

AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

Fondata nel 1983 da Paolo BISOGNO

N. 1-2 — Anno 35 — gennaio–giugno 2017

Proprietario della rivista
UNIVERSITÀ DELLA CALABRIA

Direttore Scientifico
Roberto GUARASCI
Università della Calabria

Direttore Responsabile
Fabrizia Flavia SERNIA

Comitato scientifico

Roberto Guarasci, UNIVERSITÀ DELLA CALABRIA
Anna Rovella, UNIVERSITÀ DELLA CALABRIA
Maria Guercio, SAPIENZA UNIVERSITÀ DI ROMA
Giovanni Adamo, CONSIGLIO NAZIONALE DELLE RICERCHE
Claudio Gnoli, UNIVERSITÀ DEGLI STUDI DI PAVIA
Ferruccio Diozzi, ASSOCIAZIONE ITALIANA DOCUMENTAZIONE AVANZATA
Gino Roncaglia, UNIVERSITÀ DELLA TUSCIA
Laurence Favier, UNIVERSITÉ CHARLES-DE-GAULLE LILLE 3
Madjid Ihadjadene, UNIVERSITÉ VINCENNES-SAINT-DÉNIS PARIS 8
Maria Mirabelli, UNIVERSITÀ DELLA CALABRIA

Comitato di redazione

Antonietta Folino, UNIVERSITÀ DELLA CALABRIA
Erika Pasceri, CONSIGLIO NAZIONALE DELLE RICERCHE
Maria Taverniti, CONSIGLIO NAZIONALE DELLE RICERCHE
Maria Teresa Chiaravalloti, CONSIGLIO NAZIONALE DELLE RICERCHE
Assunta Caruso, UNIVERSITÀ DELLA CALABRIA

Segreteria di Redazione

Valeria Rovella, UNIVERSITÀ DELLA CALABRIA

ARACNE

AIDAinformazioni

RIVISTA SEMESTRALE

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti le Scienze dell'Informazione, la Documentazione, la Gestione Documentale e l'Organizzazione della Conoscenza. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria.

La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi possono riguardare *topics* quali Documentazione, Scienze dell'informazione e della comunicazione, Scienze del testo e del documento, Organizzazione e Gestione della conoscenza, Terminologia, Statistica testuale e Linguistica computazionale e possono illustrare studi sperimentali in domini specialistici, casi di studio, aspetti e risultati metodologici conseguiti in attività di ricerca applicata, presentazioni dello stato dell'arte, ecc.

«AIDAinformazioni» è censita dall'ANVUR per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali, così come dall'AERES (Agence d'évaluation de la recherche et de l'enseignement supérieur) che la annovera tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE – Bielefeld Academic Search Engine; EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; Letteratura Professionale Italiana – Associazione Italiana Biblioteche; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Summon™ – by SerialsSolutions; Ulrich's; UniCat – Union Catalogue of Belgian Libraries; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti dal comitato scientifico sono inviati in forma anonima a due *referee*, selezionati sulla base della loro comprovata esperienza nei *topics* specifici del contributo in valutazione.

Condizioni di acquisto

(spese di spedizione in Italia incluse)

Abbonamento annuale euro 42,00

Fascicolo euro 26,00

Per ordini

telefono/fax: 06 45551463

e-mail: info@giocchinoonoratieditore.it

online: <http://www.aracneeditrice.it/>

Skype: aracneeditrice

Modalità di pagamento

Bonifico bancario intestato a:

Gioacchino Onorati editore S.r.l. unip.

IBAN: IT 28 B 03069 38860 10000003170

presso Banca Intesa Sanpaolo

(filiale di Ariccia)

Causale: *Abbonamento «AIDAinformazioni»*

Editore

Gioacchino Onorati editore S.r.l. – unip.

via Vittorio Veneto, 20

00020 Canterano

(06) 45551463

www.aracneeditrice.it

info@aracneeditrice.it

info@giocchinoonoratieditore.it

Copyright ©

Gioacchino Onorati editore S.r.l. – unip.

ISBN 978-88-255-0627-3

edizione a stampa ISSN 1121-0095

edizione elettronica ISSN 1594-2201

Indice

- 5 Editoriale
La riscoperta affascinante del passato lontano o recente attraverso
le scienze documentali
Fabrizia Flavia Sernia

Contributi

- 13 Alla ricerca della soluzione definitiva al problema dell'obsolescenza dei formati elettronici. Lo standard ISO 19005-3 (PDF/A-3)
Stefano Allegrezza
- 33 Strumenti e terapie nelle cure psichiatriche: estratti dall'archivio storico dell'ospedale psichiatrico di Girifalco
Maria Teresa Chiaravalloti, Maria Taverniti
- 49 Politiche e regolazione nel cambiamento del centro storico di Cosenza
Antonella Coco
- 71 Analyser l'autorité dans les publications scientifiques
Evelyne Broudoux
- 83 Un'architettura scalabile per la real-time streaming analytics. Un'applicazione per l'Energy Management & Forecasting
Angelo Martone, Gaetano Zazzaro, Gianpaolo Romano

- 103 La citazione giurisprudenziale nella sentenza amministrativa come
atto comunicativo e cognitivo
Renato Rolli, Angelo Costa

Note e Rubriche

- 123 David Hume, *Contro Rousseau*, a cura di Spartaco Pupo, Bietti,
Milano 2016
Annabella d'Atri
- 129 *Discutere di futuro*, a Napoli. Un approccio e un convegno
Ferruccio Diozzi
- 131 Autodefinizione di un dominio. L'Enciclopedia della ko
Claudio Gnoli
- 137 L'Archivio Comunale dall'unità al 1897. Note in margine al volume
di Dimitri Brunetti
Roberto Guarasci
- 141 Terminologia e internazionalizzazione
Maria Teresa Zanola

La riscoperta affascinante del passato lontano o recente attraverso le scienze documentali

C'era una volta il *floppy disk*

FABRIZIA FLAVIA SERNIA*

C'era una volta il floppy disk, la cui “scomparsa” insieme a tanti altri tipi di supporti per la conservazione dei documenti ha creato problemi inaspettati e aperto nuovi campi di indagine. C'era una volta il centro di Cosenza, con una fisionomia diversa da quella attuale, che si è modificato per tutta una serie di circostanze esogene ed endogene del territorio. E c'era, una volta, anche l'ospedale psichiatrico di Girifalco, un piccolo paese della Calabria in provincia di Catanzaro che dal 22 luglio 1881 fino al 15 maggio 1978 ospitò quasi 15.800 pazienti, le cui pene, cure e terapie sono custodite nelle cartelle cliniche dell'archivio storico oggi ospitato nell'ex convento dei Padri Riformati di Girifalco. C'era una volta... il fascino delle scienze documentali consiste nel regalare inaspettate correlazioni fra domini di investigazione anche platealmente distanti, come si verifica in questo numero di «AIDainformazioni» che offre orgogliosamente ai suoi lettori il risultato di sei ambiziosi progetti scientifici, dalla forte valenza trasversale in svariati ambiti professionali e sociali. Fra questi, tre in particolare ci accompagnano nella lettura sia di un passato recente, come accade con i lavori di Stefano Allegrezza e di Antonella Coco, sia di uno più remoto, come fanno Maria Teresa Chiaravalloti e Maria Taverniti, evidenziando tutti i chiaroscuri attraverso i quali si è giunti al presente. Stefano Allegrezza, dell'Università degli Studi di Udine, ha scandagliato il tema della *Ricerca della soluzione definitiva al problema dell'obsolescenza dei formati elettronici: lo standard ISO 19005-3 (PDF/A-3)*. Il nodo della conservazione a lungo termine per l'archiviazione di varie tipologie documentarie riconduce alla minaccia, sempre in agguato, di non poter avere più possibilità di accesso, a distanza di anni, ai dati memorizzati secondo i formati elettronici in uso in epoche precedenti. Allegrezza cita l'esempio dell'antenato degli attuali programmi

* Giornalista scientifica.
fabrizia.sernia@gmail.com

di elaborazione testi, *Wordstar*. I documenti editati venivano memorizzati su *floppy disk* da 5 pollici e $\frac{1}{4}$ o da 3 pollici e $\frac{1}{2}$. Oggi, dice Allegrezza, se volessimo visualizzarne il contenuto, l'operazione non sarebbe semplicissima, dal momento che occorrerebbe «trovare degli strumenti di conversione che consentano di trasformare quei documenti in altri oggi accessibili». Quanti di noi hanno vissuto questa incombenza! Chi scrive ne ha avuto esperienza diretta nel tentativo di migrazione, non riuscita, da *floppy disk* a PC, della tesi del Master di specializzazione, datata fine anni Ottanta. Mantenendo il ritmo di un racconto investigativo, avvincente e inusuale per un lavoro scientifico, Stefano Allegrezza sottolinea come la comunità archivistica concordi sul punto di adottare, «fin dal momento della formazione di documenti digitali, formati elettronici che possiedano caratteristiche tali da fornire garanzie in termini di conservazione a lungo termine», citando casi, anche clamorosi, di perdita di documenti digitali da parte di enti e istituzioni. Emblematica, sottolinea l'autore, la vicenda che riguardò la perdita dei dati raccolti dalle due sonde Viking che la NASA inviò su Marte nel 1975, al fine di scoprire l'eventuale presenza di forme di vita. I dati furono registrati su nastri magnetici che, alla fine degli anni Novanta, quando se ne decise il riutilizzo, si mostrarono parzialmente deteriorati, nonostante le precauzioni per la corretta conservazione adottate dall'Agenzia spaziale americana. Per di più, nel momento in cui gli scienziati tentarono di riesaminare alcune di quelle informazioni, scoprirono che non erano in grado di decodificare i formati elettronici utilizzati. Soltanto grazie a vecchie stampe su carta fu possibile recuperare i dati, che tuttavia furono interamente ridigitati. Passando per l'aspetto della *migrazione* o *riversamento sostitutivo* dei contenuti digitali, Stefano Allegrezza presenta quindi la soluzione definitiva all'obsolescenza dei formati elettronici — il formato PDF/A-3 — esplorandone tutte le peculiarità e i vantaggi. Altri due lavori pubblicati in questo numero gettano un ponte fra passato e presente. L'analisi e la "lettura" del centro storico di una città come Cosenza di Antonella Coco, dell'Università della Calabria, condotta in *Politiche e regolazione nel cambiamento del Centro storico di Cosenza* e quella di una realtà di ricovero e cura per le malattie mentali del passato, l'ex ospedale psichiatrico di Girifalco, che Maria Teresa Chiaravalloti e Maria Taverniti, dell'IRT del CNR, presentano in *Strumenti e terapie nelle cure psichiatriche: estratti dall'archivio storico dell'ospedale psichiatrico di Girifalco*. Antonella Coco indaga sul cambiamento che ha contrassegnato il centro storico di Cosenza e la fisionomia dei suoi abitanti, focalizzandosi sul ruolo svolto dalle politiche locali, al crocevia fra modernità e salvaguardia delle tradizioni. Maria Teresa Chiaravalloti e Maria Taverniti ci conducono in un ambito "pieno di mistero e leggenda", come è quello di un ex ospedale psichiatrico, grazie all'opportunità offerta dal progetto d'interesse "Invecchiamento: innovazioni tecnologiche e molecolari per un miglioramento della salute dell'anziano",

condotto in collaborazione con il Centro regionale di Neurogenetica di Lamezia Terme (cz). Fra gli obiettivi del progetto, spiegano le due ricercatrici, figurava la digitalizzazione delle cartelle cliniche dell'ex ospedale psichiatrico di Girifalco (cz), «al fine di utilizzare la memoria storica per rintracciare aspetti antesignani delle patologie dementigene e ricostruirne, laddove possibile, un profilo di ereditarietà». La scansione e la parziale trascrizione delle oltre 5200 cartelle cliniche, riferite al periodo 1881-1931, ha permesso di indagare su molti aspetti, «fra i quali la terminologia in uso per gli strumenti e le terapie usate nella pratica quotidiana delle cure psichiatriche all'interno del manicomio». Il quadro che emerge delle cure e dei trattamenti sanitari cui erano sottoposti i pazienti ricorda alcuni film *noir* dei primi Novecento. Grazie all'analisi terminologica condotta dalle due autrici, si perviene a uno spaccato sia degli strumenti in uso, sia delle terapie adottate sul finire dell'Ottocento per le patologie mentali. Morfina, iniezioni ipodermiche, idroterapia, Marsala, Idrato di corallo e corrente elettrica, compasso di Weber, ago o spillo ricorrono nelle cartelle cliniche, analogamente al termine "camicia di forza". Dalla documentazione medica emerge come fossero in uso prevalentemente «metodi di contenimento di vario tipo, sia mediante sostanze sedative svariate, sia mediante soluzioni fisiche come la camicia di forza», a discapito di terapie finalizzate alla cura vera e propria della malattia. La lettura di questo lavoro presenta aspetti toccanti. Viene spontaneo pensare a quanta sofferenza, dolore e, talvolta, menomazioni o morte procurate dalle stesse terapie, abbiano celato questi luoghi di cura, in uno scenario che oggi appare inimmaginabile. Tuttavia, non soltanto grazie alla legge Basaglia nel 1978 gli ospedali psichiatrici sono stati chiusi, ma anche, nel corso degli anni, in tutto il perimetro delle scienze mediche, non soltanto quelle neurologiche, è cresciuta la collaborazione fra gli specialisti. In tutti i campi, la scienza medica oggi considera il malato nella sua centralità di persona che soffre, il cui benessere psicologico è fondamentale per il migliore esito delle cure. Le iniziative meritorie che alcuni ospedali promuovono per il benessere psicologico del paziente, grande o piccolo che egli sia, sono un emblema di civiltà e una testimonianza di umanità. L'iniziativa "Curarsi ad arte" del Gruppo Multimedita di Milano in collaborazione con l'Accademia di Brera, prevede un percorso espositivo di cura e cultura in cui immergersi camminando nell'Ospedale San Giuseppe. La cineterapia per pazienti adulti e bambini, promossa e realizzata grazie alla onlus Medicinema Italia e a Walt Disney Company Italia, lo scorso anno presso il Policlinico universitario Gemelli di Roma e quest'anno al Grande Ospedale Metropolitano Niguarda di Milano, prevede la realizzazione di grandi sale cinematografiche HD sensoriali all'interno dei nosocomi, raggiungibili da ogni parte della struttura ospedaliera, dove utilizzare anche vibro-acustica a basse frequenze quale innovazione di intervento terapeuti-

co. Parallelamente, si implementa lo sviluppo di uno *Studio sulla terapia del sollievo con il cinema*, condotto in collaborazione con il Policlinico Gemelli, di cui sono stati presentati nello scorso aprile i primi risultati.

Il numero di «AIDAinformazioni» non si esaurisce qui. La realizzazione di una piattaforma software di *Big Data Analytics* di supporto alle decisioni strategiche di enti, aziende, professionisti e comunità, che punta a offrire soluzioni di approcci innovativi nel settore delle politiche energetiche, in rapporto alle politiche pubbliche, è al centro del contributo scientifico di Angelo Martone, Gaetano Zazzaro e Gianpaolo Romano del CIRA, Centro Italiano Ricerche Aerospaziali, dal titolo *Un'architettura scalabile per la real time streaming analytics. Un'applicazione per l'Energy Management & Forecasting*. I tre ricercatori si soffermano sull'architettura progettata e implementata «per il monitoraggio, l'analisi e la correlazione, in tempo reale, di fenomeni fisici e/o parametri ricevuti da stream di dati eterogenei, come le reti di sensori wireless ed i *web services* nel dominio dell'*Energy Management*», per la cui peculiarità dei dati sono state utilizzate tecniche appartenenti al paradigma dei *Big Data*. Il tema affrontato in quest'ultimo lavoro di ricerca è molto più vicino alla nostra quotidianità di quanto non si immagini, per gli effetti che ragionevolmente dispiega sulla mitigazione dei cambiamenti climatici globali. La progettazione e lo sviluppo di architetture basate sull'analisi e il monitoraggio di parametri fisici e chimici ambientali, sebbene su scala assai più ridotta rispetto a quella presentata nel progetto, è un'attività su cui si stanno sviluppando svariate start-up di supporto all'attività agricola di precisione. Nel recente *Food Innovation Summit*, svoltosi a Milano nello scorso maggio per celebrare l'eredità di Expo2015, il ricercatore Paolo Bassi, responsabile Ricerca e Sviluppo della start up innovativa Auroras, incentrata sull'IOT in soluzioni verticali per l'agricoltura di precisione, ha spiegato gli effetti dell'utilizzo concreto di una tecnologia implementata *ad hoc* dalla società. Sensori capaci di rilevare sul campo parametri quali umidità e temperatura solare, trasmettono i dati in tempo reale ad un data base che li elabora, consentendo all'agronomo o al coltivatore di ricevere sul proprio dispositivo mobile un report con tutte le indicazioni sullo stato di salute della coltivazione, quale strumento di supporto alle decisioni, per valutare il ricorso o meno al trattamento antiparassitario per la coltivazione.

Ancora un cambio di registro, passando dall'ambito energia e ambiente a quello legale, con il lavoro scientifico di Renato Rolli e Angelo Costa dell'Università della Calabria, *La citazione giurisprudenziale nella sentenza amministrativa come atto comunicativo e cognitivo*. Si tratta del primo studio, in campo scientifico, che analizza la citazione giurisprudenziale nella sentenza amministrativa come momento di comunicazione giuridica e di cognizione. Infine, Evelyne Broudoux, del DICEN-IDF-CNAM di Parigi ha indagato sulle *Modalità di analisi dell'autorevolezza (autorità) nelle pubblicazioni scientifiche*,

ponendosi l'obiettivo di testare un modello di apprezzamento dei lavori a vocazione scientifica che potrebbe servire sia agli autori, sia ai valutatori della comunità scientifica *prima* della pubblicazione, con ciò affrontando un aspetto congenito del lavoro specifico della ricerca.

CONTRIBUTI

Alla ricerca della soluzione definitiva al problema dell'obsolescenza dei formati elettronici

Lo standard ISO 19005–3 (PDF/A–3)

STEFANO ALLEGREZZA*

ABSTRACT: With the publication of the ISO 19005–3:2012 standard on October 15, 2012, the International Organization for Standardization (ISO) officially recognized the PDF/A–3 file format as the new standard for long-term digital preservation. This file format seems to be a good candidate as the definitive format, suitable for storing and long-term preservation of various types of documents. However, it has not failed to give rise to a broad debate; on the one hand, it has been widely accepted and has been proposed as the definitive file format for the preservation of digital documents, on the other hand it has been criticized and has been seen as an evolution of PDF/A format that betrays the original intent that led to its proposal. A few years after the publication of the standard, this article wants to conduct a smooth and balanced analysis of the format, exploring its real potential and trying to figure out if and to what extent it can contribute to the solution of at least one of the digital preservation problems, that of obsolescence of electronic formats.

Keywords: Digital preservation, File formats, Migration, Obsolescence, PDF/A.

I. Introduzione

Con la pubblicazione della norma ISO 19005–3:2012, *Document management — Electronic document file format for long-term preservation — Part 3: Use of ISO 32000–1 with support for embedded files (PDF/A–3)* avvenuta il 15 ottobre 2012, l'International Organisation for Standardization (ISO) ha ufficialmente riconosciuto il formato PDF/A–3 come il nuovo standard per la conservazione digitale a lungo termine. Questo formato sembra avere tutte le carte in regola per proporsi come il formato definitivo, adatto sia per l'archiviazione di svariate tipologie documentarie (dai documenti di testo, ai fogli di calcolo, agli elaborati grafici, ecc.) sia per la loro conservazione a lungo termine. Tuttavia, esso non ha mancato di dare origine ad un vasto dibattito; se da una parte ha riscosso un ampio consenso ed è stato suggerito come il

* Università degli Studi di Udine, Udine, Italia.
stefano.allegrezza@uniud.it

formato definitivo per la conservazione dei documenti digitali (qualcuno ha parlato di una svolta nella conservazione digitale), dall'altra ha ottenuto invece delle critiche ed è stato visto come una evoluzione del formato PDF/A che sembra tradire lo spirito originario del formato, ovvero la sua capacità di garantire la conservazione nel tempo. A distanza di qualche anno dalla pubblicazione dello standard, è giunto il momento di condurre un'analisi pacata ed equilibrata del formato, esplorando le sue reali potenzialità e cercando di capire se ed in quale misura possa contribuire alla soluzione di almeno un problema della conservazione digitale, quello dell'obsolescenza dei formati elettronici.

2. Il problema dell'obsolescenza dei formati elettronici

Come è noto, una delle principali criticità che minacciano la conservazione a lungo termine dei documenti digitali è quella dell'obsolescenza dei formati elettronici. Sono numerosi i casi di perdita di contenuti digitali che si sono verificati a causa sua. Molti ricorderanno che solamente una trentina di anni fa il *software* di elaborazione testi più utilizzato era *Wordstar*, un programma con interfaccia a carattere (nulla a che vedere con l'interfaccia grafica degli attuali programmi) grazie al quale era possibile produrre documenti di testo in un formato proprietario identificato dall'estensione *.ws*. Le funzionalità di *Wordstar* non sono neanche lontanamente paragonabile a quelle dei moderni programmi di elaborazione testi, ma per l'epoca si trattava comunque di un programma assolutamente avveniristico, soprattutto se si tiene conto del fatto che fino ad allora lo strumento più utilizzato per redigere testi in maniera non manuale era costituito dalla macchina da scrivere. Viene spontaneo domandarsi: se si avessero ancor oggi dei documenti di testo in quel formato, memorizzati magari sui *floppy disk* in uso all'epoca (*floppy disk* da 5 pollici e $\frac{1}{4}$ o da 3 pollici e $\frac{1}{2}$), sarebbe ancora possibile visualizzarne il contenuto? La risposta non è delle più confortanti: anche ipotizzando che i supporti siano ancora in buone condizioni di conservazione, sicuramente si incontrerebbero serie difficoltà per accedere a tali contenuti dal momento che il formato *.ws* utilizzato dal programma *Wordstar* non è più gestito dagli attuali programmi di elaborazione testi (ad esempio, *Microsoft Word*) ed è difficile trovare strumenti di conversione che consentano di trasformare quei documenti in altri oggi accessibili. Per risolvere questo genere di problema sono state proposte diverse strategie, ma la comunità archivistica è concorde nel ritenere che il punto di partenza sia costituito dall'adozione, fin dal momento della formazione dei documenti digitali, di formati elettronici che possiedano caratteristiche tali da fornire concrete

garanzie in termini di conservazione a lungo termine¹. Non si tratta, tuttavia, di una scelta semplice se è vero che numerosi sono i casi di perdita di documenti digitali che riguardano enti ed istituzioni che dovrebbero disporre delle risorse tecnologiche ed organizzative per evitare qualsiasi tipo di disastro². Uno dei casi certamente più conosciuti è quello che ha riguardato la perdita dei dati scientifici raccolti durante la spedizione spaziale delle sonde Viking su Marte. Nel 1975 la NASA inviò due sonde Viking sul pianeta rosso con l'obiettivo di scoprire se vi fossero forme di vita; all'epoca non si aveva alcuna coscienza delle problematiche legate alla conservazione digitale e si diede per scontato che i dati scientifici accuratamente raccolti dagli scienziati sarebbero rimasti disponibili per le future generazioni di scienziati semplicemente registrandoli su nastri magnetici. Purtroppo non fu così, perché solamente qualche decennio più tardi, alla fine degli anni Novanta, nonostante gli sforzi dell'agenzia spaziale americana di mantenere i nastri in ambienti con temperatura ed umidità controllata, essi si erano già parzialmente deteriorati. Ma l'aspetto più preoccupante della vicenda è un altro: quando gli scienziati tentarono di riutilizzare alcuni di quei dati, scoprirono che non erano più in grado di decodificare i formati elettronici utilizzati. Alla fine riuscirono a rintracciare delle vecchie stampe su carta e fu così possibile recuperare i dati, ma fu necessario ridigitare tutto³.

Un altro esempio famoso⁴ è quello del Censimento USA del 1960, i cui dati erano stati archiviati su nastri magnetici (di tipo UNIVAC III-A e II-A, o compatibili); quando, alla fine degli anni Settanta, si decise di migrare quei dati su supporti di memorizzazione più moderni, i nastri UNIVAC — e con essi le unità a nastro in gradi di leggerli — erano già diventati obsoleti, il che ha reso il progetto una vera e propria sfida tecnologica. Nel 1996 erano rimaste al mondo solo due macchine in grado di leggere quei dati; gran

1. Le caratteristiche che un formato elettronico deve possedere per essere compatibile con un processo di conservazione digitale sono state oggetto di approfonditi studi scientifici; possiamo ricordare quelle di non proprietà, apertura, standardizzazione, trasparenza, robustezza, stabilità, auto-contenimento, auto-documentazione, indipendenza dal dispositivo, assenza di meccanismi tecnici di protezione, assenza di limitazioni sull'utilizzo, ecc. Per una disamina approfondita della questione si rimanda a S. ALLEGREZZA, S. PIGLIAPOCO, *Produzione e conservazione del documento digitale. Requisiti e standard per i formati elettronici*, Edizioni EUM, Macerata 2008.

2. Si veda il sito <<http://www.atlasofdigitaldamages.info>> (ultima consultazione: 18/05/2017), che riporta una lista di casi di perdita di dati digitali che hanno riguardato amministrazioni pubbliche o grandi istituzioni (dal caso appena citato delle sonde Viking al sito web dell'amministrazione Clinton). Il sito è gestito da Barbara Sierman, Digital Preservation Manager presso The National Library of the Netherlands.

3. Si veda, M. WALLER, R. SHARE, *Mind the gap. Assessing digital preservation needs in the UK*, Digital Preservation Coalition, 2006, <<http://www.dpconline.org/docman/miscellaneous/advocacy/340-mind-the-gap-assessing-digital-preservation-needs-in-the-uk/file>> (ultima consultazione: 18/05/2017).

4. Si veda *Conservazione delle memorie digitali rischi ed emergenze. Sei casi di studio*, a cura di A. Ruggiero, <<http://www.iccu.sbn.it/upload/documenti/emergenze.pdf>> (ultima consultazione: 18/05/2017).

parte di essi fu recuperata dai nastri magnetici e convertita in un formato più moderno, ma per ricostruire il censimento nella sua interezza fu necessario recuperare i dati grezzi conservati su microfilm (facendo ricorso anche in questo caso ad informazioni di tipo analogico, che, nel frattempo, si erano conservate perfettamente).

Un ultimo ed esplicativo esempio riguarda il *Domesday Book*, il primo esempio di “catasto” in Europa, fatto compilare da Guglielmo il Conquistatore nel 1086 per censire i propri possedimenti in Inghilterra ed assoggettarli al pagamento delle tasse. Si tratta di un documento di 913 pagine che censisce oltre 13000 località dell’Inghilterra e del Galles. Nel 1983, la BBC e il governo britannico pensarono di celebrare il novecentenario del *Domesday Book* avviando un progetto per la realizzazione di una versione multimediale che avrebbe dovuto fotografare l’Inghilterra come appariva agli occhi di un inglese negli anni ’80⁵. Il “moderno” BBC *Domesday Book* venne pubblicato nel novembre 1986: era contenuto in due *laserdisc* (il *Community Disc* e il *National Disc*)⁶ leggibili mediante un computer BBC *Master* modificato con l’aggiunta di diversi componenti hardware speciali (una scheda SCSI, un coprocessore, una trackball e un tipo particolare di lettore di *Laserdisc* detto LVROM): l’intero apparato costava 5.000 sterline, una cifra certamente non alla portata di tutti, tanto che ne vennero venduti soltanto un migliaio. All’epoca, il BBC *Master* era un computer molto diffuso in Gran Bretagna e in Australia, ma nemmeno quindici anni dopo, il “moderno” *Domesday Book*, costato 2,5 milioni di sterline, era già illeggibile: i particolari lettori di *laserdisc* e i microcomputer necessari per interpretare i formati elettronici nei quali erano stati codificati i contenuti presenti su tali supporti non esistevano più⁷. Di recente il *Domesday Book* “digitale” è stato recuperato e trasferito *in extremis* su computer moderni grazie all’eroico lavoro di un

5. Il BBC *Domesday* è stato uno dei più importanti prodotti interattivi del suo tempo ed è stato realizzato su un’ampia scala, come mai si era visto fino ad allora. I due *laserdisc* contenevano più di 250.000 luoghi geografici, 25.000 mappe, 50.000 immagini, 3.000 serie di dati, 60 minuti di video e un numero imprecisato di parole. Più di un milione di persone, tra studenti e ricercatori provenienti da tutto il paese, hanno preso parte a questo ambizioso progetto; si narra che sarebbero necessari più di sette anni per visionare l’intero contenuto dei dischi. Sebbene sia stato un grande insuccesso commerciale, chi lo ha realizzato lo ha ritenuto un trionfo per quanto concerne la progettazione interattiva.

6. La differenza fondamentale tra i due sta nel fatto che nel *Community Disc* le informazioni sono organizzate su base geografica e mostrano la Gran Bretagna come appariva agli occhi degli abitanti dell’epoca, mentre nel *National Disc* le informazioni sono organizzate per argomenti e forniscono una panoramica generale della nazione. Il *National Disc* raccoglie un’ampia varietà di dati statistici, immagini del Regno Unito e della sua cultura in una serie di fotografie, passeggiate virtuali (un primo esempio di realtà virtuale) e una gran quantità di video.

7. Il *Domesday Book* originale, quello del 1086, è, invece, ancora perfettamente leggibile dopo quasi mille anni, e senza alcun requisito hardware (basta conoscere il latino, le cui “specifiche tecniche” sono ben documentate).

gruppo di ricercatori⁸. Grazie ad un ulteriore intervento di recupero che ha avuto luogo in tempi ancor più recenti, a distanza di 25 anni dalla sua pubblicazione, l'intero contenuto del *Community Disc* è stato recuperato e reso disponibile direttamente sul web⁹.

3. La strategia della migrazione

La strategia di conservazione digitale che è stata finora proposta per contrastare il problema dell'obsolescenza dei formati è quella della migrazione (*migration*) o "riversamento sostitutivo"¹⁰; essa consiste (v. Fig. 1) nella conversione di un documento digitale da un formato elettronico fruibile in un determinato ambiente tecnologico (*hardware* e *software*) che sta diventando obsoleto in un formato elettronico fruibile in un altro ambiente, non obsoleto e di solito temporalmente successivo rispetto al primo. Si tratta di una strategia che è necessario mettere in atto subito non appena i cambiamenti nell'*hardware* e nel *software* cominciano a minacciare la possibilità di fruizione di un documento digitale perché il formato secondo cui esso è codificato inizia a diventare obsoleto, e che va ripetuta nel tempo, almeno fino a quando si vuole conservare quel documento¹¹. Dal momento che il riversamento sostitutivo comporta una modifica nella sequenza di *bit* che co-

8. Il recupero rientra in un più ampio studio sulle strategie di conservazione del patrimonio digitale basato sull'emulazione portato avanti dal progetto CAMILEON (*Creative Archiving at Michigan and Leeds Emulating the Old on the New*) il quale ha utilizzato il *BBC Domesday Book* per verificare l'affidabilità delle strategie di emulazione sviluppate.

9. Si veda, I. MACKENZIE, *Domesday Project reborn online after 25 years*, in "BBC News", disponibile all'indirizzo <<http://www.bbc.co.uk/news/technology-13367398>> (ultima consultazione: 18/05/2017).

10. La Deliberazione del Centro nazionale per l'informatica nella pubblica amministrazione (CNIPA) n. 11 del 19 febbraio 2004 *Regole tecniche per la riproduzione e conservazione di documenti su supporto ottico idoneo a garantire la conformità dei documenti agli originali — Art. 6, commi 1 e 2, del testo unico delle disposizioni legislative e regolamentari in materia di documentazione amministrativa, di cui al decreto del Presidente della Repubblica 28 dicembre 2000, n. 445*, ora abrogata, definiva il riversamento sostitutivo come il «processo che trasferisce uno o più documenti conservati da un supporto ottico di memorizzazione ad un altro, modificando la loro rappresentazione informatica». Accanto a questa strategia menzionava quella del riversamento diretto definita come il «processo che trasferisce uno o più documenti conservati da un supporto ottico di memorizzazione ad un altro, non alterando la loro rappresentazione informatica».

11. Ad esempio, per conservare fino ad oggi un documento di testo realizzato alla fine degli anni Ottanta con *Wordstar*, sarebbe stato necessario un primo riversamento sostitutivo nel formato di *WordPerfect* agli inizi degli anni Novanta; poi altri due riversamenti sostitutivi, prima nel formato di *Microsoft Word 2.0* e poi in quello di *Microsoft Word 6.0/95* a metà degli anni Novanta; successivamente un riversamento sostitutivo nel formato di *Microsoft Word 97/2003* alla fine degli anni Novanta; infine, un ulteriore riversamento sostitutivo nel formato di *Microsoft Word 2007 (.docx)* nel 2007; e così via. La serie di riversamenti sostitutivi che occorrerà eseguire su un determinato documento digitale per garantirne la conservazione nel tempo non avrà termine fintantoché lo si vorrà conservare.

stituisce il contenuto digitale, il formato di destinazione deve assicurare che il contenuto informativo e le proprietà significative (*significant properties*)¹² del contenuto digitale siano mantenute salvaguardandone nel contempo l'autenticità e l'integrità¹³. Così come il riversamento diretto, anche il riversamento sostitutivo è un processo molto critico nei confronti del tempo: non ammette alcuna dilazione e deve essere eseguito prima che il corrente formato divenga obsoleto. Se viene "saltata" una generazione di formati, i documenti digitali potrebbero essere difficili da recuperare; se vengono perse più generazioni, il loro recupero potrebbe diventare impossibile. I cicli di migrazione devono essere relativamente frequenti, dal momento che pochi documenti digitali sono oggi in grado di sopravvivere più di 5-10 anni senza essere sottoposti ad alcuna operazione di migrazione¹⁴.

La strategia di conservazione digitale basata sulla migrazione incontra serie difficoltà dal punto di vista operativo se si considera l'enorme quantità di formati che oggi esistono. Uno studio condotto nel 2004 nell'ambito del progetto DELOS aveva censito oltre 15.000 formati diversi¹⁵. Dal 2004 ad oggi le cose sono notevolmente cambiate: molti di quei formati sono diventati obsoleti e contemporaneamente sono stati proposti migliaia nuovi formati. Gestire la migrazione ripetuta nel tempo di migliaia di formati elettronici è un'impresa non semplice e certamente non alla portata di tutti. Di conseguenza è urgente individuare uno o più formati che abbiano

12. Per *significant properties* si intendono le proprietà dei formati elettronici che devono essere conservate e che non possono essere perdute, ad esempio a seguito di operazioni di migrazione. Esse sono state oggetto di accurata analisi prima nell'ambito del progetto INSPECT (*Investigating the Significant Properties of Electronic Content over Time*), promosso nel Regno Unito dall'Arts and Humanities Data Service (AHDS) in collaborazione con The National Archives, e poi nell'ambito del progetto PLANETS (*Preservation and Long-term Access through Networked Services*).

13. Per cercare di conservare inalterato nel tempo il *bitstream* ed evitare qualsiasi tipo di migrazione sono state proposte altre strategie di conservazione che, come quella dell'emulazione, si sono rivelate, tuttavia, non sostenibili nel tempo perché quasi sempre molto costose e di non facile applicabilità.

14. Il Decreto del Presidente del Consiglio dei Ministri 3 dicembre 2013 *Regole tecniche in materia di sistema di conservazione ai sensi degli articoli 20, commi 3 e 5-bis, 23-ter, comma 4, 43, commi 1 e 3, 44, 44-bis e 71, comma 1, del Codice dell'amministrazione digitale di cui al decreto legislativo n. 82 del 2005 all'articolo 7, comma 1, lettera f)* stabilisce che il Responsabile della conservazione «assicura la verifica periodica, con cadenza non superiore ai cinque anni, dell'integrità degli archivi e della leggibilità degli stessi». Analoga previsione in termini temporali era stabilita anche dalla Deliberazione CNIPA n. 11 del 19 febbraio 2004 (già cit.), ora abrogata; infatti, l'articolo 5, comma 1, lettera h) stabiliva che il Responsabile della conservazione verifica periodicamente, con cadenza non superiore a cinque anni, l'effettiva leggibilità dei documenti conservati provvedendo, se necessario, al riversamento diretto o sostitutivo del contenuto dei supporti».

15. Cfr. M. GUERCIO, C. CAPIELLO, *File formats typology and registries for digital preservation*, (Deliverable reference number: WP6, D 6.3.1), DELOS Project no.507618, Università degli Studi di Urbino — Istituto di Studi per la tutela dei Beni Archivistici e Librari (ISTBAL), Urbino 2004. Lo studio proponeva una classificazione dei formati, ne discuteva gli aspetti salienti ed i requisiti fondamentali e suggeriva criteri di valutazione.

caratteristiche tali da assicurare la loro riproducibilità nel lungo periodo riducendo al minimo il numero di migrazioni necessarie. Uno di questi formati è senza alcun dubbio il PDF/A.

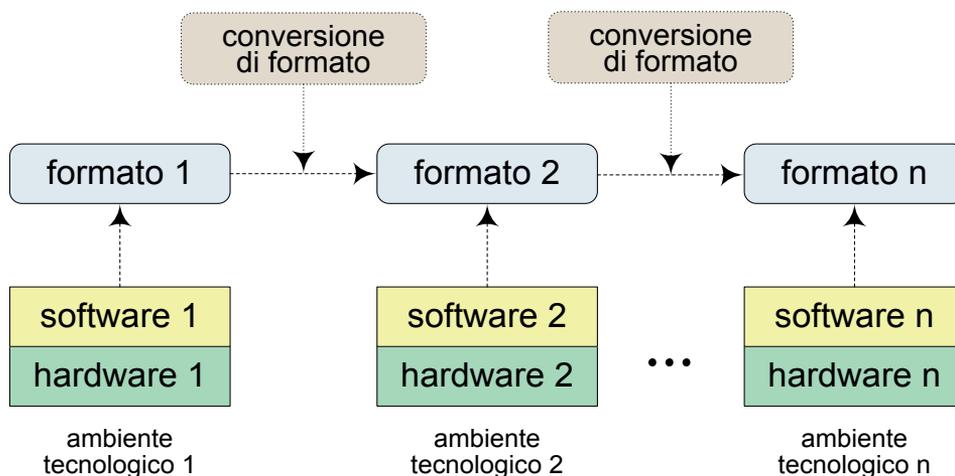


Figura 1. Schema di funzionamento della strategia di conservazione digitale basata sulla migrazione.

4. La famiglia di formati PDF/A

Il formato PDF/A è definito dallo standard ISO 19005. Si tratta di una norma pensata in maniera modulare: è costituita da più parti¹⁶ che sono indipendenti tra di loro e possono essere singolarmente sviluppate per seguire l'evoluzione nel tempo delle varie versioni del formato PDF a cui il PDF/A fa riferimento. Lo sviluppo dello standard in senso all'ISO è seguito dal Sottocomitato 2 "Application issues" del Comitato Tecnico 171 "Document management applications" (ISO/TC171/SC2 – Document management applications – Application issues) per il quale opera come segretariato l'Association for Information and Image Management (AIIM).

16. Si noti che la stessa filosofia è stata seguita nella definizione dello standard ISO che descrive le specifiche del formato PDF; la prima parte, denominata ISO 32000-1:2008 *Document management – Portable document format – Part 1: PDF 1.7*, descrive le specifiche del formato PDF nella versione 1.7. Seguiranno nel tempo le altre parti della norma, come la ISO/DIS 32000-2.4, *Document management – Portable document format – Part 2: PDF 2.0*, che si trova attualmente nello stato di *Draft International Standard* (DIS).

Al momento sono state pubblicate le prime tre parti dello standard¹⁷. La prima parte, quella che definisce le specifiche del PDF/A-1, è stata pubblicata il 28 settembre 2005 con la denominazione di ISO 19005-1:2005 “*Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1)*”; si basa sul formato PDF nella versione 1.4 (PDF Reference Version 1.4)¹⁸. La seconda parte dello standard è stata pubblicata il 20 giugno 2011 con la denominazione di ISO 19005-2:2011, *Document management — Electronic document file format for long-term preservation — Part 2: Use of ISO 32000-1 (PDF/A-2)*; essa definisce le specifiche del formato PDF/A-2 basandosi sul formato PDF nella versione 1.7 (che nel frattempo è stato riconosciuto come standard ISO 32000-1:2008). Infine, come già anticipato, il 15 ottobre 2012 è stata pubblicata la terza parte dello standard, il PDF/A-3, con la denominazione di ISO 19005-3:2012, *Document management — Electronic document file format for long-term preservation — Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)*.

Nel momento in cui si scrive è in fase di sviluppo una nuova versione: si tratta formato del PDF/A-4, di cui il relativo standard si trova nella fase *Committee Draft* (CD) con la denominazione di ISO/CD 19005-4, *Document management — Electronic document file format for long-term preservation — Part 4: Use of ISO 32000-2 (PDF/A-NEXT)*. Il progetto di creazione del nuovo standard era stato approvato il 21 settembre 2016 ed era stato registrato il relativo *Committee Draft*. Il 30 marzo 2017 si è chiusa la fase di commento e di voto sullo standard. Il comitato tecnico che ne sta seguendo lo sviluppo è, come per gli altri standard, il Sottocomitato 2, *Document file formats, EDMS systems and authenticity of information*, del Comitato Tecnico 171, *Document management applications, (ISO/TC171/SC2 — Document management applications — Document file formats, EDMS systems and authenticity of information)*¹⁹. La Figura 2 riepiloga sinteticamente le varie “parti” che compongono lo standard ISO 19005.

17. Le norme citate nel presente articolo sono disponibili, a pagamento, sul sito dell’iso, <<http://www.iso.org>> (ultima consultazione: 18/05/2017).

18. Il PDF Reference è il documento che contiene le specifiche del formato PDF. Il PDF Reference 1.4 contiene le specifiche del formato PDF versione 1.4, utilizzato in Adobe Acrobat 5. Analogamente il PDF Reference 1.5 contiene le specifiche del formato PDF versione 1.5, utilizzato in Adobe Acrobat 6, e così via. I PDF Reference sono disponibili gratuitamente sul sito di Adobe Systems. Ad esempio, l’ultima versione, la 1.7, è disponibile all’indirizzo <www.adobe.com/devnet/pdf/pdf_reference.html> (ultima consultazione: 18/05/2017).

19. Si noti che la denominazione del Sottocomitato 2 del Comitato Tecnico 171 è cambiata (da “*Application issues*” a “*Document file formats, EDMS systems and authenticity of information*”) per riflettere le nuove competenze attribuite a tale Sottocomitato.