

DOCUMENT MANAGEMENT

Manuale di conservazione digitale

Alessandro Sinibaldi
Paolo Bartolomeo Buongiorno



FRANCOANGELI

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio “Informatemi” per ricevere via e.mail le segnalazioni delle novità.

Alessandro Sinibaldi
Paolo Bartolomeo Buongiorno

Manuale di conservazione digitale



FRANCOANGELI

Copyright © 2012 by FrancoAngeli s.r.l., Milano, Italy

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito www.francoangeli.it.

Indice

Presentazione	pag. 11
Ringraziamenti	» 16
Parte prima Informazioni e documenti	
1. Cenni di archivistica	» 19
1. Documenti e archivi	» 19
2. L'archivio corrente e l'archivio di deposito	» 24
3. Il protocollo informatico	» 29
4. Caratteristiche fisiche di un archivio cartaceo	» 34
Riferimenti bibliografici	» 44
2. Lo standard OAIS	» 45
1. Il modello informativo	» 48
2. Il modello funzionale	» 55
Riferimenti bibliografici	» 64
3. Processi e documenti	» 65
1. Cosa è un processo?	» 65
2. Ripensare i processi	» 70
3. <i>Business process management system</i>	» 74

4. Dal processo al documento	pag. 77
Riferimenti bibliografici	» 82
4. Dal cartaceo al digitale e ritorno	» 83
1. Lettura e riconoscimento ottico	» 85
2. I glifi	» 94
3. Un esempio di servizio: la mailroom	» 96
Riferimenti bibliografici	» 99
5. La conservazione a norma	» 101
1. La conservazione sostitutiva	» 103
Riferimenti bibliografici	» 108
6. Il documento digitale	» 109
1. I formati	» 110
2. Modalità di identificazione	» 113
3. Il fascicolo elettronico	» 120
4. La fatturazione elettronica	» 132
Riferimenti bibliografici	» 136
7. Metadati e standard descrittivi	» 137
1. Dublin core	» 138
2. MAG	» 144
3. MODS/MADS	» 148
4. METS	» 149
5. ISAD	» 153
6. ISAAR	» 154
7. EAC/EAD	» 156
8. ISDIAH	» 159
9. ISDF	» 160
10. MPEG 21-DIDL	» 161

11. PREMIS	pag. 165
12. MARC	» 169
13. Dal metadato alla biblioteca digitale	» 173
Riferimenti bibliografici	» 177
8. La conservazione a lungo termine dei documenti digitali	» 181
1. Conservazione e ciclo di vita del documento	» 189
2. <i>Le significant properties</i>	» 190
3. La standardizzazione dei formati di archiviazione	» 196
4. Architetture di conservazione	» 202
5. L'accreditamento dei conservatori	» 207
6. Biblioteche digitali	» 213
7. Archiviare il web	» 218
8. E se tutto va storto?	» 224
Riferimenti bibliografici	» 226

Parte seconda Sicurezza

9. Introduzione alla sicurezza delle informazioni	» 229
1. I requisiti di sicurezza	» 229
2. La gestione del rischio	» 235
3. ISO 17799 e ISO 27000	» 240
4. Il segreto di Stato	» 243
Riferimenti bibliografici	» 244
10. Identity management	» 245
1. Introduzione	» 245
2. <i>Single sign on</i>	» 252
3. LDAP	» 259

4. I problemi di sicurezza	pag. 266
Riferimenti bibliografici	» 267
11. Certificati e firme	» 271
1. Introduzione	» 271
2. La normativa sui certificatori	» 275
3. Storia della firma digitale in Italia	» 276
4. Tipi di firme elettroniche	» 281
5. Problemi di sicurezza	» 284
6. Gli standard europei	» 285
7. La marca temporale	» 288
Riferimenti bibliografici	» 290
12. La posta elettronica certificata	» 291
1. Il modello dati	» 297
2. PEC e gestione documentale	» 298
Riferimenti bibliografici	» 300
13. Proprietà intellettuale e DRM	» 301
1. Copyright e copyleft	» 303
2. <i>Digital rights management</i>	» 304
Riferimenti bibliografici	» 317

Parte terza
Strumenti di supporto

14. Modellazione	» 321
1. Tecniche diagrammatiche	» 323
2. Introduzione a XML	» 331
Riferimenti bibliografici	» 353

15. Tecnologie	pag. 355
1. Architettura di rete	» 355
2. Architettura applicativa	» 360
3. Storage	» 362
4. Backup	» 369
5. Virtualizzazione	» 372
6. SOA	» 373
7. <i>Cloud computing</i>	» 386
Riferimenti bibliografici	» 389

Glossario	» 391
Riferimenti bibliografici	» 400

Appendici

Appendice A. Principali normative	» 403
Appendice B. Standard documentali	» 407
Appendice C. Progetti	» 413
Appendice D. Software	» 419
Indice analitico	» 423

Presentazione

Il nostro tempo è arrivato a un importante punto di svolta. L'enorme progresso dell'informatica negli ultimi decenni ha dato all'uomo la possibilità di trasmettere grandi quantità di informazione, anche complessa, da una parte all'altra della Terra in pochi istanti, con vantaggi così evidenti da portare al rapido superamento delle tecnologie analogiche, come l'uso della carta e della pellicola fotografica. Questo cambiamento epocale non è stato però completamente indolore. Si è sottovalutato, infatti, un aspetto importante e cioè quello legato alla conservazione, soprattutto se a lungo termine, dell'informazione. Se, infatti, un foglio di carta o una lingua possono durare centinaia di anni, così non è per il formato di un file, il tempo di vita di un hard disk o di un CD o l'utilizzo di un linguaggio informatico o un applicativo.

Ci stiamo di fatto rendendo conto che il mondo rischia di attraversare, se non lo ha già fatto, un periodo di buio nella storia umana semplicemente perché non sarà possibile trasmettere ai nostri figli le testimonianze della nostra epoca storica. Anzi, a dire il vero, il problema è così grave da rischiare di non riuscire neppure a trasmetterle a noi stessi tra qualche tempo.

La velocità con cui tutto ciò sta accadendo porta una serie di conseguenze. Il primo è un salto psicologico che non è presente solo nella forma di "analfabetismo digitale" ma anche di difficoltà di stare al passo con le tecnologie che si evolvono e susseguono rapidamente. Nel mondo digitale il cambiamento è all'ordine del giorno e, perfino per chi lavora nel settore, è difficile aggiornarsi in tempo utile, anche per l'elevato tasso di specializzazione richiesto. Ciò che si sa usare o si conosce oggi è già vecchio in un arco di tempo di cinque anni. In più va considerato il fatto che le competenze informatiche sono diventate così pervasive da diventare irrinunciabili per chiunque.

Questo crea anche una carenza di competenze, dal momento che, per esempio, gli archivisti, che hanno le conoscenze necessarie alla conservazione, alla classificazione e alla disseminazione dell'informazione sono sempre state figure sostanzialmente "umanistiche" mentre gli informatici, dotati delle competenze tecniche necessarie nell'ambito della conservazione digitale, sono invece all'oscuro anche dei più elementari concetti di gestione degli archivi.

Lo stesso problema riguarda le normative, che si trovano sempre più costrette, anche per il numero crescente di reati commessi nel mondo digitale, a occuparsi dei dettagli tecnici con regole quanto mai precise, fatto, questo, che farebbe rivoltare nella tomba un giurista "purista" per il quale una legge dovrebbe essere espressa in una forma il più possibile generale e svincolata dagli aspetti concreti. Basta seguire tutto l'iter normativo per l'equivalenza tra documento digitale e quello cartaceo per rendersi conto di quale sia la difficoltà. Ciò crea anche pericolosi "vuoti" normativi che costringono sempre di più il giudice a piegare e interpretare le leggi nate per ambiti diversi in attesa di norme più precise.

Tutta questa frammentazione di competenze, esacerbata da una buona dose di diffidenza per il nuovo, dalla difesa a oltranza di ruoli storici e di cultura acquisita, e a volte anche dalla mancanza di umiltà nell'accostarsi a professionalità diverse, provoca una pericolosa mancanza di confronto e comunicazione tra le discipline, in particolar modo quelle connesse alla conservazione digitale, che dovrebbe veder lavorare a braccetto informatici, archivisti, giuristi, manager e quant'altro necessario.

La prima grande conseguenza di questo problema è la mancanza di una visione d'insieme, di una lungimiranza e di una regia condivisa. Anche quando si assiste agli sforzi di persone illuminate in determinati ambiti questi rimangono isolati e incompresi perché non si sa o non si è voluto comunicare con gli altri attori, ognuno dei quali presiede una parte del problema.

Tutto ciò che abbiamo detto porta alla necessità di creare nuove figure professionali capaci di muoversi al bordo tra le varie discipline ed è a esse che è rivolta principalmente quest'opera la quale nasce con il desiderio di "fare il punto" e di inserirsi in un dibattito in corso. Abbiamo concepito il testo come un manuale che possa essere letto per intero per avere una visione d'insieme dell'argomento, ma anche consultato puntualmente su argomenti specifici quando serve.

Il libro inizia con la trattazione del documento cartaceo tradizionale e l'archivistica "classica" e prosegue con le problematiche relative alla transizione al digitale per poi approdare al tema della conservazione digitale a lungo termine.

Il testo è diviso in tre parti.

La prima è quella più propriamente dedicata ai documenti.

Il cap. 1 è un'introduzione all'archivistica e, in particolare alle varie fasi di vita degli archivi e a come vengono gestiti i documenti, soprattutto cartacei. Particolare spazio è dedicato al protocollo informatico o, meglio, alla gestione informatica del protocollo. Quest'ultimo rappresenta l'atto con cui un documento entra a far parte del sistema documentale di un ente ed è così importante da costituire, secondo la legge, "atto pubblico di fede privilegiata", cioè non ha un semplice valore amministrativo ma può essere usato come prova in ambito giudiziario e lo stesso operatore di protocollo è un pubblico ufficiale.

Il cap. 2 racconta il modello OAIS che descrive gli aspetti funzionali e informativi di un qualsiasi archivio. La sua importanza è tale da essere assunto come riferimento mondiale e da diventare standard ISO.

Dal momento che un documento è definito come la rappresentazione di un atto o fatto giuridicamente rilevante (definizione tratta dalla normativa vigente), è impossibile parlare di informazione senza parlare dei processi che la generano, consumano o gestiscono. Lo stesso concetto di vincolo archivistico che unisce documenti che sono parte dello stesso procedimento o affare esprime questo legame. Per questo il cap. 3 è dedicato ai processi, a cosa sono, a qual è il loro ciclo di vita, a come si descrivono tramite i diagrammi IDEF0, che sono lo strumento più diffuso nei progetti archivistici, ma anche a come possono essere ridisegnati per renderli più efficienti, più automatizzati e "dematerializzati". Introduciamo anche il concetto di *Business process management system* che è uno strumento software che permette di disegnare, simulare, eseguire e monitorare flussi di attività logicamente correlate quali possono essere quelle relative alla conservazione digitale.

Il cap. 4 è invece dedicato ai problemi della transizione da cartaceo a digitale e viceversa. Si inizia con la descrizione delle tecnologie di lettura e riconoscimento ottico, partendo da quelle più semplici come l'OMR, utilizzato per esempio nella lettura automatizzata delle schedine del Totocalcio, fino ad arrivare a quelle più complesse di riconoscimento della scrittura manuale.

La necessità di far convivere il mondo cartaceo e quello digitale, con continui rimandi dall'uno all'altro, ha portato alla creazione dei QR-codice e dei glifi. Questi ultimi, in particolare, sono previsti nel nuovo CAD (D.lgs. n. 235/2010) come mezzo per trasportare la firma digitale su carta.

Un ulteriore problema è quello legato alla conservazione sostitutiva: in un mondo che sta migrando dall'analogico al digitale cosa e come conservare? In quali condizioni una copia digitale può sostituire in tutto un originale analogico e viceversa? Nel cap. 5 affrontiamo il problema della conservazione a norma, comunemente detta "conservazione sostitutiva".

Il cap. 6 è completamente incentrato sui documenti digitali, di come essi possano essere identificati univocamente e di cosa sia il formato di un file. In quest'ambito rientrano progetti interessanti relativi alla sanità elettronica, o *e-Health*, e alla fatturazione elettronica, entrambi argomenti che sono al centro dell'Agenda digitale europea e delle priorità di dematerializzazione di molti governi.

Il cap. 7 è dedicato ai metadati, che sono "dati riguardo ai dati", nel senso che costituiscono informazione strutturata che descrive, spiega, localizza o in qualsiasi modo facilita il reperimento, l'uso e la gestione di qualunque altra risorsa. Essi sono fondamentali per l'archivista perché è attraverso di essi che i documenti vengono posti in relazione gli uni con gli altri e con l'agente che li ha prodotti e viene mantenuta nel tempo la loro comprensione e intellegibilità. Nel corso del capitolo vedremo gli standard internazionali più diffusi. Infine, descriveremo come creare una biblioteca digitale che sia in grado di attingere a più sorgenti documentali con caratteristiche differenti allo scopo di presentare all'utente un'unica interfaccia di interrogazione e fruizione.

Il cap. 8 è uno dei capitoli centrali del libro, in quanto affronta il problema della conservazione digitale e, in particolare, quali siano le strategie consigliate, dal momento che non esiste ancora una soluzione definitiva. L'argomento è quanto mai attuale dato che la quantità di documenti digitali aumenta a vista d'occhio ma l'invecchiamento rapido dei formati e dei supporti rischia di creare un vero e proprio buco nella memoria storica dell'umanità soprattutto quando si parla di archivi privati. Se un documento cartaceo può durare, in condizioni ottimali di conservazione, anche centinaia di anni, così non è per uno digitale che già nel giro di poco tempo rischia di diventare illeggibile nonostante, ironicamente, la facilità con cui esso può essere duplicato e trasmesso in un attimo da una parte all'altra del globo. Lo sforzo economico e tecnico necessario per mettere in piedi un sistema di conservazione digitale è così elevato da favorire la nascita di *outsourcers* o, se vogliamo, provider di servizi cui affidare la conservazione dei propri documenti. In tal caso, d'altronde, come essere certi che questi stiano svolgendo correttamente il proprio lavoro? La risposta a questa domanda ha portato alla nascita di schemi di certificazione, attraverso i quali sia possibile accertare le reali capacità sia di attuare quanto

promesso con i livelli di servizio offerti sia di far fronte agli onnipresenti rischi da parte del conservatore, nonché di valutare la sostenibilità del suo modello economico.

Sempre in questo capitolo vedremo come archiviare il web che appare come una necessità fondamentale visto che gran parte della conoscenza umana sta ormai transitando per siti, blog e social network.

La seconda parte del libro è dedicata alla sicurezza.

Il cap. 9 costituisce un'introduzione all'argomento e, in particolare, a cosa voglia dire esprimere dei requisiti di sicurezza e cosa sia il risk management.

Il cap. 10 affronta l'argomento dell'Identità nel mondo digitale che è centrale per l'autenticità dei documenti.

Nel cap. 11 parleremo invece di certificati e firme. L'Italia è stato uno dei primi Paesi al mondo ad affrontare il tema della firma digitale ma ha dovuto successivamente fare i conti con una direttiva europea il cui recepimento all'interno della normativa nazionale ha creato una situazione ingarbugliata in cui coesistono più tipi di firma, il cui uso non è sempre ovvio.

Il cap. 12 è dedicato alla PEC che è stata individuata come modalità privilegiata per la trasmissione di documenti con la pubblica amministrazione.

Infine, il cap. 13 affronta il problema della tutela della proprietà intellettuale quando riferita a oggetti digitali e della relativa tecnologia del *Digital rights management*.

La terza parte è un compendio di tecniche, strumenti e tecnologie che un architetto dell'informazione o un conservatore digitale usa normalmente.

Nel cap. 14 diamo uno sguardo alle tecniche di modellazione dell'informazione e, in particolare, ai diagrammi entità-relazione, ai diagrammi UML e a XML. Gran parte degli standard archivistici, soprattutto quelli relativi ai metadati, usa queste strumenti.

Il cap. 15 è invece un mini compendio di elementi relativi ai sistemi informativi che va dalle architetture di rete a quelle di storage, a SOA, alla virtualizzazione e fino al *cloud computing*. È una raccolta di argomenti necessari a comprendere le architetture tecniche dei centri di conservazione digitale.

Il panorama di argomenti trattati nel libro è decisamente ampio perché lo stesso argomento della conservazione digitale è complesso e tuttora irrisolto in molti aspetti. Per questo le migliori menti archivistiche, giuridiche

e informatiche a livello mondiale stanno cercando di portare velocemente risposte a problemi urgenti. È in gioco la stessa memoria umana e la capacità di trasmettere in modo efficace ai posteri le nostre conoscenze e la storia del nostro tempo.

Da parte nostra, ci auguriamo che questo lavoro di raccolta possa aiutare a facilitare la comprensione, la discussione e l'assimilazione degli argomenti relativi e possa contribuire alla formazione di nuove figure professionali con le giuste competenze.

*Alessandro Sinibaldi
Paolo Buongiorno*

Ringraziamenti

Questo lavoro non sarebbe stato possibile senza l'aiuto di amici e colleghi che hanno gentilmente messo a disposizione il loro tempo e la loro esperienza per utili confronti, spunti e suggerimenti e per la lettura delle bozze.

In particolare gli autori desiderano ringraziare Ruggero Rinaldi, Alessandra Aversa, Andrea Barulli, Paolo Verrecchia, Silvia Sinibaldi e Giovanni Ciccaglioni.

Parte prima

Informazioni e documenti

1. Documenti e archivi

Un **documento**, nella sua accezione più ampia, è un'informazione registrata su un supporto. In questo modo l'informazione viene resa trasmissibile nel tempo e nello spazio e questa capacità dipende strettamente dalle qualità di durevolezza e facilità di trasporto del supporto stesso. In realtà, come sarà ancora più chiaro quando parleremo di conservazione a lungo termine, ciò non è sufficiente per garantire l'intelligibilità dell'informazione, come è ben chiaro a un archeologo che trova in uno scavo una tavoletta di argilla con incisi caratteri in una lingua morta da millenni. In questo caso il supporto si è conservato ma, ahimè, l'informazione contenuta è incomprensibile. È quindi chiaro che il lavoro dell'archivista, inteso come *conservatore dell'informazione*, più che come conservatore del documento visto come mero supporto (che, a quel punto, è più lavoro di un magazziniere), va inserito in un *contesto* storico e culturale e comporta non solo la conservazione dell'informazione primaria ma anche di tutte quelle informazioni accessorie che ne possano garantire la corretta comprensione anche a distanza di molto tempo o anche semplicemente per uno studioso di una cultura diversa dalla nostra.

Capire e tramandare il contesto entro cui è stato prodotto un determinato documento è quindi un compito di primaria importanza. Se, quindi, cerchiamo di allargare la nostra visuale oltre i limiti fisici del supporto, la prima cosa che troviamo è ovviamente il *soggetto produttore* di quel documento, che può essere una persona fisica o una persona giuridica. Ecco che sorgono le prime domande: “perché è stato scritto quel documento? Cosa si cercava di conservare o tramandare?”, “è stato preceduto o seguito da altri documenti utili per capire il contesto in cui è stato scritto?”. Per