet ISU_PRDOCL

ISU_BILLZ (lignes de document de calcul de facturation) et ISU_BILL (en-tête de document calcul de facturation)

Avant de pouvoir établir une facture, un décompte ou un acompte,[48] des calculs de facturation sont exécutés dans le système. Ces derniers sont stockés et peuvent être archivés pour les lignes par l'objet ISU_BILLZ et ISU_BILL pour les en-têtes. Cependant, l'archivage est possible en respectant la dépendance montrée dans la Figure 39).

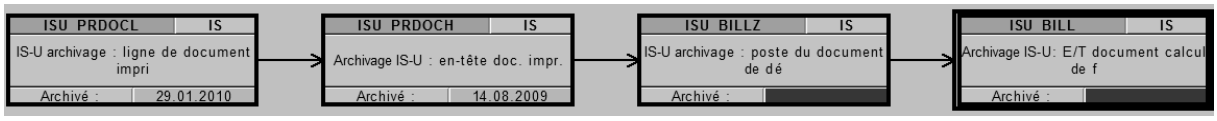


Figure 39: dépendances objet ISU_BILLZ et BILL

ISU_FACTS - Valeurs fixes pour l'installation

Lors du calcul de facturation, un algorithme très sophistiqué détermine le tarif de chaque installation.[49] Les valeurs fixes influencent la détermination de ce tarif et l'objet ISU_FACTS permet de les archiver si les dépendances sont respectées (Figure 40).

Figure 40: dépendances objet ISU_FACTS

ISU_EABL - Résultats de relevé

Périodiquement, en général annuellement, les compteurs des installations sont relevés pour pouvoir établir une facture basée sur un décompte exact. Les données concernant ces relevés peuvent être archivées par l'objet ISU_EABL si les dépendances sont respectées (voir Figure 41).

Figure 41: dépendances objet ISU_EABL

[48] Pour échantillonner le payement de la vente d'électricité, les sociétés facturent, à cadence régulière (mensuellement, bimensuellement, etc.), des acomptes. Ensuite, de manière moins régulière (généralement annuellement), les compteurs sont relevés, ce qui permet d'établir une facture basée sur un décompte exact.

[49] Une installation est un ensemble homogène de produits et services à facturer.

ISU_BBP - Plan de paiement

L'objet ISU_BBP permet d'archiver les données concernant les plans de payement si les dépendances sont respectées (voir Figure 42).

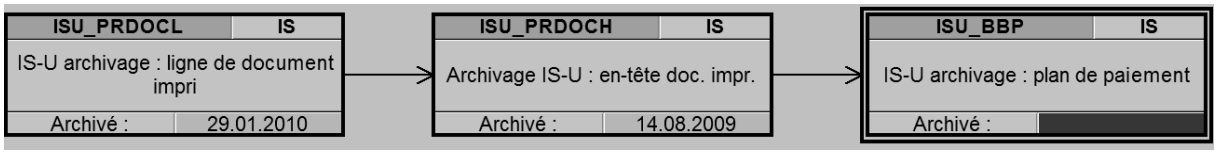


Figure 42: dépendances objet ISU_BBP

1.3.2.5 Logistique

SAP journalise tous les mouvements des articles dans les différents magasins de l'entreprise. Ces journaux peuvent être archivés grâce à l'objet MM_MATBEL (Gestion des articles : documents article).

1.3.2.6 Ressources humaines

Après le calcul de paie, des index sont créés et enregistrés dans le module FI-CO. Ces index peuvent être archivés par l'objet PA_PIDX HR (fichiers index pour enregistrement dans FI/CO) si les résultats de calcul de paie l'ont préalablement été (objet PA_CALC : HR : résultats du calcul de paie).

1.3.2.7 Objets techniques

Dans son fonctionnement, SAP mémorise plusieurs journaux consultés surtout en cas de dysfonctionnement. Ces derniers sont générés à l'attention des administrateurs SAP. Ils sont souvent inconnus par les utilisateurs, car ils n'ont pas d'impact direct sur les transactions métiers.

Normalement, ces objets sont plutôt utilisés pour faire de la purge, c'est-à-dire que le processus d'archivage s'est déroulé normalement et que les fichiers archives sont ensuite supprimés.

IDOC - Document intermédiaire

Un document intermédiaire (Idoc - *Intermediate document*) est un conteneur destiné à l'échange de données entre des systèmes SAP et tout autre système. A l'heure actuelle, le protocole d'échange le plus pertinent et le plus judicieux est XML. Malheureusement les IDOC sont tellement utilisés par les applications que pour le moment reste un grand standard de communication dans le milieu de l'éditeur allemand

CHANGEDOCU - Documents de modification

Beaucoup de documents tels que les commandes, partenaires, etc. peuvent subir des modifications. Ces dernières sont mémorisées dans des journaux et l'objet CHANDEDOCU permet de les archiver.

BC_SBAL - Objet d'archivage pour journal des applications

Lorsqu'un programme s'arrête brusquement de s'exécuter pour des raisons anormales, un journal d'erreurs est sauvegardé. Ces journaux peuvent être archivés par l'objet BC_SBAL.

BC_DBLOGS - Archivage des modifications des tables de paramétrage

Pour s'adapter aux spécificités de chaque implémentation, SAP met à disposition un immense éventail de paramètres. Les modifications de ces derniers sont journalisées et l'objet BC_DBLOGS permet de les archiver.
N.B. : parmi l'ensemble des objets techniques mentionnés, il est le seul qui ne doit pas être traité comme de la purge.

1.4 Temps de rétention des données

Ce chapitre propose un temps de rétention en ligne des données avant l'archivage. Il s'intéresse également au temps de conservation des fichiers archives selon les directives du code des obligations évoquées au chapitre II3.4 qui se traduit par une conservation de 10 ans pour l'ensemble des objets sauf ceux qui ont une vocation technique (voir Tableau v).

Objet	Description objet	Temps de résidence en ligne [jours]	Temps résidence système archivage [jours]	Volume libéré estimé [Go]
EC_PCA_ITM	Facture centre de profit : postes réels/de budget	1095	2555	2,82
COPAC_1000	COPA analytique - objet résultat, périmètre résult	1095	2555	1,04
COPA1_1000	COPA analytique, périm. résultat 1000	1095	2555	0,68
FI_MKKPAYL	FI-CA : lot de paiement	1095	2555	1,35
FI_MKKDUNN	FI-CA : historique de relance	1095	2555	0,71
FI_MKKDOC	FI-CA : pièce	1095	2555	21,34
FI_DOCUMNT	FI : pièces de la comptabilité	1095	2555	1,50
ISU_PRDOCL	IS-U archivage : ligne de document imprimé	1095	2555	10,00
ISU_PRDOCH	Archivage IS-U : en-tête doc. impr.	1095	2555	1,69
ISU_FACTS	Valeurs fixes pour l'installation	1095	2555	0,93
ISU_EABL	IS-U archivage : résultats de relevé	1095	2555	1,55
ISU_BILLZ	IS-U archivage : poste du document de décompte	1095	2555	42,77
ISU_BILL	Archivage IS-U: E/T document calcul de facturation	1095	2555	1,05
ISU_BBP	IS-U archivage : plan de paiement	1095	2555	8,81
MM_MATBEL	Gestion des articles : documents article	1095	2555	1,02
PA_PIDX	HR : fichiers index pour enregistrement dans FI/CO	1095	2555	0,72
IDOC	IDOC - Document intermédiaire	100	purge	*)
CHANGEDOCU	Documents de modification	100	purge	*)
BC_SBAL	Objet d'archivage pour journal des applications	100	purge	*)
BC_DBLOGS	Archivage des modifications des tables de paramétrage	1095	1)	0,80
TOTAL VOLUME				**103,19**

Tableau v: temps de résidence

*) Objets déjà archivés

1) Ces fichiers archives seront conservés tant que le progiciel SAP est utilisé

1.4.1 Durée de rétention en ligne

La durée de rétention en ligne permet de savoir combien de temps les données sont conservées dans la base de données avant qu'elles ne soient archivées. Ce temps est donné de manière indicative car il doit être déterminé par le responsable du domaine fonctionnel.

Grâce à cette durée, il est possible d'estimer le volume théorique que l'archivage va pouvoir libérer de la base de données.

1.4.1.1 Calcul du volume libéré

Le calcul du volume libéré peut être estimé par la formule suivante :

Volume objet = Volume_table1 + Volume_table2 + …

Volume_table[50] = Croissance_moyenne_journalière * Nombre_jours_résidence

1.5 Visualisation des archives

Pour un meilleur confort des utilisateurs, ces derniers peuvent consulter les archives généralement avec les mêmes transactions qui permettent de visualiser les données en ligne (voir Tableau vi).

Objet	Description objet	Code transaction	Même pour données en ligne?
EC_PCA_ITM	Facture centre de profit : postes réels/de budget	9KE9	Oui
COPAC_1000	COPA analytique - objet résultat, périmètre résult	KE24	Oui
COPA1_1000	COPA analytique, périm. résultat 1000		
FI_MKKPAYL	FI-CA : lot de paiement	FP05	Non
FI_MKKDUNN	FI-CA : historique de relance	FPM3	Non
FI_MKKDOC	FI-CA : pièce	FPE3	Oui
FI_DOCUMNT	FI : pièces de la comptabilité	fb03	Oui
ISU_PRDOCL	IS-U archivage : ligne de document imprimé	EA40	Oui
ISU_PRDOCH	Archivage IS-U : en-tête doc. impr.	EA40	Oui
ISU_FACTS	Valeurs fixes pour l'installation	ES32	Oui
ISU_EABL	IS-U archivage : résultats de relevé	EL31	Oui
ISU_BILLZ	IS-U archivage : poste du document de décompte	EA22	Oui
ISU_BILL	Archivage IS-U: E/T document calcul de facturation	EA22	Oui
ISU_BBP	IS-U archivage : plan de paiement	EA63	Oui
MM_MATBEL	Gestion des articles : documents article	MB51	Oui
PA_PIDX	HR : fichiers index pour enregistrement dans FI/CO	PCP0	Non

Tableau vi: consultation des archives

1.6 Equipe projet

En se basant sur l'analyse faite et en fonction des domaines fonctionnels concernés par l'archivage de données, il est désormais possible de définir l'équipe projet qui est composée

[50] Les tables prises en considération pour ce calcul sont uniquement celles qui apparaissent dans le classement.

des personnes suivantes : chef de projet, administrateur SAP, techniciens d'exploitation, utilisateur clés par domaine et juriste d'entreprise.

- Le chef de projet a bien entendu la charge de planifier, diriger et coordonner le projet. Ce rôle est celui qui m'a été imparti.
- L'administrateur SAP analyse la croissance de la base de données, établit le classement des tables les plus volumineuses, paramètre l'ArchiveLink et réorganise la base de données.
- Le technicien d'exploitation est en charge de la préparation et de la surveillance de l'exécution des programmes du processus d'archivage.
- Les utilisateurs clés des modules fonctionnels (finance, comptabilité, ressources humaines, logistique et vente d'énergie) se limitent à valider, grâce à leur expérience professionnelle, le temps de résidence dans la base de données, la formation de l'ensemble des usagers sur le moyen de consultation des fichiers archives et les autorisations pour les objets d'archivage concernant leur domaine de compétence.
- Le juriste d'entreprise valide le temps de résidence des données d'archivage dans la base de données et supervise de manière globale que les exigences légales soient respectées.

2 *Business blueprint* (conception générale)

Le *business blueprint* est, dans une certaine mesure, le cahier des charges de l'application à intégrer dans le progiciel. Ce dernier est établi en général par l'intégrateur en collaboration avec le client. Parfois, pour certains projets très conséquents et volumineux, la conception générale est accompagnée par un autre document rédigé uniquement par l'intégrateur appelé *technical blueprint*. Dans ce cas, le *business blueprint* décrit les processus, les règles de gestion des métiers, etc. alors que le *technical blueprint* précise les spécifications des programmes, les points de paramétrage, etc. En l'absence de *technical blueprint*, ces informations sont englobées dans le *business blueprint*.

Malheureusement, aucun de ces documents n'a été établi de manière officielle dans le cadre de ce projet car le client n'a jamais validé l'offre de mise en œuvre. Cependant, s'il fallait l'établir, voici sa structure :

- Définition de l'archivage (voir chapitre III1)
- Motivations (voir chapitre III2)
- Principe de l'archivage (voir chapitre III3)
- Liste des objets archives (voir chapitre IV1.3)
    - Description de leur contexte fonctionnel
    - Temps de rétention en ligne et dans le système d'archivage
    - Dépendance entre objets
- Architecture technique
- Visualisation des archives (voir chapitre IV1.5)
- Autorisations
- Planification

Ci-après seront davantage développés certains points de cette structure de *blueprint* qui n'ont pas été encore décrits.

2.1 Architecture technique

L'architecture technique, c'est-à-dire l'infrastructure physique qui va accueillir les archives, s'appuie principalement sur celle déjà existante. Elle n'est donc pas détaillée, sauf quelques spécifications concernant trois catégories: fichiers archives, purge et documents.

2.1.1 Fichiers archives

Les fichiers archives sont dans un format propriétaire qui ne peut être lu que par un environnement SAP. De ce fait, ces documents sont généralement stockés directement dans un emplacement du gestionnaire de fichiers. Pour ce projet, ils sont déposés dans le SAN[51] déjà existant dans l'infrastructure de Romande Energie SA.

Nota : lors de l'étape de suppression des données (voir III5.2.3) du processus d'archivage, il est fortement recommandé de sauvegarder au préalablement les fichiers archives.

[51] SAN (Storage Area Network) est un réseau spécialisé permettant de mutualiser des ressources de stockage. Les baies de stockage n'apparaissent pas comme des volumes partagés sur le réseau. Elles sont directement accessibles en mode bloc par le système de fichiers des serveurs. En clair, chaque serveur voit l'espace disque d'une baie SAN auquel il a accès comme son propre disque dur.

2.1.2 Purge

Le processus d'archivage pour les objets définis comme techniques (voir IV1.3.2.7) est utilisé pour faire de la purge. Dans ce but, l'emplacement des fichiers est purement temporaire car ces derniers seront supprimés après leur création.

2.1.3 Documents

Les factures et les rappels au format pdf sont archivés dans un gestionnaire de contenu, mais cette démarche sera développée plus loin. Originalement, cette gestion devait être faite par le programme Documentum[52] car il était déjà installé et opérationnel chez Romande Energie SA. Cependant, pour pouvoir s'intégrer dans ce projet, le module Documentum Content Services for SAP aurait dû être ajouté.

Mais pour les raisons exposées dans le chapitre IV3.2, le gestionnaire HTTP Content Server de SAP a été utilisé en lieu et place.

2.2 Visualisation des archives

Comme décrit dans le chapitre IV1.5, la plupart des transactions permettent de visualiser aussi bien les données en ligne que celles archivées. Par contre, les informations archivées ne peuvent pas être imprimées, contrairement à celles en ligne. Pour pallier ce problème, il a été proposé d'archiver au préalable une copie de l'impression au format pdf dans Documentum et créer un lien avec SAP par l'ArchiveLink.

Le client a manifesté un vif intérêt pour cette proposition car elle apportait également une solution à un autre problème qu'il rencontrait depuis plusieurs années : reproduire à l'identique les factures et les rappels :

- Les factures : leur mise en page est en perpétuelle mutation et, bien souvent, lors de la réimpression, la mise en page diffère quelque peu de l'original. En plus, lorsque le client a changé d'adresse, il n'est pas possible de réimprimer le document avec l'ancien domicile. En clair, dans tous les cas, le document est imprimé avec la dernière adresse connue.
- Les relances : non seulement la mise en page et l'adresse peuvent changer, mais surtout, dans la version précédente de SAP (forte heureusement ce n'est plus le cas), il était impossible de la réimprimer dans un niveau précédent. Par

[52] Documentum est un logiciel édité par la société ECM pour gérer tous les types de contenu de l'entreprise au sein d'une infrastructure d'informations communes.

exemple, si une facture avait le statut « interruption de fourniture » (troisième niveau de relance), il était impossible de réimprimer l'état « invitation au paiement » (deuxième niveau de relance).

Ces problèmes avaient de fâcheuses conséquences, particulièrement lors de la mise en contentieux. Effectivement, les juges peuvent exiger l'ensemble des factures et rappels tels qu'ils ont été envoyés au client.

Pour réaliser l'archivage des documents, il faut modifier le processus d'édition. En fait, après le calcul de facturation et la création des documents d'impression, un fichier d'échange est envoyé de SAP vers le logiciel JetForm de la société Cisel SA. JetForm est un logiciel d'éditique qui permet de faire la mise en page de documents et de piloter des imprimantes et des machines de mise sous pli (voir Figure 43).

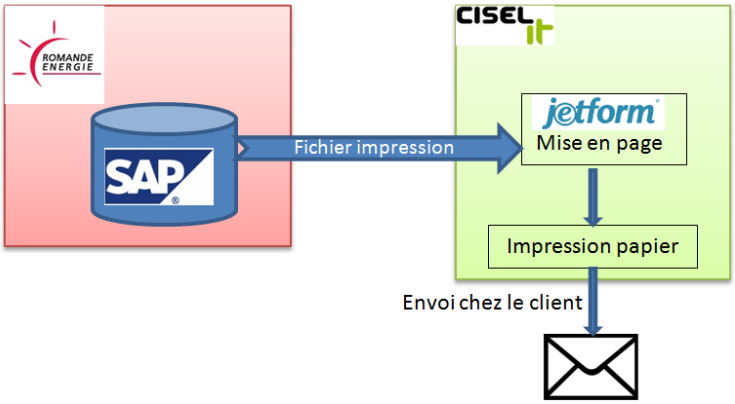


Figure 43: processus d'impression avant modification

Ce processus a dû être modifié par la société Cisel SA pour que lors de la génération du fichier de spool, un fichier pdf soit créé pour chaque document et stocké dans un répertoire temporaire. Un automate dans le serveur de gestion de contenu Documentum lit les fichiers, les archive et les supprime du répertoire temporaire. Ensuite, un autre programme crée une référence du fichier archivé dans SAP afin que ce dernier puisse être lu comme pièce jointe depuis les transactions standards du document (voir Figure 44).

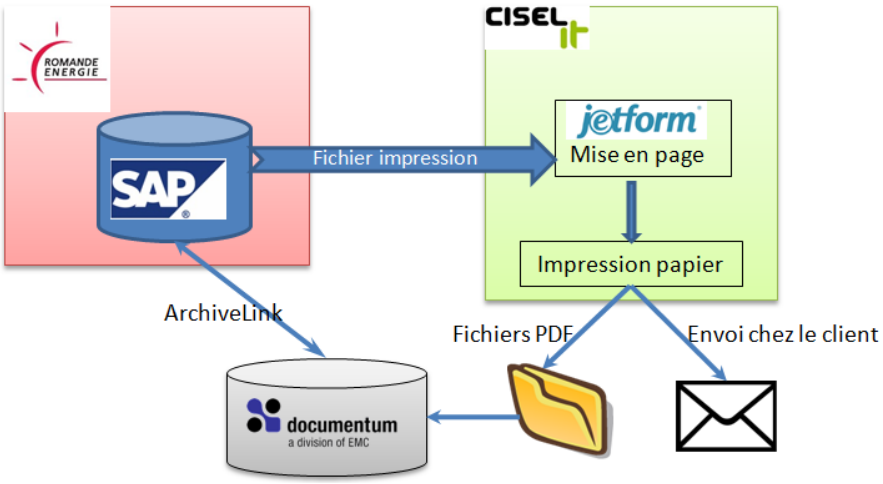


Figure 44: processus de facturation après modification

2.2.1 Spécification de nommage du fichier pdf

Le nom de fichier pdf créé par JetForm doit respecter la nomenclature suivante :

mandant_système_CodeTypeDocument_noDocument_TimeStamp.pdf

- **Mandant** : se réfère au code mandant SAP, par exemple 100 pour la société Romande Energie du système productif.
- **Système** : se réfère au système SAP, par exemple P01 pour le système productif.
- **Code type document** :
  - **F** : pour les facture ISU
  - **R** : pour les relances
- **Numéro du document** : numéro de facture
- ***Time stamp*** : la date, l'heure, les minutes et les secondes de la création du document.

2.2.2 Protection des fichiers pdf

Les fichiers pdf doivent être protégés de la modification, de la copie du contenu, etc., mais ils doivent pouvoir être imprimables (voir Figure 45).

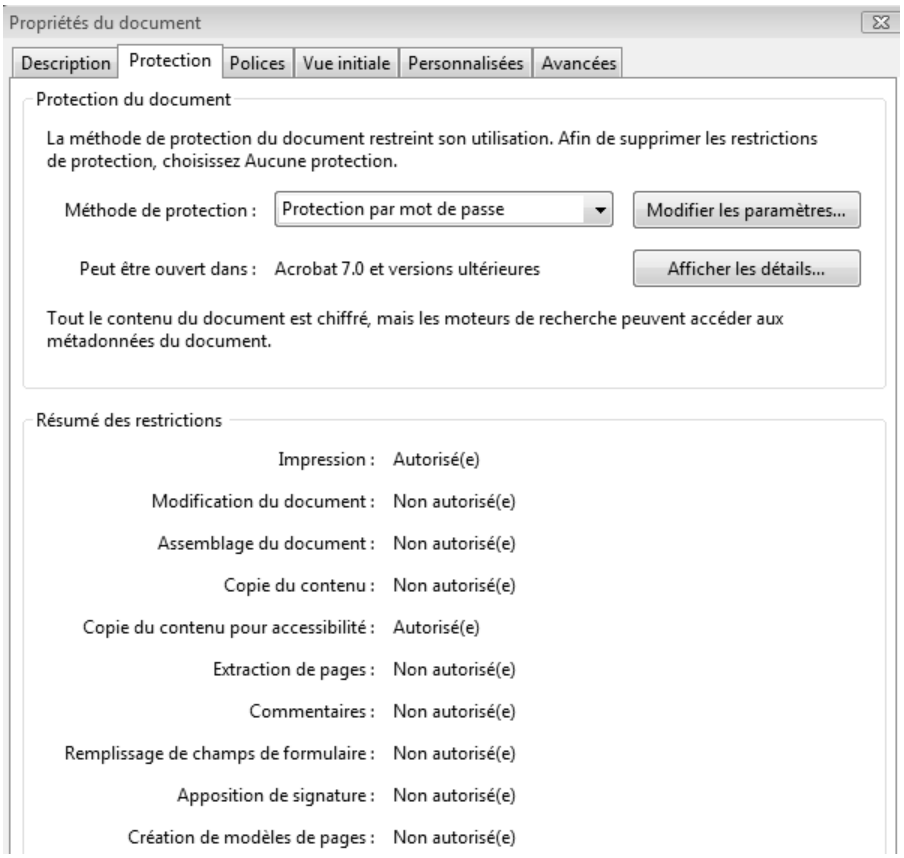


Figure 45: protection des fichiers pdf

2.2.3 Spécification pour des métadonnées des fichiers pdf

Les fichiers pdf doivent être enrichis d'un certain nombre de métadonnées qui se trouvent dans les documents de factures et relances :

- Système source (pour rappel, c'est le système SAP d'origine du document)
- Mandant (pour rappel, c'est une unité indépendante d'un point de vue commercial, organisationnel et technique dans un système SAP)
- N° du compte de contrat
- N° du client
- N° de la facture
- Date comptable de la facture
- Montant à payer ou dû

Grâce à ces métadonnées, un automate à l'intérieur de Documentum est en mesure de classer et indexer les fichiers de sorte que ces derniers puissent être consultés depuis une interface fournie par ce logiciel.

2.2.4 Spécifications du programme d'intégration des fichiers pdf dans Documentum

La société Sword SA intégratrice de la solution Documentum chez Romande Energie SA est en charge d'adapter la structure et de développer un programme d'intégration des fichiers pdf dans ce gestionnaire de contenu. Cette recherche mentionne uniquement les grandes lignes et les propriétés principales, car les spécifications exactes et détaillées doivent être fournies par la société Sword SA.

2.2.4.1 Structure physique

La structure physique du plan de classement des factures et des relances doit être la selon la Figure 46 ci-dessous :

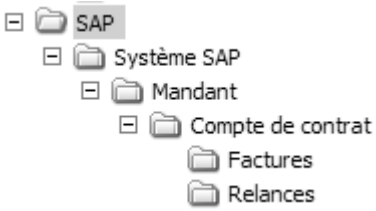


Figure 46: structure du plan de classement

2.2.4.2 Programme d'intégration

Le programme d'intégration devra prendre en considération les points suivants :

- Lire les fichiers stockés dans le répertoire temporaire.
- Contrôler les fichiers :
  - Contrôle si le fichier pdf n'est pas corrompu,
  - Contrôle des doublons,
  - Contrôle si le type de document est reconnu.

  Si une de ces erreurs est détectée, cette dernière doit être notée dans un journal et le fichier doit être déplacé dans un répertoire dédié aux erreurs.
- Classer le fichier selon son nommage et lire les métadonnées.
- Enrichir les caractéristiques par la lecture des métadonnées du fichier. Ces dernières permettent d'indexer le document dans Documentum afin qu'il puisse être consulté facilement par les utilisateurs.
- Définir la date de conservation en fonction de la date comptable qui se trouve dans les métadonnées : date comptable + 10 ans.
- Supprimer les fichiers du répertoire après intégration.

2.2.5 Lien entre la facture SAP et Documentum

Une fois que les fichiers sont intégrés dans Documentum, il faut créer un lien pour que les utilisateurs puissent les consulter depuis la transaction de visualisation des documents imprimés.

Pour ce faire, il faut créer une communication entre SAP et Documentum par le paramétrage de l'ArchiveLink explicité dans le chapitre III6 ArchiveLink. En plus, il faut développer un service web[53] afin qu'un automate créé par Sword puisse l'utiliser pour lier le fichier avec le document imprimé.

2.2.5.1 Service web pour la création du lien fichier – document imprimé

- Paramètre d'entrée :
  - Numéro de facture
  - Type de document : F pour facture et R pour relance
  - Numéro d'indentification unique : ce numéro alphanumérique est l'identifiant unique dans Documentum d'un fichier
- Retour :
  - 0 : traitement terminé avec succès
  - 4 : doublon
  - 8 : document inexistant
  - D'autres messages de retours sont supportés directement par les couches d'échange
- Contrôles :
  - Contrôle si le numéro de facture existe et, si ce n'est pas le cas, retourner 8.
  - Contrôle des doublons par la combinaison du numéro de facture et du numéro d'identifiant unique, si c'est le cas renvoyer 4.
- Création :
  - Création du lien dans la table des liens TOA01 (voir chapitre III6.3.3 Table TOA01 pour plus de détails).

[53] Le service web est une technologie permettant à des applications de dialoguer à distance via Internet, et ceci indépendamment des plates-formes et des langages sur lesquels elles reposent.

2.3 Autorisations

Comme déjà mentionné à plusieurs reprises, la consultation des fichiers archives ainsi que des documents liés se fait depuis la transaction standard de visualisation des données en ligne. Des droits d'accès sont déjà associés à cette dernière, mais si des restrictions supplémentaires sont nécessaires à la consultation des archives ou des documents liés, cela doit être spécifié par le client.

A noter par contre que la consultation des fichiers pdf depuis Documentum requiert la mise en place des autorisations.

3 Réalisation

Selon la méthode ASAP, la réalisation concerne le paramétrage, le développement et les tests unitaires dans le système de développement (voir chapitre I2.3.7.4).

Comme le client a uniquement validé la mise en place de l'archivage des documents pour les factures et les relances, la direction de neo technologies SA a décidé de mettre en place gracieusement l'archivage des données pour les objets techniques.

Il faut savoir que Romande Energies SA a signé, avec neo technologies SA, un contrat de niveau de service, plus connu sur la terminologie anglaise SLA[54]. Bien que ce dernier ne prévoie pas de prestation concernant l'archivage des données, il a été admis que les objets techniques peuvent faire partie de l'administration courante.

3.1 Archivage des données

La réalisation de l'archivage comprend le paramétrage global, le paramétrage spécifique, des traitements particuliers et un test de déroulement du processus pour chaque objet technique : document intermédiaire, document de modification, journal des applications et journal des protocoles des workflows[55].

---

[54] Le SLA (Service Level Agreement) est un document qui définit la qualité de service requise entre un prestataire et un client.

[55] Le workflow est une automatisation d'un processus (partiel ou complet), au cours duquel des documents, des informations et des tâches passent d'un participant à un autre, au sein d'un groupe de travail, en conformité avec un ensemble de règles prédéfinies. Un système de workflow définit, crée et gère l'exécution de tels processus (voir http://www.alaide.com/dico.php?q=Workflow).

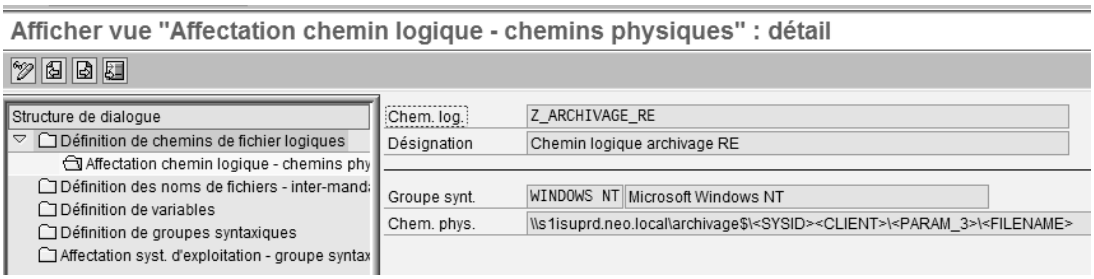
Certains ont pu remarquer que les protocoles des workflows n'ont pas été définis lors de l'analyse de Pareto (voir chapitre 1.2). Cependant, lors de l'analyse effectuée en 2009, ces derniers faisaient partie des objets dont les tables étaient parmi les plus volumineuses. Ensuite, après la purge de septembre 2009, la croissance de ces tables n'a pas été aussi importante.

3.1.1 Paramétrage global

Les paramétrages globaux sont l'ensemble des paramètres communs à tous les objets d'archivages mentionnés au chapitre III5.1.1.1. En réalité, aucun de ces paramètres n'a été modifié car les options standard conviennent.

Par contre, un fichier logique commun à tous les objets a été défini (voir Annexe 2 pour plus de détails) :

- Chemin logique :



- Définition du nom de fichier :

3.1.2 Paramétrage spécifique

Pour l'ensemble des objets techniques seront utilisés les mêmes paramètres gérés par la transaction SM30 vue V_ARC_USR. Ces derniers sont (encadrés en rouge dans la copie d'écran ci-dessous voir Figure 47) : le nom du fichier logique, la définition de la taille des

fichiers archives, la variante d'exécution en mode productif et le lancement automatique de la suppression.

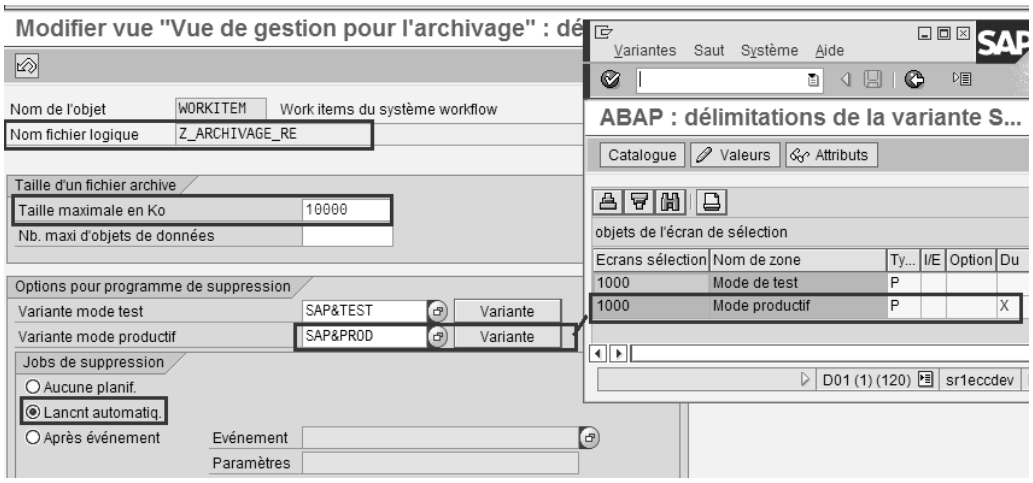


Figure 47: paramétrage spécifique

3.1.3 Traitement particulier

Associés au processus d'archivage, certains objets tels que l'IDOC bénéficient des traitements particuliers. Ces derniers servent à inhiber certains contrôles de cohérences typiques du programme de création des fichiers archives ou à compléter le processus.

3.1.3.1 Document intermédiaire (IDOC)

Dans la génération des documents intermédiaires, un certain nombre de liens sont créés et mémorisés dans la table SRRELROLES. Lorsque les IDOCs sont supprimés, ces liens n'ont plus d'utilité et le programme RSRLDREL les supprime.

Dans le contexte de ce projet, ce programme doit être exécuté pour le type de relation IDC8 (voir Figure 48) et IDCA.

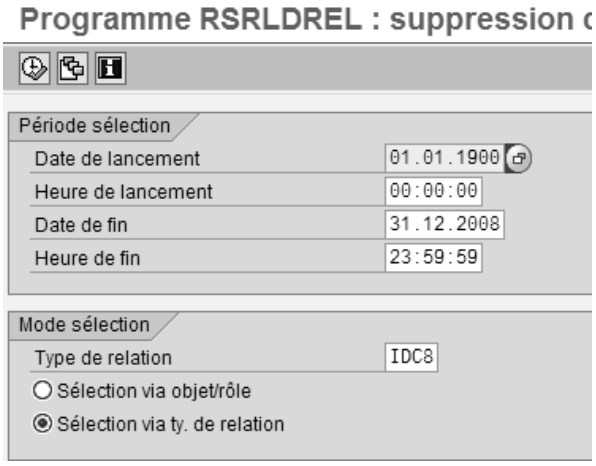


Figure 48: suppression des relations IDC8

3.1.4 Tests

Les tests dans l'environnement de développement ont été effectués par mes soins sur un petit échantillon de données. Ces derniers, bien que sommaires, ont permis de vérifier que le processus est exact et les programmes s'exécutent correctement.

NB : des tests plus approfondis sont effectués dans le système de qualité (voir I2.3.7.4).

3.2 Archivage des documents

Avant d'entrer dans le vif du sujet concernant l'archivage des documents (voir IV2.2), il est important de relever une substantielle modification qui est survenue avant la réalisation. Le client, pour des raisons de rapidité de mise en œuvre, a souhaité que les documents soient intégrés dans le gestionnaire de contenu HTTP Content Server de SAP (voir l'Annexe 3) au lieu de Documentum, et ce n'est que dans un deuxième temps qu'ils migreront du HTTP Content Server vers Documentum comme défini dans le cahier des charges initial (voir IV2.2).

Ce changement a entrainé quelques modifications dans le processus de facturation prévu. Pour la première partie, c'est-à-dire la création du fichier pdf, le stockage de ces derniers dans un répertoire partagé et le paramétrage de l'ArchiveLink, il n'y a pas eu de changements. Par contre, toute la partie d'intégration du pdf dans le Content Server et le lien avec SAP a dû être adaptée par la création d'un programme spécifique pour gérer les fichiers, les intégrer dans le serveur de contenu et faire le lien avec SAP (voir la représentation de la Figure 49).

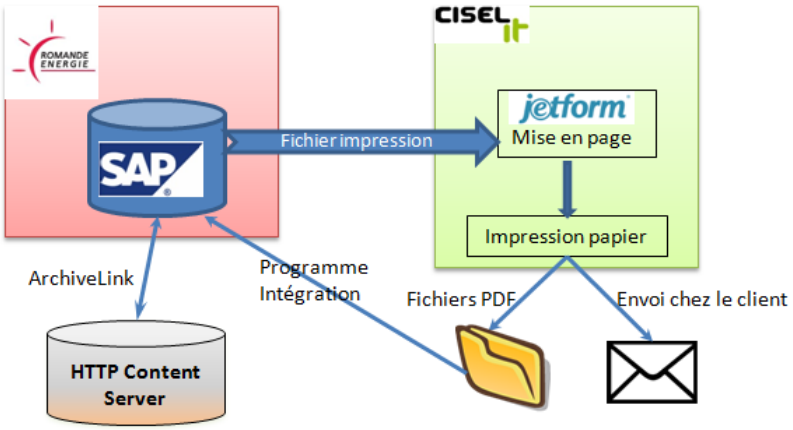


Figure 49: intégration des pdf dans le HTTP Content Server

3.2.1 Création des fichiers pdf

La société Cisel SA s'est chargée d'ajouter à leur processus interne la mise en page des documents, la création des fichiers pdf et leur sauvegarde dans un répertoire partagé.

3.2.2 Réalisation du paramétrage de l'ArchiveLink

La première partie de la réalisation concerne le paramétrage de l'ArchiveLink. La suite de sous-chapitre explicite pas à pas les étapes de ce paramétrage.

1. **Création du référentiel de contenu (transaction OAA3) :**

   Cette transaction permet de définir les protocoles pour piloter, en fonction du type de document, la communication entre le système SAP et les gestionnaires de contenu. La Figure 50 montre les paramètres saisis.

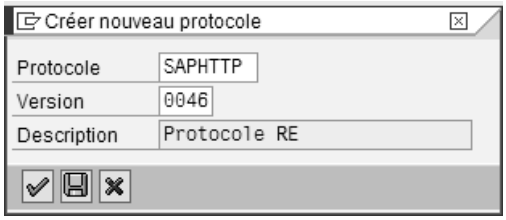


Figure 50: référentiel de contenu

2. **Création d'un repository (transaction OAC0)**

   Comme expliqué dans le chapitre III6.3.2, (voir Figure 51) définit le moyen de communiquer avec le serveur de contenu. Physiquement, ce dernier correspond à un répertoire (voir Figure 62).

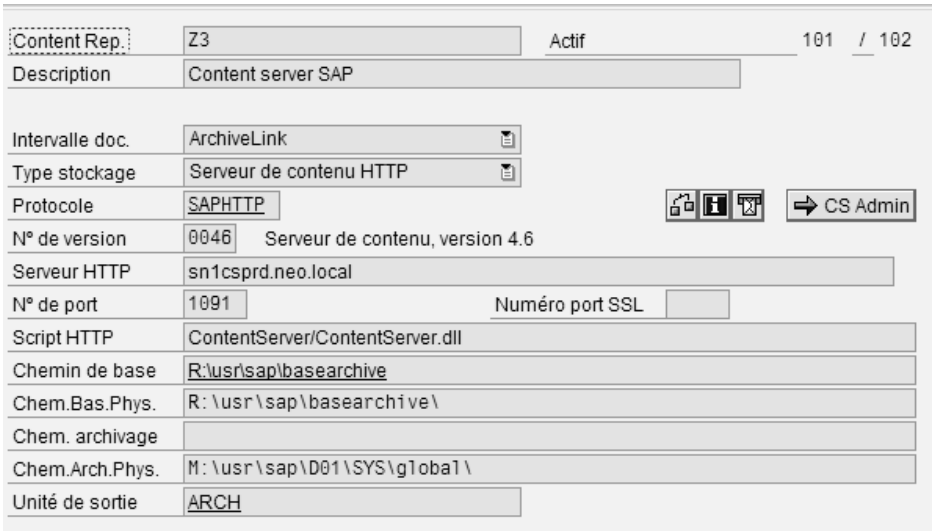


Figure 51: paramétrage du content repositorie

   Nota : lors de la mise en place de Documentum, il faudra changer uniquement les paramètres : Serveur HTTP, N° de port et Script HTTP.

3. **Création des types de document (transaction SM30 vue TOAVE)**

Le type de document permet de regrouper tous les documents du même type par rapport à un objet métier. Pour ce projet, il a été créé le type ISU_BILL pour les factures et ISU_DUNN pour les relances (voir Figure 52).

Afficher vue "Tous types de documents" : synthèse

Tous types de documents

Type de do...	Désignation complète	Format de document
ISU_BILL	Facture IS-U	PDF
ISU_CASH	Reçu IS-U	PDF
ISU_CHG_CA	IS-U : écriture de modification de cpte	PDF
ISU_DUNN	Lettre de relance IS-U	PDF

Figure 52: type de document

4. **Création d'une table des liens (transaction SE11)**

La table des liens est le pivot entre l'écran de visualisation des documents imprimés et les fichiers pdf stockés dans le HTTP Content Server. La création de cette dernière n'est pas indispensable car il est possible d'utiliser celle fournie en standard. Cependant, dans le cadre de ce projet, je préconise d'en créer une nouvelle tous les trois ans pour maîtriser le nombre d'enregistrements et préserver les temps d'accès. La création de cette nouvelle table est faite par copie de la TAO01 (voir **Figure 53**).

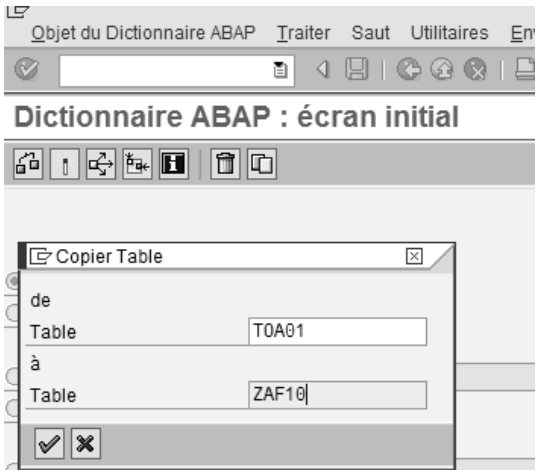


Figure 53: Création de la table des liens

5. **Enregistrement la table des liens (transaction SM30 vue V_TOACO)**

Après la création de la table des liens, il faut l'enregistrer pour qu'elle soit reconnue dans le traitement des liens.

6. **Traitement des liens (transaction OAC3)**

Le paramétrage de traitement des liens met en relation (voir Figure 54) l'objet métier (PRINTDOC : document d'impression), le type de document (voir point 3), le *content repository* (voir point 2) et la table des liens (voir points 4 et 5).

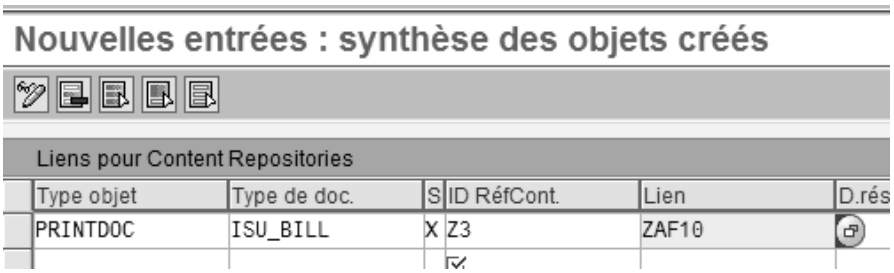


Figure 54: traitement des liens

NOTA : Le X du 3ème champ « S » signifie que le référentiel est actif, c'est-à-dire que c'est dans ce dernier que les données vont être stockées.

3.2.3 Réalisation du programme d'intégration

Le programme ZIP_ISU_LOAD_FACTURE_PDF (le code source se trouve dans l'annexe 4) permet de lire les fichiers stockés dans le répertoire partagé, de contrôler la cohérence des ces derniers, de poster les fichiers dans le serveur de contenu et de créer les liens dans SAP.

3.2.3.1 Format du fichier attendu

<NuméroDeFacture><Date>.pdf

a. <NuméroDeFacture> : 12 positions numériques

b. <Date > : format anglais -> AAAAMMJJ

Exemple fichier: 00100045667020060528.pdf

3.2.3.2 Paramètres

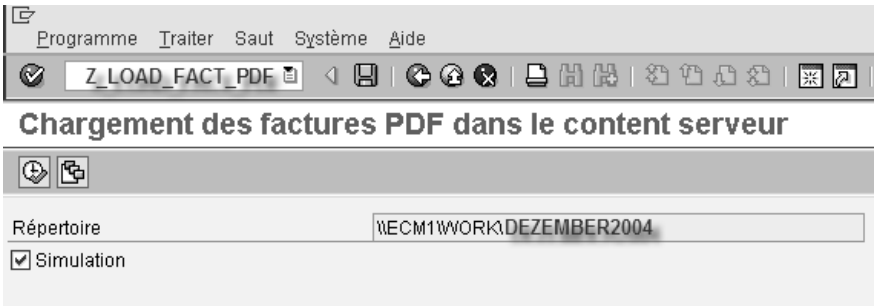


Figure 55: paramètres du programme ZP_ISU_LOAD_FACTURE_PDF

Répertoire : saisir le chemin complet du répertoire contenant les fichiers à intégrer.

Simulation : si la coche est active, les fichiers ne sont pas intégrés. Cependant, les contrôles suivants sont effectués :

a. Erreur si le fichier n'a pas l'extension pdf

b. Erreur si le numéro de facture n'est pas au bon format

c. Erreur si la date n'est pas au bon format

d. Erreur si le fichier existe déjà dans le Content Server

3.2.3.3 Messages d'erreurs

Message d'erreur	**Explication**
Numéro de facture incohérent !	Le test est fait sur la base du nom du fichier : à savoir les douze premières positions doivent être numériques.
Date incorrecte !	Le test est fait sur la base du nom du fichier : à savoir les huit dernières positions avant le '.'.
L'extension du fichier doit être pdf !	Test si l'extension du fichier est pdf
Fichier déjà existant dans le content serveur !	Message d'erreur si le numéro de facture et la date (20 digits du nom du fichier) existent déjà dans la table des liens (ZAI08 ou ZAS08).
Erreur lorsque le fichier est entreposé dans CS !	Il y un problème dans le Content Server, car SAP n'as pas pu déposer le fichier.
Erreur création lien Archivelink !	Erreur de communication entre SAP et CS.

4 Préparation finale

Selon la méthode ASAP, la préparation finale comporte tout d'abord le transport du paramétrage et des programmes dans le système de qualité. Des tests approfondis sont ensuite effectués pour s'assurer que lorsque le processus sera installé dans le système productif, il n'y aura pas de surprises majeures. Enfin, l'équipe de support est formée pour assurer la future maintenance dans l'environnement productif, alors que les utilisateurs le sont pour la consultation des documents archivés.

4.1 Test archivage des données

Le Tableau vii ci-dessous résume les tests réalisés dans cet environnement, qui se sont déroulés sans aucun problème. Cela témoigne que cette configuration peut être transportée et exécutée dans le système productif.

Objet	Libellé	Nb objets supprimés	Temps de traitement [h]	Espace disque total [Mo]
IDOC	Document intermédiaire	1 093 219	6.20	2692,920
CHANGEDOCU	Documents de modification	365 051	6.66	1997,262
BC_SBAL	Journal des applications	185 118	0.25	457,679
WORKITEM	Journal des protocoles des Workflows	64 545	0.42	355,524
TOTAL				5503,38

Tableau vii: résultat après l'archivage de test

4.2 Tests archivage des documents

Concernant l'archivage des documents, des tests d'intégration de plusieurs milliers de fichiers pdf ont été intégrés pour tester d'une part la stabilité du programme ZP_ISU_LOAD_FACTURE_PDF et, d'autre part, le temps d'exécution. Ces tests ont porté entière satisfaction.

D'autre part le client a également fait des tests en ouvrant plusieurs de ces fichiers par les transactions standard SAP, comme le témoigne la copie d'écran ci-dessous (voir Figure 56).

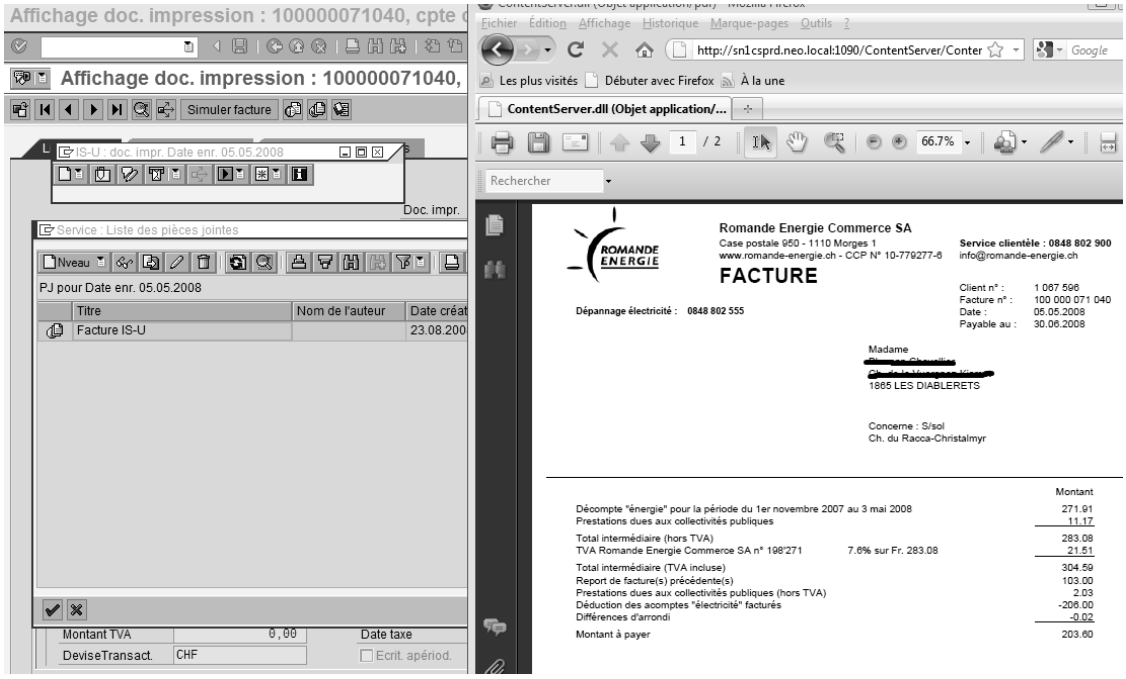


Figure 56: exemple de visualisation d'une facture pdf

5 *Go live* et support

Selon la méthodologie ASAP, cette dernière étape marque la fin du projet, le début de la vie de l'archivage (d'où le terme anglais *go live*) et le début du support.

Concrètement, les paramétrages et tous les programmes sont transportés dans l'environnement productif. Les processus d'archivages sont exécutés pour la première fois en production et les premières factures pdf sont intégrées dans le serveur de contenu.

5.1 Résultat après le premier cycle d'archivage des données

Le premier cycle d'archivage a eu lieu en septembre 2009. Le Tableau ii du chapitre IV1.1.1 montre que le gain d'espace a été de 189 Go, presque un tiers du volume total. Cet important gain est dû en petite partie à l'archivage (voir Tableau viii), mais surtout grâce à une réorganisation complète de la base de données[56].

Objet	**Libellé**	**Nb objets supprimés**	**Temps de traitement [h]**	**Espace disque total [Mo]**
IDOC	Document intermédiaire	1 401 415	4.30	3452,094
CHANGEDOCU	Documents de modification	502 784	4.26	2751,835
BC_SBAL	Journal des applications	3 341 695	6.13	7423,300
WORKITEM	Journal des protocoles des Workflows	1 322 137	15.22	3456,819
TOTAL				17084.05

Tableau viii: résultat après le premier archivage productif

La périodicité de cycles est définie en fonction du volume et de la courbe de croissance. En clair, plus le volume est important et/ou la croissance est rapide, plus les cycles doivent être rapprochés. Dans le cas de ce projet, il a été décidé de faire un cycle tous les ans. Le prochain aura donc lieu en septembre 2010.

[56] Il faut savoir que cette opération (la réorganisation) n'avait jamais été entreprise depuis l'installation, donc environ 7 ans.

5.2 Résultat après l'archivage des documents

L'archivage des documents a débuté en octobre 2007. A l'heure actuelle, il y a plus de 4 millions de documents, entre les factures et les relances, qui ont été intégrés dans le serveur de contenu.

Le traitement qui permet d'intégrer ces documents est fait de manière automatique tous les soirs et un technicien d'exploitation contrôle tous les matins le résultat de cette exécution. En clair à chaque impression un fichier pdf est automatiquement généré et le programme ZP_ISU_LOAD_FACTURE_PDF est planifié tous les soirs pour intégrer l'ensemble des documents créés dans la journée.

5.2.1 Reprise

Concernant les factures émises avant octobre 2007, Romande Energies a décidé de ne pas entreprendre un projet de reprise pour plusieurs raisons :

1. Le temps de réimpression des factures est estimé à environ 3 mois de travail machine, sans compter le temps de préparation et d'intégration.
2. Compte tenu de cet important temps de traitement il est utopique d'exécuter la reprise dans le système productif. Il a été étudié l'éventualité de créer un système parallèle issu d'une copie de celui productif, mais cette alternative porte préjudice sur l'intégrité des fichiers générés.
3. La mise en page des fichiers ne pourra pas être identique à l'original

En résumé l'effort et les coûts de reprise ne la justifie pas.

5.2.2 Gestion du cycle de vie

Le gestionnaire de contenu HTTP Content Server de SAP ne prévoit pas la gestion du cycle de vie des documents. En clair, lorsque les factures et les rappels auront 10 ans, aucune fonction existante ne permettra de les détruire.

Dans ce but, il faudra soit migrer les fichiers pdf dans le gestionnaire de contenu Documentum comme cela a été prévu, soit développer des programmes spécifiques pour ajouter cette fonctionnalité. Le client a environ 7 ans pour se décider.

V Conclusion

En conclusion de ce rapport, je souhaite faire un bilan de ce travail, partager quelques considérations personnelles concernant l'outils SAP et enfin mentionner certains projets analogues qui m'ont permis d'approfondir et consolider les connaissances acquises.

1 Bilan du projet

Le Tableau ix ci-dessous établit un bilan en fonction des objectifs et des attentes de ce travail de diplôme d'ingénieur (voir chapitre I3).

Objectif	Bilan
Découvrir le milieu de l'archivage électronique, son langage, ses pratiques et son application de manière générale.	**Réalisé** : la découverte s'est concrétisée par l'étude de plusieurs ressources mises en annexe et par la participation à plusieurs séminaires.
Décrire les outils propres au système SAP dans le domaine de l'archivage de données, ainsi qu'un système d'archivage pour le stockage des fichiers d'archive.	**Réalisé partiellement** : du point de vue du projet, aucun document n'a été remis au client ; cependant, la partie III Principe de l'archivage SAP de ce rapport de mémoire peut être utilisée pour faire cette description.
Cibler les objets d'archivage potentiels par une analyse des besoins techniques et fonctionnels tels que la volumétrie des données et les problèmes de performances.	**Réalisé** : plusieurs analyses ont été faites pour cibler les objets d'archivage au cours de ces années. La dernière en date est celle décrite dans le présent rapport.
Définir les besoins en termes de rétention de données et les contraintes légales d'archivage de données.	**Réalisé partiellement** : en ce qui concerne la définition de la rétention des fichiers archives selon les exigences légales, cela peut se résumer, dans le contexte, à dix ans. Cependant, la définition du temps de rétention en ligne des données non techniques devra être définie ultérieurement au cours du projet.
Présenter une structure de projet complète avec les intervenants ainsi que les responsabilités dans un projet d'archivage SAP.	**Réalisé partiellement** : la définition des profils a été décrite dans le chapitre IV1.6. Cependant, l'attribution devra être faite lors du lancement du projet.

Réaliser dans un premier temps une maquette dans le système de qualité pour valider les concepts théoriques étudiés. Réaliser ensuite la mise en œuvre de l'archivage dans le système productif en orchestrant tous les intervenants de la structure du projet présenté.	Pour ce qui concerne la maquette elle a été **réalisée** il y a 4 ans dans le système de qualité. Cependant elle a dû être détruite, car tous les ans ce système est réinstallé par une copie de celui productif. Concernant la mise en œuvre dans le système productif, seul l'archivage des objets techniques a été implémenté et de ce fait il peut être considéré comme **partiellement réalisé.**
Planifier l'exécution périodique des lots de programmes d'archivage et l'archivage à long terme.	**Partiellement réalisé** : la planification de l'exécution périodique des objets techniques a été définie par une exécution par an. Par contre, pour les autres objets, cette planification devra être définie.

Tableau ix: bilan du projet

En résumé, l'ensemble des objectifs concernant la réalisation du projet a été partiellement atteint car le client, pour des raisons que lui sont propres, n'a toujours pas validé la mise en place de l'archivage des données pour tous les objets non techniques. Dans une certaine mesure, je comprends sa position : alors que SAP préconise l'archivage pour améliorer le temps de réponse des transactions, la réduction des coûts d'exploitation et une diminution d'indisponibilité en cas de problème, certains de ces points ne se sont pas vérifiés :

- La dégradation du temps de réponse des transactions est un levier très significatif pour que les utilisateurs acceptent un projet d'archivage. Malheureusement ou heureusement, cela dépend du point de vue, au cours de ces dernières années, le temps de réponse moyen est resté pratiquement constant se situant entre 300 et 600 ms.
- Bien qu'il soit assez difficile à mesurer, il n'y a pas eu une conséquente augmentation de temps d'exploitation dû à la croissance de la base de données. Cependant, il a été relevé que le temps de sauvegarde, il y a 5 ans, était d'environ 5 heures pour une base de données de 350 Go. A l'heure actuelle, pour une base de données de 600 Go, cela prend théoriquement 7 heures. Ce temps est divisé par deux car le processus d'écriture sur bande magnétique a été parallélisé, c'est-à-dire que l'écriture de la copie de sauvegarde est faite sur deux bandes simultanément.
- Le temps d'indisponibilité en cas de problèmes est tributaire, d'une certaine mesure, du système de sauvegarde. En effet, le temps de restaurer la base de

données est pratiquement identique à celui de sauvegarde. Comme vu dans le point précédent, grâce à la parallélisation sur deux bandes, le temps s'est amélioré.

Cet ensemble de constatations tend à prouver que dans le contexte de ce projet, les préconisations de l'éditeur allemand confortent la position du client. Cependant, il faudra tout même archiver les données pour purger l'ensemble des objets dont les contraintes légales n'obligent plus la conservation, c'est-à-dire à partir de la onzième année. Cela n'est pas une contrainte clairement exprimée dans le code des obligations, mais selon les dires de Maître Kronbichler,[57] lors d'un procès, le juge va exiger des justificatifs à l'entité si elle conserve des documents au-delà du temps légal.

2 Considérations personnelles

En conclusion, dans l'ensemble des environnements que j'ai pu analyser, l'archivage des données reste utile surtout pour la purge. Par contre, l'archivage des documents du point de vue de SAP reste un atout très intéressant pour lier des fichiers à des transactions. Toutefois, j'estime que l'HTTP Content Server de SAP n'est pas adapté pour cette fonctionnalité car il n'est pas utilisable dans un autre environnement que celui de SAP, il ne gère pas le cycle de vie du document et surtout il n'intègre pas des fonctionnalités d'archivage légal.

Un bon remplaçant de ce gestionnaire de contenu pourrait être Alfresco. Ce dernier, bien que Open Source, a toutes les fonctionnalités essentielles d'un progiciel commercial tel que Documentum (d'ailleurs il a été conçu par une équipe de ce dernier) ou OpenText. Dans ce but, je souhaite proposer à l'ensemble de la direction de neo technologies SA, dont je fais désormais partie, une solution clé en main. Cette dernière serait ciblée pour des petites et moyennes entreprises du secteur public et de la distribution et la vente d'énergie qui souhaitent gérer des documents autour de SAP. La solution devrait contenir un ensemble de fonctionnalités de base telles que le connecteur ArchiveLink, des paramètres et programmes spécifiques pour SAP ainsi que des programmes, un plan de classement, etc. pour Alfresco.

Désormais ce projet-ci est de la douce musique d'avenir et seul le futur en connaît le refrain…

[57] Monsieur Stephan Kronbichler, avocat du cabinet Kronbichler & Tourette, spécialiste dans les questions du droit commercial et de la tenue et la conservation des documents.

3 Autres projets

L'étude et la formation de l'environnement de l'archivage et de la gestion des documents ont donné le jour à d'autres projets. Certains sont en extension directe comme celui de PayNet, alors que d'autres sont tout simplement de nouvelles implémentations comme l'archivage de données chez Alcan, l'archivage des documents chez ESB et la gestion des documents aux services industriels de Lausanne.

3.1 PayNet

Romande Energies SA a souhaité ajouter les services offert par la société PayNet[58] à la solution d'archivage des factures. Cela permet à ses clients de payer leur consommation d'électricité via le portail de la banque virtuelle.

Dans une certaine mesure, PayNet est une boîte aux lettres électronique annexe aux services du portail de la banque virtuelle. Dans cette dernière est géré l'ensemble des factures dématérialisées que le fournisseur a envoyé à son client. Le client peut en quelques clicks de souris payer et gérer les factures reçues par ce biais (voir Figure 57). A l'heure actuelle, en Suisse, ce service est proposé par environ 90 banques et près de 200 émetteurs de factures et tous semblent témoigner que cette tendance va augmenter.

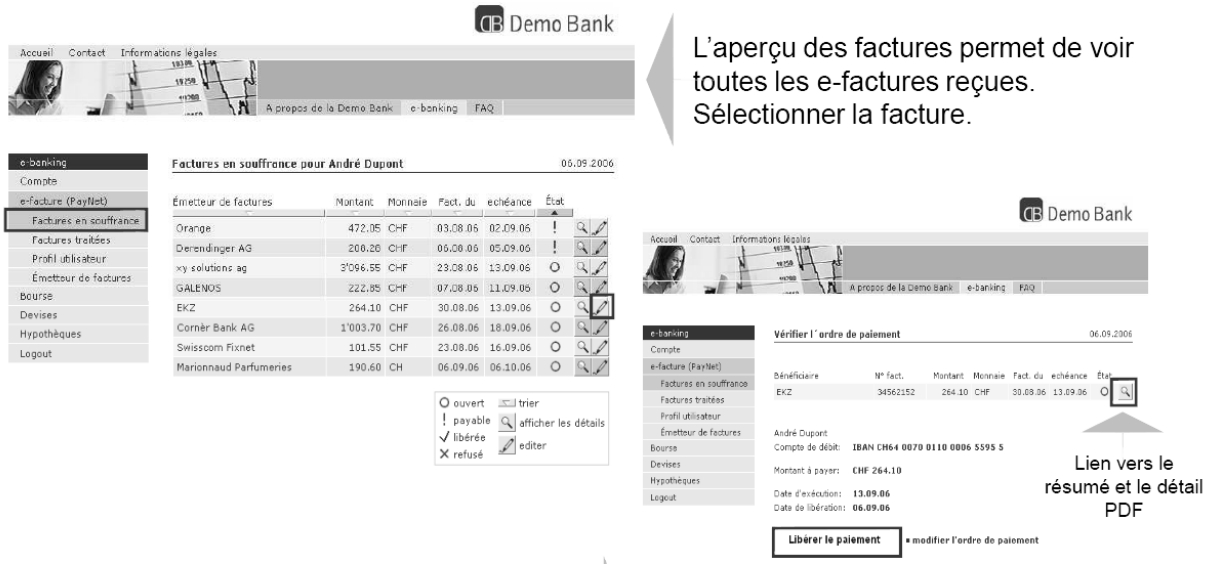


Figure 57: exemple d'un système de payement par internet

[58] PayNet (Schweiz) AG est une société du Telekurs Group.

Sans entrer trop dans les détails de la solution PayNet, le processus de facturation tel que décrit dans le chapitre IV3.2 reste identique (voir Figure 58). En effet, après le calcul de facturation, un fichier d'échange est envoyé de SAP vers le logiciel JetForm de la société Cisel SA (1). Par ce logiciel, les factures sont mises en page (2) et un fichier pdf est généré et stocké dans un répertoire partagé (3). Un programme SAP contrôle périodiquement ce répertoire et intègre les fichiers dans le serveur de contenu (4) et (5). Un programme spécifique a ensuite été ajouté pour envoyer la facture au format pdf et un ensemble de métadonnées sur le portail de PayNet (6). PayNet publie cette facture pour que les clients puissent faire le payement directement par leur accès bancaire sur internet (7). Une fois que le payement est effectué, la banque envoie un acquittement à SAP par le biais d'un fichier (8).

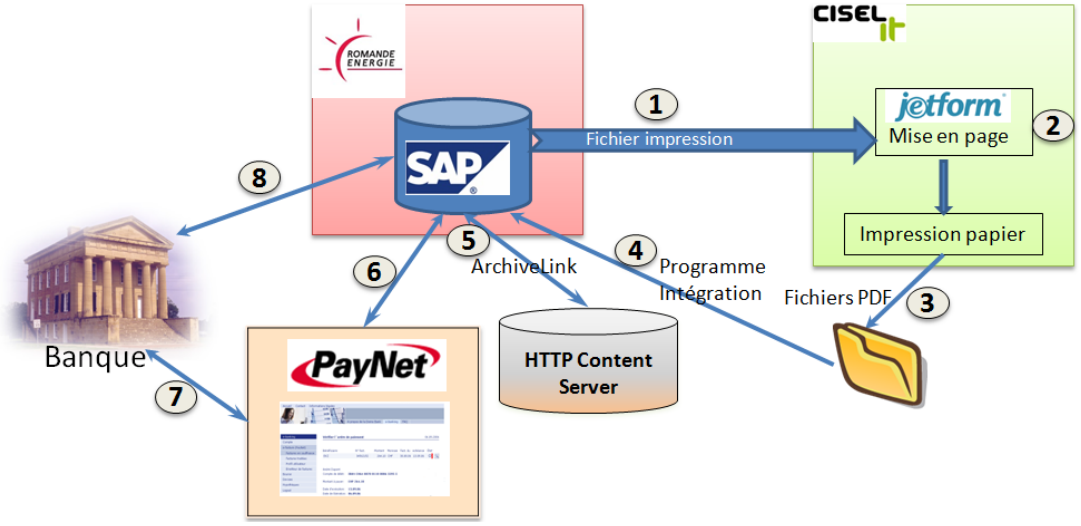


Figure 58: intégration PayNet

3.2 Archivage de données SAP chez Alcan Aluminium Valais SA

Alcan Aluminium Valais SA est une société appartenant au groupe Rio Tinto Alcan, spécialisé dans la transformation de l'aluminium. Leur site de production basé à Sierre utilise le progiciel de gestion d'entreprise SAP. Après plus de 10 ans d'utilisation, ils ont constaté d'importantes pertes de performances. Pour pallier ce problème, on m'a mandaté pour définir les périmètres d'un projet d'archivage de data. Après analyse, j'ai pu constater que la base de données n'avait pas atteint une taille critique par rapport à leur système. Il s'est avéré que le problème était causé par des dysfonctionnements d'un relais réseau entre le site de production et l'hébergeur des systèmes.

Malgré le résultat de l'analyse, le client a souhaité réaliser un projet d'archivage en trois phases : purge, réactivation des processus déjà réalisés dans le passé et développement de nouveaux processus.

- Comme pour Romande Energie SA, j'ai mis en place plusieurs processus d'archivage pour purger les données techniques.
- Lors de l'analyse, à la surprise générale, j'ai remarqué que certains processus d'archivages avaient déjà été mis en place par une ancienne équipe du service informatique. Ma mission a été de documenter ces derniers, former et planifier le lancement de ces processus.
- En ce qui concerne la troisième phase, c'est-à-dire paramétrer des nouveaux objets, le client a annulé cette partie.

3.3 Archivage des documents ESB Bienne

Le service ESB (Energie Service Bienne) de la ville de Bienne distribue et vend de l'électricité, du gaz et de l'eau. Dans leur utilisation de SAP, ils ont souhaité mettre en place la gestion par le gestionnaire de contenu HTTP Content Server de SAP des documents comme les factures, les relances et les courriers reçus des clients. Ma mission a été de paramétrer et créer des programmes spécifiques pour l'intégration de ces derniers.

3.4 Gestion des documents aux services industriels de Lausanne

Les services industriels de la ville de Lausanne commercialisent et vendent l'électricité, le gaz, le chauffage à distance et l'eau. Dans leur utilisation de SAP, ils ont souhaité mettre en place la gestion par le gestionnaire de contenu Alfresco des documents comme les factures, les relances et les photos de leurs articles en stock.[59] Ma mission a été de paramétrer l'ArchiveLink, de définir le cahier des charges pour la création d'un connecteur ArchiveLink ainsi que celui pour les programmes d'intégration.

[59] Alfresco offre une vraie alternative Open Source pour la Gestion de Contenu d'Entreprise (ECM) - Gestion de Documents, Collaboration, Gestion des Archives/Enregistrements légaux, Gestion de Contenu Web et Gestion des Documents Numérisés (voir http://www.alfresco.com/fr/).

VI Annexes

1 ILM ou gestion du cycle de vie de l'information.

L'ILM (*Information Lifecycle Management*) est un concept lancé en 2003 par Legato dans une optique marketing et qui recèle une portée bien plus grande et bien plus profonde qu'une appellation commerciale.

1.1 Définition

La SNIA (*Storage network industry association*) définit l'ILM comme « l'ensemble de règles, processus, pratiques et outils que l'on déploie pour aligner au mieux la valeur métier des informations et l'infrastructure qui les héberge, et ce depuis leur création jusqu'à leur destruction ! »

1.2 Objectif

L'objectif fondamental de l'ILM est de différencier les règles de stockage selon la valeur de l'information et d'ajuster au mieux les coûts de stockage aux exigences de sécurité, de conservation et de disponibilité des données. En plus, il cherche la convergence entre une gestion plus performante des infrastructures et des moyens de stockage, à l'aide de la virtualisation et par l'organisation logique de l'information avec une structuration et une hiérarchisation des données.

1.3 Démarche

La démarche consiste à prendre en compte chaque information avec ses caractéristiques afin de définir des règles communes et des « classes de service » pour l'ensemble des informations présentant les mêmes caractéristiques. Ces caractéristiques sont par exemple :

- Consultation : quand ? Par qui ? A quelle fréquence ? Avec quelle urgence ?
- Confidentialité : droits d'accès.
- Etat du document : figé, validé, signé etc.
- Durée de conservation : quelle est la durée de conservation ? Est-elle révisable ?
- Unicité de l'information : quelle originalité de l'information ? Existe-t-elle ailleurs ? Pourrait-on la reconstituer ?

Il est important de remarquer que ces caractéristiques évoluent dans le temps et l'ILM doit en tenir compte pour proposer la meilleure formule à chaque étape du cycle de vie de l'information.

1.4 Cycle de vie de l'information

L'information numérique est malléable, c'est-à-dire plus facile à produire, plus facile à modifier et plus facile à diffuser. Cependant, elle est aussi plus difficile à protéger et à conserver. Malgré tout, elle a un cycle de vie avec les mêmes étapes que le document traditionnel :

- **Etape 1** : c'est la phase de l'élaboration, donc modifiable jusqu'à son approbation.
- **Etape 2** : c'est la validation et elle devient figée.
- **Etape 3** : à ce stade elle est dite en maintenance dans le sens qu'elle est conservée ou/et archivée pendant la durée de validité.
- **Etape 4** : c'est la fin de l'utilité qui se traduit par une destruction ou par un transfert vers des collections patrimoniales.

2 Nom du fichier logique

Le nom du fichier logique permet de créer des fichiers indépendants de la plate-forme (Windows, Mac OS, etc..) et dans un emplacement du système des fichiers. Ce dernier est défini par la transaction FILE.

2.1 Création d'un chemin logique

Avant de pouvoir déterminer le nom du fichier logique, il faut au préalable déterminer un chemin logique. Ce dernier définit un chemin (ou emplacement) indépendant de la plate-forme où les fichiers logiques sont enregistrés. Les principaux paramètres sont (voir Figure 59) :

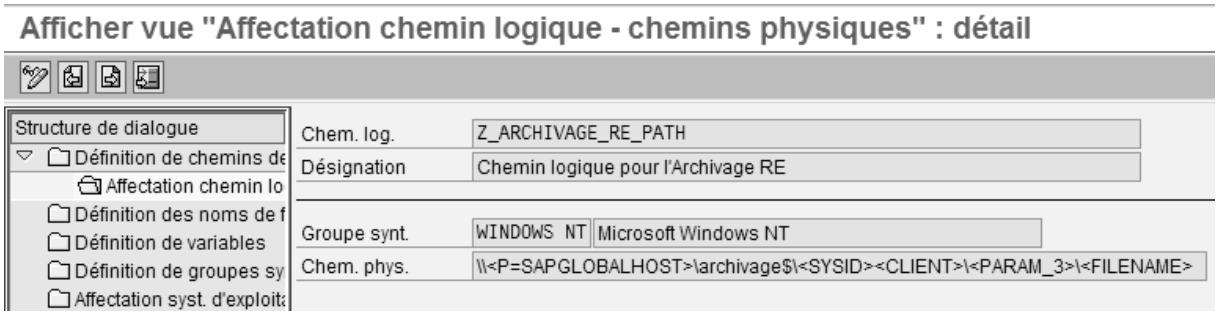


Figure 59 : Définition d'un chemin logique

- **Chem. log.** : c'est le nom du chemin logique.
- **Groupe synt**. : il permet de choisir dans une liste déroulante les systèmes d'exploitation supportés par SAP.
- **Chem. phys**. : c'est le chemin physique spécifique à la plate-forme dans laquelle des fichiers sont enregistrés. Il doit contenir le mot réservé <FILENAME> comme caractère générique pour le nom de fichier. Par ailleurs, il peut contenir d'autres mots réservés pour indiquer dynamiquement des sous-répertoires.

2.2 Définition du nom de fichier logique

La définition du nom de fichier logique est faite par la définition des paramètres suivants (voir Figure 60) :

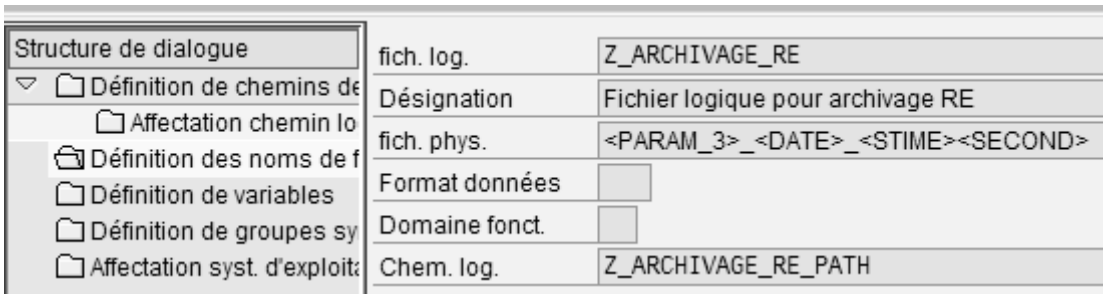


Figure 60 : Définition d'un fichier logique

- **Fich. log. :** nom du fichier logique.
- **Fich. phys. :** définition de la structure pour la création du nom du fichier lors de l'exécution. Comme pour le chemin logique, il est possible de lui associer un certain nombre de mots réservés.
- **Format données :** il est possible de « forcer » le format du fichier généré.
- **Domaine fonct**. : le domaine fonctionnel indique dans quelle application un nom de fichier logique est utilisé. Au niveau fonctionnel, il ne joue aucun rôle pour la création des noms de fichiers spécifiques de la plate-forme.
- **Chem. Log**. : c'est le nom du chemin logique précédemment créé. Par la concaténation de ce dernier avec la définition du fichier logique, le système peut créer un fichier dans un placement défini.

2.3 Mots réservés

Voici une liste non exhaustive de mots réservés pour définir dynamiquement une partie du chemin ou le nom du fichier:

Mot réservé	Description
<SYSID>	Est le nom de l'instance SAP.
<CLIENT>	Est le nom du mandant
<PARAM_1>	Domaine fonctionnel (par exemple FI pour finance)
<PARAM_3>	Est le nom de l'objet d'archivage
<DATE>	Date lors de l'exécution
<TIME>	Heure minute lors de l'exécution
<SECOND>	Seconde lors de l'exécution

3 HTTP Content Server de SAP

Le HTTP Content Server fourni par SAP est un gestionnaire de contenu basé sur un serveur http[60] (voir Figure 61).

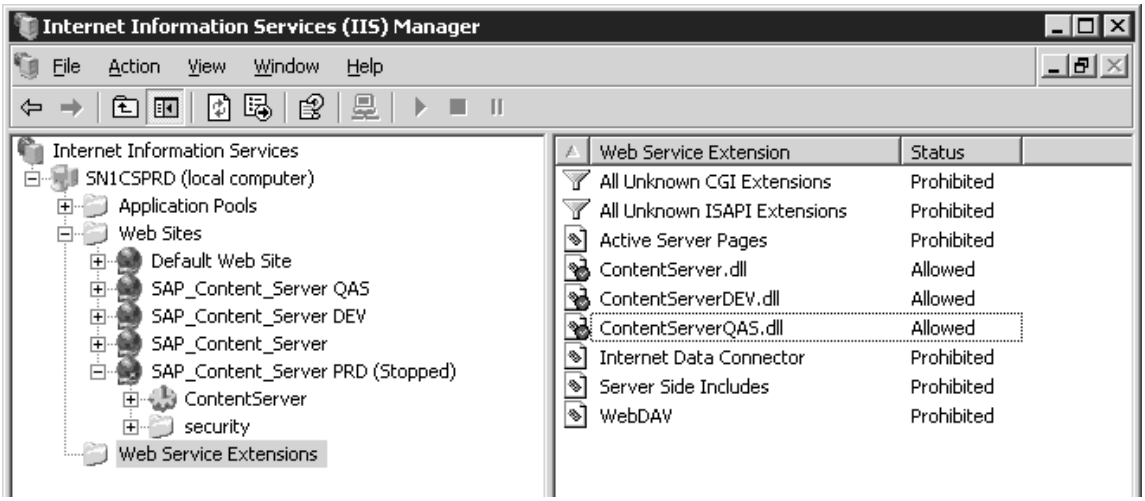


Figure 61: exemple d'implémentation server HTTP content Server

Ce gestionnaire gère uniquement le flux des fichiers entrants et sortants en les classant dans une arborescence de répertoires pour un accès rapide (voir Figure 62).

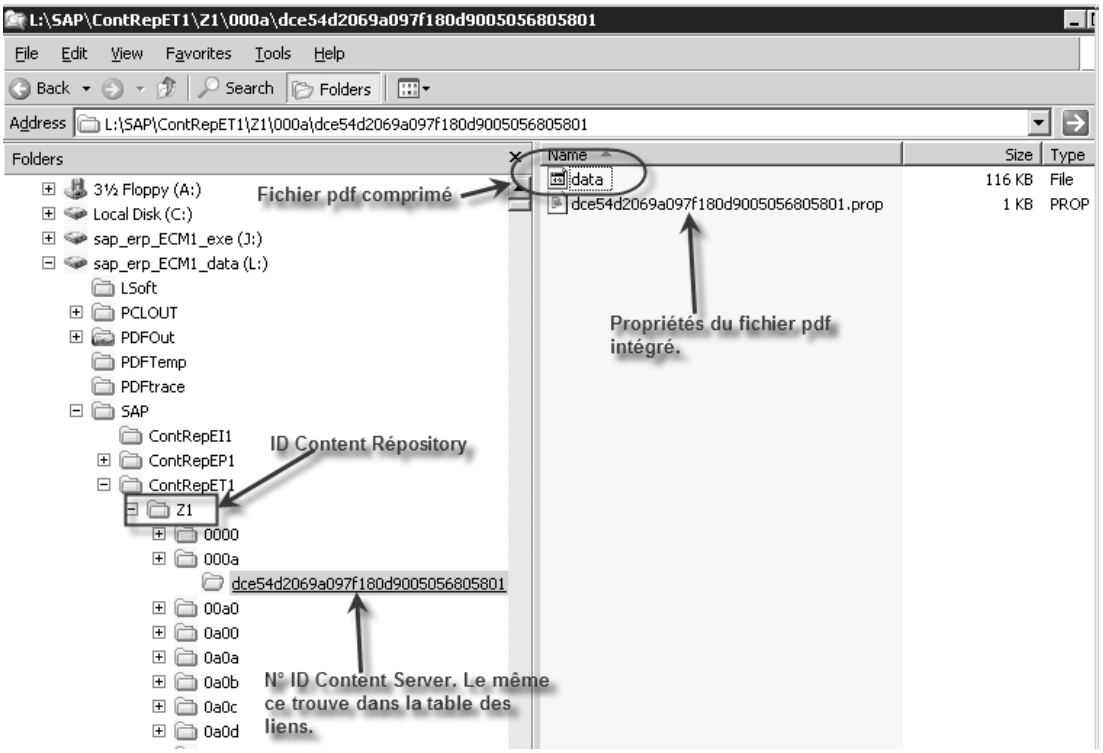


Figure 62: exemple d'arborescence du HTTP Content Server

[60] Est un logiciel implémentant le protocole HTTP et ouvert aux connections provenant de l'extérieur

Les fichiers peuvent être comprimés, cryptés et un fichier data accompagne ces derniers avec un certain nombre d'informations (voir **Figure 63**).

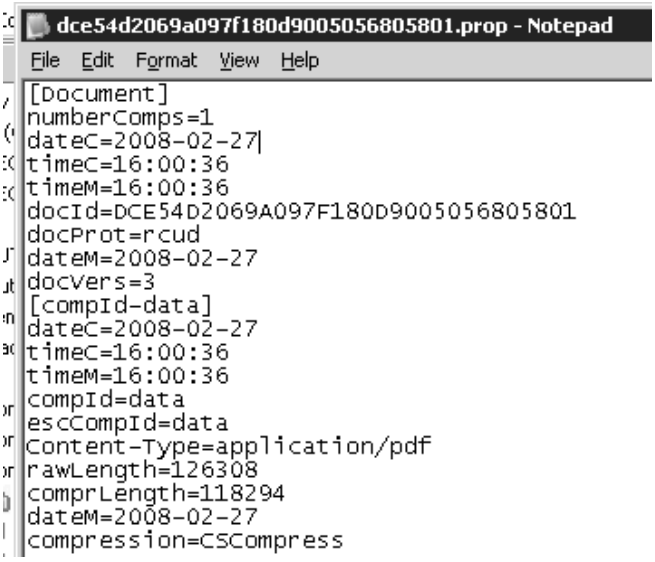


Figure 63: exemple de fichier data

En conclusion, l'HTTP Content Server de SAP reste un outil très rudimentaire et économique pour lier de fichiers annexes à SAP. Ses principaux atouts sont : économique (car il est fourni gratuitement avec les licences SAP), facile à installer et administrer et parfaitement intégré avec l'environnement SAP par le protocole ArchiveLink. D'ailleurs, il est complètement inutilisable en dehors de cet environnement car, d'une part, il est impossible de lui associer des métadonnées et, d'autre part, il est complètement dépourvu d'outils indispensables comme la gestion du cycle de vie du document, le classement, les outils de recherche et de gestion des versions.

Compte tenu de ces carences, l'HTTP Content Server n'est pas adapté à la gestion électronique des documents.

4 Code source ZP_ISU_LOAD_FACTURE_PDF

Ce programme écrit en ABAP[61] (Advanced Business Application Programming) permet de lire les fichiers stockés dans le répertoire partagé, de contrôler la cohérence des ces derniers, de poster les fichiers dans le serveur de contenu et de créer les liens dans SAP.

```
*&---------------------------------------------------------------------*
*& Report  ZP_ISU_LOAD_FACTURE_PDF
*&
*&---------------------------------------------------------------------*
*& Auteur:   Michel BOLOGNESI
*& Société:  neo technologies SA
*&---------------------------------------------------------------------*
*& Description:
*&    Intégration dans le Content Serveur des factures au format pdf.
*&    - Format des fichiers : <NuméroDefacture><Date >.pdf
*&        <NuméroDeFacture> : 12 positions numériques
*&        <Date > : format anglais -> AAAAMMJJ
*&    - Exemple fichier: 00100045667020060528.pdf
*&---------------------------------------------------------------------*

report  zp_isu_load_facture_pdf
 no standard page heading
 line-size  150
 line-count 65(0)
 message-id e9.

*---------------------------------------------------------------------*
* Définition des variables
*---------------------------------------------------------------------*
types: begin of ty_table,  "Pour récupérer les fichiers stockée dans le répertoire
       line(200) type c,
       end of ty_table.

data: t_file type epsfili occurs 0 with header line.
data: w_file type epsfili.
data: w_path like  epsf-epsdirnam.  "répertoire
data: w_count like epsf-epsfilsiz.  "Nombre de fichier lu dans le répertoire
data: w_date_ref like toav0-ar_date. "Date
data: w_no_facture like  sapb-sapobjid. "Numéro de la facture
data: w_err(1).   "Retour de la form f_test_nom_file
data: w_err_cs(1).  "Retoutd e la form f_ctrl_doublon_tb_lien
data: w_no_err type i.
data: t_toaom like toaom occurs 0 with header line.
data: w_file_arch(128) type c.  "path + file pour archivage
data: w_doc_id type toav0-arc_doc_id.  "Numéro du document du content server
data: w_cont type i.
```

[61] L'ABAP est l'un des successeurs du COBOL et est apparu dans les années 1980 dans la vague des langages de quatrième génération (4GLs). Il s'agit d'un dérivé du langage permettant de réaliser des sorties de données (appelées rapports) de l'application SAP R/2, sur lequel de nombreuses multinationales avaient bâti leur architecture d'application professionnelle. L'ABAP a par la suite été maintenu comme langage de référence pour les applications SAP R/3 qui sont apparues en 1992.

Le langage a été par la suite étendu pour englober un modèle de données orienté objet (ABAP Objects) à partir de sa version 4.5, pour être finalement intégré comme langage d'un produit plus général appelé NetWeaver. Ce dernier utilise aussi bien l'ABAP que le java.

C'est un langage interprété.

(voir wikipedia : http://fr.wikipedia.org/wiki/ABAP)

```
*---------------------------------------------------------------------*
* Variables de sélection
*---------------------------------------------------------------------*

parameter p_path like  epsf-epsdirnam obligatory default '\\ECM1\WORK\'.
parameter p_text as checkbox default 'X'.
parameter p_sapobj like toaom-sap_object no-display default 'PRINTDOC'.
parameter p_arobj like toaom-ar_object no-display default 'ISU_BILL'.

*---------------------------------------------------------------------*
* Start
*---------------------------------------------------------------------*

start-of-selection.

* Lecture du répertoire et récupération des fichiers
  w_path = p_path.
  translate w_path to upper case.
  call function 'EPS_GET_DIRECTORY_LISTING'
    exporting
      dir_name               = w_path
    importing
      file_counter           = w_count
    tables
      dir_list               = t_file
    exceptions
      invalid_eps_subdir     = 1
      sapgparam_failed       = 2
      build_directory_failed = 3
      no_authorization       = 4
      read_directory_failed  = 5
      too_many_read_errors   = 6
      empty_directory_list   = 7
      others                 = 8.
  if sy-subrc <> 0.
    message i899 with 'Erreur lecture du répertoire'.
    stop.
  endif.

* Recherche des tables de liens: voir transaction AOC3.
  call function 'ARCHIV_METAINFO_GET'
    exporting
      ar_object             = p_arobj
      sap_object            = p_sapobj
    tables
      toaom_fkt             = t_toaom
    exceptions
      error_connectiontable = 1
      error_parameter       = 2
      others                = 3.
  if sy-subrc <> 0.
    message e052 with 'TOAOM'.
  endif.
*
* Traitement de tous les fichiers lu
  loop at t_file into w_file
    where rc = '0000'.

*   Test si le fichier est correct
    perform f_test_nom_file using w_file-name
                                changing w_err.
    if w_err = 'N'.
*     Test si le fichier n'est pas un doublon
      perform f_ctrl_doublon_tb_lien
            using w_file-name
            changing w_err_cs.

      if w_err_cs = 'N' and p_text <> 'X'.
        clear w_doc_id.
        concatenate p_path '\' w_file-name into w_file_arch.

*       Stocker le fichier dans le Content Server
        call function 'SCMS_AO_FILE_CREATE_PATH'
          exporting
```

```
            mandt        = sy-mandt
            arc_id       = t_toaom-archiv_id
            path         = w_file_arch
            doc_type     = 'PDF'
            no_delete    = 'X'
          importing
            doc_id       = w_doc_id
          exceptions
            error_http   = 1
            error_archiv = 2
            error_kernel = 3
            error_config = 4
            others       = 5.
        if sy-subrc <> 0.
*         text-005 : Erreur lors dépose fichier dans CS:
          write: / w_file-name(24), text-005, sy-subrc.
          add 1 to w_no_err.
        else.
          w_no_facture = w_file-name(12).
          w_date_ref = w_file-name+12(8).

*         Création lien Archivelink
          call function 'ARCHIV_CONNECTION_INSERT'
            exporting
              archiv_id             = t_toaom-archiv_id
              arc_doc_id            = w_doc_id
              ar_date               = w_date_ref
              ar_object             = p_arobj
              mandant               = sy-mandt
              object_id             = w_no_facture
              sap_object            = p_sapobj
              doc_type              = 'PDF'
            exceptions
              error_connectiontable = 1
              others                = 2.

          if sy-subrc <> 0.
            add 1 to w_no_err.
*           text-006: Erreur création lien Archivelink !
            write:/ w_file-name(24), text-006, sy-subrc.
          else.
            add 1 to w_cont.
          endif.
        endif.
      else.
        if w_err_cs <> 'N'.
          add 1 to w_no_err.
        endif.
      endif.
    else.
      add 1 to w_no_err.
    endif.
  endloop.

* Ecriture des résultats
  new-line.
  uline.
  new-line.
  write: / text-008,'       ', w_count.        "Text-008: Nombre de fichier lu:
  write: / text-009,'             ', w_no_err. "Text-009: Nombre d''erreur:
  write: / text-010,' ',w_cont.                "Text-010: Nombre de factures intégrés:

end-of-selection.


*---------------------------------------------------------------------*
* FORMS : sous-programmes
*---------------------------------------------------------------------*


*&---------------------------------------------------------------------*
*&      Form  f_test_nom_file
*&---------------------------------------------------------------------*
*       Premier test de vérification de pertinence si le nom du fichier
*       est pertinent.
*---------------------------------------------------------------------*
```

```
form f_test_nom_file using w_nom
                     changing w_err.
* Variables locales
  data: w_long type i.
  data: w_no_fact(12) type c.
  data: w_date type d.
  data: w_ext(4) type c.

  w_long = strlen( w_nom ).
  w_no_fact = w_nom(12).
  w_date = w_nom+12(8).
  w_ext = w_nom+20(4). translate w_ext to upper case.

* Si le num facture est alphanumérique -> erreur
  if w_no_fact co '0123456789'.
*   Vérification si la date est valide
    call function 'DATE_CHECK_PLAUSIBILITY'
      exporting
        date                      = w_date
      exceptions
        plausibility_check_failed = 1
        others                    = 2.
    if sy-subrc <> 0.
      w_err = 'O'.
*     Text-002: Date incorrecte !
      write: / w_nom(24), text-002.
    else.
*     Vérification si le fichier est un pdf
      if w_ext = '.PDF'.
        w_err = 'N'.
      else.
        w_err = 'O'.
*       Text-003 : Ce n''est pas un fichier pdf !
        write: / w_nom(24), text-003.
      endif.
    endif.
  else.
    w_err = 'O'.
*   Text-001 : Numéro de facture incohérent !
    write: / w_nom(24), text-001.
  endif.
endform.                    "f_test_nom_file


*&---------------------------------------------------------------------*
*&      Form  f_ctrl_doublon_tb_lien
*&---------------------------------------------------------------------*
*       Vérification si le fichier est déjà présent dans le content
*       serveur.
*----------------------------------------------------------------------*
form f_ctrl_doublon_tb_lien
      using w_nom
      changing w_err.

* Variables locales
  data: w_no_fact like toa01-object_id.
  data: w_date type d.
  data: w_date_ref type d.
  data: w_toa01 like toa01.
  data: w_tb_err like toaom-connection.

  w_no_fact = w_nom(12).
  w_date = w_nom+12(8).
  clear w_tb_err.

* Vérif si le fichier existe dans les tables de liens
  loop at t_toaom.
    select single * from (t_toaom-connection) into w_toa01
      where sap_object = p_sapobj
        and object_id = w_no_fact.
    if sy-subrc = 0.
      clear w_toa01.
      select single * from (t_toaom-connection) into w_toa01
      where sap_object = p_sapobj
        and object_id = w_no_fact
        and ar_date = w_date.
```

```
      if sy-subrc = 0.
        w_tb_err = t_toaom-connection.
      endif.
    endif.
  endloop.

  if w_tb_err is initial.
    w_err = 'N'.
  else.
    w_err = 'O'.
*     Text-004 : Fichier déjà existant dans le content serveur !
    write: / w_nom(24),' ' , text-004, ' ', ' (',w_tb_err,')'.
  endif.

endform.                    "f_ctrl_doublon_tb_lien
```

5 Exemple de programme de lecture des archives

Le code source ci-dessous est un exemple de lecture des fichiers archives issus de l'objet BC_SBAL. A noter que dans le cadre de ce projet, ce programme n'a qu'un intérêt pédagogique.

```
*---------------------------------------------------------------------
* neo technologies SA
* Michel Bolognesi
*---------------------------------------------------------------------
* Date: 28.10.2005
*---------------------------------------------------------------------
* Description : Ce programme affiche les données archives des protocoles
*               d'applications (objet BC_SBAL).
*---------------------------------------------------------------------

REPORT z_archive_bcsbal .

DATA: w_handle LIKE sy-tabix,
      w_record_type LIKE arc_buffer-rname,
      w_buffer LIKE arc_buffer-segment,
      w_balhdr_wa LIKE balhdr,
      w_number_of_records_read TYPE i.

* Tables internes ..........................................................
TYPES:
  BEGIN OF w_donnee_type,
    lognumber LIKE balhdr-lognumber,  "N° de protocole
    object LIKE balhdr-object,        "Nom de l'objet
    subobject LIKE balhdr-subobject,  "Sous-objet
    extnumber LIKE balhdr-extnumber,  "Identification externe
    aldate LIKE balhdr-aldate,        "Date
    aluser LIKE balhdr-aluser,        "Utilisateur
    altcode LIKE balhdr-altcode,      "Code transaction
  END OF w_donnee_type.
DATA: t_donnee TYPE w_donnee_type OCCURS 0 WITH HEADER LINE.

***********************************************************************
*   PROGRAMME PRINCIPALE
***********************************************************************
START-OF-SELECTION.

* Lecture des archives
  PERFORM f_lecture_archives.

* Affichage des valeurs lues
  LOOP AT t_donnee.
    WRITE: / t_donnee.
  END LOOP.

END-OF-SELECTION.


***********************************************************************
*   FORMS
***********************************************************************

*&---------------------------------------------------------------------
*   FORM f_lecture_archives
*&---------------------------------------------------------------------
*   Lecture des données archivées
*---------------------------------------------------------------------
FORM f_lecture_archives.
* Ouverture du fichier archive
  CALL FUNCTION 'ARCHIVE_OPEN_FOR_READ'
       EXPORTING
            object         = 'BC_SBAL'
       IMPORTING
            archive_handle = w_handle
       EXCEPTIONS
```

```
            OTHERS         = 1.

  IF sy-subrc <> 0.
    WRITE: / 'Aucun file archive trouvé'(001).
    EXIT.
  ENDIF.
  CLEAR w_number_of_records_read.
* Boucle pour trouver le prochain objet du fichier archive
  DO.
    CALL FUNCTION 'ARCHIVE_GET_NEXT_OBJECT'
         EXPORTING
              archive_handle = w_handle
         EXCEPTIONS
              end_of_file    = 1
              OTHERS         = 2.

    IF sy-subrc <> 0.
      EXIT.
    ENDIF.
*   Retour des valeurs lues
    DO.
      CALL FUNCTION 'ARCHIVE_GET_NEXT_RECORD'
           EXPORTING
                archive_handle   = w_handle
           IMPORTING
                record_structure = w_record_type
                record           = w_buffer
           EXCEPTIONS
                end_of_object    = 1
                OTHERS           = 2.
      IF w_record_type = 'BALHDR'.
        w_balhdr_wa = w_buffer.
        MOVE-CORRESPONDING w_balhdr_wa TO t_donnee.
        APPEND t_donnee.
        CLEAR: t_donnee.
      ENDIF.

      IF sy-subrc <> 0.
        EXIT.
      ENDIF.
      ADD 1 TO w_number_of_records_read.
    ENDDO.
  ENDDO.

* Clôture du fichier archive
  CALL FUNCTION 'ARCHIVE_CLOSE_FILE'
       EXPORTING
            archive_handle = w_handle.

ENDFORM.
```

VII Bibliographie

1. RIETSCH Jean-Marc, CHABIN Marie-Anne et CAPRIOLI Eric, *Dématérialisation et archivage électronique*, Dunod, 2006, 207 pp.
2. STEFANI Helmut, Archiving Your SAP Data, SAP Press, 2003, 334 pp.

1 Liens internet

L'ensemble des liens internet ont été consultés au cours des mois d'avril – juin 2010.

- http://www.archives.ch/ : c'est le site de la société Archive & Plus SA. Cette société est spécialisée dans l'archivage. Elle propose un certain nombre de documents concernant la législation et des recommandations sur l'archivage du point de vue de l'archiviste.
- www.adobe.com/fr/products/acrobat/pdfs/pdfarchiving.pdf : Livre blanc « Adobe PDF, norme d'archivage ».
- http://www.guideinformatique.com/dossier-archivage_et_sauvegarde-25.htm : Il y a plusieurs articles concernant l'archivage surtout du point de vue Hardware.
- http://www.fedisa.eu/fedisa2007/ : ce site de la FedISA (Fédération de l'ILM, du Stockage et de l'Archivage) propose plusieurs livres blancs très intéressants concernant :
    - Identification et authentification à l'ère numérique - Enjeux et approches méthodologiques
    - Conserver les courriers électroniques, ou comment résoudre la problématique de l'archivage des e-mails
    - Guide Pratique "Comprendre et utiliser les normes dans le domaine de l'archivage numérique"
    - Guide pratique du PDF/A
- http://www.ssi.gouv.fr : ce site de l'ANSSI (Agence nationale de la sécurité des systèmes d'information) propose plusieurs documents concernant l'archivage électronique sécurisé (excellent référentiel des bonnes pratiques).
    - Archivage électronique sécurisé – Etat de l'art
    - Cahier des charges pour un système d'archivage électronique (sphère publique)

  - Etc.
- http://www.bakom.admin.ch/themen/internet/00467/index.html?lang=fr Ceci est un lien vers le site de l'administration fédérale (suisse) sur le thème de la signature électronique.
- http://www.digitech-sa.ch/presse.php: dans ce site de Digitech SA on peut trouver plusieurs articles très intéressants consacrés à différentes problématiques de l'archivage :
  - Conservation électronique des documents dans les entreprises
  - La loi et la conservation électronique des documents en Suisse
  - Conservation des factures et pièces justificatives au regard de la TVA
  - Etc.
- http://www.cines.fr/spip.php?article131 : brève description du modèle OAIS (Open Archival Information System).
- http://www.figoblog.org/document1089.php : brève description du modèle OAIS.
- http://www.aristote.asso.fr/PIN/presentations/2006/ILM.pdf : brève présentation concernant la notion de ILM.
- http://www.aproged.org/index.php/Voir-details/Publications/8-Memo-La-maitrise-du-cycle-de-vie-du-document-numerique.html : Memo - La maîtrise du cycle de vie du document numérique
- http://help.sap.com/saphelp_erp2005/helpdata/fr/7c/aea63a9c69612fe10000000a114084/frameset.htm : aide en ligne de SAP concernant les objets d'archivage SAP.
- http://help.sap.com/saphelp_erp2005/helpdata/fr/d7/ad7c38ec757932e10000009b38f8cf/frameset.htm : aide en ligne SAP concernant le processus d'archivage
- http://www.sdn.sap.com/irj/sdn/inf?rid=/webcontent/uuid/ef1b00f7-0a01-0010-74ab-e9260e80a441 : plusieurs documents mis à disposition par SAP concernant l'archivage de data.
- http://www.developpez.net/forums/d507979/logiciels/solutions-dentreprise/erp/sap/methode-asap-questionnaire/ : forum autour de la méthodologie ASAP.

2 Support de cours SAP

- **BC 660, Data Archiving, 2001, 347 pp** : notion de base concernant l'archivage de données SAP.
- **BC 615, SAP ArchiveLink, 2001, 358 pp**: bases techniques concernant l'ArchiveLink de SAP et un survol des applications possibles.
- **BC 670, Programming Display Fonctions, 2002, 225 pp** : donne les bases nécessaires pour la création de programmes personnalisés d'archivage de données.

VIII Liste des figures

Figure 1 : intégration des solutions mySAP 13
Figure 2 : modèle en trois couches. 15
Figure 3 : paysage système 16
Figure 4 : fonctions du système d'archivage électronique 45
Figure 5 : Exemple d'architecture technique du SAE 46
Figure 6 : définition 51
Figure 7 : Principe d'archivage 52
Figure 8 : Exemple facture de décompte 54
Figure 9 : Copie d'écran d'un objet d'archivage (transaction AOBJ) 55
Figure 10 : structure objet archivage 56
Figure 11 : Liste des tables supprimées uniquement 57
Figure 12 : exemple de dépendance entre objets 58
Figure 13 : liste des programmes pour la lecture des fichiers archives 58
Figure 14 : liste des transactions permettant de définir le temps de résidence 59
Figure 15 : exemple de définition temps de résidence 59
Figure 16 : Liste des tables index 59
Figure 17 : Copie d'écran de la transaction SARA 60
Figure 18 : paramètres processus archivage 61
Figure 19 : Options techniques pour tous les objets archivage 62
Figure 20 : Options techniques spécifiques objets archivage 63
Figure 21 Options de routage 65
Figure 22: étapes processus annexes 67
Figure 23: gestion des archives 68
Figure 24 : documents entrants et sortant entre SAP et d'autres applications. 69
Figure 25 : copie d'écran de la vue « lien pour Content Repositories ». 70
Figure 26 : exemple d'entrée de la table type de document. 71
Figure 27 : Vue de paramétrage du Content Repositories. 72
Figure 28 : exemple d'un tuple de la table d'une table de lien. 73
Figure 29: Table des liens 75
Figure 30: feuille de route ASAP 79

Figure 31 : croissance de la base de données 81
Figure 32: Transaction DB15 83
Figure 33: graphique de réseau..... 84
Figure 34: résultat d'analyse par le programme RARCCOA2 86
Figure 35: analyse de la table GLPCA par la transaction TAANA 87
Figure 36: Dépendance objet FI_MKKDUNN 89
Figure 37: dépendance objet FI_ACCRECV 90
Figure 38: dépendance objet ISU_PRDOCL 90
Figure 39: dépendances objet ISU_BILLZ et BILL..... 91
Figure 40: dépendances objet ISU_FACTS 91
Figure 41: dépendances objet ISU_EABL 91
Figure 42: dépendances objet ISU_BBP 92
Figure 43: processus d'impression avant modification..... 99
Figure 44: processus de facturation après modification 100
Figure 45: protection des fichiers pdf..... 101
Figure 46: structure du plan de classement 102
Figure 47: paramétrage spécifique 106
Figure 48: suppression des relations IDC8..... 106
Figure 49: intégration des pdf dans le HTTP Content Server 107
Figure 50: référentiel de contenu..... 108
Figure 51: paramétrage du content repositorie..... 108
Figure 52: type de document 109
Figure 53: Création de la table des liens 109
Figure 54: traitement des liens 110
Figure 55: paramètres du programme ZP_ISU_LOAD_FACTURE_PDF..... 110
Figure 56: exemple de visualisation d'une facture pdf 112
Figure 57: exemple d'un système de payement par internet 120
Figure 58: intégration PayNet 121
Figure 59 : Définition d'un chemin logique..... 127
Figure 60 : Définition d'un fichier logique 128
Figure 61: exemple d'implémentation server HTTP content Server 129
Figure 62: exemple d'arborescence du HTTP Content Server 129
Figure 63: exemple de fichier data 130

IX Liste des tableaux

Tableau i: Temps de conservation 28
Tableau ii: croissance de la base de données 80
Tableau iii: classement des 50 tables les plus volumineuses 82
Tableau iv: correspondance table / objet d'archivage 85
Tableau v: temps de résidence 94
Tableau vi: consultation des archives 95
Tableau vii: résultat après l'archivage de test 112
Tableau viii: résultat après le premier archivage productif 113
Tableau ix: bilan du projet 118