



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
DIPARTIMENTO DI BENI CULTURALI

Paul Gabriele Weston - Emmanuela Carbè - Primo Baldini

*Se i bit non bastano:
pratiche di conservazione del contesto di origine
per gli archivi letterari nativi digitali*

Se la raccolta e la gestione di documentazione nativamente digitale comportano già un notevole impegno, assai maggiore è lo sforzo richiesto per archivarlo in modo tale da garantirne una fruizione costante, specialmente quando tale fruizione si intenda prostrarla nel lungo periodo. In quest'ultimo caso la difficoltà non è rappresentata tanto dall'entità dei file da archiviare, anche se passando da un archivio tradizionale a uno digitale i numeri tendono in genere a crescere esponenzialmente, ma dalle conseguenze di quell'evoluzione delle tecnologie, sia hardware che software, che sembra inarrestabile, spinta non soltanto dal comprensibile desiderio di sostituire un prodotto meno evoluto e performante con un altro più affidabile e funzionale, ma dall'esigenza di smuovere il mercato e favorire la sostituzione del parco-macchine e degli applicativi, anche a costo di cannibalizzare un prodotto esistente con l'arrivo di uno nuovo, prima ancora che il precedente abbia saturato le vendite ed espresso pienamente le proprie potenzialità.¹ Nel caso degli archivi digitali l'evoluzione ha riguardato

¹ La storia delle tecnologie presenta numerosi casi di prodotti potenzialmente in grado di offrire grandi prestazioni, il cui sviluppo si è arrestato per ragioni di

di volta in volta il sistema operativo, gli applicativi, i supporti di archiviazione, i dispositivi di scrittura. Preservare a lungo termine le memorie collettive e personali degli ultimi decenni è un'impresa resa particolarmente complessa dalla necessità di integrare competenze appartenenti ad ambiti considerevolmente diversi: discipline letterarie, tecniche archivistiche, tecnologia dell'informazione, questioni giuridiche, aspetti amministrativi. Inoltre, la gestione dell'archivio digitale presuppone l'aggiornamento costante dei modelli di dati, degli standard e delle procedure per far fronte alla crescente varietà delle fonti documentarie.²

Nel definire cosa sia un oggetto "nativamente digitale", Erway (2010), dopo aver fornito un elenco di tipologie di materiali, indica i rischi ai quali con maggior frequenza gli oggetti digitali sono esposti, nonché le questioni per le quali chi si impegna ad archiviare e rendere disponibili risorse digitali deve trovare soluzioni adeguate e sostenibili. Tra queste ultime, ricordiamo la stratificazione delle versioni, la tutela dei dati sensibili, la gestione dei diritti, le licenze d'uso degli applicativi software (nell'evenienza di trasferimenti di pacchetti software dal soggetto utilizzatore a quello che si occupa della conservazione), l'organizzazione e le responsabilità della curatela digitale e, infine, il rapporto con l'utenza. L'allestimento di un deposito finalizzato alla preservazione a lungo termine dei materiali nativamente digitali creati da scrittori, giornalisti e scienziati in seguito e a supporto della propria attività letteraria ed editoriale vede proprio in tali problematiche

natura commerciale, a beneficio di prodotti meno avanzati, ma per i quali esisteva un più ampio magazzino di opere da pubblicare. Possiamo qui ricordare il sistema di videoregistrazione magnetica Betamax, sviluppato dalla Sony, e il Video2000 della Philips, entrambi tecnicamente superiori allo standard VHS della JVC, presto abbandonati in quanto non in grado di conquistare una fetta significativa del mercato, lasciando alla JVC il monopolio delle videocassette che durò circa un ventennio e si concluse all'apparire dei DVD.

² Si veda: Allegrezza 2016, p. 42-58.

l'aspetto più delicato e di maggiore complessità, dalla cui soluzione dipende in buona sostanza il successo dell'iniziativa.

Ancor prima che l'archivio incominciasse a delinearci, PAD - Pavia Archivi Digitali³ ha riflettuto a lungo sulla possibilità che un progetto con queste ambizioni fosse realizzabile e sostenibile, e una volta effettuate le prime scelte riguardanti la conformazione del sistema e l'allestimento delle funzionalità di caricamento dei dati, ha ritenuto di doversi confrontare con gli esperti che in progetti di altra natura, ma con caratteristiche comparabili, avevano affrontato problemi analoghi.⁴ Se il confronto sui metodi e la condivisione dei dati dovrebbe essere una procedura ovvia in tutte le circostanze, lo è a maggior ragione quando la strada che si sta iniziando a percorrere è piena di incognite e di insidie.

La nascita del progetto PAD – Pavia Archivi Digitali

È certamente noto che presso l'Università di Pavia esiste il Fondo Manoscritti, una felice intuizione di Maria Corti, che nel 1969, dopo aver ricevuto in dono alcuni taccuini di Montale, comprese l'importanza di dare vita a un progetto che si ponesse l'obiettivo di raccogliere e conservare le carte di scrittori, artisti e uomini di cultura per mettere a disposizione dei ricercatori quel prezioso patrimonio documentario, rappresentato da manoscritti e dattiloscritti, epistolari, materiale fotografico e altro, che rischiava, altrimenti, di andare

³ <<http://pad.unipv.it>>, cfr. in particolare, per un primo quadro introduttivo, la sezione "Progetto".

⁴ PAD ha dedicato nel 2016 un primo seminario tecnico ad un confronto su questi temi il 4 marzo 2016 con Fabio Ciotti e Salvatore Vassallo. Inoltre il progetto, nei suoi diversi aspetti è stato illustrato e discusso in alcuni workshop in Italia e all'estero. Lo studio di PAD è stato, inoltre, oggetto dell'apprezzata tesi di laurea magistrale di Michele Stefani (*Gli archivi degli scrittori nativi digitali: verso la conservazione e oltre*, tesi di laurea in Storia e Gestione del Patrimonio Archivistico e Bibliografico, Università Ca' Foscari Venezia, a.a. 2015/16, relatore prof. Paul Gabriele Weston, correlatore prof. Riccardo Ridi).

disperso. Proprio con questa finalità, nel 1980 è stato costituito il Centro per gli studi sulla tradizione manoscritta di autori moderni e contemporanei.

Nel 1997, rievocando nel libro *Ombre dal Fondo* gli inizi del Centro, Maria Corti (1997, p. 87) assegna alla sua creatura una missione che si applica allo stesso modo (ma forse con un supplemento di difficoltà) al progetto PAD:

Al di là degli eventi che passano, le Carte durano, ciascuna con la sua minuscola storia e vivono in quella che Borges chiama la nostra «quarta dimensione, la memoria». E quando anche noi ce ne andremo, loro le Carte resteranno lì e non sapranno mai che non ci siamo più.

Nei suoi primi anni di esistenza, PAD si configura come una struttura autonoma rispetto al Centro Manoscritti. Quando nel 2009 Beppe Severgnini, allora Presidente dell'Associazione Alumni dell'Ateneo pavese, pone la domanda: “Quanti anni può durare una memoria digitale?”, la sfida di progettare un archivio digitale di scrittori, giornalisti e intellettuali contemporanei è lanciata. Le motivazioni non sono molto diverse da quelle sottese all'allestimento di un archivio di manoscritti, ma tengono conto del fatto che negli ultimi anni il panorama della comunicazione scritta è radicalmente mutato: il supporto cartaceo sta gradualmente lasciando il campo all'informatica e sempre più persone utilizzano computer, tablet e cellulari. Scrittori, giornalisti e ricercatori non fanno eccezione. Se in un primo momento questi dispositivi vengono utilizzati sostanzialmente nello stesso modo in cui prima era stata utilizzata la macchina da scrivere,⁵ con il tempo

⁵ Interessante a tale proposito la lettera di Franco Fortini all'amico Carlo Fini (18 settembre 1985), cit. in (Lenzini 2016): «non è senza singolare emozione che mi accingo a scriverti questa letterina. Essa è il primo frutto del premio Librex-Montale perché, con la fanciullaggine che si addice alla mia età, la prima cosa che mi sono precipitato a fare con quei soldi [...] è stato di acquistare un da lungo

gli autori aprono siti, gestiscono blog, e, in seguito alla diffusione delle pratiche di social networking, intervengono su Facebook o Twitter,⁶ solo per citare i due esempi più significativi. La componente social diventa così parte delle loro abitudini di scrittura, luogo di confronto con i lettori, fonte di ispirazione per nuove opere, e così via.⁷ Tutte queste attività determinano lo spostamento di una parte dei loro archivi dalla memoria del computer in loro possesso al cloud condiviso nello spazio del web.

Il digitale modifica le più o meno consolidate proporzioni esistenti negli archivi tradizionali tra materiale testuale e altre tipologie di documenti. Gli archivi digitali nativi si popolano così di fotografie, interviste radiofoniche, videoregistrazioni di presentazioni, conferenze e premi, scansioni di recensioni, di copertine e di altri materiali, copie di siti e di altre risorse elettroniche, in una parola di tutto ciò che lo scrittore nel corso degli anni ha potuto trasferire nella capace memoria del proprio computer semplicemente agendo su un clic. Tra le conseguenze della maggiore facilità di trattare gli oggetti digitali alcune hanno un grande rilievo quando si tratta di organizzare l'archivio ai fini della sua conservazione: molti file vengono copiati, integralmente o in parte, per confluire, magari con un nome diverso, in più di una cartella; procedure di copia e incolla determinano la presenza di materiali creati da più soggetti (parti di lettere,

tempo convoité, o, se vuoi, appetito word processor: ossia computer per scrittura; che da tre giorni mi tiene occupato come una amante giovane tiene un vegliardo».

⁶ Sul tema si veda l'inchiesta *Scrittori e Facebook* (Lombardi 2016), pubblicata a puntate su «Le parole e le cose» tra febbraio e maggio 2016: tra gli scrittori intervistati anche Francesco Pecoraro e Franco Buffoni, che hanno conferito il loro archivio a PAD rispettivamente nel 2015 e 2016.

⁷ PAD ha dedicato nel 2016 due seminari per un confronto su questi temi: il 31 marzo 2016 con Gilda Policastro (per cui cfr. Policastro 2012) e Francesco Pecoraro; il 10 maggio con Paolo Costa, Alessandro Gazoia e Mirko Volpi (cfr. <<http://pad.unipv.it/content/seminari-letterari>>).

commenti, recensioni, ecc.) all'interno di un file, la cui presenza, oltre a costituire un elemento spesso estraneo all'archivio, potrebbe rivelarsi problematica nel momento in cui venissero pubblicati; file privati, come lettere e fotografie, si trovano all'interno di cartelle di lavoro e debbono pertanto essere rimossi prima che sia conclusa la procedura di conferimento. Sono tre situazioni differenti che tuttavia richiedono che ogni file sia sottoposto a qualche forma di screening, il che ovviamente presuppone una stretta collaborazione tra l'archivista e lo scrittore nel momento in cui i file vengono riversati sul server, ma che al tempo stesso possono richiedere un notevole dispendio di tempo se i file che costituiscono l'archivio sono nell'ordine delle decine di migliaia.

Le differenze esistenti tra la struttura di un archivio cartaceo e quella dell'omologo digitale e la necessità di dar vita a una infrastruttura e a una organizzazione in grado di assicurare la corretta gestione delle questioni appena ricordate, ha convinto l'Università a fare di PAD una struttura autonoma, almeno temporaneamente distinta da quella del Centro, garantendole in questo modo quella flessibilità che le sarebbe potuta altrimenti mancare. È del tutto ragionevole che, una volta stabilizzata la struttura di PAD, portati a compimento il caricamento e la descrizione di un paio di archivi e individuata l'organizzazione occorrente a svolgere efficientemente le procedure occorrenti al funzionamento del sistema, PAD e il Centro Manoscritti si consolidino in un'unica struttura amministrativa che possa rappresentare il luogo nel quale si raccolgano le Carte, indipendentemente dal supporto che le ospita. Occorre anche osservare che nel prossimo futuro saranno sempre più numerosi gli archivi ibridi, anche da parte di scrittori che abbiano già conferito al Centro Manoscritti una parte del proprio archivio. Un primo esempio di questo intreccio tra carta e bit è rappresentato dal fondo Franco Buffoni, poeta e anglista, che ha donato al Centro a partire dal 2012 un cospicuo corpus di materiali manoscritti e dattiloscritti relativi alla sua opera poetica e di

traduttore, al quale ha fatto seguire nel 2016 il conferimento della sua multiforme produzione digitale.⁸

Alcune specificità degli archivi letterari nativi digitali

Già da queste prime considerazioni appare evidente che gli archivi letterari costituiti da documentazione nativamente digitale ricadano in una zona grigia da più punti di vista. Da quello puramente archivistico, considerato che si tratta di materiale corrente prodotti da soggetti privati, che continuano a utilizzarli e non di rado a modificare forma e contenuti dei medesimi documenti e che potrebbero persino decidere di riacquisire i propri file rimuovendoli dall'archivio di conservazione qualora, come è avvenuto in un caso, abbiano precisato una data di scadenza del contratto. Poiché non si tratta di documentazione della pubblica amministrazione, essa non rientra tra quella gestita all'interno della filiera della conservazione di lungo termine, secondo le disposizioni prescritte per la documentazione informatica. Infine, non essendo intervenuta la dichiarazione di interesse storico, questa tipologia di archivio non rientra neanche sotto la tutela dei beni culturali.

La letteratura riguardante il trattamento degli archivi nativamente digitali privati è di gran lunga meno abbondante di quella relativa agli archivi di soggetti istituzionali e di imprese, e di conseguenza sono assai meno numerose le indicazioni, sotto forma di linee guida e di buone pratiche, alle quali potersi rivolgere. Nel 2012 OCLC ha avviato un progetto di ricerca denominato *Demystifying Born Digital*,⁹ le cui finalità sono state illustrate da Jackie Dooley (2013a; 2013b), che ha visto la collaborazione di numerosi specialisti nei diversi ambiti del digitale nativo e che ha prodotto una serie di rapporti

⁸ Sulla questione dell'archivio ibrido e delle implicazioni in merito alla necessità di far convivere tradizioni, procedure e standard differenti si veda il caso dell'Archivio di Massimo Vannucci (Allegrezza - Gorgolini 2016).

⁹ <<http://www.oclc.org/research/themes/research-collections/borndigital.html>>.

volti a individuare strategie, obiettivi da perseguire, risorse hardware e software disponibili, competenze tecniche (Erway 2012; Barrera-Gomez, Erway 2013; Erway, Goldman, McKinley 2014; Dooley 2015).

Assai più articolati sono lo scenario e il modello organizzativo elaborati nell'ambito del progetto PARADIGM (Personal Archives Accessible in Digital Media),¹⁰ condotto congiuntamente dalle biblioteche delle Università di Oxford e Manchester, con l'obiettivo di studiare le problematiche connesse all'archiviazione di carte private in formato digitale, nel rispetto dei requisiti archivistici e di quelli relativi alla conservazione del digitale. La ricerca ha goduto di finanziamenti dal 2005 al 2007 nell'ambito del programma "Supporting Institutional Digital Preservation and Asset Management", coordinato dal JISC. Il progetto ha preso le mosse dalla scelta dei personaggi politici contemporanei appartenenti ai due maggiori partiti inglesi, la cui documentazione sarebbe stata utilizzata come banco di prova nell'applicazione delle procedure sviluppate. La Bodleian Library, che ospita i Conservative Party Archives, ha curato le carte personali dell'esponente conservatore, mentre la John Rylands University Library, che collabora con il Labour History Archives and Study Centre (LHASC) di Manchester, ha fatto lo stesso con l'esponente laburista. In entrambi i casi, il progetto si è svolto al di fuori dell'attività corrente dei due archivi, anche perché nessuno dei due conserva la documentazione privata dei parlamentari. Il lavoro è consistito nel condurre in parallelo due procedure di archiviazione, la prima basata sull'applicazione del modello OAIS e l'altra secondo il modello della tradizione archivistica. Le due procedure sono state quindi confrontate per elaborare una serie di linee guida volte ad armonizzarne i criteri e a individuare gli ostacoli che si frappongono a una loro perfetta sovrapposizione. Tra i risultati, uno dei più

¹⁰ Paradigm (Personal Archives Accessible in Digital Media) website, <<http://www.paradigm.ac.uk/index.html>>.

significativi è il *Workbook for Digital Private Papers*,¹¹ nel quale criteri, buone pratiche, riferimenti alla letteratura e ad altri casi di studio prendono la forma di un prezioso strumento di lavoro, non soltanto per coloro che si occupano di carte private, ma per tutti quei casi in cui la documentazione digitale sia stata creata con l'uso di standard, strumenti e tecnologie estranei al controllo degli archivisti, ma la cui autenticità debba essere comunque garantita nel lungo termine.

Il caso degli archivi nativi digitali letterari presenta, rispetto ai casi presi in considerazione nei due precedenti studi, una serie di specificità che richiedono soluzioni ad hoc.¹² Una prima singolarità è costituita dal fatto che difficilmente si tratta esclusivamente di manoscritti digitali. La fase pretestuale, infatti, implica spesso la produzione di una grande varietà di documentazione, comprendente fotografie, materiale cartografico, interviste registrate, appunti, il cui contenuto non viene sempre immediatamente trasferito sulla pagina scritta. Alcuni scrittori hanno preso l'abitudine di organizzare questo materiale all'interno di cartelle denominate "Nuclei embrionali" o simili; nei casi più fortunati le cartelle sono associate a quelle relative alla singola produzione letteraria e ciò rappresenta, al momento del conferimento e dell'individuazione dei nuclei documentali dell'archivio, un vantaggio di non poco conto. Purtroppo questo modo ordinato di procedere non è particolarmente diffuso. Di gran lunga prevalenti sono gli archivi al cui interno una simile documentazione è sparsa e il collegamento con la specifica pubblicazione è andato perduto. Inoltre, l'archivio letterario contiene opere non finite, che vengono riaperte di tanto in tanto per aggiungere un capitolo, un

¹¹ *Workbook on Digital Private Papers*, <<http://www.paradigm.ac.uk/workbook/index.html>>. Il Workbook è integrato da *Guidelines for creators of personal archives*, <<http://www.paradigm.ac.uk/workbook/appendices/guidelines.html>>.

¹² Tra i progetti pionieri si segnalano quelli dell'Harry Ransom Center (Stollar Peters 2006) e dell'Emory University's Manuscript, Archives and Rare Book Library (Carroll - Farr - Hornsby - Ranker 2011).

brano poetico, un personaggio, o anche semplicemente per apportare qualche correzione al testo esistente. Il manoscritto finisce, dunque, per essere generato attraverso la sovrapposizione di strati successivi di scrittura, ciascuno dei quali andrebbe documentato e preservato, una procedura incredibilmente complessa.

Idealmente un autore dovrebbe fare uso di una piattaforma in cloud computing per tutti i propri documenti, avvalendosi delle funzioni di *versioning* automatiche che offre, ad esempio, Google Docs. In alternativa, il curatore digitale dovrebbe affiancare lo scrittore per prevenire l'insorgere del problema, informandolo sul modo migliore di selezionare e scartare, organizzare e denominare i file, raccomandandogli i formati e i supporti più adeguati alle necessità della conservazione, segnalandogli la necessità di conservare i dispositivi utilizzati (computer, tablet, smartphone, lettori di floppy e dvd, ecc.) accanto all'insieme dei file (Weston 2013, p. 56-65). La definizione che Lee e Tibbo (2007; vedi anche Lee, Tibbo, Schaefer 2007, p. 49-50) forniscono della digital curation

[it] involves the management of digital objects over their entire lifecycle, ranging from pre-creation activities wherein systems are designed, and file formats and other data creation standards are established, through ongoing capture of evolving contextual information for digital assets housed in archival repositories

è senz'altro troppo restrittiva nel caso degli archivi letterari. Nessun archivistà si sognerebbe, infatti, di dettare a un autore i criteri cui attenersi nella fase di scrittura del suo libro. Più realistica nel caso di PAD, e assai più adeguata a rappresentarne le attività, appare la posizione espressa da Cragin (2007):

active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education; [...] activities [which] enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time.

Ne consegue che una delle fasi fondamentali per PAD sia la preservazione del contesto di provenienza sia dal punto di vista logico che informatico. Gli archivi fino a oggi pervenuti¹³ si sono rivelati molto diversi tra loro, e hanno contribuito via via a precisare e perfezionare il processo di acquisizione, la cui fase preliminare risulta imprescindibile per la corretta conservazione e fruizione degli archivi.

Le fasi preliminari del conferimento dell'archivio

Prima di esaminare le diverse fasi nelle quali si articola la procedura di conferimento, nella quale massima attenzione è data alla raccolta di tutte quelle informazioni che consentano di conoscere l'organizzazione che l'autore ha conferito al proprio archivio, anche in relazione alle sue pratiche di scrittura, un accenno alla struttura informatica di PAD è opportuno, perché sia più agevole seguire i passaggi a cui i file vengono sottoposti, dal momento della raccolta alla loro archiviazione definitiva.

L'architettura degli archivi di PAD è basata su sei aree (*staging, info, deposit, working, permanent e database*) che gestiscono fasi distinte, a partire dalla consegna del materiale all'ateneo fino al suo utilizzo per attività di natura didattica o di ricerca. Tutti i processi tecnici e amministrativi, dall'acquisizione dei fondi al controllo del ciclo di vita di ciascun archivio, vengono controllati da un modulo software appositamente sviluppato – QUANDO (Quality control for Archiving and Networking of Digital Objects)¹⁴ – che integra le informazioni inserite dagli operatori con quelle automatiche generate dagli applicativi PAD. Per ragioni di sicurezza i singoli archivi sono tenuti separati l'uno dall'altro.

¹³ Oltre ai già citati Beppe Severgnini, Francesco Pecoraro e Franco Buffoni, hanno conferito i propri archivi Silvia Avallone, Gianrico Carofiglio e Paolo Di Paolo.

¹⁴ Per una più puntuale descrizione del modulo QUANDO si rimanda a (Weston - Carbè - Baldini 2016).

Come già accennato, nella preservazione del contesto e della raccolta dei dati, la preparazione del conferimento è una delle fasi più delicate del processo di acquisizione, ed è la prima procedura gestita da QUANDO, che si avvia con l'apertura di una nuova scheda, nella quale vengono inserite le informazioni relative all'autore e a eventuali altre persone e istituzioni coinvolte (agenti letterari, editor, studi legali, case editrici, Fondazioni).

L'autore riceve quindi una bozza standard del contratto: si tratta di un'offerta al pubblico rivolta «a ogni autore interessato che abbia già pubblicato almeno un'opera con un editore iscritto a una associazione nazionale (italiana o straniera) o “appartenente” a una università». Attraverso il contratto è possibile stabilire quali file debbano essere secretati (ed eventualmente per quanto tempo) e indicare quali dovranno essere le modalità di consultazione dei materiali (esclusivamente presso la sede pavese o anche da remoto). È prevista la possibilità di incrementare l'archivio in seguito a successivi conferimenti,¹⁵ di richiedere un backup in caso di perdita dei dati sul proprio computer, e di ricevere assistenza qualora venissero rilevati dei virus. L'autore può chiedere di personalizzare il contratto mediante clausole specifiche, che verranno sottoposte al comitato scientifico di PAD ed eventualmente vagliate dall'ufficio legale dell'ateneo. Contestualmente riceve una *Guida al conferimento*, che spiega come verrà acquisito e utilizzato il fondo. Nel documento si pone l'attenzione al problema della fluidità intrinseca di questa tipologia di materiale, cercando di sensibilizzare l'autore sulle questioni che dovranno affrontare gli archivisti e gli studiosi: naturalmente, più la donazione è ampia e la copia dell'archivio fedele alla situazione originale, tanto migliore potrà essere la fruizione dei materiali raccolti.

¹⁵ A oggi non si sono ancora verificati casi di incremento di un fondo. È chiaro però che questa eventualità pone diversi problemi di natura archivistica e tecnica: l'archivio dovrà essere aggiornato con un sistema che mantenga separati le strutture e i file acquisiti nel corso di ciascun conferimento.

Già in questa fase gli operatori iniziano a raccogliere le risorse bibliografiche relative al lavoro dell'autore, risorse che vengono almeno in parte presentate sul sito del progetto e poi sistematicamente inserite in un applicativo in fase di perfezionamento, PADManager,¹⁶ che lavora sull'area *database* del sistema PAD, ed è pensato per la descrizione archivistica dei fondi e ispirato al modello FRBRoo¹⁷ e al CIDOC object-oriented Conceptual Reference Model (CRM).¹⁸

Il questionario informativo

Strumento decisivo per la raccolta di informazioni preliminari, il questionario informativo viene proposto all'autore in forma digitale.¹⁹ Il questionario, insieme con tutta la documentazione relativa all'acquisizione, viene inserito nell'area *info* del fondo; in QUANDO sono annotati i nomi degli operatori coinvolti in questa fase, e i dati cronologici relativi all'invio e alla ricezione.

¹⁶ Il software è stato sperimentato da un gruppo di studenti durante le esercitazioni del corso "Lineamenti di Biblioteca digitale" (a.a. 2016/17) dell'Università di Pavia. Come caso di studio è stato scelto l'archivio di Franco Buffoni, la cui organizzazione, eccezionalmente ordinata e strutturata già in origine, si presta molto bene ad attività didattiche e laboratoriali. Gli studenti hanno preparato una bibliografia dell'autore inserita in PADManager servendosi di un template (esportabile) che segue le linee guida previste in Wikipedia. La scelta, che per taluni aspetti presenta qualche inconveniente ed è pertanto sottoposta ad ulteriori riflessioni, è stata operata in funzione della condivisione e del riuso dei dati e nella prospettiva di una pubblicazione sotto forma di Linked Open Data.

¹⁷ <http://www.ifla.org/files/assets/cataloguing/frbr/frbroo_v2.2.pdf>.

¹⁸ <<http://www.cidoc-crm.org/node/202>>. Vedi anche <<http://www.cidoc-crm.org/frbroo/home-0>>.

¹⁹ L'idea di ricorrere ad un questionario è stata di Simone Albonico, componente del Comitato di coordinamento di PAD, che ha attivamente partecipato alla sua formulazione. Quanto alle modalità di compilazione, l'autore può decidere di procedere in autonomia o di ricorrere all'assistenza di un operatore. Quest'ultimo può inserire ulteriori note utili per acquisire informazioni sul contesto d'origine.

Il questionario è suddiviso in quindici aree tematiche. Le domande prevedono perlopiù risposte multiple, con la richiesta di aggiungere commenti e precisazioni, e riguardano anche argomenti non strettamente connessi al materiale, che concernono in modo più generico il rapporto tra il lavoro di scrittura e gli strumenti tecnologici. È suddiviso in quindici aree tematiche:

1. *Rapporto tra archivio personale e archivio conferito*: com'è organizzato l'archivio che ha intenzione di conferire e quali le differenze, se sono rilevanti, con quello originale?
2. *Struttura e contenuti dell'archivio*: saranno apportate modifiche all'archivio prima del conferimento (ad es. nella struttura delle directory, nei nomi delle cartelle, nei contenuti o nei nomi dei file)?
3. *Consultazione*: parte dei contenuti non saranno consultabili per un certo numero di anni? L'archivio sarà consultabile anche da remoto o vi si potrà accedere solo dalla sede di PAD?²⁰
4. *Privacy*: l'archivio contiene dati sensibili, materiali generati da terzi o che potrebbero violare la privacy di altre persone?
5. *Ulteriori conferimenti*: ha intenzione di incrementare l'archivio con successivi conferimenti? Acconsente a che PAD lo contatti per chiedere aggiornamenti?
6. *Altri conferimenti*: ha intenzione di conferire materiale speciale (e-mail, pagine web, social network, newsletter, etc.)?
7. *Computer principale*: che computer utilizza per lavorare? Quando è stato acquisito? È lo stesso da cui provengono i file che saranno conferiti?
8. *Macchine precedenti*: con che frequenza vengono cambiati gli strumenti informatici, e in particolare il computer principale? Che computer venivano utilizzati in precedenza? Le macchine sostituite

²⁰ Le modalità di consultazione, come già detto, vengono stabilite nel contratto che regola l'acquisizione dell'archivio.

- sono ancora nella sua disponibilità? I file che verranno conferiti sono stati generati anche da computer precedenti?
9. *Sistema operativo*: qual è il sistema operativo installato sul computer principale? Utilizza o ha utilizzato anche altri sistemi operativi?
 10. *Altri computer e dispositivi*: utilizza altri computer oltre a quello principale? Utilizza altri dispositivi per l'attività di scrittura e lavoro (ad es. tablet, smartphone)? Se li utilizza, provvede poi a trasferire sul computer principale i testi elaborati altrove? Pensa che alcuni suoi testi siano ora soltanto su dispositivi diversi dal computer principali? I file che ha intenzione di conferire sono stati generati anche da altri dispositivi (ad es. app di tablet e smartphone, macchine fotografiche)?
 11. *Organizzazione dei dati*: conosce, all'incirca, la disposizione all'interno del computer dei documenti? Quando salva un file sceglie una collocazione particolare?
 12. *Software di scrittura e revisione*: quali programmi di scrittura utilizza abitualmente? È una scelta pensata/consigliata/che dipende dal tipo di computer? Tende a salvare a più riprese i suoi file di lavoro, a cambiare il nome dei file per ottenere versioni diverse e progressive? Elimina i file di sue scritture senza tenerne copia? Quando lavora sui testi con i redattori delle case editrici, utilizza strumenti particolari (ad es. commenti su bozze in pdf, documenti in cui vengano evidenziate parti di testo, discussioni via e-mail, etc.)? Nell'archivio sono presenti file frutto di un lavoro condiviso con i redattori?
 13. *Processi di scrittura*: parte del lavoro è cartaceo? Che rapporto c'è tra il lavoro condotto avvalendosi di un supporto cartaceo e quello svolto in ambiente informatico?
 14. *Scrittura e web*: qual è il rapporto con le nuove tecnologie e il web? Legge e-book, si documenta via internet mentre scrive? Ritiene che i nuovi strumenti a disposizione stiano influenzando il suo lavoro? Ha mai tenuto un sito internet o un blog? Se sì, è ancora attivo? Ha mai scritto su siti/blog collettivi?

15. *Social network*: ritiene che i social network siano oggi rilevanti per promuovere al pubblico l'attività di un autore? È iscritto a social network? Se li utilizza abitualmente, pensa che questa attività influisca sui processi di scrittura? Ritiene che i materiali elaborati per i social network possano entrare nell'archivio PAD?

La fase di acquisizione

Il conferimento prevede di norma l'acquisizione del materiale attraverso un hard disk dotato di crittografia hardware.²¹ Un incaricato di PAD si reca dall'autore con un primo quadro di riferimento ottenuto grazie al questionario, che può essere ulteriormente integrato attraverso un colloquio. I casi finora affrontati, come già detto, si sono rivelati molto diversi tra loro: talvolta l'autore aveva già individuato con sicurezza il materiale, in altri casi all'operatore è stata richiesta assistenza per effettuare la selezione tra ciò che fa parte dell'archivio letterario e quei file che sono invece di natura privata. Nel momento in cui si copiano le cartelle e i file si richiede la possibilità di acquisire (attraverso screenshot e/o appunti, nei limiti del rispetto della privacy) la situazione di partenza, in modo da inquadrare nella documentazione il "tavolo di lavoro" dell'autore, soprattutto nei casi in cui si verificano forti differenze con l'archivio che verrà conferito.

Tutte le informazioni relative al processo di acquisizione vengono inserite nel software QUANDO: data, nomi degli operatori presenti, luogo, orario di inizio e di fine del conferimento, hardware utilizzato, eventuali problemi tecnici, nome dell'operatore che prende in custodia

²¹ Sono tuttavia previste altre modalità di acquisizione: il salvataggio di siti internet (sperimentato sul sito personale di Franco Buffoni) o di file provenienti da supporti obsoleti (è il caso di Francesco Pecoraro, che ha consegnato a PAD trentaquattro floppy disk, restituiti all'autore dopo il salvataggio dei dati ivi contenuti). Sono stati effettuati dei test per l'acquisizione di e-mail e di materiali provenienti da Facebook e Twitter, ma su queste tipologie di conferimento naturalmente gravano, più che negli altri casi, una serie di questioni legali che attendono ora serie valutazioni.

l'hard disk, data e ora di arrivo del materiale presso l'ateneo e nomi degli operatori che ricevono l'hard disk e la documentazione legale.

All'arrivo a Pavia i dati sono provvisoriamente trasferiti su un server protetto (area *staging*) in attesa di essere trattati. Sono successivamente attivate, nell'area di deposito, le prime procedure di controllo. Se all'interno dell'archivio fossero presenti file o cartelle compressi (come ad esempio .zip, .rar) il programma PadCleaner li decomprime per visualizzare in via preliminare il quadro generale del conferimento. Lo stesso programma elimina automaticamente cartelle e file nascosti, che potrebbero essere stati generati automaticamente dal computer e contenere, ad esempio, istruzioni per il funzionamento del sistema operativo e dei programmi. PAD parte dal principio che cartelle e file nascosti non siano stati, almeno in parte, generati dall'autore e potrebbero contenere dati sensibili conferiti inconsapevolmente. Eventuali virus sono messi in quarantena; sono eliminati (ma segnalati nella documentazione) solo se compromettono irrimediabilmente un documento: in tal caso si attivano le procedure per tentare di salvare il contenuto del file con conversioni in diverso formato.

Vengono generati degli algoritmi SHA-1, che rimarranno associati per sempre ai singoli file e alle cartelle. Il programma PadPrint genera una lista di tutti i file conferiti, che viene inviata all'autore per avere una convalida dei file. L'autore può segnalare quali file e cartelle debbano essere eventualmente secretate, secondo le condizioni decise in precedenza dal contratto; in caso di ripensamenti può decidere di eliminare un file o un gruppo di file: se l'archivio ha un elevato numero di documenti, o se contiene numerose cartelle compresse, potrebbe infatti aver conferito alcuni materiali inconsapevolmente. Con la firma del documento, poi allegato al contratto che viene protocollato dall'ufficio amministrativo e conservato nell'area *info* dell'archivio, si conclude la fase di acquisizione. Si avviano dunque le procedure per il trattamento dei materiali (area *working*) e per la conservazione a lungo termine (area *permanent*).

Riflessioni conclusive

È evidente, da quanto esposto, quale opportunità offra la possibilità di dialogare direttamente con l'autore e ricevere di prima mano informazioni che non sono importanti soltanto per meglio risolvere eventuali problemi di natura tecnologica e per una più efficiente consegna dei file. La conoscenza dei criteri seguiti dall'autore nell'organizzare i file all'interno del proprio computer, del ruolo che il digitale ha nell'attività di scrittura e nei rapporti intessuti con colleghi, istituzioni e lettori, nonché del modo in cui frammenti embrionali vengono arricchiti e rielaborati, lasciati e ripresi, e il testo si stratifica per interventi successivi, è un patrimonio che il sistema di gestione, sul quale questi elementi vengono riportati, offrirà allo studioso assieme ai file. Questo aspetto può essere senz'altro considerato un valore aggiunto ai fini della ricerca.

Si comprende anche che PAD è un progetto ben lungi dall'essere considerato concluso, sia in relazione alla sua struttura informatica ed organizzativa, sia per quanto attiene agli usi che si potranno fare dei suoi dati, a fini di studio e di attività didattiche. Da un lato gli standard e le buone pratiche offerte dal panorama delle realizzazioni esistenti non sembrano sufficientemente consolidate,²² dall'altro lato alcune questioni, quelle giuridiche in particolare, sono state sollevate troppo di recente per aver prodotto interpretazioni univocamente accettate.

Nonostante ciò, per offrire il proprio contributo nel far sì che non si avveri la profezia espressa da Vint Cerf, Chief Evangelist di Google, ad una conferenza stampa nel febbraio 2015, «We are nonchalantly throwing all of our data into what could become an information black hole without realising it», PAD sta già individuando i prossimi obiettivi, e tra questi l'auspicio di condividere con altri dipartimenti e centri di ricerca afferenti ad atenei italiani la ricerca di metodologie

²² Sulla questione, si veda il paragrafo *Iniziativa di ricerca e "Best practices"* in (Guercio 2016, p. 139-145).

e standard, lo sviluppo di componenti software e la realizzazione di funzioni e servizi per la ricerca e la didattica. Possiamo immaginare che, in un'ottica di partecipazione e condivisione, ossia col mettere a fattor comune le specificità, l'esperienza e le competenze di ciascuno, ciò che può in questo modo derivare da PAD possa trasmettere alle future generazioni queste testimonianze dell'ingegno creativo e dell'attività scientifica nel primo scorcio dell'era digitale.

BIBLIOGRAFIA

- Allegrezza 2016 = Stefano Allegrezza, *Le criticità nella conservazione degli archivi di persona tra passato, presente e futuro*, in *Gli archivi di persona nell'era digitale. Il caso dell'archivio di Massimo Vannucci*, a cura di Stefano Allegrezza e Luca Gorgoglino, Bologna, Il Mulino, 2016, p. 41-72.
- Allegrezza - Gorgoglino 2016 = *Gli archivi di persona nell'era digitale. Il caso dell'archivio di Massimo Vannucci*, a cura di Stefano Allegrezza e Luca Gorgoglino, Bologna, Il Mulino, 2016.
- Barrera-Gomez - Erway 2013 = Julianna Barrera-Gomez - Ricky Erway, *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house*. Dublin, OH, OCLC Research, 2013. Web: <<http://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf>> (ult. cons.: 09.03.2017)
- Carroll - Farr - Hornsby - Ranker 2011 = Laura Carroll - Erika Farr - Peter Hornsby - Ben Ranker, *A Comprehensive Approach to Born-Digital Archives*, «Archiviara», 72 (2011), p. 61-92.
- Corti 1997 = Maria Corti, *Ombre dal Fondo*, Torino, Einaudi, 1997.
- Cragin [et al.] 2007 = Melissa H. Cragin [et al.], *An Educational Program on Data Curation*, Graduate School of Library and Information Science University of Illinois at Urbana-Champaign, 2007. Web: <<http://www.ala.org/acrl/sites/ala.org.acrl/files/content/aboutacrl/directoryofleadership/sections/sts/stswebsite/programs/annual2007programs/annual2007posters/Cragin2.pdf>> (ult. cons.: 09.03.2017)
- Dooley 2015 = Jackie Dooley, *The Archival Advantage: Integrating Archival Expertise into Management of Born-digital Library Materials*, Dublin, Ohio, OCLC Research, 2015.

- Erway 2010 = Ricky Erway, *Defining "Born Digital". An essay*, Dublin, Ohio, OCLC Research, 2010. Web: <<http://www.oclc.org/research/activities/hiddencollections/borndigital.pdf>> (ult. cons.: 09.03.2017)
- Erway 2012 = Ricky Erway, *You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media*. Dublin, Ohio, OCLC Research, 2012. Web: <<http://www.oclc.org/research/publications/library/2012/2012-06.pdf>> (ult. cons.: 09.03.2017)
- Erway - Goldman - McKinley 2014 = Ricky Erway - Ben Goldman - Matthew McKinley, *Agreement Elements for Outsourcing Transfer of Born Digital Content*, Dublin, Ohio, OCLC Research, 2014. Web: <<http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-born-digital-content-transfer-2014.pdf>> (ult. cons.: 09.03.2017)
- Guercio 2016 = Mariella Guercio, *Archivi personali: la sfida del digitale. Una riflessione*, in *Gli archivi di persona nell'era digitale. Il caso dell'archivio di Massimo Vannucci*, a cura di Stefano Allegrezza e Luca Gorgolini, Bologna, Il Mulino, 2016, p. 133-148
- Guidelines for creators of personal archives* 2008, <<http://www.paradigm.ac.uk/workbook/appendices/guidelines.html>> (ult. cons.: 09.03.2017)
- Lee - Tibbo 2007 = Christopher A. Lee - Helen R. Tibbo, *Digital Curation and Trusted Repositories: Steps Toward Success*, «Journal of Digital Information» [Online], 8.2, 2007. Web: <<https://journals.tdl.org/jodi/index.php/jodi/article/view/229/183>> (ult. cons.: 09.03.2017)
- Lee - Tibbo - Schaefer 2007 = Christopher A. Lee - Helen R. Tibbo - John C. Schaefer, *Defining What Digital Curators Do and What They Need to Know: The DigCCurr Project*, proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries, Vancouver, British Columbia, Canada, June 18-23, 2007, edited by R. Larson [et al.], New York, NY, ACM Press, 2007

- Lenzini 2016 = Luca Lenzini, *Don't Save! Fortini, un Mac e le ultime cose*, «L'Ulisse. Rivista di poesia, arti e scritture», 19 (2016), p. 96-101. Web: <<http://www.lietocolle.com/cms/wp-content/uploads/2016/04/ULISSE-19.pdf>> (ult. cons.: 09.03.2017)
- Lombardi 2016 = Andrea Lombardi, *Scrittori e Facebook*, «Le parole e le cose», agosto-maggio 2016, <<http://www.leparoleelecose.it/?p=24083>> (ult. cons.: 09.03.2017)
- PAD, Pavia Archivi Digitali, <<http://pad.unipv.it>> (ult. cons.: 09.03.2017)
- Policastro 2012 = Gilda Policastro, *Polemiche letterarie. Dai Novissimi ai Lit-blog*, Roma, Carocci, 2012.
- Stollar Peters 2006 = Catherine Stollar Peters, *When Not All Papers are Paper: A Case Study in Digital Archivy*, «Journal of the Society of Georgia Archivists», 24 (2006), p. 22-34.
- Weston 2013 = Paul Gabriele Weston, *Digital curation: nuove opportunità per bibliotecari e archivisti*, in *Il libro, gli archivi e la memoria digitale*, a cura di Paul Gabriele Weston, Milano, Unicopli, 2013, p. 9-66.
- Weston - Carbè - Baldini 2016 = Paul Gabriele Weston, Emmanuela Carbè, Primo Baldini, *Hold it All Together: a Case Study in Quality Control for Born-Digital Archiving*, «Qualitative and Quantitative Methods in Libraries (QQML)», 5 (2016), p. 695-710. Web: <http://www.qqml.net/papers/September_2016_Issue/5313QQML_Journal_2016_Westonetal_695-710.pdf> (ult. cons.: 09.03.2017)
- Workbook on Digital Private Papers* 2008, <<http://www.paradigm.ac.uk/workbook/index.html>> (ult. cons.: 09.03.2017)

ABSTRACT

Nato nel 2009 per affiancare lo storico Fondo Manoscritti dell'Università di Pavia, PAD - Pavia Archivi Digitali si occupa di acquisire, conservare e rendere accessibili gli archivi nativi digitali di scrittori e giornalisti contemporanei. I fondi fino a oggi acquisiti si sono rivelati molto diversi tra loro, e hanno contribuito via via a precisare e perfezionare le procedure di *ingest* e la gestione del ciclo di vita dei materiali conservati. Una delle questioni a nostro avviso più importanti riguarda l'analisi preliminare della struttura degli archivi – sia dal punto di vista logico, che informatico – e il colloquio con lo scrittore, a cui viene somministrato un questionario, che si sta rivelando molto utile per comprendere il metodo di lavoro e l'organizzazione dei materiali. Non sono molte le indicazioni che la letteratura offre al riguardo, essendosi quest'ultima prevalentemente occupata di archivi tradizionali o di archivi digitali caratterizzati da una maggiore stabilità di quelli letterari. L'intervento mira a mettere a disposizione della comunità le riflessioni maturate nel corso delle acquisizioni.

Digital curation, Digital preservation, Born-digital archiving, Private Papers Archiving, Born-digital literary papers

PAD – Pavia Archivi Digitali was created in 2009, alongside the Centre for Research in the Manuscript Tradition of Modern and Contemporary Authors, with the aim of acquiring, preserving and giving access to born-digital papers of contemporary writers and journalists. The archives which were bestowed to PAD so far show a variety of features involving a considerable effort in defining and fine-tuning both the ingest procedures and the management of the preserved papers life-cycle. One of the crucial issues is the preliminary analysis of the logical and technical structure of these archives. For this purpose the writer is interviewed

and more information is collected through a specific questionnaire. By combining the two one gains a valuable insight into his or her working methods and files arrangement logic. The literature on the subject is still quite scarce, since researchers have been more keen to focus either on traditional archives or on digital archives though characterized by a greater stability. To share reflections emerged in response to the various acquisitions is therefore the aim of this paper.

Digital curation, Digital preservation, Born-digital archiving, Private Papers Archiving, Born-digital literary papers