



**Livre blanc**

---

# ***GED open source***

*Thomas CHOPPY – Consultant ECM*

*Nicolas RICHTON – Consultant technique*

---

## GED Open source

---

### PREAMBULE

#### Smile

Smile est une société d'ingénieurs experts dans la mise en œuvre de solutions open source et l'intégration de systèmes appuyés sur l'open source. Smile est membre de l'APRIL, l'association pour la promotion et la défense du logiciel libre.

Smile compte 290 collaborateurs en France, 320 dans le monde (septembre 2009), ce qui en fait *la première société en France spécialisée dans l'open source*.

Depuis 2000, environ, Smile mène une action active de veille technologique qui lui permet de découvrir les produits les plus prometteurs de l'open source, de les qualifier et de les évaluer, de manière à proposer à ses clients les produits les plus aboutis, les plus robustes et les plus pérennes.

Cette démarche a donné lieu à toute une gamme de *livres blancs* couvrant différents domaines d'application. La gestion de contenus (2004), les portails (2005), la business intelligence (2006), les frameworks PHP (2007), la virtualisation (2007), et la gestion électronique de documents (2008), ainsi que les PGIs/ERPs (2008). Parmi les ouvrages publiés en 2009, citons également « Les VPN open source », et « Firewall est Contrôle de flux open source », dans le cadre de la collection « Système et Infrastructure ».

Chacun de ces ouvrages présente une sélection des meilleures solutions open source dans le domaine considéré, leurs qualités respectives, ainsi que des retours d'expérience opérationnels.

Au fur et à mesure que des solutions open source solides gagnent de nouveaux domaines, Smile sera présent pour proposer à ses clients d'en bénéficier sans risque. Smile apparaît dans le paysage informatique français comme le prestataire intégrateur de choix pour accompagner les plus grandes entreprises dans l'adoption des meilleures solutions open source.

Ces dernières années, Smile a également étendu la gamme des services proposés. Depuis 2005, un département consulting accompagne nos clients, tant dans les phases d'avant-projet, en recherche de solutions, qu'en accompagnement de projet. Depuis 2000, Smile dispose d'un studio graphique, devenu en 2007 Agence Media Interactive, proposant outre la création graphique, une expertise e-marketing, éditoriale, et interfaces

## GED Open source

---

riches. Smile dispose aussi d'une agence spécialisée dans la Tierce Maintenance Applicative, le support et l'exploitation des applications. Enfin, Smile est implanté à Paris, Lyon, Nantes, Bordeaux et Montpellier. Et présent également en Espagne, en Suisse, en Ukraine et au Maroc.

### Quelques références

#### Gestion documentaire

EDF R&D, Conseil Régional de Picardie, Leroy Merlin, Primagaz, Renault F1, INRIA, Véolia Propreté, Centre d'Information de la Jeunesse (CIDJ), Pierre Audoin Consultants, Union de la Coopération Forestière Française, Ministère Belge de la Communauté Française

#### Intranets et Extranets

Société Générale, Caisse d'Épargne, Bureau Veritas, Commissariat à l'Energie Atomique, Visual, CIRAD, Camif, Lynxial, RATP, Sonacotra, Faceo, CNRS, AmecSpie, INRA, CTIFL, Château de Versailles, Banque PSA Finance, Groupe Moniteur, Vega Finance, Ministère de l'Environnement, Arjowiggins, JCDecaux, Ministère du Tourisme, DIREN PACA, SAS, CIDJ, Institut National de l'Audiovisuel, Cogedim, Diagnostica Stago, Ecureuil Gestion, Prolea, IRP-Auto, Conseil Régional Ile de France, Verspieren, Conseil Général de la Côte d'Or, Ipsos, Bouygues Telecom, Prisma Presse, Zodiac, SANEF, ETS Europe, Conseil Régional d'Ile de France, AON Assurances & Courtage, IONIS, Structis (Bouygues Construction), Degrémont Suez, GS1-France, DxO, Conseil Régional du Centre, Beauté Prestige International, HEC, Veolia

#### Internet, Portails et e-Commerce

Cadremploi.fr, chocolat.nestle.fr, creditlyonnais.fr, explorimmo.com, meilleurtaux.com, cogedim.fr, capem.fr, Editions-cigale.com, hotels-exclusive.com, souriau.com, pci.fr, odit-france.fr, dsv-cea.fr, egide.asso.fr, Osmoz.com, spie.fr, nec.fr, vizzavi.fr, sogeposte.fr, ecofi.fr, idtgv.com, metro.fr, stein-heurtey-services.fr, bipm.org, buitoni.fr, aviation-register.com, cci.fr, eaufrance.fr, schneider-electric.com, calypso.tm.fr, inra.fr, cnil.fr, longchamp.com, aesn.fr, bloom.com, Dassault Systemes 3ds.com, croix-rouge.fr, worldwatercouncil.org, Projectif, credit-cooperatif.fr, editionsbussiere.com, glamour.com, nmmedical.fr, medistore.fr, fratel.org, tiru.fr, faurecia.com, cidil.fr, prolea.fr, bsv-tourisme.fr, yves.rocher.fr, jcdecaux.com, cg21.fr, veristar.com, Voyages-sncf.com, prismapub.com, eurostar.com, nationalgeographic.fr, eau-seine-normandie.fr, ETS Europe,

## GED Open source

---

LPG Systèmes, cnous.fr, meddispar.com, Amnesty International, pompiers.fr, Femme Actuelle, Stanhome-Kiotis, Gîtes de France, Bouygues Immobilier, GPdis, DeDietrich, OSEO, AEP, Lagardère Active Média, Comexpo, Reed Midem, UCCIFE, Pagesjaunes Annonces, 1001 listes, UDF, Air Pays de Loire, Jaccede.com, ECE Zodiac, Polytech Savoie, Institut Français du Pétrole, Jeulin, Atoobi.com, Notaires de France, Conseil Régional d'Ile-de-France, AMUE

### Applications métier et applications décisionnelles

Renault, Le Figaro, Sucden, Capri, Libération, Société Générale, Ministère de l'Emploi, CNOUS, Neopost, Industries, ARC, Laboratoires Merck, Egide, Bureau Veritas, ATEL-Hotels, Exclusive Hotels, CFRT, Ministère du Tourisme, Groupe Moniteur, Verspieren, Caisse d'Epargne, AFNOR, Souriau, MTV, Capem, Institut Mutualiste Montsouris, Dassault Systèmes, Gaz de France, CAPRI Immobilier, Croix-Rouge Française, Groupama, Crédit Agricole, Groupe Accueil, Eurordis, Mindscape, Xinek, Institut National de l'Audiovisuel, CDC Arkhineo

## **Ce livre blanc**

Ce document a pour objectifs de présenter notre approche de la gestion documentaire par rapport à la gestion de contenu et d'apporter une aide au choix pour vos projets.

Il ne s'agit ni d'un répertoire de solutions ni d'une approche théorique de la gestion de contenu documentaire, mais bien d'une confrontation de la réalité des besoins avec le marché de la GED (Gestion Electronique de Documents) open source.

En effet, si les solutions open source répondent très bien à de nombreux besoins, il n'en demeure pas moins que le savoir-faire spécifiquement documentaire reste, avant les outils, le principal enjeu des projets de Gestion Electronique de Documents et de Gestion de Contenus d'Entreprise (ECM).

Comme les autres livres blancs publiés par Smile, cet ouvrage s'efforce de réunir :

- Une approche générale des notions sous-jacentes à la gestion documentaire, fondement méthodologique de nombreux projets de GED.
- Une description des grandes fonctions attendues dans ce type de projet et de leurs enjeux
- Un recensement des principales solutions open source de gestion de contenu documentaire
- Une présentation des meilleurs outils, une évaluation de leurs forces, de leurs limites, de leur maturité.

Les informations transmises dans ces pages sont le résultat de retour d'expériences collectifs mais non exhaustifs et nous serions très heureux d'échanger avec vous sur le sujet.

## **Les bénéfices d'une GED open source**

L'open source gagne chaque année de nouveaux domaines d'application, dans une extraordinaire dynamique. De nouveaux acteurs apparaissent, les éditeurs open source, et la pertinence de ce business model est aujourd'hui démontrée. Les offres de solution sont de plus en plus matures, et sont de vraies alternatives aux solutions historiques, propriétaires.

Avant de faire un focus sur les fonctionnalités, voyons quels sont les

## GED Open source

---

bénéfices de ces solutions qui sont liés à leur caractère open source.

Bien sûr, les bénéfices économiques sont parmi les premières raisons dans le choix de solutions open source. Même si implémenter une solution open source ne signifie pas un coût de projet nul, ces solutions sont toujours sensiblement moins chères que leurs équivalents propriétaires. Le rapport de prix peut être de 1/5 à 1/10.

D'autant que les prix de prestations tendent aussi à être moins élevés, car l'ouverture du produit tend à faciliter la diffusion de la connaissance.

Mais au fur et à mesure que ces solutions arrivent à maturité, le moindre coût n'est plus le premier critère de choix.

Les principaux arguments sont alors :

- La non-dépendance, ou moindre dépendance, par rapport à un éditeur. On sait que changer d'outil peut coûter très cher, et les éditeurs peuvent être tentés de profiter de la vache à lait que constituent ces clients devenus captifs. En anglais, on parle de vendor lock-in, le verrouillage par le fournisseur.
- La pérennité est un autre critère de choix fort. Certes, les solutions open source n'ont pas une garantie d'éternelle jouvence. Mais la diffusion de leur source (leur code non compilé) à une communauté de client et de partenaires garantie une large diffusion du savoir-faire associé à leur implémentation et donc leur évolutivité
- L'ouverture est également un argument de poids. Les solutions open source sont en général plus respectueuses des standards, et plus ouvertes vers l'ajout de modules d'extension.

En matière de pérennité, le pire qu'il puisse arriver pour une solution open source est une désaffection progressive de la part des communautés, généralement au profit d'une solution plus prometteuse. Ainsi, il est possible qu'il faille un jour changer de produit. Mais du moins le phénomène est toujours lent, et le client a le temps d'organiser la migration.

Il faut souligner aussi que, même si l'éditeur original était un jour défaillant, il resterait toujours possible pour une communauté de reprendre en main le produit et ses évolutions, c'est le principe des licences open source.

Un mot également sur la question de l'ouverture. La possibilité de faire des modifications dans les sources est fondamentale sur le plan théorique, mais risquée sur le plan pratique. Ce n'est donc pas en ces termes qu'il faut apprécier l'ouverture, mais plutôt dans la capacité à accepter des extensions, ou à s'interfacer à d'autres applications.

Sur le fond, il faut comprendre qu'un éditeur à vocation commerciale n'a

## GED Open source

---

pas que des intérêts convergents avec ceux de ses clients. Certes, il évolue dans un marché concurrentiel, et son produit doit être meilleur que ses concurrents. Mais une fois sa position bien assise, l'éditeur peut faire l'analyse que :

- Son produit doit être performant, mais pas trop, car s'il faut plus de serveurs, ce sera plus de licences vendues.
- Son produit doit être robuste, mais pas trop, car il faut continuer à vendre du support.
- Son produit doit être ouvert, mais pas trop, pour garder la maîtrise du client.

Nous ne disons pas que les éditeurs propriétaires seraient machiavéliques au point de dégrader ces qualités dans leur produit, nous disons seulement que la priorité stratégique n'est pas nécessairement mise sur ces qualités.

## GED Open source

### Sommaire

<b>PRÉAMBULE.....</b>	<b>2</b>
Smile.....	2
Quelques références.....	3
Ce livre blanc.....	5
Les bénéfices d'une GED open source.....	5
Sommaire.....	8
<b>GÉNÉRALITÉS.....</b>	<b>9</b>
Pourquoi une solution de gestion documentaire ?.....	9
Présentation des concepts.....	10
De la GED à l'ECM.....	14
Enjeux principaux de la GED.....	16
<b>LES SOLUTIONS DE GESTION DOCUMENTAIRE.....</b>	<b>21</b>
Alfresco.....	21
Nuxeo EP.....	23
Knowledge Tree.....	24
Jahia.....	25
Maarch.....	27
Contineo.....	28
Freedom.....	30
Les autres.....	31
<b>LES FONCTIONNALITÉS.....</b>	<b>32</b>
Métadonnées.....	32
Gestion des versions.....	36
Référentiel de classement.....	38
Moteur de recherche.....	43
Intégration bureautique.....	46
Numérisation.....	48
Gestion des droits.....	51
Workflow.....	52
Règles de gestion.....	55
Gestion du cycle de vie.....	57
Import/export.....	59
Gestion des emails.....	60
Gestion de dossiers.....	62
Intégration à l'infrastructure du SI.....	64
Configuration de l'interface.....	66
Synthèse.....	67



## GENERALITES

### Pourquoi une solution de gestion documentaire ?

La décision de mettre en œuvre une solution de gestion documentaire résulte encore souvent d'une crise, d'une tension dans la gestion des documents. Une situation qui nécessite la mise en place d'une **organisation plus structurée du travail** et d'une **traçabilité** autour des documents que seule l'informatique est à même de fournir.

Les crises peuvent se manifester de bien des façons, par exemple par l'impossibilité de retrouver un document important en version électronique, lorsque sa version papier n'est plus disponible. **Il y a donc des pertes de mémoire au niveau de l'organisation.** Alors que les tensions peuvent résulter de difficultés chroniques de réutiliser la version électronique d'un document papier ou de temps perdu à des tâches de recherche (mémoire), de réutilisation de documents (modèle)... **Il s'agit ici de perte de productivité.**

La gestion documentaire fait partie des **processus qualité** (ISO 9001, 14001) et si l'utilisation d'un logiciel de gestion documentaire n'est pas requise pour s'y conformer, elle permet d'appliquer les normes en matière de gestion documentaire au-delà de la stricte application des critères qualité.

Le retour sur investissement est difficile à calculer pour ce type d'application de gestion car, d'une part il repose sur l'évaluation de nombreux critères immatériels et d'autre part l'outil vient presque toujours en remplacement d'un mode de fonctionnement existant, et seule une évaluation avant/après, sur des critères difficiles à quantifier, permettrait une évaluation du différentiel.

## Présentation des concepts

### Définitions de base

#### ♦ Document électronique

La notion de document, si elle ne pose guère de problème dans un univers matériel, est largement remise en question par la dématérialisation et la reprographie.

Notre approche de la gestion documentaire prend l'acception suivante du terme « document » : « **Ensemble constitué par un support et par l'information qu'il porte** ».

Et en matière de documents électroniques, le *support* est un *fichier*. Le document-fichier est créé, modifié et rendu lisible par une *application*. Typiquement, un document Word est créé, modifié et rendu lisible par l'application Microsoft Word, et de la même manière pour les documents ODF (OpenOffice), Autocad, ou Adobe Première.

Cette association document / application est omniprésente dans la GED. Ceci d'autant plus que beaucoup des documents manipulés utilisent des formats *fermés*, c'est à dire qu'ils ne peuvent être créés, modifiés et même lus *que* par leur application associée. De sorte que l'application de GED gère finalement des objets qui lui sont le plus souvent opaques, impénétrables.

Un document électronique est avant tout un fichier, mais n'importe quel fichier n'est pas un document. Typiquement, un fichier exécutable n'est pas un document. Un document est un fichier porteur d'une information, une information produite par l'homme et destinée à l'homme.

Nous verrons plus loin que les applications de GED introduisent la notion de document *objet*, qui réunit conceptuellement le document *fichier* et l'ensemble des méta-données qui lui sont attachées.

#### ♦ Contenu

« Contenu » est employé pour désigner **un composant informatif cohérent**.

Le terme de contenu s'applique à toute information numérique structurée ou non. Un *document* est donc un *contenu*.

Un contenu peut être constitué d'un ensemble de contenus élémentaires, incluant des documents fichiers (images, vidéos, documents bureautiques,

## GED Open source

---

...), ainsi que des champs d'information, que ce soit textuelle, numérique, ou d'autres types. On parle alors de contenus *structurés*.

Quelques exemples de contenus : Une brève sur un site d'actualité, une image et sa légende, une fiche descriptive de produit, une notice de document, un fichier PDF...

Il est donc parfois difficile de faire la différence entre un contenu, un ensemble de contenus, un contenu rédactionnel et un document électronique. **C'est souvent le contexte d'utilisation et les traitements que l'on souhaite réaliser** qui permettent de distinguer la nature de l'information et le domaine fonctionnel concerné ; sur cette base le meilleur type d'outil pour la gérer.

### Domaines fonctionnels

Nous nous efforçons ici de positionner la gestion documentaire vis-à-vis d'un certain nombre de domaines connexes. Ces domaines ont des intersections avec la gestion documentaire, et peuvent parfois être complémentaires. Il est donc utile de bien les situer.

#### ♦ Gestion documentaire

**Objectif** : Prendre en charge le stockage, le partage et la restitution de documents électroniques. Il s'agit majoritairement de systèmes gérant des fichiers incluant un fond et une forme.

**Priorité** : Gestion de documents électroniques, donc de fichiers.

**Orientation** : Stockage, partage, recherche

**Termes rencontrés** : GED - Gestion Electronique de Document, EDM - Electronic Document Management, DMS – Document Management System

#### ♦ Gestion de contenu Web

**Objectif** : prendre en charge la rédaction, la validation et la mise en ligne de contenus à destination de sites web

**Priorité** : Mise en ligne d'information

**Orientation** : Diffusion, restitution de contenus structurés et/ou éditoriaux, contribution au second plan.

**Termes rencontrés** : WCM - Web Content Management, CMS – souvent employé pour désigner la gestion de contenu Web

#### ♦ Workflow

## GED Open source

---

**Objectif** : permettre de fluidifier des processus, via l'enchaînement de tâches unitaires entre différents acteurs

**Priorité** : Gestion des processus

**Orientation** : gestion de tâches attachées à des profils d'utilisateur (« Taskflow », par exemple Organisation d'un séminaire) ou à des documents (« Docflow », typiquement la validation d'un document)

**Termes rencontrés** : BPM – Business Process Management

### ♦ Collaboratif

**Objectif** : Faciliter le travail de groupes de personnes au moyen d'outils dédiés, souvent de communication ou d'interaction en temps réel ou différé (par exemple : email, agenda partagé...)

**Priorité** : Facilitation du travail en équipe

**Orientation** : Partage d'information opérationnelle

**Termes rencontrés** : Groupware – travail en équipe (outil ou méthode), Chat – messagerie instantanée, Blog – Mode de communication personnel via des billets thématiques, auxquels il est possible de réagir, Wiki – Mode de partage d'information via la coédition de pages, avec traçabilité des modifications, Agenda partagé – possibilité de consulter, voire d'interagir avec des plannings

## Applications intégrées

### ♦ Portail

**Objectif** : Faciliter l'accès à l'ensemble des applications du système d'information.

On trouve des portails spécialisés tels que des « portails documentaires » qui offrent un point central pour accéder à des ressources homogènes dans le cadre de sources d'informations très réparties.

**Priorité** : Point d'accès unique

**Orientation** : Agrégateur de contenus et d'applications

**Termes rencontrés** : Portail ou Portail d'entreprise, Portail documentaire, Intranet – ils sont parfois dotés de ces caractéristiques, Hub – point de regroupement d'information thématique

### ♦ Gestion de contenu

**Objectif** : Permettre la gestion des contenus numériques, aux travers des

---

## GED Open source

---

étapes de conception, diffusion, utilisation et fin de vie. On peut considérer la gestion de contenu (Content Management) comme un sur-ensemble pour la GED et le WCM, on parlera le plus souvent de Gestion de contenu d'entreprise (ECM – entreprise content management). Il inclut des fonctions de gestion, de type workflow et travail collaboratif par exemple.

**Priorité** : Tous contenus numériques

**Orientation** : Prise en compte de contenus au sens large, c'est-à-dire l'intégration des outils de GED, de WCM, parfois même de portail.

**Termes rencontrés** : CMS – Content Management System, ECM – Entreprise Content Management

La gestion de contenu d'entreprise (ECM) est donc le concept le plus abouti d'intégration de gestion d'information numérique puisqu'elle adresse à la fois l'information structurée et non structurée, sur l'ensemble des étapes de leur cycle de vie.

A noter qu'on rencontre parfois le terme de gestions de contenu dans un sens plus limité, en voulant dire « gestions de contenu web ».

### ♦ Photothèque et Gestion du Patrimoine Numérique

**Objectif** : Gérer des documents numériques principalement constitués d'images, musiques ou vidéos.

**Priorité** : Contenus images et sons, navigation et recherche au sein du référentiel, gestion spécifique des droits d'auteurs

**Orientation** : Métadonnées, navigation, recherche,

**Termes rencontrés** : DAM – *Digital Asset Management*

Il s'agit d'une spécialité au sein de la GED. De nombreuses fonctionnalités sont communes, quant à la catégorisation, la gestion des droits, des workflows et cycles de vie. Les métadonnées spécifiques à la gestion des droits d'auteurs peuvent en général être modélisées avec n'importe quel outil de GED. Néanmoins, des fonctionnalités spécifiques concernent la restitution et la navigation.

## De la GED à l'ECM

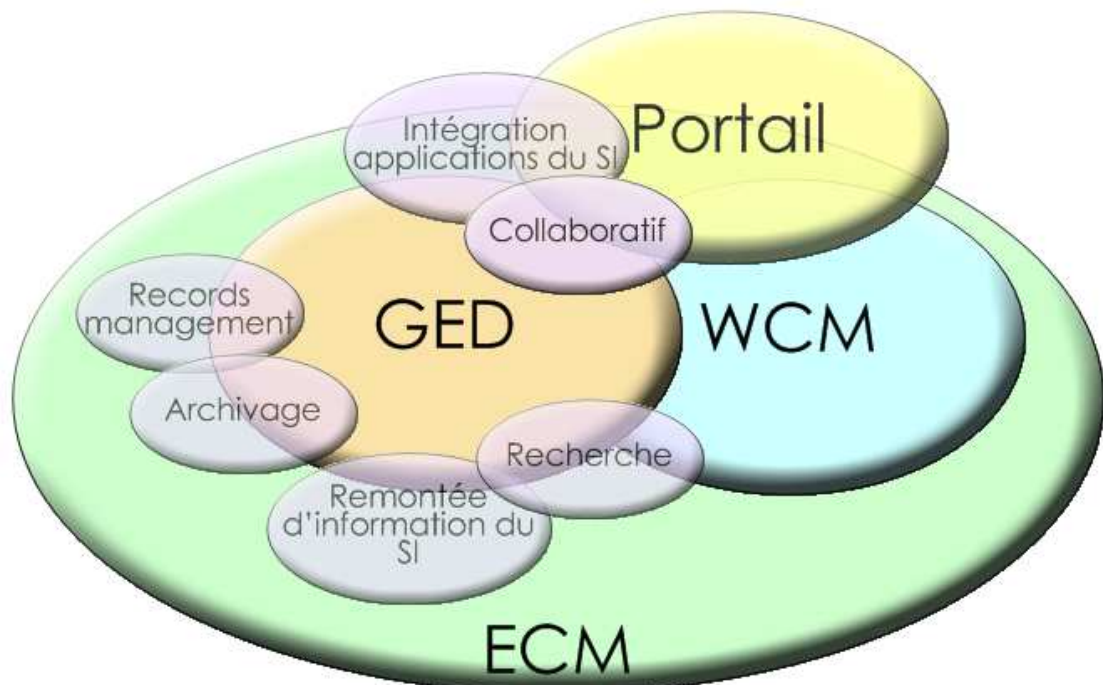
C'est-à-dire : de la « gestion documentaire » à la « gestion de contenu d'entreprise ».

### Le périmètre de la GED

La Gestion Electronique de Documents est un ensemble aux contours imprécis, qui peut recouvrir :

- **Des outils** : pour la capture des fichiers documents, leur stockage, leur circulation, leur mise à disposition, la recherche...
- **Des métiers** : traitement sur les documents, description documentaire, application des procédures qualité, gestion du cycle de vie de l'information, interaction avec des applications métier.
- **Des domaines d'application** : documents bureautiques, fichiers multimédias, pages Web, contenus structurés, groupe de travail...

Les applications de GED « pures » n'existent quasiment plus : elles intègrent presque toujours des fonctions empruntées aux domaines évoqués plus haut ou à des domaines métier.



Ce graphique vise à positionner les domaines de la GED, du WCM et de l'ECM les uns par rapport aux autres et vis à vis d'un certains nombres de domaines connexes.

---

## GED Open source

---

Certains des domaines cités n'ont pas encore été définis et sont abordés plus loin.

### Evolutions du concept de GED

Les solutions de GED étant orientées vers la gestion de fichiers, puis vers la prise en compte spécifique des documents, leur évolution naturelle est de s'intégrer de mieux en mieux aux processus de *gestion* documentaire, notamment par une meilleure prise en compte du contexte d'édition et du contenu des fichiers document (MS Word, Open Office, PDF...

La gestion des contenus d'entreprise (ECM) couvre le périmètre fonctionnel de la gestion documentaire (GED) et y ajoute des outils complémentaires, tels que la gestion des contenus structurés (WCM), la prise en compte de circuits de validation élaborés (BPM) et des aspects collaboratifs (Groupware).

Par exemple :

La sollicitation d'experts géographiquement éloignés passe souvent par des demandes de contribution sur des documents. Les échanges de mails, largement adoptés depuis une dizaine d'années, révèlent un certain nombre de failles, notamment pour la consolidation des remarques et le stockage. Les fonctions de travail en groupe intégrées aux applications de GED permettent de faciliter et de fiabiliser ce type d'utilisation.

L'évolution de la couverture fonctionnelle des solutions de GED tend à résoudre ces différents problèmes. D'une part en intégrant la gestion des contenus structurés en parallèle de celle des contenus documentaires et d'autre part en permettant de réaliser des actions collaboratives poussées.

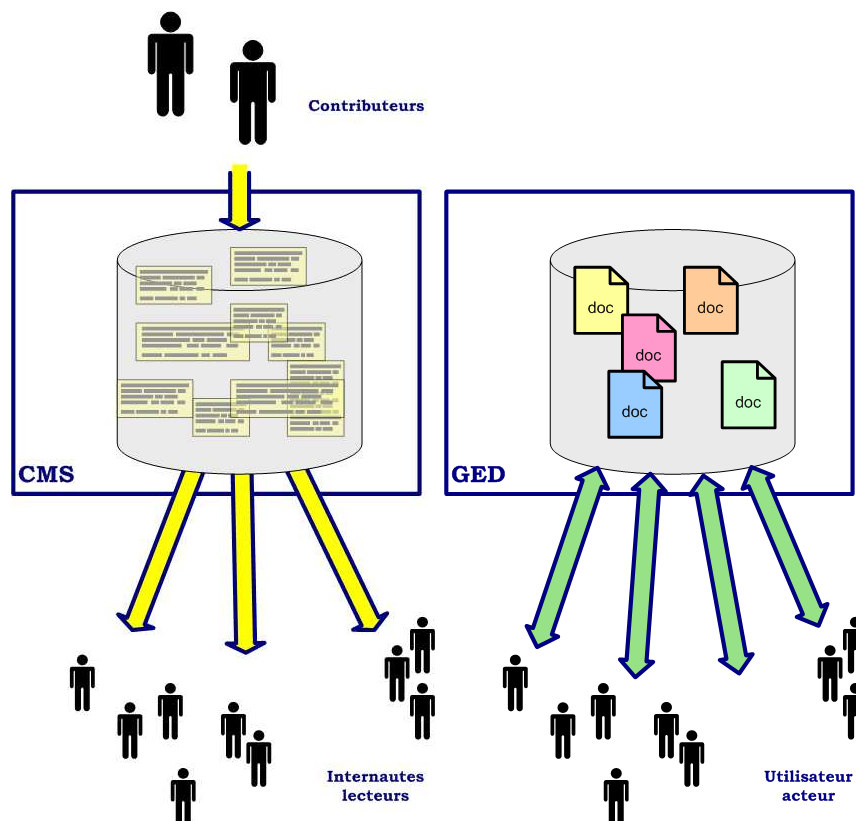
La **notion de portail d'entreprise** est fréquemment associée à celle d'ECM. Fonctionnellement, les intersections sont faibles, mais les solutions d'ECM sont souvent intégrées au sein d'un portail d'entreprise. Elles bénéficient alors d'une gestion centrale des utilisateurs et profils, avec dispositif de Single Sign-On. Le portail permet de réaliser des pages agrégées réunissant des contenus issus de la GED d'une part, des contenus de type WCM d'autre part, et enfin des contenus issus de différentes applications métier. La capacité à s'intégrer à un portail est donc une des qualités d'une solution ECM ou GED.



## Enjeux principaux de la GED

La gestion documentaire (GED) est un domaine avant tout organisationnel, alors que la gestion de contenu Web (WCM) relève essentiellement de problématiques techniques.

Les **outils de gestion de contenu Web** s'orientent essentiellement vers la diffusion de contenus structurés : leur mise en forme, leur publication, leur facilité d'accès par un large public, etc.. Alors que les **outils de gestion documentaire** mettent l'accent sur les processus de gestion en amont comme la possibilité de réaliser une indexation riche ou d'intégrer des règles de traitement (workflow, transformation, alertes conditionnelles), par exemple.



La mise en place d'une solution de GED repose donc principalement sur les fonctions de gestion, et sa bonne utilisation sur un certain nombre d'outils et de principes décrits ci-dessous.



## GED Open source

---

### Structuration des métadonnées

La gestion des documents électroniques s'apparente encore souvent à une gestion en aveugle. En l'absence de formats ouverts (voir plus loin), le document-fichier est un objet opaque et impénétrable pour toutes les applications autres que celle qui l'a créé, et donc pour l'application de GED, qui ne connaît au départ du document que son nom (le nom du fichier) et son type (type MIME).

Dès lors et pour permettre de disposer de fonctions avancées, un système de GED doit permettre d'associer aux documents des informations structurées exploitables. C'est tout l'enjeu des métadonnées.

Les métadonnées sont les informations associées aux documents gérés. Elles servent à qualifier les documents, leur apporter une information complémentaire qui soit exploitable. La métadonnée est directement utile en tant qu'information, mais surtout elle est la base des fonctions de recherche, ou de traitements sélectifs appliqués aux documents.

Les métadonnées typiques incluent par exemple le titre, l'auteur, la description, la langue, la date de publication, etc.

Il existe un standard de nomenclature des métadonnées, le Dublin Core, qui définit 15 champs principaux de métadonnées. C'est en général insuffisant, mais du moins le respect du standard est un gage d'interopérabilité.

Afin d'appréhender un maximum de besoins, les solutions doivent donc gérer différents jeux (ou *sets*) de métadonnées en fonction des types de documents et prévoir un large éventail de types de métadonnées, simple stockage d'information textuelle (champ texte), mais aussi utilisation de listes de valeurs issues de tables de références mises à jour dynamiquement depuis une autre application.

### Classement

La plupart des solutions de gestion de contenu (CMS), que ce soit de gestion documentaire (GED) ou de gestion de contenu Web (WCM), intègrent des **outils de gestion de référentiel de classement**.

Dès lors que l'on parle d'organisation professionnelle, de métiers spécifiques et de procédures de gestion documentaire, **il est indispensable de disposer d'une structure** qui serve de **colonne vertébrale** au système de gestion documentaire – c'est le rôle du référentiel de classement. Il s'agit de l'un des outils les plus importants de la gestion documentaire, il ne devrait jamais être négligé dans le choix d'une solution.

Les outils sont plus ou moins élaborés, mais permettent a minima de définir un plan de classement c'est-à-dire une arborescence hiérarchique par

---

## GED Open source

---

rapport à laquelle sont classés les documents. Ils peuvent aller jusqu'à permettre de gérer des vocabulaires métiers avec synonyme, liens de proximité sémantique, hiérarchie et traduction de termes, on parlera alors de thésaurus.

L'utilité d'un référentiel de classement concerne différentes fonctions : aide à l'indexation des documents, mise à disposition de vocabulaires communs (glossaires), recherche avancée (avec prise en compte des synonymes, de la proximité sémantique, des liens entre concepts...), navigation par catégorie.

### Gestion des droits

La gestion des droits ou habilitations des utilisateurs doit pouvoir être réglée finement, le plus souvent en interaction avec l'annuaire d'entreprise. En effet, les utilisateurs d'un système de GED peuvent avoir différents profils, en fonction de leur appartenance à un service, à un niveau hiérarchique, mais également de leur rôle par rapport à un processus documentaire particulier, par exemple contributeur, responsable d'un dossier...

Les habilitations portent sur des actions possibles vis-à-vis des processus, mais aussi sur la consultation des documents, et donc les questions de confidentialité.

La gestion des droits doit donc pouvoir être fine, a minima par nœud d'arborescence (répertoire dans un système de fichiers), et de préférence au niveau de chaque document, voire selon les métadonnées afin de gérer des problématiques de confidentialité pointues.

### Format de fichier

Le format des fichiers gérés dans une application documentaire revêt une importance particulière. En effet, la plupart des applications de GED peuvent stocker n'importe quel type de fichier, cependant certaines fonctions ne sont disponibles que pour certains formats.

Parmi ces fonctions, on compte notamment :

**La recherche en texte intégral**, c'est-à-dire dans les fichiers eux-mêmes. On l'a vu, les formats documentaires sont majoritairement fermés et opaques. Néanmoins, il existe souvent la possibilité d'en extraire un contenu textuel, qui pourra être indexé. Les moteurs d'indexation ne savent indexer que des formats de fichier spécifiques. Les formats les plus courants (PDF, MS Office, HTML) figurent presque toujours parmi ceux reconnus, mais pour certains besoins, il peut être nécessaire d'avoir un moteur sachant indexer des formats particuliers, par exemple DWG (AutoCAD), mp3 (pour les tags textuels) ou d'autres formats de fichiers propriétaires.

## GED Open source

---

**La prévisualisation**, c'est-à-dire la possibilité de visualiser un fichier sans l'ouvrir avec son application associée.

Cette notion est particulièrement utile pour les fichiers très lourds (image, vidéo) puisque l'application permet de visualiser rapidement une partie du contenu. Cette fonction peut être étendue par des plug-ins (pour navigateur internet) ou des applications dites de visualisation universelle.

**L'extraction d'information**, en particulier des métadonnées. Chaque format de fichier utilise une façon spécifique de gérer les métadonnées qu'il embarque.

### ♦ **Formats ouverts**

Les formats ouverts sont une question d'actualité avec la montée en puissance de la standardisation des formats bureautiques. En effet, ce phénomène que l'on observe depuis plusieurs années dans les domaines techniques, où le besoin d'interopérabilité est très fort, devient une question à la convergence de nombreux projets.

Il existe deux enjeux principaux pour un projet de GED.

Le premier est de disposer d'un format libre et très largement répandu pour garantir la **pérennité des fichiers**.

Le second est de disposer d'un format qui permet la manipulation du contenu (l'édition), parce qu'il est ouvert et documenté. Ceci permet d'offrir **des fonctions très riches** comme l'import et l'export de contenu, des conversions et des transformations, des combinaisons de plusieurs documents, l'extraction de métadonnées, de tables des matières, de paragraphes spécifiques...

### ♦ **Formats de conservation**

Si les documents stockés dans le système documentaire doivent être conservés plusieurs années (au-delà de 5 ans) leur conservation doit garantir à minima qu'ils seront lisibles, voire même intelligibles.

En effet, la problématique de la conservation se trouve de plus en plus posée pour les documents électroniques dans la mesure où ils remplacent le papier alors qu'ils ne faisaient que le compléter il y a encore quelques années. Le **format choisi pour cette conservation est donc important**, cf. « Cycle de vie et archivage », page 20.

Dans le cas où ces documents sont porteurs de **signatures électroniques**, la problématique s'en trouve complexifiée (voir « Archivage électronique à valeur probante », page 58).

Les formats de fichier à même de garantir la lisibilité des documents sur plus de dix années sont soit des formats simples et ouverts, par exemple les

---

## GED Open source

---

fichiers XML pour les données structurées, soit les formats normalisés comme le PDF/A, le (futur) PDF/E ou l'ODF (Open Document Format).

### **Cycle de vie et archivage**

Pour optimiser la gestion des ressources documentaires, il faut gérer leur cycle de vie qui s'étend de leur création jusqu'au-delà de leur utilité opérationnelle.

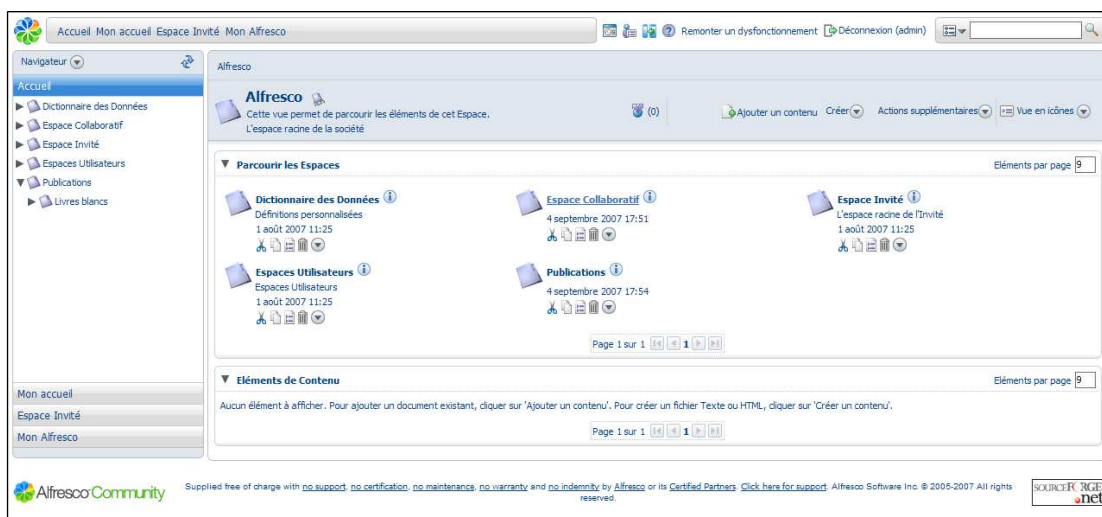
Un document est d'abord conçu, éventuellement de façon collaborative, en mettant en œuvre un ensemble de processus avant de pouvoir être utilisé. Par exemple, un contrat est rédigé, validé, signé puis envoyé, souvent sous forme papier. Le cycle de vie inclut la gestion de l'applicabilité des documents.

## LES SOLUTIONS DE GESTION DOCUMENTAIRE

### Alfresco

Alfresco<sup>1</sup> est un éditeur anglais fondé en 2005 par d'anciens dirigeants de Documentum et Business Object.

C'est une solution complète qui propose l'ensemble des fonctionnalités attendues : métadonnées, types de documents, workflow, collaboration, gestion des versions, recherche, intégration bureautique...



A noter que Alfresco propose également un composant puissant de gestion de contenu web. Cette partie est cependant un peu moins mature que le reste de l'application et ne permet pas d'interaction poussée entre les documents et les contenus web.

Le comportement d'Alfresco est entièrement paramétrable et l'ajout de nouvelles fonctionnalités est possible au moyen d'APIs de différents niveaux : du simple script au composant Java, en passant par des web services.

Alfresco est distribué en deux versions : Une version « Community »,

<sup>1</sup> <http://www.alfresco.com>

---

## GED Open source

---

gratuite sous licence GPL, et une version « Entreprise », qui nécessite une souscription annuelle et donne accès à la garantie de l'éditeur ainsi qu'aux mises à jour intermédiaires.

Cette solution dispose d'une forte dynamique de développement et de communication ainsi que d'une communauté d'utilisateurs et de développeurs importante. On notera l'existence du site Alfresco Forge<sup>2</sup>, qui héberge de nombreux plug-ins.

De nombreuses améliorations sont prévues à court terme dans la feuille de route, parmi lesquelles :

- La notion de réseaux pour les groupes de travail et la collaboration
- Des espaces projets : avec tableaux de bord et calendriers par projet
- Un mécanisme de discussion par mail
- Des indicateurs de présence des acteurs du projet
- Un suivi de l'activité (qui fait quoi dans mes groupes d'intérêt)
- Le support de OpenSocial
- Des blogs par espace
- Une nouvelle interface en technologie Flex, accompagnée d'un SDK

Domaines : **GED**, **WCM**

---

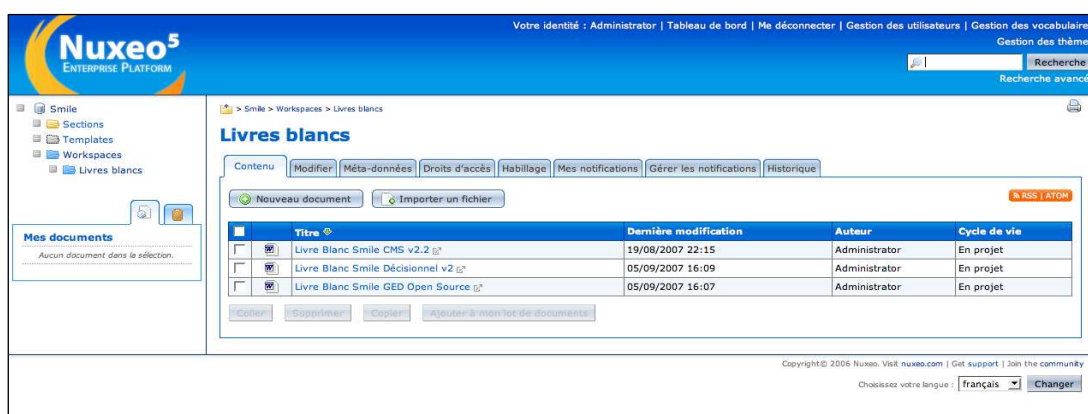
<sup>2</sup> <http://forge.alfresco.com/>

## GED Open source

### Nuxeo EP

Nuxeo est une société française qui produit une solution de GED open source depuis 2001.

Autrefois éditeur de la solution Nuxeo CPS en Zope/Python, Nuxeo a migré sa solution de GED en Java début 2007. Il s'agit d'une solution complète, en environnement J2EE : métadonnées, types de documents, workflow, collaboration, gestion des versions, recherche, avec cependant une intégration bureautique assez légère pour le moment.



L'outil se distingue par son éditeur de thèmes, entièrement graphique qui permet de personnaliser son interface, et par son architecture totalement modulaire qui facilite le développement, la maintenance et la réutilisation de fonctionnalités additionnelles.

Cette solution est relativement jeune, mais bénéficie des années d'expérience du produit Nuxeo CPS. La première version de Nuxeo EP (5.0), à laquelle il manquait certaines fonctions importantes, est sortie début 2007. La version 5.1 bien plus complète est sortie fin août 2007.

De nombreuses évolutions sont prévues à court terme dans la feuille de route, telles que :

- la mise à disposition d'une technologie de widgets (gadgets Web)
- des outils de gestion des formats vidéo
- l'introduction d'outils collaboratifs tels que les blogs et wiki qui permettront donc de gérer des contenus Web et d'outil de communication instantanée
- la détection de doublons
- la nouvelle version de LiveEdit qui améliorera l'intégration à MS Office et OpenOffice

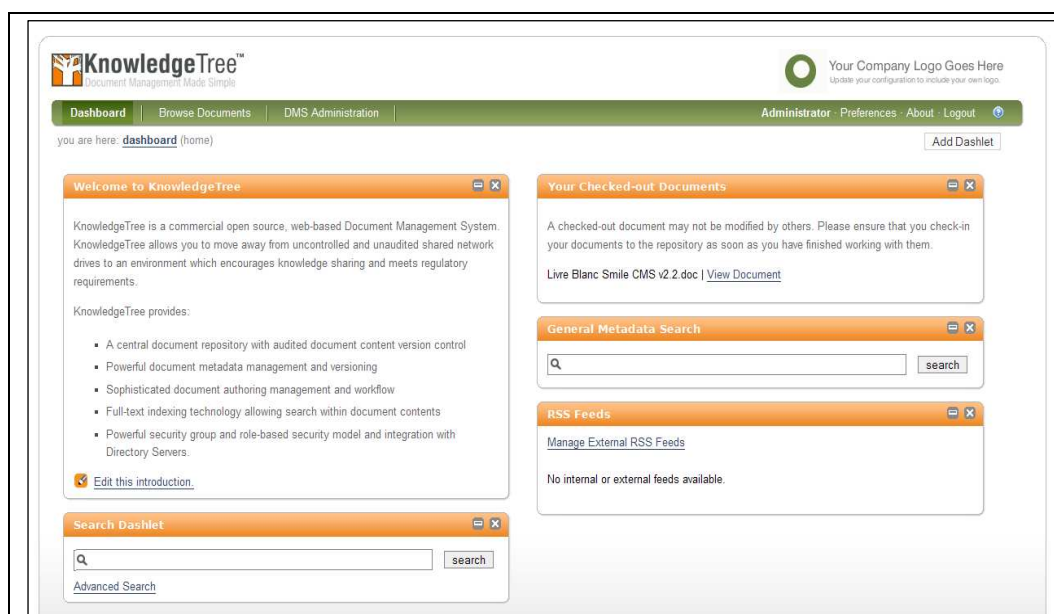
## GED Open source

- la synchronisation entre bases documentaires (avec SyncML)

Domaine : **GED**

### Knowledge Tree

Knowledge Tree<sup>3</sup> est une solution de GED développée par la société Sud Africaine JamWarehouse. Elle dispose d'un ensemble complet de fonctionnalités : métadonnées, workflows, gestion des versions, recherche ainsi que de plusieurs applications qui permettent une bonne intégration dans l'environnement bureautique : explorateur de référentiel, intégration à Microsoft Office, gestion des mails depuis Outlook, et application de gestion de scanners.



Cette solution est disponible sous plusieurs licences, selon les fonctionnalités souhaitées et le nombre d'utilisateurs.

Une première édition, open source et gratuite, est proposée sans support, sans l'intégration aux outils bureautiques et sans la gestion des scanners.

Les deux autres éditions disposent de la totalité des fonctionnalités, de support de la part de l'éditeur, et se différencient uniquement par le nombre d'utilisateurs et de processeurs autorisés.

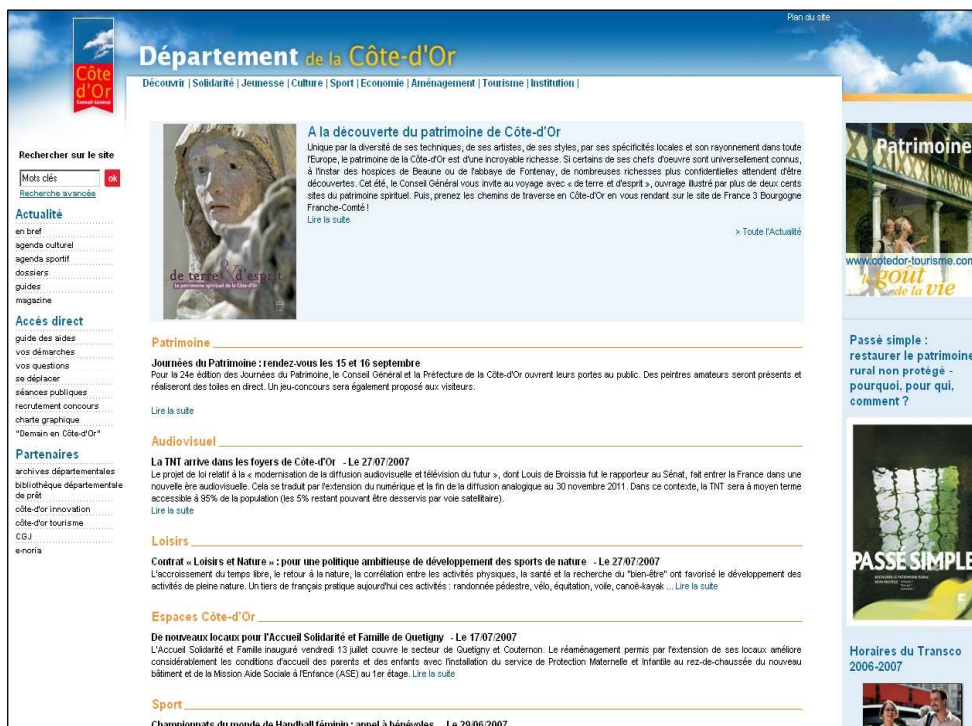
Domaine : **GED**

<sup>3</sup> <http://www.knowledgetree.com/>



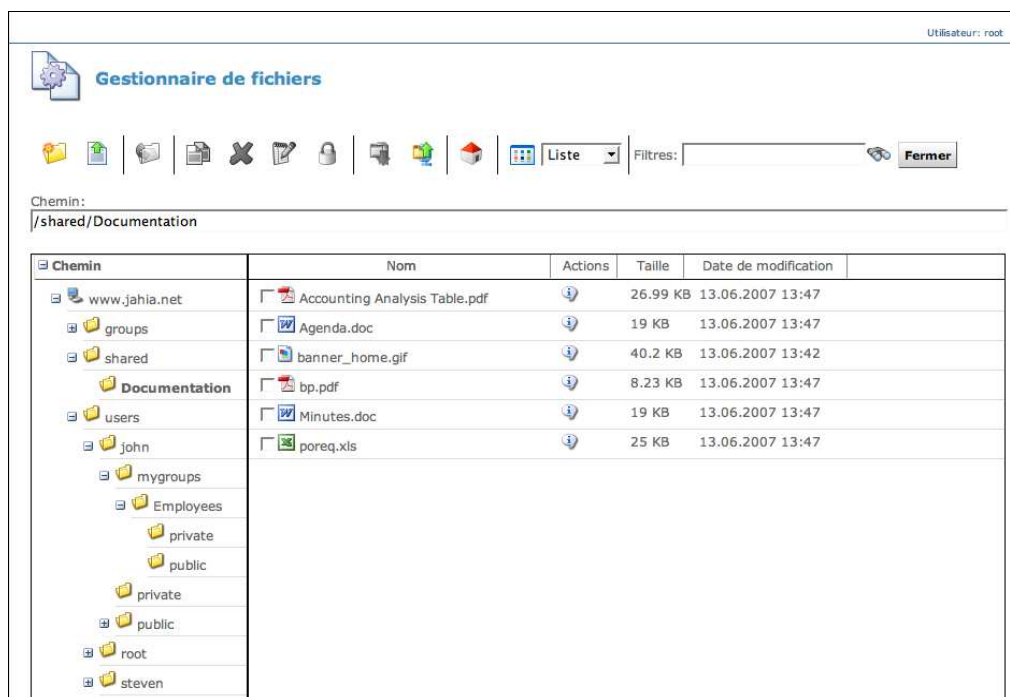
## Jahia

Jahia est une solution intégrée de portail web et de gestion de contenu. Proposée sous forme de produit commercial, cette solution répond à la fois aux besoins de portail (agrégation de différentes briques applicatives), notamment avec la norme JSR 168, et aux besoins poussés de gestion de contenu web (contenus structurés, workflows, édition directe des contenus, multilinguisme, catégorisation...)



Jahia permet également de gérer un référentiel de fichiers qui peuvent être liés aux différents contenus web, le tout avec une bonne intégration à l'environnement bureautique grâce à l'utilisation d'un accès WebDav. Cependant, les fonctionnalités de GED proprement dites sont rares (pas de gestion des versions ou de réservation, métadonnées différenciées pour les fichiers, par exemple). La médiathèque est différenciée de la base des contenus Web et la prééminence est nettement à la seconde.

## GED Open source



De nombreuses évolutions autour de la gestion documentaire sont prévues pour 2008, notamment :

- intégration d'un repository JSR-170 qui permettra l'accès via CIFS et FTP
- amélioration des fonctions de recherche sur les documents, possibilité d'enregistrer des requêtes
- intégration de composants collaboratifs plus complets
- intégration progressive des problématiques de record management
- navigation multidimensionnelle parmi les fichiers
- moteur de règles évoluées, orientés vers la gestion documentaire

D'ici là, les domaines d'excellence sont le portail d'entreprise, incluant de la gestion des contenus web avancée, incluant des documents plutôt que la GED proprement dite.

Domaines : **WCM, Portail**

## Maarch

Maarch<sup>4</sup>, est une solution développée par la société Maerys. L'offre comporte deux applications : Maarch, dédiée à l'archivage, et Maarch Letter Box, dédiée à la gestion dématérialisée du courrier.

Ce sont des outils de GED spécialisés et simples, mais qui remplissent bien leurs objectifs.

**Maarch** permet de stocker différentes catégories de documents et inclut des fonctions de recherche, d'export et de restauration.

The screenshot shows the 'Maarch v2.1' web interface. At the top left is the logo 'Maarch v2.1 Maerys Archive'. To its right is a 'Recherche rapide' (Quick search) field with an 'OK' button. Below this is a navigation bar with a 'Menu' button and a link 'Maarch > Ajouter un nouveau document'. On the right of the navigation bar, it shows the user 'John Smith (informatique)' and a 'Déconnexion' (Logout) button. The main content area is titled 'Indexation d'un fichier' (Indexing a file). It contains several form fields: 'Collection documentaire' (Documentary collection) with a dropdown menu showing 'Accounting documents'; 'Choisissez le fichier' (Choose the file) with a text input and a 'Parcourir...' (Browse...) button; 'Titre' (Title) with a text input; 'Description' with a larger text area; 'Type du document' (Document type) with a dropdown menu showing 'Choisissez un type de document'; 'Date' with a text input showing 'AAAA-MM-JJ' and a calendar icon; and 'Référence' (Reference) with a text input. At the bottom of the form is a button labeled 'Ajouter un nouveau document' (Add a new document).

**Maarch Letter Box** s'interface avec un scanner et permet de gérer le cycle de vie du courrier dans l'entreprise : réception, validation, traitement et recherche.

<sup>4</sup> <http://www.maarch.org/>

## GED Open source

Domaines : Archivage, dématérialisation du courrier

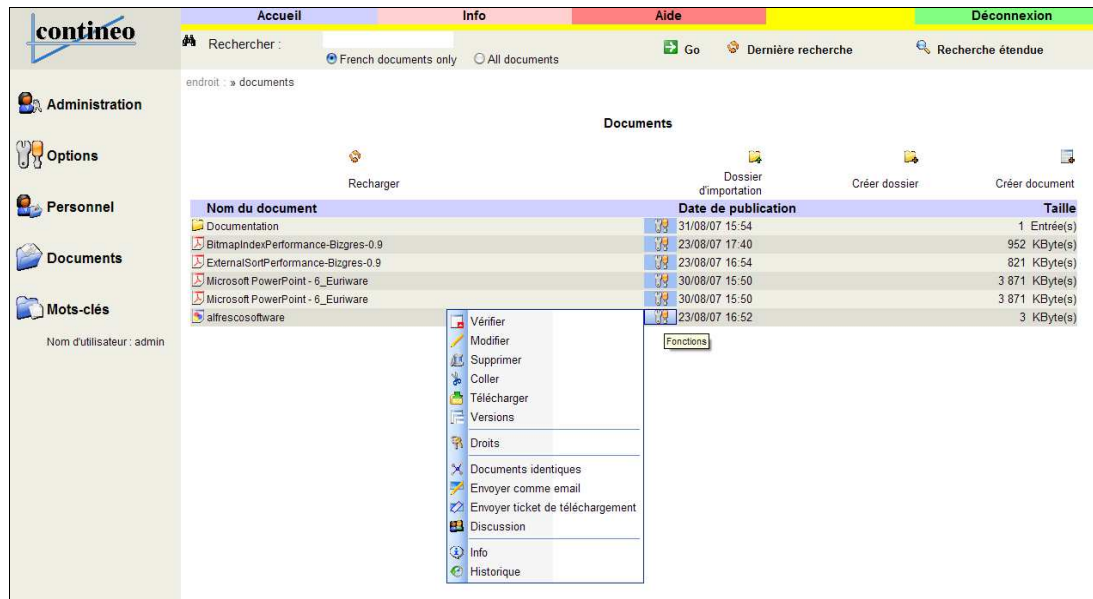
## Contineo

Contineo<sup>5</sup> est un outil de gestion documentaire dont le développement est en partie supporté par la société Logical Objects ainsi que par quelques développeurs indépendants.

Parmi les fonctionnalités proposées, on peut citer : gestion des versions, envoi de document par mail ou par lien généré, intégration d'un webmail, gestion des langues de documents, discussions.

<sup>5</sup> <http://contineo.sourceforge.net/>

## GED Open source



Si Contineo propose les fonctionnalités de base d'un outil de GED ainsi que quelques fonctions intéressantes, il présente très vite des limites :

- Il n'est pas possible d'utiliser des métadonnées différentes en fonction des types de documents, ni d'en ajouter au jeu existant.
- Pas de verrous sur les documents, ni de workflow.
- Très peu de documentation, notamment aucune documentation utilisateur
- Pas d'intégration bureautique : des web-services permettent d'interagir avec l'application, mais l'intégration aux outils du poste client reste à développer.
- Communauté réduite, peu de support et faible activité (développement, correction de bugs).

Domaine : **GED**.

## GED Open source

### Freedom

Freedom est un outil de GED, développé par la société française Anakeen, qui dispose également de fonctions collaboratives : calendrier partagé, carnet d'adresses...

Malgré une interface plutôt sommaire, l'outil est assez complet : métadonnées, types de documents, collaboration, gestion des versions, recherche, plan de classement ...

The screenshot displays the Freedom Master GED interface. On the left is a sidebar with a hierarchical tree structure including folders like 'racine', 'import', 'la poubelle', 'les cycles', 'les familles', 'les maisons', 'les profils', 'Fiches en retard', 'Freedom Master', 'Mon espace', 'Mon porte-document', 'Comptes-rendus', 'Documents verrouillés par Freedom Master', 'Les personnes', and 'Les services portail'. The main window is titled 'Gestion documentaire' and shows a list of documents under 'Comptes-rendus - 3 documents'. The selected document is 'CR/20070208/2111/6', which is a 'Compte Rendu' (Meeting Report) titled 'Revue de proposition'. The document details include: Reference: CR/20070208/2111/6, Client: Test Tony, Date de réunion: 07/02/07, Lieu: , Rédacteur: Master Freedom, Durée: , Client: , Service: , Contexte: , Date de diffusion: 08/02/2007. The 'Objet' is 'Test'. Below this, there is a table for 'Ordre du jour' (Agenda) with columns for 'Inv', 'Prés', 'Diff', 'Participants', and 'Sociétés'. The table shows 'non' for 'Inv' and 'Prés', and 'Alonso Veronique' for 'Participants'. The 'Diff' column is empty. Below the table is a 'Synthèse' (Summary) section with a 'Compte Rendu' (Meeting Report) section.

La communauté autour de cet outil est cependant très réduite, uniquement française et seules quelques références sont affichées.

Domaines : **GED, Portail, Collaboration**

## GED Open source

---

### Les autres

Plusieurs autres solutions de GED existent dans le monde open source. Leur niveau de qualité est variable et en général ces outils sont trop limités et trop peu actifs pour être viables et pour que nous les recommandions.

Dans cette catégorie, on peut citer : **DocMgr**, **OpenGED**, **myDMS**, **eDMS**, ...

## LES FONCTIONNALITES

Nous présentons dans ce chapitre les principales fonctionnalités des solutions de gestion documentaires, et pour chacune, nous indiquons comment les outils cités se positionnent.

Il s'agit de décrire non seulement la fonction en elle-même, mais aussi les impacts qu'elle a dans le contexte d'un projet.

Les meilleures applications de GED sont aujourd'hui très abouties et tendent à intégrer de nombreuses fonctions à la marge de leur domaine fonctionnel strict. De même, certaines applications d'autres domaines intègrent des « briques GED » en relation étroite avec les fonctions métiers.

Les concepteurs de solutions capitalisent donc sur des expériences acquises dans différents domaines : documentation, moteur de recherche, bureautique, numérisation, représentation de processus... Les fonctions le plus souvent intégrées aux solutions de gestion documentaire sont : la collaboration, le portail, la gestion de contenus structurés, le workflow...

Puisque les solutions assurent convenablement des fonctions essentielles, c'est donc sur leur optimisation et leur sophistication, ainsi que sur la mise à disposition de fonctions de haut niveau que va porter l'essentiel de notre attention, en vue de choisir une solution adaptée à chaque contexte.

### Métadonnées

L'indexation est une fonction centrale des outils de GED, elle consiste en premier lieu à attacher des métadonnées aux documents.

#### Types de documents

**Chaque type de document peut être qualifié par un ensemble de métadonnées** et être associé éventuellement à des règles de gestion particulières, s'appuyant sur ces métadonnées.

**Chaque document sera donc doté de métadonnées propres.** Dans certains cas, il peut être intéressant d'indexer plusieurs fichiers avec un seul jeu de métadonnées.



## GED Open source

---

### ♦ Structure des notices

On appelle parfois *notice* l'ensemble de métadonnées associées à un document, saisies souvent au moyen d'un formulaire web.

Le degré de finesse de la qualification des documents doit être ajusté au cas par cas, afin de trouver un compromis entre la richesse d'information d'une part et le travail de saisie d'autre part.

En effet, s'il est parfois contre-productif d'obliger à saisir 15 métadonnées pour un document courant, certaines utilisations peuvent au contraire nécessiter des jeux de métadonnées importants.

Ceci afin de disposer :

- des informations utiles **pour la recherche et la gestion** de chaque type de document dans le cadre du système de GED
- du maximum de métadonnées renseignées (sans les rendre toutes obligatoires)
- d'une indexation pertinente, c'est-à-dire qui reflète précisément et utilement les documents qu'elle décrit

Attention à ne pas tomber dans la surindexation (trop de métadonnées pour un type de document) ou l'indexation libre (pas assez de contrôles entraînant des métadonnées non renseignées ou mal renseignées).

### ♦ Fonctions relatives aux types de documents

Toutes les fonctions permettant de fiabiliser l'indexation et de réaliser des traitements documentaires propres à un type de document, par exemple :

- Champs multivalués, c'est-à-dire permettant de saisir plusieurs valeurs séparées
- Règles de cohérence sur un champ et entre champs, par exemple un format de date pour une date, ou un chiffre pour un prix...
- Possibilité de disposer de champs calculés, par exemple, l'URL d'accès direct au document par rapport à son titre et son emplacement
- Liaison des documents et « typage » de ces associations, par exemple, un courrier qui « a pour annexe » ou un contrat qui « concerne la même affaire » que tels autres documents...

## Indexation manuelle

### ♦ Interface de saisie

La notice (ou fiche) décrivant un document est souvent le premier point

## GED Open source

---

d'entrée des utilisateurs. C'est également ce formulaire qui sollicitera le plus l'utilisateur puisqu'il devra saisir de l'information.

Pour ce faire, et dans un souci d'ergonomie, de fiabilité et de productivité, les solutions de GED doivent permettre différentes facilités, directement ou par le biais d'une intégration légère, par exemple :

- **Copier/coller** entre le document électronique et les métadonnées. Ceci est particulièrement pertinent lorsque le document électronique est issu d'une numérisation et qu'il est possible de le visualiser sur le même écran que la fiche d'indexation
- **Aide à la saisie** dès que possible, c'est-à-dire notamment :
  - des listes de contrôle facilement accessibles qui permettent de fiabiliser le fond (sémantique) et la forme (orthographe notamment) de la saisie
  - de cases à cocher ou de bouton radio
  - des boîtes de dialogue adaptées aux tables de référence. Ceci pouvant aller de la simple liste déroulante à une fenêtre (pop-up) disposant à la fois d'une recherche et d'un outil (abécédaire, arbre) de navigation dans les valeurs de la table de référence
- **Lien entre documents** avec une interface de recherche des documents liés simple et adaptée au contexte. Ce type de lien peut être simple ou « typé », c'est-à-dire être porteur de sens comme un lien « annule » ou « est la traduction de »

### ♦ Table de référence

L'objectif est de proposer, dès que c'est possible, des listes de valeurs pour éviter que l'utilisateur ait des questions à se poser par rapport à l'indexation, mais surtout pour mieux structurer la saisie, et donc mieux utiliser les métadonnées.

Parmi les tables de références, on trouve :

- les référentiels de classement (cf « Référentiel de classement », page 38)
- les listes de valeurs enrichies par la saisie, par exemple le nom des auteurs des documents déjà dans le système

L'objectif de ces tables est d'offrir des aides à la saisie, et éventuellement des que des contraintes et contrôles sur les saisies.

## GED Open source

---

### Indexation automatique

#### ◆ **Extraction de métadonnées**

Il existe plusieurs modes d'extraction de métadonnées.

- Récupération d'information

Chaque fichier informatique est automatiquement doté d'informations issues du système de gestion de fichiers, qui peuvent être pertinentes comme métadonnées : le nom du fichier, les dates de création et de modification, sa taille, son emplacement...

Les solutions de GED proposent de récupérer certaines de ces informations. Il est également possible, notamment dans le cadre de reprise d'historique, de les travailler de façon à reconstituer une indexation par mot clé, par exemple en décomposant l'emplacement ou la structure d'un nom sous forme d'une succession de termes qui serviront à l'indexation.

- Extraction de données structurées

Certains formats de fichier disposent d'une structure lisible. C'est notamment le cas pour nombre de formats ouverts, ODF par exemple.

Lorsque ces structures sont connues et documentées, il est possible d'utiliser des automates pour extraire les informations pertinentes directement depuis le fichier et les utiliser pour alimenter une fiche d'indexation.

#### ◆ **Induction de métadonnées**

Certaines solutions de GED, souvent les plus élaborées, permettent de déterminer les métas informations les plus pertinentes pour indexer un document.

Ce type d'induction de métadonnées est souvent réalisée par :

- des outils de reconnaissance qui retrouvent dans un document les chaînes de caractère présentes dans un référentiel afin de permettre à l'utilisateur de les ajouter à la fiche descriptive
- des outils statistiques, qui analysent les chaînes de caractère les plus présentes et donc potentiellement les plus représentatives
- des outils sémantiques, capables d'extraire automatiquement les mots et expressions les plus pertinents, voire de reconnaître s'il s'agit de mot clé, de date, du titre...

Ces différentes approches sont souvent combinées.

Les outils les plus évolués permettent, au prix d'une diminution de la

## GED Open source

---

pertinence d'indexation, par rapport à l'indexation manuelle, de prendre en compte très rapidement de très grandes quantités d'informations.

**Alfresco** utilise la notion d'*aspect* pour gérer les différents types de documents. Un aspect contient un ensemble de métadonnées typées, avec ou sans contraintes, que l'on peut appliquer sur n'importe quel document. L'ajout d'un aspect et la saisie des métadonnées peuvent être manuels ou bien automatiques en se basant sur le contenu du fichier, son nom, ou encore son emplacement. Un aspect peut également modifier le comportement d'un document dans le référentiel : les fonctions d'audit et de gestion des versions sont activées par l'aspect correspondant. Les aspects sont ajoutés par configuration XML.

**Nuxeo EP** permet d'ajouter de nouveaux types de documents via des *facets*. La saisie des métadonnées est manuelle ou automatique en utilisant le contenu des fichiers. Les nouveaux types de documents ou *facets* sont créés à partir de schémas XML (XSD) et sont ajoutés en créant un nouveau plug-in (extension).

**Knowledge Tree** permet à l'administrateur de créer des types de documents et des ensembles de métadonnées directement depuis l'interface web. Les possibilités en termes de types de données et de contraintes sont plus limitées que pour les outils précédents. La saisie des métadonnées se fait manuellement.

**Maarch** permet de créer différents types de documents, qui utilisent tous le même jeu de métadonnées de base. La saisie est manuelle.

**Contineo** utilise le même jeu de métadonnées pour tous les documents. La saisie est manuelle.

**FreeDom** permet de créer des familles de documents qui disposent chacune d'un jeu de métadonnées dédié. De nombreux types sont disponibles, et les contraintes se définissent par des fonctions PHP. La saisie des informations reste manuelle.

**Jahia** permet également d'avoir des types de contenus, porteurs de métadonnées différentes, la nature de métadonnée des contenus Web et celles relatives aux documents sont sensiblement différentes. La saisie des informations est manuelle, via l'interface web, elle peut être automatisées par rapport aux fichiers PDF et MP3.

## Gestion des versions

La gestion des versions d'un document est l'un des domaines où la GED apporte un bénéfice important.

En effet, dans un contexte de travail en groupe, qui plus est avec des échanges d'emails multiples, il devient rapidement complexe de connaître

## GED Open source

---

avec certitude la dernière version d'un document.

La gestion des versions permet à la fois de tracer les évolutions d'un document et, par un système de réservation (le check-in/check-out), de garantir qu'un utilisateur peut prendre la main sur un document et le modifier dans la base documentaire sans qu'il n'y ait de modification concurrente, simultanée.

### Réservation (check-in/check-out)

Le check-in/check-out peut fonctionner techniquement de différentes façons, mais doit dans tous les cas garantir qu'un utilisateur qui fait une réservation verrouille le document jusqu'à la réalisation d'une condition. Cette condition est en général un check-in, on peut aussi envisager qu'un check-out soit libéré automatiquement au bout d'un certain temps.

Techniquement, le check-out peut être réalisé automatiquement dès qu'un utilisateur ouvre un document et le check-in réalisé dès qu'il le ferme ; mais la réservation d'un document peut également impliquer que l'utilisateur sorte le document du système le temps de la modification et donc nécessiter un système déclaratif de la part de l'utilisateur.

### Incrémentation des versions

En général, l'incrémentation des versions est un compteur automatique qui définit un numéro séquentiel aux versions successives d'un document.

On peut également avoir des méthodes de calcul de l'incrémentation plus élaborées, par exemple pour prendre en compte les notions de versions mineures/majeures, ou des règles métiers pré établies, qui définissent à quoi correspondent les numéros de version.

Dans tous les cas, l'historique des versions est conservé et il doit être possible de visualiser une version antérieure. Chaque modification de document doit donner lieu à une nouvelle version. Ces deux points sont essentiels pour la traçabilité globale du système de gestion documentaire.

Des règles de gestion particulières doivent pouvoir être déclenchées par rapport à des conditions issues du versioning, par exemple la création d'une version majeure, le temps écoulé depuis la dernière version majeure...

**Alfresco, Nuxeo EP et Knowledge Tree** proposent des fonctions avancées de réservation et de gestion de versions pour l'ensemble des documents. Notamment la distinction entre versions mineure et majeure, les commentaires de version, voir la réservation automatique lors de l'édition en ligne d'un document

## GED Open source

---

**FreeDom et Contineo** proposent également réservation et gestion de versions, mais les possibilités sont en retrait par rapport aux solutions précédentes : gestion des verrous limitée, pas de versions mineures...

**Maarch** ne propose pas de ce type de fonctionnalités.

**Jahia** propose ces fonctionnalités, mais uniquement pour les contenus web. Il n'y a pas de gestion des versions pour les documents.

### Référentiel de classement

Un référentiel de classement est avant tout un ensemble structuré de mots clés ou d'expressions, on parle aussi de « vocabulaires » métiers.

Les solutions GED proposent des implémentations plus ou moins profondes, à la fois en terme de complexité des référentiels (hiérarchie, liens typés...) et en termes d'utilisations possibles.

L'extensibilité des solutions présente un intérêt tout particulier, puisqu'elle permet d'envisager d'étendre le modèle de référentiel en fonction des besoins (voir « Ontologie », page 40).

### Types de référentiels

#### ♦ Lexique

Il s'agit d'une liste des termes utiles dans un contexte donné et organisée le plus souvent de façon alphabétique. Il est parfois possible d'en donner une définition.

#### ♦ Plan de classement

Un plan de classement est un ensemble de termes hiérarchisés logiquement les uns par rapport aux autres.

Il s'agit du référentiel utilisé pour réaliser des classements physiques (papier ou autres).

On trouve souvent des cotes ou codes affectés à chaque terme. L'utilisation classique du plan de classement consiste à ranger physiquement les documents en affectant à chacun un et un seul terme du plan de classement.

Dans un usage électronique, il devient possible d'affecter plusieurs termes

---

## GED Open source

---

du plan de classement à un même document, ce qui donne autant de « clés » pour le retrouver. Une catégorisation multiple donc.

### ◆ **Thésaurus**

En plus de la structure logique et hiérarchique du plan de classement, le thésaurus introduit plusieurs notions permettant d'enrichir l'utilisation du référentiel documentaire. Le thésaurus est un ensemble de termes parfaitement définis, sur lesquels on s'appuie dans un domaine d'application particulier.

Plusieurs normes ISO définissent des relations entre les termes dits « descripteurs » qui constituent un thésaurus. Parmi les plus courantes, on retiendra :

- Terme « générique », c'est-à-dire un « père » dans la hiérarchie. Un terme peut avoir plusieurs « pères »
- Terme « spécifique », c'est-à-dire un « fils » dans la hiérarchie
- Terme « employé pour », c'est-à-dire les synonymes dont on ne retient pas l'utilisation dans le référentiel, mais qui pourraient être employés à la place du descripteur retenu
- Terme proche (« Voir aussi »), qui définit des liens transversaux entre termes descripteurs du fait de leur proximité sémantique

### ◆ **Tableau de gestion**

En matière de GED le tableau de gestion sert à établir les politiques de conservation, c'est-à-dire les règles qui présideront à la conservation des documents électroniques.

Le tableau de gestion est généralement construit sur la base d'un plan de classement.

Un tableau de gestion associe à chaque terme plusieurs informations :

- La durée de conservation : éventuellement une durée opérationnelle (dans le bureau) et obligatoirement une durée d'archivage
- Le « sort final » à l'issue de la durée d'archivage
- Les textes de référence qui justifient la durée de conservation, procédure qualité, loi, règlements... afin de permettre la révision

Tout l'intérêt de cet outil est de permettre d'associer des durées et des conditions de conservations à chaque élément du corpus documentaire (document ou dossier) en fonction d'une classification. Cet outil est indispensable à l'application de règles de conservation.

---

## GED Open source

---

Son utilité est spécifique au domaine des archives et de la conservation des archives vitales, c'est-à-dire le records management.

### ♦ **Ontologie**

Les ontologies fonctionnent sur le même principe que les thesaurus, c'est-à-dire par la mise en place de relations entre termes. Il s'agit d'un type de référentiel « extensible », c'est-à-dire dont la structure même prévoit qu'elle peut évoluer.

Deux exemples de liens : « a pour fournisseur » / « est fournisseur de », « est la filiale de » / « est la société mère de ».

Il s'agit d'un outil qui n'est pas implémenté dans les solutions de GED, certains permettent néanmoins, de par la structuration XML de leur référentiel de classement, de mettre en œuvre ce type d'outil.

A la différence des thesaurus, il n'y a pas de prédétermination ni de limitation dans les types de relations. Il est donc possible de créer autant de relations qu'on le souhaite entre les termes d'une ontologie et d'y associer contraintes et règles d'inférences.

Leur exploitation est encore assez limitée aujourd'hui et essentiellement basée sur les RDF (Ressource Description Framework), modèle de représentation d'ontologie, normalisée par le W3C.

Le concept d'ontologies est à la base du web sémantique, l'évolution pressentie par Tim Berners Lee pour rendre le web plus « intelligent » en permettant à des machines de réaliser seules des associations pertinentes entre contenus.

Dans le cadre d'une application documentaire, la mise en place d'ontologie(s) permet de disposer d'outils extrêmement puissants et adaptables à la fois au corpus documentaire traité et aux profils des utilisateurs.

Une ontologie peut par exemple définir des relations métiers qui pourront être exploitées par un moteur de recherche en fonction du profil de l'utilisateur ou de son « optique » de recherche. Elles pourraient également permettre de définir des plans de classement contextuels en fonction de paramètre qui indique quels liens de l'ontologie sont utilisés.

### **Utilisation des référentiels**

Etudions quelques utilisations envisageables.

### ♦ **Ergonomie de restitution des contenus**

La navigation par arborescence permet à l'utilisateur de retrouver un



## GED Open source

---

document selon une logique hiérarchique, voire plusieurs logiques hiérarchiques différentes.

Ainsi, si un contenu est référencé, catégorisé selon /Europe/France/Paris, mais également selon /Etablissements/Restaurants/Trois-Etoiles, l'utilisateur pourra parvenir à ce contenu par une navigation correspondant à la logique géographique, ou bien à la logique de classement gastronomique.

Cette utilisation permet donc de donner à chaque utilisateur le même référentiel documentaire, mais avec des approches différentes.

### ♦ **Gestion multi référentiels**

Dans un contexte professionnel, on est souvent amené à présenter une même information à différents publics. On a parfois à gérer plusieurs ensembles de vocabulaires contrôlés, par exemple une liste de mots-clés et une liste d'unités organisationnelles.

La gestion multi-référentiels impose de gérer certains cas difficiles tels que les homonymies, les liens entre référentiels... La plupart des applications GED étudiées disposent d'un seul référentiel de classement complexe (lorsqu'elles en disposent), il est donc important de valider la cohabitation de plusieurs référentiels si besoin.

### ♦ **Gestion des synonymes**

Il s'agit, pour un vocabulaire donné de pouvoir non seulement gérer les synonymes et les acronymes, mais également de rendre leur utilisation aussi transparente que possible à l'utilisateur.

Les fonctions les plus utiles sont :

- Enrichissement facile du référentiel, éventuellement par les utilisateurs eux-mêmes
- Utilisation automatique des synonymes lors d'une recherche sur le terme principale (voir expansion de requête ci-après)
- Détection automatique et remplacement des synonymes par le terme retenu lors de l'indexation

### ♦ **Exploitation par le moteur de recherche**

Les relations des référentiels peuvent être utilisées de différentes façons par le moteur de recherche.

#### Expansion de requêtes

L'un des intérêts des référentiels riches est qu'ils permettent de définir

---

## GED Open source

---

l'environnement sémantique d'un terme.

Certains moteurs de recherche sont capables « d'étendre la recherche » automatiquement en utilisant les relations indiquées dans le référentiel.

Par exemple, en indiquant automatiquement avec un « OU booléen » qu'une recherche faite sur le terme « voiture » concerne également tout ce qui parle « d'automobile » et de « bagnole ». Ainsi, tous les documents qui contiennent ces termes seront dans les résultats de recherche.

Le moteur peut également étendre la requête par proximité sémantique, par exemple entre pêcheur, bateau, poisson, ...

Les « expanseurs » de requête peuvent également reposer sur des analyseurs syntaxiques ou sémantiques qui construisent automatiquement des dictionnaires de correspondance à partir de l'analyse des documents indexés.

### Suggestion à l'utilisateur

L'analyse des mots-clés du référentiel trouvés dans les fiches des réponses à une requête donne une autre occasion d'exploiter les référentiels existants.

Par exemple, en proposant à l'utilisateur des termes proches de ceux trouvés le plus fréquemment dans les résultats de sa recherche.

Ce type d'utilisation est souvent plus riche lorsqu'on utilise un moteur syntaxique ou sémantique pour réaliser ces associations de mots-clés.

### Représentation en cluster (ou catégorie)

Il s'agit ici de ranger un ensemble de documents selon une arborescence de mots-clés en rapport avec ces documents.

Lorsque le regroupement est fait à partir d'une liste de valeurs préétablie dans un ou plusieurs champs, on parle de catégorisation.

Le terme de cluster est généralement réservé aux solutions qui proposent un calcul dynamique basé sur l'analyse de termes présents dans les documents eux-mêmes.

Ce type de fonction est par exemple utilisé pour représenter des résultats de recherche sous la forme d'une arborescence. Celle-ci est générée par rapport aux termes du référentiel présent dans chaque résultat. Ainsi, l'utilisateur verra immédiatement les thématiques représentées dans les documents trouvés et pourra le cas échéant préciser sa requête en ajoutant ou excluant certains termes du cluster.

## GED Open source

**Alfresco** propose un plan de classement sous forme d'une hiérarchie de catégories. Les fonctions de recherches permettent de filtrer sur une ou plusieurs de ces catégories.

**Nuxeo** propose un système de mots-clés organisé dans un plan de classement. Ceci est utilisé pour l'indexation des documents, la navigation et dans le moteur de recherche.

**Knowledge Tree** permet de créer aussi bien des lexiques que des plans de classement.

**Maarch** propose uniquement de créer des types de documents sur lesquels il est possible d'effectuer des recherches.

**Contineo** utilise un lexique qu'il est possible d'explorer lettre par lettre pour retrouver les documents associés.

**Freedom** utilise un plan de classement qu'il est possible d'explorer via une arborescence. La recherche dans ce plan de classement est également possible.

**Jahia** permet de disposer de plan de classement. Ce dernier sera utilisable pour indexer les documents uniquement dans la mesure où ceux-ci sont associés à des contenus.

## Moteur de recherche

Brique essentielle des applications de gestion d'information, les moteurs de recherche associés aux solutions de GED présentent comme caractéristique principale de pouvoir faire des recherches sur le contenu de la base documentaire.

### Fonctions de base

A minima, le moteur de recherche d'une application de gestion documentaire doit :

- Indexer les fiches documentaires, c'est-à-dire toutes les métadonnées ou la fiche d'un document
- Indexer les documents électroniques en texte intégral (ou « full text »), c'est-à-dire en prenant en compte le contenu des documents pour tous ceux qui incluent du contenu textuel
- Permettre des recherches aussi bien sur l'intégralité d'un document que sur un ou plusieurs champs spécifiques de métadonnées.

## GED Open source

---

- Gérer les accents et caractères spéciaux et plus largement des textes multilingues
- Permettre les recherches utilisant plusieurs critères associés avec des opérateurs booléens (ET, OU, SAUF), ainsi que des parenthèses.
- Permettre de mixer les approches documentaires, sur les métadonnées, et texte intégral
- Permettre l'utilisation des tables de référence dans les interfaces de recherche
- Permettre l'enregistrement de résultats de recherche, sous forme de « panier » par exemple, pour permettre aux utilisateurs de visualiser rapidement une sélection de documents
- Offrir la possibilité de personnaliser les pages de recherche, si possible plusieurs pages différentes, pour correspondre aux attentes des utilisateurs
- Autoriser les troncatures centrales et à droite, c'est-à-dire l'utilisation de caractères génériques pour remplacer un ou plusieurs caractères quelconques (classiquement l'étoile « \* » ou le « ? ») en milieu et en fin la chaîne de caractère qui sert à la recherche
- Le moteur de recherche doit **filtrer ses résultats de recherche selon les habilitations en lecture de l'utilisateur**, c'est-à-dire ne lui proposer que des documents qu'il a le droit de lire. C'est souvent ce croisement, entre résultats de recherche et habilitations, qui est le plus difficile à mettre en œuvre avec des moteurs autres que celui nativement intégré à l'application.

### Fonctions avancées

L'utilisation de moteurs élaborés est souvent envisagée pour prendre en charge des fonctions avancées comme :

- Les recherches successives ou transversales qui autorisent les utilisateurs à effectuer des recherches sur les documents, puis à affiner leur requête par rapport à des critères d'autres référentiels (la nationalité des auteurs par exemple)
- L'expansion de recherche, par rapport à des synonymes par exemple
- Les approximations, la suggestion orthographique, la lemmatisation (c'est-à-dire la recherche sur les racines des mots)
- La recherche en langage naturel, c'est-à-dire possibilité de soumettre une phrase interprétée par le moteur comme une équation de recherche
- La recherche par l'exemple, c'est-à-dire un texte sert de requête pour trouver les documents approchants. Le texte de référence peut être la

## GED Open source

---

valeur d'un critère d'indexation ou un document complet

- Suggestion de résultats approchant à l'issue d'une recherche, par différents moyens :
  - explicites, c'est-à-dire via des liens entre documents spécifiés dans l'indexation de chacun
  - implicite, c'est-à-dire des rapprochements issus de calculs statistiques ou sémantiques
- La détection automatique des « entités nommées », c'est-à-dire des noms propres (personnes, lieux) et de leurs relations
- La fédération de recherche, c'est-à-dire la possibilité de rechercher en une fois sur plusieurs bases documentaires
- L'enregistrement de stratégie de recherche, c'est-à-dire les critères utilisés pour la recherche, afin de permettre aux utilisateurs de relancer simplement des requêtes

La fonction de recherche est centrale dans les projets de GED et souvent le bénéfice le plus important pour les utilisateurs, sous réserve qu'elle soit adaptée pour répondre à leurs attentes spécifiques.

Parmi les besoins fréquemment évoqués, on retrouve par exemple : la navigation par arborescence, la recherche sur intervalle de dates, la recherche simple (pour les utilisateurs de Google), la recherche par référence, la recherche fédérée couvrant à la fois la GED et les applications tierces.

Il est donc fondamental d'analyser les attentes des utilisateurs pour définir leurs besoins de recherche. C'est un bon point de départ pour bâtir l'architecture (structure des données et ergonomie) de l'application cible.

Tous les outils proposent une fonction de recherche à minima sur les métadonnées.

**Alfresco, Nuxeo, et Knowledge Tree** permettent également une recherche en texte intégral sur la majorité des documents bureautiques. Les recherches peuvent porter sur chaque métadonnée de façon indépendante et il est également possible de les sauvegarder.

**Jahia** permet également les recherches intégrales sur tous les documents bureautiques.

**Alfresco et Nuxeo** permettent d'enregistrer des stratégies de recherche pour les ré exécuter à la demande.

## GED Open source

---

**Nuxeo** propose également la génération d'un plan de classement dynamique sur les métadonnées, module non activé par défaut.

Pour **Contineo** ou **Freedom**, le texte intégral se limite aux fichiers les plus simples. **Contineo** va cependant plus loin en proposant un affichage des résultats de recherche sous forme de graphe.

Dans **Maarch** la recherche se limite aux métadonnées des documents.

Tous ces outils utilisent par défaut des moteurs de recherche performants, souvent basés sur le moteur open source **Lucene**, mais ils ne proposent pas d'emblée certaines fonctions évoluées telles que la gestion des synonymes, l'expansion des requêtes, la suggestion, etc.

Les fonctions de recherche plus avancées ne seront pas disponibles nativement et demanderont une étape d'intégration avec des solutions de recherche sémantique plus complètes et généralement non open source, tel que Sinequa, Antidot, Exalead, Autonomy, Fast, etc.

### Intégration bureautique

Lorsque l'application de GED vise en grande partie la production de documents bureautiques, l'intégration aux applications bureautiques est importante.

Dans ce contexte, l'ergonomie et l'efficacité du travail, la fluidité du passage de la GED à la bureautique, seront des critères déterminants de l'acceptation de l'outil et donc de la réussite du projet.

Plusieurs modes d'intégration sont envisageables, et le choix impacte à la fois la complexité de l'environnement technique et la qualité de l'ergonomie.

### Espace de stockage

La GED peut être considérée comme un espace de stockage, similaire à l'utilisation d'un disque réseau : possibilité de gestion de dossiers, de navigation entre les différents niveaux, voire de déplacement d'élément directement par « glisser-déposer ».

L'intérêt essentiel de cette approche est que le changement est mineur pour les utilisateurs, par rapport aux simples serveurs de fichiers auxquels ils étaient habitués, ce qui facilite l'acceptation de l'outil de GED.

Technologiquement, cette approche est rendue possible par des interfaces telles que WebDav, CIFS ou même FTP. Il s'agit de technologies

## GED Open source

---

implémentées comme surcouche d'une base documentaire et qui permettent d'accéder à la GED par les outils standard sur le poste client.

L'utilisation du référentiel documentaire est proposée dans l'environnement habituel des utilisateurs, prenant en compte la gestion des habilitations, mais ne tire pas pleinement partie de l'application de gestion documentaire sous-jacente.

Notamment, l'indexation avec des métadonnées spécifiques, la recherche et généralement les fonctions qui nécessitent une interface utilisateur spécifique ne sont pas prises en compte par cette approche. Elle est donc le plus souvent complétée par une interface Web qui permet de donner une vue complète des fonctionnalités.

### Accès depuis l'application associée

Il s'agit de permettre l'accès à l'application documentaire à partir des applications associées aux documents (par exemple MS Office, AutoCAD, Photoshop...) et d'optimiser la fluidité du processus de création de documents.

Les rédacteurs doivent, au minimum, pouvoir réaliser depuis leur outil habituel les tâches : d'indexation, d'ouverture et d'enregistrement dans la base documentaire.

Tout l'intérêt par rapport à l'approche précédente est sur le premier point, l'indexation : une interface incluse dans l'outil de production permet de renseigner les métadonnées de la notice d'indexation. Ceci permet, d'une part d'enrichir les fichiers enregistrés d'une indexation et également de supprimer l'étape d'enregistrement intermédiaire que les utilisateurs réalisent souvent avant de déplacer leur fichier sur un lecteur réseau.

### Intégration à l'application

La meilleure intégration des outils de gestion documentaire est obtenue en proposant un panel de fonctions directement depuis les applications.

Parmi les fonctions les plus intéressantes à mettre à disposition des utilisateurs depuis leur outil de production, citons : indexation, ouverture, enregistrement, gestion des versions et recherche. C'est-à-dire **les fonctions de création et de modification des documents**.

Cette approche, lorsqu'elle n'est pas implémentée par l'application documentaire, nécessite un investissement important. Et le plus souvent une compétence des deux applications qui doivent communiquer, celles-ci pouvant être dans des technologies très différentes. Le pré requis est bien entendu de pouvoir ajouter un composant dans l'outil de production et de permettre une communication bidirectionnelle.

## GED Open source

---

L'intégration bureautique est la force d'**Alfresco**. Un partage réseau CIFS permet un accès totalement transparent au référentiel, ainsi que l'utilisation directe du processus de réservation de documents ou l'accès aux métadonnées. Un plug-in pour la suite Microsoft Office est également disponible, il permet directement depuis chaque application : ouverture, édition, réservation (c'est-à-dire verrouillage des modifications), accès aux versions, aux espaces ainsi qu'aux workflows. Les accès WebDav et FTP offrent un accès facile aux documents à travers un réseau IP.

**Nuxeo** propose une interface WebDav pour l'accès aux documents. L'ajout de document est possible via l'interface web, y compris par cliqué glissé sur la fenêtre du navigateur, via WebDav et via le plug-in LiveEdit qui automatise le processus de téléchargement depuis la suite Office.

**Knowledge Tree** dispose d'un accès WebDav, et d'une suite logicielle permettant une très bonne intégration dans Windows (exploration du référentiel), Microsoft Office (édition des documents) et Outlook (indexation des emails). Cependant, ces extensions sont une exclusivité de la version commerciale.

**Contineo** et **Freedom** proposent seulement un accès WebDav au référentiel.

**Maarch** ne permet aucune intégration au poste bureautique.

**Jahia** propose un accès WebDav, limité à l'ajout et la suppression de documents. La gestion des versions et la notion de réservation ne sont pas prises en charge.

## Numérisation

Il s'agit de la transformation d'un document sur un support matériel en données électroniques. Il peut s'agir de la numérisation d'un document papier vers un document bureautique ou d'un ensemble de formulaires vers une base de données, de la transformation d'un film ou d'un enregistrement sonore sur bande en fichier multimédia...

Les applications de gestion documentaire ne prennent en général pas en charge la numérisation, ni les traitements associés. Ceci conduit donc à utiliser des solutions de numérisation dédiées qui traitent les phases initiales depuis la conversion du support, jusqu'à l'injection du document dans la GED. En effet, ce type de solution est installé sur les postes de numérisation et communique avec la GED pour l'alimenter avec les documents numériques.

L'application documentaire reprend la main sur les traitements métiers,



## GED Open source

---

notamment d'indexation ou de manipulation des formats de fichiers.

Les fonctions suivantes entrent dans le spectre de la gestion documentaire, mais ne sont en général pas prises en compte directement par les solutions de GED.

### **Pilotage de scanner et préindexation**

Dans une chaîne de dématérialisation, il s'agit de la première fonction à prendre en compte.

C'est le plus souvent via une application installée sur le poste connecté au scanner qu'est réalisée la numérisation, mais cela peut également être mis en œuvre via une interface Web.

L'objectif étant de faciliter le processus qui comprend la numérisation et la saisie des premières métadonnées (on parle de préindexation), on s'efforce d'éviter à un opérateur d'utiliser plusieurs interfaces pour numériser et préindexer les documents dans la GED.

Bien que souvent perçue comme à faible valeur ajoutée par les opérateurs, la fiabilité et la précision des premières étapes de dématérialisation sont cruciales pour la valeur ajoutée globale du système de GED. C'est souvent à ce moment que sont saisies les métadonnées principales (date d'entrée, expéditeur, destinataire, montant...), leur exactitude et leur complétude vont déterminer pour une bonne part l'efficacité des traitements ultérieurs.

Des procédés de gestion de lot ou l'introduction de reconnaissance automatique permettent d'améliorer ces processus.

### **Reconnaissance automatique**

On trouve plusieurs catégories d'outils de reconnaissance automatique. Leur finalité est toujours de retranscrire l'information issue du support physique (papier) en son équivalent électronique, en minimisant les pertes et les interventions humaines.

Tous ces outils sont largement dépendants de la qualité de la numérisation : inclinaison, définition, visibilité de la trame du papier... Ces considérations doivent donc être prises en compte d'autant plus précisément qu'une reconnaissance automatique est prévue.

- OCR : Optical Character Recognition ou reconnaissance optique de caractère

Les « bons OCR » atteignent des taux de reconnaissance de 95%, et savent remettre en forme des tableaux, reconnaître les styles, etc.

- ICR : Intelligent Character Recognition ou reconnaissance intelligente

## GED Open source

---

des caractères

Essentiellement pour reconnaître l'écriture cursive (manuscrite), ce type d'application est notamment très utilisé par les banques pour les chèques.

- LAD : Lecture Automatique de Document

Pour l'acquisition de données structurées dans une base de données par rapport à des champs préalablement identifiés.

Basés sur la reconnaissance de forme, ces outils sont largement dépendants de la qualité de la reconnaissance (OCR) et de l'utilisation de dictionnaires complets. Les meilleurs utilisent des notions d'apprentissage et de suggestion de reconnaissance par association.

- RAD : Reconnaissance Automatique de Document

L'objectif est d'orienter les documents vers le bon processus de reconnaissance (plus haut). Ceci permet notamment d'appliquer des règles particulières en fonction du type de document reconnu.

Les solutions de GED embarquent parfois des moteurs dédiés et commerciaux pour ces fonctions. Leur intégration consiste souvent en un « injecteur » des documents numérisés dans la GED.

**Alfresco** s'interface avec Kofax Ascent Capture, solution commerciale de pilotage de scanner et de reconnaissance automatique du contenu.

La société **Nuxeo** a déjà intégré le support de scanners dans des solutions basées sur CPS. Pour Nuxeo EP, ce type de fonctionnalités ne fait pas encore partie de l'offre. Cependant, l'utilisation d'un client lourd (basé sur Eclipse RCP) constitue un atout pour bâtir une chaîne de numérisation homogène.

**Knowledge Tree** intègre, dans ses éditions commerciales, un outil spécifique qui permet à la fois le pilotage et l'OCR.

**Maarch** s'interface avec Kofax Ascent Capture ou directement le scanner Fujitsu ScanSnap.

**Contineo, FreeDom et Jahia** ne proposent pas nativement d'intégration avec de telles solutions.

## **Gestion des droits**

Dans la GED comme dans une majorité d'applications de gestion susceptibles d'avoir une diversité d'utilisateurs, la gestion des droits est une fonctionnalité très importante.

### **Gestion des profils**

La gestion des droits des utilisateurs par rapport à l'application passe par l'association d'autorisations sur des sous-ensembles définis de la base documentaire souvent des répertoires ou des arborescences.

Ces associations peuvent être réalisées par groupe ou individuellement et reposer sur une métadéfinition des droits au niveau d'un annuaire central (de type LDAP) ou non.

Les solutions présentées se différencient les unes des autres par rapport à la granularité des autorisations et la précision des sous-ensembles qu'elles permettent de définir.

Pour gérer des droits, on définit les rôles de chaque utilisateur sur les différents sous-ensembles de la base documentaire. Cette gestion peut elle-même faire l'objet d'une autorisation, ce qui permet de disposer d'administrateurs locaux, en charge de la gestion des droits déléguée sur une section de la base documentaire.

Les autorisations s'appliquent sur des objets documentaires, en fonction des applications, ces objets peuvent être définis différemment et comprendre notamment : une zone de la base documentaire, un dossier, un document voir un champ.

Elles sont donc relatives à la structure de la base documentaire.

Elles permettent d'indiquer ce qu'un utilisateur peut faire sur ces objets, notamment : voir, saisir, modifier, exporter, soumettre à validation, supprimer, affecter des droits, versionner, modifier fiche, modifier document... Chaque application définit ses droits unitaires.

C'est donc la combinaison de la granularité des objets documentaire et le nombre de droits qui va définir le niveau de granularité de la finesse des autorisations que l'on peut affecter.

### **LDAP et SSO**

Les applications documentaires peuvent reposer de la même façon que les autres sur des annuaires LDAP et dispositifs de Single Sign-On. Leur utilisation en est plus ou moins riche, notamment par rapport à la gestion des groupes et à la possibilité de créer des utilisateurs spécifiques, hors

## GED Open source

---

annuaire.

**Alfresco, Nuxeo, Knowledge Tree, Jahia** et **FreeDom** permettent une gestion fine des droits utilisateurs. La définition des droits pour un dossier (ou espace) donné peut être déléguée à un responsable. Ces outils supportent l'utilisation d'un annuaire LDAP, mais seuls **Alfresco, Nuxeo,** et **Jahia** supportent la configuration d'un système de SSO.

**Maarch** et **Contineo** ont une gestion plus sommaire, avec peu de choix dans les autorisations.

### Workflow

Deux types de workflows (processus) sont mis en œuvre dans le cadre de solution de GED/ECM. On se trouve ici à l'une des frontières entre le domaine de la GED et celui de l'ECM. En effet, dans les solutions de GED, on ne trouvera que des workflows appliqués à des documents. Dans les solutions d'ECM, on trouvera parfois des automatisations de procédures en dehors de tout contexte documentaire.

Différents niveaux de processus de gestion sont adressés par les solutions :

- Workflow documentaire, par exemple : validation, approbation, diffusion de document
- Workflow métier, par exemple : instruction de dossier, intégration de donnée dans une application métier

#### Workflow documentaire ou « Docflow »

Il s'agit de faire suivre aux documents un circuit de validation, souvent pour les publier.

L'outil de workflow mis en œuvre repose en général sur des fonctions standard, notamment la gestion des droits, l'envoi d'email, éventuellement le versioning. Certains sont des développement simple et donc peu paramétrable, notamment en termes de nombre d'étapes de validation, de paramétrage des notifications ou de résultat du workflow. Une structure classique pourrait être, par exemple soumission, déclenchement, validation/refus, publication

Les outils qui permettent d'aller plus loin, et donc de créer des workflows différents ou de les paramétrer finement sont le plus souvent basées sur des règles juxtaposées les unes aux autres (voir la notion de règle de gestion ou un moteur de règles). Ce type d'outil permet d'ajouter différentes fonctions, à partir d'une structure de base. Ces fonctions sont dépendantes

---

## GED Open source

---

des étapes du workflow, par exemple : notification par email, tableau de bord de suivi, ajout d'étape de workflow, transformation du document (en PDF par exemple), modification des métadonnées, voire signature électronique...

### Workflow métier ou BPM (Business Process Management)

L'objectif est de permettre la dématérialisation des procédures, qu'elles aient ou non un lien avec des documents. Les outils mis en œuvre sont des moteurs de workflow, c'est-à-dire des applications dédiées au paramétrage et à l'exécution de processus complexe.

Les utilisateurs sont vus uniquement par rapport à leur rôle dans tel ou tel processus.

Il existe de nombreux projets de moteur de workflow open source de qualité, le plus souvent en technologie Java. Ce sont des outils relativement complexes, parmi lesquels on peut citer **jBPM**, **Bonita**, **Orchestra** et **OSWorkflow**.

De même que pour une solution de gestion documentaire open source, il convient de :

- Se poser la question de leur pérennité par rapport aux critères de communauté, d'utilisateurs, de niveau fonctionnel...
- Envisager d'intégrer un autre outil que celui proposé en standard par la solution documentaire, en général, les intégrations entre applications open source sont réalisées de façon suffisamment générique pour permettre d'interchanger les briques concernées. C'est toutefois une voie sensiblement plus complexe.

#### ♦ Représentation de workflow

Il existe plusieurs méthodes de représentation de processus, disposant le plus souvent d'une partie graphique.

Les moteurs les plus élaborés s'appuient sur une représentation graphique des workflows pour en réaliser le paramétrage, tandis que la modélisation sous-jacente est souvent en XML. La transcription dans le moteur peut être plus ou moins riche, depuis la simple interprétation des étapes jusqu'à la possibilité de paramétrer graphiquement des alertes, des conditions, voire des scripts aux différentes étapes.

Les normes ou méthodes de représentation graphique de processus que l'on rencontre dans les moteurs de workflow sont notamment : BPML, formalismes BPEL, OSSAD.

## GED Open source

---

### ◆ **Implémentation**

L'intégration du moteur de workflow à l'outil de gestion documentaire est plus ou moins forte, notamment en ce qui concerne la gestion des droits et rôles, l'interaction entre les objets documentaires (document, dossier) et les objets de workflow (processus, étapes...).

Elle se traduit également par une intégration des interfaces qui ne donne pas à l'utilisateur l'impression d'utiliser deux applications différentes, le moteur de workflow doit être masqué.

### ◆ **Tableau de bord et suivi**

L'un des objectifs du BPM est de permettre un suivi en temps réel et l'audit des processus mis en place. Toutes les fonctions de « log » et leur exploitation doivent donc faire l'objet d'une attention particulière au moment de choisir sa solution de BPM.

A minima, chaque utilisateur doit pouvoir visualiser ses tâches en cours. Pour aller plus loin, les indicateurs disponibles comprendront également :

- La liste des tâches des subordonnés
- L'historique des tâches réalisées
- La liste des tâches des groupes auxquels appartient l'utilisateur
- Des statistiques d'utilisation au niveau de l'individu, du groupe...

### **Relations entre les rôles et les droits**

Le lien entre la gestion des droits et celle des rôles dépend fortement du moteur de workflow utilisé et de l'implémentation qui en est faite.

Les utilisateurs ont des rôles dans les processus, c'est-à-dire qu'ils peuvent être destinataires de certaines tâches. Parallèlement, ils ont des droits sur la base documentaire et sont donc autorisés à réaliser des actions sur certains objets documentaires.

L'intégration d'un moteur de workflow doit donc bien appréhender ce lien pour éviter les situations où un utilisateur serait destinataire d'une tâche, mais qu'il ne disposerait pas des autorisations pour la réaliser.

**Alfresco** disposent d'un outil très complet basé sur un moteur de workflow (JBPM). Il est bien intégré à l'application et permet de réaliser des workflows documentaires ou métiers.

**Nuxeo** disposent d'un outil très complet basé sur un moteur de workflow (JBPM). Notons que l'ensemble des processus, y compris ceux paramétrés simplement par

## GED Open source

l'interface graphique reposent sur jBPM, ce qui permet de disposer des fonctions de monitoring et d'historisation de façon globale.

**Jahia** embarque un « BPM server » dans les versions Professional et Entreprise.

**Ces trois outils** proposent également d'un niveau de workflow simple basé sur des règles de gestion facilement configurables via l'interface graphique.

**Knowledge Tree** utilise son propre moteur de workflow, moins complet qu'un BPM, mais simple à configurer. Dans cet outil, les processus sont uniquement documentaires.

**FreeDom** propose un workflow par développement de routines PHP. Cette approche est donc puissante, puisque chaque workflow est un développement, mais assez lourde à mettre en place. Une représentation graphique de workflow est proposée aux utilisateur.

**Maarch, Contineo** ne disposent pas de système de workflow.

## Règles de gestion

### Fonctions de gestion

Gérer des documents implique de disposer de fonctions permettant de prendre en compte les règles de gestion inhérentes aux métiers adressés, notamment en termes d'acquisition, de notification, de conservation et d'évolution.

On trouvera selon les applications des fonctions diverses, comme :

- Diffusion Sélective d'Information (DSI), il s'agit de faciliter la diffusion d'information en l'automatisant en fonction de différents critères :
  - Une stratégie de recherche préenregistrée
  - Un mot-clé, par exemple « tous les documents qui sont indexés avec le mot-clé GED »
  - Un dossier ou une arborescence, par exemple « les nouveaux documents dans un dossier »
  - La mise à jour d'un document particulier ou d'un ensemble de documents
- Notification par email, pour les échanges d'information avec les utilisateurs concernant leur compte notamment. Dans un projet de

## GED Open source

---

GED, le fait que l'application sache envoyer et éventuellement recevoir des emails est souvent un atout

- Comptabilisation des consultations, par exemple pour calculer le montant d'une redevance dans le cadre d'une consommation à la consultation
- Définition des formats d'affichage des fiches d'indexation, par exemple : par profil d'utilisateur, selon des optiques métiers ; ou par type de document, pour mettre en valeur les champs les plus pertinents pour chaque type
- Affectation conditionnelle de métadonnées à un document, par exemple, par rapport à la personne qui l'indexe, ou le dossier dans lequel il est placé
- Manipulation d'image, dans une approche photothèque, pour permettre de ne pas sortir l'image de l'application pour des actions simples : recadrage, rotation, balance de couleur, par exemple
- Echange collaboratif autour d'un document ou d'un dossier, par exemple par des fonctions de type forum

### Moteur de règles

Sur le même concept que les fonctions de gestion ci-dessus, les outils reposant sur un moteur de règles permettent de paramétrer des actions plus complexes et combinant les règles les unes aux autres.

Ces règles constituent un enchaînement d'action unitaire, par exemple, copie, transformation, alimentation d'une métadonnée, envoi d'email... Elles peuvent également être plus complexe et refléter des besoins métiers.

De plus, il est souvent plus facile d'implémenter de nouvelles fonctions dans un moteur de règles.

**Alfresco et Nuxeo EP** sont les seuls outils de ce comparatif qui intègrent un moteur de règles.

**Dans les deux cas**, les règles permettent une interaction poussée avec la base documentaire : modification des métadonnées, transformation des fichiers, envoi d'emails, etc. Un certain nombre de fonctions sont livrées en standard.

**Alfresco** permet une configuration directement depuis l'interface web. Pour **Nuxeo EP**, les règles sont définies au moyen de modules d'extension.



## **Gestion du cycle de vie**

### **Politique de conservation**

La conservation, la préservation et la sécurisation de l'information entrent de plus en plus dans les projets de gestion documentaire, à mesure que les entreprises prennent conscience de la valeur des documents qu'elles brassent.

Les politiques de conservation existent depuis longtemps dans le domaine papier (physique) mais ne sont que depuis peu appliquées aux données électroniques. C'est l'archivistique, discipline recouvrant toutes les fonctions liées à l'archivage, qui permet de répondre aux questions liées à la politique de conservation.

On trouve également le vocable de « *records management* » dont l'optique est plus adaptée au domaine numérique dans la mesure où il prend en compte une sous-sélection de l'ensemble des documents à archiver, sous sélection qui correspond aux « documents vitaux », c'est-à-dire nécessaires à l'activité de l'organisation.

La définition d'une politique de records management dans une application de gestion documentaire permet :

- de définir des ensembles de documents, par rapport à des types de documents ou des critères d'indexation
- de relier chaque ensemble à des actions de conservation

### **Archivage**

L'archivage peut être analysé de différentes façons en fonction du contexte de l'application. Du simple tag « archive » à la mise hors ligne de sélections de contenu sur des supports optiques, par exemple.

Sa nécessité et sa complexité vont croître avec le volume d'informations gérées.

Les applications intégreront donc des fonctions qui permettent :

- de gérer des informations de conservation dans les métadonnées des documents
- de déclencher automatiquement un traitement d'archivage
- de choisir de rechercher ou non dans les éléments archivés

En fonction des besoins réels, on envisagera différents types de traitements pour répondre à des problématiques de pérennité, de complétude, de

## GED Open source

---

fiabilité et de traçabilité des archives.

Parmi les fonctions envisageables, on citera :

- figer les éléments archivés, en calculant des empreintes cryptographiques, éventuellement signées, de façon à garantir l'inaltérabilité des archives
- faire valider, via un workflow, l'archivage (« le versement en archive ») d'un ensemble de documents par leur auteur ou le responsable du service concerné
- extraire les documents archivés et leurs métadonnées pour les stocker sur des supports à longue durée de vie (disque optique par exemple)

### Archivage électronique à valeur probante

Depuis quelques années, aux questions d'archivage viennent s'ajouter/juxtaposer les questions « d'archivage légal » ou « à valeur probante », c'est-à-dire la conservation sous forme électronique des documents pouvant servir de preuve. Cet aspect de l'archivage n'est en général pas couvert directement par la solution de GED, mais par un module annexe qui gère les signatures électroniques, les relations avec le tiers de confiance et la traçabilité de conservation.

Pour l'implémentation d'un archivage légal, il est nécessaire de garantir :

- L'intégrité : l'archive ne doit pas avoir été modifiée
- L'authenticité : signature nominative et vérifiable
- La traçabilité : la vie de chaque document doit être connue sans faille possible (horodatage certifié)
- L'auditabilité : le système doit pouvoir être vérifié pour prouver que ces processus ne peuvent être altérés

A noter que Smile a travaillé pour **CDC Arkhinéo**, numéro un de l'archivage électronique à valeur probante en France, pour définir l'architecture des plateformes d'archivage.

**Alfresco et Nuxeo EP** proposent des métadonnées pour la politique de conservation et l'archivage, mais ni l'un ni l'autre ne permettent d'automatiser la mise hors ligne des contenus par le pilotage de support de stockage ou l'archivage légal. **Nuxeo EP** prend nativement en charge la gestion du cycle de vie des documents.

**Maarch**, outil d'archivage, permet la mise hors ligne de documents, mais ne supporte qu'un jeu limité de métadonnées pour gérer une politique de conservation complexe.

## GED Open source

---

**Les autres solutions** ne proposent pas de fonctions d'archivage.

### Import/export

Ces fonctions ont un statut particulier dans les applications documentaires. En effet, en plus de leur utilité pendant l'exploitation de la solution, **l'import et l'export servent à entrer et sortir d'une solution.**

#### Import / export en masse

L'objectif est ici de permettre de faire entrer ou sortir rapidement un grand nombre de documents.

Ces fonctions peuvent prendre en compte ou non les métadonnées et la gestion des autorisations.

Idéalement, l'import permet d'intégrer plusieurs documents et de leur affecter des métadonnées en lot. L'export doit permettre de restituer l'ensemble des documents sélectionnés (voire l'ensemble de la base documentaire) ainsi que toutes les méta-informations rattachées et relatives aux autorisations.

#### Export sur support physique

Deux optiques différentes sont à envisager ici.

L'export à des fins d'**archivages**, c'est-à-dire la possibilité de « sortir » certains documents et leurs métadonnées sur des supports lents (disque optique, bande DAT...). Idéalement, le système conserve la mémoire des documents ainsi exportés de façon à permettre des recherches sur les métadonnées. C'est-à-dire, que les index de recherche sont conservés de façon à rester accessibles à la demande

L'export à des fins de **consultation hors ligne**. Il consiste à exporter une partie de la base documentaire sur un support numérique doté de sa propre interface de consultation

Les accès réseau ou WebDav présents dans la plupart des solutions présentées permettent un import/export massif des documents. Cependant, les métadonnées ne sont pas conservées avec cette méthode.

**Alfresco** permet d'importer et exporter sous forme d'archives des espaces entiers du référentiel, en incluant les métadonnées ainsi que les règles et les droits au format XML.

## GED Open source

---

**Nuxeo EP** dispose d'un module I/O en charge des imports et export de l'ensemble des données (fichiers, métadonnées et droits associés), il s'agit d'un module complémentaire.

**Jahia** dispose d'un module d'import export de l'ensemble des données (contenu et droits en XML ainsi que les fichiers liés).

**Knowledge Tree** permet l'import automatique de données à partir du fichier lui-même et des métadonnées associées stockées dans le format XML. Il est également possible d'importer et exporter des fichiers en bloc, mais sans gestion de métadonnées.

Pour **Maarch**, l'export de données est un module commercial.

Enfin, il est toujours possible, pour les autres outils, de développer un programme d'import/export avec métadonnées qui utilise les API de l'outil.

## Gestion des emails

La gestion des mails est une fonction très demandée en matière de gestion documentaire, à la fois pour que ces éléments d'information participent à la mémoire collective et pour des raisons réglementaires (notamment Sarbanes-Oxley ou Bâle II qui imposent des règles de conservation).

Les problématiques de gestion des connaissances impliquent de partager et de capitaliser les informations, même si leur degré de formalisation est faible, ce qui est typiquement le cas des emails. Leur stockage et leur partage facilitent le travail en équipe. Les contrôles, qu'ils soient réglementaires ou qualité, s'imposent également de plus en plus. Ils poussent les organisations à conserver tous les échanges avec leurs employés et leurs partenaires commerciaux. La mutualisation semi-automatique des emails peut donc apparaître comme une solution à cette problématique.

### Sélection des emails à archiver

Le processus d'archivage d'un mail ne peut être totalement automatique, sous peine de voir stocker une multitude de pourriels (spam), de messages sans intérêt voire de messages privés (qui doivent rester en dehors des mécanismes de mutualisation).

Les solutions de GED doivent donc proposer des processus de sélection pour l'archivage des emails.

La sélection peut passer par exemple par l'intégration d'un plug-in dans le

## GED Open source

---

gestionnaire de courrier (Mozilla ou Outlook...) ou par l'intégration automatique d'emails envoyés à une adresse mail spécifique (stockage\_ged@votreorganisation.com, par exemple).

### Processus de gestion et de conservation

La gestion des emails pose plusieurs problèmes à la fois techniques et méthodologiques.

D'une part, il n'existe pas de contrôle sur le nombre ou le format des pièces jointes. Le format des emails eux-mêmes est variable (HTML ou texte, voire formats propriétaires selon le client de messagerie utilisé).

Par ailleurs, les échanges par emails sont le plus souvent multipartites (à plus de deux interlocuteurs), le stockage d'un email en plusieurs exemplaires génèrerait alors des doublons inutiles.

L'application de gestion documentaire devra donc proposer des solutions pour :

- gérer des transformations de formats afin d'assurer une conservation et une restitution optimale des informations transmises
- détecter les doublons des objets documentaires de type email et pièces jointes pour éviter les stockages multiples et faciliter le travail collaboratif
- permettre d'indexer à la fois le contenu des emails et des pièces jointes, pour permettre de les retrouver par le moteur de recherche

**Alfresco** gère les emails via simple glisser-déposer depuis Outlook. Les informations de base (expéditeur, destinataires, sujet) sont extraites comme métadonnées et le contenu du mail et des pièces jointes est disponible en recherche texte intégral.

**Nuxeo EP** dispose d'un module qui permet de réaliser du glisser-déposer depuis Outlook.

**Knowledge Tree**, dans sa version commerciale, dispose d'une intégration à Outlook pour la gestion des emails : depuis Outlook il est possible d'archiver les emails desquels les métadonnées sont automatiquement extraites, mais aussi d'attacher des fichiers du référentiel lors de l'envoi d'un mail. Le contenu des mails n'est cependant pas indexé.

**Les autres outils** n'ont pas de telles fonctions.

## Gestion de dossiers

### Notions relatives à la « gestion de dossiers »

A l'image de la gestion des fichiers sur un poste de travail, elle-même issue de la gestion de documents papier dans des classeurs, chemises... l'ergonomie des solutions de GED en matière de « gestion de dossier » se rapproche de la logique des utilisateurs. Ici, un *dossier* n'est pas un simple répertoire, mais un ensemble de documents liés par une finalité commune.

La notion de dossier est cependant fluctuante et il convient de préciser les deux acceptions que nous lui donnerons :

- **Dans l'optique professionnelle**, il s'agit d'un ensemble de documents ayant une cohérence à être regroupés, et à partir duquel des utilisateurs accompliront des tâches (au sein d'un workflow le plus souvent)
- **Dans l'optique documentaire**, un dossier réunit des documents ayant des caractéristiques communes, une métadonnée, par exemple la thématique, un type de document
- **Dans l'optique du système d'information d'entreprise**, le dossier peut regrouper virtuellement des informations issues de divers composants du système d'information pour en faciliter la consultation

Notons que ces approches peuvent être complémentaires.

Il existe deux différences fondamentales entre les dossiers papier et électroniques :

- L'électronique permet de gérer les accès simultanés pour un même document
- La gestion électronique permet de réaliser plusieurs traitements simultanément, notamment la mise à disposition pour information et l'instruction d'un dossier

### Les outils de la gestion de dossiers

Les deux moyens principaux de cette gestion sont : d'une part les **métadonnées pour partager des informations** synthétiques et d'autre part la possibilité d'**associer des tâches ou des séries de tâches** à un dossier.

#### ♦ La constitution d'un dossier

- Définir les pièces constitutives d'un dossier dans le contexte envisagé
- Définir comment sont agencées les pièces d'un dossier

## GED Open source

---

- Prévoir les règles applicables à chaque type de dossier, par exemple contrôle de complétude, validation de certaines pièces, envoi de courriers prédéfinis...

### ♦ **Le partage d'information**

- Définition de méta-informations par dossier
- Possibilité de structurer un dossier par regroupement de documents via une (des) métadonnée (s)
- Possibilité de structurer les documents en « pochette », sous classification des dossiers, y compris lorsque la notion du dossier repose sur une structuration virtuelle (sur une métadonnée par exemple)
- Liens entre dossiers (même niveau, hiérarchie...)
- Gestion des droits et des rôles des utilisateurs sur un dossier (initialisation, agent traitant, superviseur, décideur) et leurs prérogatives respectives

### ♦ **Les règles, par exemple de gestion, de conformité, d'alerte...**

- avec des déclencheurs automatiques ; par exemple pour l'archivage ou les rappels dans le cadre d'une procédure, ou à la réception d'une pièce particulière
- Interaction entre documents et données d'une application de gestion tierce ou une base de données (pour des formulaires notamment)
- Application de règles sur des sous-ensembles de documents constitutifs du dossier
- Vérification de la complétude du dossier
- Possibilité de générer des documents à partir de modèles et de conditions, par exemple un formulaire pré rempli par rapport aux métadonnées extraites d'un courrier
- Gestion de workflow de dossier pour automatiser les transitions et permettre un suivi automatique des acteurs de l'instruction d'un dossier

### ♦ **Le classement**

- La relation entre les pièces physiques et numériques peut être importante dans le cas où des pièces en format papier ont un caractère légal
- Identification du dossier : règles de constitution, sémantique du nom...

### ♦ **Traçabilité**

## GED Open source

---

- Toute action sur chaque dossier doit être tracée pour permettre des audits sur le traitement des dossiers

**Alfresco** permet de gérer des dossiers en combinant répertoire, métadonnées, règles de gestion et fonctions d'audit.

**Nuxeo EP** gère également la notion de dossier et apporte une notion de section. Dans la pratique on combinera répertoire, métadonnées, règles de gestion et audit. La notion de section autorise une abstraction supplémentaire, transversale par rapport à la publication de sélections de documents.

**Jahia** permet de gérer des dossiers dynamiques en fonction de métadonnées, mais n'intègre pas nativement de règles de gestion.

Les autres outils ne proposent pas nativement de gestion de dossiers au sens où on l'entend ici, ils se contentent de gérer des répertoires. Elle se limite à une gestion d'arborescence de type explorateur de fichiers.

### Intégration à l'infrastructure du SI

Un outil de GED, de par son rôle central dans l'entreprise, se doit d'être communicant.

A cette fin, plusieurs standards sont disponibles :

#### Les services Web (web services)

Norme d'appel de méthodes entre applications de technologies semblables ou différentes. La création et l'utilisation d'un service web se font de façon pratiquement transparente dans la plupart des cas.

#### Services REST

Plus simples et performants que les services web, les services REST se limitent à l'utilisation du protocole HTTP pour les entrées-sorties en format XML.

#### Interfaces simplifiées

Une autre façon simple d'intégrer l'outil de GED dans le Système d'Information est de proposer une ou plusieurs interfaces utilisateurs simplifiées et limitées à un domaine ou une utilisation précise. Cas d'utilisations : insertion dans des portails ou des applications clients lourds, échange de données, AJAX, mini applications utilisant la base de



## GED Open source

---

GED.

### Portail

L'outil peut fournir un certain nombre de « portlets », composants qui s'installent sur un portail J2EE (intranet par exemple) et qui donnent directement accès à ses fonctionnalités. Consultez le livre blanc des solutions de portail pour plus de détails.

### Brique applicative

L'outil peut également permettre d'intégrer le référentiel de GED comme une brique fonctionnelle au sein d'une application spécifique. Les fonctionnalités de stockage de gros volume de données, gestion des versions, métadonnées, workflow... sont alors directement fournies par l'outil de GED.

**Alfresco** dispose d'un grand éventail d'API : web-services, REST et également webscripts, pour la réalisation d'interfaces simplifiées à base d'HTML et de javascript.

**Nuxeo EP** dispose d'un ensemble de web-services, technologies de remoting EJB, SOAP et REST, ainsi qu'un système de plug-ins qui permet d'étendre facilement les fonctionnalités de l'outil. La disponibilité de l'interface client lourd Eclipse RCP facilite le développement d'application nécessitant des traitements lourds.

**Knowledge Tree** dispose uniquement de web-services.

**Contineo, Freedom, Maarch** sont plus limités et ne proposent pratiquement pas de possibilités d'interaction.

**Jahia**, de par son statut de portail, met peu de fonctions à disposition pour accéder et manipuler ses propres contenus. Il est possible de récupérer les contenus en XML via un web service et les fichiers par leur adresse WebDav.

## Configuration de l'interface

Un outil de GED peut être très complet, mais offrir une interface par défaut qui ne convient pas aux utilisateurs ciblés, ou qui doit être mise aux normes graphiques de l'entreprise. S'il est toujours envisageable, avec des solutions open source, de faire les adaptations directement sur le code de l'application elle-même, c'est au risque de rendre les mises à jour plus complexes. Certaines solutions ont bien compris ce problème et permettent de modifier leur interface par configuration.

**Alfresco** dispose d'une interface, modifiable par configuration : les menus et actions disponibles, les différents assistants, la navigation et les réglages par défaut sont définis par simple configuration. La modification de l'interface graphique nécessite cependant de modifier les fichiers de l'outil. Ces modifications peuvent être assemblées sous la forme d'un module, qui devra être appliqué à nouveau lors des mises à jour.

**Nuxeo** dispose d'un éditeur de thèmes puissant qui permet de configurer de façon interactive l'aspect graphique de l'outil. Pour les modifications plus poussées, le système de plug-in permet de modifier l'application à volonté.

**Knowledge Tree** permet de modifier facilement le texte des différents messages de son interface.

**Contineo**, **Freedom** et **Maarch** ne permettent pas réellement de modification de leur interface.

**Jahia**, de par son orientation pour la publication de contenus, offre une très grande souplesse dans l'adaptation des interfaces aux besoins utilisateurs.

## Synthèse

Remarquons que dans le domaine de la gestion documentaire, les solutions réellement matures sont portées par des éditeurs qui s'appuient sur un modèle open source qui implique le plus souvent une contractualisation avec l'éditeur. L'étendue des attentes fonctionnelles pour ce type d'application et leur caractère stratégique expliquent largement cette caractéristique. En effet, il s'agit de traiter l'information à la fois opérationnelle et vitale des organisations, le support d'un éditeur est donc largement justifié dans ce contexte.

Trois produits sortent du lot : **Alfresco**, **Nuxeo EP** et **Knowledge Tree**.

S'ils ne disposent pas tous des mêmes fonctionnalités, ils rivalisent pour répondre à la majorité des problématiques d'entreprise et disposent d'un très bon niveau de support, à la fois de leur éditeur, des intégrateurs et de leur communauté. Le choix s'effectuera en fonction du contexte, sur les fonctionnalités nécessaires, les possibilités d'intégration et le coût du support.

Lorsque les projets de gestion de contenu ont les deux composantes, documentaire et Web, on pourra se tourner vers des outils de CMS qui disposent de fonctions de GED, tels que **Jahia** ou **eZ publish**. En effet, à l'heure actuelle les solutions présentées dans ce livre blanc disposent de fonctions limitées à ce niveau et sont donc moins bien placées pour assumer les deux rôles. Citons néanmoins le module WCM d'Alfresco qui devrait sous peu apporter une réponse intéressante. Nous vous invitons à vous référer à notre livre blanc sur la Gestion de contenu open source.

Comme indiqué au début de ce livre blanc, notre objectif était non seulement de présenter les solutions, mais également notre vision de la gestion documentaire.

Nous espérons vous avoir éclairé sur cette problématique et mettons notre expertise à la disposition de vos projets.