

Informatica

La morte della memoria (digitale)

di Nicola Bortolotti

Un allarme di enorme portata ha risuonato solo recentemente, nel gennaio di quest'anno, e in pochi giorni è rimbalzato sui media mondiali facendo, in pochi istanti, il giro di tutto il globo; un potente richiamo autorevolissimo, non inaspettato e ben più pervasivo di quanto riportato dai titoli dei giornali, quantunque sostanzialmente tardivo, se si considera il fatto che tali problematiche sono ben note da sempre agli addetti ai lavori e, anche limitandosi a questa rivista, sono state toccate – direttamente e indirettamente – almeno altre tre volte dal 2000 ad oggi.

È stata Lynne Brindley, proprio dal 2000 direttrice (*chief executive*) della prestigiosa biblioteca nazionale inglese, la British Library, docente universitaria, recentemente insignita del titolo di "*Dame of the British Empire*", a concentrare le ben riposte ansie e preoccupazioni che riguardano la nostra memoria, sempre più digitalizzata. La Brindley è stata giustamente incurante, dall'alto del suo ruolo istituzionale e dei suoi numerosi riconoscimenti accademici, delle possibili critiche che le sarebbero (e le sono) piovute addosso in quanto il suo appello non è inedito (ma la Brindley è sempre stata una delle più acute osservatrici del mondo dell'informazione digitale, ponendo sovente problemi inediti e fondamentali, anche in connessione col diritto d'autore) e non è nemmeno senza risposte (seppure parziali). La direttrice della British Library si è posta un problema nodale, che è stato riassunto dalla stampa globale in molti titoli simili e, quasi sempre, assai riduttivi.

“Wère in danger of losing our memories”

Qualche esempio: “*Is Our Future History Safe Within the Web?*” (La nostra storia futura è sicura all'interno del web?); “Memoria sul web a rischio”; “*Wère in danger of losing our memories*” (Siamo in pericolo di perdere le nostre memorie).

È proprio quest'ultimo titolo, tratto dall'articolo dell'*Observer* – settimanale del *Guardian* – scritto dalla stessa Brindley (<http://www.guardian.co.uk/technology/2009/jan/25/internet-heritage>), ad essere più rispettoso della portata (enorme) del “warning” lanciato dall'istituzione inglese, in quanto non cita la rete Internet, che potrebbe essere erroneamente considerata l'unico “luogo” ove possa avvenire la perdita di informazione o la causa stessa dei problemi; al contrario, la perdita dei dati diffusi sul web è solo un piccolo aspetto della problematica, tanto che un'opportuna raccolta dei dati tramite web ne potrebbe anzi rappresentare la possibile soluzione o – quantomeno – mitigazione.

Un allarme che viene da lontano

Per capire quanto il grido virtuale della Brindley sia tutto fuorché uno sterile arroccamento istituzionale su posizioni conservatrici (il che, tra l'altro, sarebbe in totale contrasto con la storia professionale del chief executive della British Library) è sufficiente una serie di banali ma calzanti esempi.

In un'epoca in cui la “dematerializzazione dei documenti” rappresenta un imperativo categorico, soprattutto per le amministrazioni pubbliche, si pensi

a un caso concreto: il classico esempio di un generico documento risalente a svariati anni prima; la situazione vale sia in ambito aziendale che familiare, in quanto può trattarsi di una pratica, di un contratto, di una scrittura privata, di un rogito, di una fattura, di una bolletta... Fermo restando il fatto che il documento cartaceo non sia stato cestinato, esso – una volta recuperato con maggiore o minore difficoltà – a seconda delle condizioni di conservazione sarà in tutto o in parte leggibile senza troppi problemi.

Si trasferisca, ora, la situazione allo stesso documento in formato elettronico. Il primo problema è il “dove” il documento sarà stato archiviato, perché un flusso di dati “smaterializzati”, ossia un flusso di bytes, per essere effettivamente memorizzato abbisogna comunque di un supporto “fisico”. Già in questa prima fase la situazione comincia a complicarsi non poco: in caso di memorizzazione “locale” (azienda, casa), infatti, si porrà dapprima il problema del supporto fisico: nastro magnetico? Floppy disk? ZIP? Hard disk fisso? Hard disk rimovibile? Magneto ottico? CD? DVD? Flash drive?

Il deterioramento dei supporti

Anche limitandosi ai citati supporti, i potenziali problemi sono già ingentissimi. Pur senza contare la probabilità di deterioramento dell’informazione memorizzata (che, in taluni casi, può essere elevatissima), vi è anche la questione legata all’obsolescenza del supporto utilizzato, che potrebbe essere sì potenzialmente integro ma sostanzialmente illeggibile a causa dell’assenza (o rarità) di dispositivi in grado di leggerlo; e, pur senza andare troppo lontano, è sufficiente un floppy disc da 5,25 pollici per mandare in crisi la maggior parte delle aziende... Da notare il fatto che le tecniche più robuste nei confronti delle ingiurie del tempo (quelle su nastro magnetico) siano anche quelle per le quali l’obsolescenza dei lettori sia stata più rapida.

Il deterioramento non va, infatti, sottovalutato: molti CD e DVD masterizzati – soprattutto se maldestramente conservati – risultano in parte o totalmente irrecuperabili già dopo pochi anni, contraddicendo nei fatti l’illusione di grande stabilità data dalla tecnologia ottica. Molti Hard Disk si guastano se lasciati inattivi per alcuni anni, ed hanno una vita comunque limitata ad alcune centinaia di migliaia di ore anche in caso di funzionamento continuativo. Le Flash Memory spesso cessano di funzionare per un semplice choc elettrico durante l’inserzione nella porta USB. I dischi ZIP erano famosi per soffrire del tristemente noto “*click of death*”. Si pensi alla

differenza con un fascicolo cartaceo, che può sopravvivere senza guasti per secoli e secoli...

La babele dei supporti

Il passo successivo, limitandosi a supporti “rimuovibili”, dunque pensando di avere superato anche problemi di interfaccia “elettrica” come quella degli hard disk, è legato al “file system” utilizzato per la memorizzazione permanente dei dati: non di rado, negli anni passati, hanno avuto grande diffusione, anche in supporti come floppy e CD, delle strutture di memorizzazione “proprietarie” (si pensi al mondo Apple Macintosh) che possono non essere correttamente riconosciute.

Infine, anche superando tutte queste barriere, quasi ci si trovasse in un film di avventura, si porrà al termine la questione forse più insidiosa: il “formato” del file sarà ancora decodificabile?

Anche in questo caso, la semplice “decodificabilità” nasconde, in realtà, un *mare magnum* di grattacapi. Per la memorizzazione delle informazioni potrà essere stato utilizzato un formato cosiddetto “proprietario” o, al contrario, uno “aperto”. Quasi sicuramente si tratterà del primo caso, poiché solo recentemente è nata una grande sensibilità nei confronti della necessità di “apertura” dei formati elettronici.

La babele dei formati

I dolori saranno pochi se il formato proprietario si sarà, nel frattempo, affermato come standard di mercato (come, ad esempio, il “doc” di Microsoft, che pure nasconde un elevato numero di sottotipi compatibili solo verso l’alto). Ma se, invece, ci si sarà affidati a una scelta meno fortunata (di aziende nel frattempo scomparse, ma anche di società leader che abbiano – nel frattempo – deciso, più o meno arbitrariamente, di abbandonarlo) il problema potrebbe anche risultare insormontabile, e questo anche limitandosi al caso “banale” di un semplice testo dattiloscritto. Ben diverso e ancor più complicato l’averne che fare con un file audio o video, o CAD, o di grafica ad oggetti...

Una volta che si sia riusciti ad “aprire” il file, potrebbero tuttavia sorgere altri impedimenti: in particolare, il documento potrebbe apparire assai diverso da quello che si ricorda essere l’originale per l’assenza o la modifica di alcuni “font” (caratteri), per l’impaginazione differente, per l’assenza di alcuni oggetti “embedded” (ossia contenuti all’interno ma che richiedono, per la corretta visualizzazione, di programmi esterni non più disponibili).

Tutto ciò senza prendere in esame un'eventualità ancor più bloccante, ossia la possibilità che – per buone o cattive ragioni – il file fosse stato protetto da una password andata, nel frattempo, perduta.

La babele delle protezioni

Un ulteriore problema, messo in risalto alcuni anni fa proprio dalla Brindley e dalla British Library anche andando coraggiosamente contro la “lobby” delle major discografiche e cinematografiche (si veda, ad esempio, <http://punto-informatico.it/1666777/PI/News/british-library-scomunica-questo-drm.aspx>), è legato alla presenza di tecnologie “DRM”, ossia di protezione del diritto d'autore, che potrebbero – anch'esse – rendere non fruibile il contenuto.

Il caso PDF

Anche formati assurti a fondamentale importanza nel mondo globale, come il PDF (Portable Document Format) di Adobe – sviluppato nel 1993, prima affermatosi come standard di mercato, poi divenuto aperto nel 2008 e, conseguentemente, incoronato standard ISO nello stesso anno (ISO 32000 http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51502) – non sono scevri da grattacapi in tal senso.

Prova ne sia che, come già segnalato in questa rubrica in passato, per una archiviazione “di lungo termine” che preservi intatte tutte le caratteristiche del documento atte a una visualizzazione corretta anche nel futuro, la specifica del formato PDF avesse avuto bisogno di una notevole “ripulitura” e revisione critica (<http://www.pdfa.org>) culminata nello standard PDF/A, standard ISO 19005 pubblicato nel 2005 (dopo tre anni di lavoro) e basato sulla versione 1.4 del PDF, quella di Adobe Acrobat 5.

La prima curiosità è costituita dal fatto che la standardizzazione del PDF/A preceda – non di poco – quella del PDF. Anche se la questione è legata in primis al tardivo rilascio come “open standard” delle specifiche PDF più recenti da parte di Adobe, il punto nodale risiede nella necessaria inerzia richiesta a standard di lungo periodo, il cui scopo è affidabilità e compatibilità, spesso non associabili alla rincorsa della “novità”.

Il PDF/A

La vera curiosità risiede invece sempre nella inerzia (ma, forse, sarebbe meglio scrivere indifferenza) – questa volta colpevole – con la quale – sia il legislatore sia i produttori di software indipendenti – han-

no accolto il formato PDF/A, “*un formato basato sul PDF che fornisce un meccanismo per rappresentare documenti elettronici in un modo che preservino nel tempo la loro veste visuale, indipendentemente dagli strumenti e dai sistemi utilizzati per creare, memorizzare e visualizzare i files*” (“*a file format based on PDF, known as PDF/A, which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files*”).

Il risultato è che, a tutt'oggi, anche le amministrazioni statali più avanzate (ad esempio il Ministero dell'Economia e delle Finanze) inviano in posta elettronica cedolini stipendio e CUD ai propri dipendenti – ad esempio quelli del comparto istruzione – in un formato che non è PDF/A compliant, come facilmente verificabile mediante uno qualsiasi dei tools utilizzabili gratuitamente in rete (<http://www.intarsys.de/produkte/pdf-a-live/pdf-a-check-1> valida on-line, mentre con il servizio <http://www.validatepdfa.com/it/> la risposta arriva via email). Quasi beffardamente, anche la scheda della Brindley presente sul sito ufficiale della Unione Europea, quantunque creata nel 2006, reperibile all'indirizzo

http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/cv/lynne_brindley.pdf, non ottempera allo standard PDF/A!

Il DPCM 10 dicembre 2008

Il legislatore italiano, pur tradizionalmente attento alle questioni legate al mondo digitale, ha preso una posizione netta e inconfutabile al proposito solo con il DPCM 10 dicembre 2008 (“Specifiche tecniche del formato elettronico elaborabile (XBRL) per la presentazione dei bilanci di esercizio e consolidati e di altri atti al registro delle imprese”) pubblicato sulla Gazzetta Ufficiale n. 304 del 31 dicembre 2008 laddove recita all'art 5 comma 4: “*l'interessato allega all'istanza di cui all'art. 4 e al bilancio elaborabile un ulteriore documento informatico contenente il bilancio approvato, in formato PDF/A senza immagini ottenute dalla scansione di documenti cartacei*” e all'art. 6 comma 3 “*Nelle more della definizione delle specifiche di cui al comma 1, l'interessato allega all'istanza di cui all'art. 4 un documento informatico in formato PDF/A con il contenuto dell'atto, anche senza immagini ottenute dalla scansione di documenti cartacei*”.

Il citato DPCM, per effetto domino, ha portato Infocamere (il primo motore di digitalizzazione della

vita aziendale italiana, tuttavia curiosamente poco attenta alla questione PDF/A) a correre ai ripari, anche consigliando esplicitamente l'utilizzo della suite open-source OpenOffice, più volte segnalata su queste pagine, a tutt'oggi l'unico strumento che consenta di produrre gratuitamente documenti in formato PDF/A.

Sia chiaro: un documento PDF (non PDF/A) – di per sé – non è detto che non sarà correttamente visualizzato tra qualche anno sulle macchine che saranno allora disponibili, ma non fornisce alcuna garanzia al proposito. Si pensi, allora, ad una amministrazione pubblica che dovesse trovarsi, tra qualche anno o decennio, ad avere archivi in formato elettronico più o meno inutilizzabili... Le conseguenze sarebbero facilmente immaginabili.

Se la questione è così delicata anche per un formato consolidato, “portabile” e affermato come il PDF, è facile capire cosa possa avvenire per la babele degli altri formati.

Delocalizzare non vuol dire risolvere

Ai problemi legati alla memorizzazione “locale” si è già accennato. Si potrebbe pensare che un servizio di conservazione dei dati “in rete”, su appositi spazi dotati di strategie di backup professionali, pur con una minore velocità di accesso e risolte le problematiche legate alla privacy, costituisca la risposta vincente, fatta salva la già discussa questione dei formati, che – comunque – non riguarda solo il mondo dell'informatica (si pensi agli archivi RAI) e non è un problema recente (il restauro delle foto lunari degli anni '50 e '60).

Anche la conservazione “in rete”, tuttavia, non è immune da grattacapi, e non è (né può essere) minimamente paragonabile alla silente conservazione di un documento – per decenni, se non secoli – in un tranquillo archivio cartaceo. Difficile pensare, ad esempio, a cosa sarà diventato un servizio gratuito come Gmail tra cinquant'anni: esisterà ancora? E attenzione: tutti i servizi online (anche quelli privi di costi) prevedono – ovviamente – la cancellazione dell'account dopo un periodo di inattività (tipicamente alcuni mesi, sei più tre nel caso di Google). Cosa ne sarebbe, dunque, di un archivio di una persona – ad esempio – deceduta per morte violenta? Pagine e pagine di letteratura, di epistolari e di opere inedite finirebbero nel nulla, nel devastante “buco nero” evocato dalla Brindley.

Sempre rimanendo al “web”, paradigmatico è il caso ricordato dalla direttrice della British Library e ripreso da tutti gli organi di stampa internazionali: “(...) oltre 150 siti riguardanti le Olimpiadi del

2000 a Sidney, sono svaniti istantaneamente al termine dei Giochi. È la memoria delle nazioni che se ne va”, ammonisce Lynne Brindley. Paradossalmente, mentre un antico documento scritto su carta di papiro può essere preservato per sempre da un museo o da una biblioteca, molti aspetti dell'informazione digitale, una volta cancellato un sito o diventato obsoleta una tecnologia, diventano irrecuperabili”

(<http://www.repubblica.it/2009/01/sezioni/tecnologia/web-memoria/web-memoria/web-memoria.html>, in parte traduzione letterale dell'articolo dell'Observer).

Vero è che esiste la cache di Google (fatto salvo il “diritto all'oblio”); vero è che esistono iniziative come archive.org (<http://www.archive.org>), ma tutto ciò è sicuramente non sufficiente se non ci sarà un impegno serio delle istituzioni all'uopo preposte, in primis le biblioteche. Il grido di allarme della Brindley (“*We are in danger of creating a black hole for future historians and writers*”) è, dunque, assai appropriato.

Il ruolo delle istituzioni

Non solo la British Library è attiva su questo fronte: a livello internazionale l'attenzione a queste tematiche è massima da almeno un quinquennio, come testimoniano le annuali Conferenze internazionali sulla conservazione degli oggetti digitali (*International Conference on Preservation of Digital Objects*, iPRES), la cui quinta edizione è stata ospitata l'anno scorso proprio in terra inglese (<http://www.bl.uk/ipres2008>).

La via da seguire è senz'altro quella indicata dalla Brindley: “*Lo sforzo di catturare il nostro patrimonio intellettuale on-line e preservarlo per lungo tempo ricade, giustamente, sulle stesse biblioteche e archivi che, nel corso dei secoli, hanno sistematicamente raccolto libri, periodici, giornali e registrazioni che rimangono disponibili per sempre, proprio grazie a queste istituzioni*”.