

Content Life Cycle: un modello di riferimento.

Vittorio Viganò¹

Premessa

Con questo contributo si vogliono approfondire alcuni concetti chiave relativi all'Enterprise Content Management (ECM) e più in particolare gli aspetti peculiari di ciò che si può, a tutti gli effetti, chiamare il ciclo di vita di un documento all'interno di un'organizzazione. Documento è un termine utilizzato non nella sua accezione tipica, quale quella di supporto cartaceo, bensì per rimandare ad un concetto più ampio che è quello di content. Le informazioni sono sicuramente contenute in un documento, inteso come mero raccoglitore di segni o di dati, ma assumono un valore più alto se si prende in considerazione il contesto entro il quale queste sono inserite e talvolta annegate. Le conoscenze, le competenze e le persone sono quindi le chiavi di volta per gestire l'informazione lungo tutto il suo ciclo di vita. Prima di procedere è opportuno meglio definire cosa si intende per organizzazione in questo contesto. Pertanto il termine organizzazione verrà utilizzato con l'accezione non solo di assetto organizzativo di un istituto e quindi componente di un'azienda ma anche di azione organizzativa, ossia specifica attività o funzione tesa a definire l'assetto organizzativo di un'azienda. Secondo Airoldi, Brunetti e Coda (1984), l'assetto organizzativo è la risultante del combinarsi della struttura organizzativa - modalità di distribuzione tra i vari organi aziendali, in insieme coordinato, dei compiti e delle responsabilità - e dei sistemi operativi - meccanismi che governano la dinamica e la remunerazione dei prestatori di lavoro e l'assegnazione ai vari organi aziendali degli obiettivi e delle risorse. E' proprio infatti in un ambiente così complesso ed articolato che il ciclo di vita di un documento si cala e si interseca sia con altri processi, anche ma non solo di natura transazionale, sia con le persone che danno vita all'organizzazione ed ai processi stessi.

Enterprise Content Management

Poiché diverse sono le definizioni di Enterprise Content Management che si trovano in letteratura, in questo scritto ci si rifarà a quella che riporta l'AIIM, un'organizzazione non-profit finalizzata ad assistere individui o aziende, nel comprendere le sfide associate al Document, Record e Content Management nonché, più in generale, ai processi di business. Ad oggi AIIM rappresenta, in maniera del tutto indipendente e su scala mondiale, tutta l'industria del settore, comprendendo quindi utenti, fornitori e canali di vendita. Nata in America nel lontano 1943 come "*National Microfilm Association*", diventata successivamente "*Association for Information and Image Management*", da cui peraltro prende l'acronimo, muta definitivamente il proprio nome in "*The Enterprise Content Management Association*" pur mantenendo la sigla AIIM. Già semplicemente osservando il cambiamento del nome dell'associazione stessa si può intuire la continua evoluzione nonché l'importanza dei concetti sottesi all'enterprise content management. Dopo questa parentesi di carattere puramente storico e di presentazione, facciamo un piccolo passo indietro e, riallanciandoci a quale definizione di ECM sia opportuno dare ed utilizzare, riportiamo quella elaborata appunto dall'AIIM:

¹ Nota depositata, Bocconi School of Management, 2007

“Enterprise Content Management is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists”.

Tale definizione è riportata in inglese per non perdere il significato dei termini usati che peraltro andremo a commentare puntualmente. Prima però di passare in rassegna i vocaboli usati in quella che vuole essere l'analisi grammaticale del periodo, è opportuno fare l'analisi logica. Rimasta tale fino al 2005, la definizione appena riportata ha subito nell'inverno del 2006 una modifica sostanziale. In quell'anno infatti è stato aggiunto il secondo periodo col quale ci si vuole espressamente riferire alle discipline del Document Management e del Record Management, divenute ormai di enorme importanza come meglio avremo modo in seguito di commentare. Tale frase vuole inoltre riferirsi in particolare a tutte le problematiche relative alla conversione in digitale del content ed in generale alla gestione dello stesso (Content Management).

Sul piano dell'analisi grammaticale la prima parola chiave è *technologies*, il termine che per primo viene utilizzato nella definizione di Enterprise Content Management. Essa attiene sicuramente alla tecnologia in senso stretto e quindi agli strumenti e ai metodi propri di un sistema ECM ma anche e soprattutto all'abilità di governare i processi di business. Strumenti e metodi si riferiscono quindi alla parte più prettamente informatica che comprende sia il lato di infrastruttura sia quello relativo agli applicativi.

In questo contesto, l'uso della tecnologia è riferito al ciclo di vita delle informazioni che, per essere gestite, devono scorrere attraverso un processo le cui fasi sono: *capture, manage, store, preserve e deliver*. Utilizziamo una metafora, quella dell'azienda manifatturiera, per meglio comprendere questi cinque passi. Immaginando quindi di dover produrre un bene qualsiasi, *capture* si riferisce all'ingresso in azienda delle materie prime, di fatto un documento può provenire da diverse fonti e con formati differenti. *Manage* è il processo produttivo durante il quale le materie prime, siano esse parti elementari per l'azienda manifatturiera o documenti, vengono trasformate, elaborate, utilizzate e poi, alla fine della fase operativa, archiviate con un'ottica di breve-medio periodo in magazzino (*store*) ovvero con un'ottica di lungo periodo (*preserve*). La logistica di distribuzione, che comprende la fase di recupero dal magazzino e di destino dei lavorati ai clienti - ovvero ai destinatari dell'informazione nel caso documentale - è quella che la definizione indica come *deliver*.

E' importante sottolineare che per astrarsi dal supporto cartaceo al quale un documento istintivamente rimanda, l'AIIM inserisce il termine *content* che può essere definito come un fenomeno aggregato, non oggettivo e reinterpretabile. Esso si pone in maniera diametralmente opposta al dato che è un elemento atomico e quindi per definizione un fenomeno puntuale, oggettivo e non interpretabile. Un esempio può risultare chiarificatore in questo senso. Il numero di determinati beni a magazzino è un dato non interpretabile in quanto si riferisce ad un'informazione puntuale: quantità X del bene Y. Chiunque legga questo dato vince la stessa informazione che risulta essere univoca. La valorizzazione contabile del magazzino, invece, utilizza questo dato come elemento imm modificabile e lo riaggrega secondo una logica specifica e pertanto soggettiva - supponiamo L.I.F.O. - dando quindi ad esso una particolare interpretazione che risulta diversa da quella che se ne avrebbe se si fosse utilizzata una metodologia differente - ad esempio F.I.F.O. Il content può dunque essere visto sia come un supporto per i segni (dati) sia la chiave di lettura del significato che dipende dal contenuto entro il quale i segni sono inseriti.

Un'altra parola chiave è *process* che si riferisce al flusso organizzativo lungo il quale scorrono le informazioni e dentro il quale un dato si arricchisce di significato, utilizzando magari proprio un documento. Ricorrendo sempre alla metafora dell'azienda manifatturiera, questo termine riconduce al processo attraverso il quale una materia prima in ingresso subisce diverse fasi di lavorazione necessarie per la produzione del bene finito. Riagganciandosi, a questo punto, al tema della tecnologia e dell'abilità di governare i processi di business - intesa come la capacità di coordinare un insieme di attività *“appartenenti a professionalità differenti, legate tra di loro da flussi significativi di informazione, la cui combinazione permette di ottenere un output rilevante per l'impresa nella sua interezza”* (A. Dossi, 2001) - possiamo affermare che tali processi siano elementi strettamente legati al buon funzionamento economico dell'azienda. Parafrasando dunque un vecchio adagio secondo cui la

tecnologia è utile solo se al servizio del business e dei processi che si articolano nel suo insieme, risulta evidente la necessità di ottimizzare tali processi che in un'impresa moderna sono molteplici e di sovente caratterizzati da un'elevata componente di tipo documentale.

Per concludere l'analisi grammaticale risulta fondamentale dissertare su *unstructured information*. Secondo uno studio condotto da Merryll Lynch nel 2003 più dell'85% delle informazioni relative al business sono presenti in azienda sottoforma di dati non strutturati, si pensi, ad esempio, alle e-mail, alle note, agli appunti dei call center, alle notizie, alle ricerche, alle chat, ai blog, alle presentazioni, alle pagine web, etc. In senso aggregato possiamo dunque affermare che le informazioni non strutturate appartengono ad un mondo che è diametralmente opposto a quello dei database che si poggiano su una logica molto strutturata. Basti pensare, ad esempio, alle informazioni contenute nelle tabelle di un database relazionale. Ogni singola tabella è frutto di un'analisi condotta con una logica metodica che vede nella rappresentazione del Diagramma E-R (entity-relationship diagram) la sua coerenza generale. Le informazioni destrutturate non hanno un layout e un form predefiniti e tantomeno una frequenza temporale preordinata, soprattutto se si pensa ad un possibile utilizzo in un processo. Come esempio per chiarire i concetti sopra esposti utilizziamo un documento amministrativo di uso comune: la fattura. Se in particolare ci soffermiamo sul ciclo passivo, tali documenti relativi ai fornitori arrivano potenzialmente tutti diversi nel formato. Ogni fattura ha una sua specifica gabbia grafica (layout) e una sua specifica disposizione dei dati (form); è infatti discrezionale il punto in cui inserire, all'interno del documento, il dato relativo alla quantità e alla descrizione dei beni forniti, all'importo parziale e totale della fattura, alla ragione sociale e alla partita iva dell'azienda fornitrice. Una fattura potrebbe inoltre arrivare in amministrazione prima del documento che attesti l'ingresso merci e quindi interrompere il processo di registrazione (frequenza temporale). L'impossibilità di sapere con anticipo la tempistica di arrivo dei documenti impatta negativamente su due processi differenti ma fortemente intersecati. Anche quest'ultimo può essere considerato a tutti gli effetti un elemento di non strutturazione. Immaginando ora di ordinare su di un asse il grado di strutturazione di una informazione avremmo su un lato il dato - il segno elementare – e su quello opposto il contesto – il significato che il dato prende in relazione appunto al contorno. Graficamente:



Figura 1: Il grado di strutturazione dell'informazione.

Prima di passare alla descrizione del ciclo di vita di un documento, con particolare riferimento all'accezione di content, è utile razionalizzare i tipi di materie prime che possono entrare nella nostra ideale azienda manifatturiera. Cataloghiamo quindi con l'aiuto della figura 2 quali possono essere gli input di un sistema di Enterprise Content Management. Tali input sono di fatto le sorgenti dell'informazione.

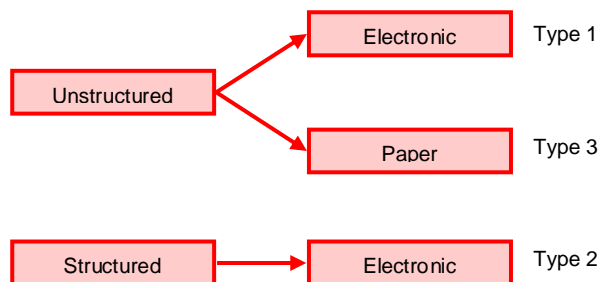


Figura 2: Le sorgenti dell'informazione.

Il Ciclo di Vita del Documento

La prima fase del ciclo di vita del documento è quella relativa al *capture* (figura 3), che si può a sua volta suddividere in due sottofasi che racchiudono le operazioni relative ad un momento di preparazione (*set up*) ed a un altro di organizzazione (*organize*) delle informazioni.

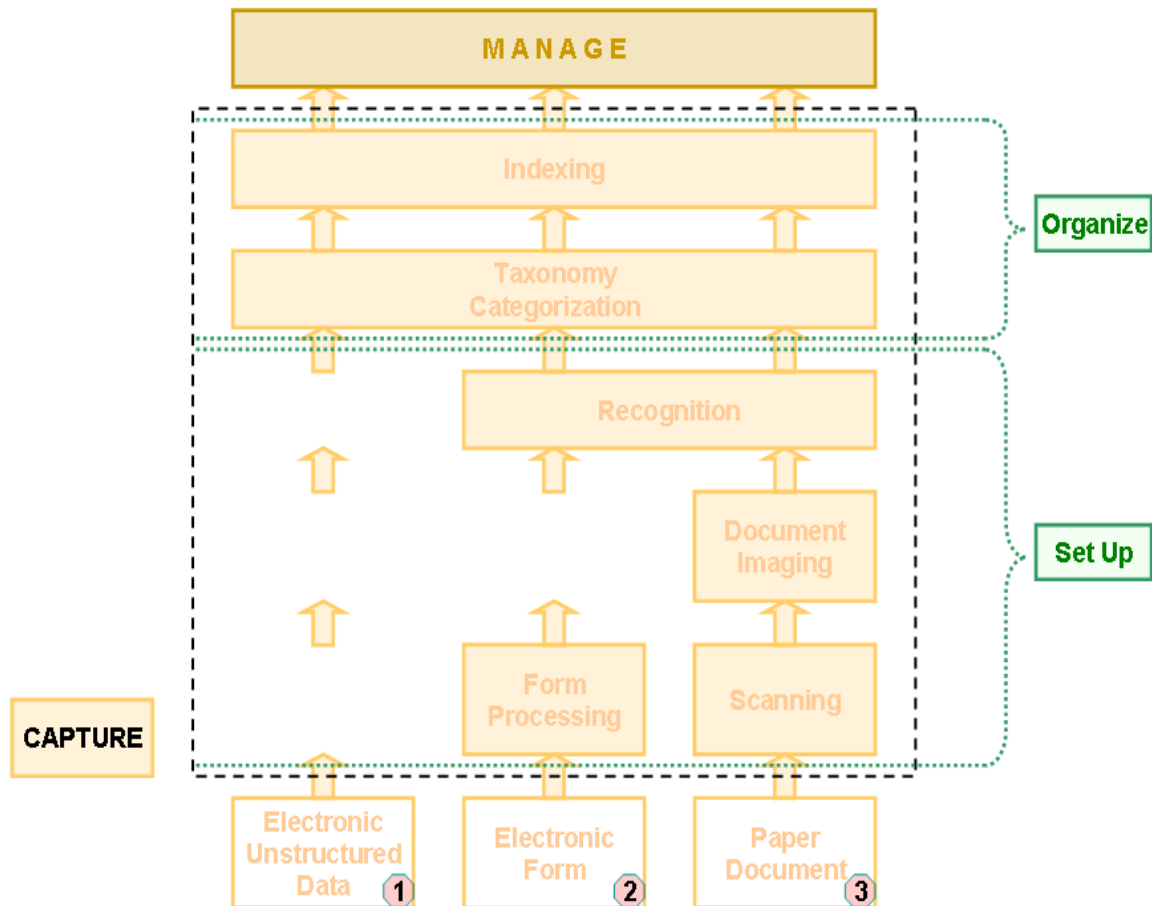


Figura 3: La fase di Capture.

A prescindere dal tipo di documento o di contenuto che entra in un sistema di ECM, esso, dopo una preparazione specifica che lo trasforma in formato digitale, accede alla fase di *Categorization*, relativa alle logiche, che possono essere applicate sia a strumenti manuali che automatici, di collocazione all'interno della tassonomia preordinata di tutti i content elettronici. *Taxonomy* è dunque una struttura formale dell'informazione basata sulle specifiche esigenze di business. Dopo aver categorizzato ed in seguito ordinato in uno schema predefinito il content è necessaria una fase di indicizzazione (*indexing*) che attiene alla creazione dei metadati necessari per la ricerca. Tale momento è una parte essenziale e non opzionale del processo di capture. L'indicizzazione, che si può basare sia su Parole Chiave – keyword – sia su Chiavi Testuali Complete – full-text – può essere sia manuale, quando è l'utente che crea la chiave per quel documento, sia automatica, quando è un software che gestisce in autonomia questo aspetto. Si pensi, in questo senso, all'utilizzo del bar code sui documenti. E' importante sottolineare come tali operazioni influiscano in modo determinante sull'efficienza e sull'efficacia delle successive fasi di store e retrieval. Per concludere la presentazione di questo primo stadio del ciclo di vita del documento, facciamo osservare che solo i documenti di tipo 2 (structured electronic) e di tipo 3 (unstructured paper) accedono alla fase di *recognition* prima di essere organizzati. Tale fase è l'insieme degli strumenti software che permettono di trasformare in modo automatico l'informazione cartacea in dati elettronici, esempi di ciò sono i sistemi di OCR (optical

character recognition) che fanno il riconoscimento dei caratteri “machine-printed”, i sistemi di ICR (intelligent character recognition) che fanno il riconoscimento dei caratteri “hand-printed” o i sistemi di OMR (optical mark recognition) che fanno il riconoscimento dei caratteri quali “check box”, “filled-in bubble”, etc.

Ora che le informazioni sono state sistematizzate, il content può essere gestito con strumenti relativi a diversi domini di competenza quali il Document Management (DM), il Record Management (RM), il Digital Asset Management (DAM), l'E-Mail Management (EMM) ed il Web Content Management (WCM). In figura 4 si può meglio comprendere da dove arrivino gli input della fase di *manage* e dove confluiscono gli output prodotti.

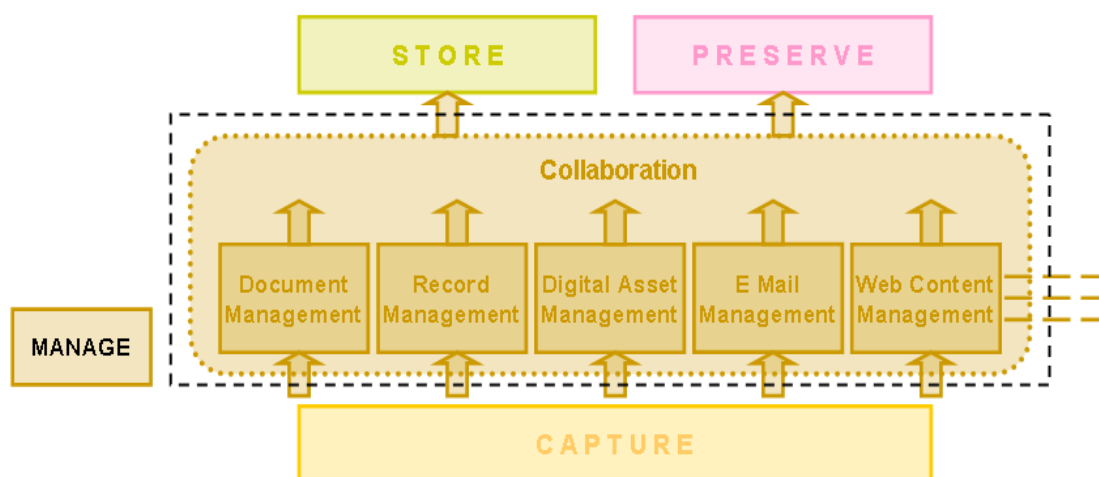


Figura 4: La fase di Manage.

Non potendo purtroppo fare all'interno di questo articolo una trattazione esaustiva circa le discipline di management del content, riporteremo un elenco di definizioni che siano il più possibile evocative dell'ampio territorio alle quale afferiscono.

Document Management: insieme di strumenti per la creazione, la revisione, l'approvazione e l'utilizzo del documento elettronico. Alcune caratteristiche peculiari sono, ad esempio:

- Retrieval: tecnica automatica di ricerca e di recupero delle informazioni contenute in archivi elettronici, indipendentemente dalla loro organizzazione logica-sistematica;
- Versioning: meccanismo automatico che tiene traccia di tutte le modifiche di un content condiviso ed utilizzato da più persone. Si ricordano a tal proposito le tecniche di check-in e di check-out;
- Document Security: aspetto relativo non solo alla sicurezza puramente informatica - diritti di accesso e backup - ma anche di conformità alla normativa – compliance;
- Workflow Engine: motore che gestisce il flusso del documento.

Record Management: insieme di strumenti per la gestione di tutti i content relativi ai business value aziendali che riguardano non solo valori monetizzabili ma anche valori intangibili – i cosiddetti intangible asset - quali strategie, tattiche e principi ispiratori. I criteri di archiviazione dei record sono relativi ad un'ottica di lungo periodo e sono conformi sia alle regole interne all'azienda sia alla normativa vigente. Poiché il termine record, molto comune nella lingua inglese, è impiegato in diversi contesti e con differenti significati risulta utile darne una definizione in riferimento ai sistemi di record management. In questo ambito useremo quella data dallo Standard Internazionale ISO 15489-2001: “un record è l'informazione creata, ricevuta e conservata come prova o dato sensibile da una persona o da un'organizzazione, nel perseguire obblighi legali o in una transazione di business”. Usando un linguaggio meno formale potremmo definire un record come una cellula di informazione aziendale. Il RM stabilisce le pratiche e le direttive da seguire per assicurare che ogni record creato od usato in una

organizzazione nel corso delle normali attività di business sia conservato in modo appropriato. Troppo spesso si dimentica che i record e le informazioni che si usano a scopo professionale sono un patrimonio del nostro business da salvaguardare come risorsa di grande valore. Ed è proprio per questo che entra in gioco il RM per assicurare che nei riguardi del patrimonio informativo si prendano le stesse necessarie precauzioni adottate per il patrimonio dei beni tangibili come, ad esempio, il denaro.

E-mail Management: insieme di strumenti per la gestione dei messaggi di posta elettronica che permette la classificazione, il salvataggio e la distruzione degli stessi secondo le regole di business utilizzando alcune delle logiche utilizzate sia per il Document sia per il Record management.

Digital Asset Management: insieme di strumenti che permettono di digitalizzare, archiviare, tracciare (log & tracking) e distribuire grossi volumi di “rich media” attraverso canali diversi quali immagini, audio e video.

Web Content Management: insieme di strumenti e tecnologie per la creazione, la revisione, l’approvazione e la pubblicazione di content “web-based”. L’utilizzo in questo ambito di strumenti quali i CSS (Cascading Style Sheets) permette infatti di tener separato il content vero e proprio dalla rappresentazione dello stesso o dalla sua presentazione sulle pagine di un portale aziendale.

Sebbene idealmente separati, questi sistemi di gestione dei vari tipi di content convivono all’interno di un ambiente o, per meglio dire, all’interno di un frame work condiviso che mette a disposizione strumenti il cui utilizzo è trasversale agli stessi. Basti pensare, ad esempio, alle strutture dati sottostanti ed ai motori di retrieval che sono sicuramente necessari a tutti i sistemi di management. Un’altra caratteristica peculiare di questo frame work è quella di racchiudere un insieme di tecnologie che permettono ai diversi utenti coinvolti, secondo il loro profilo, di condividere l’ambiente virtuale di lavoro e mantenere il team di progetto. Questo aspetto va spesso sotto il nome di *collaboration environment*. Gli strumenti di questo ambiente che risultano essere fondamentali ed essenziali per facilitare la creazione di “team-based content”.

Le fasi di *store* e *preserve* (figura 5) possono, di fatto, essere presentate assieme poiché sono caratterizzate da aspetti comuni sebbene la prima consideri un orizzonte temporale di breve-medio periodo e la seconda un orizzonte temporale di lungo periodo. Si ricorda in tal senso la metafora dell’azienda manifatturiera.

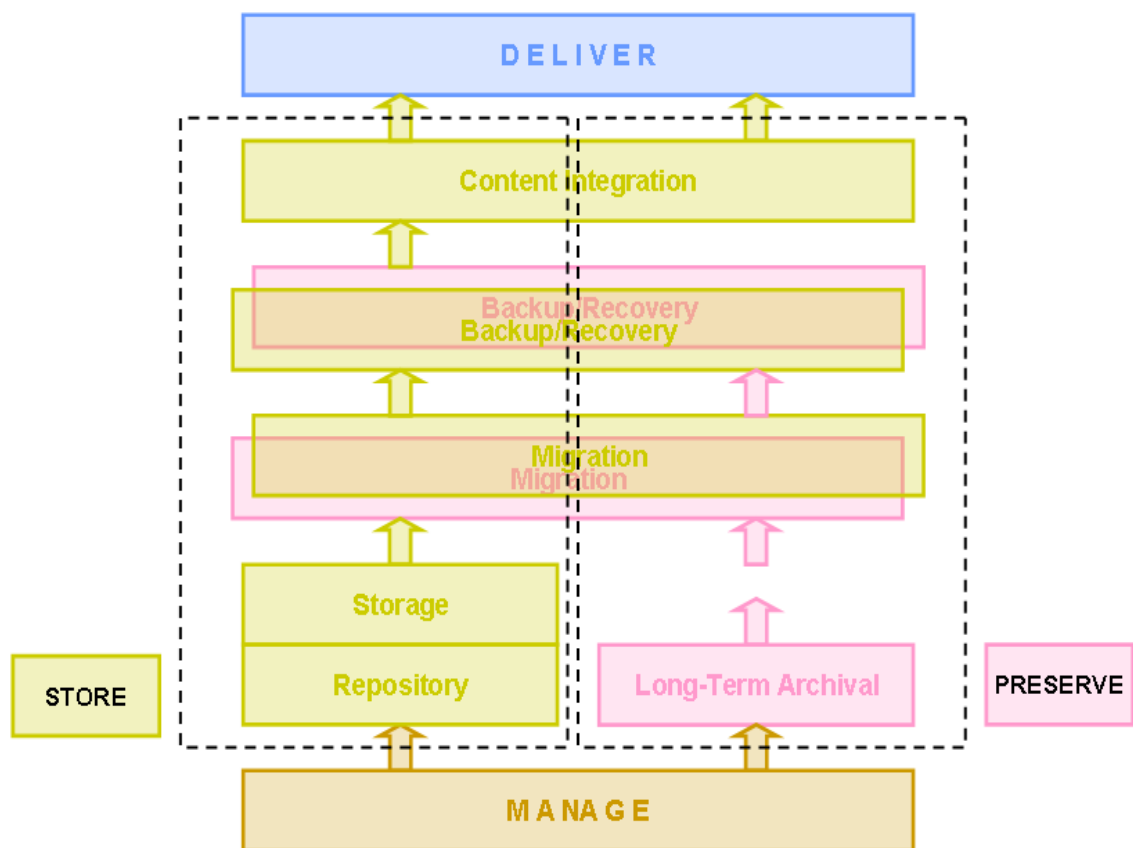


Figura 4: Le fasi di Store e di Preserve.

Repository e Storage sono visti come un unico elemento poiché attengono alla struttura logica ed all'ambiente fisico relativi all'archiviazione dei dati. Se il repository è il magazzino del content, lo storage è allora il magazzino dei repository. Backup/Recovery e Migration sono invece metodi e tecnologie che garantiscono la durabilità del dato. La parte di content integration richiama invece il livello di integrazione logica tra i possibili diversi repository di content. E' utile in questo caso pensare alla prospettiva dell'utente che vede ed utilizza un unico archivio dati, sebbene ce ne siano o ce ne possano essere differenti.

Per concludere il nostro percorso di descrizione del ciclo di vita del documento, ci manca da esaminare quella che abbiamo definito, nella metafora dell'azienda manifatturiera, la logistica di distribuzione (figura 5).

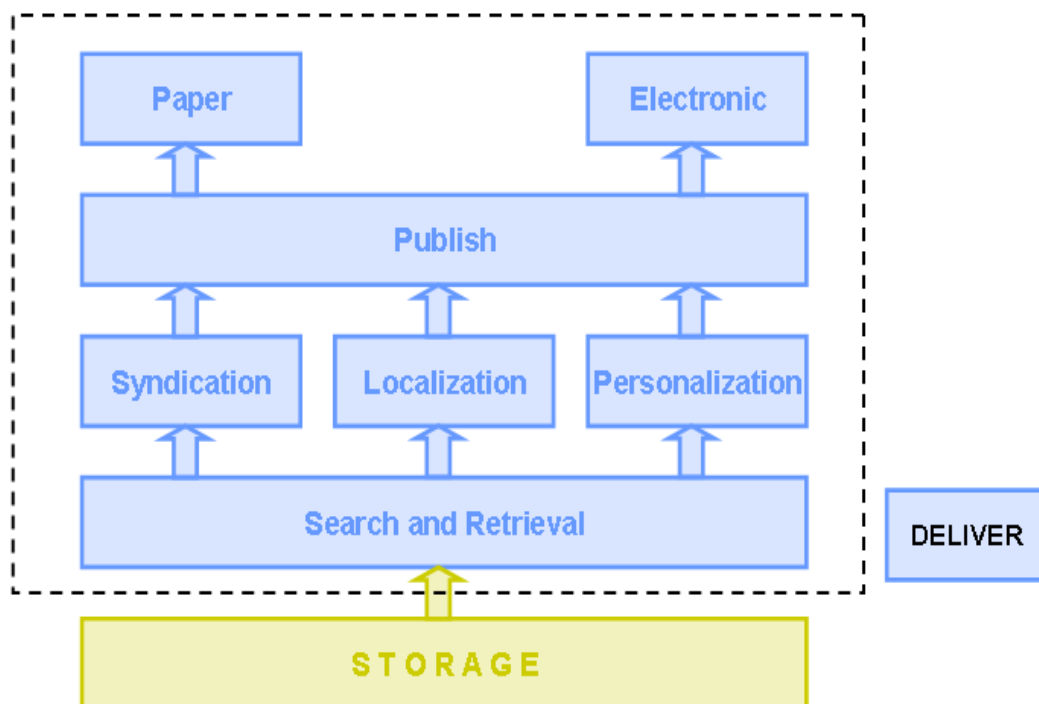


Figura 5: La fase di Deliver.

Il cuore della fase di *delivery* sono le operazioni di search e di retrieval: la prima riguarda gli strumenti per la ricerca delle informazioni archiviate nei diversi repository mentre la seconda attiene alle metodologie per l'estrazione di informazioni mediante tecniche algoritmiche sofisticate. Il text mining è un esempio di come possano essere estratte informazioni apparentemente nascoste utilizzando metodi di tipo statistico. Attraverso poi specifiche tecnologie di trasformazione del content, come ad esempio quelle di personalization, lo stesso content è pubblicato nelle forme più svariate. Tali forme possono essere elettroniche - e-mail, portali, intranet, internet, mobile device - ovvero tradizionali come il semplice foglio di carta.

La descrizione fin qui fatta di un generico flusso documentale è solo un aspetto, quello di natura più prettamente tecnologica, del più ampio tema del Knowledge Management (KM). Poiché sono molte le accezioni con le quale si può guardare al KM, noi ci riferiremo a quella che indaga tutti i processi di trasformazione del dato in informazione, dell'informazione in conoscenza e infine della conoscenza in saggezza. Il termine saggezza può essere poco comprensibile se riferito all'impresa: con quali criteri e sotto quali aspetti potremmo definire che una azienda è saggia? Sarebbe possibile trovare degli indicatori misurabili di saggezza? Solo forse rimandando all'organizzazione aziendale nonché alle persone che le danno vita è possibile comprendere come questi due pilastri possano essere certamente definiti la saggezza dell'impresa. Il knowledge management, e nello specifico il document management, può dunque essere visto come la leva tecnologica che, solo grazie alle persone e a un modello organizzativo di tipo antropocentrico, può realmente innovare.

La chiave di lettura del KM/DM come leva tecnologica di innovazione è un aspetto strettamente correlato alla cultura dell'organizzazione. E' solo in grado l'avvedutezza dei board composti da manager illuminati, e quindi alle persone, che la gestione dell'informazione crea un nuovo vantaggio competitivo, un vero asset strategico. Poiché a volte risulta complesso calcolare a priori il ROI di information management è importante comprendere che la direzione degli investimenti si debba spostare sulla cultura che le persone portano in azienda (Learning Organization).