

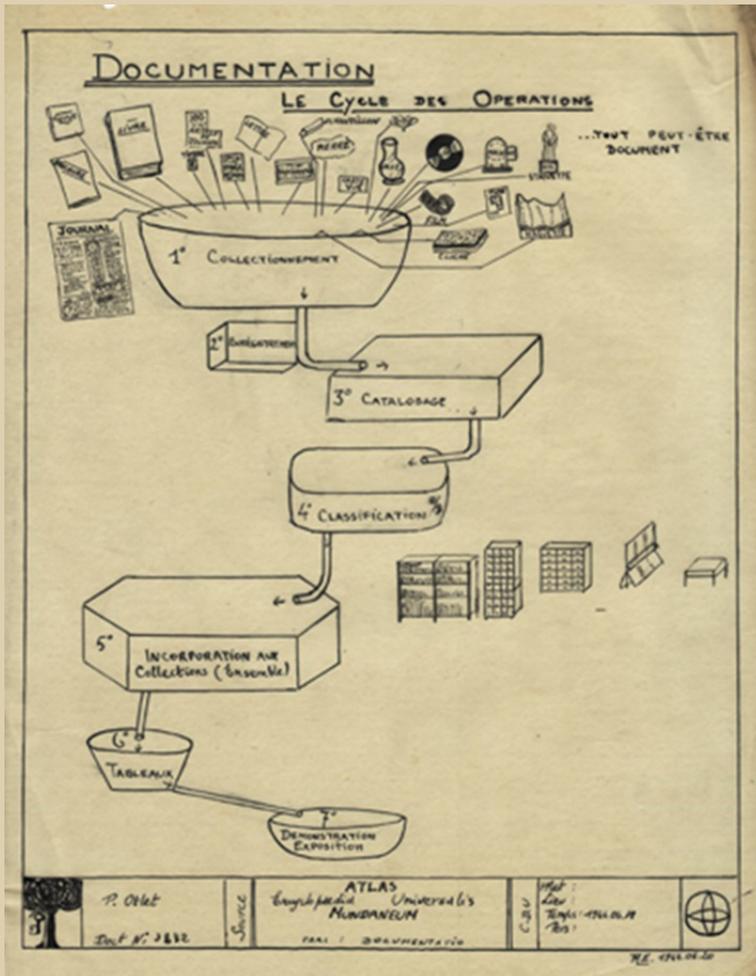
AIDinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 1-2

ANNO 40

GENNAIO-GIUGNO 2022



AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

Fondata nel 1983 da Paolo Bisogno

Proprietario della rivista:

Università della Calabria

Direttore Scientifico:

Roberto Guarasci, *Università della Calabria*

Direttore Responsabile:

Fabrizia Flavia Sernia

Comitato scientifico:

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, *Centro Italiano Ricerche Aerospaziali*;

Gino Roncaglia, *Università della Toscana*;

Laurence Favier, *Université Charles-de-Gaulle Lille 3*;

Madjid Ihadjadene, *Université Vincennes-Saint-Denis Paris 8*;

Maria Mirabelli, *Università della Calabria*;

Agustín Vivas Moreno, *Universidad de Extremadura*;

Douglas Tudhope, *University of South Wales*;

Christian Galinski, *International Information Centre for Terminology*;

Béatrice Daille, *Université de Nantes*;

Alexander Murzaku, *College of Saint Elizabeth, USA*;

Federico Valacchi, *Università di Macerata*.

Comitato di redazione:

Antonietta Folino, *Università della Calabria*;

Erika Pasceri, *Università della Calabria*;

Maria Taverniti, *Consiglio Nazionale delle Ricerche*;

Maria Teresa Chiaravallotti, *Consiglio Nazionale delle Ricerche*;

Assunta Caruso, *Università della Calabria*;

Claudia Lanza, *Università della Calabria*.

Segreteria di Redazione:

Valeria Rovella, *Università della Calabria*

Editrice: Cacucci Editore S.a.s.

Via D. Nicolai, 39 – 70122 Bari (BA)

www.cacuccieditore.it

e-mail: riviste@cacuccieditore.it

Telefono 080/5214220

AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

«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 riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Settore 11/A4 e censita 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'ARES (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; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – 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.

AIDAinformazioni

Anno 40

N. 1-2 – gennaio-giugno 2022

CACUCCI  EDITORE
BARI

PROPRIETÀ LETTERARIA RISERVATA

© 2022 Cacucci Editore – Bari

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220

<http://www.cacuccieditore.it> e-mail: info@cacucci.it

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.

Sommario

Contributi

STEFANO ALLEGREZZA, La conservazione degli archivi di posta elettronica: sviluppi recenti e prospettive future	9
ANDREA BELLANDI, Le Risorse Linguistiche nell'era del Web Semantico. Un insieme di servizi informatici per la gestione di lessici e terminologie	31
GIORGIA DI MARCANTONIO, E se l'archivio non rispecchia l'istituto? Pavone e il rispecchiamento: analisi di una bozza preliminare	51
GERARDO GIARDIELLO, CHIARA FIORAVANTI, FRANCESCO ROMANO, MARIASOLE RINALDI, Strumenti per il reperimento di informazioni rilevanti in Rete. Applicazioni nel dominio della comunicazione pubblica in materia di immigrazione in Italia	69
CLAUDIA LANZA, Termini e testi nella comunicazione dei gruppi No-vax	89
ANNA ROVELLA, La posta elettronica negli archivi di persona: conservazione e accesso	113
GIOVANNI SALUCCI, Il Blog di qualità. Proposta di un nuovo format di pubblicazione nell'editoria scientifica	129

Note e rubriche

ROBERTO GUARASCI, Paul Otlet e Suzanne Briet. Note in margine al libro di Paola Castellucci e Sara Mori	147
ERIKA PASCERI, L'evoluzione del <i>Nuovo Glossario di biblioteconomia e scienza dell'informazione</i> ai tempi dello <i>Smart Working</i>	151
CLAUDIO GNOLI, Come mi vuoi, canonica o rivoluzionaria?	153
CLAUDIO GRIMALDI, L'universo affascinante del colore: le sociocromie	157

Contributi

La conservazione degli archivi di posta elettronica: sviluppi recenti e prospettive future

STEFANO ALLEGREZZA*

ABSTRACT: This paper aims at providing an outline of the unsolved email preservation problem. Starting from a discussion of the major preservation challenges (such as the massive amount of emails sent and received, the issue of attachments in a variety of file formats, the issue of links to external resources, the issues of privacy and digital legacy, and so on), it moves on to the email preservation strategies proposed over the last twenty years (from printing to paper, to normalization to XML-based formats, to the use of PDF and PDF/A and its various profiles). Finally, the most interesting recent developments on email are discussed. In particular, great attention is given to the EA-PDF file format that seems to become the ultimate solution to the problem of email archiving and preservation.

Keywords: Email preservation, Email archiving, Digital preservation, PDF, EA-PDF.

1. Introduzione

Il tema della conservazione degli archivi di posta elettronica rientra tra quelli sui quali, da almeno una ventina di anni, ferve una approfondita riflessione da parte della comunità scientifica e professionale degli archivisti, alla ricerca delle strategie più idonee per assicurare un futuro alle email che quotidianamente vengono inviate e ricevute sia dalle organizzazioni pubbliche e private che dai singoli individui. A distanza di cinquanta anni dalla sua nascita¹, la posta elettronica è cresciuta fino a diventare uno dei mezzi di comunicazione più utilizzati al mondo sia per le interazioni di tipo lavorativo che personale. Tra la fine degli anni Ottanta e l'inizio degli anni Novanta ha iniziato progressivamente a sostituirsi alla corrispondenza cartacea (le tradizionali 'lettere') e oggi è diventata uno strumento irrinunciabile per la maggior parte degli individui, andando a sostituire in molti casi anche le tradizionali comunicazioni

* Dipartimento di Beni culturali, Università degli Studi di Bologna (Italia). stefano.allegrezza@unibo.it.

¹ La nascita della posta elettronica si fa risalire al 1971, quando Ray Tomlinson inviò a sé stesso un elementare messaggio di prova – qualcosa come “QWERTYUIOP”.

telefoniche. I messaggi inviati e ricevuti costituiscono la testimonianza di pensieri, decisioni, azioni, attività (Council of Library and Information Resources 2018) e spesso devono essere conservati per lunghi periodi di tempo – se non per sempre – sia per motivazioni di ordine giuridico o amministrativo che per motivazioni contingenti, personali, affettive.

Ciò nonostante, la questione della conservazione della posta elettronica è tutt'altro che risolta: ad oggi le strategie che sono state proposte si sono rivelate frammentarie, incomplete e non sempre concretamente realizzabili all'atto pratico, con la conseguenza che numerosi archivi di posta elettronica sono ormai andati perduti, soprattutto – ma non solo – nell'ambito degli archivi di persona. Tuttavia, recentemente ci sono stati degli interessanti sviluppi sui quali vale la pena soffermarsi, come si vedrà nel seguito.

2. L'importanza della conservazione degli archivi di posta elettronica

Conservare gli archivi di posta elettronica è fondamentale sia che si tratti di email ricevute ed inviate in ambito lavorativo che di email personali. Dalle chiacchiere tra amici ai pettegolezzi di famiglia, dalle decisioni aziendali alle azioni governative, tutto oggi è documentato dalle email che si sedimentano nelle caselle di posta elettronica². Attraverso la posta elettronica vengono trasmessi documenti di varia natura; anzi in molti casi, la mail è diventata essa stessa documento e il contenuto informativo, che nel passato era normalmente presente nell'allegato, oggi tende ad essere spostato nel corpo del messaggio³. Per molte organizzazioni, le caselle di posta elettronica costituiscono una delle fonti più importanti di prove documentali di decisioni, attività e transazioni, e sono state accettate come prova nelle aule dei tribunali; come tali, devono essere conservate (InterPARES3 Project 2009). Le email possono rappresentare anche una fonte storica e sociale significativa, importantissima per documen-

² Per la verità in questi ultimi anni si sta assistendo ad un utilizzo sempre più frequente dei sistemi di messaggistica istantanea (WhatsApp, Messenger di Facebook, messaggi su LinkedIn, etc.) che vanno a sostituirsi alle email, in considerazione del fatto che richiedono un minor tempo per la risposta. Purtroppo, a volte anche le comunicazioni istituzionali vengono veicolate attraverso questi sistemi, con conseguenze, dal punto di vista archivistico, piuttosto preoccupanti (si pensi, ad esempio, alle conversazioni con un determinato corrispondente che iniziano utilizzando la posta elettronica e poi proseguono su WhatsApp o altri sistemi di messaggistica, con la conseguente difficoltà – se non addirittura l'impossibilità – di riunire all'interno di una unica aggregazione documentale tutti i messaggi scambiati con quel corrispondente).

³ The National Archives, *Managing emails*, <https://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/managing-email> (ultima consultazione: 26/05/2022); si veda anche University of South Wales, *How to Manage Emails as Records*, <https://uswvarious1.blob.core.windows.net/uswvarious-prod-uploads/documents/HowTo-Manage-EmailsAsRecords.pdf> (ultima consultazione: 26/05/2022).

tare la storia e poterla ricostruire in futuro, così come possono costituire una fonte straordinaria per ricostruire il profilo e la vita personale e lavorativa di un personaggio illustre.

Tuttavia, nonostante il valore storico o culturale della posta elettronica sia ormai riconosciuto, solo poche istituzioni hanno compiuto progressi sostanziali nell'individuare le metodologie, le procedure e gli strumenti più idonei per la sua conservazione a lungo termine (Prom 2019).

È importante riflettere anche sul concetto di email come documento in senso archivistico («email as a record»). È evidente che non tutte le email inviate, ricevute o comunque scambiate all'interno di una organizzazione (o da una persona fisica nel caso delle email personali) sono da considerare documenti in senso archivistico; tuttavia alcune di queste lo sono e come tali sono meritevoli di particolare attenzione e devono essere conservate per periodi più o meno lunghi (in alcuni casi per sempre). Uno dei progetti che ha affrontato con maggiore rigore metodologico questo aspetto è stato InterPares⁴. Dal punto di vista della diplomatica del documento contemporaneo esso ha individuato cinque caratteristiche che un oggetto digitale deve possedere affinché possa essere considerato un documento in senso archivistico (InterPARES3 Project 2011, 8-10): un contenuto stabile ed una forma fissa; un legame con un atto a cui il documento fornisce supporto o proceduralmente o come parte di un processo decisionale; la presenza di un vincolo archivistico; il coinvolgimento di alcuni soggetti (un autore, uno scrittore, un destinatario); la collocazione in uno dei seguenti cinque tipi di contesto: giuridico-amministrativo, finanziario, procedurale, documentale, tecnologico. Non è certamente questa la sede opportuna per approfondire ulteriormente l'argomento e si rimanda per una sua trattazione esaustiva ai numerosi studi che sono stati pubblicati come risultati del progetto⁵; preme, comunque, sottolineare come la questione del riconoscimento del valore archivistico dei messaggi di posta elettronica sia di primaria importanza, perché da esso discendono tutta una serie di adempimenti ed obblighi – anche di legge – in ordine alla loro tenuta e conservazione da parte dei soggetti che trasmettono o ricevono tali messaggi e, in prospettiva,

⁴ Si tratta della terza fase del progetto internazionale di ricerca InterPARES (acronimo di INTERnational Research on Permanent Authentic Records in Electronic Systems), avviata nel 2007 e completata nel 2012. L'obiettivo primario del progetto, rivolto alle «small and medium sized public and private archival organizations and programs, which are responsible for the digital records resulting from government, business, research, art and entertainment, social and/or community activities» è stato quello di renderle capaci «to preserve over the long-term authentic records that satisfy the requirements of their stakeholders and society's needs for an adequate record of its past» (Duranti 2007). Sulla base di studi generali e di casi, nonché dei prodotti e risultati delle precedenti fasi del progetto, InterPARES 3 è stato in grado di sviluppare una serie di strategie concrete per la conservazione dei documenti digitali oltre a policy e raccomandazioni.

⁵ Disponibili all'indirizzo www.interpares.org (ultima consultazione: 26/05/2022).

da parte dei soggetti conservatori che riceveranno i versamenti di archivi di posta elettronica.

3. Le sfide della conservazione degli archivi di posta elettronica

Gli archivi di posta elettronica costituiscono una categoria particolarmente difficile da conservare, e questo per tutta una serie di motivi.

Innanzitutto, i messaggi di posta elettronica sono oggetti digitali complessi: infatti, dal punto di vista tecnologico, un messaggio di posta elettronica è composto da tre parti:

1) l'intestazione (*header*), che contiene, sotto forma di metadati, le informazioni sul mittente o sul destinatario, sulla data di creazione, sull'oggetto, tipicamente visualizzate a schermo. Oltre a queste, vi sono altri metadati, più tecnici e che solitamente non vengono mostrati all'utente, ma che sono fondamentali al fine di assicurare l'autenticità del messaggio, come le informazioni sui vari server di posta elettronica attraverso cui il messaggio è transitato nel suo percorso dal server della posta in uscita del mittente al server della posta in arrivo del destinatario, insieme con i rispettivi orari⁶.

2) il "corpo del messaggio" (*message body*), cioè il contenuto vero e proprio dell'email, che viene visualizzato in modo diverso a seconda delle impostazioni definite dall'utente nel software di posta elettronica o nella webmail;

3) facoltativamente, uno o più allegati (*attachments*), che spesso sono documenti di testo, fogli di calcolo, presentazioni, disegni tecnici, immagini, documenti scansionati, ecc. Non ci sono limitazioni al tipo di oggetto digitale che può essere incluso come allegato (se non per questioni legate alla sicurezza, per cui alcuni sistemi impediscono, ad esempio, di allegare programmi, codice eseguibile, script o altre tipologie di allegati potenzialmente pericolosi); di conseguenza possono essere presenti come allegati anche documenti codificati secondo formati elettronici scarsamente utilizzati o addirittura sconosciuti.

È evidente che conservare un messaggio di posta elettronica significa conservare non solo il corpo del messaggio, ma ciascuna di queste parti; inoltre, le tre parti non devono essere conservate in maniera indipendente l'una dall'altra ma in maniera da mantenere le relazioni tra loro.

A ben vedere, il vero problema della conservazione a lungo termine delle email non è tanto la conservazione dell'intestazione o del corpo del messaggio (che sono ovviamente importanti, e presentano anch'essi delle difficoltà di tipo tecnico), quanto la conservazione degli allegati. Mentre conservare l'intestazione o il corpo del messaggio è relativamente semplice, conservare gli allegati può presentare dei livelli di difficoltà non facilmente superabili – ad esempio, a

⁶ Queste informazioni sono comunque visualizzabili andando ad agire sulle impostazioni del client di posta elettronica o del sistema di webmail in uso.

causa della presenza di formati elettronici obsoleti o in procinto di diventarlo – oppure di formati proprietari e non standard⁷. La conservazione di un archivio di posta elettronica può quindi richiedere la conservazione di una quantità potenzialmente elevata di oggetti digitali in una grande varietà di formati.

Un'altra criticità è rappresentata dall'enorme quantità di email che vengono inviate e ricevute ogni giorno. La produzione di email è cresciuta in modo esponenziale e ogni giorno vengono scambiate centinaia di miliardi di email in tutto il mondo (The Radicati group 2015). Di conseguenza, la dimensione dell'archivio di posta elettronica che un'organizzazione deve gestire raggiunge facilmente l'ordine delle decine di terabyte e contiene centinaia di migliaia, se non addirittura milioni, di email. Questo pone seri problemi sia per la gestione che per la conservazione delle email.

Un altro problema è costituito dall'inserimento nel corpo del messaggio di *link* che puntano verso risorse esterne. Ad esempio, in un messaggio di posta elettronica può essere presente un collegamento ad un documento archiviato sul cloud, come sempre più spesso si verifica quando si utilizzano sistemi di posta elettronica integrati in una suite di *office automation* on-line⁸. Ma cosa succederà tra cinque, dieci o venti anni, quando magari quel documento sarà stato cancellato o spostato dallo spazio condiviso? Ovviamente, accedendo dalla email risulterà impossibile recuperare il documento di testo, perché quest'ultimo non è stato allegato al messaggio ma solo "linkato".

Analogo è il caso di un messaggio di posta elettronica che contiene riferimenti a contenuti esterni necessari per comprendere il messaggio stesso. Ad esempio, in un'email potrebbe essere presente un'immagine che non è incorporata nell'email stessa, ma che viene semplicemente recuperata dal Web durante la visualizzazione dell'email. Se in futuro l'immagine venisse cancellata, rinominata o spostata in un'altra cartella, non potrebbe più essere visualizzata nell'email. I *link* tendono a durare poco e dopo poco tempo risultano spesso interrotti; quindi, per preservare l'aspetto visivo dell'email, i link a contenuti

⁷ Il tema della scelta dei formati elettronici per la conservazione dei messaggi di posta elettronica riveste grande importanza, dal momento che è anche da una loro accurata selezione che dipendono, da una parte, la possibilità di conservare nel tempo i documenti prodotti in tali formati, dall'altra, la possibilità di ridurre al minimo le operazioni di riversamento sostitutivo (*migration*) che si renderanno necessarie nel tempo. Come osserva Maria Guercio «non vi è dubbio che i formati costituiscano un nodo importante del processo di conservazione a lungo termine: una scelta corretta di formati non proprietari, aperti, standard *de jure* e orientati ai dati facilita l'individuazione di soluzioni adeguate e riduce i costi» (Guercio 2019).

⁸ Il classico esempio è costituito dall'invio ad un destinatario di un documento di testo redatto on line e disponibile, ad esempio, su Sharepoint, lo spazio virtuale basato sul cloud fornito da Microsoft; il documento in questione non viene effettivamente allegato al messaggio ma è disponibile su tale spazio virtuale, dove può essere 'recuperato' semplicemente 'cliccando' sul link presente nel messaggio.

esterni devono essere “risolti” ed è necessario che i contenuti esterni vengano incorporati nell’email stessa.

Bisogna poi tenere presenti i problemi legati alla privacy e alla protezione dei dati personali. Un messaggio di posta elettronica non di rado contiene nel corpo del messaggio o nei suoi allegati delle informazioni personali o dei dati sensibili, riferibili al mittente o al destinatario – o a più destinatari nel caso di messaggi inviati a più corrispondenti: non è raro avere conversazioni multiple che coinvolgono molte persone contemporaneamente (Digital Preservation Coalition 2018). Le caselle di posta elettronica possono contenere quantità rilevanti di informazioni personali e dati sensibili, anche a seconda del tipo di organizzazione (si pensi alle banche, agli ospedali o ad alcuni tipi di pubbliche amministrazioni) o del ruolo rivestito dal soggetto produttore (si pensi al personale delle risorse umane, ai dirigenti sindacali, etc.). Si consideri poi che gli individui che donano le caselle di posta elettronica agli archivi possono aver usato la loro posta elettronica per scambiare messaggi con un numero elevato di corrispondenti che non sono a conoscenza del fatto che le loro comunicazioni vengono conservate e un domani potrebbero diventare pubblicamente consultabili. Ci possono essere anche complicazioni relative al copyright, dato che l’autore dell’email non è il destinatario della stessa. Il riutilizzo e la conservazione possono essere influenzati anche dai diversi quadri normativi presenti nelle giurisdizioni dei vari Paesi a cui appartengono i corrispondenti di uno scambio di email (Prom 2019).

Un altro aspetto da considerare è il problema della cosiddetta “eredità digitale”, di cui si sta acquisendo una maggiore consapevolezza proprio in questi ultimi anni: alla morte di un soggetto produttore, è spesso molto difficile – se non impossibile – avere accesso al suo archivio di posta elettronica, perché questo è contenuto nell’account che è protetto da credenziali di autenticazione (tipicamente una username ed una password) che di solito nessuno condivide con altri e quindi si perdono con la morte del *de cuius*. Inoltre, i fornitori di servizi di posta elettronica conservano le email solo per un determinato periodo di tempo, variabile da pochi mesi a uno o due anni al massimo dall’ultimo accesso alla casella di posta. Quindi, se si aspetta troppo tempo prima di tentare di recuperare l’archivio delle email di una persona deceduta (o di qualcuno che, per i motivi più vari – una malattia, un incidente, etc. – da qualche tempo non effettua l’accesso al suo account), si avrà l’amara sorpresa di trovare l’archivio di posta elettronica completamente vuoto (Allegrezza 2021, 352-400)⁹.

Un altro problema che si riscontra quasi ovunque è la totale assenza di *policy*. La maggior parte delle organizzazioni non dispone di una policy chiara su come gli utenti di quella organizzazione debbano gestire e conservare le email, il che significa che ognuno può adottare un approccio diverso e quasi mai si tratta di un approccio che fornisce garanzie sufficienti sulla loro conservazione.

⁹ Per una trattazione completa ed esaustiva dei problemi legati alla successione del patrimonio digitale si veda (D’Arminio Monforte 2020; D’Arminio Monforte 2018).

Infine, dobbiamo considerare la mancanza di un formato elettronico specifico per la conservazione delle email. Di solito, i formati utilizzati dai vari software per gestire le email (i cosiddetti *client* di posta elettronica) non sono standard né sono pensati per la conservazione a lungo termine. Ciò significa che non esiste un formato di conservazione “nativo” e qualsiasi strategia per conservare le email richiede un’operazione di riversamento, anche se non c’è ancora consenso su quale sia il formato di conservazione da utilizzare.

Infine, occorre tenere bene a mente che l’archivio dei messaggi di posta elettronica non dovrebbe essere considerato come un archivio a sé stante, ma come una parte dell’archivio del soggetto produttore: l’archivio è unico. Ad esempio, i messaggi di posta elettronica relativi ad un determinato affare dovrebbero confluire nel fascicolo relativo a quell’affare e non costituire un fascicolo a sé stante nel sistema di posta elettronica, nel rispetto del mantenimento del vincolo archivistico¹⁰. Purtroppo, quello che invece si riscontra nella quasi totalità dei casi è che si formano due archivi distinti – e, per certi versi, “paralleli” – : da una parte, l’archivio vero e proprio, contenente i documenti inviati, ricevuti o comunque formati dal soggetto produttore nel corso della sua attività; dall’altra, l’archivio dei messaggi di posta elettronica, contenente almeno in parte i documenti presenti nel primo archivio (ad esempio, quelli trasmessi o inviati tramite email). Si tratta di una duplicazione che certamente non contribuisce all’organicità ed unitarietà¹¹ dell’archivio e rende più difficile la gestione degli affari correnti e, nel futuro, la consultazione dell’archivio per finalità di ricerca.

Quelli appena visti sono solo alcuni dei problemi che si devono affrontare quando si vuole attuare una strategia di conservazione degli archivi di posta elettronica e che rendono questo compito alquanto difficile. Nel seguito si cercherà di comprendere quali siano le soluzioni che sono state proposte fino ad oggi e quelle sulle quali si sta discutendo in questo momento.

4. Le strategie per la conservazione degli archivi di posta elettronica

Negli ultimi venti anni sono state proposte diverse strategie per risolvere il problema dell’archiviazione e conservazione dei messaggi di posta elettronica.

¹⁰ Il concetto di vincolo archivistico come elemento essenziale di un archivio fu identificato da Giorgio Cencetti nel suo famoso articolo “Sull’archivio come «universitas rerum»”: «Ben diversamente stanno le cose per ciò che riguarda l’archivio. Qui i singoli componenti, le carte [...] non solo provengono dal medesimo individuo, aggregato familiare o ente [...] ma poiché costituiscono niente altro che uno fra i mezzi usati dall’ente o individuo per raggiungere i propri scopi, portano in loro stessi fin dall’origine il vincolo della destinazione comune, sintetizzato nell’adempimento delle funzioni dell’ente o individuo medesimo». (Cencetti 1937).

¹¹ Il concetto era stato più volte enunciato da Giorgio Cencetti. Nell’articolo “Il fondamento teorico della scienza archivistica” egli ribadiva l’unitarietà dell’archivio nelle sue tre fasi di vita e l’impossibilità di «differenziare teoricamente l’ufficio di protocollo dall’archivio, l’archivio corrente da quello di deposito: tutto è semplicemente archivio». (Cencetti 1939).

Una delle prime soluzioni proposte è quella che viene denominata “printing to paper”: essa consiste nello stampare su carta e archiviare in fascicoli o in faldoni tutti i messaggi di posta elettronica che si intende conservare. Per quanto possa sembrare incredibile, questa strategia è tutt’ora ampiamente utilizzata sia da individui singoli – magari per avere la certezza di preservare quella mail che si ritiene particolarmente importante – sia dagli uffici di alcune pubbliche amministrazioni, dove il direttore dell’ufficio incarica l’impiegato che si occupa della gestione della posta elettronica di stampare tutte le email che arrivano giornalmente. Stampare le email su carta può sembrare una soluzione semplice e diretta, ma presenta notevoli controindicazioni. Lasciando da parte le ovvie conseguenze sul piano ecologico, che pur sono rilevanti a fronte dell’enorme quantità di messaggi di posta elettronica ricevuti ed inviati e che quindi verrebbero stampati, merita rilevare come la stampa su carta comporti la perdita di una quantità significativa di informazioni, e soprattutto di quei metadati tecnici che sono fondamentali per assicurare l’autenticità, come si è già detto. Si pensi, poi, alla stampa degli allegati, spesso costituiti da corposi documenti testuali (in formato DOCX o PDF, ad esempio), o da fogli di calcolo che, per le loro dimensioni, mal si prestano alla stampa sui normali foglio A4, oppure da elaborati tecnici (come quelli inviati da studi di architettura e ingegneria) che richiederebbero, per poter essere stampati, la disponibilità di apparecchiature particolari – come i plotter – non sempre presenti negli uffici. Con il progredire delle competenze e delle esperienze nella conservazione del digitale, questo tipo di approccio è ormai unanimemente riconosciuto come inadatto, in quanto, oltre a comportare la perdita di informazioni fondamentali (come i metadati di cui si è parlato in precedenza), modifica l’aspetto visivo delle email e l’associata esperienza utente e dissocia i messaggi dai loro allegati: al contrario, il messaggio di posta elettronica deve essere conservato nella sua interezza come unità documentaria.

Un’altra strategia ampiamente utilizzata si basa sulla conversione dei messaggi di posta elettronica in un formato indipendente dal sistema, utilizzando uno dei numerosi strumenti specializzati oggi disponibili. La maggior parte di questi formati neutrali si basa sull’XML, ma anche questi formati presentano dei limiti, come la mancata o, comunque, difficoltosa conservazione dell’aspetto visivo del messaggio.

Altre strategie si basano sull’adozione di alcuni formati di archiviazione raccomandati, come il formato EML per le singole email e il formato MBOX per le aggregazioni di email (tipicamente il contenuto di una cartella)¹². Questo è

¹² Di questo avviso sono le *Linee guida sulla Formazione, gestione e conservazione di documenti informatici* pubblicate il 9 settembre 2020 dall’Agenzia per l’Italia Digitale (AgID) ed applicabili a partire dal 1° gennaio 2022. Infatti, l’*Allegato 2 – Formati elettronici e riversamento*, raccomanda di utilizzare «il formato EML per archiviare un singolo messaggio di posta elettronica» e «il formato MBOX per l’archiviazione di più messaggi ovvero di un’intera casella di posta elettronica». (Agenzia per l’Italia digitale 2020, 55).

l'approccio seguito, ad esempio, dalla Library and Archives Canada (LAC) e dalla National Archives and Records Administration (NARA) degli Stati Uniti. Tuttavia, entrambi i due formati non sono specifici per la conservazione; inoltre, a causa del grande numero di email che può essere contenuto in una singola cartella, i file MBOX raggiungono con facilità delle grandi dimensioni e, quindi, sono più sensibili al fenomeno della corruzione; ciò significa anche che, in alcuni casi, il danneggiamento di un singolo messaggio può impedire il caricamento o l'apertura dell'intero archivio. Il formato MBOX risente anche del fatto che non è, in realtà, un formato unico ma una famiglia di formati con almeno quattro diverse varianti (MBOXO, MBOXRD, MBOXCL e MBOXCL2), tra l'altro non completamente compatibili tra loro. Altri formati, come quelli sviluppati da Microsoft – l'MSG per i singoli messaggi e il PST per le aggregazioni di email – sebbene siano ampiamente utilizzati e le loro specifiche siano pubbliche, sono di fatto formati proprietari e soggetti a frequenti modifiche, e come tali non compatibili con processi di conservazione a lungo termine. Ad ogni modo, ad oggi non è stato ancora definitivamente individuato un formato ideale per la conservazione delle email ma, come si dirà più avanti, c'è una forte convergenza verso un nuovo profilo del formato PDF, denominato EA-PDF, che sembra avere caratteristiche tali da costituire la soluzione perfetta per la conservazione a lungo termine degli archivi di posta elettronica.

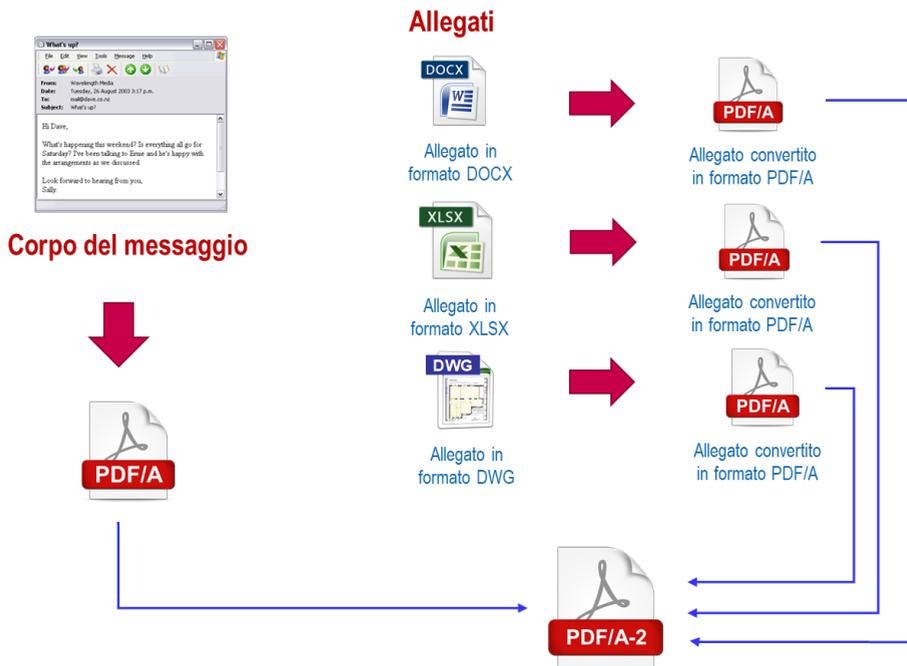


Figura 1: Esempio di utilizzo del formato PDF/A-2 per la conservazione di un messaggio di posta elettronica.

I problemi della strategia basata sulla ‘stampa su carta’, di cui si è trattato in precedenza, sono stati parzialmente risolti con l’introduzione di un’altra strategia, per certi versi analoga, che consiste nell’effettuare una stampa del messaggio di posta elettronica e dei suoi allegati, non più su carta ma nel formato PDF, utilizzando una delle tante stampanti virtuali oggi disponibili¹³. Sebbene sia diffusa la convinzione che la conversione delle email e dei loro allegati nel formato PDF sia in grado di garantire la conservazione a lungo termine, occorre constatare che, purtroppo, il PDF non è un formato idoneo per la conservazione, e questo per tutta una serie di motivazioni: ad esempio, non richiede l’incorporazione dei *font* (per il testo) o di profili colore ICC (per le immagini), quindi non può garantire la conservazione della rappresentazione visiva né del corpo del messaggio né dei suoi allegati. Inoltre, tutte le informazioni che garantiscono l’autenticità (tipicamente contenute nella parte più tecnica della intestazione, come si è già detto) non vengono catturate.

Per superare la criticità del formato PDF è stato proposto il formato PDF/A (PDF for archiving) che è stato riconosciuto standard ISO 19005-1 nel 2005¹⁴ e può essere utilizzato per l’archiviazione e la conservazione delle email e degli allegati (PDF Association 2021). Questo formato è da tempo utilizzato per scopi generali di archiviazione ed è universalmente riconosciuto come il formato di elezione per la conservazione a lungo termine di documenti a prevalente contenuto testuale. Il formato richiede l’incorporazione di font o profili ICC per i colori al fine di garantire la riproducibilità delle email nel corso degli anni e vieta qualsiasi contenuto dinamico; pertanto è in grado di garantire che gli allegati convertiti nel formato PDF/A rimarranno riproducibili e leggibili per decenni a venire (PDF Association 2015) e verranno visualizzati esattamente allo stesso modo su qualsiasi sistema e nel corso degli anni.

Con il rilascio del formato PDF/A-2, riconosciuto standard ISO 19005-2 nel 2011, si è introdotta una importante novità, ovvero la possibilità di

¹³ Si tratta della cosiddetta “stampa in formato PDF”, che si ottiene utilizzando una ‘stampante virtuale’ come PDF creator, PDF24 Creator, PDF Printer, o la stampante PDF disponibile di *default* nel sistema operativo in uso. Ovviamente questa strategia è applicabile solo se gli allegati sono ‘stampabili’, mentre non lo è nel caso di allegati ‘non stampabili’ (come un programma eseguibile, uno script, etc.).

¹⁴ La famiglia di standard ISO che definiscono i vari profili PDF/A è composta come segue: PDF/A-1 (ISO 19005-1:2005 – Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4); PDF/A-2 (ISO 19005-2:2011 – Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1); PDF/A-3 (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-4 (ISO 19005-4:2020 – Document management – Electronic document file format for long-term preservation – Part 4: Use of ISO 32000-2).

incorporare, all'interno di un documento in formato PDF/A-2, altri documenti purché nel formato PDF/A (PDF/A-1 o PDF/A-2). Questa caratteristica si è rivelata fondamentale in tanti settori, consentendo, ad esempio, la creazione di unità documentarie – costituita dal documento principale e dai suoi allegati – corrispondenti ad un unico documento PDF/A-2. Ma la sua utilità si è resa evidente in modo particolare nel settore della conservazione degli archivi di posta elettronica. Infatti, considerando ad esempio una email con tre allegati (Fig. 1) è possibile convertire in formato PDF/A-2 il corpo del messaggio e poi incorporare al suo interno i tre allegati dopo averli convertiti in formato PDF/A. In questo modo si è ottenuto un oggetto digitale essendo costituito solamente da documenti in formato PDF/A e come tale conservabile a lungo termine.

Il formato PDF/A-3, riconosciuto come standard ISO 19005-3 nel 2012, ha esteso questa possibilità: con questo nuovo formato è possibile incorporare all'interno di un documento in formato PDF/A-3 (che funge da documento “principale”) non solo altri documenti in formato PDF/A ma documenti in qualsiasi altro formato elettronico. Questo significa che, con riferimento all'esempio visto poco fa, è possibile convertire in formato PDF/A-3 il corpo del messaggio e poi incorporare al suo interno non solo i documenti in formato PDF/A ottenuti dalla conversione degli allegati, ma gli stessi allegati nel loro formato nativo (Fig. 2). In questo modo si ottiene un oggetto digitale che è idoneo alla conservazione a lungo termine perché tutte le componenti sono presenti *anche* nel formato PDF/A, ma nello stesso tempo è idoneo per la gestione corrente. Ad esempio, se si avesse bisogno di visualizzare o riutilizzare uno degli allegati nel loro formato nativo (DOCX, XLSA o DWG nell'esempio in questione) è possibile “estrarli” dal documento PDF/A-3. Il problema della conservazione a lungo termine è risolto perché tra cinquanta anni quegli allegati molto probabilmente non saranno più leggibili, ma si avrà la possibilità di leggere al loro posto gli equivalenti ottenuti dalla trasformazione in formato PDF/A.

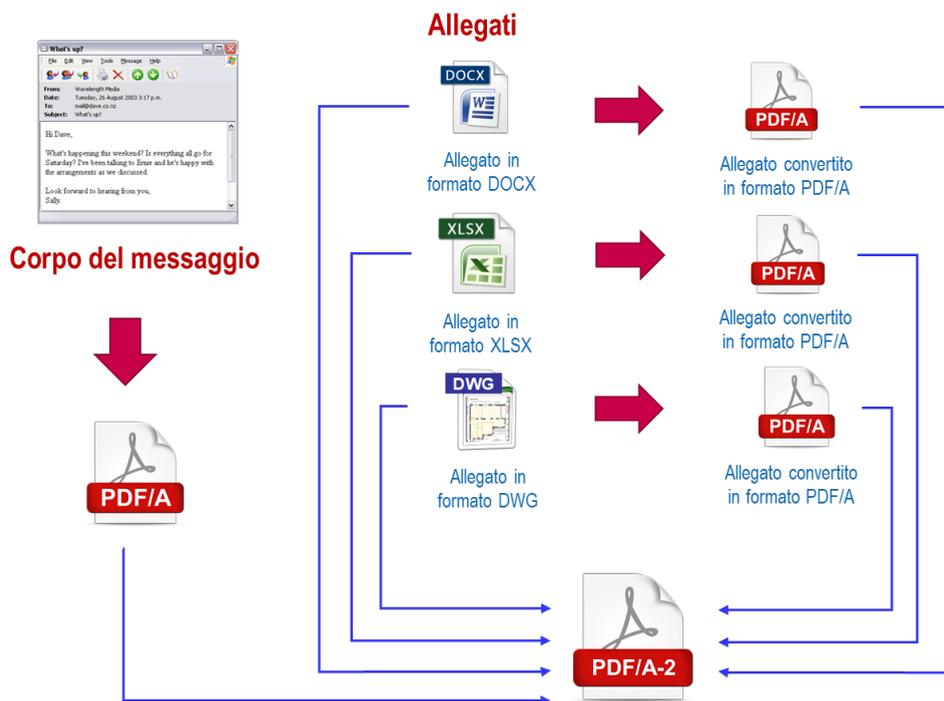


Figura 2: Esempio di utilizzo del formato PDF/A-3 per la conservazione di un messaggio di posta elettronica.

Alla fine del 2020 è stato riconosciuto standard ISO 19005-4 il profilo PDF/A-4, il successore del PDF/A-3. Uno dei suoi sotto-profili, il PDF/A-4f consente di incorporare qualsiasi tipologia di documento, svolgendo di fatto le funzioni che aveva in precedenza il formato PDF/A-3.

La maggior parte dei sistemi di posta elettronica offre una funzione di esportazione delle email nel formato PDF o PDF/A. Il corpo del messaggio viene convertito in base al tipo di codifica utilizzata (ASCII semplice, testo formattato, HTML), mentre i *link* vengono integrati incorporando le immagini a cui “puntano”. Purtroppo, però, questo approccio si rivela insufficiente, perché di solito viene preso in considerazione solo il corpo del messaggio e non l’intestazione o gli allegati.

In breve, la strada verso una soluzione efficace è ancora lunga, ma cominciano ad essere proposte alcune soluzioni interessanti (Von Seggern 2021). È appena il caso di far notare che la migrazione nel formato PDF/A potrebbe richiedere qualche riflessione in più sul mantenimento delle caratteristiche di integrità ed autenticità dei messaggi di posta elettronica (si tratta, a tutti gli effetti, di un riversamento di formato) che, almeno in Italia, possono essere assicurate ricorrendo alla procedura prevista dall’Allegato 2 alle “Linee Guida

sulla formazione, gestione e conservazione dei documenti informatici” dell’Agenzia per l’Italia Digitale¹⁵.

5. Recenti sviluppi nella conservazione della posta elettronica

La comunità internazionale degli archivisti si occupa del tema della conservazione degli archivi di posta elettronica da almeno un paio di decenni, ma è negli ultimi anni – grossomodo dal 2015 in poi – che si è assistito ad una spinta vigorosa verso la ricerca di soluzioni solide sotto il profilo teorico ma anche concrete, forse anche per l’importanza che la posta elettronica ha assunto in ogni organizzazione oltre che nella vita privata degli individui. Questo ha portato alla nascita di diversi progetti, alcuni dei quali meritano decisamente di essere conosciuti.

Uno dei più interessanti è quello condotto dal 2016 al 2018 dalla *Task Force on Technical Approaches for Email Archives*, finanziata dalla Andrew W. Mellon Foundation e dalla Digital Preservation Coalition (Prom 2020). Nell’agosto 2018 la *task force* ha pubblicato un approfondito rapporto, dal titolo “The Future of Email Archives” (Council of Library and Information Resources 2018), destinato non solo alla comunità archivistica e ai professionisti della conservazione digitale, ma anche agli sviluppatori di software, ai fornitori di servizi, agli storici e agli studiosi, agli amministratori di istituzioni pubbliche e private. Il rapporto fornisce un’analisi dettagliata delle sfide che la conservazione delle email pone e stabilisce un’agenda di lavoro per migliorare le conoscenze sul tema, compreso il supporto, la sperimentazione e lo sviluppo di strumenti interoperabili per colmare le lacune mancanti. Una delle principali lacune individuate dalla *task force* è stata la mancanza di indicazioni sui formati elettronici adatti all’archiviazione e conservazione a lungo termine della posta elettronica. È interessante notare, per gli sviluppi di cui si parlerà nel seguito

¹⁵ In particolare, il paragrafo 3.3 specifica che «quando si effettua un riversamento finalizzato alla conservazione, il file riversato è una copia digitale di un documento digitale e, come tale, la conformità della copia è attestata in base alla normativa vigente, inclusa la certificazione di processo come riportata nell’Allegato 3 delle presenti Linee guida». In sostanza la conformità del documento riversato rispetto all’originale deve essere attestata da un notaio o da altro pubblico ufficiale a ciò autorizzato (ottenendo così un valore probatorio ‘privilegiato’) oppure da un soggetto privato (in questo caso il valore probatorio è ‘semplice’). L’Allegato 2 specifica altresì che «quando si effettua un riversamento finalizzato alla conservazione del file si può e, in certi casi previsti dalla legge, si deve, conservare anche la copia del file nel formato originario. In entrambi questi casi il file originario è conservato indipendentemente dal suo formato di file originario, purché sia conservata anche – in una forma logicamente e univocamente legata ad esso – copia conforme del medesimo in un formato adatto alla conservazione». In questo senso la soluzione basata sul PDF/A-3 vista poc’anzi, capace di conservare gli allegati nel loro formato originario, sembra costituire una buona soluzione sotto il profilo tecnologico, anche se sarebbero auspicabili ulteriori valutazioni sotto il profilo diplomatico.

del paragrafo, che la *task force* individua nel formato PDF e, in particolare, nel profilo PDF/A, uno dei migliori candidati al ruolo di formato di elezione per la conservazione degli archivi di posta elettronica.

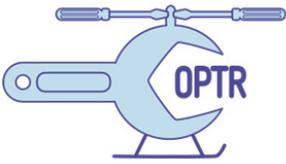
Un altro progetto che merita una segnalazione è RATOM (Review, Appraisal, and Triage of Mail) dell'Università della Carolina¹⁶. RATOM ha l'obiettivo di sviluppare un'applicazione software per assistere le istituzioni della memoria nelle attività di valutazione, selezione e scarto degli archivi di posta elettronica. Infatti, nonostante i notevoli progressi compiuti nello sviluppo di sistemi di conservazione digitale, sono relativamente pochi gli studi e le soluzioni che sono state proposte, anche a livello di applicazioni software, per supportare queste attività fondamentali¹⁷.

Altro progetto degno di nota è COPTR (Community Owned digital Preservation Tool Registry)¹⁸. Si tratta di un registro di strumenti di conservazione digitale pensati per supportare la conservazione della posta elettronica basato su una piattaforma wiki. Il suo scopo principale è quello di offrire una raccolta di strumenti che possono essere utili per affrontare particolari sfide legate alla conservazione in generale. Può essere consultato e ricercato direttamente dagli operatori, oppure interrogato da altri sistemi tramite un'API. Ogni strumento in COPTR è classificato in base alla fase del ciclo di vita in cui opera, alla funzione che svolge e al tipo di contenuto su cui è in grado di lavorare (Fig. 3). Sono descritti poco meno di seicento strumenti, 33 dei quali sono utilizzabili per l'archiviazione e la conservazione della posta elettronica (come Aid4Mail, DArcMail, Emailchemy, MailStore Home).

¹⁶ RATOM (Review, Appraisal, and Triage of Mail) Project, <https://ratom.web.unc.edu> (ultima consultazione: 26/05/2022).

¹⁷ Il progetto estende le capacità di elaborazione delle email attualmente disponibili nello strumento software TOMES e nell'ambiente BitCurator, sviluppando moduli aggiuntivi ed anche soluzioni software autonome in grado di supportare flussi di lavoro più avanzati, come l'identificazione di email che devono essere sottoposte a revisione a causa della presenza di informazioni potenzialmente sensibili o lo sviluppo di moduli software per assistere gli archivisti nella preparazione dei materiali in vista dell'apertura alla consultazione da parte del pubblico.

¹⁸ COPTR (Community Owned digital Preservation Tool Registry) Project, https://coptr.digipres.org/index.php/Main_Page (ultima consultazione: 26/05/2022).



	Create or Receive (Acquire)	Ingest	Preservation Planning	Preservation Action	Access, Use and Reuse	Store	Dispose	Cross-Lifecycle Functions
3D	1	3		5	3	2		3
Audio	5	15	3	17	3	2	1	14
Binary Data	1	3	1	4	1	2		6
Container	1	1		5		1		
Database	1	5		16	2	1		1
Disk Image	12	6		4	1	3		8
Document	4	29	1	24	8	1	1	22
EBook		2		2				1
Email	5	5	2	5	6	3		7
Geospatial		1						
Image	9	31	1	31	6	1		18
Metadata	5	21	1	8	6	3		5
Project Management Data	1			1	1			
Research Data	10	5	9		2	16		6
Software	1	2		3		1		2
Spreadsheet	1	5		4				2
Video	4	14	3	19	2	2	1	11
Web	34	8	3	9	5	4		8
Not content-type specific	28	42	101	14	91	84	29	75

Figura 3: L'elenco di strumenti per la conservazione digitale censiti in COPTR.

Molto interessante è il progetto ePADD, condotto dalla Special Collections & University Archives della Stanford University¹⁹. Esso si prefigge lo scopo di sviluppare un software libero e *open source* per supportare le difficoltà che i vari soggetti coinvolti (donatori di archivi di posta elettronica, archivisti, ricercatori, etc.) si trovano ad affrontare nelle operazioni di valutazione, acquisizione, ricerca e fruizione di archivi di posta elettronica con un valore storico e culturale. Ciò comprende, ad esempio, il vaglio dei messaggi di posta elettronica alla ricerca di informazioni confidenziali, riservate o protette dalla legge, la preparazione delle email per la conservazione e la possibilità di rendere i file risultanti (che incorporano le azioni di conservazione intraprese

¹⁹ ePADD Project, <https://library.stanford.edu/projects/epadd> (ultima consultazione: 26/05/2022).

dall'archivio) accessibili ai ricercatori. Per raggiungere queste finalità ePADD si basa su tecniche della linguistica computazionale e dell'intelligenza artificiale, tra cui il *machine learning*, il *natural language processing* e il *named entity recognition*. Il progetto è iniziato nel 2010 e ha ricevuto finanziamenti dalla National Historical Publications & Records Commission dal 2012 al 2015 per sviluppare la prima versione completa del pacchetto software. Dal 2015 al 2018, il progetto ha ricevuto finanziamenti dall'Institute of Museum and Library Services per sviluppare altre sei versioni di ePADD. Nel biennio 2020-2021 è stato finanziato dalla Andrew W. Mellon Foundation per assicurare lo sviluppo del software e nel 2021 ha ricevuto un finanziamento dal programma "Email Archiving: Building Capacity and Community", gestito dall'Università dell'Illinois a Urbana-Champaign, per integrare in ePADD le funzionalità di conservazione a lungo termine degli archivi di posta elettronica e supportare così l'intero ciclo di vita dell'archivio. Il nuovo progetto è stato intitolato "Integrating Preservation Functionality into ePADD" (ePADD+) ed è attualmente in corso di svolgimento.

L'ultimo progetto che merita una segnalazione è l'Email Archiving PDF Project (EA-PDF), che sembra il più promettente. Finanziato dalla Andrew W. Mellon Foundation nel 2019, è il frutto della collaborazione tra l'Illinois University, i National Archives and Records Administration (NARA) e la Library of Congress degli Stati Uniti, la PDF Association²⁰ e altre istituzioni (Duff 2019). Il progetto intende sviluppare un profilo del formato PDF specificatamente pensato per la conservazione dei messaggi di posta elettronica. A tal fine è stato condotto uno studio per comprendere come i messaggi di posta elettronica possano essere convertiti in "contenitori" PDF mantenendo le loro proprietà significative (*significant properties*)²¹ in modo da creare documenti elettronici autentici e completi che soddisfino le esigenze di conservazione a lungo termine (PDF Association 2021). Ciò ha portato alla pubblicazione di un rapporto finale, intitolato "A Specification for Using PDF to Package and Represent Email" (EA-PDF Working Group 2021), che stabilisce i requisiti funzionali di alto livello per l'utilizzo della tecnologia PDF ai fini della con-

²⁰ Il sito della PDF Association, <https://www.pdfa.org>; in particolare si veda la pagina dedicata al progetto EA-PDF, <https://www.pdfa.org/resource/ea-pdf/> (ultima consultazione: 26/05/2022).

²¹ Uno studio approfondito delle proprietà significative delle varie categorie di formati elettronici è stato condotto nell'ambito del progetto INSPECT (Investigating the Significant Properties of Electronic Content Over Time). Il progetto è stato finanziato dall'ente britannico JISC tra il mese di marzo 2007 e il mese di marzo 2009 nell'ambito del programma "Repositories and Preservation". È stato guidato dall'Arts and Humanities Data Service (AHDS) Executive fino alla sua scomparsa nel marzo 2008, e poi dal Centre for e-Research (CeRch) del Kings College di Londra, con la collaborazione degli Archivi Nazionali Britannici (The National Archives). I risultati del progetto sono tutt'ora disponibili all'indirizzo: <https://significantproperties.kdl.kcl.ac.uk> (ultima consultazione: 26/05/2022).

servazione a lungo termine di messaggi di posta elettronica, sia che si tratti di email singole che di intere cartelle o intere caselle di email. Va fatto notare che nella redazione del rapporto sono state recepite le osservazioni e i suggerimenti provenienti dalla comunità scientifica, dalle istituzioni governative e dal mondo delle imprese. Sotto il profilo concettuale, il formato EA-PDF presenta lo stesso grado di complessità del messaggio di posta elettronica da cui deriva, ma è capace di rappresentare tale complessità in un modo formalmente definito, utilizzando le strutture del formato PDF (che è uno standard riconosciuto a livello mondiale)²², al contrario di altri formati che – come l’EML e l’MBOX – sono definiti più dalle implementazioni dei client di posta elettronica che da specifiche autorevoli. Inoltre, il fatto che EA-PDF sia un profilo del PDF, un formato dalle caratteristiche eccellenti (si tratta del formato più diffuso al mondo, dotato di specifiche aperte e ben documentate, interoperabile e ricco di funzioni, fruibile senza alcuna difficoltà utilizzando qualsiasi piattaforma hardware e software) gioca senz’altro a suo favore nel confronto con le altre proposte. Basti pensare al fatto che un archivio di posta elettronica in cui i messaggi sono codificati secondo il formato EA-PDF è fruibile oggi – e lo sarà anche in futuro – semplicemente utilizzando un qualsiasi lettore PDF²³, compreso quello nativamente integrato in qualsiasi browser. Alla fine del 2021, nell’ambito della PDF Association si è costituito il gruppo di lavoro “Email Archiving Liaison Working Group (EA-LWG)” con l’obiettivo di continuare il lavoro precedentemente svolto e arrivare alla definizione delle specifiche tecniche del profilo EA-PDF per l’archiviazione a lungo termine degli archivi di posta elettronica²⁴. Dati i presupposti, EA-PDF sembra poter essere una soluzione concreta e affidabile – e, forse, definitiva – al problema della conservazione a lungo termine delle email, replicando in qualche modo la strada che il formato PDF/A ha percorso nel campo dei documenti a prevalente contenuto testuale, dove dopo essere stato proposto nell’ormai lontano 2005, ha conquistato nel tempo la fiducia della comunità archivistica ed è diventato il formato di elezione per la conservazione a lungo termine di tali tipologie di documenti sia nel mondo delle pubbliche amministrazioni che in quello delle imprese.

²² Si tratta dello standard ISO 32000, di cui attualmente sono state rilasciate due versioni: la prima, pubblicata nel 2008, è intitolata “ISO 32000-1:2008 – Document management – Portable document format – Part 1: PDF 1.7”; la seconda, pubblicata dapprima nel 2017 e poi aggiornata nel 2020, è intitolata “ISO 32000-2:2020 – Document management – Portable document format – Part 2: PDF 2.0”.

²³ Infatti, tutti i profili del formato PDF (come il PDF/A per l’archiviazione a lungo termine, il PDF/E per l’ingegneria, il PDF/X per la stampa, il PDF/R per la digitalizzazione, il PDF/UA per l’accessibilità universale, etc.) sono fruibili utilizzando un qualsiasi lettore PDF.

²⁴ Il sito di riferimento del gruppo di lavoro è <https://www.pdfa.org/community/ea-pdf-lwg> (ultima consultazione: 26/05/2022).

6. Conclusioni

Con l'adozione del profilo EA-PDF, la questione della scelta del formato per la conservazione a lungo termine dei messaggi di posta elettronica sembra avviarsi verso una soluzione, anche se rimangono ancora da affrontare altre questioni. Ad esempio, sotto il profilo organizzativo – sia dal punto di vista dei soggetti produttori di archivi di posta elettronica (pubbliche amministrazioni, imprese, singoli individui, etc.) che da quelle delle istituzioni deputate alla conservazione della memoria – i nodi da sciogliere sono ancora numerosi: come considerare i messaggi di posta elettronica inviati e ricevuti dagli impiegati di una organizzazione, sia essa pubblica o privata? Si tratta di messaggi di proprietà dei singoli individui e, come tali, devono confluire nei loro archivi di persona, oppure di messaggi che devono confluire nell'archivio dell'organizzazione stessa – dal momento che sono associati alla casella 'istituzionale' assegnata al dipendente – e, come tali, da tutelare e conservare, ai sensi del Codice dei Beni culturali? Va approfondito anche il tema dell'uso "promiscuo" della posta elettronica istituzionale: il dipendente di una organizzazione deve utilizzare la casella di posta elettronica assegnatagli (email istituzionale) solo per finalità legate al lavoro o può utilizzarla anche per finalità che fuoriescono da quell'ambito²⁵? E come gestire le situazioni "ibride" in cui nell'archivio di posta elettronica confluiscono sia messaggi afferenti alla sfera personale che messaggi afferenti alla sfera lavorativa, situazione questa ormai ampiamente diffusa²⁶? Vanno attentamente esplorate anche le questioni legate alla protezione dei dati personali, alla privacy, al diritto all'oblio e all'eredità digitale, che possono condizionare pesantemente il futuro degli archivi di posta elettronica.

Per affrontare in maniera decisa il problema della gestione e conservazione della posta elettronica il primo passo da fare è quello di definire delle *policy* basate su regole chiare, puntuali e condivise da tutti, come quella di mantenere separate le email personali da quelle professionali o quella di organizzare l'archivio di posta elettronica creando cartelle per l'archiviazione delle email importanti (ad esempio, tutte le email inviate e ricevute con un particolare corrispondente o tutte le email relative a un particolare progetto o attività), evitando, cioè, di lasciare i messaggi nelle cartelle della posta in arrivo e della

²⁵ Su questo si veda l'interessante indagine in ordine alla gestione e conservazione dei messaggi di posta elettronica rivolta alle pubbliche amministrazioni italiane elettronica nell'ottobre 2019 pubblicata in (Guarasci 2020).

²⁶ In alcune situazioni non è così netta la distinzione tra la sfera professionale e quella personale: si pensi, ad esempio, alla posta elettronica di personaggi che ricoprono cariche politiche, sia a livello centrale (deputati, senatori) che a livello periferico (sindaci, governatori, presidenti di consigli comunali e regionali), nelle cui caselle di posta elettronica possono sedimentarsi non solo email di 'lavoro' ma anche email che hanno contenuti più strettamente personali, anche se in qualche modo attinenti anch'esse all'incarico rivestito. Su questo si veda (Allegrezza e Gorgolini 2016).

posta in uscita. Queste attività di gestione della posta elettronica devono essere svolte regolarmente anche si tratta di operazioni che indubbiamente richiedono molto tempo. Le policy devono definire anche quali sono le categorie di email significative per l'amministrazione – vuoi per ragioni amministrativo-giuridiche, vuoi per motivazioni di ordine storico-culturale – e che quindi devono essere conservate.

Dopo la definizione delle policy, il passo successivo è la scelta degli strumenti tecnologici (Prom 2020) da utilizzare per le varie operazioni, come la conversione dal formato nativo delle email al formato di conservazione (ad esempio, l'EA-PDF).

Da ultimo, è importante cominciare a ragionare sulle possibili applicazioni delle *disruptive technologies* – come quelle basate sulle tecniche di intelligenza artificiale – che potrebbero avere un ruolo di primaria importanza nella gestione e conservazione degli archivi di posta elettronica, ad esempio per automatizzare parzialmente o del tutto operazioni come la classificazione, la creazione delle aggregazioni di messaggi di posta elettronica o le operazioni di selezione e scarto. È questo forse il campo più promettente, come dimostrano i progetti di ricerca che, a livello internazionale, stanno indagando sul rapporto tra gli archivi e tali nuove tecnologie²⁷.

Come si è visto, conservare gli archivi di posta elettronica è un compito complesso e richiede la collaborazione di tutte le figure a vario titolo coinvolte: dai soggetti produttori agli archivisti, agli informatici, ai giuristi, ai professionisti dell'informazione. La strada verso la soluzione definitiva è ancora lunga, ma si cominciano finalmente ad intravedere soluzioni promettenti.

Riferimenti bibliografici

- Agenzia per l'Italia Digitale. 2019. *Formati di file e riversamento. Allegato 2 al documento "Linee Guida sulla formazione, gestione e conservazione dei documenti informatici"*. https://www.agid.gov.it/sites/default/files/repository_files/allegato_2_formati_di_file_e_riversamento.pdf.
- Allegrezza, Stefano. 2021. "Il problema dell'eredità digitale nella trasmissione di archivi e biblioteche personali." *Bibliothecae.it*, 10 no. 1: pp. 352-400. <https://bibliothecae.unibo.it/article/download/13074/12626>.
- Allegrezza, Stefano, e Luca Gorgolini. 2016. *Gli archivi di persona nell'era digitale. Il caso dell'archivio di Massimo Vannucci*. Bologna: Il Mulino.

²⁷ Tra gli altri si segnala il progetto InterPARESTrustAI, il progetto di ricerca internazionale con durata sessennale (2021-2016) finanziato dal Social Sciences and Humanities Research Council of Canada (SSHRC), che ha l'obiettivo di indagare le applicazioni dell'intelligenza artificiale agli archivi. <https://interparestrustai.org> (ultima consultazione: 26/05/2022).

- Artefactual Systems e Digital Preservation Coalition (DPC), Preserving Email Data Types Series. <https://www.dpconline.org/docs/technology-watch-reports/2472-preserving-email/file>.
- Cencetti, Giorgio. 1937. "Sull'archivio come «universitas rerum»." *Archivi*, IV: 7-13.
- Cencetti, Giorgio. 1939. "Il fondamento teorico della scienza archivistica." *Archivi*, VI: 7-13
- COPTR (Community Owned digital Preservation Tool Registry) Project. https://coptr.digipres.org/index.php/Main_Page.
- Council of Library and Information Resources (CLIR). 2018. *The future of email archives, A Report from the Task Force on Technical Approaches for Email Archives*, Agosto 2018. <https://www.clir.org/pubs/reports/pub175>.
- D'Arminio Monforte, Alessandro. 2018. *L'amministrazione dei beni ereditari. Chiamato all'eredità, curatore dell'eredità giacente ed esecutore testamentario*. Pisa: Pacini Editore.
- D'Arminio Monforte, Alessandro. 2020. *La successione nel patrimonio digitale*. Pisa: Pacini Editore.
- Digital Preservation Coalition (DPC). 2018. *Preserving Email. Digital Preservation Topical Note 7*. <https://www.dpconline.org/docs/knowledge-base/1868-dp-note-7-preserving-email/file>.
- Duff, Johnson. 2019. *Archiving email into PDF containers: A Mellon Foundation project*. <https://www.pdfa.org/archiving-email-into-pdf-containers-a-mellon-foundation-project>.
- Duranti, Luciana. 2007. "An Overview of InterPARES 3 (2007-2012)." *Archives & Social Studies: A Journal of Interdisciplinary Research* 1, no. 1 (September 2007): 586.
- EA-PDF Working Group. 2021. *A specification for using PDF to package and represent email*, University of Illinois at Urbana-Champaign. <https://www.ideals.illinois.edu/handle/2142/109251>.
- ePADD Project. <https://library.stanford.edu/projects/epadd>.
- Guarasci, Roberto, Francesca Parisi, e Anna Rovella. 2020. "Gestione e conservazione dei messaggi di posta elettronica." *AIDAinformazioni* 38, no. 1-2 : 149-75.
- Guercio, Maria. 2019. *Archivistica informatica. I documenti in ambiente digitale*. Roma: Carocci Editore.
- InterPARES3 Project. 2009. *General Study 05 – Keeping and Preserving E-mail*. http://www.interpares.org/ip3/display_file.cfm?doc=ip3_italy_gs05a_final_report.pdf.

- InterPARES3 Project. 2011. *General Study 05b – Guidelines and Recommendations for E-Mail Records Management and Long-Term Preservation*. http://www.interpares.org/ip3/display_file.cfm?doc=ip3_italy_gs05b_final_report.pdf.
- PDF Association. 2015. *Email archiving with PDF/A*. <https://www.pdfa.org/email-archiving-with-pdfa>.
- PDF Association. 2021. *Packaging email archives using PDF*. <https://www.pdfa.org/packaging-email-archives-using-pdf>.
- Prom, Chris. 2011. *Preserving Email, DPC Technology Watch Report 11-01*. <https://www.dpconline.org/docs/technology-watch-reports/739-dpctw11-01-pdf/file>.
- Prom, Chris. 2019. *Preserving Email (2nd Ed), DPC Technology Watch Report 19-01*. <https://www.dpconline.org/docs/technology-watch-reports/2159-twr19-01/file>.
- Prom, Chris. 2020. *The future of past email is PDF*. <https://www.pdfa.org/the-future-of-past-email-is-pdf>.
- RATOM (Review, Appraisal, and Triage of Mail) Project. <https://ratom.web.unc.edu>.
- The National Archives. *Managing emails*. <https://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/managing-email>.
- The Radicati group. 2015. *Email Statistics Report, 2015–2019*. <https://www.radicati.com/wp/wp-content/uploads/2015/02/Email-Statistics-Report-2015-2019-Executive-Summary.pdf>.
- Von Seggern, Dietrich. 2021. *Emails for eternity*. <https://www.pdfa.org/emails-for-eternity>.

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 40

N. 1-2 – gennaio-giugno 2022

Contributi

STEFANO ALLEGREZZA

La conservazione degli archivi di posta elettronica: sviluppi recenti e prospettive future

ANDREA BELLANDI

Le Risorse Linguistiche nell'era del Web Semantico. Un insieme di servizi informatici per la gestione di lessici e terminologie

GIORGIA DI MARCANTONIO

E se l'archivio non rispecchia l'istituto? Pavone e il rispecchiamento: analisi di una bozza preliminare

GERARDO GIARDIELLO, CHIARA FIORAVANTI,
FRANCESCO ROMANO, MARIASOLE RINALDI

Strumenti per il reperimento di informazioni rilevanti in Rete. Applicazioni nel dominio della comunicazione pubblica in materia di immigrazione in Italia

CLAUDIA LANZA

Termini e testi nella comunicazione dei gruppi No-vax

ANNA ROVELLA

La posta elettronica negli archivi di persona: conservazione e accesso

GIOVANNI SALUCCI

Il Blog di qualità. Proposta di un nuovo format di pubblicazione nell'editoria scientifica

Note e Rubriche

ROBERTO GUARASCI

Paul Otlet e Suzanne Briet. Note in margine al libro di Paola Castellucci e Sara Mori

ERIKA PASCERI

L'evoluzione del Nuovo Glossario di biblioteconomia e scienza dell'informazione ai tempi dello Smart Working

CLAUDIO GNOLI

Come mi vuoi, canonica o rivoluzionaria?

CLAUDIO GRIMALDI

L'universo affascinante del colore: le sociocromie



mundaneum

In copertina

Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-163-1



9 791259 651631

ISSN 1121-0095



9 770112 100950