

Gestion documentaire

Le meilleur des solutions open source

Smile
OPEN SOURCE SOLUTIONS

www.smile.fr • +33 (0)1 41 40 11 00 • contact@smile.fr
www.smile-oss.com • blog.smile.fr • [@GroupeSmile](https://twitter.com/GroupeSmile)

Thomas **CHOPPY**
Directeur de l'offre ECM

Hélène **ZYSMAN**
Consultante

Ioana **LETI**
Consultante

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, du **PLOSS** – le réseau des entreprises du Logiciel Libre en Ile-de-France et du **CNLL** – le conseil national du logiciel libre.

Smile compte près de 700 collaborateurs dans le monde, dont plus de 500 en France (mars 2013), ce qui en fait *le premier intégrateur français et européen de solutions 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), la virtualisation (2007), la gestion électronique de documents (2008), les PGIs/ERPs (2008), les VPN open source (2009), les Firewall et Contrôle de flux (2009), les Middleware orientés messages (2009), l'e-commerce et les Réseaux Sociaux d'Entreprise (2010) et plus récemment, le Guide de l'open source et NoSQL (2011). 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 Smile Digital – agence interactive, proposant outre la création graphique, une expertise e-marketing, éditoriale, et interfaces riches. Smile dispose aussi d'une agence spécialisée dans la TMA (support et l'exploitation des applications) et d'un centre de formation complet, Smile Training. Enfin, Smile est implanté à Paris, Lille, Lyon, Grenoble, Nantes, Bordeaux, Poitiers, Aix-en-Provence et Montpellier. Et présent également en Espagne, en Suisse, au Benelux, en Ukraine, au Maroc et en Côte d'Ivoire.



**open
source**

**QUELQUES REFERENCES DE
SMILE**

SMILE est fier d'avoir contribué, au fil des années, aux plus grandes réalisations Web françaises et européennes. Vous trouvez ci-dessous quelques clients nous ayant adressé leur confiance.

Gestion documentaire

Generali, HEC, JCDecaux, Serimax, Pierre Audoin Consultant, Alstom Power services, NetasQ, CS informatique, SNCF - Direction du matériel, Mazars, EDF R&D, EDF Nucléaire, Conseil Régional du Centre, Leroy Merlin, Primagaz, Renault F1, INRIA, Ministère belge de la Communauté Française, APAVE, CNIL, Services du Premier Ministre...

Sites Internet

EMI Music, Salon de l'Agriculture, Mazars, Areva, Société Générale, Gîtes de France, Patrice Pichet, Groupama, Eco-Emballage, CFnews, CEA, Prisma Pub, Véolia, NRJ, JCDecaux, 01 Informatique, Spie, PSA, Boiron, Larousse, Dassault-Systèmes, Action Contre la Faim, BNP Paribas, Air Pays de Loire, Forum des Images, IFP, BHV, ZeMedical, Gallimard, Cheval Mag, Afssaps, Benetaux, Carrefour, AG2R La Mondiale, Groupe Bayard, Association de la Prévention Routière, Secours Catholique, Canson, Veolia, Bouygues Telecom, CNIL...

Portails, Intranets et Systèmes d'Information

HEC, Bouygues Telecom, Prisma, Veolia, Arjowiggins, INA, Primagaz, Croix Rouge, Eurosport, Invivo, Faceo, Château de Versailles, Eurosport, Ipsos, VSC Technologies, Sanef, Explorimmo, Bureau Veritas, Région Centre, Dassault Systèmes, Fondation d'Auteuil, INRA, Gaz Electricité de Grenoble, Ville de Niort, Ministère de la Culture, PagesJaunes Annonces...

E-Commerce

Krys, La Halle, Gibert Joseph, De Dietrich, Adenclassifieds, Macif, Furet du Nord, Gîtes de France, Camif Collectivité, GPdis, Projectif, ETS, Bain & Spa, Yves Rocher, Bouygues Immobilier, Nestlé, Stanhome, AVF Périmédical, CCI, Pompiers de France, Commissariat à l'Energie Atomique, Snowleader, Darjeeling...

ERP et Décisionnel

Veolia, La Poste, Christian Louboutin, Eveha, Sun'R, Home Ciné Solutions, Pub Audit, Effia, France 24, Publicis, iCasque, Nomadventure, Gets, Nouvelles Frontières, Anevia, Jus de Fruits de Mooréa, Espace Loggia, Bureau Veritas, Skyrack, Lafarge, Cadremploi, Meilleurmobil.com, Groupe Vinci, IEDOM (Banque de France), Carrefour, Jardiland, Trésorerie Générale du Maroc, Ville de Genève, ESCP, Sofia, Faiveley Transport, INRA, Deloitte, Yves Rocher, ETS, DGAC, Generalitat de Catalunya, Gilbert Joseph, Perouse Médical...

Infrastructure et Hébergement

Agence Nationale pour les Chèques Vacances, Pierre Audoin Consultants, Rexel, Motor Presse, OSEO, Sport24, Eco-Emballage, Institut Mutualiste Montsouris, ETS, Ionis, Osmoz, SIDEL, Atel Hotels, Cadremploi, SETRAG, Institut Français du Pétrole, Mutualité Française...

Consulter nos références, en ligne, à l'adresse : <http://www.smile.fr/clients>.

CE LIVRE BLANC

Ce document a pour objectif de présenter notre approche de la gestion documentaire, notamment par rapport à la gestion de contenu, et de vous apporter une aide au choix d'une solution logicielle pour vos projets.

Pour cela, nous vous proposons une analyse des réponses apportées par une sélection de solutions open source à des problématiques concrètes, ainsi qu'une approche en termes de méthodes pour vous orienter dans la mise en œuvre de votre projet.

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 entre la réalité des besoins et le marché de la GED (Gestion Electronique de Documents) open source.

Si les solutions open source répondent très bien à de nombreux besoins, le savoir-faire spécifiquement documentaire demeure, avant les outils, le principal enjeu des projets de Gestion Electronique de Documents ou de Gestion de Contenus d'Entreprise (ECM).

Ainsi, cet ouvrage s'efforce de réunir :

- une approche générale des notions sous-jacentes de la gestion documentaire, fondements méthodologiques des projets de GED
- une description des grandes fonctionnalités attendues pour 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

Ce livre blanc est le résultat de retours d'expériences collectifs. Nous serions très heureux d'échanger avec vous sur le sujet.

N'hésitez pas à nous transmettre vos avis et évaluations sur les produits présentés dans ce livre blanc. Une seule adresse : contact@smile.fr

Version 1.0

Paru peu après que Smile se soit intéressé au domaine de la GED open source, notamment avec l'arrivée d'une équipe dédiée. Nous avons voulu proposer une vision de la gestion documentaire à la fois structurante et opérationnelle pour permettre à tout un chacun de se faire une opinion et disposer des atouts nécessaires à la mise en place de son projet.

Version 2.0

Une mise à jour importante, réalisée entre octobre 2009 et janvier 2010. Mise à jour des parties méthodologiques, des retours d'expérience de notre pôle d'expertise et des évolutions du marché.

Version 3.0

La version 2011 visait à intégrer de nouveaux retours d'expérience et les nouveautés des applications. Nous avons également modifié la liste des applications concernées. Une tendance en notable avec l'arrivée du SaaS dans la GED.

Version 4.0

Nous avons modifié la liste et les pondérations des applications comparées. Autre fait notable, l'arrivée de Jahia Wise et de solutions ou de fonction de DAM (digital asset management) sur un marché maintenant mature en open source. Nous avons également mis en retrait plusieurs solutions dont l'évolution ou le caractère open source ne nous semble plus répondre aux attentes du marché.

L'OPEN SOURCE EN ENTREPRISE

Gestion documentaire non open source

De nombreuses solutions existent et apportent leurs lots de fonctions et de réponses aux problématiques de gestion documentaire.

Parmi la multitude de solutions propriétaires connues dans le domaine, on peut citer Documentum d'EMC, FileNet d'IBM, SharePoint de Microsoft, Open Text de la société éponyme, ou encore Ever.

Le marché des solutions propriétaires de gestion documentaire suit la même tendance que celui des outils de gestion de contenu Web, il y a quelques années. Les solutions qui demeurent sont soit sur des marchés de niches et incorporent un savoir-faire métier important, soit disposent d'un positionnement très haut de gamme, associé à une notoriété qui leur permet de continuer à trouver des clients.

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

- son produit doit être ouvert, mais pas trop, pour garder la maîtrise du client (c'est le fameux « vendor lock-in »)
- son produit doit être performant, mais pas trop, car l'accroissement du nombre de serveurs rime avec l'augmentation du nombre de licences vendues
- son produit doit être robuste, mais complexe, car il faut continuer à vendre des prestations de service et du support
- son produit doit être utile, mais surtout être attractif. En effet, les décideurs sont parfois éloignés de la réalité du terrain et un produit très bien packagé, mais peu évolutif ou ne répondant pas de façon assez précise aux attentes opérationnelles, peut alors apparaître comme la meilleure alternative, alors que sa valeur ajoutée sera faible finalement.

Nous ne disons pas que les éditeurs propriétaires seraient machiavéliques au point de rendre leurs produits moins bons qu'ils ne pourraient l'être, mais leurs priorités ne portent pas exclusivement sur la valeur ajoutée du produit.

Le choix de l'open source

L'open source gagne chaque année de nouveaux domaines d'application, dans une extraordinaire dynamique. De nouveaux acteurs sont apparus, les éditeurs open source et la pertinence de leur 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.

Voyons quels sont les critères de choix, liés au caractère open source des solutions.

Les bénéfices économiques sont parmi les premières raisons du choix de solutions open source. Même si implémenter une solution ne signifie jamais un coût de projet nul, ne serait-ce qu'en temps de montée en compétence et d'implémentation. De nombreuses études constatent que le coût total des projets open source est significativement inférieur sur la durée. **Les gains constatés pouvant aller de 20 à 80% en fonction du degré de maturité de l'open source dans le domaine considéré.**

Le coût de licence des solutions propriétaires est bien entendu le premier facteur évoqué : il constitue un investissement initial important, avant même que l'adéquation de la solution ne soit prouvée. Ceci est accentué par un prix de prestations qui tend à être moins élevé dans le domaine de l'open source, car l'ouverture du produit facilite la diffusion de la connaissance. Enfin, le coût de possession (maintenance, évolution) tend également à être moins important.

Leur évolution ne dépend pas de leur rentabilité ou de considérations marketing. Tant qu'une communauté s'y intéresse, le produit vivra.

La diffusion du savoir-faire associé à leur implémentation est garantie par des barrières quasi nulles à son acquisition (téléchargement et documentation sont libres et accessibles).

La standardisation, le respect des normes et standards, et **l'ouverture** sont dans les gênes des développeurs open source, qui cherchent avant tout à ne pas réinventer la roue, et à mettre en œuvre des solutions performantes.

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 avec d'autres applications.

La pérennité : la libre disposition des sources est un gage fondamental de pérennité. Cela même si – il faut le souligner – il n'est nullement nécessaire que l'entreprise cliente les maîtrise.

En matière de pérennité, le pire qui puisse arriver à une solution open source, est une lente désaffection 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 le phénomène est toujours lent et les utilisateurs ont le temps d'organiser la migration.

Dans le cas d'un éditeur open source, il faut souligner aussi que, même s'il était un jour défaillant, il reste toujours possible pour une communauté de reprendre en main le produit et ses évolutions, c'est le principe des licences open source.



**open
source**

Ainsi au fur et à mesure que ces solutions arrivent à maturité, le moindre coût n'est plus le premier critère de choix. Et ce sont bien les autres qualités de l'open source que nous mettons le plus en avant.

SOMMAIRE

PREAMBULE	1
SMILE.....	2
QUELQUES REFERENCES DE SMILE.....	3
CE LIVRE BLANC.....	5
L'OPEN SOURCE EN ENTREPRISE.....	7
SOMMAIRE	10
INTRODUCTION A LA GESTION DOCUMENTAIRE	12
POURQUOI UNE SOLUTION DE GESTION DOCUMENTAIRE ?.....	12
PRESENTATION DES CONCEPTS	13
VERS LA GESTION DE CONTENU D'ENTREPRISE	17
ENJEUX PRINCIPAUX DE LA GED	20
LES PRINCIPALES SOLUTIONS DE GESTION DOCUMENTAIRE OPEN SOURCE	25
ALFRESCO	25
NUXEO	30
KNOWLEDGE TREE.....	34
JAHIA	36
MAARCH	39
SOLUTIONS SPECIFIQUES POUR LA GESTION DE MEDIA.....	41
AUTRES SOLUTIONS.....	42
LES FONCTIONNALITES	48
METADONNEES.....	48
GESTION DES VERSIONS	52
REFERENTIEL DE CLASSEMENT	53
MOTEUR DE RECHERCHE	60
NUMERISATION.....	63
GESTION DES DROITS.....	65
FONCTIONS COLLABORATIVES.....	66
WORKFLOW	68
REGLES DE GESTION	71
GESTION DU CYCLE DE VIE	73
GESTION DES DOSSIERS.....	77
IMPORT/EXPORT	79
GESTION DES EMAILS.....	81

**open
source**



INTEGRATION DE LA GED	82
INTEGRATION TECHNIQUE	85
SYNTHESE.....	89

INTRODUCTION A LA GESTION DOCUMENTAIRE

POURQUOI UNE SOLUTION DE GESTION DOCUMENTAIRE ?

La décision de déployer une solution de GED résulte encore souvent d'une crise ou d'une tension dans la gestion de l'information ; une situation qui nécessite la mise en place d'une organisation **plus structurée**, d'une **traçabilité** ou d'une meilleure **utilisabilité**. L'informatique est le plus souvent le meilleur moyen de fournir une réponse adaptée.

Les crises peuvent se manifester de diverses façons : l'impossibilité de reconstituer un dossier client complet dans un délai raisonnable, de retrouver la dernière version d'un document électronique qui a nécessité des jours de travail, etc. Bien qu'insidieuse dans notre ère hautement numérique, **la crise la plus fréquente est probablement la « perte par dilution »**. Dilution d'une information importante dans un ensemble beaucoup trop vaste. Il y a perte de connaissance ou de mémoire au niveau de l'organisation.

Les tensions résultent de difficultés chroniques dans l'usage des documents ou de temps perdu. Le temps perdu en recherche est le facteur le plus fréquemment cité. La réutilisation de documents existants est également problématique, lorsque des ensembles documentaires autonomes croissent dans l'organisation ; l'efficacité du travail peut alors être largement dégradée du fait d'un manque d'organisation, de traçabilité ou simplement de visibilité dans l'organisation documentaire. **Il s'agit ici de perte de productivité.**

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

Le retour sur investissement peut être complexe à calculer pour ce type d'application. Il dépend de critères souvent immatériels dont l'évaluation la plus fine est dans la comparaison avant/après, critères qui plus sont souvent difficiles à quantifier numériquement. Néanmoins, il s'agit encore d'une méthode incontournable pour prouver l'intérêt d'un projet et on trouve souvent des critères d'évaluation de bon sens et contextuels qui donnent toute satisfaction.

Dès que c'est possible, il est intéressant d'évaluer des données objectives et de fixer des objectifs d'amélioration pour ce type d'application. On peut citer par exemple : le temps de traitement des courriers, dans le cas d'une GED de correspondance ; le taux de bonne application des normes et procédures, dans le cas d'une GED qualité ; le temps de traitement du renouvellement des contrats pour une contrathèque ou le taux de doublons d'une même image dans l'organisation pour une application de gestion de médias...

PRESENTATION DES CONCEPTS

Nous nous efforçons ici de positionner la gestion documentaire vis-à-vis d'un certain nombre de domaines ou de notions connexes.

Les concepts présentés ci-dessous ne représentent pas l'exhaustivité de l'environnement de la gestion documentaire, mais nous semblent nécessaires pour bien appréhender nos explications.

Document électronique

La notion de document, si elle ne pose guère de problème dans un univers matériel, doit être précisée dès lors qu'il s'agit de documents électroniques, de fichiers.

Notre approche de la gestion documentaire prend l'acception issue de la norme ISO : « Un document est l'ensemble constitué d'un support d'information et des données enregistrées sur celui-ci, sous une forme en générale permanente et lisible par l'homme ou par une machine. »

En matière de documents électroniques, le support est un format de fichier. Un fichier est créé, modifié et rendu lisible par une application. Par exemple, un document « .doc » est créé, modifié et rendu lisible par l'application Microsoft Word. Les informations quant à elles sont des données informatiques comprises dans l'enveloppe du fichier.

Nous verrons plus loin que les applications de GED introduisent la notion d'objet documentaire, qui réunit conceptuellement le fichier et les métadonnées qui lui sont attachées.

Contenu

Le terme « contenu » est employé pour désigner un **composant informatif cohérent**. L'objet documentaire, comme cité plus haut, est un contenu. Le terme s'applique à toute information numérique porteuse de sens.

Les contenus sont, le plus souvent, structurés, c'est-à-dire constitués d'un ensemble de champs d'information. Dans ce cas, on parle de contenus documentaires ou structurés.

Il est parfois difficile de faire la différence entre un ensemble de contenus et un document électronique. C'est souvent le contexte d'utilisation et les traitements réalisés qui permettent de distinguer la nature de l'information et le domaine fonctionnel concerné, et de définir le meilleur type d'outil pour la gérer.

Quelques exemples qui peuvent être considérés comme un document et/ou comme un contenu, pour illustrer notre propos : une brève sur un site d'actualité, une revue d'articles de presse, une image et sa légende, une fiche descriptive de produit, une notice de document, un catalogue de produits en format PDF, un module documentaire en format SCORM.

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 (le contenu informatif) et une forme (l'enveloppe de présentation).
Priorité	Gestion de documents électroniques.
Orientation	Stockage, partage, recherche.
Termes rencontrés	GED - Gestion Electronique de Document, EDM(S) - Electronic Document Management (System), 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 – Content Management System, souvent employé pour désigner la gestion de contenu Web.

Gestion de contenu

Objectif	Permettre la gestion des contenus numériques, au travers de fonctions de conception, diffusion, utilisation, recherche et archivage. On peut considérer la gestion de contenu comme un surensemble pour la GED et le WCM. Les solutions de gestion de contenu incluent le plus souvent des aspects procéduraux de workflow (BPM – Business Process Management) et de travail collaboratif (Groupware).
Priorité	La gestion de contenu d'entreprise (ECM) est le concept le plus récent de la gestion d'informations numériques, puisqu'elle adresse

	à la fois l'information structurée et non structurée, sur l'ensemble des étapes du cycle de vie des contenus numériques.
Orientation	Gestion des 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, BPM – Business Process Management, EIM - Enterprise Information Management.

A noter qu'on rencontre parfois le terme de gestions de contenu dans un sens plus limité, en voulant dire « gestions de contenu web » du fait de l'amalgame historique avec les CMS, précurseur de ce type d'outil, dédié aux sites internet.

Gestion de contenu multimédia

Objectif	Gérer les spécificités des « actifs multimédias numériques » (les Digital Assets, anglo-saxons) : images, musiques et vidéos.
Priorité	Contenus images et sons, navigation et recherche au sein du référentiel, gestion spécifique des droits d'auteurs.
Orientation	Métadonnées, navigation, recherche, DRM (digital right management), gestion des volumes importants.
Termes rencontrés	DAM – Digital Asset Management ou gestion des actifs numériques.

Il s'agit d'une spécialité au sein de la GED. De nombreuses fonctionnalités sont communes quant à la catégorisation, la gestion des droits et cycles de vie. Mais si les métadonnées spécifiques à la gestion des droits d'auteurs peuvent être modélisées avec n'importe quel outil de GED, certaines fonctionnalités spécifiques sont propres au DAM, manipulation d'images, gestion des droits d'auteurs, panier de sélection, extraction des imagettes (miniature ou thumbnail), prévisualisation de vidéos, gestion des descripteurs EXIF ou IPTC, notamment.

Workflow - processus

Objectif	Permettre de formaliser et de fluidifier des processus, via l'enchaînement de tâches unitaires entre différents acteurs.
Priorité	Gestion de processus et de formulaires. Le BPM couvre les besoins d'interactions formelles entre des acteurs de l'activité de l'organisation, qu'il s'agisse d'humains ou de systèmes, puisque les

	processus font souvent interagir les deux.
Orientation	Gestion de tâches attachées à des profils d'utilisateur. Cependant, il s'agit le plus souvent d'outils qui permettent « d'orchestrer » des webservices produits par différentes applications en vue d'organiser des actions unitaires simples pour produire un résultat complexe. On notera que certains outils sont spécialisés par rapport à des objets auxquels ils se rapportent : par exemple le « Docflow » à un document.
Termes rencontrés	BPM – Business Process Management, gestion de formulaire, workflow, flux de travail, cycle de vie, instance de processus.

Il est à noter que les processus rencontrés dans l'univers de la gestion documentaire sont très liés au document. Ils sont exploités à des fins d'orchestration de tâches de validation ou de publication essentiellement. Cela constitue une petite part de ce qu'il est possible de faire avec un outil de BPM autonome.

Archivage – Record Management

Objectif	Permettre de gérer les phases de vie des documents après leur période d'utilité (la Durée d'Utilité Administrative)
Priorité	La préservation des enregistrements (pièces) d'archives.
Orientation	L'ensemble du processus de gestion archivistique à même de garantir la prise en compte, la sécurisation, la fiabilité et la pérennité des archives de l'organisation. Notons que les applications de RM sont de plus en plus fréquemment « fondues » dans les applications de GED. En effet, il a été démontré que l'archivage est mieux pris en compte, lorsqu'il est géré dès la création de la pièce d'archive (dossier, document...).
Termes rencontrés	SAE – système électronique d'archivage, RM – Record Management, archivage, archivistique, gestion du cycle de vie.

Travail collaboratif

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

	partagé ...
Priorité	Faciliter le travail en équipe.
Orientation	Partage d'informations opérationnelles
Termes rencontrés	groupware – travail en équipe, Chat (messagerie instantanée), Blog (mode de communication personnel via des billets thématiques), Wiki (mode de partage d'information via la coédition de pages), Agenda partagé, Document multi-auteurs (comme Google Docs, par exemple, qui permet la collaboration synchrone et asynchrone).

Moteur de recherche

Objectif	Permettre de retrouver une information existante.
Priorité	Indexation ou interrogation de bases de contenu.
Orientation	L'objectif est de permettre aux utilisateurs de retrouver une information. Les outils de recherche se concentrent, d'une part, sur l'indexation de ces contenus ou l'usage d'index externes dans le cas des métamoteurs et, d'autre part, sur les outils d'amélioration de la pertinence (adéquation de ce qui est cherché, avec ce qui est retourné par le moteur).
Termes rencontrés	Moteur de recherche, Search Engine, SEO (Search engine optimisation), Findability, Meta moteur, Crawler, Opérateur de recherche.

VERS LA GESTION DE CONTENU D'ENTREPRISE

Depuis quelques années déjà, la prédominance du terme d'ECM (Enterprise Content Management), pour gestion de contenu d'entreprise, tend à supplanter les acronymes DMS et GED.

Cette évolution de la « gestion documentaire » à la « gestion de contenu d'entreprise » traduit une certaine réalité, au-delà de l'aspect marketing évident qui la caractérise.


open
source

Le périmètre de la GED

La Gestion Electronique de Documents traite des documents numériques. A ce titre une solution de GED peut intégrer :

- **des outils** : de numérisation, de stockage, de circulation, de diffusion et de recherche, notamment
- **des spécificités métiers** : application des procédures qualité, gestion du cycle de vie, règles de gestion, interaction avec des applications métier, etc.
- **des spécificités techniques** : transformation de format de fichier, prévisualisation de fichiers multimédias, pages Web, contenus structurés, groupe de travail, etc.

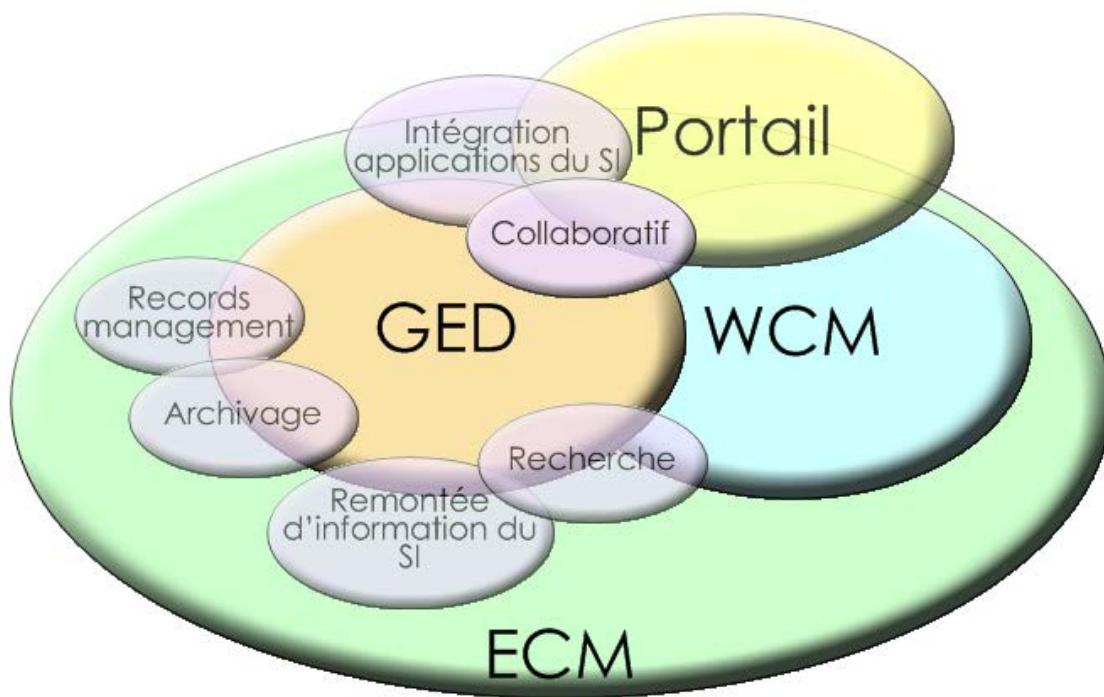
Les applications de GED « pures » n'existent quasiment plus : elles intègrent presque toujours des fonctions empruntées aux domaines évoqués plus haut, ou à des domaines métier.

Le concept d'ECM

La gestion de contenu d'entreprise (traduction d'Enterprise Content Management) regroupe les solutions à même de gérer l'ensemble des contenus numériques d'une organisation.

Il s'agit de prendre en compte (acquérir, capturer) les informations électroniques (structurées ou documentaires) pour les gérer (stockage, édition, diffusion) en répondant aux exigences des utilisateurs (ergonomie, fonctionnalité) et aux processus de l'organisation (sécurité, fiabilité, processus).

Bien que couvrant essentiellement le domaine technique, l'ECM intègre ou se réfère donc à une part importante de méthodes et de modélisation.



Ce schéma 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 certain nombre de domaines connexes.

Notons que la notion de portail d'entreprise est fréquemment associée à celle d'ECM. Bien que fonctionnellement, les intersections soient faibles, l'aspect globalisant des solutions d'ECM ne doit pas masquer la finalité fondamentalement différente des deux outils. Exposer les outils/services de l'organisation (portail) versus gérer les contenus de l'organisation.

La capacité pour une solution d'ECM ou de GED à s'intégrer à un portail est une qualité importante. Notamment, lorsque le service informatique de votre organisation ne souhaite pas agréger dans une seule application l'ensemble des solutions d'accès à l'information.

Evolution des outils

Les solutions de GED étaient initialement orientées vers la gestion de fichiers, puis vers la prise en compte de documents spécifiques, avec une typologie de plus en plus large (mail, formulaire, image, vidéo...). Leur évolution naturelle est donc de s'intégrer de mieux en mieux aux processus de gestion documentaire, notamment par une meilleure prise en compte du contexte d'édition et de la structure des fichiers document (MS Word, Open Document, PDF...) qui permettra, demain, de traiter les documents comme une agrégation de contenus.

L'évolution de la couverture fonctionnelle des solutions de GED tend à résoudre différents problèmes liés aux usages. D'une part en intégrant la gestion des contenus structurés en parallèle de celle des contenus documentaires (on parle parfois de contenu semi-structuré) et d'autre part en permettant la réalisation d'actions collaboratives poussées.

L'ECM recouvre le périmètre de la GED et l'étend aux autres contenus d'entreprise, notamment les contenus Web (WCM), voire les contenus gérés par des applications tierces, on parle alors de fédération.

Translation des méthodes

L'évolution dans la gestion des contenus suit deux logiques bien distinctes : d'une part, un élargissement en termes de contenus gérés et, d'autre part, une adaptation des méthodes de gestion à ces contenus, dérivées de la gestion documentaire.

On constate, en effet, que nombre des méthodes de l'ECM sont issues de la GED. Un contenu est souvent considéré comme un fichier (ou un ensemble de fichiers) et les actions que l'on applique s'apparentent à celles qu'un utilisateur est familiarisé à appliquer sur des documents.

Deux facteurs expliquent cette translation de méthodes entre GED et ECM.

1. Les solutions d'ECM sont souvent issues des éditeurs de GED, ceux-ci tentent d'élargir leur marché en élargissant le périmètre de leur solution.
2. Et c'est surtout les utilisateurs, en quittant le paradigme issu de leur vision naturelle (culturelle) du document papier - séparation *entre document (forme) et contenu (fond)* - lorsqu'ils gèrent des documents ou des contenus électroniques qui conduisent naturellement les fournisseurs de solution à vouloir eux aussi s'affranchir de cette dichotomie et gérer ensemble tous les contenus électroniques.

Il est en effet logique de constater que les utilisateurs envisagent plus facilement de gérer un contenu considéré comme un ensemble cohérent, que les morceaux d'un « quelque chose » qui n'a pas de forme et dont il est complexe de s'approprier le rendu (la forme).

En clair, une page de site web est plus facile à considérer et à manipuler que plusieurs contenus (par exemple, trois images, une bannière de publicité et plusieurs blocs de texte), qui une fois agencés sur l'écran, forment une page Web.

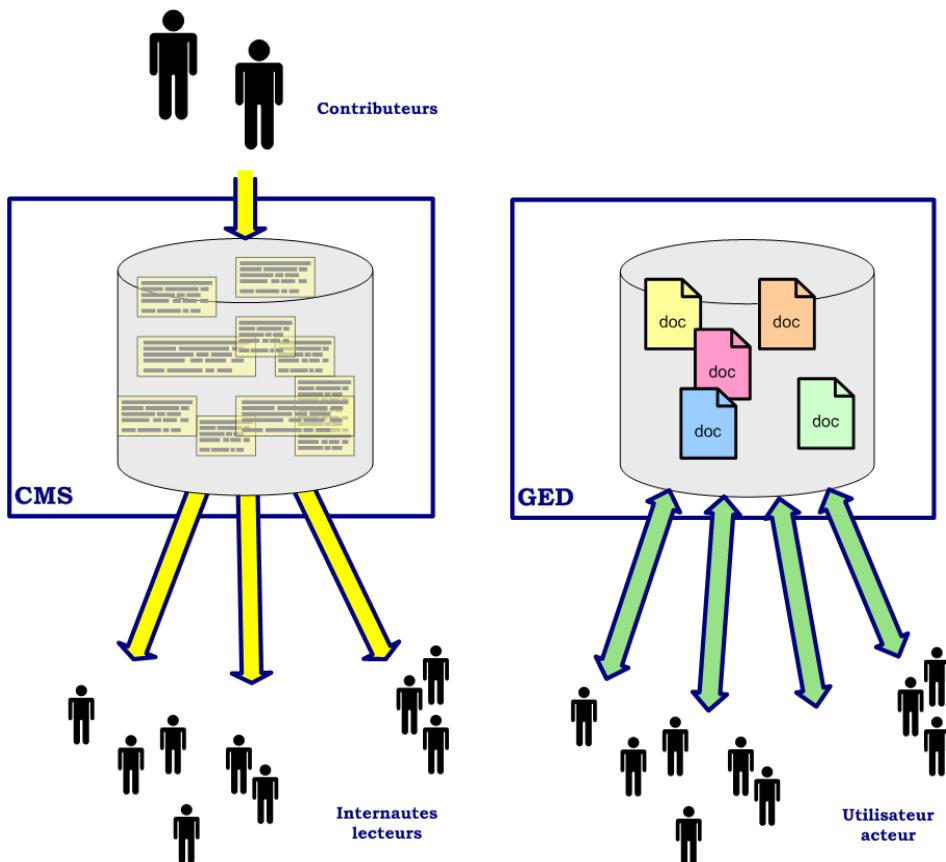
ENJEUX PRINCIPAUX DE LA GED

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



open
source

Les **outils de gestion de contenu Web** s'orientent 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 doit donc impliquer les futurs utilisateurs de façon importante.

Nous constatons que le succès des déploiements d'applications de GED repose essentiellement sur la bonne utilisation d'un certain nombre d'outils et de principes décrits ci-dessous.

L'ergonomie de l'application doit s'adapter à l'usage d'une application de gestion, utilisée par un nombre important de personnes, au même titre qu'une application de messagerie, par exemple.

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 celles 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).

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.

Pour 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 typiques incluent, par exemple, le titre, l'auteur, la description, la langue et la date de publication.

Il existe des nomenclatures standard de métadonnée, notamment le Dublin Core, qui définit 15 champs principaux de métadonnées. Différents métiers disposent de leur propre standard (musique, architecture, santé ou archivistique, par exemple).

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, information textuelle (champ texte), liste de valeurs issues de tables de références, dates ou nombres sur la base desquels des règles de gestion pourront être calculées...

Référentiel de classement

La plupart des solutions de gestion de contenu intègrent un ou **plusieurs référentiels 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. Les usages qui reposent sur cet outil constituent un facteur non négligeable dans le choix d'une solution.

Les outils de GED sont plus ou moins élaborés, mais permettent à minima de définir un plan de classement c'est-à-dire une arborescence hiérarchique par rapport à laquelle sont classés les documents. Ils peuvent aller jusqu'à permettre de gérer des vocabulaires métiers avec synonyme, liens de proximité sémantique, multihiérarchies et traductions de termes. On parlera alors de thésaurus ou d'ontologie.

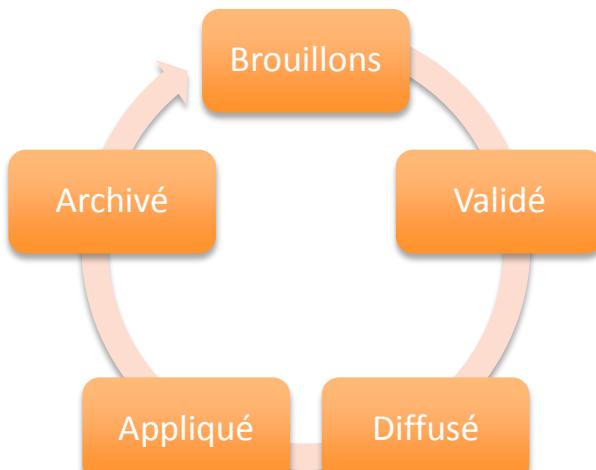
L'utilité d'un référentiel de classement concerne différentes fonctions : aide à l'indexation, vocabulaires communs (glossaires), recherche avancée (avec prise en compte des synonymes, de la proximité sémantique, par exemple), ou encore navigation par facettes.

Prise en compte du cycle de vie

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 en mettant en œuvre un ensemble de processus avant de pouvoir être utilisé. Par exemple, un contrat est rédigé, modifié, validé, imprimé, signé puis envoyé, souvent sous forme papier. Le cycle de vie inclut toutes ces phases, depuis son élaboration jusqu'à son archivage définitif ou sa destruction.

La gestion du cycle de vie des documents électroniques doit être réalisée au moins aussi finement que celle des documents papier. En effet, du fait de la facilité de duplication, de transmission et de stockage des documents électroniques, leur nombre est nettement plus important. Ceci induit différents problèmes, notamment de traçabilité entre versions, de capacité à trouver une information précise ou encore de volume de stockage.



Exemple de cycle de vie simple

Dématérialisation et transformation numérique

Les solutions de GED sont à même de gérer différents types de documents, qu'ils soient issus d'applications bureautiques ou d'un processus de numérisation.

La numérisation ou dématérialisation (également appelée « démat ») consiste en la transformation d'un document d'un support matériel en données électroniques. Il peut s'agir par exemple de la numérisation d'un document papier vers un document bureautique, 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, etc.

Depuis de nombreuses années, le « zéro papier » est présenté comme un objectif, même si tous s'accordent sur le fait que le zéro ne sera probablement jamais atteint. La dématérialisation permet néanmoins de bénéficier des avantages du numérique sur l'ensemble des contenus dématérialisés.

Certaines organisations ont intégré cela et leurs projets de gestion d'informations s'orientent vers le « tout-en-un » qui permet de bénéficier d'accès et de modes de gestion unifiés pour tous les contenus, qu'ils soient physiques et numériques.

LES PRINCIPALES SOLUTIONS DE GESTION DOCUMENTAIRE OPEN SOURCE

ALFRESCO

Alfresco (www.alfresco.com) est un éditeur anglais, fondé en 2005 par d'anciens dirigeants de Documentum et Business Object. La société est aujourd'hui présente en Angleterre, en France et aux Etats-Unis et compte plus de 3000 clients de par le monde.

Alfresco est une solution J2E (java) qui propose un large socle de fonctionnalités : métadonnées, typage des documents, workflow documentaire et BPM (avec le module **ACTIVITI** (www.activiti.org), gestion de catégories, outils collaboratifs, recherche (y compris un module SOLR), gestion de bases indépendantes, stockage et publication de contenus Web, transformation (bureautique et multimédia), pour citer les principales.

Tableau de bord d'un utilisateur

Alfresco se distingue par différents apports :

- une orientation collaborative à la fois au niveau des fonctions et de l'interface native

**open
source**

- une architecture fonctionnelle et technique de très bon niveau, lui permettant de se positionner en tant que socle de GED pour de très grandes organisations. Alfresco est très ouvert et ses APIs et ses outils de développement sont nombreux et bien documentés
- les règles de gestion, directement accessibles aux utilisateurs, permettent de déporter une part des aménagements fonctionnels au niveau des gestionnaires et non plus des développeurs ou des administrateurs
- un composant de gestion des contenus web, parallèle à la gestion documentaire. Ce module est dédié à la gestion des contenus structurés à leur distribution sur des frontaux de consultations distants
- les facilités d'intégration avec le poste de travail. On citera notamment une bonne utilisation du protocole Sharepoint™ et Webdav pour l'édition en ligne ou encore la synchronisation avec le poste de travail (via un outil de type DropBox)
- un module de Record Management, c'est-à-dire de gestion d'archives, basé sur la norme américaine DoD 5015.02 (Smile a développé des modules dédiés à la gestion d'archives, orientés sur les règles européennes - Moreq notamment)
- une interface d'administration permet de garder une vue sur les fonctions principales et notamment les modules de la version « entreprise »
- des modules techniques avancés, tels que le load balancing ou la gestion de plusieurs instances dans une seule installation (multi tenancy)
- l'interface Share apporte une interface prête à l'emploi pour la collaboration
- une visibilité très importante sur la scène internationale, du fait de sa communication, de son activité marketing et du grand nombre de clients
- Enfin, la double orientation Cloud et mobile accentuée avec les dernières versions apporte une valeur ajoutée complémentaire - « l'hybridation » - qui autorise des architectures d'information innovantes pour le partage et la valorisation des contenus.

Alfresco Share Preview | My Dashboard | My Profile | Sites | Help | Search This Site | Logout

Share Project Site

Site Dashboard | Wiki | Blog | Document Library | Calendar | Discussions | Members

New Folder | Upload | Selected items... | RSS Feed

Up | Documents > Specifications | Select... | Show Folders | Simple View

Documents

- I'm Editing
- Others are Editing
- Recently Modified
- Recently Added

Tags

Library

- Documents
- Analysis
- Personnel
- Plans
- Prototypes
- Specifications
- UI Mock-ups
- Use Cases

Alfresco 3 PR Brief.ppt
Modified on: Thu Jul 31 2008 16:58:52 | Modified by: John Newton | Version: 1.0
Description: (None)
Tags: (None)

Alfresco Labs 3 draft v12a.doc
Modified on: Thu Jul 31 2008 16:58:44 | Modified by: John Newton | Version: 1.0
Description: (None)
Tags: (None)

Budget Plan 20080617.xls
Modified on: Thu Jul 31 2008 16:58:45 | Modified by: John Newton | Version: 1.0
Description: (None)
Tags: (None)

Download | Edit offline | Details | More >

Alfresco Share | My Dashboard | My Profile | Sites | People | Help | Logout | Search This Site

Linton's Test Site

Site Dashboard | Wiki | Blog | Document Library | Calendar | Discussions | Members

Location: Documents > Reqs

Seamless Edit.txt | Full Screen | Prev | Page 1 of 2 | Next | Jump to page: []

Project: Jira_3d
Functional Area: Document Edit
Note: The behaviour is different when a document has just been opened, see Seamless Open.
TBD: Is this Share only or should we support Alfresco Explorer?
Story: None
Requirements: A user needs to update an Alfresco managed document. Using Share, they locate the document they would like to edit either by browsing a Site Doclib or searching. The user selects a document and selects 'Edit' from the context menu. The document opens directly in Word and the user makes their updates. When the user saves the document, Alfresco manages the document in a normal way. User checks in there change via the Share UI using a 'Done Editing' option. Alternatively, user can also checkin their changes via a lightweight desktop client (LOC) which shows a list of documents I'm Editing with Done Editing option against each doc.
Functional overview: User can edit Alfresco managed documents directly via their browser. They do not need to 'download' and manually save a copy locally. A desktop component, possibly AIR based, will automatically manage reading/writing between the browser and the client. For example, it may save the document to a well-known location on the client, e.g. My Documents\Alfresco Documents\SiteName.
TBD: How is the local working area defined, by an administrator, by the end user, by client?
Requirements: Edit Option against content in Share client
Done Editing option for content that has selected for Edit
Option to enter version information and version comments when done editing
LOC (assume AIR based for now) shows list of Documents I'm Editing and highlights documents that have been changed locally but not updated.
UI Hint: See Eric's AIR demo

Metadata

Name: Seamless Edit.txt
Content Type: Plain Text
Title: Seamless Edit.txt
Description:
Size: 2 KB

Creator: Linton Baddeley
Created on: Wed 15 Oct 2008 17:39:13
Modifier: Linton Baddeley
Modified on: Wed 15 Oct 2008 17:39:13

Document Actions

Download | Edit Metadata | Assign Workflow... | Copy to... | Move to... | Delete

Share

Download File URL: http://eng.alfresco.com:8080/s

Tags

(None)

Permissions

Managers: Manager
Collaborators: Collaborator
Consumers: Consumer
Everyone: Read

Document URL

http://eng.alfresco.com:8080/s

This Page URL

http://eng.alfresco.com:8080/s

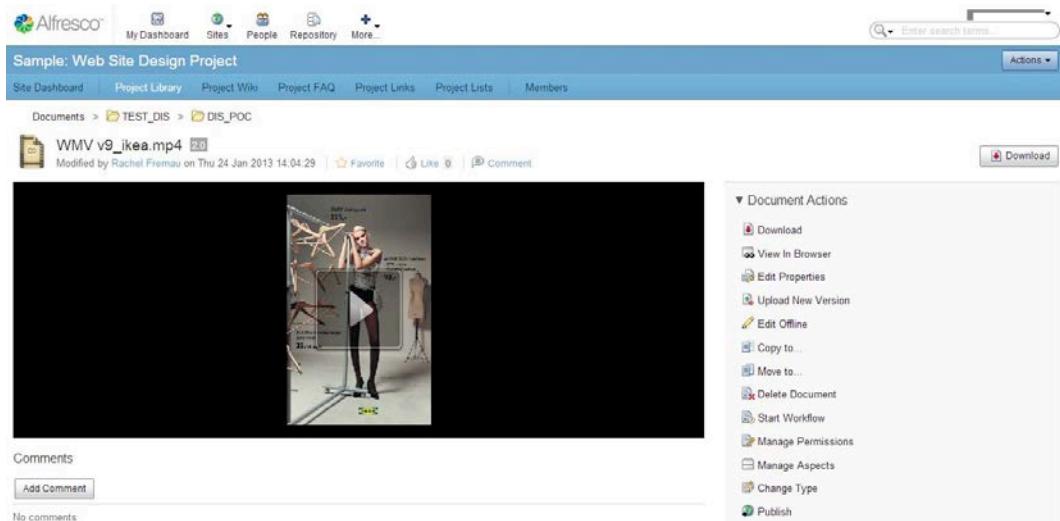
open
source

Alfresco est disponible en deux versions : une version Community (ou Labs), gratuite et sous licence GPL, et une version « Entreprise », dont le code est ouvert sous une licence commerciale, qui donne accès au support éditeur, aux différents modules. Le montant de cette souscription dépend de différents facteurs : le niveau de service de support attendu (SLA), les modules mis en œuvre et le nombre de processeurs utilisés.

Alfresco dispose d'une forte dynamique de développement, ainsi que d'une communauté d'utilisateurs et de développeurs importante. On notera l'existence du site Alfresco Forge (<http://forge.alfresco.com>) qui héberge de nombreux plug-ins.

La version 4, sortie en décembre 2011, intègre de nombreuses améliorations, parmi lesquelles nous citerons :

- Les améliorations notables de l'interface collaborative Alfresco Share avec notamment le glisser/déposer, plus de finesse de personnalisation et de gestion des droits pour les espaces collaboratifs, enrichissement de la console d'administration, la catégorisation de résultats
- l'intégration d'un serveur de transformation amélioré pour les fichiers bureautiques basés sur MS Office™
- un plug-in intégré au poste de travail de l'utilisateur permettant à tout un chacun de disposer d'une sélection des documents d'Alfresco sur son poste de travail (fonction telle que la DropBox)
- l'amélioration des fonctions sociales (suivi des personnes, des contenus les plus aimés, des contenus modifiés, partage et publication vers des sites sociaux)
- **l'architecture hybride** : synchronisation vers des espaces en mode « Cloud » et accès Mobile. Il s'agit sans doute d'une percée majeure dans le domaine de la gestion de contenu puisque pour la première fois, la mise à disposition à l'extérieur du firewall de l'entreprise est largement simplifiée et accessible à l'utilisateur final ; en toute sécurité
- l'amélioration de la gestion des images et vidéos.
Ceci passe par la dernière nouveauté d'Alfresco, non encore publié au moment où nous écrivons ses lignes, mais que nous avons eu l'occasion de tester et qui viendra compléter les fonctions de gestions des médias (images et vidéos). Ce module (Média Management) viendra s'intégrer dans l'interface standard et y apportera des fonctions non seulement dans l'interface mais également côté serveurs, comme des transformations.



NUXEO

Nuxeo (www.nuxeo.fr) est une société française, éditrice de la solution de GED open source éponyme depuis 2001. La société est aujourd'hui présente en France et aux Etats Unis.

Nuxeo est une solution J2E (java), transcrive du Python (Nuxeo CPS) en 2007, et qui reprend le périmètre complet de la gestion de contenu d'entreprise : métadonnées, types de documents, workflow avancé, gestion de catégories, fonctions de collaboration, recherche, gestion de contenu complexes, multibases, outils de transformation bureautique et multimédia, paramétrage via un outil graphique.

The screenshot shows the Nuxeo Documentation interface. At the top, there's a navigation bar with links for Home, Document Management, DAM, Collaboration, Admin Center, Studio, and Administrator. Below the navigation is a search bar with options for Recherche avancée and Recherche intelligente. The main area is titled "Documentation" and contains a sub-menu with Contenu, Modifier, Notifications, Historique, and Administration. On the left, there's a sidebar showing a tree view of workspaces, including Default domain, Sections, Templates, Workspaces (Finance, Headquarters, Human Resources, IT Department, Expertise Center, IT Mailbox, Maintenance Team), Projects (Database Creation, Discussions, Documentation, Planning, Requirements), Marketing, Retail, and Template usage samples. The main content area displays a list of files with columns for Titre, Dernière modification, Dernier contributeur, and Version État. The files listed are Bon de commande.pdf, Image1.jpg, Java_softwarePlatform.doc, Nuxeo_BootCampProgram.odt, and project_report.xls. At the bottom of the list are buttons for Copier, Coller, Ajouter à la sélection, Supprimer, and Comparer. A "RSS ATOM" link is located in the bottom right corner of the main content area.

L'outil se distingue par différents aspects :

- **une notion de « relations »** avancée qui permet de typer à la volée des liens entre contenus. Par exemple « est la traduction de / est traduit de » ou « a pour pièce jointe / est la pièce jointe de »
- une interface standard facilement utilisable pour des projets de GED simple et dotée d'une ergonomie satisfaisante (cliquer-glisser, clic droit, présentation en onglet, arborescence multiple),
- la notion de **publication** qui permet de dé-corréler l'espace de travail des utilisateurs de ce qui est donné à voir aux différents publics
- la structure technique de la solution permet de l'intégrer aussi bien dans de petites structures, que de grands groupes internationaux, et même d'être intégrée comme brique de gestion documentaire par d'autres projets (le Portail-ESUP par exemple)
- « l'Admin center » est une interface de gestion (backoffice) riche qui se révèle suffisante à l'usage

**open
source**

- Nuxeo Studio est un outil de configuration graphique qui permet de mettre en œuvre de façon opérationnelle la plupart des fonctions disponibles dans l'outil

Enfin, notons la « Marketplace », site de dépôt de plug-ins prêts à intégrer et facilement installable via l'Admin Center.

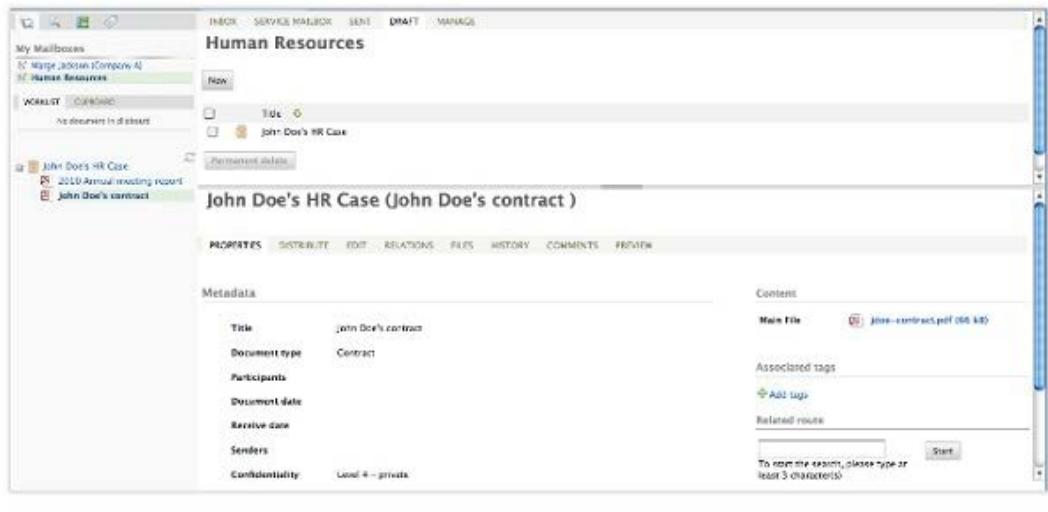
Plusieurs distributions complète l'offre de Nuxeo. Une est destinée à la gestion des médias (images et vidéos), Nuxeo DAM. Et une est essentiellement destinée à la gestion de dossier et de courrier, Nuxeo CMF et Nuxeo Collaboration dédié au collaboratif d'équipe.

Nuxeo CMF est un framework de Case Management pour gérer les dossiers au sein de la Nuxeo Platform

Principales fonctionnalités :

**open
source**

- **content routing** pour la configuration des workflows
- capture des mails
- diffusion par mailing list des dossiers à traiter.



Nuxeo DAM, pour la gestion des contenus multimédia (vidéo, audio, image).

Principales fonctionnalités :

- acquisition des contenus et modification des métadonnées en masse
- annotation des images
- prévisualisation des contenus
- encodage, conversion des formats
- gestion des droits.

**open
source**

Détail d'une image

Planche contact

Nuxeo Collaboration, orienté vers la gestion d'équipe et le partage « social ».

La version 5.6 est sortie en septembre 2012 et apporte de nouvelles fonctionnalités :

- Pour Nuxeo Studio
 - un nouvel outil de création graphique et de gestion de workflow - **Content Routing** - il est destiné à la fois aux utilisateurs finaux de niveau avancé et aux développeurs
 - le **Tab Designer** qui permet désormais aux utilisateurs finaux de niveau avancé de gérer le paramétrage du contenu et des vues de chaque onglet d'un objet
 - le **rechargement à chaud** de la plateforme pour une intégration immédiate des modifications effectuées depuis Nuxeo Studio a été amélioré

**open
source**

- la **définition de propriétés complexes** qui sera mise à profit pour le paramétrage dans les formulaires et les vues au travers des schémas de contenu
- La synchronisation bidirectionnelle de fichiers via **Nuxeo Drive**, sur les postes utilisateurs des utilisateurs nomades (équivalent de Dropbox) est *en release candidate*
- Le partitionnement ou **multi-tenant** pour optimiser l'utilisation des ressources système, notamment dans les projets qui utilisent plusieurs bases documentaires sur un serveur unique
- **La comparaison automatique des documents** pour assurer un suivi encore plus fin des modifications du contenu, des métadonnées et des propriétés
- Le développement du mode collaboratif avec l'évolution du module **Social Collaboration** qui intègre un nouveau « mur » (Wall), le vote (Like) sur les documents ainsi que le déclenchement des notifications via un bouton dédié
- Une interface utilisateur améliorée
 - la modification de l'ergonomie de la page résumé
 - une barre d'actions personnalisables
- Enfin, le Framework d'application mobile reprenant les fonctionnalités les plus utilisées (recherche, navigation, prévisualisation, fil d'activités) permettra de déployer simplement des applications mobiles basées sur Nuxeo

KNOWLEDGE TREE

Knowledge Tree (www.knowledgetree.com) est une solution de GED développée par la société sud-africaine JamWarehouse. Knowledge Tree existe en deux versions, Enterprise et Community clairement différenciée dans [cette page](http://tinyurl.com/c9zv5ry) (<http://tinyurl.com/c9zv5ry>). La solution dispose d'un ensemble complet de fonctionnalités de GED et plusieurs modules qui permettent une bonne intégration dans l'environnement bureautique.

La version open source de Knowledge Tree semble de plus en plus en replis par rapport à la version commerciale, bien qu'elle intègre une bonne partie de l'offre. Plusieurs modules ou fonctions importantes ne sont pas inclus dans la version Community, notamment : l'intégration à MS Office, des fonctions de personnalisation et de confort, la comparaison de document, les alertes, le reporting, la signature électronique ou la traduction de l'interface dans différentes langues (notamment en français !).

Notons que la licence logicielle de KnowledgeTree est spécifique, il s'agit d'une licence open source basée sur le Mozilla Public License avec des clauses supplémentaires relatives à la marque et la rétention copyright. Ceci dénote d'une volonté de l'éditeur de verrouiller, peut être un peu trop, sa solution.

The screenshot displays the KnowledgeTree GED application. At the top, there's a navigation bar with links like 'Invite Users', 'Settings', 'Peter Mollins', 'Applications', 'Get Help', and 'Logout'. Below the navigation is a search bar. The main content area shows a document titled 'Banking Customer Proposal' with a preview, version history (0.2), and modification details (Peter Mollins on Feb 5, 2013). It also lists document properties such as type (Word Document - 155.44KB), ID (25340), and tags. To the right of the document view are several sidebar modules: 'My Recent Documents' (listing 'Sales Proposal Temp...', 'Banking Case Study', etc.), 'My Recent Folders' (listing 'Proposal Collateral', 'Finance Industry Pro...', etc.), 'Pending Document Alerts' (no alerts found), and 'Workflow Transitions' (document not in a workflow). Below the main document view are two smaller windows showing file management and analytics. The left window shows a tree view of documents like 'Contracts', 'Correspondence', and 'Proposals'. The right window shows charts for 'Views, Comments and Likes', 'Uploads per Month', and 'Document Length', along with a 'Most Analyzed' section.

L'application est globalement de bonne facture et permet de disposer d'une GED simple et efficace. Nous apprécions particulièrement plusieurs points :

- une interface standard épurée et rapidement opérationnelle
- une recherche avancée qui satisfera les plus exigeants
- des fonctions d'administration, accessibles pour une large part du paramétrage
- des modes de navigation virtuelle implémentés par défaut, notamment par type de document
- l'ergonomie du module intégré à MS Office qui permet, contrairement à ses concurrents, de gérer les métadonnées de la GED depuis la bureautique.

La dernière version propose plusieurs avancées importantes :

**open
source**

- une verticalisation pour les équipes de vente et de marketing orientée vers un usage SaaS (nouvelle stratégie de la société)
- la refonte graphique et ergonomique très importante qui fait suite à la refonte technique de la version précédente (basé sur Zend)
- l'intégration des protocoles : CMIS, permettant d'interroger la solution via des webservice normalisés et Open Search permettant de rechercher dans les contenus et de récupérer des réponses de son moteur de recherche via des méthodes normalisées

Notons enfin que l'éditeur continue de s'éloigner sensiblement de l'open source, non seulement par la licence, mais également par son approche SaaS ou encore par le renforcement de la version Entreprise au détriment da la version communautaire.

JAHIA

Jahia est une plateforme intégrée de portail web et de gestion de contenu, proposée par l'éditeur suisse du même nom.

Elle est proposée sous deux distributions : Jahia Community et Jahia Enterprise. La première est open source sous licence GPL v2, la seconde est sous licence commerciale, et intègre notamment des fonctions à destination des organisations et est supportée par l'éditeur.

Cette solution répond principalement aux besoins de gestion de contenus et documentaires Web, ainsi qu'à une partie des besoins de portail (agrégation), notamment avec le support de la norme JSR 168.

Jahia permet également de gérer un référentiel de fichiers à la norme JCR, ceux-ci peuvent être utilisés par les différentes interfaces de Jahia.

Jahia est bâti sur un socle regroupant le stockage des contenus et les fonctions liées : **Jahia Content Platform**. Sur ce socle, deux solutions principales sont proposées :

- **Jahia xCM** (eXtended Content Management) pour la gestion de contenu
- **Jahia Wise**, une solution de partage documentaire, sociale et collaborative.

Dans cette nouvelle édition du livre blanc, une place particulière est donnée à Jahia Wise, nouveau produit, orienté GED et collaboration de Jahia.

L'intérêt des fonctionnalités GED de Jahia xCM réside à la fois dans l'usage qui peut être fait des contenus documentaires dans un contexte de site (internet, intranet, application web...) ; mais aussi avec Jahia Wise sur des aspects plus collaboratifs. *Notons que Jahia Wise est disponible uniquement aux clients de la version entreprise.*

Jahia permet des fonctionnalités appréciables sur les images : rotation, recadrage en ligne, redimensionnement, traitements simples, vignettes automatiques, extraction des données EXIF.

Des évolutions importantes sont arrivées avec la version 6.5, notamment autour de la gestion documentaire :

- possibilité d'accès aux fichiers via CIFS, SMB
- amélioration des fonctions de recherche sur les documents : possibilité d'enregistrer des requêtes, autocomplétion, support des facettes y compris des facettes hiérarchiques
- gestion de versions sur les documents au sein du gestionnaire de fichiers
- intégration automatique des mails
- utilisation graphique de workflow avec visualisation de l'avancement
- intégration dans l'interface Wise de fonction de blog, wiki, forum et de gestion des tâches, outils pratiques voire indispensables pour le travail collaboratif

Jahia Wise, qui nous intéresse donc plus particulièrement dans ce livre blanc, présente une approche collaborative du partage de documents orientée sur les fonctionnalités clés pour les utilisateurs, notamment suivi et historique des versions, verrouillage et possibilité de restaurer des versions antérieures ; drag-and-drop web; gestion des tags et des catégories ; import en masse de documents et auto-tagging ; recherche multi-facettes ; affichage des images sous forme de galerie.

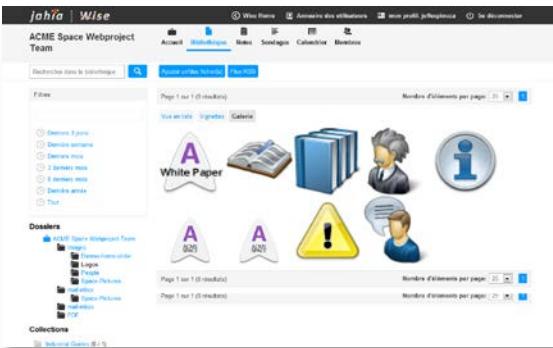
Jahia Wise est structuré autour de la notion d'espaces collaboratifs, qui peuvent refléter tout type d'organisation souhaitée par le client : les métiers de l'entreprise, les directions et services, les projets ou autres...

The screenshot shows the Jahia Wise library interface. At the top, there's a navigation bar with links for Spaces Home, Users directory, Profile, Logout, and various administrative icons. Below the navigation is a search bar and a toolbar with options like Upload a file, Create Folder, Delete, Cut, Copy, Rename, Permissions, and RSS Feed. A sidebar on the left contains a 'Tags' section with items like 'satellite' (6), 'space shuttle' (6), 'Mars' (5), and 'Jahia' (1). It also has a 'Folders' section listing 'Web project team space', 'Official pictures', 'Press Release', and 'mail-inbox'. The main content area displays a list of files with columns for TYPE, FILE NAME, DOWNLOAD, CREATION DATE, LAST MODIFICATION, AUTHOR, and WORKFLOW. The first item listed is 'mail-inbox' (May 21, 2012, by bill). The second item is 'Official pictures' (May 21, 2012, by bill). The third item is 'Press Release' (May 21, 2012, by irina).

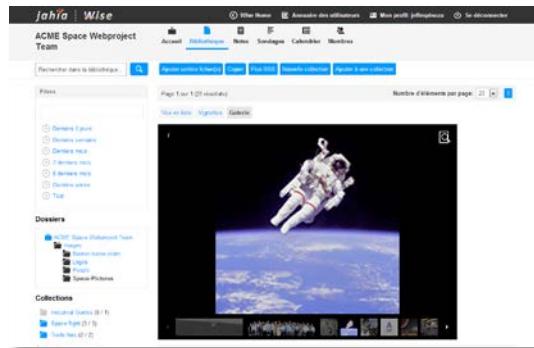
TYPE	FILE NAME	DOWNLOAD	CREATION DATE	LAST MODIFICATION	AUTHOR	WORKFLOW
file	mail-inbox		May 21, 2012	May 21, 2012	bill	
file	Official pictures		May 21, 2012	May 21, 2012	bill	
file	Press Release		May 21, 2012	May 21, 2012	irina	

Un dossier de la bibliothèque et les tags associés.

**open
source**

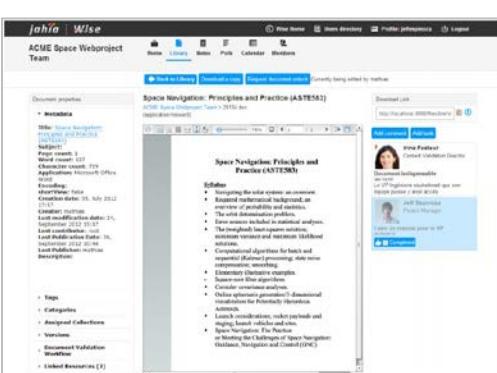


Une galerie d'images dans Jahia Wise en mode « vignettes ».

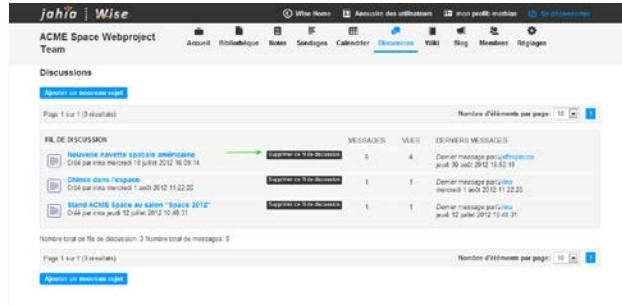


Une galerie en mode « diaporama ».

Collaboration sociale : les espaces collaboratifs permettent aux utilisateurs d'échanger des commentaires sur des projets ou des documents, de soumettre des documents à un workflow de validation ou de publication, d'affecter des tâches et de suivre leur avancement. Les fonctionnalités sociales telles que les blogs, les sondages, les forums, les wikis, la recherche d'utilisateurs et de compétences, les notifications et les collections partagées offrent aux utilisateurs le meilleur des pratiques du web 2.0.



Collaboration autour d'un document
(édition, commentaire et tâche).

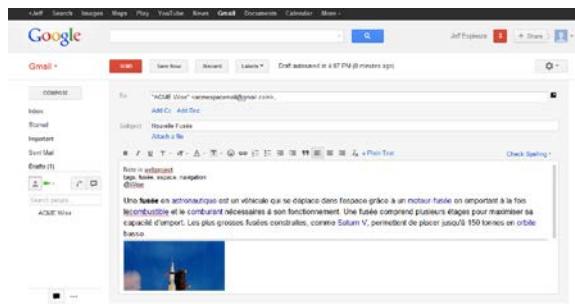


Forum de discussion.

Contenu Web : Jahia Wise peut être installé de façon autonome dans un système d'information ou en synergie avec Jahia xCM et être rapidement opérationnel sans développement spécifique. L'intégration de Jahia Wise avec Jahia xCM permet aux entreprises d'abolir les barrières entre conception et partage de documents collaboratifs d'un côté et publication de documents ou ressources validées sur leurs sites internet de l'autre.

**open
source**

De plus, Jahia Wise offre la possibilité innovante d'envoyer du contenu dans un espace collaboratif à partir d'un simple e-mail, qu'il s'agisse de contenu trouvé sur le web, ou d'une conversation par messagerie électronique, afin de l'intégrer aux ressources collaboratives de l'espace.



Jahia a donc pris un net virage avec Jahia Wise et propose désormais des fonctions et une interface résolument orientée vers le travail collaboratif autour des documents. Sans toutefois tourner le dos à ce qui a fait son succès, notamment la publication des contenus sur différentes interfaces (site Web, Intranet, interface Wise...) et l'utilisation des fonctions de Portail et de CMS pour enrichir l'expérience utilisateur.

MAARCH

Maarch (www.maarch.org), est une solution PHP développée par la société française éponyme. La solution regroupe l'ensemble des fonctions sous une unique distribution open source : Maarch Enterprise.

Les fonctions de **Letter Box** sont maintenant intégrées à Maarch. Ceci permet notamment de gérer les interfaces avec un scanner ou de gérer le cycle de vie du courrier dans l'entreprise : réception, traitement de la réponse et recherche.

Enregistrer un document

Fichier chargé

>> Informations générales

Achats	Facture fournisseur
Catégorie	
Classement	
Objet	
Fournisseur	THOMSON-CSF (139)
Référence	F2007-002
Date du document	04-03-2013
Devise	EURO €
Montant HT	1300 €1.300,00
Montant TVA	130 €130,00
Montant TTC	1430 €1.430,00
Service traitant	Choisissez un service
Liste de diffusion	Choisissez une liste de diffusion
Activer la date limite	<input checked="" type="radio"/> Oui <input type="radio"/> Non
Date limite de traitement	05-04-2013

<< Index du type de document

<< Informations complémentaires

Actions : Enregistrer les modifications Annuler

Invoice - Facture
Invoice No : F2007-002
Invoice Date : 07/01/07

Sold To:
THOMSON-CSF
Purchasing Department
Michel BECK
1, Ave. Aristide Briand
94117 ARCEUIL CEDEX
France

Ship to (if different address):

SALESPERSON	YOUR P.O. #	DATE SHIPPED	SHIPPED VIA	F.O.B. POINT	TERMS
JLE	PO2007-002	07/01/07			

QUANTITY	DESCRIPTION	UNIT PRICE	AMOUNT
2	Maarch Professional Services days	650,00	1300,00
			0,00
			0,00
			0,00
			0,00
			TAX
			0,00
			SHIPPING & HANDLING
			0,00

**open
source**

Maarch Entreprise est structuré comme un SAE (Système d'Archivage Electronique) et est dédié à la conception d'applications de gestion de flux documentaire. Cette distribution inclut les fonctions nécessaires à une GED d'entreprise.

Solution de plus en plus mature et intéressante, notamment pour ces capacités dans les contextes dans lesquels elle excelle, en particulier : gestion de courrier, de dossier, document finaux, ou gestion d'archives (SAE – système d'archivage électronique).

Maarch propose un mode multi instance « *SaaS Factory* », installation unique permettant de gérer plusieurs environnements compartimentés. Le service « *SaaS Factory* » prend en compte la mise en place d'une infrastructure haute disponibilité et dispose d'un outil complet d'administration des divers environnements.

Notons enfin que la communauté autour de la solution est de petite taille, concentrée autour de son éditeur.

Parmi les nouveautés des deux dernières versions (dont la 1.4 qui sortira dans quelques semaines), les évolutions suivantes nous ont paru particulièrement intéressantes :

- un moteur de notifications d'événements, qui historise l'activité et permet des notifications par email ou par flux RSS
- un module de capture de documents qui permet d'ingérer des documents de différentes natures et provenances, avec des outils de traitement (OCR, parsing de mail, gestion de spool, LAD) et différents connecteurs de versement vers Maarch
- intégration et contribution à la norme SEDA, norme de gestion française de gestion des archives électroniques
- intégration bureautique permettant de piloter certains outils bureautiques sur le poste client pour générer des documents (publipostage)
- portabilité de l'interface sur supports nomades (Mac et Androïd).

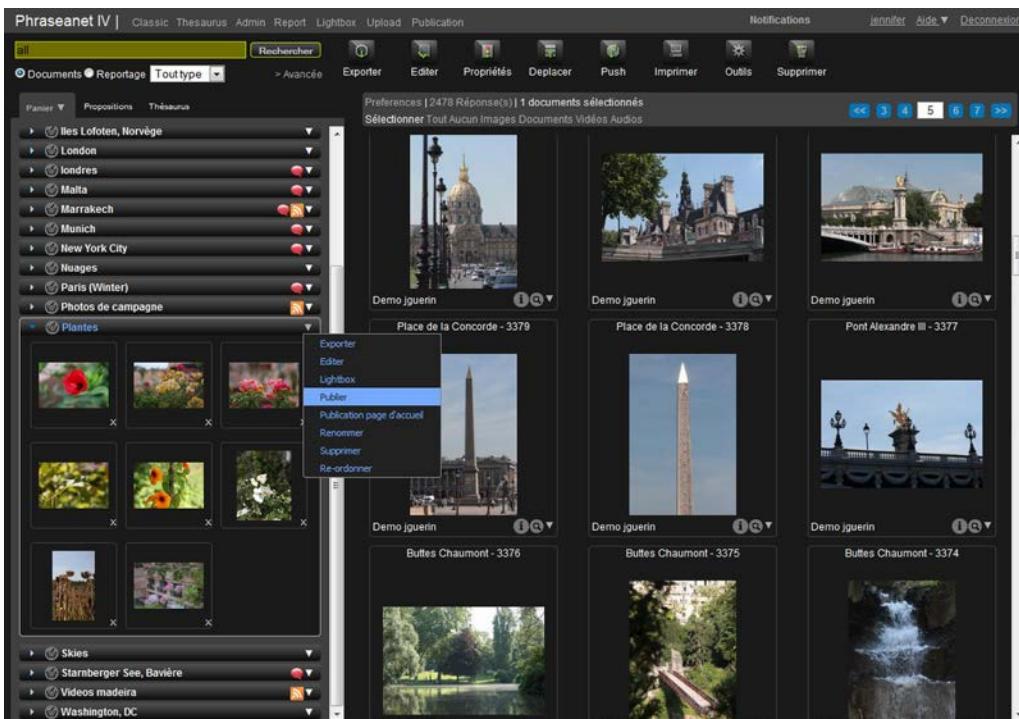
**SOLUTIONS SPECIFIQUES POUR LA
GESTION DE MEDIA**

PhraseaNet

Application de DAM (Digital Asset Management) orientée vers la gestion de média (image, vidéo), Phraseanet fait partie de ses (trop) rares solutions qui sont converties à l'open source (en 2010) après une expérience longue et fructueuse dans le monde propriétaire.

L'application apporte l'ensemble des fonctions essentielles d'un outil de gestion de médias, notamment l'upload intelligent, les paniers et reportage, la prévisualisation, les alertes (push d'info) etc. L'ensemble agencé de façon très cohérente devant une interface d'administration et de paramétrage complète.

Et dispose également de composants avancés, tels que la gestion de thésaurus, une API de développement d'interface, un reporting et une gestion des workflows (tâches) adaptés au flux de production dans le monde de l'image (agence de presse, musée, presse...)



**open
source**

RessourceSpaces

Dans la catégorie des solutions de DAM, citons également RessourceSpace (<http://www.resourcespace.org>), qui bien que plus modeste, permet également d'adresser le besoin spécifique de gestion d'images et de vidéos.

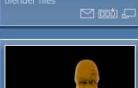
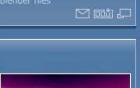
ResourceSpace

Admin User | Log Out | Contact Us

You found: 34 resources

Display: Large | Small | List Sort Order: Relevance | Popularity | Date | Colour

< Previous | Page 1 of 1 | Next >

				
Atmosphere View Edit	monkey_cornelius_passes View Edit Download Preview	monkey_cornelius_bake View Edit Download Preview	compo_map_uv_cubes View Edit Download Preview	ram_gsl View Edit Download Preview
				
blender files View Edit	blender files View Edit Download Preview	blender files View Edit Download Preview	blender files View Edit Download Preview	blender files View Edit Download Preview
				
My Collection Current Collection: My Collection 3 New Collection	monkey_com... View Edit	Atmosphere View Edit	blender files View Edit	blender files View Edit

Simple Search
Search using descriptions, keywords and resource numbers

Photo
 Document
 Video
 Audio

[Clear](#) [Search](#)

By Date
Any year Any month

Results Display
48 per page
 Large Small
 List

[> Go to Advanced Search](#)

AUTRES SOLUTIONS

Plusieurs autres solutions de GED existent dans le monde open source. Leur niveau de qualité est variable et ces outils sont comparativement limités, par rapport aux solutions phares du marché. Cette limitation peut être soit entendue, soit par rapport à leur adaptabilité au monde de l'entreprise soit par rapport à leur périmètre fonctionnel qui ne couvre pas l'ensemble de celui de la GED.

Ces solutions pourraient néanmoins correspondre à des besoins bien spécifiques, dans des contextes particuliers, notamment pour les besoins souvent simples de PME.

Ces outils ne sont pas systématiquement intégrés dans les comparatifs présents dans chaque chapitre fonctionnel de ce livre blanc.

On peut également citer : DocMgr, OpenGED, myDMS, eDMS et bien d'autres dont la maturité, la pérennité ou le niveau fonctionnel nous semblent aujourd'hui trop faibles et que nous ne préconisons pas.

open source

SilverPeas

SilverPeas est avant tout une solution de portail J2E, avec une forte connotation vers la collaboration. SilverPeas est déjà ancien, puisque développé entre 2001 et 2003. C'est un autre exemple de passage en open source puisque la solution a connu une première vie dans le monde propriétaire avant de devenir open source.



SilverPeas est à la croisée des univers du portail, du CMS et des outils de travail collaboratifs. C'est un portail, car il utilise la norme JSR-168 et est ainsi capable d'agréger des ressources hétérogènes. Il permet ainsi de proposer à l'utilisateur une vue synthétique de l'ensemble de ses ressources. De plus, SilverPeas fournit à chaque utilisateur un espace personnel extensible et personnalisable.

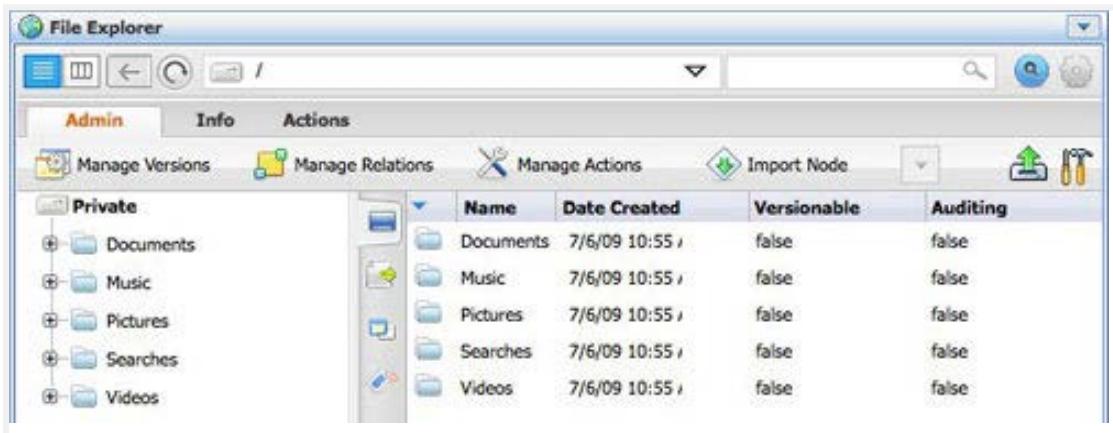
SilverPeas est également un CMS permettant de construire simplement des minisites avec une interface WYSIWYG simple. Le CMS intégré à SilverPeas est efficace dans le cadre de petites sections de publications à la disposition des utilisateurs. Cet outil permet également de créer rapidement en quelques clics des flashes infos, des sondages, des enquêtes en ligne. Cette fonction est couplée à un outil de GED proposant des classements paramétrables par l'utilisateur.

Enfin, SilverPeas aligne un ensemble d'outils collaboratifs répondant à l'essentiel du besoin. On y trouve un agenda partagé, un forum, un blog, un annuaire commun ou encore outil de gestion de projet avec tâches et visuel Gantt intégré.

**open
source**

Exo DMS

Exo est un éditeur français fondé en 2003 par le français Benjamin Mestralet. La société est aujourd’hui présente en France, aux Etats-Unis, au Vietnam, en Ukraine et en Tunisie.



L’application initiale, un portail d’intégration, a rapidement évolué vers la suite eXo Platform, étoffée de nombreux modules et briques techniques, couvrant de façon intégrée l’ensemble des besoins d’un ECM moderne. Citons la gestion d’un portail d’intégration (eXo Portal), la gestion de contenu web (eXo WCM), la gestion documentaire (eXo DMS – pour Document Management System), la gestion des workflows avec notamment l’intégration du projet Bonita et même un WebOS, sorte de portail en forme de poste de travail dédié à fournir une solution de bureau virtuel.

La partie DMS (c'est-à-dire la GED) ne constitue pas un axe majeur de la suite, même si cette brique sert au stockage de tous les fichiers. Raison pour laquelle nous avons choisi de ne pas la faire figurer systématiquement dans les comparatifs de solution.

Nous ne parlerons, dans ce livre blanc, que de cette brique, en prenant en compte son appartenance à un ensemble plus vaste.

L’application eXo DMS dispose de caractéristiques différenciantes importantes, parmi lesquelles nous retenons notamment :

- une interface standard qui intègre nativement les facilités de l’explorateur Windows : cliquer-glisser, raccourcis clavier, affichage selon plusieurs modes...
- la base de contenu (JCR) globale pour l’ensemble des contenus (Web et document) peut être utilisée via Webdav, FTP et CIFS
- l’intégration bureautique relativement poussée, avec un plug-in spécifique pour MS Office et une intégration Open Office
- la possibilité d’adapter, de façon très importante, les interfaces utilisateurs en fonction de leurs profils. En effet, eXo DMS, comme toutes les applications eXo, se base sur eXo Portal, application de gestion de portail

**open
source**

- il s'agit d'une application orientée technique, et non fonctionnelle. L'utilisation d'un package fourni par l'éditeur ou une intégration soignée est nécessaire pour en apprécier la qualité et en tirer pleinement parti.

Quotero

Jeune solution en JAVA (la v1.0 date de mars 2009) éditée et portée par la société de service, également éditeur : Core-techs.

L'offre est centrée sur quatre composants : le serveur documentaire, l'interface web d'interrogation, l'intégration à MS Office™ et le client lourd. Certains composants sont liés à une licence commerciale, la version open source étant limitée aux fonctions de bases.



Parmi les fonctionnalités présentes, nous citerons : indexation et recherche plein texte, réservation (check-in / check-out), alertes email et RSS, workflows documentaires avec jBPM, liens entre documents, paramétrage graphique des métadonnées.

On trouve de plus dans la version commerciale des fonctions intéressantes :

- ouverture et modification de documents depuis MS Office et OpenOffice
- drag & drop de documents depuis Windows et Linux via une application cliente.

La communauté autour de la solution semble principalement portée par l'éditeur et n'est plus active. Le caractère open source est, dans ce cadre, d'un intérêt plus limité.

La solution Quotero est permet de répondre à des besoins modestes (PME, par exemple) pour lesquelles une solution plus riche ne se justifie pas, mais qui souhaite néanmoins disposer des fonctions essentielles d'une GED pour stocker et partager ses documents.

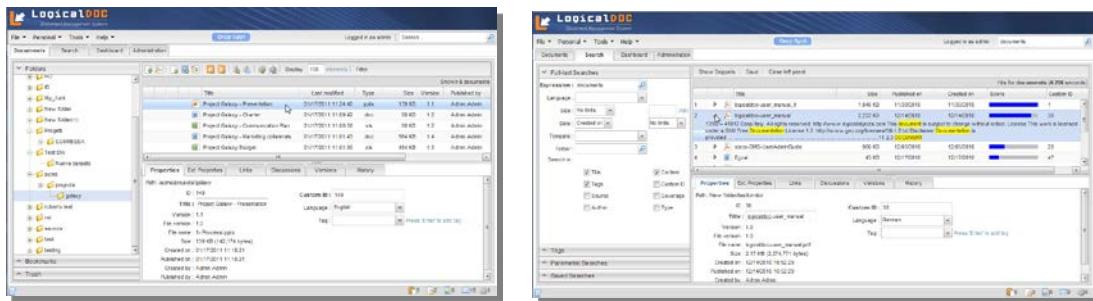
LogicalDOC

LogicalDOC (www.logicaldoc.com), autrefois Contineo, est un outil de gestion documentaire, édité par la société italienne Logical Objects. La solution est aujourd'hui en version 6.6.

La solution est distribuée d'une part dans une distribution open source ainsi que sous une édition commerciale, non libre, qui ajoute des fonctions et permet de bénéficier du support de l'éditeur.

**open
source**

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.



LogicalDoc propose les fonctionnalités essentielles d'un outil de GED ainsi que quelques fonctions intéressantes comme une gestion des droits graphiques, via des cases à cocher sur une grille.

Mais l'usage de la version open source est limité :

- pas d'intégration bureautique, la solution dans son ensemble n'est pas axée vers les fonctions collaboratives
- les fonctions d'import/export sont absentes
- l'authentification sur un annuaire d'entreprise n'est disponible que dans la version entreprise
- la communauté autour de la version open source est quasi inexistante
- l'architecture documentaire est très simple, ce qui limite grandement son extensibilité.

Des modules complètent la version open source, avec quelques fonctions intéressantes, parfois indispensables :

- la reconnaissance optique de caractères (OCR) et intégration aux outils de numérisation
- l'exploitation des formats AutoCAD et MS Office récents
- un workflow documentaire, paramétrable via une interface graphique
- des fonctions de traçabilité (audit)
- la possibilité d'intégrer des mails comme des documents, directement sur le serveur.

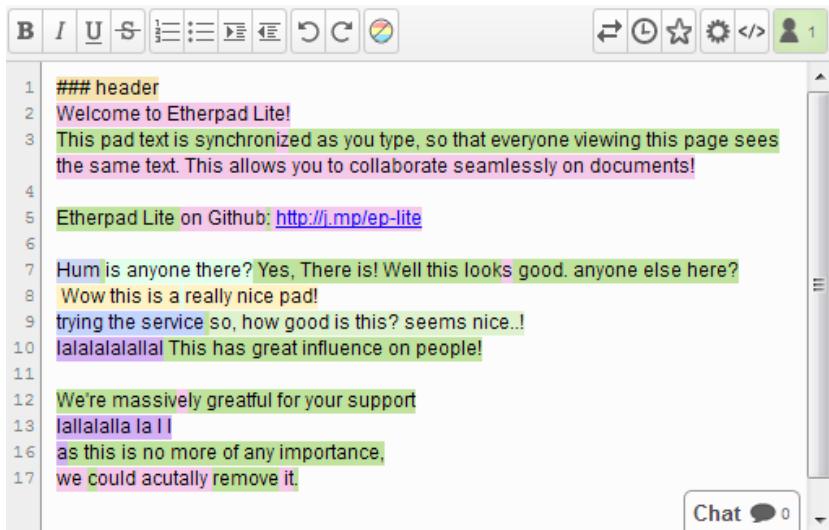
L'usage de LogicalDoc est donc réservé à des besoins simples et très orientés vers les quelques points forts de la solution. La solution ne bénéficie pas suffisamment de la dimension open source de sa distribution (communauté, extension, etc.).

Etherpad

Enfin, Etherpad (<http://etherpad.org>), qui n'est PAS une solution de GED, mais qui nous semble intéressante de présenter dans ce livre blanc tant les demandes d'édition collaborative (en temps réel) sont aujourd'hui fréquentes.

**open
source**

Cette solution open source a en effet pour unique vocation à venir concurrencer Google sur sa chasse gardée de l'élaboration collaborative. Etherpad est intégrable à la plupart des outils de GED de ce livre blanc pour y ajouter cette fonction d'élaboration collaborative en temps réel.



LES FONCTIONNALITES

Nous présentons dans ce chapitre les principales fonctionnalités des solutions de gestion documentaire, et, pour chacune, nous indiquons comment les outils cités précédemment 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 open source sont aujourd'hui très abouties et tendent à intégrer d'autres fonctions que celles historiquement réservées au domaine de la GED. De même, certaines applications d'autres domaines intègrent des fonctions de 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, la gestion de contenu structurée et le workflow.

Puisque les solutions assurent les fonctions essentielles, c'est donc sur leur optimisation et leur sophistication, ainsi que sur la mise à disposition de fonctions de hauts niveaux, que va porter l'essentiel de notre attention, en vue du choix d'une solution adaptée à chaque contexte.

METADONNEES

L'indexation est une fonction centrale des outils de GED, qui consiste en premier lieu à attacher des métadonnées aux documents et/ou dossiers.

Types de document

Chaque type de document peut être qualifié par un ensemble de métadonnées. Chaque document sera donc doté de métadonnées propres et sera éventuellement associé à des règles de gestion s'appuyant sur ces métadonnées.

Notons qu'il peut être intéressant d'indexer plusieurs fichiers avec un seul jeu de métadonnées. La relation entre notice documentaire et fichiers stockés doit donc idéalement être souple pour permettre d'avoir de zéro à « n » fichiers attachés à une seule fiche.

Structure des notices

On appelle notice l'ensemble de métadonnées associées à un document, le contenu de cette notice est le plus souvent saisi au moyen d'un formulaire.



open
source

Il est fondamental que le degré de finesse dans la qualification des documents puisse être ajusté à chaque projet. Il faut trouver le bon compromis entre la richesse d'informations, l'exploitation qui doit pouvoir en être faite, et la quantité de travail ou de développement que représente l'alimentation des métadonnées.

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.

La structure des notices des types de document doit :

- permettre une indexation pertinente, qui reflète précisément les documents décrits tout en incitant les contributeurs à alimenter
- couvrir l'ensemble des informations utiles à l'exploitation. La recherche bien entendu, mais également les processus et traitements appliqués aux documents tels que les alertes, la traçabilité, l'affichage...

Attention à **ne pas tomber dans la sur indexation** (trop de métadonnées pour un type de document) ou **l'indexation libre** (nombre insuffisant de contrôles) qui entraînent soit des coûts d'indexation trop importants par rapport à la valeur ajoutée, soit des défauts dans l'exploitation qui peut en être faite (bruit et silence en recherche, notamment).

Fonctions relatives aux typages de l'information

Il s'agit des 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
- règles de cohérence sur un champ et entre champs. Par exemple un format de date pour un jour, ou un nombre positif pour un prix...
- champs calculés permettant de faire dépendre la valeur d'un champ d'une ou plusieurs autres valeurs ou conditions
- liaisons entre documents et « typage » de ces associations, par exemple, un courrier qui « a pour annexe » ou un contrat qui « concerne » un dossier désigné.

La définition et la structuration des types de document constituent une phase fondamentale de la mise en place d'une GED. Elle ne doit en aucun cas être négligée.

Indexation manuelle

Interface de saisie

La notice (ou fiche) décrivant un document est souvent le premier point d'entrée des utilisateurs. C'est également ce formulaire qui est le plus utilisé par les contributeurs, puisqu'ils devront saisir de l'information.



**open
source**

Pour ce faire, et dans un souci d'ergonomie, de fiabilité et de productivité, les solutions de GED doivent faciliter la contribution, 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, par exemple :
 - des listes de contrôle qui permettent de fiabiliser le fond (sémantique) et la forme (orthographe notamment) de la saisie
 - des cases à cocher ou des boutons radio pour les choix multiples
 - des boîtes de dialogue adaptées aux tables des valeurs de référence (listes des choix possibles). Ceci pouvant aller de la simple liste déroulante, à des interfaces disposant d'une recherche avec autocomplétion ou d'une navigation (abécédaire, arbre) dans les valeurs de référence
 - suggestions pour un champ donné (cf « Induction de métadonnées »).

Table de référence

L'objectif est de proposer, dès que c'est possible, des listes de valeurs. Ceci pour limiter les questions que l'utilisateur peut se poser lors de la contribution, mais également pour fiabiliser la saisie et ainsi améliorer l'usage des métadonnées.

Parmi les tables de références, on trouve par exemple :

- les référentiels de classement (voir le chapitre Référentiel de classement, page 53)
- les listes de valeurs enrichies par la saisie ou figées, par exemple le nom des auteurs des documents (déjà présent dans le système) ou la liste des services d'une organisation.

L'objectif de ces tables est d'offrir des aides à la saisie et d'apporter des contraintes et contrôles lors de celles-ci.

Indexation automatique et extraction de métadonnées

Il existe plusieurs modes d'extraction de métadonnées.

- Récupération d'informations

Chaque fichier informatique est doté d'informations issues du système de gestion de fichiers. Certaines peuvent être pertinentes comme des 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 d'une 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, mais également des formats MS Office, par le biais de leurs propriétés.

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, les plus élaborées, permettent de déterminer automatiquement les métadonnées les plus pertinentes pour indexer un document.

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

- des outils de reconnaissance, qui retrouvent dans un document les chaînes de caractères présentes dans un référentiel pouvant être ajoutées à la notice
- des outils statistiques, qui analysent les chaînes de caractères 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 pertinence d'indexation par rapport à l'indexation manuelle, de prendre en compte très rapidement, de très grandes quantités d'informations.

Alfresco gère les types de documents par une combinaison d'aspects. 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, son emplacement ou ses propriétés. Notons qu'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.

eXo DMS propose une gestion avancée des métadonnées. Plusieurs types d'objets peuvent être gérés, dont certains peuvent être complexes. Cette gestion peut notamment exploiter des taxonomies. La structuration des métadonnées peut être gérée directement dans l'interface d'eXo DMS.

Jahia permet également d'avoir des types de contenus. La nature des métadonnées des contenus Web et celles relatives aux documents sont sensiblement différentes. La saisie des informations peut être partiellement automatisée, que ce soit par l'extraction pour certains types de document pour l'usage de règles (avec Drools).

Knowledge Tree permet à l'administrateur de créer des types de documents et des ensembles de métadonnées depuis l'interface web d'administration. La saisie des métadonnées se fait manuellement, mais il est possible de paramétriser des filtres pour automatiser certaines extractions.

LogicalDoc utilise des modèles pour différencier les documents. Ces modèles sont définis via l'interface graphique. La saisie des métadonnées se fait manuellement.

Maarch permet de créer facilement différents types de documents. Ces types de documents s'inscrivent dans des collections documentaires dont la logique est proche de celle de l'archivage physique pour faciliter la gestion et les traitements qui peuvent y être rattachés.

Nuxeo utilise la notion de facette pour ajouter de nouveaux types de documents. Les nouveaux types de documents sont créés à partir de schémas XML (XSD) ou de Nuxeo Studio, de même que les écrans de saisie ou de restitution. Ils sont ajoutés à l'architecture de la solution en ajoutant de nouveaux plug-ins (extension). Nuxeo Studio, service associé au support de l'éditeur, permet de disposer d'une interface graphique pour ces paramétrages.

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 vite complexe de connaître, avec certitude, la dernière version d'un document ou la version applicable.

La gestion des versions permet donc à la fois de tracer les évolutions d'un document, et, par un système de réservation (check-in/check-out), de garantir qu'un utilisateur ne puisse prendre la main sur un document et le modifier s'il est déjà en cours d'édition.

Réservation (check-in/check-out)

La réservation peut être implémentée 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 également envisager qu'un check-out soit libéré automatiquement au-delà d'un certain délai.



open
source

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.

Incrémantion des versions

En général, l'incrémentation des versions est un compteur automatique qui définit un numéro séquentiel à chaque version successive 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 préétablies, qui définissent la construction des numéros de version.

Dans tous les cas, l'historique des versions est conservé et il doit être possible de visualiser une version antérieure d'un document. 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 versionning, par exemple lors de la création d'une version majeure, ou selon le temps écoulé depuis la dernière version majeure ou encore selon l'identification des évolutions entre deux versions.

Alfresco, Nuxeo, eXo DMS et Knowledge Tree proposent des fonctions avancées de réservation et de gestion de versions. Ils proposent aussi : la distinction entre versions mineures et majeures, les commentaires d'une version donnée, voire la réservation automatique lors de l'édition en ligne d'un document. Les systèmes de versionning peuvent être surchargés pour prendre en compte des contraintes métier particulières comme dans la santé ou le bâtiment.

LogicalDoc propose également la réservation et la gestion de versions, mais les possibilités, notamment de paramétrage, sont en retrait par rapport aux autres solutions.

Jahia propose ces fonctionnalités, mais essentiellement pour les contenus web. Il existe dans la version 6.5 des possibilités de versionning des fichiers, toutefois relativement en deçà de ce que propose un outil de GED.

Maarch ne propose ce type de fonctionnalités que sur les courriers sortants. Cette fonction est associée à la production du document sur la base de modèle bureautique.

REFERENTIEL 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 termes de complexité des référentiels (hiérarchie, liens typés...) et 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 de ses besoins (voir « Ontologie »).

Lexique

Il s'agit d'une liste des termes utiles dans un contexte donné et organisé le plus souvent par ordre 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 côtes 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 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 lequel 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 termes les plus courants, on retiendra :

- terme « générique », qui correspond au « père » dans la hiérarchie. Un terme peut avoir plusieurs « pères »
- terme « spécifique », qui correspond à un « fils » dans la hiérarchie
- terme « employé pour », qui correspond aux 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

Aucune solution de GED de ce livre blanc, à l'exception de Phraseanet n'implémente de gestion de Thésaurus.

Notons qu'une solution de gestion de multi-Thésaurus est en cours de réalisation, initiée par le Ministère de la Culture et réalisée par Smile.

Référentiel de conservation

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

Ce référentiel est généralement construit sur la base d'un plan de classement, qui 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, règlement... afin de permettre la révision.

Tout l'intérêt de cet outil est de permettre l'association 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.

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

Ontologie

Les ontologies fonctionnent sur le même principe que les thésaurus, 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 souvent 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 thésaurus, 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 aujourd'hui encore assez limitée, et essentiellement basée sur les RDF (Ressource Description Framework), modèle de représentation d'ontologie, normalisée par le W3C.

Le concept d'ontologie 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 paramètres qui indiquent quels liens de l'ontologique sont utilisés.

Notons que des poussées intéressantes dans le monde du Web et de la GED, notamment avec Nuxeo font jour ces derniers mois dans le domaine de la sémantique, dont les ontologies sont l'un des supports principaux. On notera par exemple l'usage d'ontologies publiques comme DBpedia pour enrichir les requêtes (une initiative de Wikipedia <http://en.wikipedia.org/wiki/DBpedia>).

Utilisation des référentiels

Etudions quelques utilisations envisageables.

Les objectifs principaux des référentiels de classement visent à faciliter l'usage de la base documentaire, notamment en réduisant le bruit et le silence lors de recherche ou d'autres types d'exploitations.

Restitution des contenus

La navigation par arborescence permet à l'utilisateur de retrouver un document selon une logique hiérarchique, voire plusieurs logiques hiérarchiques différentes.

Nous parlerons alors de « facettes ». Pour illustrer, c'est comme si on proposait à l'utilisateur les mêmes objets (le fonds documentaire) selon plusieurs points de vue (les facettes).

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 correspondante à la logique (ou « facette ») géographique, ou bien à la logique (ou « facette ») gastronomique.

Cette approche est très puissante, puisqu'elle propose à chaque utilisateur une vision et une exploitation de la base documentaire qui correspond à son usage.

Gestion multi référentiels

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

La gestion de multi référentiels impose de gérer certains cas complexes tels que les homonymies, les liens entre référentiels, etc. 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 ce point est important dans votre contexte projet.

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 :

- l'enrichissement facile de la base de synonymes, éventuellement par les utilisateurs eux-mêmes
- l'utilisation automatique des synonymes lors d'une recherche sur le terme principal (voir Outil de recherche)
- la détection automatique et le remplacement des synonymes par le terme retenu lors de l'indexation, afin d'homogénéiser l'indexation.

Outil de recherche

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

- Expansion de requêtes

Une exploitation possible de référentiels riches est de permettre la définition de l'environnement sémantique d'un terme. Certains moteurs de recherche sont ainsi 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 de « automobile », de « bagnole »... Ainsi, tous les documents qui contiennent ces termes seront effectivement dans les résultats de recherche (réduction du silence).

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

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 préétablie de valeurs 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 structurée par thème ou par domaine. 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.

La recherche est l'un des domaines où l'énergie est la plus palpable, à la fois en terme d'innovation que d'attentes des utilisateurs. Les solutions, y compris open source, ne s'y trompent pas et propose de plus en plus de fonctions avancées sur cet aspect.

Alfresco propose un plan de classement sous la forme d'une hiérarchie de catégories. Il propose également d'afficher le contenu du référentiel de classement selon l'arborescence des catégories. L'intégration de SolR permet d'affiner les résultats selon des facettes administrables.

eXo DMS recommande d'utiliser sa gestion de taxonomie pour catégoriser les documents. Le classement des documents est par ailleurs très similaire à celui d'un système de fichier.

Jahia permet de disposer d'un plan de classement et des tags. Ce dernier est utilisable pour indexer les documents lorsqu'ils sont associés à des contenus. Une notion d'auto tagging permet d'automatiser facilement la création de tags en fonction des projets sur Jahia Wise.

Knowledge Tree permet de créer aussi bien des lexiques que des plans de classement. La navigation par facette est générée automatiquement pour certains types de contenus.

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

Maarch propose un plan de classement lié à la typologie des dossiers. Un module d'AUTOFOLDING permet de constituer à la demande des arborescences basées sur des métadonnées (navigation virtuelle).

Nuxeo propose un système de gestion de référentiel qui permet de structurer l'ensemble des référentiels. Ils sont notamment utilisés pour l'indexation des documents, le typage des liens entre documents ou la navigation par facette.

Quotero propose un plan de classement navigable. Il est possible d'effectuer des recherches au sein de ce plan de classement.

MOTEUR DE RECHERCHE

Il s'agit bien entendu d'une brique essentielle des applications de gestion d'informations.

Trois tendances non exclusives existent dans les solutions de GED :

- l'intégration d'un moteur de recherche, les solutions open source intègrent souvent Lucene, moteur de référence du monde open source.
- l'exploitation de la recherche sur de la base de données, ceci peut limiter les fonctions disponibles (notamment lemmatisation, troncatures...)
- l'ouverture vers des solutions de recherche du marché (propriétaire) via des connecteurs.

Il est donc nécessaire de savoir **comment** on souhaite rechercher, notamment en termes de fonctions et de périmètre de recherche.

Fonctions de base

A minima, le moteur de recherche d'une GED doit :

- indexer les fiches documentaires, c'est-à-dire toutes les métadonnées d'un document
- indexer les documents électroniques en texte intégral (ou « full text »), c'est-à-dire en prenant en compte le contenu des fichiers électroniques, pour tous ceux qui incluent du contenu textuel
- 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 consulter
- permettre des recherches aussi bien sur l'intégralité d'un document que sur un ou plusieurs champs spécifiques de métadonnées.
- permettre l'utilisation des tables de référence dans les interfaces de recherche
- 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 le texte intégral
- autoriser les troncatures centrales et à droite, implicitement ou par l'utilisation de caractères génériques qui remplacent un ou plusieurs caractères (classiquement l'étoile « * » ou le « ? »)
- permettre de personnaliser les pages de recherche, si possible plusieurs pages différentes, pour correspondre aux attentes des utilisateurs

Fonctions avancées

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

- permettre l'enregistrement de résultats sous forme de « panier » par exemple, pour permettre aux utilisateurs de constituer une sélection
- permettre l'enregistrement des stratégies de recherche, c'est-à-dire les critères utilisés pour la recherche, afin de pouvoir les relancer simplement
- rechercher par étapes successives de façon à affiner une requête par rapport aux résultats de la recherche précédente
- étendre une recherche, par rapport à des synonymes par exemple
- fédérer la recherche sur plusieurs bases documentaires
- automatiser les approximations, suggestion orthographique et la lemmatisation (c'est-à-dire la recherche sur les racines des mots)
- permettre la recherche en langage naturel ou recherche par l'exemple, c'est-à-dire déduire l'équation de recherche d'une phrase ou d'un texte interprété par le moteur
- suggérer des résultats approchant à l'issue d'une recherche, par différents moyens :
 - explicites, via des liens entre documents spécifiés dans l'indexation de chacun
 - implicites, des rapprochements issus de calculs statistiques ou sémantiques
- détecter automatiquement des « entités nommées », c'est-à-dire des noms propres (personnes, lieux), et leurs relations.

La fonction de recherche est centrale dans les projets de GED. Il s'agit souvent de la fonctionnalité la plus utilisée et la plus importante 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 aficionados de Google), la recherche par référence, la recherche fédérée couvrant à la fois la GED et 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 utilise le moteur Lucene ou SolR pour indexer ses contenus, documents et métadonnées. Cette intégration permet de disposer nativement d'une solution puissante et largement configurable. L'écran de recherche avancée de l'interface standard est paramétrable. OpenSearch est également disponible.

eXo DMS propose une fonction de recherche complète sur le contenu des fichiers et des métadonnées. Des paramétrages avancés permettent de définir des synonymes ainsi qu'un premier niveau de configuration de Lucene.

Jahia permet également de rechercher sur les documents et les métadonnées par l'intégration de SolR. Les fonctions du moteur sont donc puissantes et largement configurables. OpenSearch est également disponible.

KnowledgeTree intègre un moteur de recherche simple et puissant, qui propose de nombreuses options pratiques (historique de recherche, recherche par type de contenu, par localisation...). Mais la caractéristique la plus intéressante se trouve dans l'interface de recherche avancée qui permet de bâtir très simplement des recherches extrêmement sophistiquées croisant potentiellement l'ensemble des critères de la base documentaire.

LogicalDoc ne prend en compte que les types de documents les plus simples dans la recherche de sa version communautaire. Les fonctions de recherche présentent néanmoins deux fonctions intéressantes : la configuration de l'écran de recherche avancée par l'utilisateur et la recherche de documents par similarité (basée sur la fréquence des mots).

Maarch intègre un module de recherche basé sur Lucene pour la recherche full text couplé à une recherche sur la base de données pour la recherche documentaire.

Nuxeo utilise par défaut le moteur de la base de données. Il est également possible de d'utiliser d'autres moteurs. Les fonctions de l'interface standard sont de bon niveau, mais limitées aux fonctions de recherche de la base de données sous-jacente.

Quotero propose un configateur graphique de type de documents avec métadonnées personnalisées.

Les fonctions de recherche plus avancées ne seront pas disponibles nativement et demanderont, la plupart du temps, une intégration ou une configuration.

Parmi les moteurs non open source qui apportent une valeur ajoutée non négligeable, notamment sur le plan sémantique, citons des outils tels que Sinequa, Antidot, Exalead, Polyspot, Autonomy ou Fast.

NUMERISATION

Les applications de gestion documentaire ne prennent, en général, pas en charge la numérisation, ni les traitements associés – ceux-ci sont généralement couverts par des solutions dédiées. Elles traitent les phases initiales depuis la conversion du support, jusqu'à « l'injection » du document dans la GED. L'application documentaire reprend la main sur les traitements métiers, notamment d'indexation, une fois le fichier « injecté ».

Les fonctions évoquées ci-dessous 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 pré indexation

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 au maximum le processus qui comprend la numérisation et la saisie des premières métadonnées du document (on parle de pré indexation), on s'efforce d'optimiser l'interface proposée à l'opérateur pour numériser et pré indexer les documents dans la GED.

Encore souvent perçue comme fastidieuse par les opérateurs, cette première phase est cruciale pour garantir la prise en main et la réussite de l'application de dématérialisation.

En effet, c'est là que sont saisies les métadonnées explicites (date de numérisation, expéditeur, destinataire, montant d'une facture...). Leur exactitude et leur complétude vont déterminer une part importante de l'efficacité des traitements ultérieurs et donc, de la valeur ajoutée du système.

Des procédés de gestion par lots et l'introduction de la 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, les erreurs et les interventions humaines.

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.

Les acronymes du domaine de la reconnaissance automatique :



open
source

- **OCR** : Optical Character Recognition ou reconnaissance optique de caractères. Les « bons OCR » atteignent des taux de reconnaissance du document numérisé de plus de 95% et savent remettre en forme des tableaux, reconnaître les styles, etc.
- **ICR** : Intelligent Character Recognition ou reconnaissance intelligente de caractères. Dédiées à la reconnaissance de l’écriture cursive (manuscrite), ces applications sont notamment très utilisées 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.
- **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.

Basés sur des algorithmes complexes, ces outils sont largement dépendants de la qualité de la reconnaissance et de l’utilisation de dictionnaires linguistiques et métiers. Les meilleurs utilisent des notions d’apprentissage et de suggestion de reconnaissance par association.

L’intégration de ces fonctions dans une GED consiste le plus souvent en l’adjonction d’une application dédiée et la mise en place d’un « injecteur » (ou connecteur) pour récupérer dans la GED le résultat des flux numérisés.

Notons qu’il n’existe pas de solution réellement satisfaisante de reconnaissance automatique distribuée sous licence open source.

Alfresco s’interface avec Kofax Ascent Capture, solution commerciale de pilotage de scanner et de reconnaissance automatique de contenus. Notons qu’un partenariat a été récemment conclu avec **Ephesoft** (www.ehesoft.com)- éditeur de la première solution open source totalement dédiée à la dématérialisation.

LogicalDoc propose ce type d’intégration dans sa version commerciale.

Knowledge Tree intègre, dans son édition commerciale, un outil spécifique qui permet à la fois le pilotage de scanner et l’OCR. Cet outil est une chaîne de numérisation simple qui permet de piloter le scanner, de réaliser des manipulations simples du fichier généré et d’indexer ce fichier dans la GED.

Maarch propose une interface standard avec Kofax Ascent Capture avec les scanners **Fujitsu** et **Kodak** ainsi que nombreux de scanners de productions. Des chaînes de numérisation ergonomiques et adaptées à la gestion de courrier peuvent être mises en place simplement avec Maarch.

Nuxeo dispose de points d'extension pour une intégration avec des solutions de numérisation, notamment dans le cadre de son axe vertical de gestion de courrier, basé sur Nuxeo Case Management Framework (Nuxeo CMF). A noter la sortie de la dernière version 1.7 en juin 2011.

Jahia ne propose pas nativement ce type d'intégration.

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 pierre angulaire.

La gestion des droits passe, le plus souvent, par l'association d'autorisations sur des éléments de la base documentaire. Ces autorisations permettent de définir un ensemble de droits pour un utilisateur ou un groupe d'utilisateurs.

Niveaux de gestion

La gestion des droits doit pouvoir être suffisamment fine, à minima par nœud d'arborescence (répertoire dans un système de fichiers), et si possible au niveau de chaque document. Certains systèmes vont jusqu'à offrir une gestion au niveau de chaque métadonnée, afin de gérer des problématiques de confidentialité extrêmement pointues. La gestion des droits prend également en compte les actions possibles en fonction du rôle (profil) de chaque utilisateur (voir le paragraphe ci-dessous).

La gestion des droits ou habilitations des utilisateurs doit pouvoir être réglée finement et de façon décentralisée. C'est-à-dire que ce sont les gestionnaires de la GED et non des techniciens qui sont en capacité de les administrer. Elle permet principalement de définir comment est implémentée la notion de confidentialité dans la GED.

Gestion des profils

Un utilisateur peut cumuler différents profils, en fonction de son appartenance à un service, à un niveau hiérarchique, mais également de son rôle par rapport à un processus transverse, par exemple contributeur sur le site Web, responsable d'un dossier client...

La gestion des autorisations 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.

Dans la mesure du possible, il est préférable de s'appuyer sur une métadéfinition des droits au niveau d'un annuaire d'entreprise. Ceci se traduit par l'utilisation des groupes de l'annuaire central dans le système de GED. De façon optimale, on ne gère dans la GED que l'affectation de rôles à des groupes sur des objets documentaires, ceci de façon totalement impersonnelle, la gestion des individus étant déléguée à l'annuaire d'entreprise.



open
source

Annuaire et SSO

Les applications documentaires peuvent reposer sur des annuaires LDAP et dispositifs de Single Sign-On (SSO) pour la gestion de la connexion de leurs utilisateurs.

Les différentes solutions présentées se distinguent notamment par rapport à leurs capacités de synchronisation avec l'annuaire, à leur capacité à s'interfacer rapidement avec des systèmes de SSO du marché ou à la possibilité de créer des utilisateurs pour la GED, en dehors de l'annuaire.

Toutes les solutions présentées ici permettent de s'interfacer avec un annuaire externe. Elles proposent toutes une gestion suffisamment fine des droits, au minimum par dossier.

Alfresco, eXo DMS, Nuxeo et Jahia supportent la configuration d'un système de SSO et l'authentification chaînée sur plusieurs annuaires.

Alfresco permet également de disposer d'une base d'utilisateurs secondaire dans le cadre de son interface Share, ceci afin de pouvoir inviter des utilisateurs externes.

eXo DMS permet de définir des droits sur une multitude de niveaux : les contenus, les éléments d'interfaces, certaines fonctions... Ces comportements sont notamment hérités du composant portail sur lequel est basé eXo DMS.

LogicalDoc supporte l'identification via un annuaire LDAP.

Maarch utilise un système de gestion avancée à la fois sur les modules applicatifs et par rapport aux contenus. Maarch est compatible avec une authentification LDAP et le SSO.

Knowledge Tree propose en plus du modèle standard, un mode de gestion dynamique des permissions, basées sur les métadonnées de documents.

Nuxeo permet également dans son interface standard la définition des droits en négatif, c'est-à-dire le retrait ponctuel de droits et une gestion dynamique en fonction des métadonnées.

FONCTIONS COLLABORATIVES

La GED centralise la documentation de l'entreprise et offre de facto un dépôt unique pour les documents, notamment bureautiques. Les fonctions collaboratives prennent donc tout leur sens, ceci bien au-delà de l'outil de base : l'email.

Il existe plusieurs concepts dans l'édition collaborative de documents c'est à dire, lorsque plusieurs utilisateurs souhaitent travailler sur un même document présent sur la GED :



**open
source**

- **Réservation explicite** : une copie de travail est créée afin que l'utilisateur puisse l'éditer sur son poste de travail. L'original reste en accès lecture seule sur la GED. Ce principe s'appelle « check-out » ou réservation. Une fois le document modifié, l'utilisateur met en ligne la nouvelle version et libère sa réservation. Le document est en général versionné (le numéro de version incrémenté) et remis en accès libre : il s'agit du « check-in » ou libération de réservation
- **Modification en ligne** : l'utilisateur qui souhaite modifier le document, réalise cette modification en ligne, c'est-à-dire sur la version du serveur. Tant que le document est ouvert sur son poste, le document est réservé, et donc en lecture seule. Dès que l'utilisateur ferme le document sur son poste de travail, la réservation est libérée. Ce procédé est le plus confortable pour l'utilisateur, mais il trouve ses limites, car il nécessite d'être perpétuellement « on-line »
- **Modification concurrente** : les utilisateurs éditent le document en temps réel depuis leur poste. Si le document a été ouvert par plusieurs utilisateurs en même temps, la GED gère l'édition concurrentielle en pointant les écarts et en demandant une intervention. Cette pratique est exceptionnelle, car les formats de documents bureautiques se prêtent très mal à ce type d'édition

On appréciera également d'autres fonctions telles que les alertes sur modification, l'envoi d'emails depuis la GED, les commentaires et discussions sur les documents ou encore la prévisualisation.

Les fonctions indispensables sur l'aspect collaboratif sont la réservation et la gestion de version.

Alfresco permet de réaliser les éditions collaboratives, via la réservation, via la modification en ligne ou via l'intégration à MS Office™. Alfresco propose, en standard, des outils d'alertes configurables par l'administrateur. Share fournit également des outils de gestion de contenu tels qu'un wiki, un forum, un blog et un calendrier partagé compatible iCal. Chaque utilisateur dispose d'un tableau de bord remontant toutes les informations utiles pour les sites auxquels il participe.

eXo DMS intègre peu de fonctions collaboratives directement utilisables. Un vertical dédié à la collaboration dans la suite eXo propose un ensemble d'outils, notamment calendrier, email, forum...

Knowledge Tree fournit des fonctionnalités de réservation et d'intégration à MS Office dans sa version commerciale ; notons que l'intégration à Office repose sur une intégration très ergonomique avec le poste de l'utilisateur. La solution permet également de créer des documents Word et Excel directement à partir de l'interface Web dans la GED.

Nuxeo permet de réaliser des éditions collaboratives via la modification en ligne via le module LiveEdit à déployer sur les postes clients. Nuxeo propose également un système de synchronisation poussés des métadonnées (de la GED) avec les (le contenu des) documents bureautiques. Enfin, l'interface Nuxeo Collaboration dévoile des fonctions intéressantes orientées vers le « social ».

Maarch n'est pas orienté vers ce type de fonctions.

WORKFLOW

Deux types de workflows (processus) sont mis en œuvre dans le cadre des solutions de gestion de contenu. Il s'agit de l'une des frontières entre le domaine de la GED et celui de l'ECM. En effet, dans les solutions de GED, on trouvera des workflows appliqués aux documents. Dans les solutions d'ECM, on trouvera également des procédures automatisées, en dehors de tout contexte documentaire : on parlera alors de BPM ou de workflow métier.

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, traitement de données, dématérialisation de formulaire...

Workflow documentaire

Il s'agit de mettre en œuvre un circuit de validation de documents, souvent pour les publier ou pour les archiver : on trouve parfois le terme de « Docflow ». Différentes méthodes sont implémentées par les outils sélectionnés.

Le workflow repose en général sur des fonctions standards de mises à disposition des outils : la gestion des droits, envoi d'emails, déplacement de fichiers, etc.

Les outils qui permettent de créer des workflows à la demande peuvent être basés sur une gestion de règles sous-jacente à l'application. Ils permettent d'ajouter, de juxtaposer différentes fonctions, à partir de règles simples, assemblées les unes avec les autres.

Il peut également s'agir de développements simples et souvent peu paramétrables, notamment en termes de nombre d'étapes, de notifications ou de résultats du workflow.

Workflow métier ou BPM

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.

Les applications documentaires, lorsqu'elles proposent une gestion de processus élaborée, intègrent une brique open source pour la « motoriser ».

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 **Intalio**, **Bonita**, **jBPM**, **Processmaker**, **Orchestra** ou **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...
- s'assurer que l'implémentation du moteur de BPM dans la GED est de bonne qualité, notamment en termes de gestion des utilisateurs (lorsqu'ils sont communs), de persistance des données ou de design des processus en regard de leur interaction avec les fonctions documentaires

Représentation de workflow

Il existe plusieurs méthodes de représentation de processus.

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 paramétrier graphiquement les alertes, les conditions et les interactions avec d'autres applications, voire des traitements spécifiques (scripts).

Implémentation dans la GED

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

L'intégration des interfaces proposées aux utilisateurs « masque » le plus souvent l'interface du workflow et donne à l'utilisateur l'impression d'utiliser une seule application.

La gestion des droits doit être transversale entre GED et BPM. Les utilisateurs ont des rôles au sein des processus. Ils peuvent être destinataires de certaines tâches en fonction des processus. Parallèlement, ils ont des droits sur la base documentaire et sont donc autorisés à réaliser des actions sur certains objets. 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 ne disposerait pas des autorisations nécessaires pour la réaliser, par exemple.

Tableau de bord et suivi

L'un des objectifs du BPM est de permettre le suivi et la traçabilité des processus.

Toutes les fonctions de « log » et leur exploitation doivent donc faire l'objet d'une attention particulière au moment de choisir l'implémentation de la solution de BPM dans la GED.

A minima, chaque utilisateur doit pouvoir visualiser ses tâches en cours. Pour aller plus loin, il est souhaitable de disposer des indicateurs suivants : la liste des tâches des subordonnés hiérarchiques, 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...

Ces informations sont, le plus souvent, présentées dans l'interface de la GED, sans qu'il soit besoin d'accéder à une interface tierce, sauf lorsque la solution de BPM est utilisée pour « motoriser » des processus qui n'ont pas de rapport avec la gestion documentaire. Dans ce cas, il peut être préférable de regrouper toutes les tâches dans un tableau de bord externe à la GED.

Alfresco dispose d'un outil très complet basé sur un moteur de workflow (Activity). Il est bien intégré à l'application et permet de réaliser des workflows métiers facilement. Par ailleurs, Alfresco dispose d'un outil de conception de workflow simple, basé sur l'usage de son moteur de règles intégré. Ce second type de workflow peut être généré par un utilisateur fonctionnel.

eXo DMS intègre Bonita comme moteur dédié à la modélisation et à l'automatisation des workflows documentaires. Bonita est également utilisé dans la suite eXo pour piloter différents processus. Il s'agit de l'outil de BPM le plus complet parmi ceux intégrés aux solutions présentées ici.

Jahia propose son propre mécanisme de workflow. Il est relativement souple, orienté vers la publication, que ce soit par page, celle-ci regroupant des contenus web et des fichiers, ou par type de contenu.

Knowledge Tree utilise le moteur ProcessMaker. Il est facilement paramétrable pour des processus documentaires et son intégration sous forme d'un module est pragmatique.

LogicalDoc propose, dans sa version entreprise, un système de workflow documentaire paramétrable graphiquement.

Maarch dispose d'un système de workflow faisant appel à des paramétrages complexes. Maarch dispose nativement de fonction de délégation et de gestion d'absences.

Nuxeo dispose d'un outil très complet basé sur le moteur de workflow. L'implémentation de base qui est proposée est largement paramétrable et permet de réaliser des workflows métiers de différents types. Ce module - Content Routing - permet de définir graphiquement via Nuxeo Studio, des flux de travail et des états de document correspondant.

REGLES DE GESTION

Fonctions de gestion

Gérer des documents implique de pouvoir disposer de fonctions permettant la prise en compte des règles de gestion inhérentes aux métiers concernés. Par rapport aux documents, cela se traduit notamment par des actions sur : l'acquisition, les notifications, la conservation, les transformations...



**open
source**

On trouvera, selon les applications, des fonctions diverses, par exemple :

- transformation en PDF de documents bureautiques
- calcul d'un numéro de chrono pour afficher un cartouche de document en fonction de métadonnées
- diffusion d'informations en fonction de différents critères par exemple : une stratégie de recherche préenregistrée, un mot-clé, la validation d'un document ou d'un ensemble de documents
- notification par email, par flux RSS, dans un tableau de bord...
- définition des formats d'affichage des fiches d'indexation, par exemple : par profil d'utilisateur, selon des optiques métiers ou par type de document. Ceci afin de mettre en valeur les champs les plus pertinents pour chacun
- affectation conditionnelle de métadonnées à un document, par exemple, par rapport à la personne qui l'indexe ou à son dossier d'appartenance
- manipulation d'images, dans une approche Médiathèque. Ceci permet d'éviter de sortir l'image de l'application pour des actions simples, par exemple : recadrage, rotation, balance de couleur
- échanges collaboratifs autour d'un document ou d'un dossier, par exemple par des fonctions de type forum, via les annotations sur un document

Moteur de règles

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

Ces règles constituent un enchaînement d'actions unitaires, par exemple : copie, transformation, alimentation d'une métadonnée, envoi d'emails, etc. Elles peuvent également être plus complexes et refléter des besoins métiers.

De plus, il est souvent plus facile d'implémenter de nouvelles fonctions à travers un moteur de règles (ex. : Drools). Ceux-ci sont en effet bâti sur des architectures extensibles et permettent d'ajouter facilement des cas de fonctionnement qui ne sont pas prévus nativement.

Les applications de GED se distinguent ici de façon importante selon qu'elles soient écrites en PHP ou en JAVA. Les premières tablent sur leur plus grande rapidité de développement, les secondes sur une architecture technique extensible facilitant l'ajout de nouvelles fonctions.

Alfresco dispose de nombreuses fonctions de gestion. Celles-ci sont par ailleurs largement extensibles, que ce soit par l'intermédiaire de plug-ins existants ou par des développements. Alfresco intègre également un moteur de règles. Certaines règles sont manipulables via l'interface Web standard ce qui en facilite l'usage.



**open
source**

eXo DMS ne propose pas de moteur de règles. Néanmoins, la solution intègre un large choix d'outils techniques permettant l'implémentation de règles métier.

KnowledgeTree propose peu de fonctions de gestion spécifiques. Il existe cependant de nombreux modules qui permettent l'intégration avec d'autres applications ou l'ajout de fonctions dans l'interface. Certains de ces modules sont uniquement disponibles dans la version commerciale.

LogicaDoc ne propose pas de moteur de règles, mais permet de s'abonner aux modifications d'un document ou d'un répertoire, ou de décompresser automatiquement le contenu d'un fichier zip.

Maarch permet de mettre en place des règles de gestion par l'adjonction de module spécifiques, notamment pour la réalisation de traitements particuliers.

Jahia intègre le moteur Drools, référence en la matière. Il permet l'intégration de règles relatives aux dépôts de fichiers (extraction de métadonnées, gestion des vignettes, ajouts de métadonnées en fonction de caractéristique du fichier...).

Nuxeo dispose de nombreuses fonctions de gestion qui sont presque systématiquement apportées comme des modules. C'est l'outil Nuxeo Studio qui intègre les interfaces graphiques de manipulation du moteur de règles. Bien que relativement intuitive, celui-ci nécessite une très bonne connaissance de Nuxeo.

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 quotidiennement.

Les politiques de conservation existent depuis longtemps dans le domaine papier (physique), mais ne sont que depuis peu appliquées aux données électroniques.

Ce sont les disciplines couvrant les fonctions de l'archivage qui permettent de répondre aux questions liées à la politique de conservation, on parle d'archivistique ou de « Record Management », dans le monde anglo-saxon. Le « Record Management » est parfois considéré comme plus adapté au domaine numérique, dans la mesure où il prend en compte un sous-ensemble de documents à archiver. Sous-ensemble qui correspond alors aux « documents vitaux » de l'entreprise, c'est-à-dire nécessaires à l'activité quotidienne de l'organisation.

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



open
source

- de définir des ensembles de contenus, par rapport à des types ou des critères d'indexation
- de relier chacun de ces ensembles, à des actions de conservation, processus de destruction, durée, format et support de conservation...

Archivage

L'archivage peut être appréhendé de différentes façons en fonction du contexte de l'application. Du simple tag « archive » sur les éléments considérés comme archivés, jusqu'à la mise « hors ligne » (off Line) des contenus sur des supports optiques, par exemple.

La nécessité et la complexité de la prise en compte de l'archivage numérique vont croître avec le volume et la criticité des informations gérées.

Les applications intègrent donc des fonctions qui permettent :

- de gérer des informations de conservation
- de gérer les formats de fichiers dans la GED en fonction de leur pérennité
- de déclencher automatiquement des traitements d'archivage
- de permettre la recherche ou non parmi les éléments archivés.

Différents types de traitements peuvent être envisagés pour répondre à ces problématiques de pérennité, de complétude, de fiabilité et de traçabilité des archives.

Parmi les fonctions envisageables, on citera :

- figer les éléments archivés, par exemple en calculant des empreintes cryptographiques (signature numérique) pour en garantir l'inaltérabilité
- faire valider l'archivage (« le versement en archive ») d'un ensemble de documents, en termes de date et de responsabilité
- gérer les supports de stockage des contenus et leurs métadonnées pour les conserver sur des supports adaptés (longue durée de vie, accès fréquent...).

Format de fichier

Le format des fichiers gérés dans une application documentaire revêt une importance particulière.

Contrairement aux documents format papier, dont l'uniformité permet de s'abstraire de la plupart des questions liées au support de l'information, le format des contenus numériques (le fichier) revêt plusieurs caractères critiques.



**open
source**

Fonctions liées au format

La plupart des applications de GED peuvent stocker n'importe quel type de fichier, cependant certaines fonctions ne sont disponibles que pour certains formats.

- **La recherche en texte intégral**, c'est-à-dire dans les fichiers eux-mêmes. En effet, les formats documentaires sont encore trop souvent fermés et opaques. Néanmoins, il existe, de plus en plus souvent, la possibilité d'en extraire le contenu textuel afin de l'indexer et donc de la retrouver par des moteurs de recherche. Les formats les plus courants (PDF, MS Office, OpenOffice, HTML) figurent presque toujours parmi ceux reconnus par défaut. La question se posera pour des plus formats particuliers tels que : DWG – AutoCAD ou MP3 (pour les tags textuels), par exemple
- **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, sans avoir à les télécharger.
- **L'extraction d'informations et notamment de métadonnées**. En effet, la plupart des formats de fichiers utilisent une façon spécifique de gérer les métadonnées qu'ils embarquent. Connaître ces façons ou recourir à une normalisation (open document, JPEG ou TIFF, par exemple) permet de faciliter la récupération des métadonnées renseignées par les applications ou les matériels ayant généré les fichiers.

Cette liste n'est pas exhaustive.

Formats ouverts

Les formats ouverts et/ou normalisés 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 ouvert, documenté, libre (non soumis à un brevet ou un droit d'usage) et très largement répandu. Ce qui garantit de fait la **pérennité des fichiers**
- le second est de disposer d'un **format dont la structure est connue**, ce qui permet sa manipulation (l'édition). Ceci permet notamment d'offrir des fonctions de manipulation du contenu tout comme du fichier : transformations, combinaisons et même ajout de données dans le fichier lui-même (la référence dans la GED ou une table des matières)

Formats de conservation

Si les documents stockés dans le système documentaire doivent souvent être conservés plusieurs années (au-delà de 5 ans), leur conservation doit également inclure leur préservation, c'est-à-dire à minima qu'ils seront lisibles.

En effet, la problématique de la conservation se trouve de plus en plus posée pour les documents électroniques, dans la mesure où ils remplacent de plus en plus le papier, alors qu'ils ne faisaient que le compléter, il y a encore quelques années.

Les formats de fichiers à même de garantir la lisibilité des documents sur plus de dix années sont, soit des formats simples et très répandus, les normes de fait, par exemple : les fichiers texte (.txt) dont les fichiers XML sont parmi les plus courants et les plus utiles pour la conservation de données structurées, soit les formats normalisés au niveau international comme le PDF/A (ISO 19005-1:2005), le PDF/E (ISO 24517-1 en cours) ou l'ODF (ISO/IEC 26300:2006 - Open Document Format).

Dans le cas où ces documents sont porteurs de **signatures électroniques**, la problématique s'en trouve complexifiée (cf « Archivage numérique à valeur probante », ci-dessous).

Archivage numérique à valeur probante

Aux questions d'archivage viennent s'ajouter, les questions d'archivage « légal » ou « à valeur probante », c'est-à-dire la conservation des documents électroniques pouvant servir de preuve.

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

- **l'intégrité** : les archives ne doivent pas avoir été modifiées
- **l'authenticité** : une signature nominative et vérifiable
- **la traçabilité** : la vie de chaque document doit être connue sans faille possible, notamment via l'horodatage certifié et contigu (sans discontinuité)
- **l'auditabilité** : le système doit pouvoir être vérifié pour prouver que ses processus ne peuvent être altérés

Cet aspect de l'archivage n'est en général pas couvert directement par la solution de GED, mais par des fonctions annexes qui gèrent les signatures électroniques, les relations avec le tiers de confiance et la traçabilité de la conservation.

Alfresco propose de gérer le cycle de vie selon deux approches. D'une part, un module de Records Management, certifié DoD 5015-2 (c'est-à-dire par le département américain de la défense) qui permet notamment de gérer un plan de classement d'archivage. Et d'autre part, un module de gestion du cycle de vie (ILM – Information Life Cycle) dédié à la gestion des supports de stockage en fonction des politiques de conservation.

eXo DMS et **Jahia** proposent des fonctions de gestion associée au cycle de publication des contenus, dans une optique de diffusion Web.

Maarch dispose de nombreuses fonctionnalités liées à la gestion des archives (rappelons que c'est son cœur de métier). La Politique d'Archivage est sous-tendue par la convention d'archivage passée avec les producteurs et la gestion des cycles de vie. Les fonctions différenciantes de Maarch en la matière sont les suivantes : sélection fine des candidats au cycle, indicateur d'archive figée, double sélection manuelle/automatique, fonctions de gestion des archives à valeur probante, la gestion des espaces de stockage, le contrôle périodique des archives, l'automatisation de la réPLICATION, le stockage nativement OAIS et l'intégration du protocole SEDA en version 1.0 !

Nuxeo propose de gérer le cycle de vie des contenus et la politique de conservation des contenus par le biais de fonctions standards : métadonnées, plan de classement et cycle de vie des contenus. Notons que ces fonctions ne sont pas spécifiques à un module d'archivage, mais transverses à l'application.

Les autres solutions ne proposent pas de fonctions spécifiques pour la gestion du cycle de vie ou de l'archivage.

GESTION DES 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), l'ergonomie des solutions de GED en matière de « gestion de dossiers » 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 trois acceptations que nous lui donnerons :

- **Dans l'optique professionnelle courante**, le dossier est un ensemble de documents dont le regroupement apporte de la cohérence, et à partir duquel des utilisateurs accompliront des tâches – le dossier projet ou d'instructions, par exemple
- **Dans l'optique documentaire**, le dossier réunit des documents ayant des caractéristiques communes : un thème ou une métadonnée, par exemple – dossier des articles de presse du jour ou l'analyse financière d'une société, par exemple
- **Dans l'optique du système d'information**, le dossier regroupe alors numériquement des informations issues de différentes briques du système d'information pour en faciliter l'exploitation - le dossier client, par exemple

Ces approches peuvent être complémentaires.

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

- le format numérique permet le **multipositionnement**. Un même document peut être présent dans plusieurs dossiers, utilisé par plusieurs utilisateurs avec des approches métiers différentes



open
source

- le format numérique évite **duplication et scission des dossiers**. Un même dossier peut être partagé par plusieurs entités, géographiquement ou fonctionnellement distinctes, en permettant à chacune de conserver la main sur ses contenus, sans qu'il soit nécessaire de créer plusieurs dossiers enfants. En effet, le partage d'une partie du patrimoine d'informations conduit souvent à la duplication des dossiers, ces derniers vivent alors leur propre vie.

Les outils de la gestion de dossiers

Plusieurs moyens importants existent pour gérer ces dossiers dans les outils de GED, notamment : les informations de dossier, portées par les **métadonnées** et les **tâches** associées, le plus souvent, à un dossier dans le cadre d'un workflow.

D'autres fonctions peuvent être adaptées dans un système de GED ou développées comme des modules à part entière.

Nous regroupons ci-dessous les fonctions les plus fréquemment rencontrées.

Le partage d'informations

- définition de méta-information par dossier
- possibilité de structurer un dossier par regroupement de documents, notamment via ses métadonnées ou des sous-dossiers
- liens entre dossiers (hiérarchie, thématique, réglementation, type d'affaires, type de client...)
- gestion des droits des utilisateurs (initialisation, agent traitant, superviseur, décideur) et leurs prérogatives respectives sur chaque partie (cf structuration des dossiers ci-dessus)

La gestion des dossiers

- avec des déclencheurs automatiques ; par exemple pour l'archivage ou les rappels, dans le cadre d'une procédure ou lors de la réception d'une pièce
- interaction entre documents et données d'une application de gestion tierce ou d'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 par rapport à des listes connues
- possibilité de générer des documents à partir de modèles, de conditions ou de caractéristiques du dossier. Par exemple un formulaire ou une trame de courrier
- gestion de workflow de traitements pour automatiser les transitions et permettre un suivi automatique de l'instruction

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, règles de nommage, nomenclature...

Traçabilité

- 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. La présentation par facette permet de proposer des vues distinctes aux différents acteurs des processus d'entreprise. Les notions de vue « custom » sur la visualisation de chaque dossier permettent d'ajouter facilement des informations ou des outils de gestion simple des dossiers (check-list ou état d'avancement, par exemple)

eXo DMS gère les dossiers comme un type de contenu, c'est-à-dire qu'ils peuvent disposer de métadonnées et de toutes fonctions associées. Si aucune fonction spécifique n'est disponible, tout est prévu pour permettre d'en développer à la demande.

Jahia permet de gérer des dossiers dynamiques en fonction de métadonnées, mais n'intègre pas nativement de règles de gestion paramétrables au sein des interfaces graphiques. Il est toutefois possible d'utiliser le moteur de règles intégré à Jahia pour définir des ensembles de comportements/actions lors du dépôt de fichiers.

Maarch permet de gérer des dossiers dynamiques, correspondants à des requêtes paramétrées et de les attribuer à des utilisateurs en fonction de leur rôle.

Nuxeo répond à la problématique de gestion de dossier avec une interface et des fonctions dédiées, Nuxeo CMF. L'outil propose des notions de regroupement souple permettant de positionner chaque contenu dans autant d'espaces que voulu en fonction de règles ou de caractéristiques. La présentation permet de proposer aux différents acteurs des processus d'entreprise, des vues distinctes. La notion de section (publication) autorise une abstraction supplémentaire, transversale et indépendante des métadonnées.

IMPORT/EXPORT

Les fonctions d'import et d'export sont fondamentales dans une solution de GED. En effet, elles permettent d'entrer et de sortir les contenus d'une solution à une autre et donc de rester maître de ses choix d'outils.

Import/export en masse

L'objectif est ici de permettre l'entrée ou la sortie rapide d'un grand nombre de documents. Ces fonctions permettent de prendre en compte les métadonnées, les fichiers et, si possible, la gestion des autorisations.

Les meilleures solutions proposent des outils dédiés ou des API de bas niveau permettant d'interagir avec la base documentaire pour notamment : injecter ou extraire des contenus, ajouter ou modifier des métadonnées et les droits associés aux contenus.

Export sur support physique

Deux optiques différentes sont à envisager ici :

- **l'export définitif**, à des fins d'archivage notamment, qui consiste à « sortir » certains documents et leurs métadonnées sur des supports dits de long terme, souvent plus lents que les disques durs (disques optiques, bandes DAT...). Idéalement, la GED conserve l'historique des documents sortis, de façon à permettre des recherches sur les métadonnées
- l'export à des fins de **consultation hors ligne**. Il consiste à exporter une partie de la base documentaire sur le poste de travail de l'utilisateur ou sur un support numérique doté de sa propre interface de consultation

Alfresco permet d'importer et d'exporter sous forme d'archives des espaces entiers du référentiel, en incluant les métadonnées ainsi que les règles et les droits associés au format XML. Alfresco gère également la synchronisation complète ou sélective de contenus vers le **Cloud**, et via son module WCM dédié aux contenus pour sites Web, la publication des contenus vers des serveurs distants.

eXo DMS propose des fonctions et des routines techniques d'import-export au niveau de l'implémentation de son JCR (base de contenu).

Jahia dispose d'un module d'import-export de l'ensemble des données au format XML (contenus et droits, catégories, utilisateurs ainsi que les fichiers liés).

Knowledge Tree est capable d'importer le contenu d'un fichier Zip, avec l'adjonction de métadonnées communes et d'exporter une sélection de fichiers.

Maarch permet de gérer l'import et l'export de données. Les échanges de données à la norme SEDA 1.0 sont également pris en charge.

Nuxeo dispose d'un module qui réalise des opérations d'import et d'export de documents et de métadonnées, y compris pour les droits. Ces fonctions peuvent être pilotées par une interface en ligne de commande : le « Nuxeo Shell ». Nuxeo permet également de réaliser la réPLICATION automatique de contenus entre plusieurs bases Nuxeo. On peut aussi exporter une sélection de documents au format zip. Enfin, un mode off Line (réalisé avec Smile) permet la consultation de la base documentaire sur des postes non connectés au serveur Nuxeo.

GESTION DES EMAILS

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

Les problématiques de gestion de la connaissance 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 plus ou moins automatique des emails peut donc apparaître comme une solution à ces problématiques.

Sélection des emails à archiver

Le processus d'archivage d'un email peut très difficilement 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. Celle-ci peut passer par l'intégration d'un plug-in dans le gestionnaire de courrier (Mozilla, Outlook...) ou par l'intégration automatique d'emails envoyés à une adresse email spécifique du type stockage_ged@votreorganisation.com, par exemple.

Gestion des emails

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

D'une part, il n'existe pas d'homogénéité quant au nombre ou au 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 email sont le plus souvent multipartites (à plus de deux interlocuteurs), le stockage d'un email en plusieurs exemplaires générerait alors des doublons inutiles.

Il ne s'agit là que d'un échantillon des questions soulevées par la gestion des emails.

La GED, dans une optique d'ECM, doit donc proposer des solutions pour :

- gérer les formats de conservation des emails et des pièces jointes, afin d'assurer une restitution optimale
- détecter les doublons pour éviter les stockages multiples et proposer des processus de travail collaboratif, par exemple lors de l'échange d'une pièce jointe légèrement modifiée dans plusieurs emails successifs

- indexer le contenu des emails et les pièces jointes, et gérer les liens entre ces différents contenus

Alfresco permet de gérer les emails par simple glisser-déposer depuis Outlook. Les informations de base (expéditeur, destinataires, sujet) sont automatiquement extraites. Le contenu de l'email et des pièces jointes est disponible en recherche texte intégrale. Précisons également que dans Alfresco, chaque document ou dossier dispose de sa propre adresse email, ce qui permet de lui « adresser » un email. Le comportement peut être paramétré : un dossier va ainsi stocker le mail et les pièces jointes, un document peut recevoir une annotation correspondant au contenu de l'email...

eXo DMS ne dispose pas nativement de la gestion des emails. Cependant, il existe dans la suite eXo un outil dédié à cet usage qui repose sur le même stockage d'eXo DMS.

Knowledge Tree, dans sa version commerciale, dispose d'une très bonne intégration à Outlook. Il est possible d'archiver les emails dont les métadonnées sont automatiquement extraites, mais aussi d'attacher des fichiers du référentiel lors de l'envoi d'un email.

Nuxeo permet de gérer les emails via simple glisser-déposer depuis Outlook. Le contenu de l'email et des pièces jointes est disponible en recherche texte intégrale. Nuxeo propose également un type de dossier « courriel » qu'il est possible de lier à une boîte de messagerie (récupération automatique de tous les messages de cette boîte dans Nuxeo). Les utilisations sont multiples et facilitent l'usage de boîtes emails partagées en les rendant disponibles dans la base documentaire.

Maarch propose un module d'import d'emails : Maarch Mail Capture qui permet d'extraire le contenu d'une boîte mail dans Maarch.

Les autres outils ne disposent pas de telles fonctions.

INTEGRATION DE LA GED

Lorsque l'application de GED vise en grande partie la production de documents bureautiques, son intégration aux applications du poste de travail de l'utilisateur est alors déterminante.

Dans ce contexte, l'ergonomie et l'efficacité du travail au quotidien, la fluidité du passage de la GED aux applications utilisatrices de documents numériques (applications bureautiques par exemple), sont des critères déterminants en ce qui concerne l'acceptation de l'outil et donc de la réussite du projet.

Deux principaux niveaux d'intégration sont envisageables. Notons que le choix impacte à la fois la complexité de l'environnement technique et la qualité de l'ergonomie.

Toutes les solutions présentées proposent une interface Web pour interagir avec la GED.

Espace de stockage

La GED peut être considérée comme un espace de stockage, similaire à un disque dur ou un espace réseau : possibilité de gestion de dossiers, de navigation entre les différents niveaux, voire de déplacements d'éléments directement par « glisser-déposer » depuis le poste de l'utilisateur.

L'essentiel intérêt de cette approche est que le changement est mineur pour les utilisateurs, par rapport aux simples serveurs de fichiers auxquels ils étaient habitués. Cela facilite l'acceptation de l'outil. La contrepartie immédiate est la perte d'éléments tels que les métadonnées ou l'interface de recherche. Cette approche est donc le plus souvent complétée par une interface Web ou bureautique dédiée à la GED qui apporte ces fonctions.

L'utilisation du référentiel documentaire est alors proposée dans l'environnement habituel des utilisateurs, prenant en compte la gestion des habilitations de la GED, mais sans tirer pleinement parti de l'application de gestion documentaire sous-jacente.

Technologiquement, cette approche est rendue possible par des interfaces telles que Webdav, CIFS, WSS (Microsoft) ou même FTP. Il s'agit de technologies implémentées comme surcouche d'une base documentaire et qui permettent d'accéder à la GED par des outils standard du poste client.

Accès depuis la bureautique

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

Les utilisateurs demandent à pouvoir réaliser des tâches habituelles depuis leurs outils bureautiques : ouverture, indexation, enregistrement dans la base documentaire, voire avancement d'un workflow.

Tout l'intérêt de cette approche est d'être très peu intrusive par rapport aux habitudes des utilisateurs notamment en ce qui concerne les phases de travail les plus fréquentes, tout en apportant les bénéfices d'un système de fichiers classiques.

Cette approche, lorsqu'elle n'est pas implémentée par l'application documentaire, nécessite un investissement important.

Remarque : l'ouverture par Microsoft du code source de son protocole WSS, utilisé entre SharePoint et MS Office, a permis à de nombreuses applications de GED de l'intégrer. Ceci permet d'utiliser MS Office et Windows comme s'ils étaient en relation avec un serveur SharePoint, c'est-à-dire via les panneaux intégrés à la suite bureautique et le montage de « disque réseau » sous Windows.

Ces plug-ins permettent de disposer, lorsque l'implémentation est complète, des fonctions suivantes : ouverture, édition, réservation, accès aux versions, accès aux workflows et la navigation dans les espaces.



**open
source**

Alfresco a, de longue date, été précurseur dans le domaine de l'intégration bureautique. Il propose notamment l'accès à la base par un partage réseau (en CIFS), l'accès en webdav, FTP ou l'intégration dans la suite bureautique Microsoft via un module spécifique. L'intégration WSS est aujourd'hui opérationnelle et remplace avantageusement toutes ces fonctions. Notons également l'intégration avec Google Drive.

eXo DMS propose un accès Webdav, CIFS et FTP à son référentiel de contenu. Deux modules à installer sur les postes de travail permettent l'intégration aux suites MS Office et Open Office.

Jahia propose une intégration limitée au poste de travail, via la navigation de la médiathèque depuis l'explorateur Windows ou Internet (« webfolders »). La gestion des versions et la notion de réservation ne sont pas prises en charge. Une gestion de version légère est en place depuis la version 6.5.

Knowledge Tree dispose d'un accès Webdav et de modules complémentaires (non open source) permettant une très bonne intégration dans Windows (exploration du référentiel via un navigateur dédié et notion de « hot folder »), MS Office (édition des documents) et Outlook (indexation des emails). Ces extensions sont une exclusivité de la version commerciale.

LogicaDoc propose un accès Webdav au référentiel. La solution propose également une fonction intéressante de référencement de document sur des postes réseau distants. Ce référencement comprend l'authentification sur le poste et même l'ajout de templates et de tags par défaut pour les documents référencés.

Maarch dispose d'une extension (applet JAVA) qui permet aux utilisateurs d'éditer les fichiers bureautiques directement dans leur application bureautique, sans avoir besoin de télécharger le document au préalable. Signalons également une fonction de publipostage (utilisant *TinyButStrong*) avec Open Office, des fonctions de visualisation et d'intégration avec les scanners qui facilite leur indexation. Maarch permet également de modifier un document bureautique depuis l'interface Web.

Nuxeo propose une interface Webdav et WSS pour l'accès aux documents. L'intégration WSS est opérationnelle et permet de disposer de nombreuses fonctions classiques en terme d'intégration bureautique. L'ajout de document est également possible par glisser-déposer entre le poste de travail et l'interface web. Enfin, Nuxeo propose, via le plug-in LiveEdit, l'automatisation du processus de modification en ligne depuis MS Office et Open Office.

Quotero propose une intégration Webdav. Il annonce également une intégration MS Office et OpenOffice, et un accès via client riche que nous n'avons pas pu tester.

INTEGRATION TECHNIQUE

Les outils de GED peuvent être considérés de plusieurs façons en matière d'intégration technique. D'une part, ils peuvent être considérés comme des applications à part entière, sans réel lien avec le système d'information en dehors de la gestion des utilisateurs. Ou alors, comme c'est de plus en plus souvent le cas, ils sont considérés comme l'une des briques d'un système plus global où est centralisé l'ensemble des documents d'une organisation.

Configuration de l'interface

Un outil de GED peut être très complet, et parfois délivrer des interfaces qui ne conviennent pas aux utilisateurs cibles. S'il est toujours envisageable, avec des solutions open source, de faire les adaptations directement sur le code de l'application, c'est au risque de rendre les mises à jour plus complexes. Certaines solutions répondent à ce problème en permettant de modifier leur interface par configuration ou via des outils dédiés.

Alfresco dispose de deux interfaces relativement figées et d'un framework de développement. L'interface « Alfresco UI » propose en standard un très grand nombre de fonctions, beaucoup sont adaptables par configuration XML : définition des assistants et actions disponibles ainsi que certains comportements. Les évolutions peuvent être assemblées sous la forme d'un module ou bien développées sous forme de « custom view » comme ajout à l'interface standard, ce qui permet de les appliquer facilement lors de mises à jour. Share, la seconde interface, est bâtie sur le framework SURF. Share est totalement dédiée au travail collaboratif. Notons que le principe d'une page d'accueil totalement configurable par l'utilisateur est commun aux deux interfaces standards. Il existe dans la communauté Alfresco plusieurs autres interfaces plus ou moins matures.

eXo DMS est bâti sur eXo portal, solution elle-même dédiée à la conception de portail Web. De ce fait, tous les éléments de l'interface d'eXo DMS sont paramétrables et extensibles. L'interface proposée nativement est assez technique et nécessite une adaptation. Rappelons cependant qu'eXo DMS est le plus souvent utilisé par des profils « semi-techniques ».

Jahia propose nativement une interface d'une grande simplicité d'utilisation. Etant donné son orientation vers la publication de contenus, l'interface concernée est celle du back-office. L'interface de restitution offre quant à elle une très grande souplesse dans l'adaptation aux besoins des utilisateurs. Nous sommes ici dans une logique de conception d'interfaces dédiées. Notons que le mode insite editing pour l'ajout de composant directement dans les pages des sites générés (depuis la 6.5) propose une grande souplesse de restitution.

Knowledge Tree permet de modifier facilement le texte des différents messages de son interface, mais il n'est pas prévu d'outils spécifiques de manipulation de l'interface au-delà de la page d'accueil. Les adaptations sont donc le plus souvent réalisées dans le code même de l'interface.

LogicaDoc propose quelques options d'interface, mais ne dispose pas d'outils spécifiques pour leur conception. Les adaptations sont réalisées par l'usage de thèmes ou skins.

Maarch propose plusieurs possibilités d'adaptation simple de l'interface par paramétrage.

Nuxeo dispose d'interfaces relativement complètes, ainsi que d'un éditeur de thèmes pour configurer certains aspects (disposition, thème...). Pour les modifications plus poussées, l'interface tout entière (fonctionnellement et graphiquement) est construite sur un système de plug-ins qui permet de packager les modifications développées pour faciliter leur application sur les versions ultérieures. Nuxeo Studio permet d'apporter une souplesse importante dans la configuration de certains éléments d'interface : page d'accueil, logo, écrans de saisie, de modification et de recherche avancée. Notons également que la page d'accueil est totalement configurable par l'utilisateur via l'utilisation de gadgets Open social.

Intégration applicative

L'intégration permet aux applications de gestion de contenu de s'intégrer, de converser avec d'autres applications, le plus souvent productrices ou utilisatrices de contenus.

Plusieurs standards et technologies sont disponibles pour cela :

Les Web services

Les services REST et SOAP, qui sont deux formats d'échanges et d'expositions de fonctions métiers utilisés dans la création d'API Web.

Web services SOAP

Spécification utilisant le langage XML au format standard SOAP (Simple Object Access Protocol). Les Web Services permettent de réaliser des échanges entre un client et un serveur via le protocole HTTP. Les Web Services mettent à disposition un ensemble de points d'entrée métier (Endpoint) décrit dans un contrat au format standard WSDL (Web Services Description Language).

Web services REST

Plus simple et plus performant que la spécification SOAP, les services REST (REpresentational State Transfer) se limitent à l'utilisation du protocole HTTP pour les entrées-sorties. Ils ne sont pas contraints à utiliser XML comme langage et leurs échanges entre le client et le serveur sont moins verbeux que ceux de SOAP, ce qui les rend plus performants.

CMIS pour Content Management Interoperability Services

Il s'agit d'une norme émergente, probablement l'une des plus prometteuses depuis le JCR ou le Webdav, pour homogénéiser les accès à une gestion de contenu, notamment documentaire. Cette norme consiste à proposer un accès technique homogène aux outils de gestion des contenus qui propose cette interface. En d'autres termes, cela permet qu'une application comme eZ Publish (CMS dédié à la conception de site Web) puisse exploiter les contenus stockés dans des bases de gestion documentaire telles que : Alfresco, Nuxeo ou Knowledge Tree de façon agnostique.

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 une utilisation précise. On peut citer parmi les cas d'utilisations : l'insertion dans un portail ou dans une application métier, par exemple.

Les solutions proposant des frameworks de conception d'interfaces sont les mieux à même de répondre à ce besoin.

Portlets dans un portail

L'outil de GED peut fournir un certain nombre de « portlets », composants qui s'installent sur un portail JEE et qui donnent directement accès à certaines fonctionnalités.

Certaines applications de GED proposent nativement la « portletisation » (mise en portlet standard) de certains composants ou pages (tableaux de bord, recherche ou navigation, par exemple).

Nous vous invitons à consulter le livre blanc *Portails open source* (<http://www.smile.fr/Livres-blancs>) pour plus de détails sur ce sujet.

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

eXo DMS est une solution dédiée à l'intégration. Comme nous l'avons vu plus haut, son usage est essentiellement pertinent dans le cadre de la suite eXo Platform. En effet, eXo DMS se comporte principalement comme une interface de gestion des contenus exploités par les autres briques de la plateforme. eXo DMS dispose donc de tous les éléments nécessaires pour pouvoir être intégré, éléments très largement orientés vers les autres briques de la suite eXo.

Jahia propose une architecture orientée service et une couche d'accès REST aux contenus de la plateforme. De plus, il permet le développement de composants d'accès et de manipulation de ces données.

Knowledge Tree dispose de webservices.

Maarch dispose de webservices SOAP et REST.

LogicaDoc propose quelques webservices.

Nuxeo dispose d'un ensemble de webservices, 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'applications sur le poste de travail ou nécessitant des traitements lourds.

Mode SaaS

2009 a vu l'émergence de nombreuses solutions en mode SaaS - Software as a Service – les applications de gestion de contenu, notamment de GED n'ont pas échappé à la tendance et on a vu l'éclosion du portage de différentes solutions de GED, notamment open source, sur le Cloud.

Ce mode permet d'appréhender les projets de GED sous un angle nouveau en voyant l'application comme un service et non plus comme une solution, le déploiement est plus simple, puisqu'il est industrialisé, souvent par l'éditeur lui-même. La contrepartie est que les développements sont souvent plus limités.

Ce type d'offre est encore souvent vu aujourd'hui, soit comme une façon de lancer très rapidement un projet, puisqu'une infrastructure complète peut être montée en quelques heures, soit comme la première phase d'un projet plus ambitieux en se servant de l'instance de Cloud comme d'un projet pilote pour valider.

Alfresco propose une machine Amazon EC2 pour ses versions Labs et entreprise (limitées à 30 jours).

Nuxeo propose un service de location d'instance packagé sur la plate-forme d'Amazon EC2.

KnowledgeTree propose la version KnowledgeTreeLive également hébergée chez Amazon EC2. Cette version propose notamment des plug-ins commerciaux spécifiques tels que l'intégration avec Zoho, solution de bureautique en ligne.

SYNTHESE

La gestion documentaire dans les organisations a pour finalité la gestion d'une information opérationnelle et souvent vitale. Le corollaire est que beaucoup de projets de gestion documentaire sont complexes, car ils mêlent les aspects techniques et fonctionnels propres à la gestion de contenu numérique, mais aussi métiers, propre à différentes fonctions dans l'organisation.

La GED est une brique stratégique du système d'information, aussi vital que le mail ou le partage de fichiers (partage réseau) qui existe dans la très large majorité des organisations.

Le caractère open source d'une telle solution offre des atouts importants : **robustesse**, **évolutivité** et **pérennité** en premier lieu ; mais aussi des avantages qui découlent de l'écosystème open source lui-même : des prestataires nombreux, ainsi qu'un **support de haut niveau** et un **dynamisme dans l'évolution des produits** sans égal, car ces deux aspects sont le socle du business modèle des éditeurs open source.

Les coûts éventuels associés à l'utilisation de certaines solutions reflètent deux types de besoins : la garantie du patrimoine d'information qui est hébergé (support) ou des fonctions complémentaires qui ne seraient pas contenues dans la version « communautaire » des solutions (modules).

L'open source dispose d'un coût d'entrée extrêmement compétitif puisque la souscription n'est pas obligatoire. Ceci permet de recourir à ces solutions pour prototype ou réaliser un pilote à un coût extrêmement faible. Et ce faisant, convaincre sur l'utilisation de la GED dans l'organisation, jalon encore important de l'introduction de la GED dans le système d'information centrale des organisations.

Comme vous avez pu le voir à travers ce livre, deux produits se distinguent par leur qualité et leur excellente faculté à réaliser des projets de GED : Alfresco et Nuxeo. Quatre autres correspondant à des typologies de projets un peu différentes : Jahia, Maarch, Knowledge Tree et Phraseanet.

Tous disposent d'une pérennité de très bon niveau, de services de support par leur éditeur et d'un réseau d'intégrateurs conséquent à même de réaliser des projets.

Le choix de la bonne solution pour votre projet s'effectuera donc sur trois critères principaux : les fonctionnalités natives, les questions d'architecture technique et sur le modèle économique permettant d'optimiser la valeur de chaque projet.

Notre objectif était non seulement de présenter les solutions, mais également notre vision de la gestion documentaire.

Nous espérons vous avoir éclairé sur ces questions et restons à votre disposition pour mettre notre expertise à la disposition de vos projets. Une seule adresse : contact@smile.fr

D'autres livres blancs Smile sont disponibles en ligne, en libre téléchargement, sur www.smile.fr/Livres-blancs.