ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea in Informatica per il Management

La conservazione del documento informatico

Tesi di laurea in Diritto di Internet

Relatore:

Chiar.ma Prof.ssa

Giusella Dolores Finocchiaro

Presentata da:

Maria Francesca Spagnolo

Sessione III Anno accademico 2012/2013

Indice

Introduzione

Fino a qualche anno fa, la conversione di documenti cartacei in documenti digitali appariva come una vera e propria utopia. Quest'oggi invece pensiamo al digitale di qualsiasi tipologia come fosse un fattore prioritario, scontato ed immediato, da poterlo considerare una vera a propria dipendenza.

Mentre una volta i documenti venivano depositati in archivi talvolta rudi e polverosi, spiacevoli da recuperare, sia per tempistiche sia per condizioni, oggi la quotidianità è così frenetica che occorre ottenere tutto in modo tempestivo.

Un' evoluzione socio-culturale ha portato all'elaborazione di tecniche sempre più sofisticate di ottimizzazione nel campo; si parla di "dematerializzazione dei documenti" o "processo di conservazione sostitutiva", innovazione che ha determinato una vera e propria rivoluzione nella Pubblica Amministrazione, nei rapporti tra privati, nell'Ordinamento Giuridico e nella scienza del diritto, poiché è radicalmente mutata la concezione di "documento" così com'è conosciuto da migliaia di anni, nella sua natura *res signata*, "cosa " che riporta informazioni.

Il concetto di "smaterializzazione" non è un' assoluta novità se si pensa alle transazioni finanziarie di enormi quantità di denaro, che da anni avvengono in tutto il mondo, grazie a "documenti informatici", senza che passi di mano una sola banconota.

Lo scopo di questa tesi è approfondire il processo di conservazione sostitutiva, partendo dalle figure ad esso legate, per poi continuare con le regole tecniche di adozione, la validità probatoria dei documenti informatici, i formati, i metadati utilizzati, e così via.

Capitolo 1:

Il Documento Informatico

1.1- Definizione di CAD

Il Codice dell'Amministrazione digitale (CAD) consiste in un testo unico, nato in Italia del 1997 ed approvato nel 2005, che affronta disposizioni sul documento in formato digitale trattando diverse forme ed argomenti annessi.

E' il frutto di un lavoro collegiale di grande impegno che ha visto la diretta partecipazione di enti di grande importanza, tra cui il Cnr[1] e la commissione Uni-Diam[2], portando una generica revisione delle norme precedenti relative all'informatizzazione della pubblica amministrazione italiana.

Il CAD è articolato in ben nove aree di interesse dedicate a:

- Principi generali e definizioni;
- Documento informatico e firme elettroniche;
- Formazione, gestione e conservazione dei documenti informatici;
- Trasmissione dei documenti informatici;
- Trattamento dei dati in possesso della pubblica amministrazione e ai servizi in rete;
- Sviluppo e riuso dei sistemi informativi nelle pubbliche amministrazioni
- Regole tecniche;
- Sistema pubblico di connettività;
- Norme transitorie finali e abrogative.

[1] Consiglio nazionale delle ricerche
[2] Documentazione, Informazione automatica e multimediale

Ha il compito di riunire e riordinare diverse norme inerenti alla digitalizzazione dei documenti, applicandole sia in ambito di pubblica amministrazione che in ambito privato, semplificando la comunicazione tra le due parti, evitando informazioni ridondanti e superflue.
Nell' arco degli anni il CAD ha subito alcune modifiche poiché le norme sono state adattate al progredire del processo tecnologico. Ad esempio: mentre nel 1997 veniva favorita la firma digitale, ritenuta più sicura rispetto agli altri modelli, nel 1999 venne adottata la scelta di optare per la neutralità tecnologica, approccio alternativo e funzionale secondo il quale è possibile utilizzare diverse tecnologie che siano in grado di raggiungere e garantire obiettivi prefissati dalla norma.
Il principio di neutralità tecnologica, fornendo una vasta gamma di tecnologie utilizzabili, permette di scegliere liberamente fra le proposte di mercato in base a criteri personali: quali utilità, disponibilità economiche ecc..
Successivamente, nel 2002, l' IItalia provvede ad un aggiornamento del CAD inserendo leggi relative alla firma elettronica, pur non derogando le leggi relative alla firma digitale.

Il CAD ha lo scopo di regolare ed assicurare i seguenti aspetti dell' informazione digitale:

- disponibilità,
- gestione,
- accesso,
- trasmissione,
- conservazione,
- fruibilità.

Per rendere ciò possibile utilizza le tecnologie informatiche nel modo più appropriato, sia nei rapporti tra amministrazione e privati, sia all'interno della pubblica amministrazione, sia nell'uso di documenti informatici tra privati.

1.2 - Il documento informatico secondo il CAD

Quello di documento è un concetto molto amplio, considerato socialmente noto, poiché non è semplice darne una definizione unica ben precisa. Il documento consiste nella rappresentazione di un fatto, destinato alla conservazione, relativo ad una certa realtà costituita da parole, immagini, firme, fotografie, audio, video, e così via.

Il documento informatico è una tipologia di documento che il CAD definisce nel seguente modo:

> *Art. 1 c.1., q, CAD*
>
> *Documento informatico:* rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti.

Si parla quindi di documento informatico quando ci si riferisce a qualsiasi rappresentazione digitale di un fatto.
Anche in tal caso si tratta di un concetto molto amplio poiché prende in analisi numerosi campi, si considera infatti documento informatico una video ripresa, un'immagine, una canzone, una mail, un documento scritto a pc.
Si fa fronte ad un enorme salto culturale poiché è possibile rilevare all'istante tutti i dati relativi ad un determinato documento informatico, dalla paternità, alla data di creazione, alla data di modifica, ecc.
Anche i documenti informatici sono destinati alla conservazione ma, mentre avendo a che fare con documenti classici è semplice attuare un'archiviazione, per essi è richiesto il supporto e il rispetto di alcuni requisiti.

Una caratteristica molto importante relativa al documento informatico è la forma, analizzata dal CAD come riportato:

Art. 20, CAD

1. Il documento informatico da chiunque formato, la memorizzazione su supporto informatico e la trasmissione con strumenti telematici conformi alle regole tecniche di cui all'articolo 71 sono validi e rilevanti agli effetti di legge, ai sensi delle disposizioni del presente codice

1-bis. L'idoneità del documento informatico a soddisfare il requisito della forma scritta e il suo valore probatorio sono liberamente valutabili in giudizio, tenuto conto delle sue caratteristiche oggettive di qualità, sicurezza, integrità ed immodificabilità, fermo restando quanto disposto dall'articolo 21.

3. Le regole tecniche per la formazione, per la trasmissione, la conservazione, la copia, la duplicazione, la riproduzione e la validazione temporale dei documenti informatici, nonché quelle in materia di generazione, apposizione e verifica di qualsiasi tipo di firma elettronica avanzata, sono stabilite ai sensi dell'articolo 71. La data e l'ora di formazione del documento informatico sono opponibili ai terzi se apposte in conformità alle regole tecniche sulla validazione temporale.

4. Con le medesime regole tecniche sono definite le misure tecniche, organizzative e gestionali volte a garantire l'integrità, la disponibilità e la riservatezza delle informazioni contenute nel documento informatico.

5. Restano ferme le disposizioni di legge in materia di protezione dei dati personali.

5-bis. Gli obblighi di conservazione e di esibizione di documenti previsti dalla legislazione vigente si intendono soddisfatti a tutti gli effetti di

> legge a mezzo di documenti informatici, se le procedure utilizzate sono conformi alle regole tecniche dettate ai sensi dell'articolo 71.

1.3 – Valore probatorio del documento informatico

Il valore probatorio di un documento corrisponde alla forza che questo ha in giudizio, così considerato una prova, della quale occorre valutare l'attendibilità. Per via della sua natura tecnologica il documento informatico può essere accompagnato da una firma informatica, non sempre lo è; per far sì che abbia una efficacia probatoria però è opportuno che lo sia.
La firma elettronica è un metodo di archiviazione che consiste nell'insieme di dati in forma elettronica, allegati o connessi ad altri dati elettronici.
Per firma digitale invece si intende il risultato di una procedura informatica, che consente di manifestare e di verificare la provenienza ed integrità di un documento informatico, grazie ad un sistema di chiavi asimmetriche a coppia, di cui una pubblica e una privata, utilizzate rispettivamente dal destinatario e dal sottoscrittore.
Il documento informatico al quale è stata pervenuta firma elettronica è valutabile liberamente dal giudice, in base al caso, prendendo in analisi anche il contesto della firma; di conseguenza nel momento in cui lo si firma non è possibile sapere quale sarà la sua valenza probatoria.
Tale documento è una prova se è legalmente riconosciuta attraverso perizie calligrafiche e la parte contro il quale è stato prodotto lo riconosce.
L'efficacia probatoria dei documenti con firma digitale o firma elettronica qualificata è maggiore rispetto a quelli trattati pocanzi poiché è la controparte a possedere l'onere della prova nei confronti di chi vuol far valere il documento. Chi dovrebbe aver firmato il documento, non può semplicemente dire di non averlo firmato, deve anche dimostrarlo.

Ad esempio se un dispositivo con firma elettronica qualificata o firma digitale venisse utilizzato da altri ad insaputa del titolare, questo deve dimostrare spiegazioni concrete e veritiere sull'accaduto.
E' possibile accettare tali motivazioni in caso di furto ma non nel caso in cui terzi fossero a conoscenza delle credenziali da dimostrare avendo libero accesso al dispositivo, poiché è imposto dalla legge il vincolo di segretezza.
I documenti informatici che non sono soggetti ad alcun tipo di firma possono essere considerati prove nel caso in sui la parte contro la quale vengono prodotti ne conferma la conformità.

Documento informatico	**Efficacia probatoria**
Senza firma	Se non è disconosciuto fa piena prova dei fatti, se è disconosciuto la Corte di Cassazione sostiene che potrebbe costituire un elemento di prova invece che prova piena.
Con firma elettronica	E' liberamente valutabile dal giudice
Con firma elettronica qualificata o firma digitale	Ha efficacia probatoria finché non si disconosce la firma apposta, dando prova delle dichiarazioni contenute nel documento, se non è disconosciuto. Il disconoscimento deve avvenire con la prova di non utilizzo del dispositivo di firma da parte del titolare.
Con firma digitale autenticata	Ha efficacia probatoria nel caso in cui le firme siano state approvate da particolari enti. E' una prova piena, non può essere disconosciuta, può solo essere oggetto di querela di falso. Può inoltre essere usato come atto pubblico.

Il CAD prende in analisi tale aspetto molto importante nel seguente modo:

Art. 21, CAD

1. Il documento informatico, cui è apposta una firma elettronica, sul piano probatorio è liberamente valutabile in giudizio, tenuto conto delle sue caratteristiche oggettive di qualità , sicurezza, integrità e immodificabilità.
2. Il documento informatico, sottoscritto con firma elettronica avanzata, qualificata o digitale, formato nel rispetto delle regole tecniche di cui all'articolo 20, comma 3, che garantiscano l'identificabilità dell'autore, l'integrità e l'immodificabilità del documento, ha l'efficacia prevista dall'articolo 2702 del codice civile. L'utilizzo del dispositivo di firma ((elettronica qualificata o digitale)) si presume riconducibile al titolare, salvo che questi dia prova contraria.

2-bis. Salvo quanto previsto dall'articolo 25, le scritture private di cui all'articolo 1350, primo comma, numeri da 1 a 12, del codice civile, se fatte con documento informatico, sono sottoscritte, a pena di nullità, con firma elettronica qualificata o con firma digitale. Gli atti di cui all'articolo 1350, numero 13, del codice civile soddisfano comunque il requisito della forma scritta se sottoscritti con firma elettronica avanzata, qualificata o digitale.

3. L'apposizione ad un documento informatico di una firma digitale o di un altro tipo di firma elettronica qualificata basata su un certificato elettronico revocato, scaduto o sospeso equivale a mancata sottoscrizione. La revoca o la sospensione, comunque motivate, hanno effetto dal momento della pubblicazione, salvo che il revocante, o chi richiede la sospensione, non dimostri che essa era già a conoscenza di tutte le parti interessate.

4. Le disposizioni del presente articolo si applicano anche se la firma elettronica è basata su un certificato qualificato rilasciato da un certificatore stabilito in uno Stato non facente parte dell'Unione europea, quando ricorre una delle seguenti condizioni:
    a) il certificatore possiede i requisiti di cui alla direttiva 1999/93/CE del Parlamento europeo e del Consiglio, del 13 dicembre 1999, ed è accreditato in uno Stato membro;
    b) il certificato qualificato è garantito da un certificatore stabilito nella Unione europea, in possesso dei requisiti di cui alla medesima direttiva;
    c) il certificato qualificato, o il certificatore, è riconosciuto in forza di un accordo bilaterale o multilaterale tra l'Unione europea e Paesi terzi o organizzazioni internazionali.
5. Gli obblighi fiscali relativi ai documenti informatici ed alla loro riproduzione su diversi tipi di supporto sono assolti secondo le modalità definite con uno o più decreti del Ministro dell'economia e delle finanze, sentito il Ministro delegato per l'innovazione e le tecnologie.

1.4 – I requisiti del documento informatico

Il fine primario del processo di conservazione digitale è quello di detenere inalterate nel tempo requisiti di stabilità, autenticità, leggibilità, accessibilità, riproducibilità di un contenuto digitale archiviato; per rendere ciò possibile è necessario disporre di operazioni di aggiornamento tecnologico quando necessario.

Partendo da un' analisi generale, prendiamo in considerazione il concetto di contenuto digitale, con il quale si intende un oggetto digitale in grado di

rappresentare una combinazione di dati, immagini, testi, video, audio, consistente in una sequenza binaria fissata su uno o più supporti di memorizzazione, nella quale i bit assumono un'organizzazione ed un significato specifico, delimitato seguendo un insieme di regole.

La produzione e la lettura di un contenuto digitale comportano la necessità di avere a disposizione un software in grado di formare la corrispondente sequenza binaria, memorizzarla e leggerla su un supporto, *media*, con l'appoggio del sistema di *storage management*, un insieme di componenti hardware e software volti a modificare le caratteristiche fisiche del supporto in base al valore dei bit da memorizzare.

L'efficacia dei software utilizzati per produrre, modificare o visualizzare un contenuto digitale dipende dall'architettura hardware che li ospita e dalla compatibilità che questi hanno con il sistema operativo.

Per documento informatico si intende un vero e proprio contenuto digitale rappresentativo di fatti, atti o dati giuridicamente rilevanti, proprio per tal motivo è necessario soddisfare i requisiti di cui è stato accennato all'inizio del capitolo. Analizziamo ora singolarmente tali requisiti:

- **Stabilità**, ovvero la capacità di conservare in modo inalterato nel tempo la rappresentazione del contenuto e della forma del documento, garantendo l'integrità della relativa sequenza binaria;
- **Autenticità**, ossia la possibilità di ricondurre con certezza giuridica un documento informatico originale al proprio autore, resa possibile grazie all' utilizzo di firme elettroniche;
- **Accessibilità e leggibilità**, fanno riferimento alla disponibilità di un insieme di strumenti tecnologici e metadati che permettono la ricerca dei documenti informatici in archivio e di renderli disponibili ai soggetti che hanno il diritto di accedervi, in modo comprensibile e senza particolari vincoli tecnologici. In particolare il requisito di accessibilità si occupa di valorizzare, archiviare e conservare l'insieme dei metadati descriventi il contenuto, il contesto e la struttura dei documenti;

- **Riproducibilità**, cioè la capacità di produrre copie, duplicati, o estratti di documenti informatici in archivio, su differenti tipi di supporto e garantendone la conformità agli originali, ciò rende possibile la il trasferimento di documenti digitali da un supporto all'altro senza che questi perdano la forza probatoria originaria.

L'Art.44 del CAD tratta i requisiti del documento informatico nel seguente modo:

Art. 44, CAD

1. Il sistema di conservazione dei documenti informatici assicura:
   a) l'identificazione certa del soggetto che ha formato il documento e dell'amministrazione o dell'area organizzativa omogenea di riferimento di cui all'articolo 50, comma 4, del decreto del Presidente della Repubblica 28 dicembre 2000, n. 445 ;
   b) l'integrità del documento;
   c) la leggibilità e l'agevole reperibilità dei documenti e delle informazioni identificative, inclusi i dati di registrazione e di classificazione originari;
   d) il rispetto delle misure di sicurezza previste dagli articoli da 31 a 36 del decreto legislativo 30 giugno 2003, n. 196 , e dal disciplinare tecnico pubblicato in allegato B a tale decreto.

1-bis. Il sistema di conservazione dei documenti informatici è gestito da un responsabile che opera d'intesa con il responsabile del trattamento dei dati personali di cui all'articolo 29 del decreto legislativo 30 giugno 2003, n. 196 , e, ove previsto, con il responsabile del servizio per la tenuta del protocollo informatico, della gestione dei flussi documentali e degli archivi di cui all'articolo 61 del decreto del Presidente della Repubblica 28 dicembre 2000, n. 445 , nella definizione e gestione delle attività di rispettiva competenza.

1-ter. Il responsabile della conservazione può chiedere la conservazione dei documenti informatici o la certificazione della conformità del relativo processo di conservazione a quanto stabilito dall' articolo 43 e dalle regole tecniche ivi previste, nonché dal comma 1 ad altri soggetti, pubblici o privati, che offrono idonee garanzie organizzative e tecnologiche.

Capitolo 2:

La conservazione

2.1 – Ruoli e responsabilità dei soggetti coinvolti

È possibile individuare almeno tre ruoli fondamentali e ben distinti del sistema di conservazione dei documenti, di seguito sarà approfondito ognuno di essi.

2.1.1 – Il produttore

Il ruolo di produttore del documento informatico è svolto da persone fisiche o giuridiche, enti o sistemi applicativi che possono far parte o non del sistema di conservazione. Questi forniscono all'archivio le informazioni da conservare, non è quindi necessario che il produttore delle informazioni sia il diretto autore, si può benissimo trattare di un intermediario o addirittura un altro archivio.
Egli ha la responsabilità di rappresentare nel modo più accurato possibile i dati di cui il documento tratta, renderlo quindi affidabile attribuendogli una certa completezza.
Altro compito importante del produttore è quello di garantire l'autenticità del documento, proprietà che lo accompagna per tutta la sua esistenza; è possibile stabilire se un documento è autentico sulla base della sua identità ed integrità.
L'identità di un documento è fondata da svariati attributi o caratteristiche che nell'insieme lo caratterizzano in maniera unica distinguendolo dagli altri.

Alcuni esempi di attributi includono: nomi delle persone che concorrono alla formazione del documento, la materia a cui questo si riferisce, le date di produzione e di trasmissione, l'espressione della sua relazione con gli altri documenti, la sua forma digitale e documentaria, il nome dell'ufficio competente, il riferimento ad allegati, l'esistenza di firma digitale.
Si può dire che un documento abbia integrità se questo fosse intatto e non corrotto, ovvero se l'intenzione del messaggio comunicato in principio per raggiungere il suo scopo rimanesse inalterata. Il documento può essere fisicamente modificato purché l'articolazione del contenuto e gli elementi formali necessari rimangano invariati. È possibile dimostrare l'integrità del documento da attributi relativi ad esso, espressi come metadati, argomento che sarà trattato più avanti.
Alcuni attributi di integrità del documento sono: il nome della persona competente per la pratica, nome della persona responsabile per il documento, l'esistenza di annotazioni, l'indicazione di cambiamenti tecnici, di firme digitali aggiunte o rimosse, data della rimozione pianificata dal sistema, data di trasferimento al custode designato, data di distruzione pianificata, esistenza e collocazione di duplicati.

2.1.2 – L'utente

Anche il ruolo di utente o cliente è svolto da una persona fisica o giuridica interna od esterna al sistema di conservazione, da un ente o da un sistema applicativo.
E' il destinatario delle informazioni conservate, deve quindi avere la possibilità di ricercarle e fruirne in maniera appropriata; ci si riferisce quindi ad una "comunità" in grado di capire tali informazioni.

A tal proposito è stato istituito dallo standard OAIS[3], il quale sarà trattato successivamente, il concetto di " knowledge base" ovvero l'insieme delle conoscenze e competenze necessarie per l'acquisizione e la comprensione delle informazioni da ricavare. Ad esempio se i documenti conservati in un archivio sono in lingua Russa, questi possono essere correttamente fruiti solo a chi conosce la lingua russa, in tal calo il "knowledge base" sarà la conoscenza di tale lingua.

Per acquisire i dati di suo interesse egli richiede l'accesso ai documenti al sistema di conservazione entro i limiti proposti dalla legge. L'interazione dell'utente con il sistema non ne modifica il contenuto o la forma del documento.

2.1.3 – Il responsabile della conservazione

Il responsabile della conservazione è il ruolo svolto da chi gestisce l'archivio dall'esterno controllandone la performance, finanziandolo e definendone le caratteristiche.
Stabilisce le politiche che il sistema di conservazione dovrà rispettare, mettendole in pratica e gestendole in autonomia e con responsabilità. Egli si cura dell'aggiornamento periodico del manuale di conservazione in presenza di cambiamenti normativi, organizzativi, procedurali o tecnologici rilevanti, definisce i requisiti e gli attributi del sistema di conservazione in base alla tipologia di documento da conservare, seguendo dettagliatamente le norme vigenti, gestisce la conservazione garantendone la conformità, crea e sottoscrive i pacchetti di distribuzione tramite firma digitale o elettronica qualificata.

[3] Open Archival Information System

È possibile richiedere la certificazione di conformità del processo di conservazione a soggetti privati o pubblici in grado di offrire adeguate garanzie tecnologiche ed organizzative.

Altri ruoli importanti del responsabile della conservazione sono ad esempio monitorare il sistema di conservazione garantendone il corretto funzionamento, adottando misure in grado di rilevare nell'immediato eventuali degradi dei sistemi di memorizzazione e di registrazioni, o talvolta di ripristinare il corretto funzionamento, assicurare periodicamente la verifica dell'integrità e la leggibilità degli archivi (massimo ogni cinque anni), provvedere alla copia o duplicazione dei documenti digitali in previsione del progresso tecnologico.

Nel caso in cui sia richiesto l'intervento di un pubblico ufficiale il responsabile della conservazione ne deve assicurare la presenza garantendogli assistenza e risorse adeguate per l'adempimento delle attività che deve svolgere, inoltre ha il compito di assicurare l'assistenza e le risorse necessarie per l'adempimento delle attività di vigilanza e verifica agli organismi competenti disposti dalla normativa vigente.

Egli è in grado di, sotto la propria responsabilità, delegare parte o l'intero svolgimento del processo di conservazione ad uno o più soggetti con adeguata competenza ed esperienza relative alle attività ad essi assegnate, a patto che suddetta delega sia formalizzata ed esplicitata. L'affidamento può essere assegnato a soggetti esterni, responsabili del trattamento dei dati, tramite contratto o convenzione di servizio prevedente obbligo di rispetto del manuale di conservazione improntato dal responsabile della conservazione affidata.

Il responsabile della conservazione entra inoltre in relazione con il responsabile del trattamento dei dati personali, il responsabile della sicurezza e con il responsabile dei sistemi informativi, oltre che con il responsabile della gestione documentale ovvero con il coordinatore della gestione documentale ove nominato, per quanto attiene alle pubbliche amministrazioni.

2.2 – Ciclo di vita di un documento

Il ciclo di vita di un documento informatico si può distinguere in quattro fasi fondamentali:

- La Formazione, un documento informatico può nascere in tal forma di per sé oppure può partire da un originale analogico di cui si limita ad esserne una "copia".

  Nel primo caso si parla ad esempio di registrazioni informatiche di informazioni risultanti da transazioni, da processi informatici o dalla presentazione telematica di dati attraverso moduli o formulari messi a disposizione.

  Nel secondo caso vi è una distinzione tra documenti analogici originali unici, e non unici.

  Il contenuto dei documenti originali "unici" non può derivare da documenti o scritture come cambiali, assegni, titoli dell'ordine, dei quali sia obbligatoria la tenuta.

  Si parla di documenti originali "non unici" in caso di ricevute fiscali, fatture, dichiarazioni fiscali o documenti di trasporto.

  Una copia si può ritenere un documento informatico a due condizioni:

  se l'originale cartaceo era non unico la sua conformità deve essere assicurata dal responsabile della conservazione mediante l'utilizzo della propria firma digitale e nel rispetto delle regole tecniche riportate nell'articolo 71 del CAD, se l'originale cartaceo era unico invece la conformità dell'originale deve essere autenticata da un notaio o da un altro ente pubblico ufficiale a ciò autorizzato con la dichiarazione allegata al documento informatico, osservando le regole tecniche stabilite nell'articolo 71 del CAD.

  Art 71, CAD

  1. Le regole tecniche previste nel presente codice sono dettate, con decreti del Presidente del Consiglio dei Ministri o del

> Ministro delegato per la pubblica amministrazione e l'innovazione, di concerto con i Ministri competenti, sentita la Conferenza unificata di cui all' articolo 8 del decreto legislativo 28 agosto 1997, n. 281, ed il Garante per la protezione dei dati personali nelle materie di competenza, previa acquisizione obbligatoria del parere tecnico di DigitPA.
>
> 1-ter. Le regole tecniche di cui al presente codice sono dettate in conformità (ai requisiti tecnici di accessibilità di cui all'articolo 11 della legge 9 gennaio 2004, n. 4,) alle discipline risultanti dal processo di standardizzazione tecnologica a livello internazionale ed alle normative dell'Unione europea.
>
> 2. Le regole tecniche vigenti nelle materie del presente codice restano in vigore fino all'adozione delle regole tecniche adottate ai sensi del presente articolo.

Vi sono tre modalità con cui un documento informatico, considerato sostitutivo dell'originale a tutti gli effetti, può essere creato come copia di un documento analogico digitale:

- Nel caso in cui si tratti di documento informatico copia di un documento cartaceo firmato digitalmente da parte di chi lo rilascia, che può essere un depositario pubblico autorizzato o un pubblico ufficiale;
- Nel caso in cui si tratti di una copia per immagine di un originale analogico su un supporto informatico, purché ne sia attestata la conformità da un notaio o un altro pubblico ufficiale;
- Nel caso in cui si tratti di una copia per immagine di un originale analogico su un supporto informatico, purché la conformità dell'originale non sia espressamente disonosciuta.

Le copie su supporto analogico di un documento informatico, sottoscritto con firma elettronica avanzata, qualificata o digitale, hanno uguale efficacia probatoria dell'originale da cui sono tratte se non sono

espressamente disconosciute e se la loro conformità all'originale in tutte le sue componenti è attestata da un pubblico ufficiale a ciò autorizzato.

- Il Trasferimento in conservazione, in linea con lo standard OASIS, si prevede che avvenga tramite apposito pacchetto di versamento il cui formato è definito nel manuale di conservazione.
  Per standard OASIS si intende un modello di riferimento per gli archivi, sia a livello informativo, sia funzionale, elaborato dal CCSDS[4] e successivamente adottato come standard internazionale.
  OASIS è l'acronimo di Open Archival Information System, lo standard è stato quindi sviluppato in modo aperto tramite la discussione su forum pubblici a cui potevano partecipare i soggetti interessati, per tal motivo la sua applicazione è possibile su un vasto numero di situazioni diverse fra loro. Si tratta di un modello generico, valido sia in campo di informazione digitale sia in campo di informazione di tipo fisico, con lo scopo di fornire una semplificazione definendo concetti, funzionalità e modelli da seguire.

- La Conservazione, estremamente impegnativa poiché deve essere assicurata al di la dell'esistenza fisica dell'ente che se ne prende cura poiché questo può variare nell'arco degli anni e poiché è necessario che l'archivio sperimenti cambiamenti tecnologici per rendere i documenti sempre disponibili. Gli archivi in cui sono conservati i documenti devono garantire la loro l'integrità, l'autenticità, l'usabilità e l'accessibilità "sorvegliandoli" nel tempo, è proprio per tal motivo che sono opportuni aggiornamenti continui sotto il profilo tecnologico.
  Nel caso in cui ciò non avvenga si correrebbe il rischio che i documenti informatici residenti nell'archivio non aggiornato vadano persi.
  Per conservazione sostitutiva si fa riferimento al rapporto tra il documento analogico e il documento informatico.
  Il processo di conservazione sostitutiva è costituito dalle fasi seguenti:

[4] Consultative committee for space data system

- Generazione dell' impronta (hash) dell'insieme dei documenti o dei singoli documenti costituenti l'insieme;
- Memorizzazione dell'impronta dell'insieme dei documenti nel file di chiusura;
- Apposizione della firma digitale al file di chiusura da parte del responsabile della conservazione ;
- Apposizione di un riferimento temporale associato al file di chiusura firmato;
- Memorizzazione dell'insieme di documenti, del file di chiusura firmato digitalmente e della marca temporale su un supporto con caratteristiche di alta affidabilità ed alta permanenza del dato.

  Quando si ha a che fare con una conservazione sostitutiva a lungo termine, come anticipato poche righe fa, un oggetto digitale è soggetto a due tipi di problemi: obsolescenza dei supporti ed obsolescenza dei formati.

  La legge, per conservare il valore probatorio degli oggetti in analisi, distingue due tipi di riversamento:
- Diretto, i documenti conservati sono trasferiti da un supporto ottico di memorizzazione a un altro, non alternando la loro rappresentazione informatica, realizzando una copia di sicurezza;
- Sostitutivo, i documenti conservati sono trasferiti da un supporto ottico di memorizzazione a un altro, modificando la loro rappresentazione informatica.

- L'eventuale Formazione di nuovi documenti derivati, o per esigenze di conservazione o per esifenze dell'utente.

2.3 – Misure di sicurezza adottate

Per realizzare un processo sicuro l'obiettivo principale è quello di domandarsi quanto questo debba essere sicuro ma soprattutto perché.

Occorre quindi attuare un' adeguata analisi dei benefici e dei costi da attuare, poiché la sicurezza ha un costo, ed attuare appropriate misure di sicurezza alle categorie di pericoli che consideriamo di rischio più alto.

Facendo ciò si stabiliscono i requisiti di sicurezza di un processo. Per definire dei buoni requisiti di sicurezza occorre che siano rispettate cinque proprietà fondamentali, essi devono essere:

- ✓ Necessari, è di fondamentale importanza su un requisito debba essere incluso o meno;
- ✓ Raggiungibili, è importante che un requisito sia al "passo coi tempi" ovvero compatibile con le risorse assegnate, aggiornato tecnologicamente parlando;
- ✓ Verificabili, la modalità di verifica di un requisito ne determina i criteri per l'accettabilità o meno dei risultati;
- ✓ Chiari, la forma nella quale sono espressi deve essere ben chiara e concisa;
- ✓ Tracciabili, vi sono diverse tipologie di requisito, da quelli di usabilità, a quelli di performance, a quelli di sistema, a quelli tecnologici. Tali requisiti costituiscono una vera e propria gerarchia, si prenda come esempio il fatto che un utente intenda conservare i propri dati in forma criptata, ciò comporta dei requisiti tecnologici, come l'adozione di un linguaggio di sviluppo o di un metodo di criptazione. Per tracciabilità si intende il poter determinare quali requisiti abbiano dato origine ad altri requisiti e perché.

Tra gli attributi che rientrano tipicamente tra i requisiti vi è la CIA:

- Confidenzialità: solo il legittimo proprietario dei dati o chi abbia le autorizzazioni necessarie può accedere ad essi

- Integrità: i dati possono essere modificati esclusivamente da chi ha una particolare autorizzazione per farlo.
- Disponibilità: chi possiede le autorizzazioni necessarie può avere accesso ai dati che lo riguardano nei modi e nei tempi richiesti.

Altri attributi di fondamentale importanza sono:

- Privacy, ovvero il diritto di qualunque individuo di specificare come quanto e quando comunicare informazioni relative al suo conto;
- Autenticazione, provare che l'utente che accede al sistema sia esattamente chi dichiara di essere seguendo almeno uno dei seguenti paradigmi: "cosa sai?", ad esempio la password, "cosa hai?", ad esempio un badge, "chi sei?", tali paradigmi talvolta possono entrare in conflitto con il concetto di privacy;
- Anonimato, cioè il diritto di cui favoriscono gli utenti di poter compiere un azione senza poter essere riconosciuti da terzi;
- Autorizzazione, stabilire un meccanismo grazie al quale dare agli utenti la possibilità di accedere o meno alle risorse;

Vi è una metodologia basata su 9 step con lo scopo di introdurre la sicurezza sin dalle prime fasi del processo di sviluppo di un documento informatico, è di SQUARE[5] che si parla, creata dal CERT[6] americano a fine 2005.

I 9 passi che hanno come finalità la creazione di questa lista di requisiti sono suddivisi per categorie e priorità:

1. Accordo sulle definizioni, stabilire un vocabolario ditermini comunemente utilizzati dai partecipanti all'attivita è il primo passo, tale vocabolario può essere estratto da normative, standard, terminologie di settore
2. Identificazione degli obiettivi di sicurezza,
3. Sviluppo della documentazione appropriata al supporto della definizione dei requisiti di sicurezza,

[5] security quality requirements engineering

[6] I CERT sono organizzazioni, finanziate generalmente da Università o Enti Governativi, incaricate di raccogliere le segnalazioni di incidenti informatici e potenziali vulnerabilità nei software che provengono dalla comunità degli utenti

4. Risk assessment,
5. Selezione delle tecniche di elicitazione,
6. Elicitazione dei requisiti di sicurezza,
7. Categorizzazione dei requisiti,
8. Dare una priorità ai requisiti
9. Ispezionare i requisiti.

Num	Step	Input	Tecniche	Partecipanti	Output
1	Accordo sulle definizioni	Definizioni candidate da IEEE e altri standard	Interviste strutturate e focus group	Stakeholders, requirements team	Lista delle definizioni su cui si è trovato accordo
2	Identificazione degli obiettivi di sicurezza	Definizioni, obiettivi candidati, leve di business, politicies e procedure, esempi	Sessioni di lavoro facilitato, questionari e interviste	Requirements engineer e stakeholders	Obiettivi
3	Sviluppo della documentazione appropriata al supporto della definizione dei requisiti di sicurezza	Manufatti potenziali	Sessioni di lavoro	Requirements engineer	Manufatti richiesti: scenari, misuse, cases, modelli, templtes, forms
4	Risk assessment	Misuse cases, scenari, obiettivi di sicurezza	Metodi di risk assessment, analisi dei rischi anticipati	Requirements engineer, esperto dei rischi, stakeholders	Risultati del risk assessment
5	Selezione delle tecniche di elicitazione	Obiettivi, definizioni, tecniche candidate, stili organizzativi, esperienza degli stakehlders, cultura, analisi costi-benefici	Sessioni di lavoro	Requirement engineer	Tecniche di elicitazione selezionate
6	Elicitazione dei requisiti di sicurezza	Manufatti, risultati del risk assessment,	JAD, interviste, questionari, analisi basate su modelli	Stakeholders facilitate dai Requirements engineer	Taglio iniziale dei requisiti di sicurezza

		tecniche selezionate			
7	Categorizzazione dei requisiti	Requisiti iniziali, architettura	Sessioni di lavoro usando un insieme standard di categorie	Requirements engineer e altri specialisti se richiesto	Requisiti categorizzati
8	Dare una priorità ai requisiti	Requisiti distribuiti in categorie e risultati del risk assessment	Metodi di prioririzzazione come MoSCoWtecniche di elicitazione selezionate	Stakeholders facilitati dai requirements engineer	Requisiti con priorità
9	Ispezionare i requisiti	Requisiti con priorità, tecnica di ispezione formale candidatura	Metodi di ispezione come Fagan MoSCoW	Team di ispezione	Requisiti iniziali selezionati, documentazioni dei processi decisionali e delle motivazioni

Capitolo 3:

Regole tecniche per la conservazione e la formazione del documento informatico

3.1 – I formati

Un documento informatico è costituito da una sequenza di bit in accordo con specifici formati, i quali specificano le regole sintattiche con cui si struttura un file, permettendo di leggerli, interpretarli e modificarli.
Se non si è a conoscenza del formato utilizzato per la creazione di un file è impossibile ricavarne informazioni e gestirlo, rendendolo così inutilizzabile ed indistinguibile, non è altro che una serie di bit a caso e senza senso.
L'associazione al formato da parte di un documento informatico fondamentalmente tramite tre modalità:

- Estensione, il nome del file unito ad una sequenza d lettere tramite un punto, ad esempio sappiamo che se si prende analisi il file "nomefile.docx ", si ha a che fare con un formato testo di proprietà di Microsoft;
- Metadati espliciti, dove si trova nei tipi MIME[7] l'indicazione "application/msword" si ha a che fare con un file realizzato con l'applicazione World di Microsoft;

[7] Multipart Internet Mail Extension

- Magic number, i primi byte contenuti nella sequenza di byte lo identificano, ad esempio la sequenza di byte 0xffd8 specifica che si tratta di un file immagine di tipo .jpeg.

In campo archivistico esistono vari standard in grado di generare identificatori univoci, alcuni sono:

- URN[8], il quale è utilizzato per identificare una risorsa;
- DOI[9], il quale detiene i compiti di gestione e controllo;
- PURL[10], il quale renderà disponibile il file anche nell' eventualità in cui questo fosse stato spostato nell' arco del tempo;
- ARK[11], in grado di identificare oggetti di qualunque tipologia, che si tratti di immagini, che si tratti di documenti digitali, siti web o software;
- Handle, identificatore costituito da una stringa di caratteri composta da un prefisso e da un suffisso.

Tali argomenti saranno approfonditi successivamente.

Lo sviluppo tecnologico sempre più "affamato di potere" ha condotto ad una disponibilità e complessità dell'informazione sempre più crescente, è quindi necessario avere a disposizione funzionalità sempre più mirate per la gestione delle svariate forme di informazione digitale, che si tratti di testo, di immagini, di filmati e così via.

Tali funzionalità permettono la creazione, la modifica e la manipolazione del documento in maniera più semplice.

La conseguenza di tale evento è ovviamente una netta crescita del numero dei formati a disposizione e dei rispettivi programmi utili alla relativa gestione.

Se si volesse fare una sintesi dei formati più diffusi al momento è possibile citare:

- DOC, HTML, PDF: per i file di testo e documenti;
- XLS: per i file di calcolo;
- BMP, SVG, JPG, GIF, EPS, TIF: per le immagini;

[8] uniform resurce name
[9] Uniform object identifier
[10] Persistent uniform resource locator
[11] Archivial resource key

- MP3, WAV: per i suoni;
- MPG, MPEG, AVI, WMV: per i video;
- EXE: per gli eseguibili;
- ZIP, RAR: per l'archiviazione e la compressione;
- SMTP, MIME: per le e-mail.

Tali formati consentono di memorizzare facilmente l'informazione digitale permettendo il riutilizzo, la modifica, l' elaborazione, assieme ai programmi che li gestiscono sono valutati prendendo in analisi alcune caratteristiche, tra le quali:

- La diffusione, la quale fa riferimento ad un'analisi del numero di soggetti che li adopera;
- La portabilità, ovvero il fedele utilizzo di standard documentati e accessibili;
- La funzionalità di cui l'utente dispone per elaborare l'informazione che gli interessa, permettendogli di collegarla ad altre;
- La capacità di gestire piu formati in base alle esigenze dell'utente interessato;
- La diffusione dei visualizzatori che portano all'utilizzo delle informazioni contenute nei formati indipendentemente dalla possibilità di effettuarne una rielaborazione;
- La potenzialità di occupare meno spazio possibile nella fase di memorizzazione;
- La capacità di gestire il maggior numero possibile di metadati.

La vasta gamma di scelta dei formati mette a disposizione dell' utente un gran numero di funzionalità disponibili.

Al fine di garantire la leggibilità, la reperibilità, e la modificabilità del documento, i vari formati possiedono alcune caratteristiche proprie e dei programmi che lo gestiscono, tra cui:

- Apertura, si ha a che fare con un formato "aperto" nel caso in cui questo sia disponibile a chiunque sia interessato al suo utilizzo, è conforme a specifiche pubbliche grazie alle quali è resa possibile la decodifica dei

documenti realizzati con determinate specifiche. In tal caso si ha a che fare con un formato documentato e pubblicato da un produttore o da un consorzio con lo scopo di promuoverne l'utilizzo;

- Sicurezza, la quale dipende dalla capacità di essere esonerata dall'inserimento di codice nocivo e dal grado di modificabilità del contenuto;
- Portabilità, ovvero quanto i formati possano essere "flessibili", adottati da diverse piattaforme, sia per quanto riguarda l'hardware sia per quanto riguarda il sistema operativo. La si induce dal fedele utilizzo di standard accessibili e documentati;
- Funzionalità, cioè la possibilità del formato di esser gestito da prodotti informatici prevedenti una vasta gamma di funzioni a disposizione dell'utente per effettuare la formazione e la gestione del documento informatico;
- Supporto allo sviluppo, si fa riferimento al sistema con cui si propongono le risorse utili alla manutenzione e allo sviluppo del formato e dei prodotti informatici che lo gestiscono, come società, comunità di sviluppatori, e così via;
- Diffusione, si intende quanto l' utilizzo di un formato per la formazione e gestione dei documenti informatici sia esteso, dato in grado di influire sulla probabilità che questo venga impiegato nell'arco del tempo, considerata la disponibilità dei prodotti informatici consoni alla sua gestione e visualizzazione.

Tali informazioni sono utili alla scelta del formato da adottare in base all' evenienza.

Vi sono altri dati che ne permettono la selezione, ad esempio quelli relativi alla possibilità di gestione del maggior numero possibile di metadati, comprese le modifiche del documento, e quelli relativi all' occupazione di spazio fisico che questi comportano all'interno del sistema.

Per la conservazione sono nettamente privilegiati i formati che siano standard internazionali oppure, dove se ne ha la necessità, formati proprietari con specifiche tecniche pubbliche.
Altro dato fondamentale per la scelta del formato a livello di conservazione è il dato relativo al tempo di conservazione che la normativa prevede per le singole tipologie di documenti informatici.
Nel manuale di conservazione sono riportati i formati utilizzati per la conservazione relativa alle diverse tipologie di documenti informatici, con una motivazione delle scelte eseguite.

3.2 - Standard e specifiche tecniche

Per quanto riguarda l'ambito della formazione, della gestione e della conservazione di documenti informatici e documenti amministrativi informatici sono utili alcuni standard e specifiche tecniche di riferimento, ne sono riportati alcuni in particolare:

- per la formazione e la gestione di documenti informatici:

UNI ISO 15489-1: 2006 Informazione e documentazione
- Gestione dei documenti di archivio -
Principi generali sul record management.

UNI ISO 15489-2: 2007 Informazione e documentazione
- Gestione dei documenti di archivio –
Linee Guida sul record management.

ISO/TS 23081-1:2006 Information and documentation
- Records management processes –
Metadata for records

– Part 1 – Principles, Quadro di riferimento per lo sviluppo di un sistema di metadati per la gestione documentale.

ISO/TS 23081-2:2007 Information and documentation
- Records management processes –
Metadata for records
– Part 2 – Conceptual and implementation issues, Guida pratica per l'implementazione.

ISO 15836:2003 Information and documentation
- The Dublin Core -
metadata element set, Sistema di metadata del Dublin Core.

- per la conservazione di documenti informatici:

ISO 14721:2002 *OAIS* (Open Archival Information System),
Sistema informativo aperto per l'archiviazione.

ISO/IEC 27001:2005, Information technology - Security techniques - Information security management systems – Requirements, Requisiti di un ISMS (Information Security Management System).

ETSI TS 101 533-1 V1.1.1 (2011-05) Technical Specification, Electronic Signatures and Infrastructures (ESI);
Information Preservation Systems Security;
Part 1: Requirements for Implementation and Management,
Requisiti per realizzare e gestire sistemi sicuri e affidabili per la conservazione elettronica delle informazioni.

ETSI TR 101 533-2 V1.1.1 (2011-05) Technical Report, Electronic Signatures and Infrastructures (ESI);

Information Preservation Systems Security;
Part 2: Guidelines for Assessors, Linee guida per valutare sistemi sicuri e affidabili per la conservazione elettronica delle informazioni.

UNI 11386:2010 S-Recupero degli Oggetti digitali.
Definisce la struttura dell'insieme di dati a supporto del processo di conservazione sostitutiva, individua gli elementi informativi necessari per la creazione dell'indice di conservazione descrivendone la semantica sia l'articolazione utilizzando il linguaggio formale XML. L'obiettivo è quello di consentire a chi opera nel settore l' utilizzo di una struttura-dati condivisa per giungere ad un buono grado d'interoperabilità nei processi di migrazione, grazie all' utilizzo dello Schema XML appositamente elaborato.

ISO 15836:2003 Information and documentation - The Dublin Core metadata element set, Sistema di metadata del Dublin Core.

3.3 – Specifiche tecniche del pacchetto di archiviazione

Dallo standard **UNI 11386:2010** precedentemente trattato (riguardante la conservazione e il recupero degli oggetti digitali, prende in analisi la struttura dell'insieme dei dati supportati dal sistema di conservazione) è possibile fare riferimento alla struttura descrittiva dell' Indice del Pacchetto di Archiviazione, indicato anche come indice di Conservazione.

Per Indice del Pacchetto di Archiviazione (IPdA) si intende la validità informatica attribuita ad ogni Pacchetto di Archiviazione (PdA), anche indicato come Volume di Conservazione, il quale contiene un insieme di informazioni e deve

essere accompagnato da una firma elettronica qualificata o dalla firma digitale di chi interviene nel processo di produzione del sistema di conservazione e da un riferimento temporale.

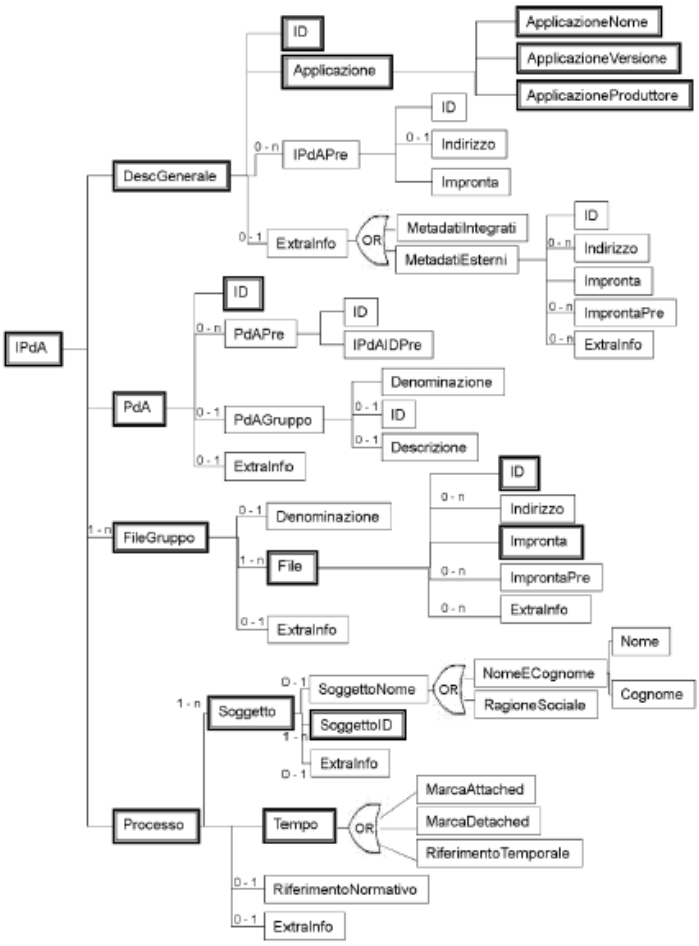
Si tratta di una struttura flessibile, la quale consente di ordinare in vari modi gli indici creandone di nuovi, unendo o dividendo le informazioni contenute negli IPdA precedenti o garantendone uno nuovo che si riferisca ad una sua precedente versione,(migrazioni causate da evoluzioni tecnologiche), quando necessario.

All' interno dell' Indice di Pacchetto di Archiviazione è possibile trovare tali strutture:

- informazioni generali del IPdA, ad esempio informazioni generali relative all'indice del pacchetto di archiviazione: un identificatore dell'IPdA, il riferimento all'applicazione che l'ha creato, eventuali riferimenti ad altri IPdA da cui deriva il presente, e un eventuale elemento "ExtraInfo" che consente di introdurre metadati soggettivi relativi all'IPdA liberamente definiti dall'utilizzatore con un proprio schema;
- informazioni inerenti il PdA, in particolare: un identificatore del PdA, eventuali riferimenti ad altri PdA da cui deriva il presente, informazioni relative a una eventuale tipologia/aggregazione (di natura logica o fisica) cui il PdA appartiene e infine un eventuale elemento "ExtraInfo" che consente di introdurre metadati soggettivi relativi al PdA;
- indicazione di uno o più raggruppamenti di uno file contenuti nel PdA, È possibile raggruppare file sulla base di criteri di ordine logico o tipologico ed assegnare ad ogni raggruppamento / singolo file le informazioni di base e un eventuale elemento "ExtraInfo" che consente di introdurre metadati definiti dall'utilizzatore. Ogni elemento file contiene l'impronta attuale dello stesso, ottenuta con l'applicazione di un algoritmo di hash e un'eventuale impronta precedentemente associata ad esso: in questo modo è possibile ad esempio gestire il passaggio da un algoritmo di hash diventato non più sicuro ad uno più robusto;

- informazioni relative al processo di produzione del PdA, l'indicazione del nome e del ruolo dei soggetti che intervengono nel processo di produzione del PdA (es. responsabile della conservazione, delegato, pubblico ufficiale ecc.), il riferimento temporale adottato (generico riferimento temporale o marca temporale), l'indicazione delle norme tecniche e giuridiche applicate per l'implementazione del processo di produzione del PdA ed, infine, anche per il processo, un elemento "ExtraInfo" che consente di aggiungere dati soggettivi relativi al processo.

Di seguito è riportata la rappresentazione grafica della struttura dell'Indice del Pacchetto di Archiviazione.



Laddove gli elementi sono racchiusi in  sono considerati obbligatori.

Analizziamo l' indice logico dello schema sopra riportato:

Nome Elemento	**Descrizione**	**Elemento Padre**	**Elementi Figli**
Applicazione	Informazioni sull' applicazione che ha generato l' IPdA	DescGenerale	ApplicazioneNome, ApplicazioneProduttore, ApplicazioneVersione
ApplicazioneNome	Nome dell'applicazione che ha generato l'IPdA.	Applicazione	
ApplicazioneProduttore	Nome del produttore dell'applicazione che ha generato l'IPdA.	Applicazione	
ApplicazioneVersione	Versione dell'applicazione che ha generato l'IPdA.	Applicazione	
Cognome	Cognome del soggetto che interviene nel processo di produzione del pacchetto di archiviazione.	NomeECognome	
Denominazione	Nome dell'eventuale tipologia o aggregazione a cui appartiene il File o il PdA.	FileGruppo, PdAGruppo	
Descrizione	Informazioni descrittive relative a una eventuale tipologia/aggregazione (di natura logica o fisica) cui il PdA appartiene.	PdAGruppo	
DescGenerale	Informazioni relative all'Indice del Pacchetto di Archiviazione, associate al pacchetto stesso.	IPdA	Applicazione, ExtraInfo, ID, IPdAPre
ExtraInfo	Ulteriori informazioni dell'elemento cui si riferisce, che non possono essere associate ad altri elementi, ad esempio per la definizione di strutture di metadati adeguate allo specifico contesto d'uso. Queste ulteriori informazioni devono essere strutturate nel formato XML, utilizzando uno schema XML. L'insieme di queste informazioni può essere inserito direttamente all'interno o all'esterno dell'elemento come file avendo quindi la stessa struttura dell'elemento <File>.	File, FileGruppo, DescGenerale, MetadatiEsterni, PdA, Processo, Soggetto	MetadatiEsterni, MetadatiIntegrati
File	Informazioni relative al file contenuto nel pacchetto di archiviazione.	FileGruppo	ExtraInfo, ID, Impronta, ImprontaPre, Indirizzo
FileGruppo	Elemento di aggregazione di più file contenuti nel pacchetto di archiviazione. È funzionale alla creazione di insiemi di file sulla base di criteri logici o tipologici.	IPdA	Denominazione, ExtraInfo, File

ID	Identificativo univoco dell'elemento cui si riferisce.	File, DescGenerale, IPdAPre, MetadatiEsterni, PdA, PdAGruppo, PdAPre	
Impronta	Informazioni sull'impronta del file cui l'elemento si riferisce.	File, IPdAPre, MetadatiEsterni	
ImprontaPre	Informazioni relative a precedenti impronte del file contenuto nel pacchetto di archiviazione o del file di metadati (esterno all'IPdA) che contiene le informazioni dell'elemento <ExtraInfo>.	File, MetadatiEsterni	
Indirizzo	Informazioni relative all'indirizzo fisico del file dell'elemento cui si riferisce, espressa come indirizzo URI.	File, IPdAPre, MetadatiEsterni	
IPdA	Indice che contiene le informazioni relative al pacchetto di archiviazione prodotto.		FileGruppo, DescGenerale, PdA, Processo
IPdAIDPre	Identificativo univoco dell'indice del pacchetto di archiviazione associato al precedente pacchetto di archiviazione oggetto della descrizione. Il valore dell'identificativo deve coincidere con il valore dell'elemento <ID> contenuto all'interno dell'elemento <IPdAPre>.	PdAPre	
IPdaPre	Informazioni relative a uno o più indici dei pacchetti di archiviazione da cui è originato quello in oggetto. Tali informazioni sono fondamentali per ricostruire la storia degli oggetti posti in conservazione. L'IPdAPre può riferirsi a: □□una precedente versione dell'IPdA attuale (ad esempio in caso di migrazione e/o modifiche del formato dei file, ove da un PdA si migri ad un nuovo PdA); □□più IPdA cronologicamente antecedenti che hanno generato per fusione l'IPdA attuale (ad esempio in caso di riorganizzazione della struttura dell'archivio, ove più PdA vengano aggregati in un singolo PdA); □□un IPdA cronologicamente antecedente che per frammentazione ha generato l'IdP attuale (ad esempio in caso di scarto di documenti da un PdA, ove a partire da un PdA si generino più PdA).	DescGenerale	ID, Indirizzo, Impronta
MarcaAttached	Data e ora di produzione dell'indice del pacchetto di archiviazione, in forma normalizzata, nel caso in cui questa sia testimoniata con una marca temporale attached all'IPdA stesso. Al contrario dell'analogo elemento <MarcaDetached>, in questo caso non ha senso indicare l'URI della marca temporale.	Tempo	

MarcaDetached	Informazioni sulla localizzazione della marca temporale detached relativa a data e ora di produzione dell'indice del pacchetto di archiviazione. Il valore dell'elemento deve essere espresso nel formato URI.	Tempo	
MetadatiEsterni	Le informazioni dell'elemento <ExtraInfo>, contenute all'esterno dell'IPdA in un file XML le cui caratteristiche sono descritte nei subelementi. Trattandosi di un file, questo elemento ha la stessa struttura dell'elemento <File>. Tale file pur essendo esterno all'IPdA è comunque contenuto nel PdA.	ExtraInfo	ExtraInfo, ID, Impronta, ImprontaPre, Indirizzo
MetadatiIntegrati	Le informazioni dell'elemento <ExtraInfo>, integrate all'interno dell'IPdA e strutturate nel formato XML.	ExtraInfo	
Nome	Nome del soggetto che interviene nel processo di produzione del pacchetto di archiviazione	NomeECognome	
NomeECognome	Nome e cognome del soggetto che interviene nel processo di produzione del pacchetto di archiviazione. Tale elemento deve essere valorizzato nel caso in cui il soggetto sia persona fisica.	SoggettoNome	Nome, Cognome
PdA	Informazioni relative al pacchetto di archiviazione.	IPdA	ExtraInfo, ID, PdAGruppo, PdAPre
PdAGruppo	Informazioni relative a una eventuale tipologia o aggregazione (di natura logica o fisica) cui il PdA appartiene.	PdA	Denominazione, Descrizione, ID
PdAPre	Informazioni relative a uno o più pacchetti di archiviazione da cui è originato quello in oggetto (ad esempio per migrazione di un pacchetto o per aggregazione di più pacchetti).	PdA	ID, IPdAIDPre
Processo	Informazioni relative alle modalità di svolgimento del processo di produzione del pacchetto di archiviazione.	IPdA	ExtraInfo, RiferimentoNormativo, Soggetto, Tempo
RagioneSociale	Ragione sociale del soggetto che interviene nel processo di produzione del pacchetto di archiviazione. Tale elemento deve essere valorizzato nel caso in cui il soggetto sia persona giuridica.	SoggettoNome	
RiferimentoNormativo	Informazioni su norme, regolamenti e standard che regolano il processo di produzione del pacchetto di archiviazione.	Processo	
RiferimentoTemporale	Informazioni relative a data e ora di produzione dell'indice del pacchetto di archiviazione, nel caso in cui non venga apposta una marca temporale. Il	Tempo	

	valore dell'elemento deve essere nel formato ISO 8601 e più precisamente nella forma		
Soggetto	Informazioni relative ai soggetti che intervengono nel processo di produzione del pacchetto di archiviazione.	Processo	ExtraInfo, SoggettoID, SoggettoNome
SoggettoID	Identificativo univoco del soggetto che interviene nel processo di produzione del pacchetto di archiviazione. Se l'identificativo è un codice con ambito nazionale, a tale codice deve essere premesso il codice Paese definito da ISO 3166 seguito dal carattere ":". Se il soggetto è colui che appone la firma digitale all'IPdA è da privilegiare l'uso di un codice identificativo presente in un campo del suo certificato digitale.	Soggetto	
SoggettoNome	Nome o denominazione sociale del soggetto che interviene nel processo di produzione del pacchetto di archiviazione.	Soggetto	NomeECognome, RagioneSociale
Tempo	Informazioni relative a data e ora di produzione dell'indice del pacchetto di archiviazione. Tale elemento è necessario a distinguere i seguenti casi: □□riferimento temporale (l'elemento <RiferimentoTemporale>)	Processo	MarcaAttached, MarcaDetached, RiferimentoTemporale

E' possibile affidare ad ogni attributo una descrizione, le caratteristiche e gli elementi a cui può essere associato.

Nome Attributo	**Descrizione**	**Elementi**	**Caratteristiche**
Altroruolo	Valorizzazione del ruolo rivestito dal soggetto nell'ambito del processo di produzione del pacchetto di archiviazione, nel caso in cui risultino non adeguati i valori previsti dall'attributo Ruolo.	Soggetto	attributo opzionale di tipo CDATA (Character data)
altroschemarif	Valorizzazione del sistema di riferimento utilizzato per identificare il soggetto nel caso in cui risultino non adeguati i valori previsti dall'attributo schemarif.	SoggettoID	attributo opzionale di tipo CDATA (Character data)
codifica	Valorizzazione del tipo di codifica utilizzato nella scrittura del file.	File, MetadatiEsterni	attributo obbligatorio, Valori ammessi: 7bit | 8 bit | base64 | binario | quotedprintable | xtoken
estensione	Estensione che caratterizza il nome del file.	File, MetadatiEsterni	attributo opzionale di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino, due punti o underscore)

formato	Informazioni sulla struttura dati del file a cui si riferisce.	File, MarcaDetached,MetadatiEsterni	attributo obbligatorio di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino, due punti o underscore)
Funzione	Specificazione della funzione di hash utilizzata.	Impronta, ImprontaPre	attributo obbligatorio di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino due punti, o underscore). Valore di default: “SHA-256”
IPdAcorrelato	Identificatore univoco dell’indice del pacchetto di archiviazione contenente la precedente impronta del file contenuto nel pacchetto di archiviazione o del file di metadati (esterno all’IPdA) che contiene le informazioni dell’elemento <ExtraInfo>.	ImprontaPre	attributo obbligatorio di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino, due punti o underscore)
lingua	Lingua in cui sono espresse le informazioni.	Descrizione, RiferimentoNormativo	attributo opzionale di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino, due punti o underscore). Deve essere espresso con un codice a due caratteri, coerentemente con lo standard ISO 639-1:2002. Valore di default: “it”
normal	Indicazione della data e dell’ora di produzione dell’indice del pacchetto di archiviazione, espressa in forma normalizzata. Il valore dell'elemento deve essere nel formato ISO 8601 e più precisamente nella forma YYYY-MM-DDT00:00:0000 (per l’Italia è di default +01).	MarcaAttached, MarcaDetached	attributo obbligatorio di tipo CDATA (Character data)
ruolo	Valorizzazione del ruolo rivestito dal soggetto nell’ambito del processo di produzione del pacchetto di archiviazione.	Soggetto	attributo obbligatorio. Valori ammessi: Delegato, Responsabile della conservazione, Pubblico ufficiale, Altro ruolo
schema	Eventuali informazioni relative al sistema di riferimento nel quale assume significato il valore dell’ identificativo univoco.	ID, IPdA_IDPre	attributo opzionale di tipo CDATA (Character data).
Schemarif	Valorizzazione del sistema di riferimento utilizzato per identificare il soggetto.	SoggettoID	attributo obbligatorio. Valori ammessi: codice fiscale, partita IVA, codice del Servizio Sanitario Nazionale, altroschemarif
schemaxml	Indirizzo URL dove è presente lo schema XML dei metadati utilizzato per descrivere le ExtraInfo.	ExtraInfo	attributo obbligatorio di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino, due punti o underscore); Deve assumere la forma di URL.
Tipo	Indicazione della natura del soggetto.	Soggetto	attributo obbligatorio. Valori ammessi: denominazione, ragione

			sociale
url	Indirizzo URL dove è presente lo schema XML dell'indice del pacchetto di archiviazione.	IPdA	attributo obbligatorio di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino, due punti o underscore); Deve assumere la forma di URL.
Versione	Indicazione della versione dello schema XML dell'indice del pacchetto di archiviazione al fine di gestire l'evoluzione dello standard.	IPdA	attributo obbligatorio di tipo NMTOKEN (ovvero esprimibile con caratteri alfanumerici, punti, trattino, due punti o underscore). Valore di default fisso: "1.0"
xmlcanonico	Indicazione se l'eventuale file in formato xml è trasformato in forma canonica.	Impronta, ImprontaPre	attributo opzionale. Valori ammessi: SI | NO

3.4 – Metadati relativi al documento informatico

La classificazione dei dati è possibile in due modi:

- utilizzando il contenuto informativo dello stesso dato;
- utilizzando i metadati, dove dati descrivono altri dati, si tratta di un' informazione strutturata volta a descrivere, spiegare e localizzare una risorsa informativa rendendo più semplice il suo recupero, utilizzo e gestione.

Nel primo caso si utilizza il contenuto del dato stesso, ad esempio ponendo tutti i documenti che contengono una determinata parola chiave nella stessa categoria.

Nel secondo caso sono utilizzate le informazioni associate al dato come ad esempio un riferimento temporale relativo alla sua creazione, il suo proprietario, la tipologia di dato, l'eventuale livello di sicurezza rispettivamente attribuito, le argomentazioni trattate, e così via.

I metadati si distinguono in tre principali categorie:

- Metadati descrittivi, volti a descrivere una risorsa con lo scopo di scoprirla ed identificarla;

- Metadati strutturali, i quali indicano la struttura di oggetti composti, ad esempio i capitoli che assemblano le pagine;
- Metadati amministrativi, con lo scopo di descrivere informazioni volte a favorire la gestione del file, come ad esempio il tipo di file, il nome dell'utente che l'ha creato, un riferimento temporale relativo alla sua creazione, e così via.

Un' altra caratteristica dei metadati è quella di poter far parte del dato stesso o di poter essere archiviati come oggetti esterni, sono spesso organizzati in gerarchie, ontologie o schemi.

Un esempio di database di metadati è quello di *Metadata registry* (MDR), come ad esempio Oracle, questo supportando la funzionalità di registrazione ne definisce tre operazioni:

- Identificazione, ad ogni oggetto registrato viene attribuito un identificativo univoco;
- Localizzazione, la quale prevede la registrazione della provenienza di ogni oggetto registrato;
- Monitoraggio della qualità, grazie al quale è possibile assicurare che il metadato corrisponda alle aspettative iniziali.

3.4.1 – Metadati minimi del documento informatico

Per quanto riguarda il documento informatico la struttura dei metadati è la seguente:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="fascicolo">
 <xs:complexType>
  <xs:sequence>
   <xs:element name="IPAtitolare" type="xs:string maxOccurs="1"/>
   <xs:element name="IPApartecipante" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="responsabile">
    <xs:complexType>
     <xs:sequence>
      <xs:element name="nome" type="xs:string"/>
      <xs:element name="cognome" type="xs:string"/>
      <xs:element name="codicefiscale" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
   </xs:element>
   <xs:element name="oggettofascicolo" type="xs:string />
   <xs:element name="documento" type="xs:string" maxOccurs="unbounded"/>
  </xs:sequence>
 <xs:attribute name="IDFascicolo" type="xs:string" use="required"/>
 </xs:complexType>
</xs:element>
</xs:schema>
```

Analizziamo le seguenti informazioni:

Identificativo

E' univoco e persistente, si tratta di una sequenza di caratteri alfanumerici associata in modo univoco e permanente al documento informatico in modo da consentirne l'identificazione. Lo standard Dublin Core raccomanda di identificare il documento per mezzo di una sequenza di caratteri alfabetici o numerici secondo un sistema di identificazione formalmente definito. Esempi di tali sistemi di identificazione includono l'Uniform Resource Identifier (URI), il Digital Object Identifier (DOI) e l'International Standard Book Number (ISBN).

Valori ammessi	Tipo di dato	xsd
Come da sistema di identificazione formalmente definito.	Alfanumerico 20 caratteri	<xs:attribute name="IDDocumento" type="xs:string" use="required"/>

Dati di chiusura

Indicano il momento nel quale il documento informatico viene reso immodificabile.

Valori ammessi	Tipo di dato	xsd
Data	Data formato gg/mm/aaaa	<xs:element name="datachiusura" type="xs:date"/>

Oggetto

Si tratta di un metadato funzionale volto a riassumere in breve il contenuto del documento o comunque a chiarirne la natura.

Lo standard Dublic Core prevede l'analoga proprietà "Description" , la quale può includere ma non è limitata solo a: un riassunto analitico, un indice, un riferimento al contenuto di una rappresentazione grafica o un testo libero del contenuto.

Valori ammessi	Tipo di dato	xsd
Testo libero	Alfanumetico 100 caratteri	<xs:element name="oggettodocumento" type="xs:string />

Soggetto produttore

Tale soggetto possiede la competenza e l'autorità di produrre il documento informatico.

Valori ammessi	Tipo di dato	xsd
Nome: testo libero	Alfanumerico 40 caratteri	<xs:element name="soggettoproduttore">

Cognome: testo libero	Alfanumerico 40 caratteri	<xs:complexType> <xs:sequence> <xs:element name="nome" type="xs:string"/> <xs:element name="cognome" type="xs:string"/> <xs:element name="codicefiscale" type="xs:string"/> </xs:sequence> </xs:complexType> </xs:element>
Codice fiscale: Codice Fiscale	Alfanumerico 16 caratteri	

Destinatario

Tale soggetto possiede l'autorità e la competenza di ricevere il documento informatico.

Valori ammessi	Tipo di dato	xsd
Nome: Testo libero	Alfanumetico 40 caratteri	<xs:element name="destinatario"> <xs:complexType> <xs:sequence> <xs:element name="nome" type="xs:string"/> <xs:element name="cognome" type="xs:string"/> <xs:element name="codicefiscale" type="xs:string"/> </xs:sequence> </xs:complexType> </xs:element>
Cognome: testo libero	Alfanumerico 40 caratteri	
Codice fiscale: Codice Fiscale **(Obbligatorio, se disponibile)**	Alfanumerico 16 caratteri	

3.4.2 – Metadati minimi del fascicolo informatico o della aggregazione documentaria

Per Fascicolo Informatico si intende l' insieme organico e ordinato di documenti Formati durante l'attività amministrativa dell'ente con lo scopo di riunire i documenti utili allo svolgimento di tale attività a fini decisionali o informativi.

Può essere organizzato sia raccogliendo documenti che differiscono per natura, contenuto giuridico, formato, sia raccogliendo documenti dello stesso tipo o qualità o forma ordinati seguendo criteri di diversa natura.
Per quanto riguarda il fascicolo informatico o l'aggregazione documentaria la struttura dei metadati minimi è la seguente:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="fascicolo">
 <xs:complexType>
  <xs:sequence>
   <xs:element name="IPAtitolare" type="xs:string maxOccurs="1"/>
   <xs:element name="IPApartecipante" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="responsabile">
    <xs:complexType>
     <xs:sequence>
      <xs:element name="nome" type="xs:string"/>
      <xs:element name="cognome" type="xs:string"/>
      <xs:element name="codicefiscale" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
   </xs:element>
   <xs:element name="oggettofascicolo" type="xs:string />
   <xs:element name="documento" type="xs:string" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="IDFascicolo" type="xs:string" use="required"/>
 </xs:complexType>
</xs:element>
</xs:schema>
```

Analizziamo le seguenti informazioni:

Identificativo

E' univoco e persistente, si tratta di una sequenza di caratteri alfanumerici associata in modo univoco e permanente al fascicolo o aggregazione documentale informatica in modo da consentirne l'identificazione.
Dublin Core raccomanda di identificare il documento per mezzo di una sequenza di caratteri alfabetici o numerici secondo un sistema di identificazione formalmente definito. Esempi di tali sistemi di identificazione includono l'Uniform Resource Identifier (URI), il Digital Object Identifier (DOI) e l'International

Standard Book Number (ISBN)

Valori ammessi	Tipo di dato	xsd
Come da sistema di identificazione formalmente definito.	Alfanumerico 20 caratteri	<xs:attribute name="IDFascicolo" type="xs:string" use="required"/>

Amministrazione titolare

Amministrazione titolare del procedimento, che cura la costituzione e la gestione del fascicolo medesimo.

Valori ammessi	Tipo di dato	xsd
Vedi specifiche Codice IPA	Codice IPA	<xs:element name="IPAtitolare" type="xs:string maxOccurs="1"/>

Per IPA si intende l'Indice delle pubbliche amministrazioni il quale costituisce l'archivio ufficiale contenente riferimenti di Enti pubblici, organizzativi, telematici e toponomastici.

mministrazione partecipanti

Amministrazioni che partecipano all'iter del procedimento.

Valori ammessi	Tipo di dato	xsd
Vedi specifiche Codice IPA	Codice IPA	<xs:element name="IPApartecipante" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>

Responsabile del procedimento

Valori ammessi	Tipo di dato	xsd
nome: Testo libero	Alfanumerico 40 caratteri	<xs:element name="responsabile"> <xs:complexType> <xs:sequence> <xs:element name="nome" type="xs:string"/> <xs:element name="cognome" type="xs:string"/> <xs:element name="codicefiscale" type="xs:string"/> </xs:sequence> </xs:complexType> </xs:element>
cognome: testo libero	Alfanumerico 40 caratteri	
Codice fiscale: Codice Fiscale	Alfanumerico 16 caratteri	

Oggetto

Si tratta di un metadato con la funzione di riassumere brevemente il contenuto del documento o comunque a chiarirne la natura. Dublic Core prevede l'analoga proprietà "Description" che può includere ma non è limitata solo a: un riassunto analitico, un indice, un riferimento al contenuto di una rappresentazione grafica o un testo libero del contenuto.

Valori ammessi	Tipo di dato	Xsd
Testo libero	Alfanumerico 100 caratteri	<xs:element name="oggettofascicolo" type="xs:string />

Documento

Elenco degli identificativi dei documenti, contenuti nel fascicolo, che ne consentono la reperibilità.

Valori ammessi	Tipo di dato	Xsd
Identificativo del documento	Alfanumerico 20 caratteri	<xs:element name="documento" type="xs:string" maxOccurs="unbounded"/>

Capitolo 4:

I formati

4.1 - I formati

Il formato ha il compito di codificare i file[12] appartenenti a determinare categorie di contenuti digitali, come documenti di testo, presentazioni, immagini, database, immagini, fogli di calcolo, registrazioni audio e video, mappe, disegni CAD, e così via.

E' sorprendente l'elevato numero di formati disponibili, adatti a diversi contesti: un'immagine può essere salvata nei formati, GIF, JPG, BMP, TIFF, PNG, un documento di testo può essere prodotto nei formati TXT, RTF, DOC, ODT, e così via per le altre categorie di contenuti digitali.

Vi sono alcune raccolte, quali *File extension collection*[13]*, Wotsit's Format*[14]*, Fileinfo*[15]*, My File Formats*[16]*, FILExt*[17], volte ad indicizzare le svariate migliaia di formati di file disponibili, fornendone descrizioni dettagliate.

Alcuni formati sono scarsamente portatili e necessitano di software specifici che permettano la loro interpretazione, altri sono scarsamente stabili nel tempo, necessitano quindi di continue trasformazioni che li integrino in funzioni sempre più specializzate e complesse.

---

[12] Un file è un insieme di bit (0 ed 1), considerati come un' entità unica dal punto di vista logico e fissati con una certa organizzazione fisica su una memoria.

[13] Mantenuta fino al 2007 dal sito http://www.icdatamaster.com,attualmente il dominio risulta in vendita.

[14] Cfr. http://www.wotsit.org

[15] Cfr. http://www.fileinfo.net

[16] Cfr. http://www.fileformats.com

[17] Cfr. http://www.filext.com

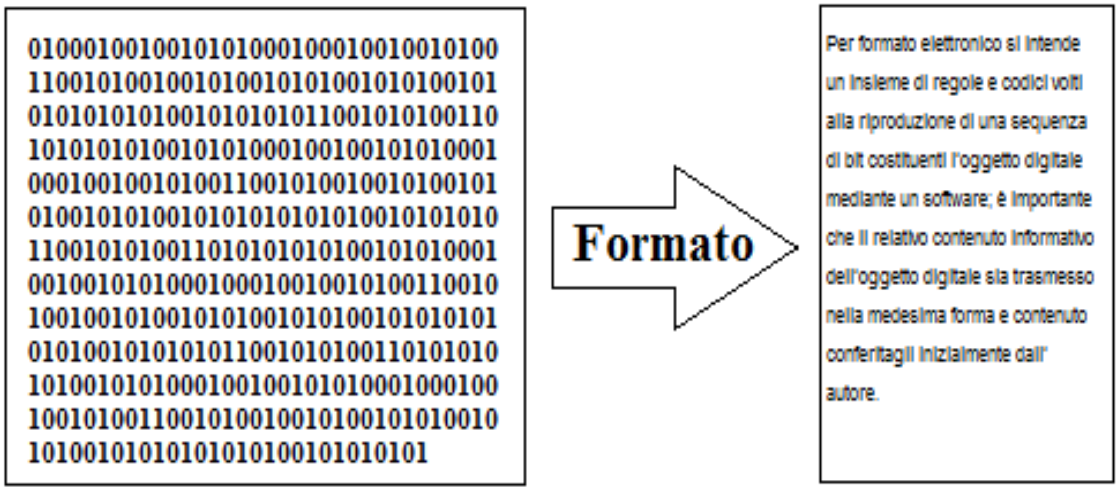
Per evitare che con l'evolversi della tecnologia diventi obsoleto è fondamentale capire cosa esattamente sia un formato e quale sia la corretta applicazione in ambito di conservazione di documenti elettronici, permettendo una corretta leggibilità dei relativi contenuti digitali.

Per formato elettronico si intende un insieme di regole e codici volti alla riproduzione di una sequenza di bit costituenti l'oggetto digitale mediante un software; è importante che il relativo contenuto informativo dell'oggetto digitale sia trasmesso nella medesima forma e contenuto conferitagli inizialmente dall' autore.

Le *specifiche* del formato forniscono le spiegazioni relative alla modalità secondo la quale il file deve essere interpretato, in modo che i programmi riescano a fornirne una corretta rappresentazione elaborandolo in modo esatto.

In parole povere i formati hanno la funzionalità primaria di fornire chiavi di rappresentazione della sequenza di bit di cui i file sono costituiti, la quale senza il formato rispettivamente utilizzato non sarebbero altro che un insignificante susseguirsi di bit.

Si consideri un documento di testo creato con Microsoft Word ad esempio, lasua dimensione convertita in bit corrisponde costituisce un file il quale non ha un significato direttamente interpretabile ne da esseri umani ne da software che non sono in grado di conoscere il formato, solo Microsoft Word è capace di interpretare la sequenza di bit riproducendo il testo ad esso legata.



Ciò è reso possibile dalle informazioni contenute specifiche del formato utilizzato, le quali identificano le informazioni di formattazione del file (margini, dimensioni della pagina, etc..), le parti della sequenza di bit contenente il testo, le informazioni relative ai metadati (autore, data di creazione, data di modifica, ecc..), e così via.
Se si fosse trattato di un immagine, invece di un documento di testo, le specifiche del formato avrebbero dovuto stabilire che il compito di alcuni bit fosse quello di indicarne l'altezza, di altri la larghezza, di altri ancora i colori, e così via.

4.2 – I nomi dei file

La forma base del nome di un file è la seguente:
`nomedelfilecreato.estensione`

- La parte antecedente al punto identifica il nome del File creato.
  Al suo interno possono comparire più punti ma non è possibile utilizzare tutti i caratteri:
  - Caratteri legali: A,….,Z a,….z 0,….9 $ # & + @ ! ( ) – { } ‘ ` _ ~ e lo spazio;
  - Caratteri illegali: | < > \ ^ = ? / [ ] “ ; , * e i caratteri di controllo
- L'insieme dei caratteri che seguono l'ultimo punto è denominato estensione, solitamente sono composte da 3 caratteri.
  Costituiscono il metodo principale per riconoscere il formato di un oggetto digitale.

Secondo il Sito *FileInfo* è possibile suddividere le estensioni in 18 categorie:

File di testo			
.doc	Microsoft Word Dodument	.text	Text File
.log	Log File	.wpd	WordPerfect Document

.msg	Mail Message	.wps	Microsoft Works Word processoe Document
.rtf	Rich Text Format		
File di data			
.123	Lotus 1-2-3- Spreadsheet	.pps	PowerPoint Slide Show
.csv	Comma Separated Values File	.ppt	PowerPoint Presentation
.dat	Data File	.sql	StructuredQuery Language Data
.db	Database File	.wks	Microsoft Works Spreadsheet
.dll	Dynamic Link Library	.xls	Microsoft Excel Spreadsheet
.mdb	Microsoft Access Database	.xml	XML File
File di immagine			
.mng	Multiple Network Graphic	.pct	Picture File
Raster image files		**Vector image files**	
.bmp	Bitmap Image	.ai	Adobe Illustrator File
.gif	Graphical Interchange Format File	.drw	Drawing File
.jpeg	JPEG Image File	.dxf	Drawing Exchange Format
.jpg	JPEG Image File	.eps	Encapsulated Post Script
.png	Portable Network Graphic	.ps	PostScript File
.psd	Photoshop Document	.svg	Scalable Vector Graphics
.psp	Paint Shop Pro Image File	**Page layout Files**	
.tif	Tagged Image File Format	.indd	Adobe InDesign File
3D images		.pdf	Portable Document Format File
.3dm	Rhino 3D Model	.qxd	QuarkXpress Document
.3dmf	QuickDraw 3D Metafile	.qxp	QuarkXpress 6 Project File

File audio			
.aac	Advanced Audio Coding File	.mp3	MP3 Audio File
.aif	Audio Interchange File Format	.mpa	MPEG Audio File
.iff	Interchange File Format	.ra	Real Audio File
.m3u	Media Plylist File	.ram	Real Audio Media
.mid	MIDI File	.wav	Windows WAVE Sound File
.midi	MIDI File	.wma	Windows Media Audio File
File video			
.3gp	3GPP Multimedia File	.mpg	MPEG Video File
.asf	Advanced Systems Format File	.qt	Apple Quick Time Movie
.asx	Microsoft ASF Redirector File	.rm	Real Media File
.avi	Audio Video Interleave File	.swf	Macromedia Flash Movie
.mov	Apple Quick Time Movie	.wmv	Windows Media Video File
.mp4	MPEG-4 Video File		
File web			
.asp	Active Server Page	.js	JavaScript File
.css	Cascading Style Sheet	.jsp	Java Server Page
.htm	Hypertext Markup Language File	.php	Hypertext Preprocessor File
.html	Hypertext Markup Language File	.xhtml	Extensible Hypertext Markup Language File
Files di font			
.fnt	Font File	.otf	Open Type Font
.fon	Generic Font File	.ttf	TrueType Font

Files di plugin			
.8bi	Photoshop Plug-in	.xll	Excel Add-In File
.plugin	Mac OS X Application Plug-in		
System files			
.cab	Windows Cabinet File	.ini	Initialization File
.cpl	Windows Control Panel	.key	Security Key
.dmp	Windows Memory Dump	.sys	System File
.drv	Device Driver		
Setting files			
.cfg	Configuration File	.reg	Registration Information
.msi	Windows Installer File		
Files eseguibili			
.app	Mac OS X Application	.exe	Executable File
.bat	DOS Batch File	.pif	Program Information File
.cgi	Common Gateway Interface Script	.vb	VBScript File
.com	Command File	.ws	Windows Script
Files compressi			
.gz	Gnu Zipped File	.sit	Stuffit archivi
.pkg	Mac OS X Installer Package	.sitx	Stuffit X archivi
.rar	WinRAR Compressed Archive	.zip	Zipped File
.sea	Self-Extracting archivi		
Files codificati			

.bin	Macbinary II Encoded File	.mim	Multi-Purpose Internet Mail
.hqx	BinHex 4.0 Encoded Fiel	.uue	Uncoded File
File di sviluppo			
.c	C/C++ Source Code File	.java	Java Source Code File
.cpp	C++ Source Code File	.pl	Perl Script
Backup Files			
.bak	Bckup File	.ori	Original File
.gho	Norton Ghost Backup File	.tmp	Temporary File
.old	Backup File		
Disk Files			
.dmg	Mac OS X Disk Image	.vcd	Virtual CD
.iso	Disc Image File		
Game files			
.gam	Server Game File	.rom	NES Game ROM
.nes	Nintendo (NES) ROM File	.sav	Saved Game
Misc files			
.lnk	File Shortcut	.yps	Yahoo! Messenger Data File
.torrent	BitTorrent File		

Nella tabella sovrastante sono riportate solo alcune delle estensioni esistenti. Esistono algoritmi matematici che trattano i dati permettendo di organizzare le informazioni in modo da minimizzare il numero di byte necessari alla loro

memorizzazione, laddove la quantità di dati da archiviare sia di dimensioni rilevanti, si parla di metodi di compressione dei file[18].

In generale questi si articolano in due categorie:

- I metodi di compressione Lossy (con perdita di informazione), i quali sono in grado di comprimere i dati al massimo eliminando però definitivamente alcune informazioni contenute nel relativo oggetto[19]. Ciò comporta l'impossibilità di ricostruire esattamente quello che era inizialmente il file compresso. Tale tecnica è di solito utilizzata per comprimere file multimediali, come audio o video, poiché questi hanno dimensioni elevate nel formato in cui vengono creati.
- I metodi di compressione Lossless, (senza perdita di informazione), i quali comprimono i file mantenendo la cronologia di modifiche mantenendo le informazioni di origine. Dal file compresso è possibile ricostruire l'originale in modo esatto.

4.3 – L' identificazione dei formati

Uno dei problemi principali della conservazione digitale è quello di identificare correttamente il formato utilizzato per la creazione di un file, alcune soluzioni sono:

- Identificazione tramite l'estensione del nome del file, ad ogni estensione viene inoltre associato una determinato programma, stabilendo quale applicazione può "aprire" un determinato file;
- Identificazione tramite magic number, una sequenza di byte che si trovano in determinate posizioni all' interno del file, la quale permette di comprendere di quale formato si tratti anche in assenza della sua estensione;

[18] Per tal motive I moderni sistemi operative incorporano un sistema di gestione dei file compressi nel formato ZIP

[19] A meno che non si sia conservata una copia del file originale

- Identificazione tramite metadati espliciti, sono fornite informazioni dettagliate sul formato o sul programma utilizzati;
- Identificazione mediante tipi MIME, permettono un' identificazione univoca.

4.4 – I registri dei formati

E' importante la presenza di database completi e liberalmente accessibili contenenti informazioni tecniche sui formati, è di registri dei formati che si parla. Questi trattano aspetti come identificazione, validazione, caratterizzazione, rappresentazione, valutazione del rischio , relativi ai formati stessi.
Alcuni esempi di registri sono:

- PRONOM[20] (PRatical Online cOMpendium of file formats Technical registry), il quale mette a disposizione on-line informazioni relative ai prodotti software coi quali i file possono essere letti e prodotti, ai formati di file , ai requisiti tecnici necessari in termini di software ed hardware, ed altre informazioni volte a garantire la conservazione a lungo termine. Oggi è reso disponibile a chiunque abbia necessità di ottenere una fonte di informazioni imparziale ma al contempo autorevole.

  E' molto utile per stabilire se vi è un percorso di migrazione da un vecchio formato ad un formato aggiornato.

  Si pensa che tale registro si svilupperà ulteriormente per fornire informazioni tecniche riguardanti le singole versioni dei formati dei file, risulta possibile, attraverso un *submission form on-line* , contribuire al so sviluppo inviando nuove informazioni;

[20] Sito web http://www.nationalarchives.gov.uk/pronom

- TOM[21] (Typed Object Model), un sistema di gestione di dati, sviluppato nel 2004 dalla *University of Pensilvenia Library,* rilasciato con licenza open source, volto a descrivere comportamento e struttura di una vasta varietà di formati fornendo servizi informativi.
  Può inoltre essere utilizzato per l' acquisizione di documentazione relativa ai formati e per ottenere assistenza in caso di conversione;
- GDFR[22] (Global Digital Format Registry), progetto internazionale sponsorizzato dalla *Digital Library* Federation, sviluppato dall' inizio del 2002, mantenuto presso la Harvard University Library, col compito di fornire informazioni affidabili ed autorevoli sui formati di File;
- FRED[23] (Format REgistry Demonstration), sistema basato su TOM, sviluppato presso la University of Pennsylvania Library, in grado di mostrare il funzionamento di un semplice registro dei formati;

La Library of Congress (U.S.) raccoglie sul proprio sto web molte informazioni relative ai formati più rilevanti per le proprie raccolte digitali, suddivisi e classificati per categorie.

4.5 – Standardizzazione dei formati

L'utilizzo di un formato standard è utile per la conservazione digitale a lungo termine di documenti informatici.
Vi sono tre principali organismi di livello internazionale:

- L' ISO[24] (International Organization for Standardization), competente in tutti i settori meno quello elettronico.
  Si tratta di un organismo internazionale composto da 148 enti di standardizzazione provenienti da tutte le parti del mondo, con lo scopo di

[21] Sito web http://tom.library.upenn.edu
[22] Sito web http://hul.harvard.edu/gdfr
[23] Sito web http://tom.library.upenn.edu/fred
[24] Sito web http://www.iso.org

sviluppare standard tecnici volontari in grado di rendere più efficiente e sicure la produzione industriale, e di salvaguardare gli utenti e i consumatori dei servizi e dei prodotti.

Non è prevista l'adesione da parte di singoli individui, la partecipazione in termini tecnici è possibile grazie all' intervento dei rappresentanti degli organismi di standardizzazione.

Sono accordi e trattati a far in modo che gli standard diventino leggi, si ha a che fare con un organizzazione non governativa, agisce autonomamente rispetto ai governi e alle rispettive politiche.

L' ISO si occupa di produrre standard internazionali caratterizzati da principi fondamentali di trasparenza, apertura globale, coerenza tecnica e consenso, le norme da esso create dall' UNI, un' associazione senza scopo di lucro che svolge il processo di normazione in Italia.

Le norme ISO sono numerati ed hanno il seguente formato:

ISO {numero} - {parte} : {anno} {titolo}

Dove:

{numero} corrisponde al numero dello standard;

{parte} corrisponde all' eventuale parte relativa allo standard;

{anno} indica l'anno di pubblicazione;

{titolo} descrive il titolo dello standard;

- L' IEC (International Electrotechnical Committee), utilizzato nel settore elettrico, è un' organizzazione che si occupa di definire standard in materia elettrica, elettronica, ed eventuali tecnologie correlate.

  Fondata nel 1906, su la prima organizzazione a proporre un sistema standard di unità di misura (il sistema Giorgi).

  Ad essa partecipano attualmente 60 Paesi, è composta da rappresentanti degli organismi di standardizzazione nazionali riconosciuti.

  Gli standard IEC sono identificati da numeri interi progressivi;

- L' ITU (International Telecommunication Union) per tante tecnologie di telecomunicazione, è stata fondata a Parigi il 17 maggio 1865 ed è suddivisa in tre settori fondamentali:
  - ITU-R, che regola le radiocomunicazioni;
  - ITU-T, che regola le telecomunicazioni;
  - ITU-D, che promuove l'accesso alle telecomunicazioni nei Paesi in via di sviluppo.

A livello europeo l' organismo principale di standardizzazione è il CEN[25], fondato nel 1961, con lo scopo di armonizzare e produrre, in collaborazione con enti nazionali ed internazionali, norme tecniche europee.
Collabora con l'unione economica europea EFTA per rendere favorevole la sicurezza dei lavoratori e dei consumatori, il libero scambio e la protezione dell'ambiente.
Gli standard da esso prodotti sono adattati, tramite organismi di normazione nazionali, dai Paesi che ne fanno uso.
In Italia invece l'organismo responsabile per lo sviluppo, la pianificazione e l'adozione di standard a livello nazionale è l'UNI (Ente Nazionale Italiano di Unificazione), che come citato poco fa svolge attività normative nei settori industriali, commerciali e del terziario escluso quello elettrico ed elettrotecnico.
L'UNI ha tra i compiti principali l'elaborazione di nuove norme, la partecipazione Italiana all'attività normativa degli organismi normativi mondiali ed europei per rendere le norme più "armoniose", la pubblicazione e la diffusione delle norme tecniche con i prodotti editoriali annessi.

4.6 – I requisiti per i formati elettronici

E' opportuno che un formato possieda requisiti particolari o generali per far sì che sia compatibile con un processo di conservazione digitale, questi si

[25] Comitè Europèen de Normalisation

articolano in due categorie: requisiti di stabilità e requisiti di qualità e funzionalità.

I primi sono applicabili a tutti i formati indipendentemente dalla categoria di appartenenza, e si suddividono in:

- Apertura, con tale requisito sono definite aperte le specifiche accessibili al pubblico ed utilizzabili senza che chi intende utilizzarle non debba corrisponder alcun onere.

  La divulgazione (possibilità di accedere alle specifiche) e l'assenza dei diritti (possibilità di utilizzare le specifiche liberamente) sono due fattori considerevoli per quanto riguarda tale requisito, di fondamentale importanza poiché permette alle aziende di sviluppare software in grado di interpretarlo eliminando a dipendenza dal produttore originario.

- Completa documentazione, requisito molto importante per la produzione di documenti informatici, poiché fornisce le informazioni necessarie che permettono di implementare applicazioni in grado di produrre, leggere o modificare file in quel formato.

  Solo quelli aperti possono esser formati completamente documentati, poiché quelli chiusi non dispongono delle specifiche liberamente accessibili.

  I formati standard aperti corrispondono al più alto livello di documentazione poiché per far si che uno standard venga approvato è necessario che le sue specifiche siano completamente documentate;

- Non proprietà, tale requisito non lega i formati all' esistenza di una specifica azienda che ne è proprietaria in grado di modificare le specifiche.

  L' utilizzo di formati non proprietari è fondamentale poiché nel caso in cui si dovesse conservare un file in un formato di proprietà di un unico fornitore, nel caso questo sparisse, sparirebbero anche il formato e le sue specifiche;

- Standardizzazione, diminuisce il rischio di obsolescenza di un formato;
- Ampia adozione, elemento fondamentale contro il rischio di obsolescenza di un formato, poiché se questo è ampiamente adottato è meno soggetto all' abbandono da parte di aziende e sarà supportato più a lungo dal mercato.
- Trasparenza, requisito che prende in considerazione il grado di semplicità con cui è possibile effettuare analisi dirette di un file.

  I formati digitali che permettono di codificare direttamente e semplicemente l'informazione sono maggiormente adatti alla conservazione.

  E' importante adottare algoritmi non aperti, non proprietari, non soggetti a licenze e vastamente documentati per sviluppare la compressione di un file poiché questo fenomeno potrebbe ridurne il requisito di trasparenza;
- Robustezza, requisito che laddove si presentasse la corruzione del file è in grado di consentire il recupero parziale o totale dei suoi contenuti.

  I formati non binari sono i più robusti poiché nel caso in cui avvenga la corruzione di un bit, la perdita riguarderebbe solo la parte di informazione interessata, rendendo il resto del file leggibile;
- L' auto-contenimento, che rispecchia la capacità di includere tutte le risorse necessarie alla sua rappresentazione, tale requisito è obbligatorio e necessario se si vuole rappresentare il file sempre nella stessa maniera;
- L' auto-documentazione, ovvero la capacità di un formato di supportare metadati che descrivono il contenuto dell' oggetto digitale;
- L'indipendenza dal dispositivo, detto anche requisito di portabilità, richiede la capacità di rappresentazione di un file in maniera attendibile ed indipendente dalla piattaforma hardware e software.

Tale requisito richiede che la rappresentazione del file venga svolta sempre nella stessa maniera, indipendentemente da quello che sia l'output;

- Assenza di meccanismi tecnici di protezione, requisito che rende possibile la replica del contenuto di un documento su nuovi supporti, adottando migrazioni e normalizzazioni che lo rendano disponibile per la diffusione;
- Assenza di limitazioni dell'utilizzo, per evitare di frenare lo sviluppo del software;
- Accessibilità, requisito che rende il formato facilmente fruibile da qualsiasi tipologia di persona, anche grazie all'utilizzo di tecnologie assistive;
- Stabilità, requisito molto importante che favorisce l' utilizzo di formati stabili piuttosto che quelli soggetti a continue modifiche nell' arco del tempo;
- Non modificabilità, requisito volto ad assicurare la stabilità e l'integrità nel tempo nell'arco della conservazione.
- Sicurezza, è importante che un formato sia immune da virus e da altri codici maligni con l'intento di danneggiare il contenuto del file;
- Efficienza, sono favoriti i formati che a parità di prestazione consentono una minore occupazione in memoria rispetto ad altri, consentendo talvolta potenziali risparmi dei costi.

I secondi sono specifici per ciascuna categoria di formato, o rispettiva categoria di appartenenza, sottolineano le proprietà che ci si aspetta di conservare nel tempo. Ad esempio per quanto riguarda i formati immagine è possibile far riferimento alla risoluzione, alla nitidezza, alla profondità di colore, alla possibilità di ingrandimento e così via.

Per i file a prevalente contenuto testuale è possibile la suddivisione dei requisiti in quattro categorie:

- Funzionalità di base, corrispondenti alle minime funzionalità da possedere per una corretta fruizione. Tra queste la possibilità di

stampa su carta, di ricerca delle parole, di lettura su schermo, di formattazione di carattere;

- Integrità della struttura, la quale considera la capacità di conservazione della struttura logica del file, di fondamentale importanza per le enciclopedie, gli elenchi, gli annuari, o comunque di documenti che necessitano diuna struttura formale;
- Integrità di layout, corrispondente al mantenimento del formato nello stesso aspetto;
- Altre funzionalità di livello avanzato.

4.7 – Classificazione delle proprietà dei formati

E' possibile classificare i formati seguendo alcune proprietà, tra le quali approfondiamo le seguenti distinzioni:

- *Formati proprietari e non proprietari*, si ha a che fare con un formato proprietario nel caso in cui questo sia stato creato da un' organizzazione privata la quale ne detiene i diritti di proprietà intellettuale e ne gestisce le specifiche, spesso presenta restrizioni, ottenute attraverso mezzi tecnici o legali, relative alla modifica delle specifiche, allo sviluppo, distribuzione e modifica delle specifiche;

  Viceversa, si ha a che fare con un formato non proprietario, o libero da restrizioni legali, nel caso in cui la sua gestione sia affidata a comunità di sviluppatori che lo condividono. A differenza dai primi non sono controllati e non sono definiti da interessi privati;
- *Formati aperti e chiusi*, Si parla di formato aperto nel caso in cui si avesse a che fare con un formato descritto da specifiche pubbliche, esaustive e accessibili liberamente. Di solito un formato aperto è gestito da organismi di standardizzazione, non prevedendo pagamento di diritti o restrizioni riguardanti il suo utilizzo;

Assicurano importanti benefici tra i quali: l'indipendenza da uno specifico prodotto o fornitore poiché è consentito a chiunque sviluppare applicazioni dato che è prevista una vasta e completa documentazione sul formato., la libertà di scelta del programma da parte dell'utente.

Viceversa un formato si dice chiuso nel caso in cui le sue specifiche non siano pubbliche;

- *Formati standard e non standard*, Si parla di fornai standard dal momento in cui le sue specifiche sono approvate o definite da un ente di standardizzazione o quando il formato è mio diffuso presso una comunità;

  Gli standard si suddividono in due categorie:

  • standard de jure, se presenta un riconoscimento ufficiale avendo le specifiche definite da un organismo di standardizzazione;

  • standard de facto, se è diventato una standard grazie alla sua diffusione, ma le sue specifiche non sono state emanate da alcun organismo di normazione, non detengono alcun riconoscimento ufficiale in gradi di garantire la loro qualità;

- *Formati binari e non binari*, si parla di formato non binario qualora la sequenza di bit che compongono il file sia direttamente interpretabile permettendo di associare ad ogni byte il corrispondente carattere. In tal caso il contenuto di un file corrisponde a caratteri appartenenti a codifiche esistenti.

  Nel caso in cui la sequenza di bit non fosse direttamente interpretabile con un susseguirsi di caratteri e risultasse illeggibile a meno che non si adoperi il programma col quale è stato creato, si ha invece a che fare con un formato di tipo binario. In tal caso è consentita soltanto la memorizzazione di informazioni di tipo testuale, non precedenti altri tipi di informazione.

  I file non binari risultano più semplici da interpretare, garantiscono quindi la loro corretta leggibilità nell' arco del tempo;

- *Formati modificabili e non modificabili*, si ha a che fare con file modificabili qualora ne sia consentita la modifica in maniera semplice, dall' applicazione che lo gestisce.

  Viceversa un file si dice non modificabile quando l'applicazione che gestisce il file non ne consente la modifica oppure ne consente modifiche solo parziali;
- *Formati variabili e fissi*, i file collegati a formati variabili hanno la possibilità di cambiare aspetto in base al dispositivo sul quale vengono visualizzati.

  Viceversa non è possibile mutare l'aspetto di un file in formato fisso;
- *Formati portabili e non portabili*, i formati portabili possono essere utilizzati con semplicità su svariate piattaforme, sia dal punto di vista software, sia dal punto di vista hardware; per tal motivo sono identificati anche come indipendenti dal dispositivo o indipendenti dalla piattaforma di utilizzo.

  Al contrario i formati non portabili possono essere utilizzati esclusivamente su un numero limitato di piattaforme.
- *Formati accessibili e non accessibili*, per accessibilità si intende la capacità di fruibilità, a qualsiasi tipologia di utenti, di un file codificato secondo uno specifico formato, è importante quindi prendere in considerazione la categoria di riferimento al file;
- *Formati stabili e non stabili*, si parla di formati stabili qualora non siano apportate modifiche e le versioni di aggiornamento si susseguono in periodi ragionevoli.

  Il requisito di stabilità è fondamentale se si fa riferimento alla conservazione.

  Infatti la non stabilità di un formato potrebbe causare l'aumento di rischio di errori dovuto alle necessarie migrazioni di file codificati, richiedendo una "vigilanza" continua;
- *Formati sicuri e non sicuri*, si tratta della capacità di immunità da parte del formato, nei confronti di attacchi da parte di codici maligni o virus.

Alcuni formati, non supportando tale requisito, potrebbero essere soggetti a perdite di contenuto parziali o totali;

- *Formati comuni e rari,* l' ampia diffusione di un formato porta migliori garanzie sulla conservazione di lungo periodo.
  E' normale che i formati nuovi sviluppino una scarsa diffusione, ma se hanno la capacità di attirare l' attenzione da parte degli utenti, potranno diventare sempre più comuni;
- *Formati compressi e non compressi*, per ridurre le dimensioni del file, quindi l'occupazione di memoria, alcuni formati prevedono la compressione sviluppata con appositi algoritmi

.

Capitolo 5:

Metadati e standard descrittivi

Tale capitolo approfondirà l'argomento "Metadati", trattato nel terzo capitolo di questa tesi, analizzando alcune tipologie di metadati esistenti con gli annessi standard descrittivi.

5.1 – Dublin Core

Si tratta di Dublin Core Metadata Element set, un vocabolario formato da 15 attributi, abbastanza generici, volti a caratterizzare una risorsa archivistica.
Ha origine nell' anno 1995, dal workshop di Dublin (Ohio), al quale contribuirono esperti pubblici e privati come editori, bibliotecari, archivisti, ricercatori, sviluppatori di software.
Inizialmente lo scopo era quello di dare alla luce uno standard condiviso per la descrizione di risorse digitali, successivamente si evolvette sempre più, fino ad arrivare alla descrizione di qualunque tipologia di risorsa.
Per giungere ad una elaborazioni dei dati di qualità più alta sono stati elaborati alcuni *qualifiers*, o qualificatori, i quali si suddividono in due tipi:

- *Element refinement,* o raffinamento dell'elemento, il quale ne specifica meglio il contenuto;
- *encoding scheme,* o schema di codifica, il quale identifica schemi utili all' interpretazione di un elemento, contenenti vocabolari controllati e nozioni formali, grazie ai quali l' elemento potrà assumere i propri valori.

Insieme all'elenco completo dei valori attribuiti ai qualificatori, i *metadata elements* costituiscono un insieme di termini più vasto, il DCMI metadata terms, nel quale ogni termine è caratterizzato almeno da:

- un nome per creare l' URI del termine, attribuito ad un namescape DCMI;
- un' etichetta che ne indichi il termine, leggibile da un essere umano e non solo da una macchina;
- un URI (Uniform resource identifier) volto ad identificare il termine;
- una definizione.

Si parla di DCMI (Dublin Core Metadata Iniziative) abstract model facendo riferimento ad un modello rappresentato da un insieme di diagrammi di classi UML, in grado di descrivere le relazioni e i concetti contenuti nei Dublin core metadata, raggruppando in una classe un insieme di risorse con attributi in comune, caratteristiche simili, quindi descrivibili da un unico concetto.

L' interoperabilità dei metadati è un argomento molto trattato da la DCMI, nasce dal fatto che utenti con competenze differenti, operanti in diversi ambiti, utilizzano vocabolari diversi per descrivere i metadati che fanno riferimento alle risorse da essi gestite.

Con l'evolversi della tecnologia nasce la necessita di collegare i metadati relativi a più campi, nel modo più automatico possibile, si fa utilizzo del modello Astratto, poco fa descritto, in grado di essere utilizzato come guida per risalire agli elementi dominanti che svolgono la stessa funzione, sottointendendo il medesimo concetto.

Da qui si faccia riferimento al concetto di linked data grazie al quale i dati presenti nel web possono essere collegati tramite metadati in grado di ricondurli a schemi concettuali.

5.2 – MAG

Parlando di MAG (Metadati amministrativi e gestionali) si intende unostandard italiano con l'obiettivo di fornire specifiche formali per la fase di raccolta, disseminazione e trasferimento di metadati e dati digitali nei rispettivi archivi, grazieall' utilizzo di uno schema XML.

MAG è in grado di fornire elementi per:

- identificare gli oggetti digitali univocamente;
- garantire l' integrità e l' autenticità dei contenuti;
- documentare la catena di custodia degli oggetti digitali;
- documentare i processi tecnici utilizzati nella conservazione permanente degli oggetti digitali;
- fornire informazioni relative alle condizioni e i diritti di accesso da parte degli utenti finali agli oggetti digitali.

L' oggetto digitale può essere la dematerializzazione di oggetti analogici o essere a sé stante, ha la possibilità di essere composto da audio, video, immagini statiche, testi nativi in digitale, e testi OCR ottenuti dalla digitalizzazione dei documenti cartacei.

Il file fondamentale del MAG è il metagit.xsd, il quale contiene elementi come ad esempio:

<gen>, elemento obbligatorio ed irripetibile contenente informazioni generiche sul progetto e sul tipo di digitalizzazione facendo affidamento su una serie di elementi figli che raccolgono le informazioni relative al progetto di digitalizzazione, all'ente che svolge il lavoro, all'accessibilità dell' oggetto digitale risultante ;

<bib>, anch' esso elemento obbligatorio e non ripetibile contenente metadati in grado di descrivere l' oggetto analogico;

<stru>, elemento nidificabile e ripetibile, contenente metadati strutturali dell' oggetto digitale, detiene inoltre l'informazione su come collegare tra loro gli elementi per formare un oggetto digitale;

<img>, elemento opzionale e ripetibile contenenti i metadati specifici per le immagini fisse;

<audio>, elemento opzionale ripetibile, contenenti i metadati specifici per file audio;

<video>, elemento opzionale ripetibile, contenenti i metadati specifici per file video;

<ocr> elemento opzionale ripetibile, contenenti i metadati specifici relativi al riconoscimento ottico del testo;

<doc>, elemento opzionale ripetibile, contenenti i metadati specifici per oggetti digitali in formato testuale;

<dis>, elemento opzionale non ripetibile, contenenti i metadati specifici per la distribuzione di oggetti digitali.

Gli elementi ripetibili <img>, <audio>, <video>, <ocr> e <doc> sono i costituenti dell' oggetto digitale.

5.3 – MODS/MADS

MODS (Metadata object description schema) e MADS (Metadata authority description schema) sono schemi complementari, possono essere utilizzati insieme o separatamente.

Il primo è volto alla descrizione di risorse bibliografiche, può essere utilizzato anche con oggetti digitali.

Possiede due elementi root possibili, <mods > che descrive un' unica risorsa, <modscollection> nel caso in cui siano contenute più risorse.

Il secondo è orientato alla descrizione degli authority element, ovvero persone, eventi, organizzazioni e termini.

Anche in tal caso si possiedono due elementi root possibili, <mads > che descrive un' unica risorsa, <madscollection> nel caso in cui siano contenute più risorse.

5.4 – METS

Lo standard METS (Metadata encoding and rasmission standard) è volto alla codifica dei metadati descrittivi, amministrativi e strutturali di un oggetto digitale in formato XML.

Tale schema è ideato in modo da :

- descrivere la gerarchia strutturale di un oggetto;
- specificare un nome identificativo ed una posizione per ogni elemento descritto nella struttura dell' oggetto digitale;
- registrare i metadati associali all' oggetto preso in analisi.

METS è un ottimo formato per lo scambio di oggetti digitali, la loro conservazione, gestione e trasmissione all' utente finale.

E' composto da sette sezioni:

1. intestazione, contenente metadati generali come l'autore la data di creazione o di modifica, e così via;
2. descriptive metadatadmdSec, per i metadati descrittivi;
3. administrative metadata amdSec, per i metadati amministrativi, contiene informazioni sulle modalità di creazione e conservazione del file, alla provenienza dell' oggetto, alla licenza d'uso;
4. file section file Sec, elenca tutti i formati comprendenti l'oggetto;
5. structural map structMap, vera e propria struttura gerarchica dell' oggetto;
6. structural links structLink, per indicare l'eventuale presenza di relazioni ipertestuali tra gli elementi che costituiscono l' oggetto;
7. behavioral behaviorSec, per associare comportamenti eseguibili al contenuto dell' oggetto.

5.5 – ISAD

ISAD(G) (General international standard archivial description) corrisponde allo standard internazionale, proposto nel 1994 dal Consiglio internazionale degli archivi, per la descrizione degli archivi.

Ha lo scopo di fornire un modello di metadati per identificare e spiegare il contenuto del materiale archivistico facilitandone l'accesso ai soggetti interessati.

Svolge tale funzione tramite 26 elementi possibili, da scegliere liberamente in accordo agli standard nazionali, suddivisi in 7 aree:

- Area relativa all' identificazione comprendente: codice di riferimento, nome, data, livello di descrizione, dimensione;
- Area relativa al contesto comprendente: nome del creatore, storia amministrativa, storia dell'archivio, informazioni relative all' acquisto dell'istituzione ospitante l' archivio;
- Area relativa al contenuto ed alla struttura comprendente: ambito, informazioni su eventuali processi di selezione o scarto, informazioni relative ad eventuali aggiunte pianificate, struttura interna e sistema di classificazione;
- Area relativa alle condizioni di accesso e consultazione comprendente: modalità di accesso, eventuali restrizioni sulla riproduzione del materiale, lingua, condizioni fisiche del materiale, presenza di inventari utili per la ricerca;
- Area relativa ai materiali collegati comprendente: esistenza, posizione e disponibilità degli organi e delle copie, relazioni con altri materiali nello stesso o in altri archivi, eventuali pubblicazioni relative all' archivio;
- Area relativa alle note;
- Area relativa alla compilazione della scheda conoscitiva dell'archivio comprendente; note sull' archivista, note su eventuali standard o regole

da seguire per la compilazione della scheda descrittiva, date di preparazione e revisione della scheda.

5.6 – ISAAR

ISAAR (CPF) lo standard internazionale per la descrizione dei soggetti produttori e curatori di archivi è stato elaborato dal 1993 al 1996 dal consiglio internazionale degli archivi per poi essere pubblicato nell' anno 1996.
Identifica 27 elementi suddivisi in 4 aree:

• area di identificazione contenente: tipologia di soggetto interessato, nome ufficiale, forme parallele del nome, altri nomi ufficiali, altre forme del nome come acronimi, identificativi numerici;

• area di descrizione composta da: data di esistenza dell' ente, storia, posti, stato legale, funzioni, descrizione delle sorgenti di autorità, struttura interna, contesto generale;

• area relativa alle relazioni con le altre entità contenente: nomi o identificativi dei soggetti in relazione, tipo di relazione, descrizione della relazione, date della relazione;

• area relativa alla compilazione della scheda conoscitiva dell'archivio, comprendente: codice identificativo, identificativi dell' istituzione responsabile della scheda, regole e convinzioni, stato della scheda, livello di dettaglio, data di creazione e revisione della scheda.

5.7 – EAC/EAD

EAC-CPF (Encoded archivial context-Corporate bodies, persons and families) consiste in un progetto per la trasmissione dei metadati relativi alla descrizione dei soggetti, mediante formati XML, nato nell' anno 1998.

Un documento presentante tale standard contiene due elementi obbligatori:

- <control>, il quale è volto alla codifica di informazioni raccolte nella control area tramite determinati elementi ed attributi;
- <cpfDescription> oppure <multipleIdentities>, il primo contiene identity, description e relations dell' ISAAR come sottoelementi, il secondo viene usato nel caso in cui ci sia più di un <cpfDescription>, cioè ad esempio quando l'autore di una risorsa è una sorta di pseudonimo che nasconde due diversi autori.

EAD (Encoded archivial description) consiste invece in uno standard de facto volto a codificare in XML i mezzi per l'aiuto alla ricerca archivistica.

Il suo elemento <ead> contiene i seguenti elementi:

• <eadheader>, il quale contiene informazioni volte a descrivere il file;

• <frontmatter>, il quale contiene informazioni come titolo, prefazione e così via, utili a mostrare o pubblicare il documenti EAD;

• <archdesc>, descrive il materiale archivistico e le informazione amministrative e di contesto.

5.8 – ISDIAH

Lo standard internazionale per la descrizione degli istituti conservatori di archivi, ISDIAH è stato pubblicato nell'anno 2008 ed identifica 31 elementi suddivisi in tre aree:

• area di identificazione comprendente: codice identificativo, forme autorizzate del nome, firme parallele del nome, altre forme del nome, tipologia di ente;

• area delle informazioni relative ai contatti, contenente: ubicazioni ed indirizzi, telefono fax o e-mail, e contatti per il personale;

• Area di descrizione comprendente: storia dell' ente, contesto territoriale e culturale, metadati e fonti normative, struttura amministrativa, politiche di gestione documentaria, edifici, patrimonio archivistico e strumenti di ricerca;

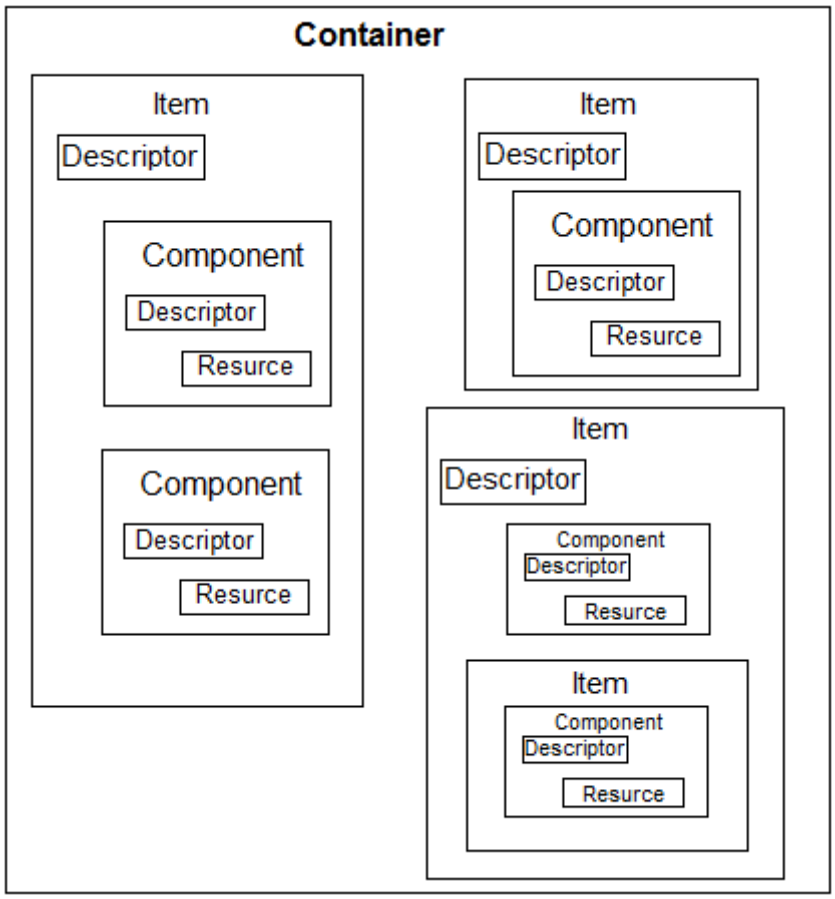
• area delle informazioni sull'accessi comprendenti: orari di apertura, condizioni e requisiti per l'accesso e l'uso, accessibilità;
• area delle informazioni riguardanti i servizi comprendente: servizi per la ricerca, servizi ti riproduzione e spazi per il pubblico;
• area di controllo a cui son associati: codice identificativo della descrizione, identificativi dell' istituzione responsabile, norma e convinzioni utilizzate bella compilazione, stato della scheda, livello di dettaglio, date di creazione aggiornamento o cancellazione della scheda, lingue e codici usati nella compilazione, note riguardo la creazione e la revisione della scheda.

5.9 – ISDF

ISDF (International standard for describing functions) consiste in uno standard internazionale pubblicato nell' anno 2007 col compito di descrivere le funzioni degli istituti produttori e conservatori di archivi.
Identifica 23 elementi suddivisi in 4 aree:
• area di identificazione che indica: tipo di funzione, forme autorizzate dal nome, forme parallele al nome, altre forme del nome e classificazione;
• area relativa al contesto comprendente: date, descrizione, storia e normative;
• area relativa alle relazioni comprendente: forma autorizzate dell' identificativo della funzione correlata, il tipo che indica se la relazione è stabilita conla funzione o con una delle sue suddivisioni, la categoria di relazione (gerarchica, temporale o associativa generale;
• area di controllo comprendente: codice che indica la descrizione della funzione, identificativi dell'istituzione responsabile della scheda, norme e convenzioni utilizzate nella compilazione della scheda, stato della scheda, livello di dettaglio, date di creazione, revisione o cancellazione della scheda, lingue e codici usati nella fase di compilazione, fonti consultate, note riguardanti la creazione e la revisione della scheda.

5.10 – MPEG 21-DIDL

Mpeg definisce DID (Digital item declaration), un modello utilizzato per la descrizione di oggetti digitali complessi, e DIDL (Digital item declaration lenguage), un linguaggio descrittivo in formato XML corrispondente.
E' possibile stabilire uno schema riassuntivo degli elementi principali del modello.



Dove:

- *Item*, corrisponde all'oggetto digitale dotato di metadati descrittivi, di identificazione e di rappresentazione.
  Gli Item possono contenere altri item, componenti con relativi descrittori e scelte attraverso le quali è permessa la loro configurazione o personalizzazione;

- *Container,* ovvero una struttura raggruppante Item o altri container e contenente una raccolta di metadati;
- *Resource,* consistente in un singolo stream di data;
- *Component,* cioè un raggruppamento di resource;
- *Descriptor,* contenente le informazioni relative ad un item, un container o un component;
- *Fragment,* in grado di identificare un punto o un intervallo preciso in una resource;
- *Anchor,* il quale congiunge dei descriptor ad un fragment;
- *Condition,* descrivente l'entità opzionale legandola ad una selection che condiziona la sua inclusione;
- *Selection,* la quale descrive una decisione in grado di attivare una o più conditions su un item;
- *Choice,* corrispondente ad un insieme di scelte, che possono essere esclusive od inclusive, in grado di cambiare la configurazione di un Item;
- *Assertion,* indica lo stato configurato in modo parziale o completo di una choice attribuendo un valore vero, falso o indeterminato ad un dato predicato di dichiarazioni identificabili, associato a tale scelta;
- *Annotation,* descrivente alcune informazioni legate ad un' altra identità del modello;
- *Statement,* corrispondente ad un valore testuale letterale contenente informazioni;
- *Declarations,* in grado di contenere un numero qualsiasi di item, component e descriptor.

5.11 – PREMIS

Il progetto internazionale PREMIS (Preservation Metadata: implementation strategies), studia un sistema di metadati orientati in modo specifico alla conservazione.

In tale modello di dati sono definite cinque tipologie di entità:

- *Entità intellettuali*, elementi di base come libri, foto, mappe, etc, in grado di concretizzare gli oggetti digitali;
- *Oggetti digitali*, obiettivo della conservazione, fondamentalmente di tre tipologie:
    - File, normali file dotati del proprio formato;
    - Representation, ovvero l' insieme di tutti i file utili a fruire correttamente di un' entità intellettuale;
    - Bitstream, cioè parti logiche di un file, ad esempio una traccia audio di un video.

  Gli oggetti possono contenere informazioni relative all' identificatore univoco per l'oggetto, all' integrità, alla dimensione dell' oggetto, al formato, al nome di origine, alla creazione, alle proprietà significative, all' ambiente operativo, sulla firma digitale e sulle relazioni con altri oggetti e tipi di entità.
- *Diritti,* contenenti informazioni relative ai permessi che l'archivio ha sugli oggetti, come il diritto di copia, di diffusione, di trasformazione, e così via;
- *Agenti,* persone, organizzazioni o applicazioni in grado di svolgere un ruolo negli eventi;
- *Eventi,* informazioni relative alle operazioni eseguite sull' oggetto dell' archivio, queste sono raccolte in:
    - Identificatore univoco per l'evento;
    - Tipo di evento;
    - Data ed ora dell' evento;
    - Descrizione dettagliata dell' evento;

- Esito dell' evento in forma codificata;
- Descrizione del dettaglio dell' evento;
- Agenti coinvolti ed eventuale ruolo rispetto all'evento;
- Oggetti coinvolti e ruoli svolti rispetto all'evento.

5.12 – MARC

Il formato di catalogazione per le biblioteche MARC (Machine- readable cataloging) è stato sviluppato a partire dagli anni Sessanta dalla Library of Concress Statunitense.
Ha lo scopo primario di rendere standard i campi di una scheda di catalogo per un documento pubblicato in modo ri renderlo utilizzabile da parte di una macchina.
Per risolvere il fatto che i vari campi di un catalogo potessero avere vari contenuti, utilizza un metodo per dire alla macchina in modo automatico quando una data informazione inizia o finisce indipendentemente dalla sua lunghezza.
Infatti ad ohni campo viene attribuito un tipo associato ad un numero a tre cifre detto *tag,* al quale segue un gruppo di numeri detto indicatore, il quale è utile per dare istruzioni variabili a seconda del tag utilizzato.
Un record MARC è caratterizzato da tre elementi:

- Una struttura;
- Una legenda;
- Il contenuto dati del record.

La struttura di un record MARC si presenta come segue:

Tag	Indicator	Delimiter	Subfield Code	Contenuto
(3 car)	(2 car)	(1 car)	(1 car)	(n car)

TTT	**II**	**DS**	**.....**	**DS**	**……**	**DS**	**……**

Oltre agli elementi descritti ve e sono altri due che precedono i singoli campi, aventi funzione di controllo:

- Il leader, il primo elemento del record con lunghezza fissa di 24 caratteri;
- La directory, immediatamente successiva al leader, contenente una serie di entries di lunghezza fissa pari a 12 caratteri.

L' evoluzione dello standard in MARC 21, presenta l'inserimento del supporto di Unicode per garantire l'utilizzo anche in paesi con alfabeti e linguaggi differenti.

Questo nuovo aggiornamento presenta il supporto di vari tipi di record:

- *Authority record,* contenente le informazioni sulle forme autorizzate per i nomi per facilitare la ricerca;
- *Bibliographic record,* contenente le informazioni vere e proprie relative alla bibliografia;
- *Classificatione record,* contenente le specifiche informazioni di classificazione dell' archivio;
- *Community information record,* contenente informazioni non bibliografiche riferite a particolari comunità di utenti;
- *Holding record,* contenente informazioni sulla posizione fisica dell' oggetto, sul numero di copie disponibili e così via.

Conclusioni

Alla luce di quanto trattato sino ad ora è possibile affermare che sono tanti i vantaggi legati al processo di conservazione sostitutiva, di seguito si affrontano i principali.

Nel momento in cui il contenuto di grandi archivi cartacei subisce un processo "Dematerializzazione Informatica", viene ridotto drasticamente il reale ingombro spaziale, necessario per contenere i documenti allo stato cartaceo. Ciò porta ad evitare sprechi di carta, salvaguardando l'ambiente.

Catalogando ogni documento informatico, seguendo un determinato ordine, è possibile effettuare ricerche nell' immediato, permettendo inoltre un fulmineo aggiornamento dei dati interessati. Viene conseguentemente di gran lunga facilitata la consultazione.

Il documento informatico può garantire maggior esaustività nel caso in cui fosse formato da più elementi multimediali, come immagini, suoni, filmati video, animazioni, e così via.

Vi è la vantaggiosa possibilità di trasmissione a distanza di un documento informatico in tempo reale grazie all' utilizzo della rete Internet, riducendo od eliminando così costi e tempi di distribuzione.

A livello economico, oltre alla riduzione dei costi sul materiale e sulla distribuzione, vi è un rilevante risparmio sul personale adibito all' amministrazione dei documenti cartacei, che porta ad un aumento della produttività ed efficienza del personale il quale impiegherà le proprie risorse in altre attività.

Talvolta i documenti cartacei, se conservati in luoghi inadeguati come umidi scantinati ad esempio, sono soggetti ad un progressivo deterioramento, tale problema si risolve nel momento in cui vengono convertiti in documenti informatici.

Se avviene una corretta conservazione sono garantiti maggior controllo e sicurezza per l'accesso alla documentazione.
Come in ogni situazione è da prendere in considerazione anche il rovescio della medaglia, vi sono difatti alcune problematiche e rilevanti aspetti negativi legati al processo di conservazione, tra i quali gli elementi di seguito trattati.
Di fondamentale importanza è organizzare un metodo di autenticazione infallibile in grado di controllare situazioni di manipolazione o di plagio del documento conservato, per garantire il più alto grado di sicurezza del processo.
Memorizzare dati o documenti su diversi supporti o con differenti formati può diventare un problema nel momento in cui non vi sia la possibilità di conversione da un supporto all' altro o da un formato all' altro.
I documenti informatici si presentano facilmente manipolabili e dinamici, possono divenire di conseguenza instabili poiché risulta difficile garantire diritti di proprietà intellettuale nel caso in cui un contenuto non fosse identificabile con certezza.
Uno dei maggiori rischi legati al processo di conservazione dei documenti informatici si presenta come una sorta di analfabetismo informatico, a prescindere dalla capacità di leggere e scrivere infatti, l'accesso all'informazione digitale presuppone non solo la conoscenza e l'esperienza dell'infrastruttura, ma anche la possibilità economica di potervi accedere.
Risulta complesso interpretare file memorizzati col supporto di un software proprietario il quale utilizza un formato del quale non sono mai state pubblicamente fornite le specifiche.
Non è possibile affrontare la conservazione con procedure valide in qualsiasi ambiente poiché occorre prendere in analisi svariati aspetti.
Si può inoltre verificare una perdita di dati importanti nel corso delle svariate modifiche apportata al documento.

Il processo di conservazione sostitutiva è molto adottato oggi, le realtà aziendali che trattano le proprie procedure modificandole con tali sistemi possono trarre grossi vantaggi competitivi rispetto ad altre aziende.

Non vi è alcun metodo esente da aspetti critici, considerando che non esistono tecniche univoche per affrontare qualunque tipologia di problema conservativo, è necessario garantire il totale controllo sulla modalità di produzione e di memorizzazione delle fonti informatiche facendo riferimento nello specifico ai formati, ai metadati utilizzati e alle informazioni rappresentate.

Bibliografia

- Stefano Pigliapoco, Stefano Allegrezza (2008), "*Produzione e conservazione del documento digitale - Requisiti e standard per i formati elettronici".* Macerata : EUM. - v.

- Vito Rizzo (2000), *"Documento informatico, firma digitale e commercio elettronico".* Napoli etc.! : Edizioni scientifiche italiane.

- Maria Guercio (2013),*"Conservare il digitale: principi, metodi e procedure per la conservazione a lungo termine di documenti digitali".* Roma [etc.] : GLF editori Laterza.

- Alessandro Sinibaldi, Paolo Bartolomeo Buongiorno (2012), *"Manuale di conservazione digitale".* Milano : Franco Angeli.

- Paolo Carretta, Antonio Cilli, Antonino Iacoviello, Alessio Grillo, Francesco Trocchi (2012), "*L' acquisizione del documento informatico : indagini penali e amministrative*". Roma : Laurus Robuffo.

- Giovanni Michetti (2007), "*OAIS. Sistema informativo aperto per l' archiviazione".* Roma, Istituto Centrale per il Catalogo Unico (ICCU).

Sitografia

http://www.sa-ero.archivi.beniculturali.it/fileadmin/template/allegati/pubblicazioni/interventi

http://www.agid.gov.it/

http://vvw.web.cs.unibo.it/wiki/images/1/1f/DI-05-documentoInformatico.pdf

http://www.indicepa.gov.it/documentale/index.php

http://www.giustizia-amministrativa.it/documentazione/studi_contributi/firma_digit.htm#3

http://www.salom.archivi.beniculturali.it/fileadmin/template/allegati/Corsi/luglio_2012/fascicolo_9luglio2012.pdf

http://www.misco.it/amministrazione/elaborazione-digitalizzazione-documenti