







Introduction à la GED et à l'ECM

Stefane Fermigier - Cours à l'Epita - 2011/10/03

**Stefane Fermigier**
@sfermigier
#18,159,539



Edit Profile...

bio

Founder of Nuxeo & OSS4Cloud; cofounder of GTLL. Interests include: entrepreneurship, open source, ECM, software craftsmanship, the cloud and the semantic web.

location

Paris, France

web

fermigier.com/

following

722

followers



1,199

tweets

2,000

favorites

35





Stefane Fermigier

@sfermigier

#18,159,539



Edit Profile...

bio Founder of Nuxeo & OSS4Cloud; cofounder of GTLL. Interests include: entrepreneurship, open source, ECM, software craftsmanship, the cloud and the semantic web.

location Paris, France

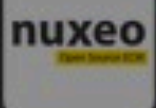
web fermigier.com/

following 722

followers 1,199

tweets 2,000

favorites 35





WARNING

**This presentation
contains lots of
TLAs!**

Le contexte

Information society

From Wikipedia, the free encyclopedia

For other uses, see [Information society \(disambiguation\)](#).

The aim of the **information society** is to gain competitive advantage internationally through using IT in a creative and productive way. An **information society** is a [society](#) in which the creation, distribution, diffusion, use, integration and manipulation of [information](#) is a significant economic, political, and cultural activity. The [knowledge economy](#) is its economic counterpart whereby wealth is created through the economic exploitation of understanding. People that have the means to partake in this form of society are sometimes called [digital citizens](#). As Beniger^[1] shows, this is one of many dozen labels that have been identified to suggest that we are entering a new phase of society.

Source:Wikipedia

“Knowledge-based Economy”

"The concept that supports creation of knowledge by organizational employees and helps and encourages them to transfer and better utilize their knowledge that is in line with company/organization goals"

Source: Wikipedia

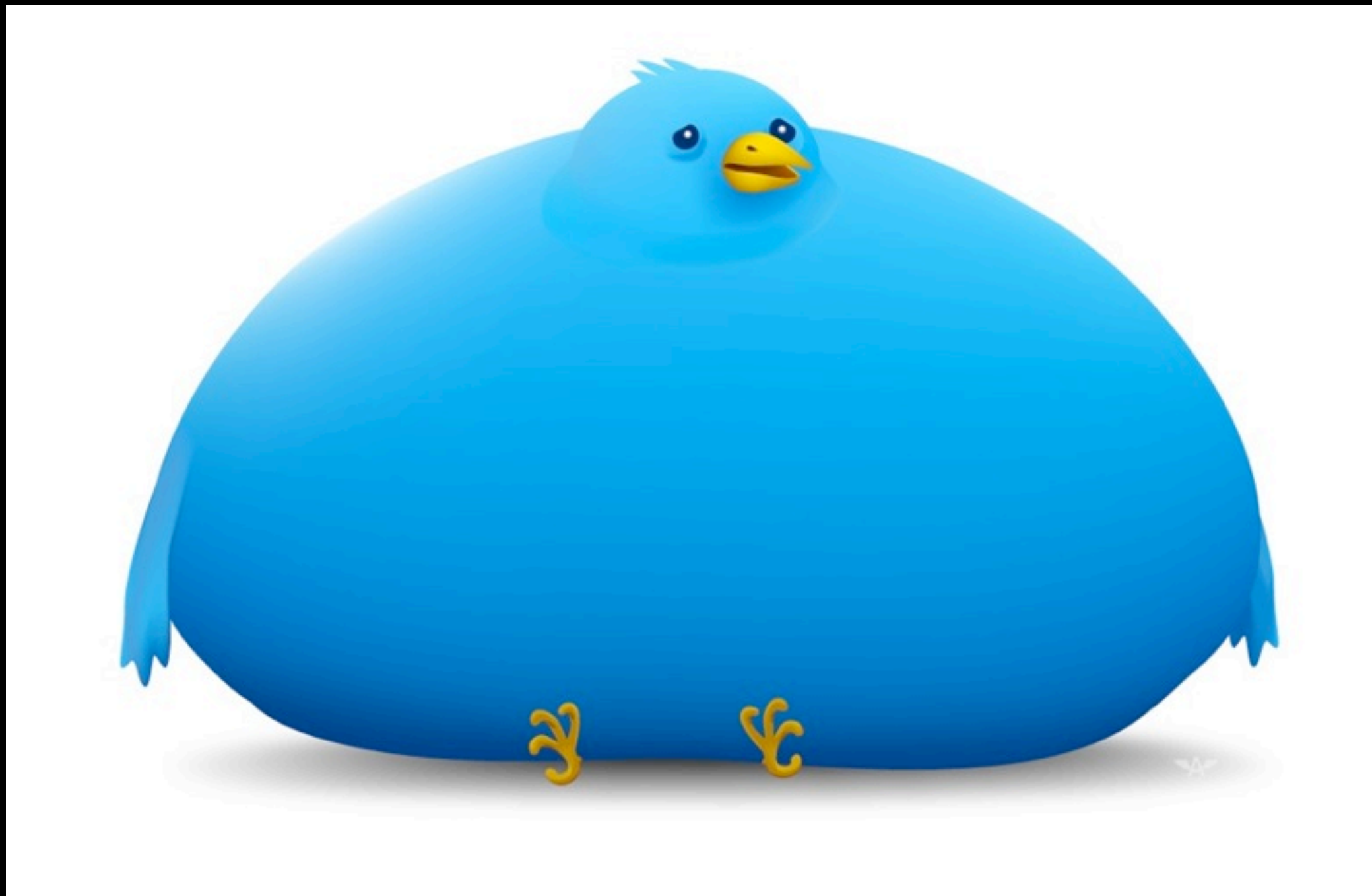
“Knowledge-based Economy”

The initial foundation for the Knowledge Economy was first introduced in 1966 in the book *The Effective Executive* by Peter Drucker. In this book, Drucker described the difference between the **manual worker** and the **knowledge worker**. The manual worker, according to him, works with his hands and produces goods or services. In contrast, a knowledge worker (page 3) works with his or her head not hands, and produces **ideas, knowledge, and information**.

Source: Wikipedia

Le revers de la médaille

L'Infobésité



Quelques chiffres

- 50% de contenu produit chaque année en plus par les entreprises
- 1.8 zettaoctets en 2011 (10^{21})
- En 2005, un employé perdait en moyenne 3.5 heures / semaine à chercher sans succès des documents existants, et 3 heures à les recréer

Un peu d'histoire

Classification de Dewey (1876)

La **classification décimale de Dewey (CDD)** est un **système visant à classer** l'ensemble du savoir humain à l'intérieur d'une **bibliothèque**, développé par **Melvil Dewey** en **1876**. Elle a été complétée et perfectionnée par la **classification décimale universelle (CDU)** développée par **Henri La Fontaine** et **Paul Otlet**.

Exemple de divisions successives

- 600 - Techniques. (l'indice est 600, car il faut au moins trois chiffres)
 - 640 - Vie domestique. (l'indice est 640, car il faut au moins trois chiffres)
 - 641 - Alimentation
 - 641.5 - Cuisine. (un point sépare le troisième et le quatrième chiffres)
 - 641.57 - Cuisine pour les collectivités

Source: Wikipedia

The Filing Cabinet (1898)

Vertical file

The **vertical filing cabinet** (**vertical file cabinet** in the United States) more or less as in use today was invented by [Edwin G. Seibels](#) in 1898.^[1] He invented a vertical filing system in 1898 that revolutionized record-keeping. Previously, businesses kept papers in envelopes in turn stored in arrays of pigeonholes often lining a wall. Finding and opening envelopes and unfolding papers was troublesome and inefficient. Seibels reasoned that folding was not necessary; papers could be kept in large envelopes standing on end vertically in a drawer.



Source: Wikipedia

Microfiches (1906)

Microfiche

La **microfiche** est l'un des plus compacts supports de stockage **analogiques** couramment utilisés. Elle est utilisée en règle générale dans les institutions (comme par exemple les **bibliothèques universitaires**) qui ne peuvent disposer de l'espace nécessaire au stockage des documents papier correspondants.

En 1906, **Paul Otlet** et **Robert Goldschmidt** inventent la microfiche normalisée pour gérer la documentation.

Chaque microfiche contient de 100 à 130 pages environ en fonction de la taille de l'original. Une collection de 20 000 microfiches, soit de 10 000 à 20 000 livres, tient dans un meuble de 1,5 x 0,5 x 2 mètres. Certaines sociétés sont spécialisées dans la fourniture de telles collections.



Microfiches du Gemeinsame Körperschaftsdatei (GKD), le fichier allemand de noms de collectivités.

Source: Wikipedia

La dématérialisation (depuis 1970)

Echanges sur les places de marchés boursières

Echanges entre entreprises (ex: EDI, EDIFACT)

Echanges entre l'état et les administrés (ex: impôts en ligne, appels d'offres...)

Dans bien des secteurs, on conserve une dualité papier / numérique, avec transformation de l'un à l'autre (scan / impression)

Le groupware (1981)

Processus humain et organisationnel du travail en groupe et outils technologiques nécessaires à l'accomplissement de ce travail

Concrètement: messagerie + agenda partagé + dossiers partagés + conférence

Ex: Lotus Notes, MS Exchange, Google Apps

Le disque partagé (1984 et 1988)

Parfois appelé “disque windows”

NFS (Sun): 1984 et SMB (Microsoft): 1988

Encore le moyen dominant de partager des documents dans les entreprises

Documents classés selon le principe du “plan de classement” hiérarchique

La GED (1980)

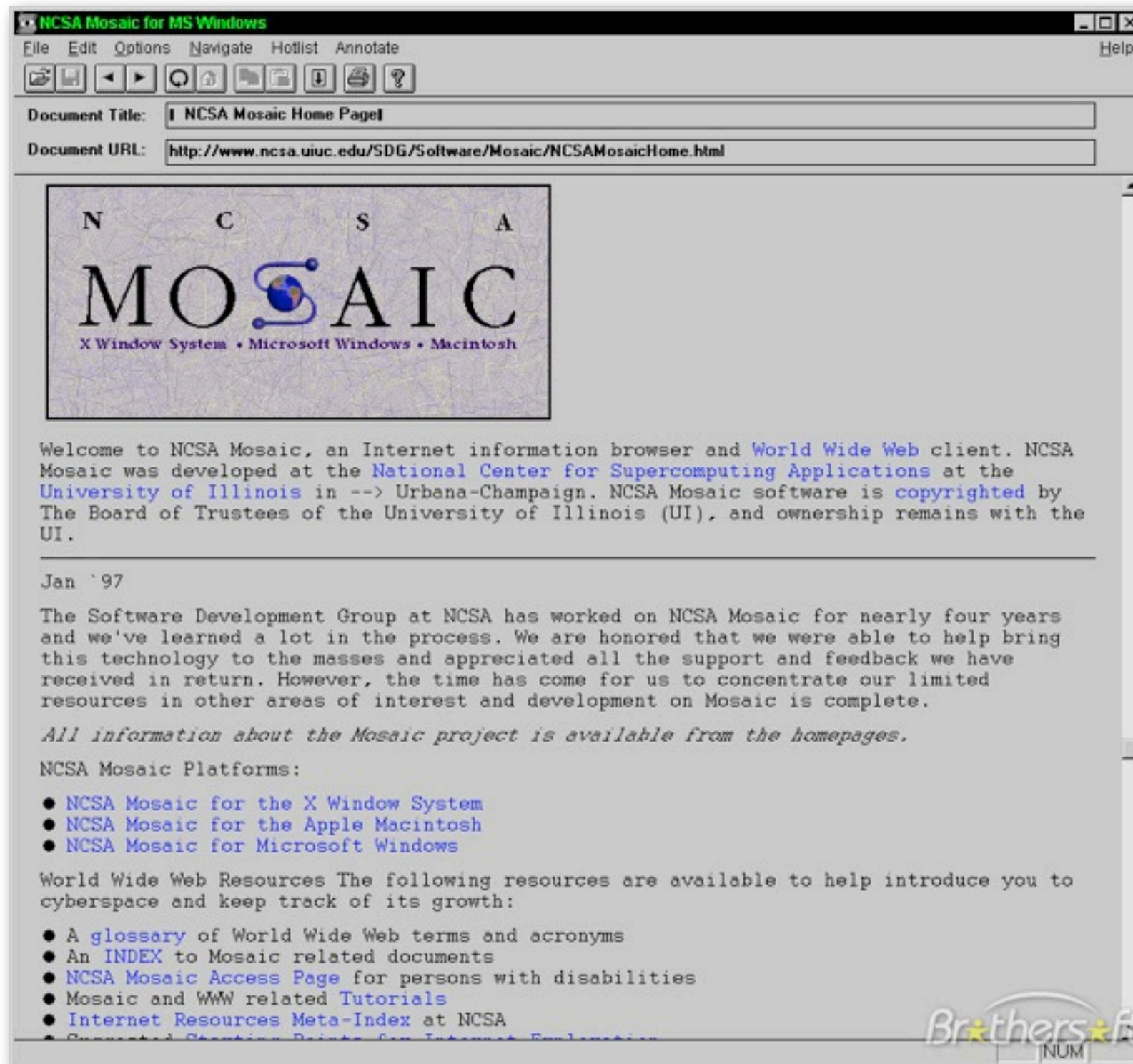
Indexation de documents stockés
physiquement hors du système informatique
(bibliothèques, archives physiques, microfiches)

Numérisation sur disques optiques (COLD)

Gestion de la production de documents
complexes et à forte valeur ajoutée (pharma,
aérospatial...) à partir de 1990

Gestion des documents de référence de
l'entreprise

Le Web (~1995)



Google (1998)



L'ECM (2000)

Apparaît comme la convergence de plusieurs technologies liées au documents: GED, WCM, DAM, Demat, etc.

Idée de remplacer des solutions ponctuelles de gestion du contenu ou de l'information par une infrastructure structurante

Le Web 2.0 (2003)

Web 1.0		Web 2.0
DoubleClick	-->	Google AdSense
Ofoto	-->	Flickr
Akamai	-->	BitTorrent
mp3.com	-->	Napster
Britannica Online	-->	Wikipedia
personal websites	-->	blogging
evite	-->	upcoming.org and EVDB
domain name speculation	-->	search engine optimization
page views	-->	cost per click
screen scraping	-->	web services
publishing	-->	participation
content management systems	-->	wikis
directories (taxonomy)	-->	tagging ("folksonomy")
stickiness	-->	syndication

Source: Tim O'Reilly - <http://oreilly.com/web2/archive/what-is-web-20.html>

L'Entreprise 2.0 (2006)

Early View of Enterprise 2.0 Elements

- S** **Search** – Discoverability of information drives reuse, leverage, and ROI.
- L** **Links** – Using URIs to forge thousands of deep interconnections between enterprise content 24/7.
- A** **Authorship** – Ensuring every worker has easy access to Enterprise 2.0 platforms.
- T** **Tags** – Allowing natural, organic, on-the-fly organization of data from every point of view.
- E** **Extensions** – Extend knowledge by mining patterns and user activity.
- S** **Signals** – Make information consumption efficient by pushing out changes.

A more refined conception of Enterprise 2.0 for 2007?



Source: Dion Hinchcliffe - <http://www.zdnet.com/blog/hinchcliffe/the-state-of-enterprise-20/143>

Les tablettes? (2010)



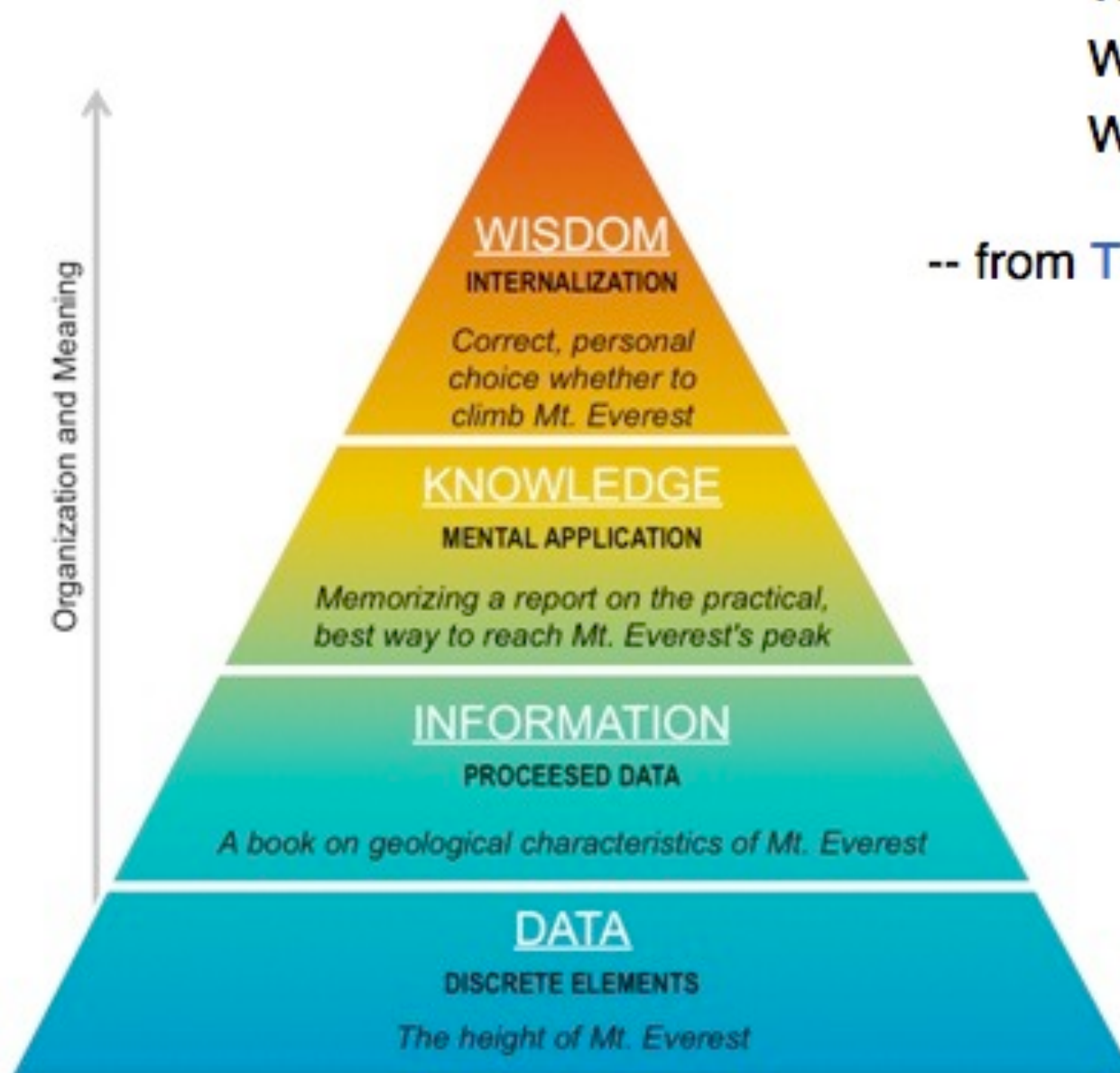
Contenu, document, etc.

Les définitions fondamentales

La hiérarchie DIKW

Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?

-- from T.S. Eliot, "Choruses from 'The Rock'"



Information is not knowledge,
Knowledge is not wisdom,
Wisdom is not truth,
Truth is not beauty,
Beauty is not love,
Love is not music,
and Music is THE BEST.

-- from Frank Zappa, "Packard Goose"

Data are discrete symbols that represent facts. You might think of them as recordings or statistics.

- There is no meaning or significance beyond the data's existence.
- Data may be clean or noisy; structured or unstructured; relevant or irrelevant.

Information is data that has been processed to be useful. I like to think of it as adding the first bit of context to data relating to "who", "what", "where", and "when".

- Information captures data at a single point in time and from a particular context; it can be wrong.

Knowledge is the mental application of data and information. Most consider this as addressing questions around "how".

- Some consider knowledge a deterministic process, which is to say the appropriation of information with the intent of use.

Wisdom is the evaluation and internalization of knowledge. It applies insight and understanding to answer "why" and "should" questions.

Information

L'information est une collection de *données* organisées pour donner forme à un *message*.

Source: “Ingénierie documentaire”, Stéphane Crozat et Bruno Bachimont (CC BY-NC-SA)

Contenu

Un **contenu** est une forme d'expression pourvue d'une valeur culturelle associée à un véhicule matériel, il exprime une signification et suscite une réception et une interprétation.

Ex: Une définition donnée oralement est un contenu.

Inscription

Une **inscription** est un contenu fixé sur un support matériel, tel qu'il lui apporte une permanence dans le temps.

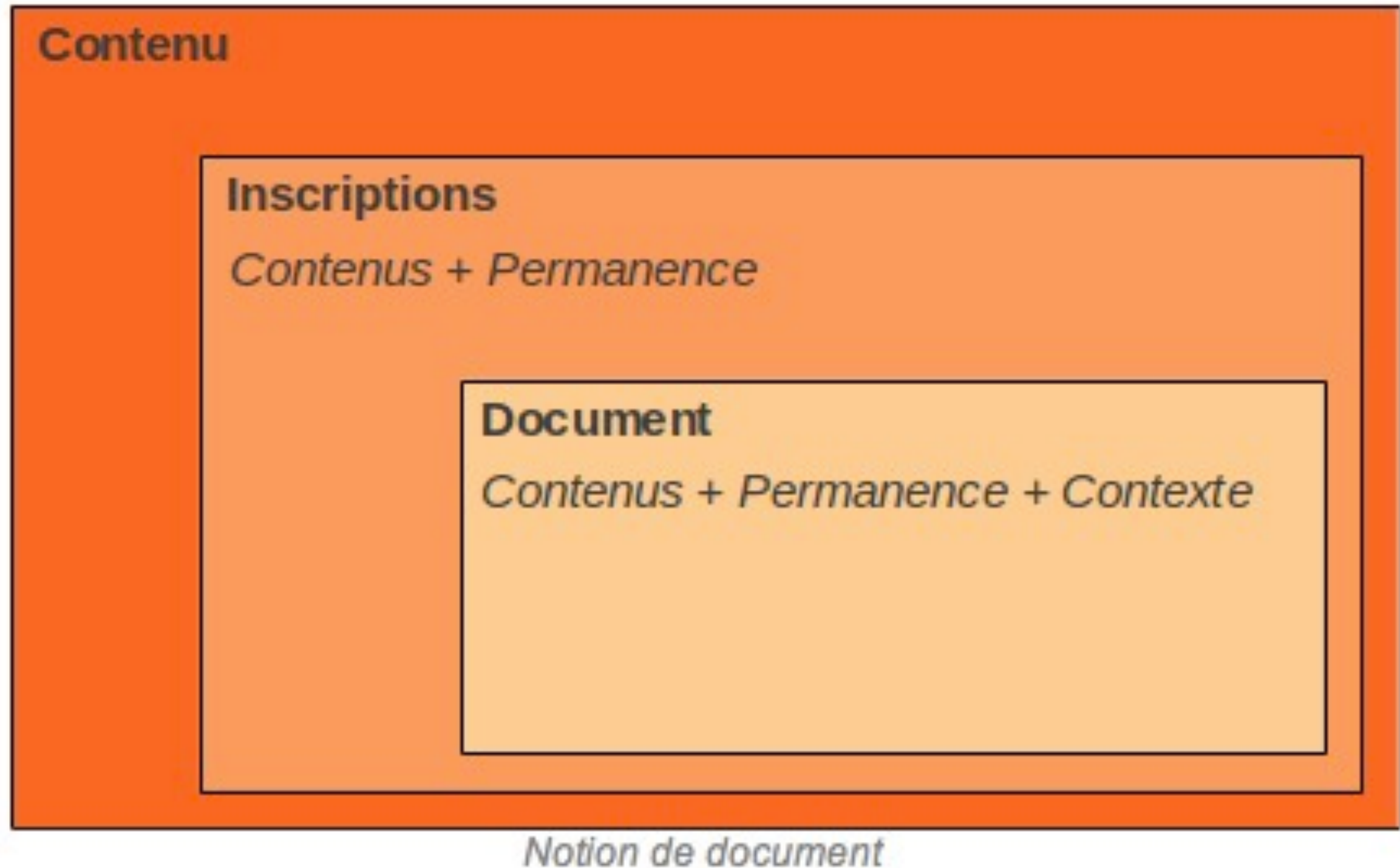
Ex: Une définition écrite est une inscription.

Document

Un **document** est une *inscription de contenus* sur un support pérenne, établie dans un contexte de production et pour un contexte de réception.

Ex: un glossaire de définitions publié sur un livret, par un enseignant pour ses étudiants est un document.

Source: “Ingénierie documentaire”, Stéphane Crozat et Bruno Bachimont (CC BY-NC-SA)



Source: “Ingénierie documentaire”, Stéphane Crozat et Bruno Bachimont (CC BY-NC-SA)

Composition d'un document (texte)

Les documents électroniques, tout comme les documents traditionnels, sont composés des quatre éléments formels : un **contenu** (le texte brut), une **structure** (l'organisation logique du texte), une **présentation** (la mise en page) et un **contexte** de création. Un document permet de conserver et de communiquer une information.

GED (1980)

Le terme de **GED** (gestion électronique de documents) apparaît dans les années 80. Son objectif initial est la gestion de l'archivage de documentations papiers, puis électronique, par la gestion de *métadonnées*.

Source: “Ingénierie documentaire”, Stéphane Crozat et Bruno Bachimont (CC BY-NC-SA)

GEIDE (1994)

Le terme de **GEIDE** est promu en 1994 pour ancrer la gestion des formats électroniques et le passage à la numérisation/dématérialisation (images scannées, PDF, bureautique via RTF notamment)

Source: “Ingénierie documentaire”, Stéphane Crozat et Bruno Bachimont (CC BY-NC-SA)

ECM (2000)

L'**ECM** (Enterprise Content Management ou gestion de contenu d'entreprise) regroupe les systèmes informatiques permettant de gérer les *processus documentaires* au sein d'une organisation

Source: "Ingénierie documentaire", Stéphane Crozat et Bruno Bachimont (CC BY-NC-SA)

ECM (def. alt.)

Enterprise Content Management (ECM) is a formalized means of organizing and storing an organization's documents, and other content, that relate to the organization's processes. The term encompasses *strategies, methods, and tools* used throughout the *lifecycle* of the content.

Les enjeux pour l'entreprise

Types d'avantages concurrentiels

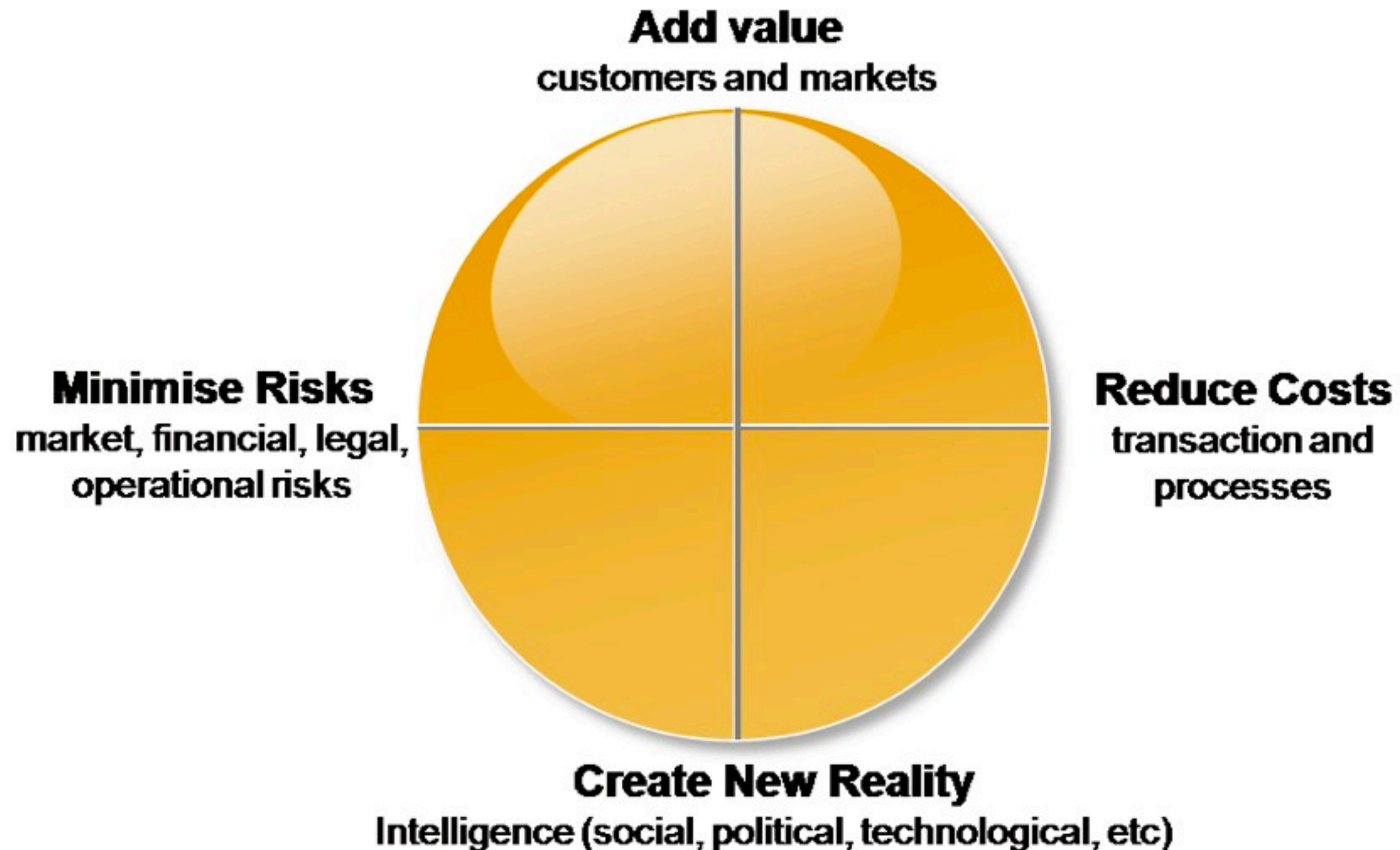
Efficienne: faire correctement, mieux, plus vite, ce qu'on fait déjà (baisser les prix, augmenter les marges)

Réduction des risques

Retour sur actifs: augmenter les gains générés par des ressources existantes

Stratégique: lancer de nouvelles offres, établir une barrière à l'entrée

Marchand's Strategic Information Alignment Framework





Source: Capitalizing on Complexity - Insights from the Global CEO Study, IBM, 2010

Exercice

Donner une liste d'avantages concurrentiels que l'on peut attendre de la GED ou de l'ECM (dans les domaines de l'efficience, du retour sur actifs et de l'avantage stratégique)

Gains de productivité

TABLE 2

Hours Wasted per Week per Task

Task	Average Weekly Hours	Cost per Worker per Week* (\$)	Cost per Worker per Year* (\$)	Annual Cost to Enterprise with 1,000 Information Workers (\$)
Reformatting from multiple formats into one document format	3.8	110	5,701	5,700,760
Search but not find	3.5	101	5,251	5,250,700
Recreating content	3	87	4,501	4,500,600
Multichannel publishing with multiple applications	2.8	81	4,201	4,200,560
Moving documents from one format to another	2.4	69	3,600	3,600,480
Acquiring archived records with little or no automation	2.3	66	3,450	3,450,460
Version control issues	2.2	63	3,300	3,300,440

Source: IDC, 2005

Agilité

Il ne faut pas pour autant figer les processus au point de ne plus pouvoir répondre au changement

Les outils “2.0” ou de collaboration émergente intègrent cette dimension, tout comme le case management adaptatif

Meilleur service

En exploitant mieux l'information disponible dans l'entreprise, et en ayant des processus de réponse mieux contrôlés, on améliore la qualité des réponses aux clients (ex: propositions commerciales, support, etc.)

In what seems a growing fog of data, CEOs have never expressed a greater need to obliterate their blind spots. Too often, they say, information based on customer interactions is trapped in organizational silos.

Organizations that are able to combine, or layer, many kinds of information from different customer channels—with frequency—are best positioned to succeed.



IBM

Source: Capitalizing on Complexity - Insights from the Global CEO Study, IBM, 2010

Maîtrises des risques

L'utilisation de l'ECM permet de contrôler les processus documentaires dans l'entreprise

Donc d'en mieux maîtriser les éventuels dérapages (retard dans des réponses, pertes de documents, conservation de documents compromettant!)

Contraintes légales

Toutes les organisation sont soumises à des contraintes légales (ex: conservation des documents comptables) ou réglementaires (ex: pharma)

“Legal discovery”

A contrario, conserver certains documents trop longtemps peut constituer un risque

Statut juridique du document électronique

Aux termes de l'article 1316-I du code civil, la reconnaissance juridique de l'écrit ne dépend ni de sa forme (format), ni du support utilisé, ni des modalités de transmission. Cette disposition s'applique également pour la validité des écrits électroniques en droit civil.

Pour que l'écrit original soit requis à titre de preuve ou à titre de validité il devra, notamment, se conformer à la double exigence fonctionnelle dont dispose l'article 1316-I du code civil, à savoir : l'imputabilité de l'acte et la garantie de son intégrité.

Gouvernance de l'information

En entreprise, l'information est souvent isolée dans des “silos” (un par département, ou un par application informatique)

Grâce à une approche globale de la gestion de ces actifs numériques, l'ECM doit permettre une meilleure circulation de l'information pertinente et donc une meilleure valorisation de celle-ci

Gestion des connaissances

L'ECM peut servir de support à une démarche de *gestion des connaissances*: percevoir, identifier, analyser, organiser, mémoriser, et partager des connaissances entre les membres de l'organisation, en particulier les savoirs créés par l'entreprise elle-même (ex: marketing, R&D) ou acquis de l'extérieur (ex: intelligence économique) en vue d'atteindre l'objectif fixé.

Intelligence collective

L'**intelligence collective** désigne les capacités **cognitives** d'une **communauté** résultant des **interactions** multiples entre ses membres (ou **agents**). Les éléments portés à la **connaissance** des membres de la communauté font qu'ils ne possèdent qu'une **perception partielle de l'environnement** et n'ont pas **conscience** de la totalité des éléments qui influencent le groupe. Des agents au comportement très simple peuvent ainsi accomplir des tâches apparemment très **complexes** grâce à un mécanisme fondamental appelé **synergie**. Sous certaines conditions particulières, la synergie créée par la **collaboration** fait **émerger** des facultés de **représentation**, de création et d'apprentissage supérieures à celles des individus isolés. L'étude de l'intelligence collective implique aussi l'étude des limites des interactions entre membres d'un groupe, limites qui conduisent à des erreurs collectives parfois catastrophiques.

L'ECM comme discipline fédératrice

Capture / Démat (*Digital Imaging*)



IntelliScan® XDS

Document Management (GED)

Gestion des documents de référence de
l'entreprise

Archivage

GED Collaborative

Gestion de la production documentaire de l'entreprise

De la bureautique simple...

...à la coordination de documents très complexes (mythique documentation d'un Airbus)

Web Content Management (WCM)

Gestion de sites web (besoin qui explose fin des années 90)

Notion de front office / back office

Plus récemment, WEM (Web Engagement Management)

Digital Asset Management (DAM)

En français: photothèque / vidéothèque /
multimédiathèque

Marché: producteurs de contenu multimédia
(ex: agence d'images), services marketing
(catalogue d'images de produits)

Enterprise Search

Techniques d'indexation appliquées aux contenus d'entreprise (“google pour l'intranet”)

Challenges:

Multiplicité des sources

Hétérogénéité des formats

Droits d'accès

Search-based application

From Wikipedia, the free encyclopedia

Search-based applications (SBA) are [software applications](#) in which a [search engine platform](#) is used as the core infrastructure for information access and reporting. SBAs use [semantic technologies](#) to aggregate, normalize and classify [unstructured](#), [semi-structured](#) and/or [structured content](#) across multiple repositories, and employ [natural language technologies](#) for accessing the aggregated information.

Record Management (RM)

En français: Système d'Archivage Electronique

Archives courantes: entre la création du document et la fin de son usage courant

Archives intermédiaires: entre la fin de l'usage courant et l'échéance légale de conservation

Archives définitives: conservation définitive des documents à valeur historique, après l'application d'une règle de sort final, et la destruction des documents sans valeur historique

Case Management

Gestion de dossiers complexes impliquant plusieurs documents et plusieurs acteurs

Ex: gestion des dossiers de sinistres (assurance), gestion de courrier pour les administrations, gestion de cas cliniques ou sociaux

Collaboration émergente

Wiki d'entreprises

Réseaux sociaux d'entreprises

CEVAs

CEVA = Content Enabled Vertical Applications

Applications *métiers* centrées sur le contenu

Les grands types de documents

Quelques tentatives de classifications

Persuasif

Transactionnel

Métier

Officiel ou legal

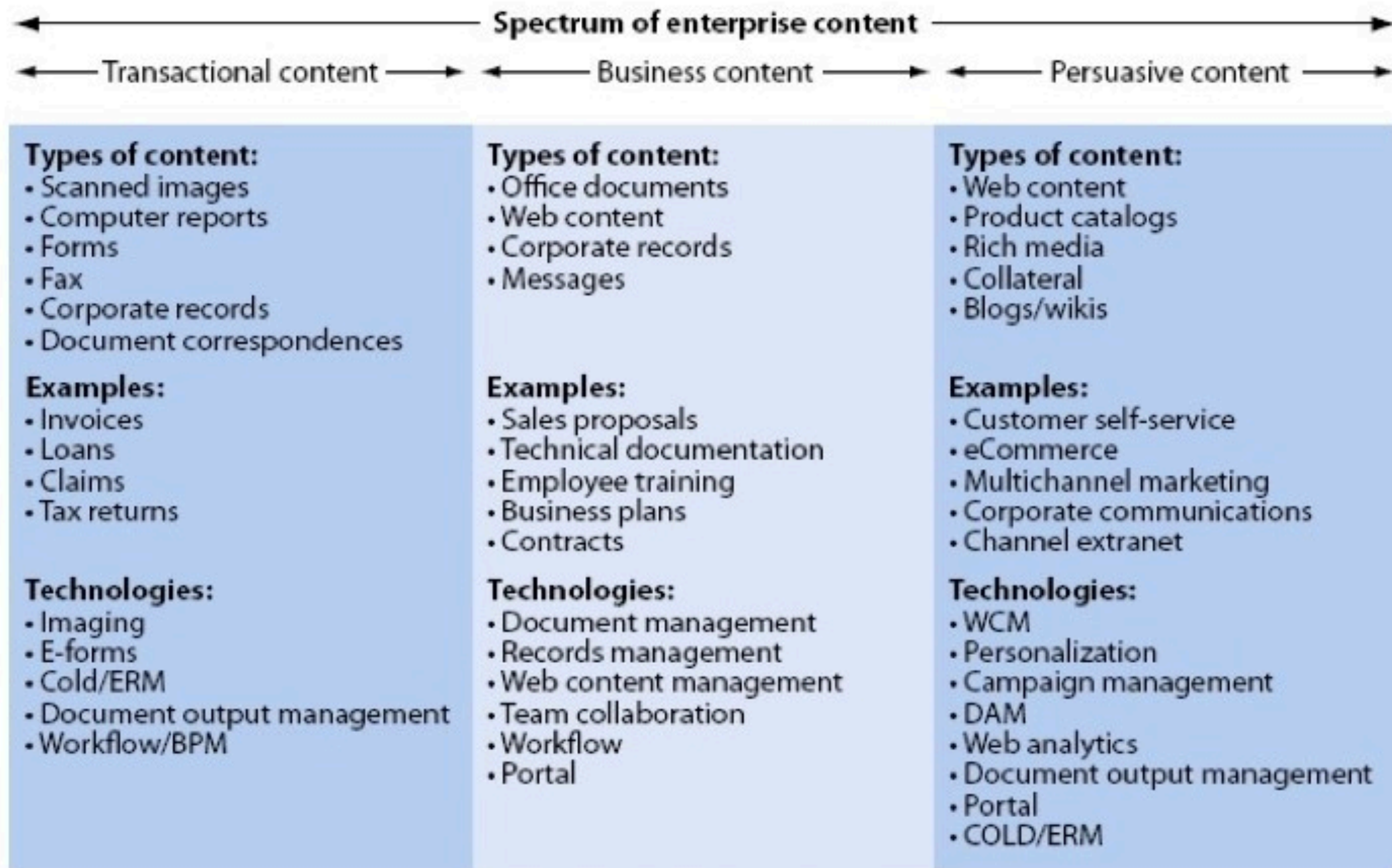
Contractuel

De savoir

De gestion interne

Source: Forrester Research (2005)

Source: Aproged (2005)



Source: Forrester Research (2005)

	DOCUMENT DE NATURE OFFICIELLE OU LÉGALE	DOCUMENT DE NATURE CONTRACTUELLE (DE PREUVE)	DOCUMENTS DE SAVOIR ET PATRIMONIAL « CAPITALISATION DES CONNAISSANCES »	DOCUMENT DE NATURE « GESTION INTERNE À L'ORGANISATION »
Exemples (*)	Cerfa (DC4,...) Déclaration IRPP Carte d'identité, Passeport Permis de conduire Acte authentique Relevé cadastral Facture Contrat de travail Contrat commercial Constat d'assurance Statuts Brevet Norme Déclaration annuelle des revenus AMM, Autorisation de mise sur le marché d'un médicament	Contractuelle (Convention de preuve – notion de négociation) Relevé bancaire Contrat de travail Certificat Code juridique Règlement intérieur Contrat commercial Nature de preuve Rapport d'audit Feuille de paie Bon de commande Bon de livraison Bilan d'entreprise	Livre/dictionnaire/ code Création musicale Création artistique Manuel d'entretien ou de réparation Guide d'utilisation Périodique Rapport de mission Revue Catalogue de VPC	Note de service, Plan de formation Rapport de mission Document « hygiène et sécurité » Manuel Qualité Annuaire téléphonique Livret d'accueil Tableaux de bord Journaux Grand livre et balance comptable

Le cycle de vie du document

Quatres grandes étapes

Création

Gestion

Diffusion

Conservation



1. CRÉATION

La « brique » création est constituée de quatre étapes : création à proprement parler du document, enregistrement, classement et indexation.

CREATION DU DOCUMENT

La création peut résulter d'un traitement automatisé (ex. production de factures) ou nécessiter l'intervention d'un agent (ex. rédaction de courrier, numérisation de document).

Recommandation : identifier la source déclarée pour créer le document ou le modèle à utiliser.

ENREGISTREMENT

L'enregistrement peut répondre à deux besoins complémentaires : la mémorisation du document pour des utilisations ultérieures (consultation, diffusion, archivage) ou le déclenchement d'un cycle de revue et validation.

Recommandation : utiliser le format d'enregistrement qui aura été déterminé pour ce document.

CLASSEMENT

Le classement consiste à ranger le document dans un espace informatique accessible aux utilisateurs prévus. L'organisation du rangement doit être compatible avec le plan de classement. Le classement peut être réalisé automatiquement en s'appuyant sur les métadonnées du document ou nécessiter l'intervention d'un agent.

INDEXATION

L'indexation consiste à déterminer des termes ou expressions qui serviront de clés d'accès pour retrouver le document dans le système de classement. Le choix des expressions peut résulter d'un traitement informatique analysant le contenu du document (ex. indexation du texte intégral) ou d'un processus humain consistant à choisir les concepts utiles du document.



2. GESTION

La « brique » gestion recouvre tout ce qui se passe sur le document après sa création. Quatre mots-clés la caractérisent : sécurité et droits d'accès, administration, variation et évolution.

SECURITE ET DROITS D'ACCES

Il s'agit de la capacité à rendre le document accessible aux personnes autorisées. Cela passe par des opérations de chiffrement ou de cryptage, de restriction d'actions sur le contenu ou d'occultation de champs.

Recommandation : bien prendre en compte la confidentialité des informations et documents. Ne crypter ou chiffrer que ce qui est absolument nécessaire.

ADMINISTRATION

Il s'agit du suivi des opérations de gestion concernant le document numérique (attribution, localisation, ...) qui peuvent se traduire par la modification des métadonnées du document liées au système de gestion.

VARIATION

Il s'agit de modifications du document numérique, sans incidence sur son contenu informationnel qui peuvent nécessiter le suivi et l'enregistrement de la traçabilité, du changement d'état ou de statut.

EVOLUTION

Il s'agit de modifications du document numérique avec incidence sur son contenu informationnel ; c'est-à-dire la gestion et le suivi du niveau de fidélité au contenu informationnel, des versions et des états intermédiaires.



3. DIFFUSION

La diffusion du document est un processus de restitution. Elle se fait après extraction depuis les bases de stockage ou de conservation selon deux modes principaux : une mise à disposition (ou mode pull : exécution après connexion et accès en ligne aux bases) ; une distribution (ou mode push : transfert systématique des documents à des destinataires préalablement identifiés).

MODES D'ACCES

Les modes d'accès sont définis et décrits dans des annuaires, réglementant les accès aux documents ou la capacité à organiser leurs transferts.

CONDITIONNEMENT

Il s'agit de la mise en forme du document pour l'adapter à son mode de diffusion, notamment : le choix du taux de compression, la maîtrise des conversions, du cryptage et du formatage. Ces choix doivent être pérennes et permettre la meilleure lisibilité pour les utilisateurs, en répondant à des normes ou à des référentiels adaptés.

Recommandation : le choix d'un format de diffusion doit permettre la lisibilité du document par le destinataire, ce qui conduit à éliminer ou à restreindre certains formats propriétaires.

TRANSMISSION

Opération obligatoire en mode push. La diffusion englobe ici la notion de réversibilité, c'est-à-dire la restitution en vue d'une réintégration dans un autre système de conservation.



4. CONSERVATION

La « brique » conservation a pour objet de maintenir dans le temps la lisibilité du document numérique et de ses composants (attributs de présentation, métadonnées, contenu). La conservation a d'importantes finalités : prouver, mémoriser, comprendre, communiquer.

Recommandations : c'est généralement la partie faible des systèmes documentaires. Conserver et sauvegarder sur plusieurs supports, détenus dans des endroits différents. La conservation et ses modalités doivent être définies lors de la conception du système. La mise en oeuvre des opérations de cette brique fonctionnelle doit conduire à des processus qui reposent sur des outils techniques qui garantissent : pérennité, intégrité, sécurité, traçabilité, lisibilité. Prendre en compte les durées de conservation déterminées.

Les grandes fonctions de l'ECM



Capturer

Numérisation / dématérialisation de documents entrants (conversion analogique / digitale)

Production de contenu depuis des applications bureautique

Production de contenu depuis des applications métiers (formulaire)

Traîter

Compression

OCR (Optical Character Recognition)

RAD (Reconnaissance automatique de documents)

LAD (Lecture Automatique de documents)

Extraction de métadonnées

Analyse sémantique

Stocker

Le stockage sur disque optique n'est pratiquement plus utilisé

Utilisations de serveurs ou de baies de stockage (NAS et SAN)






Début d'utilisation du stockage dans le cloud

Exercice





Combien de documents peut-on stocker dans
1 petaoctets de disques ?

Combien ca coûte ?

Combien de temps pour remplir l'espace ?
(ex: Banque française, AFP, Facebook,
Wikipedia)

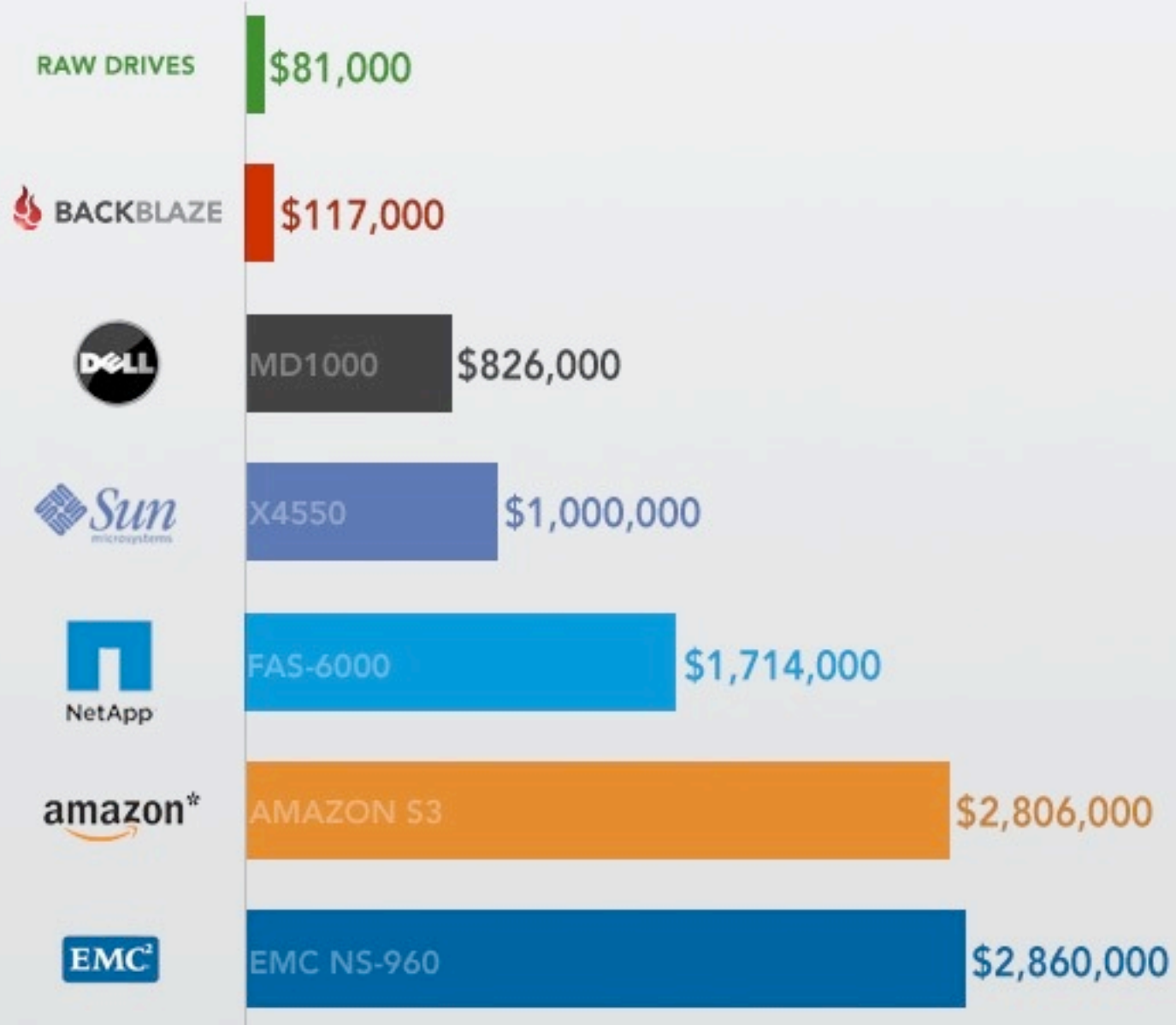
	DÉSIGNATION	NOTE	DISPO.	PRIX
HITACHI				
	Hitachi Deskstar 7K3000 3 To Hitachi Deskstar 7K3000 3 To 7200 RPM 64 Mo Serial ATA 6 Gb/s (boîte) - Voir note* + DÉTAILS		EN STOCK	156€⁷⁷
	Hitachi Deskstar 7K3000 3 To SATA 6Gb/s Hitachi Deskstar 7K3000 - 3 To 7200 RPM 64 Mo Serial ATA 6Gb/s (bulk) - Voir note* <i>Le stockage de masse rapide, fiable et efficace</i> + DÉTAILS	 3 avis	EN STOCK	156€⁷⁷
SEAGATE TECHNOLOGY				
	Seagate Barracuda XT 3 To SATA 6Gb/s Seagate Barracuda XT 3 To SATA 6Gb/s - 3 To 7200 RPM 64 Mo Serial ATA 6 Gb/s (bulk) + DÉTAILS	PCAchat 16/20	EN STOCK	182€⁰⁶
	Seagate Constellation ES 3 To 7200 RPM 64 Mo SAS 2.0 6Gb/s (bulk) + DÉTAILS		+ DE 15 JOURS	353€⁹⁹

Source: Idlc.com

DÉSIGNATION		NOTE	DISPO.	PRIX	
SEAGATE TECHNOLOGY					
	Seagate Cheetah 15K.7 SAS 6 Gbits/s 600 Go Seagate Cheetah 15K.7 - 600 Go 15000 RPM 16 Mo SAS 2.0 6Gb/s (bulk) + DÉTAILS		DISPO Fournisseur	374€ ²⁹	
	Seagate Cheetah NS.2 SAS 6 Gbits/s 600 Go Seagate Cheetah NS.2 - 600 Go 10000 RPM 16 Mo SAS 2.0 6Gb/s (bulk) + DÉTAILS		EN STOCK	546€ ²⁸	

Source: Idlc.com

COST OF A PETABYTE



* Amazon S3 Storage over three years (minus electricity, co-location and administration).

Source: backblaze.com (2009)

Cost of a Petabyte over 3 years



Hardware
Expense



Hardware plus
space and power

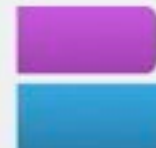


BACKBLAZE



\$56,696

\$94,563



\$378,500

\$502,000



amazon
web services™
S3 with RRS



\$2,466,360

Use cases

Banque française

Agence France-Presse (10000 dépêches /
jour, 5000 images / jour)

Facebook (700 millions de photos / mois fin
2008)

Wikipedia

Banque française

3.3 milliards de chèques / an en 2009

~500 millions pour la plus grande

100 ko par scan

-> 50 téraoctets / an (?)

Wikipedia

20 millions d'articles

~3 ko / article en moyenne

40 edits / article

15 millions de binaires (images, sons, videos)

Double tous les deux ans

Total (naïf): 1 To

Wikipedia

Cf. [http://en.wikipedia.org/wiki/
Wikipedia:Database_download](http://en.wikipedia.org/wiki/Wikipedia:Database_download)

Dump XML: 7 Go compressés, 31 Go
décompressé

Avec l'historique: 5 To (et en plus cassé!)

Classer / Cataloguer

Typologie documentaire

Organisation hiérarchique (plan de classement / taxonomie)

Métadonnées

Vocabulaire contrôlés et thesauri vs. folksonomies (*free tagging*)

Réseaux de relations (ontologies)

Métadonnées

“Données à propos de donnée”

Peut être disjointe du document lui même (ex: notice bibliographique) ou embarquées dans le document (ex: fonction proposées par les outils bureautiques, mais très peu utilisée en pratique)

Indispensable pour une bonne utilisation des documents

Droits et confidentialité

Notion de *principal*: personne ou groupe; de *rôle* (“manager”, “utilisateur”, “invité”...); et de *droit* (“lecture”, “écriture”...)

Connexion avec un SGI (ex: annuaire LDAP)

Granularité variable: niveau d'un dossier, niveau d'un document, niveau d'un champ d'un document

Indexer / Rechercher

Extraction du texte à partir de format divers

Recherches “plein texte” ou par métadonnées
(ou mixte des deux)

Recherches enregistrées, recherche et
navigation par facettes, par similarité

Difficulté technique: la gestion des droits

Recherche par facettes

Toutes nos boutiques

Rechercher Computer Software content management GO

Livres anglais et étrangers

Recherche détaillée

Nos rubriques

Nouveautés

Meilleures ventes

Bonnes affaires

Boutique

Toutes les boutiques

Livres anglais et étrangers

Computers & Internet

Software

Business (239)

E-mail (6)

Natural Language Processing (30)

Personal Finance (19)

Mathematical & Statistical (8)

Word Processors & Editors (14)

Format

Relié (76)

Poche et Broché (303)

Auteur

Greg Harvey (3)

Andy Harris (1)

Carole Matthews (1)

Ellen Finkelstein (1)

Erik T. Ray (1)

Jean-Claude Martin (1)


Nancy Muir (1)

En voir d'autres...

Livres anglais et étrangers > Computers & Internet > Software > "con

Résultats 1 - 12 sur 385

1.



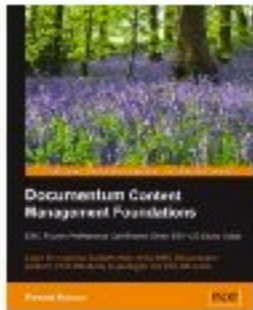
Enterprise Content Management: Using Microsoft Search Information de Ben Robb et Zlatan Dziri

Acheter neuf: EUR 25,97

À paraître - Commandez-le dès aujourd'hui! Cet article se

Livraison gratuite possible (voir fiche produit).

2.



Documentum Content Management Foundation Certification Exam E20-120 Study Guide de F

Acheter neuf: EUR 42,16


9 neufs à partir de EUR 41,78 3 d'occasion à partir de

Recevez votre article le mardi 4 octobre, si vous commandez aujourd'hui.

Plus que 1 ex. Commandez vite !

Livraison gratuite possible (voir fiche produit).

3.



Release & Control for IT Service Management ItSMF - The IT Service Management Forum (Bro

Acheter neuf: EUR 51,37

4 neufs à partir de EUR 51,37 3 d'occasion à partir de

Travail collaboratif (sur les documents)

Notion d'espaces de travail

Versionning

Check-in / Check-out

Transactions

Commentaires

Edition synchrone

“Timeline” des utilisateurs et des espaces

Workflow

Processus de validation de documents

Ex. simple: brouillon -> en attente de validation -> publié -> obsolète

Peut être beaucoup plus complexe dans les organisation très hiérarchiques

Mais besoin aussi de souplesse pour ne pas se retrouver bloqués

Publication

Mise à disposition des documents de référence dans un espace de consultation (intranet, site Web)

Transformation selon des feuilles de style (ex: XSLT)

Agrégation (ex: catalogue à partir de notices individuelles)

Impression personnalisée en masse (éditique)

Réutilisation

Ad hoc

Contenu existant indexé en vue d'une
réutilisation

Institutionnalisée

Modèles de documents

Documents composites, édition structurée

Gestion des archives

Intelligibilité (capacité d'un document à être aisément lu et compris)

Imputabilité ou authentification de l'auteur

Pérennité (capacité à garantir la permanence du support et des données, et leur stabilité dans la durée)

Gestion des archives

Intégrité (capacité à garantir qu'un document n'a pas subi d'altérations ou de modifications)

Sécurité (capacité à protéger un document et son contenu, à gérer ses accès et sa diffusion)

Traçabilité (capacité à suivre et à enregistrer les modifications apportées à un document)

Signature et coffre-fort électronique

Moyens techniques d'implémenter la non-répudiation, et la pérennité des documents archivés

Souvent fournis par des tiers de confiance

Repository ou base documentaire

Middleware entre les applications et le système de stockage physique

Service de persistance

Abstraction plus proche du domaine métier

Outil technique qui simplifie la création d'applications métiers orientées contenu (CEVAs)

Exercice final

Quelle différence en “un ECM” et un système de fichiers? Un SGBDR?

Quel effort pour implémenter un ECM au-dessus d'un FS, d'un SGBDR? Quelles sont les autres options?



Merci!