



Università
Ca' Foscari
Venezia

Corso di Laurea
in Storia e Gestione del Patrimonio
Archivistico e Bibliografico
ordinamento
ex D.M. 270/2004

Tesi di Laurea

**Gli archivi degli scrittori
nativi digitali:**
Verso la conservazione e oltre

Relatore

Ch. Prof. Paul Gabriele Weston

Correlatore

Ch. Prof. Riccardo Ridi

Laureando

Michele Stefani
Matricola 855567

Anno Accademico

2015 / 2016

GLI ARCHIVI DEGLI SCRITTORI NATIVI DIGITALI:

Verso la conservazione e oltre

- **Introduzione**
- **Lavorare con gli scrittori**
- **Il flusso di lavoro tecnico**
 - OAIS: il modello di riferimento per un sistema informativo aperto per l'archiviazione
 - La sostenibilità
 - Fasi preparatorie e di preacquisizione
 - Fasi di acquisizione
 - Fasi post acquisizione
 - Sistema di controllo qualità
- **Le questioni legali**
 - La normativa archivistica e sul documento digitale
 - La normativa sul diritto d'autore
 - Il diritto d'autore di "terze parti"
 - Le licenze aperte
- **Gli aspetti concettuali**
 - OAIS conforme oppure no?
 - Quale approccio per gli archivi d'autore
 - Come, cosa e perché conservare?
 - Metadati tecnici, amministrativi, di conservazione
 - Metadati descrittivi?
 - ...e oltre?
- **PAD – Pavia Archivi Digitali**
 - La sostenibilità di PAD
 - La struttura di PAD e le funzioni del modello OAIS
 - QUANDO e PAD Manager
 - La didattica in PAD

*A Valentina,
alla mia famiglia*

*«There are more things in heaven and heart, Horatio,
Than are dreamt of in your philosophy»*

Shakespeare, Hamlet, Act I, Scene V

INTRODUZIONE

Agli albori del nuovo millennio fu inaugurato a Ferrara un fortunato ciclo di convegni a cadenza annuale, proposto e voluto dall'Associazione Italiana Biblioteche, il cui titolo era “Conservare il Novecento”¹: a partire dal secondo anno venne aggiunto un sottotitolo tematico, che di anno in anno indirizzava l'attenzione su aspetti più specifici di questo ampio tema. Ciò che a titolo di premessa deve essere rilevato è la partecipazione a questi convegni di numerosi professionisti del settore dei beni culturali, ma non solo: bibliotecari e archivisti, storici e storici della lingua e della letteratura, docenti di letteratura, professionisti dell'editoria, specialisti della conservazione audio e del restauro di documenti². Risulta immediatamente evidente come questo non solo sia un tema di interesse e di dibattito ancora vivo e fervente in moltissimi settori tra loro differenti, benché contigui, ma, ancor più, che per essere trattato in maniera efficace debbano essere coinvolte nel dibattito diverse figure e competenze professionali, in quanto non si tratta certo del lavoro – pratico e concettuale – di un solo e unico specialista ma, piuttosto, di una serie di specialisti in diverse discipline

¹ (Messina et al. 2001)

² M. Palma, *Conoscere e salvaguardare gli archivi di persona*, in (Allegrezza e Gorgolini 2016, p. 14). In questo intervento, molto più recente di quelli contenuti nel primo volume di “Conservare il Novecento”, si ricorda come sempre di più – per la tipologia dei materiali, ma non solo – trattare della conservazione e della descrizione di questi archivi voglia «*il dialogo con le altre professioni dei beni culturali, soprattutto con bibliotecari e operatori museali*», a cui non esiterei ad aggiungere sia storici e studiosi di letteratura, che sono i destinatari principali degli esiti di questi processi, che informatici.

che dialogano e collaborano, ciascuno con le proprie competenze, per conseguire un risultato comune che tenga conto delle necessità e dei criteri propri dei campi che attraversa, e del pubblico – nel suo complesso – a cui si rivolge.

Perciò “Conservare il Novecento”: ma in quale misura un’operazione di conservazione non dissimile da quella svolta nel corso del tempo da queste e altre istituzioni aspira alla singolarità, al riconoscimento, già nominale, di istanza particolare e non ordinaria, nei suoi obiettivi?

«Come è stata pensata quest’iniziativa? Si è pensato che siamo di fronte, con quelli che ora provvisoriamente chiamerò archivî culturali in genere (non limitandomi quindi a chiamarli archivî letterari), si è pensato che sostanzialmente siamo di fronte a un fenomeno nuovo, nato nella seconda metà del secolo ormai scorso, diciamo nel secondo dopoguerra. Perché fenomeno nuovo? In realtà cose del genere esistevano anche prima, e le conosciamo abbastanza bene; tuttavia queste di cui vogliamo parlare presentano qualche caratteristica crediamo diversa e fondamentalmente diversa. [...] si è cominciato a delineare ormai da tempo un fenomeno di raccolta e archiviazione di materiali non limitati a singole persone: cercando di rappresentare ciascuno (secondo le proprie forze, naturalmente) una società letteraria, una civiltà letteraria. Credo che siamo di fronte [...] a una miniera di materiale quale gli storici della letteratura e più in generale dell’arte e della cultura, in passato non hanno mai posseduto se non per un singolo ben individuato artista o scrittore»³

In gran parte, indica Crocetti, questa singolarità è dovuta a un lento ma progressivo aumento di rilevanza del contesto in cui si opera, cioè il fatto che l’autore, l’artista, non solo non sia chiuso nella sua torre d’avorio elitariamente intesa, ma che proprio perché non è affatto un eremita – tanto più che l’autore ha *necessità* di un pubblico; sempre di più in tempi più vicini con l’arte performativa, per esempio, o gli opinionisti in televisione, o i blog dei giornalisti – e perché il paradigma sociale è diventato più fluido⁴, con più attori (e sempre meno istituzionali o istituzionalizzati) a comporre il panorama

³ Luigi Crocetti, *Parole introduttive* in (Messina et al. 2001, p. 23).

⁴ C. Del Vivo, *Accostarsi a un archivio di persona: ordinamento e condizionamento*, in (Ghersetti, Paro, e Fondazione Giuseppe Mazzotti per la civiltà veneta 2012).

informativo e mediatico, il contesto diviene sempre più indicativo ed esplicativo, si lega a doppio filo con l'informazione e con l'opera. Questo mutamento, dall'esempio in cui per contrappunto Crocetti oppone le "case d'autore" delle tre corone dell'Otto-Novecento italiano – Carducci, Pascoli e D'Annunzio – a quella di Marino Moretti⁵, si fa più insistente e pressante mano a mano che ci si avvicina alla fine del secolo, e poi ancora di più nel nuovo millennio: senza saltare a piè pari direttamente all'era del digitale, di cui avremo modo di discorrere approfonditamente più avanti, basti ricordare a titolo d'esempio che autori come Pier Paolo Pasolini, Italo Calvino, Carlo Emilio Gadda⁶ furono impegnati su più fronti (il Pasolini cineasta e dell'impegno politico, il lavoro editoriale di Calvino e quello in RAI di Gadda, solo per citarne alcuni), e comunque immersi nel *contesto* intellettuale e culturale contemporaneo, con la cura e la pubblicazione di e su numerose riviste, in una rete di contatti e scambi umani e culturali da cui è difficile prescindere per tracciarne un profilo adeguato: come immaginare un archivio di uno qualsiasi di questi autori, con tutto ciò che esso contiene, con le relazioni professionali, umane e culturali che traccia, avulso dal contesto e dai rapporti che vi sono composti⁷?

Come ricorda Mariella Guercio:

«Il tema della sopravvivenza delle memorie personali digitali è rimasto a lungo assente nella riflessione della comunità archivistica anche a livello internazionale. Eppure si tratta di documentazione che presenta molte più difficoltà e problemi di

⁵(Messina et al. 2001, p. 23): «Per esempio, tutti noi conosciamo le case abitate da scrittori e a loro consacrate dopo la loro scomparsa. [...] Tuttavia se prendiamo un caso molto più recente, Casa Moretti a Cesenatico, vediamo che l'impostazione è già diversa, perché Casa Moretti è dedicata, naturalmente, allo scrittore Marino Moretti; tuttavia vuole estendere l'ambito delle sue attività anche al contesto in cui Marino Moretti è vissuto e ha scritto.»

⁶ Un bel racconto della personalità, ma particolarmente anche dell'ambiente in cui si muoveva Gadda si può trovare in (Arbasino 2008).

⁷ Alcuni spunti a tal proposito, sulla grande carica umana che possono presentare questi archivi, possono essere ritrovati in (Corti 1997), per esempio nella vicenda della poesia *Venezia, sera di Natale* (poi *Natale al caffè Florian*) dedicata a Graziana Pentich e collegata all'acquerello *A Graziana, Natale 46, Venezia*.

quella prodotta da qualunque organizzazione sia pubblica che privata.»⁸

Le motivazioni per cui questo interesse sia quantomeno tardivo (e non solo in ambito digitale) sono molteplici, e senza dubbio non è mio compito esaurirle sistematicamente in questa sede, tuttavia è necessario fare qualche cenno a questi aspetti nel loro complesso. Nello stesso intervento di Guercio, ma come anche in quello già precedentemente ricordato di Del Vivo, si evidenzia come il mutamento del contesto sociale negli ultimi tre secoli abbia progressivamente portato l'individuo – e perciò il suo archivio – a occupare un posto di sempre maggiore attenzione, a cui ha fatto seguito di riflesso l'attenzione per ciò che appartiene all'individuo. Se però inizialmente l'attenzione era rivolta a casi non solo singoli, ma anche singolari (e sempre Del Vivo ricorda come la possibilità di recuperare un archivio di persona non legato a quello della relativa famiglia o istituzione divenga sempre più rarefatta mano a mano che si torna indietro nel tempo), con le scienze sociali acquistano interesse sempre maggiore e diventano oggetto di studio anche quegli archivi di persona non necessariamente di una qualche personalità, ma anche di personaggi minori, minimi o sconosciuti, utili però per raccogliere informazioni su molteplici aspetti della vita e della società ad ogni livello.

Possiamo quindi notare recentemente, e sempre di più, l'interesse verso archivi di persona che non siano necessariamente anche personalità, e negli ultimissimi anni la sempre maggiore diffusione di narrative del sé⁹ che alimentano indirettamente l'interesse per la condivisione (e almeno un provvisorio mantenimento) della propria storia: basti pensare che un social network come Facebook propone l'organizzazione dei post su una linea temporale, e che crea autonomamente contenuti *ad hoc*

⁸ M. Guercio, *Archivi personali: la sfida del digitale. Una riflessione*, in (Allegrezza e Gorgolini 2016, pp. 133-134).

⁹ Basti pensare al marketing tanto di colossi come Tesla, che di situazioni in scala molto piccola come i progetti in *crowdfunding* sulle piattaforme dedicate: ciascuno racconta la propria piccola storia personale, in tempi più lunghi dei meno recenti brevissimi spot televisivi, e questo aiuta estremamente l'efficacia comunicativa: tanto che il MiBACT ha impostato sullo stesso filone la campagna di promozione dei musei "L'arte ti somiglia": <http://www.beniculturali.it/lartetisomiglia>) lanciata il 20 dicembre 2016.

(video, collage di foto, ...) per le ricorrenze, per poter pubblicare o ripubblicare – e quindi portare all’attenzione della propria cerchia di contatti – la propria storia rielaborata in chiave narrativa, attraverso questi strumenti.

Non desta stupore quindi che al giorno d’oggi il portale tematico sulla prima guerra mondiale di Europeana inviti, in una sezione dedicata e ben evidenziata¹⁰, a collaborare al progetto inviando i materiali relativi a quel periodo storico di cui si è in possesso, allegando solamente poche semplici informazioni sull’oggetto e come se ne è entrati in possesso permettendo, non solo all’utente occasionale di pubblicare la propria storia, il proprio racconto o il proprio ricordo, anche non troppo succintamente (se il contributo finale viene accettato dai revisori di Europeana) e contribuire con la propria piccola parte alla crescita del progetto, ma anche a eventuali ricercatori di avere accesso a reperti, documenti, foto e altri materiali – certo non sarà all’ordine del giorno il ritrovamento del cimelio rarissimo o risolutivo, ma – che comunque con ogni probabilità non si sarebbero manifestati altrimenti. Potrebbe sembrare poca cosa nel mondo estremamente fluido in cui viviamo, e a maggior ragione rispetto alla gran parte delle esperienze a cui siamo abituati nell’ambito tecnologico: la ricerca di Google e l’integrazione con altre decine di servizi, tutti accessibili immediatamente (sia nel tempo che apparentemente senza la mediazione di altri programmi o applicazioni), o le sempre maggiori possibilità offerte da ogni tipo di strumento informatico, sono realizzati appositamente per rendere il più possibile naturale la nostra esperienza con questi strumenti e rispondere a quelle che sono le nostre esigenze immediate: sono appositamente pensati per la nostra soddisfazione. Tuttavia nell’ambito della conservazione dei beni culturali il processo è sotteso all’esperienza della comunicazione del bene e del suo contesto, non del prodotto, e implica pertanto una riflessione sul metodo e sugli

¹⁰ <http://www.europeana1914-1918.eu/>; invece la sezione dedicata ai contributi è <http://www.europeana1914-1918.eu/it/contributor>. Europeana è una biblioteca digitale per il patrimonio culturale europeo a cui partecipano numerose istituzioni nazionali e locali in tutta Europa, contribuendo con materiali digitalizzati; specifici portali tematici di Europeana sono stati dedicati a diversi progetti come quello sulla prima guerra mondiale, sul cibo e le bevande (<http://foodanddrinkeurope.eu/>), sulla moda (<http://www.europeanafashion.eu/portal/home.html>) e altri.

obiettivi, che si discostano decisamente da quelli del mercato: pertanto, si tenterà di raccogliere e consolidare i dati consistenti tralasciando quelli meno significativi o dubbi, mirando alla permanenza nel tempo. In questo senso, come ricordato da Del Vivo¹¹ «*quanto state facendo su queste carte dovrà restare*»: chi si occupa di beni culturali sa bene che il prodotto dell'esercizio intellettuale nel proprio ambito mira a fornire gli strumenti più idonei e possibilmente più duraturi, senza nessuna pretesa di indovinare il futuro, ma allo stesso tempo senza inseguire in modo acritico e pedissequo i mutamenti rapidissimi dello sviluppo della tecnica.

In supporto a queste necessità, infatti, le operazioni di descrizione dei beni culturali appoggiano su standard internazionali ben definiti, il cui scopo è quello di garantirne la consistenza nel tempo e nello spazio.

Questo non pregiudica la possibilità di utilizzare strumenti e sistemi diffusi per comunicare al pubblico – anche al grande pubblico – gli esiti del lavoro sul patrimonio culturale, purché sia sempre presente questa riflessione, che colloca i professionisti dei beni culturali tra gli utilizzatori più critici ed esigenti e non del tutto allineati con gli utilizzi più diffusi dei prodotti proposti dal mercato (perlomeno quando si tratta di un uso non personale ma strumentale rispetto al bene).

In questo contesto perciò, in modo naturalmente provocatorio, poniamo una questione: certamente “conservare il Novecento”, ma allora anche “conservare il Duemila”, con tutti i corsivi e le virgolette del caso.

Il tema certamente è vasto, ricco di sfumature e ampiamente dibattuto: ci sono le voci a sostegno dell'impossibilità che in Internet o nel web possano andare persi contenuti e ci sono le voci opposte. Un esempio emblematico, che ha un che di involontariamente umoristico, si trova in uno dei primi capitoli del manuale online della *Digital Preservation Coalition*¹² (DPC): tra le risorse suggerite per l'approfondimento in calce alla pagina, un articolo tra quelli indicati ha un titolo piuttosto evocativo: “*Don't Panic: The Archivist's Guide to Digital Preservation*”

¹¹ C. Del Vivo, *Accostarsi a un archivio di persona: ordinamento e condizionamento*, in (Gheretti, Paro, e Fondazione Giuseppe Mazzotti per la civiltà veneta 2012).

¹² <http://www.dpconline.org/handbook/getting-started>

scritto nel 2011 da Stefanie Davidson. Il collegamento che dovrebbe condurre al file in PDF della guida conduce invece a una pagina d'errore 404 (file non trovato): il titolo della guida, benché di buon auspicio, non ha scongiurato il peggio. Questo è solo uno dei molteplici esempi possibili, che parlando di conservazione del digitale in questa sede, e anche in quella della DPC, può far nascere un momentaneo sorriso, ma al contempo è una realtà con cui è necessario fare i conti. Internet è una rete che serve a permettere la comunicazione, non la memorizzazione dei documenti, permette che vengano trasferiti globalmente, attraverso vari servizi, ma non ha una memoria propria: sono i supporti collegati in rete ad avere memoria, ciascuno la propria, ma bisogna rifuggire quell'idea per cui è impossibile che oggi venga perso¹³ o dimenticato un qualsiasi documento solo perché i nostri mezzi tecnici sono troppo avanzati, piuttosto bisogna rendersi conto che nonostante i nostri mezzi tecnici siano così avanzati non per questo tutto si risolve "in automatico", o sono procedure che "vanno da sé". L'approccio dell'Università di Stanford, LOCKSS¹⁴ (*Lots Of Copies Keep Stuff Safe*) funziona nell'ambiente organizzato da Stanford, ma è molto meno consigliabile come suggerimento di portata generale: se venisse adottato da ogni persona per ogni file in proprio possesso, diciamo che in breve tempo

¹³ In un'intervista, Gilles Deleuze commenta sulle risorse tecniche rispetto al mondo culturale «[...]quando le cose spariscono, nessuno se ne accorge. Per una semplice ragione: quando una cosa scompare, non manca a nessuno. [...] Supponiamo i nuovi Beckett di oggi. Supponiamo che non vengano pubblicati. Dopo tutto Beckett stava per non essere pubblicato: se non fossero pubblicati, è evidente che non mancherebbero a nessuno. Per definizione un grande autore o un genio è colui che introduce qualcosa di nuovo; se questo nuovo non appare, non disturba nessuno, non manca a nessuno, perché non se ne aveva alcuna idea. Se Proust non fosse mai esistito, se Kafka non fosse mai stato pubblicato, non sarebbe possibile sentirne la mancanza. se avessero bruciato tutto Kafka, nessuno avrebbe potuto dire: "ci manca", perché non avremmo idea di ciò che è scomparso. Se oggi i nuovi Beckett sono fatti fuori o non sono pubblicati dall'attuale sistema editoriale, non si potrà dire "come ci mancano". Ho sentito una cosa, credo la più impudente di tutta la mia vita, non dirò di chi, in un giornale, soprattutto perché queste cose non sono mai sicure, la dichiarazione di qualcuno nell'editoria che ha detto "dovete sapere che oggi non rischiamo più di fare errori come quello di Gallimard che ha rifiutato Proust, perché con i mezzi che abbiamo oggi [sarebbe impossibile non riconoscerli]" Dovremmo avere un contatore e di fronte al nuovo Beckett, uno assolutamente inimmaginabile, che non sappiamo cosa può portare di nuovo, il contatore emetterebbe un suono.» C di cultura, in (Deleuze, Parnet, e Boutang 2005).

¹⁴ <https://www.lockss.org/>

(dieci copie sembrano un numero sufficiente, nel piccolo) si decuplicherebbe la quantità di documenti disponibili, o meglio, il volume occupato da questi, perché il numero di “risorse uniche” sarebbe lo stesso. Non si può copiare all’infinito, né questo sistema di per sé, evita che i dati possano effettivamente andare persi, poiché ha senso pensarlo all’interno di un’architettura distribuita, ma per quanto è in potere della singola persona probabilmente è al di là della portata media. Di tutto quello che è presente sul web, vari sono i settori (per cui il concetto di archivio è sempre stato fondamentale) che già provvedono o stanno intervenendo per tenere i propri documenti al sicuro: le pubbliche amministrazioni e le organizzazioni governative, e tutta una serie di imprese private che lavorano principalmente con i dati come per esempio Google (le stime parlano di qualche milione di server), Microsoft, Amazon, Facebook. Tutti questi soggetti già provvedono a mantenere al sicuro i propri dati digitali, a farne copia, a conservarli: ma oltre alle organizzazioni governative e alle pubbliche amministrazioni, per le imprese private – con buona probabilità – non è detto che i dati verranno mantenuti nel tempo assicurandone una conservazione a lungo termine, in quanto non c’è garanzia che i dati vengano tenuti oltre lo stretto necessario (viste le dimensioni e le quantità dei dati in rapporto al costo di mantenimento); normalmente infatti, dopo un periodo più o meno lungo di assenza dagli account di una determinata piattaforma l’account viene cancellato. È importante che queste imprese private conservino i dati che noi immettiamo nei loro sistemi, ma non lo faranno oltre lo stretto necessario: i dati che piuttosto conserveranno saranno i metadati statistici sull’uso e sulle attività svolte, cioè i dati che alimentano gli algoritmi di ricerca e i motori di raccomandazione ormai per moltissime realtà virtuali. I dati che rimangono fuori da questi ecosistemi ideali, i famosi *walled garden*, ci sembrano meno flessibili, meno utilizzabili, non sono integrati con i servizi e le applicazioni che i nostri dispositivi ci mettono a disposizione; quelli che sono dentro, invece, sono molto più malleabili, possiamo muoverli e modificarli con estrema agilità. Ma dove finiscono in entrambi i casi questi dati? Chi ha il compito, e in base a quale criterio, di definire quali siano tra questi dati quelli destinati alla

conservazione? Per il momento nessuno, e in futuro – se ci sarà qualcosa a decidere che cosa andrà conservato dei miliardi di documenti presenti sulle varie piattaforme – probabilmente saranno dei sistemi di intelligenza artificiale; per quanto riguarda il tempo per il quale saranno invece conservati i nostri dati, probabilmente finché avremo accesso agli account o finché le condizioni dei contratti con i *provider* dei servizi ce lo permetteranno (i limiti di spazio che ci obbligano a eliminare documenti vecchi, che altrimenti conserveremmo), ma non possiamo saperlo con certezza. Sfortunatamente, se un tempo un archivio di persona “analogico” poteva accrescere e rimanere intatto a lungo – magari in una soffitta, e non in condizioni ottimali, ma pur sempre conservandosi – con i dati digitali, che viaggiano tra dispositivi differenti che cambiamo con una relativa frequenza, non sembra un’operazione così facile. Cosa ne sarà, quindi, non dei documenti di tutti, ma almeno del lavoro degli scrittori e dei giornalisti, un lavoro a lungo “analogico” e di ricerca, un lavoro che si è sempre mosso tra le carte, e oggi invece si muove tra i *bit*? Possiamo proporci di “conservare il Duemila” per le future generazioni di linguisti, filologi, storici della lingua, dell’arte, della letteratura, del costume e della società; dovremmo almeno porci la questione – se siamo qui anche personalmente – di come conservare il nostro lavoro, la nostra memoria digitale, fermo restando che il prodotto del nostro lavoro non ha perso valore ma solo consistenza fisica.

Fondamentali nell’affrontare le questioni, i problemi e le riflessioni di questo percorso sono stati il manuale operativo del progetto PARADIGM¹⁵, il manuale della *Digital Preservation Coalition*¹⁶ e la traduzione italiana di *Open Archival Information System (OAIS) Reference Model* di Brian Lavoie a cura dell’AIB¹⁷.

¹⁵ (Bodleian Library, John Rylands University Library of Manchester, e Joint Information Systems Committee 2007; «paradigm | workbook on digital private papers» 2017)

¹⁶ (Digital Preservation Coalition 2015)

¹⁷ (Lavoie 2016)

LAVORARE CON GLI SCRITTORI

I documenti che normalmente entrano in un archivio sono documenti che in un certo senso hanno esaurito la loro funzione: stante che nello standard per la descrizione archivistica ISAD(G) un fondo è definito come

*“L’insieme organico dei documenti archivistici, senza distinzione di tipologia o di supporto, formati e/o accumulati e usati da una determinata persona, famiglia o ente nello svolgimento della propria attività personale o istituzionale”*¹⁸

e che nel d.p.c.m. 3 dicembre 2013 l’archivio è definito come

*“Complesso organico di documenti, di fascicoli e di aggregazioni documentali di qualunque natura e formato, prodotti o comunque acquisiti da un soggetto produttore durante lo svolgimento dell’attività”*¹⁹

emerge abbastanza chiaramente la natura strumentale del documento, cioè il fatto che il documento venga inteso ancillarmente rispetto allo svolgimento di una attività; perciò sembra naturale – al di fuori degli archivi correnti che solitamente vengono conservati presso i soggetti produttori dai soggetti produttori stessi – che un documento che entri

¹⁸ (International Council of Archives 2000)

¹⁹ Allegato 1, *Glossario* in (DECRETO DEL PRESIDENTE DEL CONSIGLIO DEI MINISTRI 3 dicembre 2013: *Regole tecniche in materia di sistema di conservazione ai sensi degli articoli 20, commi 3 e 5-bis, 23-ter, comma 4, 43, commi 1 e 3, 44, 44-bis e 71, comma 1, del Codice dell’amministrazione digitale di cui al decreto legislativo n. 82 del 2005. (14A02098) (GU Serie Generale n.59 del 12-3-2014 - Suppl. Ordinario n. 20) 2013)*

in archivio per essere conservato, abbia concluso la parte più cospicua delle funzioni per cui è stato formato. Nel caso degli archivi di persona, tuttavia, il *Codice dei beni culturali e del paesaggio* specifica che

“Sono altresì beni culturali, quando sia intervenuta la dichiarazione prevista dall'articolo 13: [...]

*b) gli archivi e i singoli documenti, appartenenti a privati, che rivestono interesse storico particolarmente importante.”*²⁰

L'interesse storico a cui si fa riferimento è più probabile che venga individuato al termine o in conclusione della carriera dell'autore, in caso di tentativo di vendita all'estero, o a seguito della scomparsa, e il caso dell'archivio di Dario Fo e Franca Rame²¹ conservato presso l'Archivio di Stato di Verona è emblematico in questo senso. Infatti bisogna ricordare che Franca Rame (morta il 29 maggio 2013) è sempre stata artefice (benché non da sola), durante tutto il corso della sua vita, della conservazione dei materiali di famiglia e di quelli inerenti il lavoro con il marito; il 24 luglio 2015 appare un articolo del quotidiano svedese *Dagens Nyheter* che viene ripreso almeno da *LiberoQuotidiano.it* e *ilGiornale.it*²² il 4 agosto 2015 i quali titolano simultaneamente: *“Dario Fo lascia tutto il suo archivio alla Svezia”*; il 23 dicembre 2015 l'archivio Fo-Rame viene dichiarato di interesse storico, e inaugurato in tre mesi esatti il 23 marzo 2016 (vigilia del novantesimo compleanno di Fo) alla presenza del ministro Franceschini; il 13 ottobre 2016 muore Dario Fo. È interessante notare come in questo caso coesistano tutti e tre i fenomeni: la morte di uno dei due soggetti produttori (principali) dell'archivio, una dichiarazione pubblica della volontà di spostare l'archivio all'estero (non ritengo utile

²⁰ Decreto legislativo 22 gennaio 2004, n. 42, Codice dei beni culturali e del paesaggio, ai sensi dell'articolo 10 della legge 6 luglio 2002, n. 137. (GU n.45 del 24-2-2004 - Suppl. Ordinario n. 28), articolo 10 comma 3 (*DECRETO LEGISLATIVO 22 gennaio 2004, n. 42 Codice dei beni culturali e del paesaggio, ai sensi dell'articolo 10 della legge 6 luglio 2002, n. 137. (GU n.45 del 24-2-2004 - Suppl. Ordinario n. 28)* 2004); la versione aggiornata rispetto alle modifiche successive è reperibile su <http://www.altalex.com/documents/codici-altalex/2014/11/20/codice-dei-beni-culturali-e-del-paesaggio> , tuttavia le ulteriori modificazioni non hanno alterato questo punto del codice.

²¹ Inaugurato il 23 marzo 2016: ((ICAR) 2016) ed ero presente in quell'occasione.

²²(«Dario Fo vill skänka sitt teaterarv till Sverige» 2015); («Dario Fo lascia tutto il suo archivio alla Svezia: il sospetto sul Nobel» 2015); (Mascheroni 2015).

commentare la notizia: non conoscendo la vicenda nel suo complesso trovo superfluo sollevare una polemica; è interessante però notare che l'opinione pubblica potrebbe aver contribuito alla dichiarazione d'interesse storico) e una carriera importante alla sua conclusione (diciotto anni di distanza dalla vittoria del premio Nobel e 89 anni d'età per Dario Fo).

Generalmente, quindi, l'autore tende ad associare il conferimento di documenti per la conservazione alla propria morte, e anche il conservatore è più abituato a trattare (per gli archivi di persona) non con il soggetto produttore ma con altri soggetti ad esso legati.

Progetti di questo tipo, invece, mirano alla conservazione dei documenti dell'autore durante il corso della vita e quindi della formazione dell'opera. Proviamo a fare un esempio pratico: i dispositivi tecnologici devono essere garantiti per almeno 24 mesi, ma se un autore come Dario Fo fosse nato negli anni Novanta del Novecento e ipotizzassimo una durata della vita altrettanto lunga, iniziando ad usare il computer per produrre le proprie opere attorno ai vent'anni, ammesso che venisse utilizzato un solo computer alla volta e che venisse regolarmente cambiato una volta ogni due anni, ancora in buono stato, e conservato adeguatamente (da un punto di vista fisico), ci troveremmo alla fine della carriera di questo autore con circa trentacinque dispositivi diversi che hanno attraversato settant'anni di storia e di sviluppo tecnologico: anche nelle previsioni più ottimistiche è probabile che una grande quantità dei dati e dei documenti conservati in questi dispositivi non siano più leggibili perché corrotti o per l'obsolescenza dei formati o dei supporti. Molti (certamente tra gli autori, ma non esclusivamente), ancora non sono pienamente consapevoli che la sopravvivenza del proprio materiale digitale è a rischio: ecco perché è fondamentale cominciare a occuparsi, o almeno interessarsi, della conservazione abbastanza presto; se si continuerà – per gli autori contemporanei – ad attendere la dichiarazione di interesse storico, quando questa arriverà moltissimo materiale sarà ormai irrecuperabile, e non certo per fare del sarcasmo sui tempi ministeriali. Alla luce di questa situazione ci si trova necessariamente a lavorare in una maniera che per un archivista è forse un po' insolita: fianco a

fianco col creatore dei documenti, e senza che i documenti abbiano già una loro precisa destinazione (nel caso della P.A., ad esempio, è assolutamente normale la contiguità con il creatore vivente, ma i documenti hanno delle serie precise di destinazione) o un ordinamento intrinseco già stabilito.

Confrontarsi con un archivio d'autore comporta cercare di entrare in una struttura particolare e imprevedibile in quanto segnatamente soggettiva: le informazioni che si trovano negli archivi personali forniscono una prospettiva unica, la ricchezza dei documenti prodotti dagli attori del panorama culturale contemporaneo e l'eventuale corrispondenza con altri attori forniscono dettagli che non trovano spazio nelle pubblicazioni ufficiali. Le lettere, i diari, le raccolte di articoli, le fotografie e i progetti che non sono stati realizzati o pubblicati sono tra le fonti più preziose, ma gli archivi personali sono particolarmente vulnerabili perché risiedono fuori dalle strutture che potrebbero garantire il loro mantenimento (o perché non si è mai pensato a *come* poter conservare il materiale digitale); inoltre la maggior parte delle persone non ha un supporto tecnico adeguato: spesso l'unico supporto di cui ci si avvale è qualche consiglio informale da parte di amici o conoscenti più ferrati in materia "di computer". Non di rado i creatori di documenti digitali hanno poca dimestichezza con la tecnologia che permette il loro lavoro, e non si sentono sufficientemente motivati (anche perché probabilmente non hanno esigenze che lo richiedano) ad approfondire la complessità degli strumenti che utilizzano; inoltre potrebbero non essere a conoscenza del potenziale valore dei propri documenti e gestirli in modo disordinato o incoerente, eliminando documenti per loro irrilevanti quando hanno terminato la loro funzione: non è *sempre* in questo campo che si muove l'archivista? Lavorando con i creatori di archivi di interesse storico e culturale i curatori o gli archivisti possono aumentare il tasso di sopravvivenza di queste risorse. Dunque, mantenere il contatto con gli autori nel corso del tempo è della massima importanza: possono volerci anni perché gli autori maturino una sufficiente fiducia nell'archivio da prendere in considerazione di conferire i documenti più sensibili o confidenziali.

Le tipologie di documenti più comuni sono bozze e stesure dei lavori delle varie opere, eventualmente anche bozze o minute (digitali) di corrispondenza, note contenenti idee per progetti futuri, o di progetti passati che non sono ancora stati pubblicati, ma in attesa di esserlo; in questo caso per i documenti di cui l'autore sia effettivamente creatore, il metodo di lavoro è differente per ciascuno (non è per forza detto che sia metodico) e non si può sapere a priori se un autore lavora salvando una nuova versione del documento ogni volta che apporta delle modifiche, se sovrascrive progressivamente le modifiche sullo stesso documento dall'inizio alla fine, oppure una qualsiasi oscillazione tra queste due modalità. È possibile che vengano conferiti per la conservazione anche immagini, file audio e video, grafiche per le copertine dei libri, fotografie con altri personaggi (ma inerenti l'attività autoriale), registrazioni di conferenze, interventi, interviste e materiale affine; in questo caso – essendo il materiale destinato alla conservazione – si adottano le stesse accortezze come per tutti gli altri documenti. Un caso più particolare è quello della conservazione di siti web o blog personali in cui sarà necessario accordarsi sulle modalità del conferimento relativamente almeno alle tempistiche: se rilevare automaticamente o meno le modifiche, la cadenza e la modalità con cui fare copia dei materiali sul sito, come descriverne i contenuti (all'interno del progetto PARADIGM²³ la conservazione dei siti e dei blog dei politici è stata approfondita a un buon livello, viste le necessità e la tipologia di soggetti produttori con cui si rapportava: prevedeva, almeno a livello operativo, così come indica il *Workbook*, la descrizione delle pagine del sito in maniera piuttosto minuziosa, scorporando gli elementi fino all'individualità).

Tuttavia, insieme a queste tipologie di documenti e file in generale che potremmo definire provvisoriamente *semplici*, si trovano altre tipologie di documenti più *complessi* da gestire almeno dal punto vista legale: se secondo la legge di Felson

“*Rubare idee a una persona è plagio, rubarle a molti è ricerca*”²⁴

²³ <http://www.paradigm.ac.uk/>

²⁴ (Bloch 1990)

bisogna anche considerare che – al di là dello scherzo – documenti a vario titolo (testi copiati dal web integralmente e assemblati con altri o riarrangiati dall'autore) rappresentano alcune problematiche non indifferenti da un punto di vista legale – per i diritti di terzi – e non sempre, vista la fluidità dei contenuti²⁵, si può ricostruire la loro integrità originale per l'attribuzione o la provenienza.

Oltre a ciò, e mi limiterò a farne solo un cenno a margine, se nel ciclo di “Conservare il Novecento” è stato ripreso variamente il tema delle biblioteche d'autore come connesse e facenti parte a tutti gli effetti degli archivi d'autore²⁶, come ci si dovrà rapportare in futuro con le biblioteche d'autore, qualora queste divenissero totalmente, o anche solo parzialmente, composte da libri e documenti digitali? Sarà ancora possibile conservare la biblioteca dell'autore, recuperare le annotazioni dai dispositivi, ricreare l'ecosistema creativo o “quella particolare copia” su cui si è esercitato l'intelletto dell'autore? Se da un lato le tecnologie permettono queste attività – da un punto di vista puramente tecnico – le limitazioni inerenti il diritto d'autore invece non fanno ben sperare: il fatto che sia possibile autorizzare legalmente fino a un massimo di sei dispositivi e poi sia necessario acquistare nuovamente il libro digitale; il fatto che si ritenga generalmente l'acquisto di un libro digitale non come il vero e proprio acquisto di un bene ma come l'accesso a un servizio; il fatto che non sia possibile trasmettere a terzi in eredità o in altra forma i propri documenti digitali legittimamente acquistati: sono tutti elementi su cui si discute e si continuerà a discutere²⁷.

²⁵ Paola Italia, *Editing 2.0. Quali testi leggiamo e leggeremo in rete?* in («Che lingua fa?» 2016)

²⁶ Cfr. gli interventi della seconda sessione in (Messina et al. 2001); l'intervento di Zagra in (Desideri e Zagra 2010); ma anche quello di Roncaglia in (Messina et al. 2003), che poneva allora il quesito – ad oggi ancora aperto – sulla “lettura in *bit*”, e di riflesso anche sulla scrittura. Non ho consultato tutti i volumi della serie, perciò può essere che riferimenti alle biblioteche d'autore siano presenti anche nei volumi che non cito.

²⁷ Si confronti al riguardo la campagna di EBLIDA “The right to E-Read” (ora conclusa (EBLIDA 2014)), le dichiarazioni in merito dell'AIB («Lettura digitale in biblioteca: legalizziamola!» 2014) e alcune sentenze o risposte espressi dalla corte di giustizia dell'Unione Europea (Corte di Giustizia dell'Unione Europea 2012) o altre similari come l'interrogazione di Glenis Willmott del 7 novembre 2013, pubblicata poi nell'Official Journal of the European Union (2014/C 228/01 IV; p. 191); ma in generale, sono diverse le campagne di EBLIDA sui diritti («EBLIDA Position Papers

È molto importante, comunque, sviluppare un accordo tra l'autore e l'archivio che sia in una forma giuridicamente corretta, ma facilmente comprensibile da parte del donatore: dovranno essere chiaramente definiti i diritti e i doveri, sia da parte donatore che del deposito di archiviazione, il contenuto e la portata degli archivi e, nel caso di un archivio ibrido, le proporzioni relative di materiale cartaceo e materiale digitale; condizioni diverse possono essere applicate al cartaceo e al digitale (ad esempio in materia di accesso) e, dove sia il caso, deve essere chiaramente indicato nel contratto. Alcune delle questioni che dovrebbero rientrare in un accordo includono:

- Stabilire in modo inequivocabile la proprietà dell'archivio;
- Diritti di proprietà intellettuale: bisognerebbe poter stabilire chiaramente le condizioni relative al materiale di cui gli autori sono titolari di diritti, in relazione a:
 - Conservazione: il conservatore deve chiedere l'autorizzazione esplicita del titolare principale dei diritti per intraprendere azioni di conservazione;
 - Accesso: il detentore dei diritti può aver a disposizione una vasta gamma di opzioni: concedere all'archivio di effettuare determinate azioni (riguardo l'accesso e la copia per ricercatori e utenti) o richiedere di autorizzare personalmente le richieste; gli autori possono anche scegliere di trasferire alcuni diritti all'archivio, che sarebbe responsabile di tutte le decisioni relative al materiale;
 - Diritti di terzi: è probabile che i diritti di una parte consistente del materiale in archivio non siano ascrivibili all'autore: in questo caso si può solo concedere l'accesso al materiale e fornire copie all'interno delle disposizioni della legislazione vigente;
 - Protezione dei dati e riservatezza: l'autore potrebbe desiderare di limitare l'accesso ad alcuni documenti riservati

and Statements - European Bureau of Library Information and Documentation Associations (EBLIDA)» 2017), come il più recente intervento sulla sentenza della Corte di Giustizia dell'Unione Europea riguardo l'e-lending («EBLIDA briefing on the e-lending judgement of the CJEU - European Bureau of Library Information and Documentation Associations (EBLIDA)» 2017).

per un certo periodo di tempo; è importante che l'archivio possa garantirne la sicurezza anche in questi termini;

- Diritto di esclusiva: acquisire un archivio digitale normalmente comporta la creazione di copie dei documenti, e non il prelievo degli originali che di solito restano in mano agli autori (discorso che affronteremo più nel dettaglio in seguito): è fondamentale poter garantire che l'autore conceda all'istituzione il diritto di avere l'esclusiva sul proprio archivio (anche per motivi di investimento economico da parte del conservatore);
- L'accordo dovrebbe stabilire il diritto dell'archivio di valutare il materiale che andrà a comporre le collezioni, mentre l'autore dovrebbe avere il diritto di decidere che cosa accade ai *record* eliminati;
- L'accordo dovrebbe rendere palese che l'archivio catalogherà i dati e creerà i metadati necessari per la conservazione a lungo termine: i diritti sul catalogo e sui metadati vanno separati da quelli sui documenti;
- L'archivio dovrebbe impegnarsi a sorvegliare il materiale in modo responsabile e sicuro;
- Sarebbe possibile includere nell'accordo anche la frequenza di futuri conferimenti o interventi, concordandoli fin dall'inizio con l'autore;
- Se è previsto l'affidamento a terzi esterni per alcune parti del lavoro l'autorizzazione dovrebbe provenire dall'autore;
- In ogni caso è opportuno che nell'accordo venga garantita una sufficiente autonomia tecnica all'archivio per tutte le operazioni necessarie alla conservazione, ma allo stesso tempo che rimanga sempre in capo all'autore il diritto di disporre dei propri documenti.

Lavorando a contatto diretto con gli autori, può essere utile somministrare una sorta di questionario, nelle fasi preliminari, che aiuti il conservatore a capire quali strategie dovrà mettere in atto (rispetto ai sistemi *software* e *hardware* di cui dispone l'autore, per esempio, ma

non solo), e l'autore ad acquistare fiducia nel conservatore: sapere che chi si ha di fronte si preoccupa di avere il maggior numero di informazioni per operare al meglio in tutte le fasi del conferimento potrebbe essere incoraggiante e instillare fiducia negli autori. Il questionario d'indagine sarebbe particolarmente utile qualora fosse formulato in modo da fornire alcune indicazioni, come ad esempio:

- La quantità e la tipologia di documenti che l'autore intende conferire;
- Se ci siano dei file (nel caso di *backup* del computer su cui lavora l'autore) che l'autore non intende conferire volontariamente e che pertanto andranno eliminati;
- Se l'autore ha dato una struttura consapevole ordinata al proprio archivio, oppure se le scelte di ordinamento presenti sono da ritenersi più casuali, e quale dovrebbe essere la struttura effettiva;
- Se ci sia un gran numero di documenti per cui il *copyright* non è dell'autore, ma di terze parti;
- *Hardware* e *software* utilizzati dall'autore nella creazione dei documenti conferiti, credenziali per i documenti criptati.

Una volta che sia stato instaurato un dialogo con l'autore e che questo dialogo abbia portato a un esito positivo per la conservazione del suo archivio – cosa non scontata, e soprattutto non sempre di facile attuazione, considerando che ci si rapporta sempre con delle persone, e in quanto tali, ciascuna ha le proprie esigenze e necessità, non sempre in linea con quelle del conservatore – si può dare inizio alla fase di vera e propria acquisizione dei documenti dell'autore.

IL FLUSSO DI LAVORO TECNICO

Sarà importante dare alcune nozioni tecniche, per quanto riguarda le procedure più pratiche e legate all'informatica, che non sono da intendersi come prescrittive: infatti ad oggi non esistono degli standard veri e propri che indichino chiaramente quali strumenti utilizzare per queste tipologie di archivi²⁸. Quello che si può fare, pertanto, è “guardarsi intorno” e cercare se a livello nazionale o internazionale ci siano progetti simili, consultare attentamente la documentazione dei progetti (se ce ne sono, e se c'è della documentazione) e capire quali strategie siano state messe in atto in funzione di quali necessità *locali*.

A questo punto si potrà cominciare a ragionare, in base alle *proprie* necessità, al materiale che si deve conservare, agli obiettivi e al pubblico interessato agli esiti del lavoro, su quali siano le strategie più indicate da mettere in atto e quali gli strumenti migliori.

L'ordine di presentazione degli argomenti sarà indicativo e sviluppato secondo comodità d'esposizione: talvolta si seguirà un effettivo ordine sequenziale dei processi, talvolta invece verrà arbitrariamente stabilito. Ritengo però importante sottolineare come molte delle tematiche e degli aspetti presentati in questa sede si influenzino reciprocamente, rendendo perciò artificioso l'ordine, che più spesso sarebbe invece parallelo, per cui alcuni dei punti presentati potranno essere rinviati a un approfondimento futuro, e altri ancora approfonditi e sviluppati a

²⁸ Gli standard esistenti riguardano principalmente l'utilizzo del modello OAIS (o di altri modelli), che stabilisce quali debbano essere i processi, ma non le tecnologie da utilizzare; anche gli standard o le normative inerenti la metadattazione stabiliscono solo pochi e generici criteri (cfr. i capitoli seguenti).

più riprese: proverò quindi a seguire una sorta di ordine dei passaggi del processo a cui vengono sottoposti i documenti, dai computer o dai supporti degli autori fino alla conservazione permanente (e oltre).

OAIS: il modello di riferimento per un sistema informativo aperto per l'archiviazione

ISO 14721:2012²⁹ definisce un OAIS come:

"Un archivio, costituito da un'organizzazione, che può essere parte di una più ampia organizzazione, di persone e sistemi, che ha accettato la responsabilità di conservare informazioni e renderle disponibili per una comunità di riferimento."

L'*Open Archival Information System*, solitamente indicato come OAIS o “modello OAIS”, è un modello di riferimento che è stato ampiamente accettato dalla comunità di conservazione digitale come standard chiave per gli archivi digitali. Il modello OAIS specifica in che modo le risorse digitali debbano essere conservate per una comunità di utenti dal momento in cui il materiale digitale viene acquisito (fase di *ingest*, ovvero di acquisizione), attraverso successive strategie di conservazione fino alla creazione del *dissemination package* (cioè il pacchetto di informazioni utili per fruire dell'oggetto digitale conservato) per l'utente finale. Il modello di riferimento OAIS è un modello di riferimento di alto livello, nel senso che è abbastanza flessibile da poter essere utilizzato in un'ampia varietà di contesti (basti ricordare che nasce in seno al *Consultative Committee for Space Data Systems* per la conservazione a lungo termine di dati digitali generati nel corso delle missioni spaziali): il lato negativo di questa flessibilità è che lascia alle singole istituzioni l'onere di gestire gli aspetti tecnici. Anche se proposto e progettato da chi si occupa di gestire dati spaziali (ma non solo), il modello OAIS si propone di essere il più neutro possibile rispetto al contenuto: oltre a fornire un quadro e un

²⁹ Il modello OAIS è stato adottato come standard ISO nel 2003 (ISO 14721: 2003 OAIS), e poi aggiornato nel 2012 come ISO 14721:2012.

vocabolario comuni, il modello OAIS serve come strumento di pianificazione per la progettazione di nuovi *repository* digitali e come punto di riferimento per valutare i risultati di quelli già realizzati. Per quanto riguarda l'ambiente, OAIS ha rapporti con tre entità o figure: i produttori dei dati, gli utenti finali o “consumatori” e gli amministratori del sistema OAIS.

Oltre a definire le parti coinvolte nella conservazione a lungo termine dei materiali digitali, OAIS fornisce un modello di informazioni di gestione dei materiali digitali che passano attraverso il sistema. Una componente importante di questo modello è l'*information package* (IP). Ogni IP è composto dall'oggetto digitale da conservare, dai metadati richiesti a quel livello dal sistema e dalle informazioni sull'impacchettamento (*packaging information*) che riuniscono il contenuto informativo e le informazioni sulla conservazione in un unico pacchetto logico. OAIS delinea tre tipi di pacchetti informativi:

- il pacchetto di versamento SIP (*Submission Information Package*),
- il pacchetto di archiviazione AIP (*Archival Information Package*),
- il pacchetto di distribuzione DIP (*Dissemination Information Package*).

Al livello di SIP i metadati sono idealmente forniti dal produttore che in linea di massima corrisponde al creatore originale dell'oggetto; ma nel caso degli archivi di persona non è certo così comune che i metadati vengano forniti dal produttore³⁰. In questa fase, i metadati non saranno strutturalmente definiti né completi. Un SIP può anche essere fornito a un OAIS da un altro *repository* digitale: in questo caso, l'uso di metadati interoperabili ridurrà al minimo lo sforzo richiesto per acquisire il trasferimento. L'AIP è quel pacchetto informativo che OAIS si impegna a conservare a lungo termine. La costituzione dell'AIP inizia dal *Content Data Object*, l'informazione oggetto della conservazione, a cui si associano le Informazioni sulla rappresentazione

³⁰ Come ricorda Sisti: «Se prendiamo il caso di un archivio di persona, si osservi come, con ogni probabilità, gran parte dei documenti informatici in esso contenuti (se non tutti) non saranno stati oggetto di alcuna delle operazioni sopra elencate», cioè corredati dei metadati minimi e a cui siano state applicate delle misure di immutabilità. Matteo Sisti, *La descrizione archivistica e la conservazione degli archivi di persona analogici, informatici e ibridi tra standard e normativa*, in (Allegrezza e Gorgolini 2016).

(*Representation Information*): preservare a lungo termine del contenuto informativo richiede la presenza di metadati aggiuntivi, che supportino e documentino i processi di conservazione dell'OAIS. Questo insieme di metadati è chiamato Informazioni sulla conservazione (*Preservation Description Information - PDI*) e comprende cinque componenti:

- Informazioni sull'identificazione (*Reference Information*),
- Informazioni sul contesto (*Context Information*),
- Informazioni sulla provenienza (*Provenance Information*),
- Informazioni sull'integrità (*Fixity Information*),
- Informazioni sui diritti di accesso (*Access Rights Information*).

Le informazioni descrittive permettono la ricerca e il recupero del contenuto informativo da parte del consumatore OAIS.

Il DIP è il pacchetto informativo destinato all'utente o al consumatore (in termini OAIS) che richieda un oggetto o un gruppo di oggetti dal sistema informativo. L'OAIS fornisce l'oggetto (o gli oggetti) come un DIP che comprende l'oggetto stesso e i metadati rilevanti. In questa fase i metadati riferiti all'oggetto saranno più descrittivi che tecnici: è improbabile che l'utente finale sia interessato ai *checksum* o alle informazioni di rappresentazione, ma piuttosto sarà interessato ai metadati che riguardano la descrizione dell'oggetto.

OAIS fornisce un modello funzionale, che include le informazioni sulle tipologie di attività svolte da ciascuna funzione. Le funzioni delineate dal modello OAIS sono:

- Acquisizione
- Archiviazione
- Gestione dei dati
- Pianificazione della conservazione
- Accesso
- Amministrazione
- Servizi generali

Uno schema utile delle funzioni OAIS e la transizione da SIP ad AIP e DIP è dato qui di seguito:



Primariamente, gli obiettivi di un OAIS sono quello della conservazione a lungo termine e quello di consentire agli utenti (consumatori) di avere accesso all'informazione. Ma oltre a questi due obiettivi, che non pongono una grande distinzione tra il senso del termine "archivio" in OAIS e quello inteso in altri contesti, viene aggiunta una lista delle responsabilità obbligatorie a cui un archivio OAIS dovrebbe attenersi:

- negoziare e accettare dai produttori le informazioni appropriate;
- ottenere il controllo dell'informazione archiviata a un livello tale da assicurarne la conservazione a lungo termine;
- definire la Comunità di riferimento dell'archivio;
- assicurare che l'informazione conservata sia comprensibile autonomamente dalla comunità degli utenti, e cioè che l'informazione possa essere usata senza ricorrere al produttore dell'informazione stessa;
- seguire linee di comportamento e procedure documentate tali da assicurare che le informazioni siano conservate in sicurezza rispetto a qualunque rischio prevedibile e che non vi siano cancellazioni;
- rendere l'informazione conservata disponibile per la comunità degli utenti e consentire la distribuzione di copie autentiche dell'informazione conservata nella sua forma originale o in una forma riconducibile all'originale.

Sostenibilità

A livello preliminare è importante rilevare come uno degli aspetti inevitabilmente più importanti, se non il più importante, nel momento in cui si realizza un qualsiasi progetto, ma ancor più questo tipo di progetti, sia la sostenibilità. Immediatamente si è portati a pensare alla sostenibilità economica, che è importantissima, eppure non è senz'altro l'unica voce. Realizzare questo tipo di strutture significa progettare un sistema orientato a durare nel corso del tempo, di un lungo tempo, e se in ambito informatico cinque o dieci anni sono lunghissimi³¹ bisognerebbe ragionare in tempi molto più lunghi: posto che uno degli obiettivi primari è la conservazione nel tempo, oltre che lo studio, ma che la legge sul diritto d'autore vincola l'opera fino a settant'anni dopo la morte, realizzare un progetto di questo tipo basandolo su fondi erogati saltuariamente o in un'unica soluzione, per vederlo terminare nel giro di dieci o quindici anni (in scenari estremamente ottimistici) a causa della mancanza di risorse economiche, umane e di infrastrutture, significa investire denaro ed energie in qualcosa che non potrà portare i risultati per cui lo si era realizzato. Inoltre, se fosse sufficiente tenere in considerazione come obiettivo la sola conservazione, si potrebbe pensare a un sistema più semplice rispetto a quello che vedremo, in quanto, una volta che siano stati determinati con esattezza i documenti da destinare alla conservazione e che siano stati copiati con le dovute premure (di cui parleremo più avanti) su supporti destinati alla conservazione, questo potrebbe considerarsi come il risultato finale, l'obiettivo perseguito dal progetto. Tuttavia, nel nostro caso questo non è sufficiente, poiché vi sono necessità di studio legate alla specificità dei documenti da conservare, alla loro *tipicità* ben diversa dalla quasi totalità dei materiali conservati presso gli Archivi di Stato, e che si presenta peculiarmente anche rispetto agli archivi di persona (che

³¹ L'iPhone ha quasi dieci anni e l'iPad ne ha appena compiuti sette; nessuno dei social network oggi più diffusi ha raggiunto i quindici anni di attività; Windows XP è stato distribuito dal 25 ottobre 2001 e ha ricevuto supporto esteso fino all'8 aprile 2014: alcune di queste cose, se non tutte, ci sembrano vicine nel tempo, e lo sono; ma i cambiamenti tecnologici in questo arco di tempo sono stati moltissimi e imprevedibili.

abbracciano un ambito più generico, che vede al proprio interno una fitta compagine di personalità le quali certamente *hanno scritto*, ma non per questo sono necessariamente *scrittori*, e il cui interesse è in generale di natura più storica, ma non eminentemente letteraria) e agli archivi culturali (che rappresentano un ulteriore aspetto generico, dal punto di vista della forma dei contenuti, legato cioè alla tipologia documentaria conservata: anche un archivio fotografico, d'arte, di audiovisivi si considera legittimamente culturale). Quindi la necessità del progetto è di garantire la conservazione nel tempo e contemporaneamente la fruizione dei documenti conservati almeno per motivi di studio e ricerca: possibilità che sarebbe praticabile in modo piuttosto impervio se si considerasse come risultato finale del progetto l'archiviazione dei dati su un supporto destinato alla conservazione. In questo caso infatti si dovrebbe fare un certo numero di copie prima di archiviare stabilmente il supporto principale (*master*), oppure sarebbe necessario accedere al *master* ogni volta che vi sia necessità di copiare i dati per consentirne l'accesso (è da escludere naturalmente la cessione del *master*), potendo portare nel tempo al deterioramento del supporto stesso, e quindi alla perdita parziale o a volte anche totale dei dati.

Perciò è fondamentale stabilire – o perlomeno verificare sin dalla fase di progettazione – quale sia l'incidenza dei costi per il finanziamento degli strumenti, del personale, del materiale e, soprattutto, del relativo mantenimento. Non è raro, infatti, il caso in cui vengano stanziati i fondi per iniziare nel migliore dei modi un progetto, ritenendo espletato in questo modo l'interno fabbisogno di risorse che il progetto stesso richiederà nel tempo: ogni progettualità necessita di una parte più cospicua di finanziamenti iniziali per poter essere attivata ma, allo stesso tempo, necessita anche di una quantità di risorse non trascurabili – e perciò da evidenziare – per il suo mantenimento nel corso del tempo. Pertanto sarà fondamentale interrogarsi, rispetto agli obiettivi, su quali siano le strategie e gli strumenti più idonei a essere mantenuti nel tempo.

Per esempio, affidare ad aziende esterne la gestione di alcune parti del processo vorrà dire sia dover dialogare con esse per stabilire chiaramente quali sono i risultati da ottenere, che dover fare

affidamento sulla stabilità nel tempo dell'azienda; se ci fosse del personale interno competente e preparato a seguire tutte le mansioni, bisognerà assicurarsi che possa trasferire le proprie conoscenze e competenze ad altri, e che quindi il progetto non si debba interrompere nel momento in cui venisse meno (non necessariamente per motivi tragici); nella scelta dei *software* o dell'*hardware* è opportuno considerare quali siano gli strumenti migliori anche in base al supporto fornito dalla casa madre: tutte queste sono scelte di sostenibilità che vanno necessariamente affrontate in ogni fase del progetto, avendo prima adeguata contezza delle possibilità che esistono, e valutando poi in maniera adeguata le strategie migliori, singolarmente e in rapporto tra di esse.

Fasi preparatorie e di preacquisizione

Dopo che sono stati contattati gli autori, e che abbiano manifestato l'intenzione di effettuare un conferimento, si procede al prelievo dei documenti presso l'autore; generalmente le situazioni che si presentano possono essere di quattro tipi:

- L'autore conferisce direttamente tutto il suo archivio. In questo caso, viene fatta una copia di tutti i file e le *directory* così come sono presenti sul computer dell'autore, e si potrà mantenere la struttura di origine dell'archivio stesso.
- L'autore conferisce quasi tutto l'archivio escludendo o eliminando alcuni file (potrebbe trattarsi di documenti sensibili). In questo caso viene fatta copia dei file e delle *directory*, cercando di capire con l'autore quali file sono esclusi dal conferimento e se è possibile tenerne traccia in maniera generica, per completare comunque la struttura dell'archivio.
- L'autore conferisce una minima parte del proprio archivio, anche solo un insieme di documenti rapidamente riorganizzato in una forma differente da quella originale appositamente per il conferimento. In questo caso è opportuno capire, se l'autore è disposto a collaborare, il perché di questa scelta di

riorganizzazione del materiale, e se è possibile prendere nota della struttura originaria dell'archivio personale, per poterla documentare.

- L'autore conferisce un insieme di file sciolti, che non costituiscono strutturalmente una parte dell'archivio. In questo caso non è necessario prendere nota della struttura poiché i documenti sono tra loro autonomi: rispetto alla situazione precedente, in questo caso il conferimento non riguarda tanto una minima parte dell'archivio, ma solo un insieme di file che nell'archivio dell'autore non vanno a costituire un gruppo coerente, e che verranno considerati come miscellanea.

Le modalità per poter interagire con gli autori relativamente al conferimento da parte del soggetto conservatore sono anch'esse molteplici, e dipendono dal tipo di struttura che si vuole realizzare, dalle risorse e dagli strumenti di cui si dispone, e dagli obiettivi che si vogliono conseguire. I principali approcci per l'acquisizione dei documenti sono i seguenti:

- Un primo approccio è quello di ottenere *snapshot* periodici: viene realizzata, a cadenza regolare concordata con l'autore, una sorta di fotografia istantanea dei documenti dell'autore in quel momento e a quel punto di lavorazione, a cui verranno poi affiancate le successive istantanee a formare lo storico dell'evoluzione dei vari documenti. Se questo tipo di approccio dà la possibilità di creare, a certe condizioni³², una fiducia crescente tra il creatore e il conservatore, e di ottenere nel suo complesso un quadro indicativamente molto interessante sulle scelte operate e sulle modalità di lavoro (cosa è stato ampliato e come, cosa invece è stato scartato o abbandonato), allo stesso tempo ha anche alcuni inconvenienti: un rapporto lungo e durevole tra autore e conservatore porterebbe alla formazione di una mole estremamente consistente di dati, contenente più

³² È pur vero che per creare questa fiducia e dissipare la più che naturale prudenza iniziale, ove non sia vera e propria diffidenza, ci vuole tempo e costanza: se l'autore per ogni sessione di lavoro con il conservatore si rapporta con una persona differente, sarà senz'altro meno propenso a fidarsi subito, e ci vorrà più tempo.

precisamente che l'archivio dell'autore, la vera e propria storia della formazione dell'archivio dell'autore; non sempre comunque è detto, per varie motivazioni che sia possibile mantenere un rapporto a così stretto contatto e così regolare con l'autore, con gli inconvenienti che ne derivano per questa strategia.

- Un secondo approccio alla gestione è quello del supporto nel tempo all'autore: l'autore conserverà i propri documenti, sarà "archivista di sé stesso", fino a un momento concordato con il conservatore; in questo lasso di tempo il conservatore si occuperà di fornire tutto il supporto necessario perché l'autore conservi nel modo migliore possibile i propri documenti fino al momento in cui non vengano trasferiti presso l'archivio. Questo approccio consente di limitare la mole di dati conferiti al conservatore, permettendo nel corso del tempo una miglior valutazione di quali siano i documenti da conservare e quali invece possano essere scartati (operazione ben più delicata con l'approccio per *snapshot*), tuttavia non garantisce – a meno di un preciso contratto – che a un dato momento l'archivio verrà conferito al conservatore, il quale investirebbe nel supporto all'autore senza poter avere risultati dal proprio lavoro; infine, per quanta cura possa essere messa dal conservatore nel supporto all'autore, questo non implica che l'autore segua fattivamente le linee guida stabilite né che riesca a metterle in pratica.
- Esiste quindi la possibilità di un approccio che metta insieme i due precedenti cercando di ottenere i vantaggi di entrambi e di contenerne gli svantaggi. Si cercherà pertanto di offrire un supporto durevole all'autore, avendo cura di effettuare periodicamente degli *snapshot*: in questo modo l'autore dovrebbe aver cura di organizzare i propri materiali che periodicamente verrebbero rilevati dal conservatore, limitando – ipoteticamente – le ridondanze e permettendo di ricevere anche documenti magari più sensibili a distanza di tempo. Questo rimane però un caso ottimistico, poiché questo

approccio comporta anche gli svantaggi di entrambi: ove non ci sia una certa attenzione da parte dell'autore ci si ritroverebbe con la stessa ingente mole di materiale pur avendo investito nel supporto all'autore, e senza precedenti garanzie.

Nonostante tutto però, quest'approccio, applicato con un po' di flessibilità, è probabilmente quello che meglio si adatta alle necessità reali: offrendo un supporto costante all'autore lo si sensibilizza ai problemi della conservazione; si lascia all'autore la gestione dei propri documenti, permettendo il conferimento di quelli che egli stesso ritiene possano essere conferiti, e rimandando invece al futuro l'acquisizione di quelli più sensibili. Se è pur vero che in un contesto reale – in cui ogni caso è specifico – gli svantaggi rimangono, al contempo però è anche vero che con questo approccio si possono ottimizzare alcuni risultati, ed è quello che probabilmente risulta più efficace.

- Nulla vieta – anzi, ed è senz'altro un approccio che continuerà ad essere impiegato – di ricorrere all'approccio tradizionale, ovvero quando l'autore (o chi per esso) si rivolge al conservatore alla fine della propria carriera (o della propria vita), o viceversa il conservatore si rivolge all'autore. In questo caso – vivente l'autore – ci sarà più tempo e disponibilità da parte dell'autore, nonché una maggiore sensibilità naturale a conservare la traccia del proprio lavoro; tuttavia le problematiche non sarebbero poche: se si tratta dell'archivio “completo” di tutta la carriera è impossibile sapere quanti e quali documenti siano andati perduti nel passaggio da un computer all'altro; per quanto riguarda quelli trasferiti non è detto che nel passaggio abbiano mantenuto tutti gli attributi o addirittura che siano ancora leggibili; infine, non è da sottovalutare che l'autore possa aver eliminato velleitariamente alcuni documenti rilevanti per motivi arbitrari e che perciò l'archivio possa fornirne un'immagine soltanto parziale. Quest'ultimo punto non è certo una problematica esclusiva di questo approccio: può verificarsi sostanzialmente sempre, in

qualsiasi modo si proceda – a meno di decesso improvviso del creatore e conseguente volontà degli eredi di conferire l’archivio – in quanto i documenti conferiti dipendono sempre dalla volontà dell’autore di conferirli o, comunque, quelli presenti dipendono dalla loro non cancellazione. A proposito del caso in cui l’autore morisse e la volontà di donare l’archivio dipendesse dagli eredi, bisogna notare che sarebbe importante disporre delle password, se sono presenti, e che non vi siano documenti criptati: non è perciò garantito, in questo caso, che si possa adeguatamente recuperare tutto il materiale presente.

- Un ulteriore possibile approccio è quello per cui l’autore decide di conferire al conservatore i supporti materiali mano a mano che li “esaurisce”: vecchi computer, vecchi dischi fissi, supporti che sono stati sostituiti. Questo permetterebbe al conservatore di disporre degli strumenti – a livello *hardware* e teoricamente anche a livello *software*, in caso di conferimento del computer – per poter recuperare i documenti: spetta a questo punto al conservatore decidere se sia opportuno organizzare il conferimento in termini di copia dei documenti o di emulazione integrale dell’ambiente di lavoro (e comunque dopo un’opportuna valutazione di cosa sia da destinare alla conservazione). In ogni caso, con questo approccio bisogna tenere in considerazione la possibilità che si manifestino numerose problematiche riguardo l’integrità dei dati (è piuttosto probabile che la macchina sostituita non funzionasse adeguatamente), e riguardo la possibilità da parte del conservatore di utilizzare lo stesso *software*: in caso di *software* proprietario infatti sarebbe necessario trasferire la licenza d’uso al conservatore, o trovare soluzioni che riescano a risolvere questo problema rimanendo naturalmente all’interno della legalità.
- Infine, un ulteriore possibile approccio è quello dell’auto-archiviazione: in questo caso il conservatore offre una serie di servizi più o meno ricchi (dall’*hosting* di siti, posta elettronica, blog, *weboffice*, *storage*, *backup*, etc... fino alla semplice

possibilità di caricare i file che si desidera destinare alla conservazione in uno spazio destinato dal conservatore), di cui l'autore usufruirebbe. Per quanto interessante e utile possa essere una soluzione di questo tipo, che permetterebbe di registrare sul sistema stesso lo storico dei mutamenti e delle evoluzioni dei vari documenti e, con la creazione di specifici *tool* mirati, agevolerebbe notevolmente la descrizione del contesto e la creazione di metadati (nonché la possibilità di avere delle applicazioni per il *versioning*, e altre ancora, ma a questo punto sarebbe solo una questione di limiti tecnici sulle possibilità), non bisogna dimenticare principalmente un punto: tutti questi non sono compiti del conservatore, ma piuttosto del *data manager* (e non sembra troppo prudente notare come un autore potrebbe scambiare questi servizi – pensati per la conservazione – come invece un servizio di *data management* o una suite d'applicazione per il lavoro di scrittura). Inoltre, i costi per realizzare una piattaforma di questo tipo sarebbe proibitivi per buona parte dei conservatori, senza contare la necessità di un'adeguata sorveglianza legale sulle politiche della piattaforma e sulle licenze dei *software*, nonché sulla possibilità per il conservatore di tenere traccia di tutti questi dati.

Una volta scelto l'approccio più idoneo rispetto al proprio progetto, si effettua la copia dei dati. È importante specificare un punto piuttosto importante e non così banale, né da un punto di vista concettuale (che vedremo più avanti), né da un punto di vista tecnico: mentre con gli archivi cartacei si dispone del documento originale, nel caso degli archivi digitali non si dispone che di copie (a meno che non si riceva dall'autore il computer smesso e se ne possano recuperare i dati). Questo comporta, per la natura stessa del documento informatico, che sia necessario trovare il sistema per garantire la *fissità* del documento e la sua affidabilità.

La “fissità” di un documento, o di un qualsiasi file è importante perché permette di essere certi che il documento ricevuto sia esattamente quello che ci si aspettava, permette di controllare che i dati non siano corrotti o degradati rispetto al documento originale e, infine, può

provare che il documento non sia stato alterato o manomesso in alcun modo³³.

I principali strumenti³⁴ per verificare la fissità nel caso del trasferimento di archivi digitali possono essere (in ordine di complessità, affidabilità e sicurezza):

- Il numero di file e cartelle trasferiti: se il numero non coincide, evidentemente ci sono stati dei problemi tecnici nella copia;
- Le dimensioni dei file o del trasferimento nel suo complesso: se nel trasferimento la dimensione in bytes è inferiore a quella dell'archivio copiato, alcuni file non sono stati copiati correttamente (ed è probabile che abbiano dimensione 0 byte);
- Il CRC (*Cyclic Redundancy Check*) è un codice di rilevamento d'errore molto diffuso per la verifica di errori casuali nella trasmissione di dati, il cui esito è una somma di controllo (*checksum*): se i valori non corrispondono ci sono degli errori;
- MD5 (*Message Digest 5*) è una funzione crittografica di *hash* che richiede una modesta capacità di calcolo (che aumenta in base alle dimensioni del file) e genera un *checksum* di 128 *bit*;
- SHA (*Secure Hash Algorithm*) è una famiglia di funzioni crittografiche di *hash*, molto diffuse e utilizzate; in particolare tra le più utilizzate ci sono:
 - SHA1 che produce un *checksum* di 160 *bit*;
 - SHA256 che produce un *checksum* di 256 *bit*.

Per quanto riguarda l'utilizzo del *checksum*, le funzioni (quali che siano) producono sempre lo stesso output a partire dallo stesso input, ma per un cambiamento anche minimo (una lettera maiuscola anziché minuscola, uno spazio in più, un accento) l'output sarà molto differente (è più probabile che capitino due *checksum* identici per tipologie di file differenti come un documento di testo e un file audio). A seconda di come verranno acquisiti i file, è necessario generare i *checksum* il prima possibile: se si tratta di materiale *hardware* obsoleto,

³³ (NDSA et al. 2014, p. 5)

³⁴ <http://handbook.dpconline.org/technical-solutions-and-tools/fixity-and-checksum> e (Phillips et al. 2013, p. 3) (tabella tradotta in appendice: p. 165 della tesi)

floppy disk, CD-ROM, DVD, dischi fissi esterni, *pendrive*, o altri supporti che vengono fisicamente conferiti, l'operazione andrà fatta nel momento in cui vengono copiati i documenti nella struttura informatica del conservatore; diversamente, se si procede con la copia direttamente dal computer dell'autore (possibilmente su supporti criptati) sarebbe più opportuno e consigliabile creare i *checksum* prima della copia, in modo da avere la possibilità di un confronto con gli originali (ma è comunque possibile scegliere diversamente). Si raccomanda inoltre, in genere, di utilizzare dispositivi di *write blocking* quando si lavora con i supporti originali: questi dispositivi (che possono essere sia *hardware* che *software*) permettono di evitare che il contenuto dei supporti possa venire accidentalmente o intenzionalmente manomesso, danneggiato o cancellato bloccando la possibilità di scrittura sul supporto originale.

Una volta che tutti i file siano stati copiati e sia stato generato un *checksum* – è ovviamente possibile, e forse più pratico, generare un unico *checksum* per l'intero conferimento invece che per ogni singolo file – il conferimento può essere trasferito all'interno della struttura e del sistema destinato alla conservazione per iniziare le procedure di acquisizione.

Fasi di acquisizione

Il modello OAIS individua come funzioni proprie del processo di acquisizione:

- la ricezione delle informazioni trasferite all'OAIS da un produttore;
- la validazione che le informazioni ricevute siano non corrotte e complete;
- la trasformazione delle informazioni inviate in una forma adatta alla conservazione e alla gestione all'interno del sistema di archiviazione;
- l'estrazione e/o la creazione dei metadati descrittivi per gli strumenti di ricerca e di recupero;

- il trasferimento all'archivio delle informazioni inviate e dei metadati ad esse associati.

Tendenzialmente, come già precedentemente ricordato³⁵, per quanto riguarda gli archivi di persona, e a maggior ragione gli archivi di scrittori, la possibilità di avere un SIP o anche solo delle informazioni sufficientemente complete e affidabili è piuttosto improbabile: supporre che persone esterne alla filiera della conservazione digitale producano e inoltrino al conservatore pacchetti di informazioni conformi al modello OAIS è – per il momento – assolutamente poco realistico.

Quindi, i file copiati dall'autore vengono riversati all'interno del sistema e se ne verifica l'integrità attraverso i *checksum* (nel caso in cui siano stati preventivamente generati), se invece provengono da altri supporti di memoria si provvede al recupero dei dati (ove possibile: non è detto che i supporti – CD, DVD, *pendrive*, dischi fissi, *floppy disk* – siano in uno stato di conservazione sufficientemente buono da permetterne la lettura e l'estrapolazione dei dati, o che non sorgano altri problemi) e alla creazione delle relative informazioni di fissità.

Molto spesso, per quanto alcune competenze informatiche di base possano essere diffuse – eppure molto elementari –, la reale comprensione dei processi informatici rimane oscura ai più, e gli autori non fanno certo eccezione in questo caso: è inverosimile sperare che siano stati presi degli accorgimenti nel corso degli anni volti alla conservazione dei propri documenti digitali³⁶. Così, il fatto che il computer, oltre ad essere uno strumento di lavoro, sia comunque un dispositivo *general purpose*, comporta – posto che è impossibile restringere a un numero ristretto le possibilità d'uso del proprio computer messe in atto da ciascuno – che possano esserci, tra i file

³⁵ Vd. alla nota 30 l'intervento di Sisti.

³⁶ Bisogna ricordare infatti che queste tematiche sono state presentate nel dibattito internazionale solo negli ultimi anni, e che si stanno imponendo all'attenzione collettiva e dei non addetti ai lavori ancora più recentemente; perciò non è irragionevole pensare che autori che lavorano con il computer da venti o trent'anni non abbiano preso determinate cautele fin dall'inizio. In questo senso è senz'altro lodevole l'intento di Digital Preservation (<http://www.digitalpreservation.gov/>), un'iniziativa della Library of Congress rivolta all'approfondimento delle tematiche della conservazione dei documenti digitali, che mette a disposizione – con una sezione dedicata interamente al digital personal archiving – una serie di indicazioni e materiali informativi destinati all'utente medio, che permettano di familiarizzare tanto con i concetti che con le buone pratiche di base.

destinati alla conservazione, anche dei *malware*; e questo specialmente quando venga trasferita una mole ingente di dati e non un ristretto numero di file, o quando venga fatto direttamente il *backup* dell'intero computer dell'autore; per questo stesso motivo inoltre, e per ragioni di riservatezza, è opportuno eliminare o non acquisire affatto tutti i file nascosti, in quanto possono essere stati generati in maniera automatica dal sistema (e come file di sistema non interessano il conferimento) o da altri programmi (e in questo caso è preferibile che non interessino il conferimento). Perciò è necessario che nel momento in cui i file vengono acquisiti per la prima volta all'interno della struttura (dopo aver controllato che l'acquisizione sia andata a buon fine con i *checksum*) siano accuratamente sottoposti a un controllo con un buon *software* antivirus e, possibilmente, posti in un'area di quarantena per evitare, nel caso in cui ci fossero problemi, di estenderli oltre il conferimento in acquisizione. Una volta che sia stata verificata l'assenza di programmi maligni all'interno del conferimento si può procedere con la stesura di una lista dei file conferiti: più che un vero e proprio inventario, questa lista serve per verificare con l'autore che siano stati conferiti solo ed esclusivamente quei documenti per cui è stata espressa un'esplicita volontà di conferimento; nel caso in cui ci si trovi ad acquisire grandi quantità di dati, o l'autore conferisca dischi fissi (vecchi o il *backup* di quelli ancora in uso), a causa dell'automazione delle procedure (come il trasferimento di tutta una serie di cartelle e di tutti i file e le cartelle in esse contenute o il *backup* di memorie di massa) non è affatto improbabile trovare file semplicemente non previsti o che sono finiti tra i documenti per chissà quale motivo: si può trattare di foto, di video, di file audio, di documenti di testo non indirizzati alla conservazione e di altre tipologie di file. Quindi, si stila una lista che contenga tutti i materiali dati al conservatore per l'acquisizione, affinché l'autore possa controllare che sia presente tutto ciò che era sua intenzione conferire e che non vi siano – e vengano pertanto eliminati nel caso in cui fossero presenti – file che non era nelle intenzioni dell'autore conferire.

Dal momento in cui l'autore conferma la lista, e firma quindi l'accordo perfezionato con il conservatore sul conferimento, si può procedere

con l'acquisizione definitiva dei file: dall'area di quarantena, o comunque dall'area di stoccaggio in cui i file erano stati posti temporaneamente, si può provvedere a trasferire i file nell'area destinata alla conservazione permanente. L'area di conservazione permanente, o comunque si voglia chiamare lo spazio destinato alla conservazione dei documenti, deve essere uno spazio sicuro: alcuni sono i suggerimenti³⁷ per ottimizzare l'efficacia della conservazione e non correre il rischio di perdere i dati nel corso del tempo. È opportuno pertanto fare almeno tre copie dei documenti acquisiti da destinare alla conservazione, con almeno due metodi di stoccaggio differenti, distribuiti in almeno due luoghi differenti come strategia di *disaster recovery*: un esempio potrebbe essere quello di fare due copie virtuali, conservate su due server a distanza e una terza copia su DVD per la conservazione conservato in cassette di sicurezza; oppure, una copia virtuale su server, una su DVD per la conservazione in cassetta di sicurezza e (se fosse il caso di conferimento *hardware*, ma non esclusivamente) una copia su disco fisso (criptato o meno, a seconda delle modalità di deposito).

Quest'ultima fase è quella che nel modello OAIS viene chiamata "Archiviazione": benché l'obiettivo della conservazione possa considerarsi raggiunto, non si tratta né del momento finale né di una stasi, bensì di un processo. È infatti fondamentale tenere costantemente sotto controllo i dati affinché non si corrompano nel corso del tempo (controllando i *checksum*) e sottoporli a procedure volte a prevenire o contrastare l'obsolescenza dei formati e dei supporti (*format migration*: migrazione dei formati da proprietari o obsoleti e non più supportati ad aperti, standard e supportati; e *media refreshing*: copia dei dati su supporti più idonei alla conservazione per prevenirne il deterioramento).

In questo senso si ritengono esaurite in linea di massima le operazioni dedicate alla conservazione propriamente intesa dei documenti digitali. Ulteriore e più particolareggiata attenzione andrebbe prestata nei casi in cui venga richiesto al conservatore di occuparsi dell'archiviazione

³⁷ Posto che, come già preventivamente spiegato, esistono delle linee operative, ma non degli effettivi standard pratici, si fa riferimento al già citato (Phillips et al. 2013).

permanente anche della “vita digitale” dell’autore: ovvero il conferimento, per esempio, di ciò che pubblica sui social network, su blog e siti (personali o di terzi), di corrispondenze via email, e di documenti salvati in piattaforme e sistemi di *cloud computing*; è un punto, quello dell’affidabilità e della permanenza delle risorse “affidate alla rete” che sta riscuotendo un grande interesse nel dibattito internazionale per le questioni complesse e non più derogabili che comporta; InterPARES Trust (ITrust)³⁸, è il quarto ciclo progettuale del progetto internazionale interPARES di durata quinquennale e si occupa proprio di alcuni di questi aspetti: sarebbe piuttosto pretenzioso da parte mia tentare di esaurire in così breve spazio e con nessuna esperienza un tema che sta meritando un’attenzione così ampia a livello internazionale e su cui stanno intervenendo alcuni tra i più importanti esperti dei rispettivi settori, perciò rimanderò semplicemente alla documentazione disponibile sul sito del progetto.

Fasi post acquisizione

Come si accennava più sopra, con l’archiviazione non si conclude il percorso per la conservazione: rispetto al modello funzionale di OAIS mancano ancora buona parte delle funzioni indicate:

- La “Gestione dei dati” mantiene i dati collegati al conferimento (metadati descrittivi per la descrizione e l’identificazione delle informazioni archiviate come supporto agli strumenti di ricerca OAIS) e gestisce dati amministrativi per supportare le operazioni di sistema (dati di performance, statistiche di accesso, ...). Tra le funzioni primarie della Gestione dei dati vi sono:
 - mantenimento dei database;
 - esecuzione di *query* e generazione di *report*;
 - aggiornamento dei database in caso di aggiunta, modifica o cancellazione di informazioni.
- La “Pianificazione della conservazione” indica la capacità dell’OAIS di adattarsi, e quindi aggiornare politiche e

³⁸https://interparestrust.org/trust/research_dissemination

procedure, rispetto ai mutamenti tecnologici e alle esigenze degli utenti: serve a salvaguardare, di fronte a un contesto in continua evoluzione, la capacità di un OAIS di compiere il proprio dovere;

- La funzione di “Accesso” garantisce agli utenti finali la possibilità di localizzare, chiedere e ottenere gli oggetti conservati: questo comprende l’inoltro del DIP (i cui metadati saranno differenti rispetto all’AIP), nonché tutte le modalità di controllo degli accessi, gestioni dei diritti e i meccanismi di sicurezza implementati.
- La funzione di “Amministrazione” è responsabile della gestione e del coordinamento delle attività dell’OAIS, tra cui:
 - Rapporto con i Produttori (accordi per il versamento);
 - Rapporto con i Consumatori (supporto al servizio utenti);
 - Gestione (implementare e mantenere politiche e standard).
- I “Servizi generali” sono quelli che riguardano trasversalmente tutto l’archivio, e comprendono ma non si limitano a:
 - i servizi per il sistema operativo;
 - i servizi di rete;
 - i servizi per la sicurezza (autenticazione, ...).

Tutti questi aspetti vanno attentamente considerati in rapporto con l’utente finale, il Consumatore di OAIS. La casistica di quali possano essere le varie comunità di riferimento e le rispettive necessità è ampia, nel nostro caso specifico è presumibile che possano essere interessati agli archivi digitali di scrittori e giornalisti i filologi, i linguisti e gli storici della lingua, della letteratura o dell’arte dell’età contemporanea.

Le esigenze di questa comunità di riferimento, e le riflessioni che ne nascono, verranno approfondite più avanti; per ora è importante notare come già le competenze richieste, e le prospettive che si delineano ad ogni passaggio sono piuttosto ampie: è fondamentale quindi creare una rete proficua di collaborazioni tra i professionisti dei vari settori coinvolti, e procedere ad una approfondita fase progettuale.

Sistema di controllo qualità

Lo standard ISO 9001:2015, nell'introduzione, afferma che

«L'adozione di un sistema qualità è una decisione strategica che aiuta l'organizzazione a migliorare le proprie performance e costituisce una solida base per iniziative di sviluppo sostenibile»³⁹:

cioè, in sostanza, parlare di un sistema di controllo della qualità vuol dire – semplificando – adottare ad ogni livello strategie organizzative volte a migliorare l'efficienza e ottimizzare i risultati di un processo.

Perciò, adottare un sistema di gestione e controllo della qualità, significherà mettere in atto una serie di accorgimenti che permetteranno di mantenere nel tempo e far funzionare correttamente la struttura volta alla conservazione: non si tratta, come anche per il modello OASIS, di precise indicazioni operative e dettagli sugli strumenti da utilizzare, bensì di un insieme di accorgimenti generici che possono essere applicati da qualsiasi organizzazione in un qualsiasi settore di attività, e che non sono specifici per nessuna tipologia di prodotto in particolare.

Il controllo della qualità si può applicare a una qualsiasi organizzazione che cerchi di ottenere dei vantaggi dall'applicazione di questo sistema o che miri a ottenere la fiducia dei propri clienti o utenti, ai clienti o utenti stessi, a chi certifica il sistema, a chi si occupa di formazione, in genere a chiunque sia coinvolto a qualsiasi livello nel processo: infatti lo standard rivolge l'attenzione ai processi, con una metodologia basata sul ciclo PDCA (*Plan, Do, Check, Act*):

- *Plan*: stabilire gli obiettivi del sistema, dei processi che lo compongono e delle risorse necessarie per raggiungere i risultati in accordo con le richieste del cliente e le politiche dell'organizzazione;
- *Do*: implementare quanto è stato pianificato;
- *Check*: controllare e (dove possibile) misurare i processi, i relativi risultati e i servizi rispetto a politiche, obiettivi e requisiti, e riportare i risultati;

³⁹ ISO 9001:2015

- *Act*: prendere provvedimenti per migliorare la resa dei processi, nelle modalità in cui è necessario;
- ed è organizzato intorno a sette principi:
- Orientamento al cliente/utente: incontrare le necessità del cliente/utente, attrarne la fiducia e saperla mantenere nel tempo porta a maturare un successo duraturo, al quale contribuisce il saper interpretare i bisogni futuri;
 - Leadership: una buona leadership a tutti i livelli garantisce unità di intenti, direzione e coinvolgimento, permette l'allineamento di risorse, processi, strategie e porta a raggiungere gli scopi;
 - Coinvolgimento del personale: personale competente, coinvolto e con posizioni di autorità aumenta la capacità dell'organizzazione di creare valore;
 - Approccio per processi: gestire come un sistema coerente di processi interdipendenti l'insieme delle attività porta a migliorare l'efficacia e l'efficienza dell'organizzazione;
 - Miglioramento: fondamentale per mantenere i risultati attuali, raggiungerne di nuovi e reagire ai cambiamenti;
 - Decisioni basate su evidenze: le decisioni basate sui dati effettivi e sulla resa dei processi è più probabile che diano risultati;
 - Gestione delle Relazioni: ottimizzare e migliorare le relazioni con soggetti esterni può portare a benefici per l'organizzazione.

Infine, lo standard articola i punti cardine in sette aree: comprendere il contesto dell'organizzazione e dei suoi processi; leadership, politiche e responsabilità; processi di pianificazione e considerazione di rischi e opportunità; processi per il supporto di risorse, persone e informazioni; operazioni orientate ai clienti, ai prodotti e ai servizi; processi per la valutazione della resa; processi per il miglioramento.

Nel nostro caso, un sistema di gestione e controllo della qualità, è uno strumento indispensabile: dal momento in cui l'autore viene contattato per la prima volta, fino alla fine del processo di una o molteplici acquisizioni, le persone, le operazioni, le competenze necessarie sono numerose e diversificate: è importante che il personale si senta responsabilizzato (e non controllato) e che abbia autorità ma, al

contempo, è fondamentale che l'organizzazione nel suo complesso possa tenere traccia dei vari passaggi. Infatti, a titolo esemplificativo, anche solo contattare un autore in maniera insistente perché non è stato registrato da nessuna parte che l'autore è già stato contattato diventa dannoso: l'autore, magari inizialmente interessato, potrebbe seccarsi a causa dell'insistenza e cambiare idea; oppure, un'azione come spostare prematuramente il conferimento dall'area di quarantena a quella per il deposito permanente, potrebbe causare danni a tutti i dati presenti. Nelle fasi di transizione, inoltre, come il ricambio di personale o il passaggio a nuovi *software*, è importante che non vadano persi i dati e le conoscenze acquisite in precedenza, ma che siano sempre recuperabili e utilizzabili: per la struttura nel suo complesso sarebbe un danno irreparabile – nonché un vincolo troppo pesante, anche per la sostenibilità – dovere la propria sopravvivenza, efficienza o efficacia alle competenze di una sola persona o all'adozione di un particolare sistema (come ad esempio un *software*), in quanto non sarebbe possibile cambiare personale (anche per volontà della persona stessa) né cambiare il sistema – qualora non fosse più economicamente sostenibile, o si volesse passare a nuovi sistemi – senza perdere informazioni di estrema importanza per il funzionamento di tutta l'organizzazione.

Perciò, è importante valutare, oltre alle strategie da applicare rispetto alle necessità pratiche e alla sostenibilità degli strumenti utilizzati, anche un sistema di gestione e controllo della qualità che permetta alla struttura di ottimizzare produttivamente tutti gli sforzi e le competenze del personale e dei processi che include, in favore del personale stesso – che non sarà sovraccaricato – e dei processi implementati, che potranno di volta in volta essere modificati.

ASPETTI LEGALI

«Secondo l'attuale normativa sugli archivi, l'accesso agli archivi privati dichiarati di notevole interesse storico deve essere garantito secondo modalità concordate tra i proprietari possessori o detentori della documentazione e il competente soprintendente archivistico. Sono esclusi dalla consultazione i singoli documenti dichiarati riservati dal Ministero dell'interno e quelli per i quali il proprietario possessore o detentore dell'archivio ponga la condizione della non consultabilità. Il Codice di deontologia non estende agli archivi privati conservati presso i proprietari, possessori o detentori o presso istituzioni private la procedura per l'autorizzazione anticipata dei documenti riservati, essendo stata considerata inammissibile un'estensione dei poteri del Ministero dell'interno alla sfera privata. Se ne può dedurre che sia le istituzioni private che conservano fonti archivistiche sia i proprietari possessori o detentori di archivi privati dichiarati di notevole interesse storico possono favorire la ricerca in serie riservate o mediante la "scrematura" o facendo ricorso al consenso degli interessati.»⁴⁰

In sostanza, i problemi legati alla gestione di queste strutture e di questa documentazione, ricadrebbero integralmente sotto la legislazione inerente gli archivi; o quasi. In realtà le questioni della gestione di questi documenti sono complicate, e non esclusivamente in materia di accesso, anche dall'incerta attività legislativa (allo stato attuale) relativamente all'ambito informatico nel suo complesso. A titolo d'esempio basti ricordare la gran confusione scatenata il 3 giugno 2015

⁴⁰ Paola Carucci, *La consultabilità dei documenti*, in (Pavone, Italy, e Dipartimento per i beni archivistici e librari 2006, p. 44)

a seguito della scadenza per l'attuazione delle prescrizioni da parte del garante della Privacy in materia di cookie⁴¹; il provvedimento dell'8 maggio 2014⁴² non è altro che l'adeguamento ad alcune direttive comunitarie⁴³ che sono state ampiamente criticate, particolarmente dai professionisti del settore: la contestazione più diffusa alle direttive, nonché alle relative applicazioni nazionali, accusava i legislatori di una scarsa conoscenza e competenza in merito al funzionamento delle strutture informatiche, per cui questa presa di posizione (da parte delle istituzioni) sarebbe un espediente più di facciata che sostanziale, di fronte agli effettivi problemi di privacy presenti sul web. Inoltre, la decennale vicenda giudiziaria di Google Books⁴⁴ dimostra che questa – se così vogliamo chiamarla – “lacuna legislativa diffusa” è un problema globale che si discute a un tavolo in cui importanti aziende internazionali possono portare e sostenere con grande forza le proprie opinioni, in quanto – trattandosi di un ambiente almeno parzialmente nuovo per il legislatore – questi grandi attori privati possono proporre soluzioni per questioni che hanno contribuito a sollevare e di cui conoscono la portata; non sempre così i legislatori.

⁴¹ <http://www.garanteprivacy.it/cookie>

⁴² («Individuazione delle modalità semplificate per l'informativa e... - Garante Privacy» 2014)

⁴³ *Ibidem*: «direttiva 2002/58/CE del 12 luglio 2002, del Parlamento europeo e del Consiglio; direttiva 2009/136/CE del 25 novembre 2009, del Parlamento europeo e del Consiglio, recante modifica della direttiva 2002/22/CE relativa al servizio universale e ai diritti degli utenti in materia di reti e di servizi di comunicazione elettronica, della direttiva 2002/58/CE relativa al trattamento dei dati personali e alla tutela della vita privata nel settore delle comunicazioni elettroniche e del regolamento (CE) n. 2006/2004»: tutti i documenti sono adeguatamente segnalati e riportati sul sito del Garante della privacy.

⁴⁴ Per citare solo uno degli articoli che hanno parlato degli esiti finali di questa lunga vicenda («Usa, Corte Suprema dà torto agli autori: Google Books libera di scansionare milioni di libri» 2016); ma ci sono due aspetti di questa vicenda che sintetizza bene Andrea Zanni nello stesso articolo: «*Da una parte, l'istituzione americana rafforza il principio per cui tutti, studiosi e non solo, abbiamo il diritto alla conoscenza. E Google fa di più: non si limita a scansionare, fa anche text mining, cioè indicizza i testi e analizza i dati. Allo stesso tempo, però, delegare la digitalizzazione del sapere a una multinazionale rappresenta un rischio. E se Google a un certo punto "chiudesse" quel sapere? Se utilizzasse quei dati a proprio vantaggio competitivo? Ad esempio, Big G può mettere a frutto tutti quei testi per darli in pasto ai propri sistemi di intelligenza artificiale, in modo da collaudarli, "allenarli" ed essere più competitiva di altri. La questione va al di là del copyright: il 70% delle opere di biblioteca è "orfano", lasciare questo patrimonio in mano a una multinazionale significa anche darle una sorta di delega alla cultura*»

Quindi, per quanto riguarda gli archivi di scrittori, giornalisti e intellettuali *nativi digitali*, il “terreno burocratico” sul quale ci si muove – e questo in riferimento alla gestione dell’archivio nel suo complesso, non esclusivamente in materia d’accesso – è piuttosto ampio: in quanto archivi di persona dovrebbero rientrare sotto la normativa che regola gli archivi, e – qualora intervenga la dichiarazione d’interesse storico – anche di beni culturali; in quanto materiale di autori viventi, la maggior parte dei documenti conferiti – quando non la totalità – possono essere soggetti alla normativa sul diritto d’autore (e non è da escludersi, né raro, il caso in cui tra i materiali preparatori della produzione effettivamente pubblicata possano esserci testi copiati da altri autori o altre fonti sul web: diritto d’autore di terze parti); in quanto materiale digitale destinato alla conservazione, infine, sono soggetti alla normativa per la conservazione dei documenti digitali. Le problematiche per la “vecchia” normativa legata a materiali analogici – e ora applicata a quelli digitali – riguardano alcune diversità inerenti le proprietà non comparabili, e spesso anche non compatibili, tra l’oggetto fisico e quello virtuale: solo per citare un esempio, per conservare un file digitale – almeno per il momento – si può solo considerare di fare delle copie autentiche, ma non di conservare indefinitamente un “originale”:

«i documenti elettronici possono prodursi e mantenersi esclusivamente nella forma di copie autentiche di documenti elettronici autentici, dato che mantenere l’accesso ai documenti implica necessariamente modifiche anche significative nel flusso di bit che costituisce il documento e le sue relazioni.»⁴⁵

Invece, le questioni relative alla “nuova” normativa – che comprende anche i supporti digitali – riguardano il fatto che in genere si tratta di disposizioni ancora piuttosto vaghe e poco incisive per l’ambito a cui si riferiscono, che dimostrano una padronanza delle dinamiche ancora decisamente da affinare:

⁴⁵ Maria Guercio, *I documenti informatici*, in (Pavone, Italy, e Dipartimento per i beni archivistici e librari 2006, p. 827)

«Sebbene l'optimum sarebbe definire e adottare, nella fase di formazione dell'archivio, le strategie e gli strumenti necessari a garantire ai documenti informatici, per tutto il loro ciclo di vita, le caratteristiche di autenticità, integrità, affidabilità, leggibilità e reperibilità, il caso dell'archivio Vannucci [e certo anche quello di cui stiamo parlando, viste le necessità di rapido intervento di cui si è precedentemente parlato], per il quale tali strumenti e strategie non sono evidentemente stati previsti, suggerisce l'ipotesi che gli archivisti dovranno comunque essere in grado di proporre soluzioni finalizzate a ottemperare, per quanto possibile, a tale mancanza.»⁴⁶

E ancora:

«Seppure le regole tecniche in materia di sistema di conservazione richiamino lo standard OAIS e specifichino che tale sistema assicura la conservazione dei documenti informatici e dei documenti amministrativi informatici [...], nessuna ulteriore informazione però è fornita in relazione al rapporto di tali metadati⁴⁷ con le predette categorie⁴⁸ previste dallo standard»⁴⁹

L'incastro tra questi differenti ma complementari comparti normativi non sempre è agevole quanto sarebbe opportuno per garantire una certa praticità nella gestione degli strumenti in quanto, per le motivazioni già accennate, le lacune obbligano a trovare una sorta di “zona grigia” che possa soddisfare le prescrizioni di ciascun comparto, senza per questo mancare o essere in difetto nei confronti di un altro.

Infine è da notare che la complementarità – in buona parte dei casi – è puramente ideale e non esplicitamente sancita nel nostro specifico caso

⁴⁶ Sisti, *La descrizione archivistica...*, in (Allegrezza e Gorgolini 2016, pp. 78-79).

⁴⁷ I metadati minimi del documento informatico secondo l'allegato 5 del d.p.c.m. 3 dicembre 2013 sono: *IDDocumento* (l'identificativo univoco e persistente); *Oggettodocumento* (“oggetto del documento”, riguardo il contenuto e la natura del documento); *Destinatario* (denominazione e codice fiscale); *Soggetto produttore* (denominazione e codice fiscale); *Datachiusura* (data dell'ultima modifica sul documento autentico: tutte le successive modifiche si considerano manomissioni).

⁴⁸ Le categorie a cui si fa riferimento sono i cinque componenti delle informazioni sulla conservazione (*Preservation Description Information - PDI*) di OAIS: informazioni sull'identificazione; informazioni sul contesto; informazioni sulla provenienza; informazioni sull'integrità; informazioni sui diritti di accesso.

⁴⁹ Sisti, *La descrizione archivistica...*, in (Allegrezza e Gorgolini 2016, p. 85).

di applicazione. Infatti, benché l'archivio (digitale e di persona) conferito con somma probabilità non sia affatto o sia ben lungi (temporalmente parlando) dall'essere dichiarato un bene d'interesse storico, è opportuno trattarlo in maniera conforme alle leggi relative in modo che, una volta sopraggiunta la dichiarazione o nel caso in cui venga successivamente integrato all'interno del sistema archivistico, non ci siano problemi: si tratta infatti, a tutti gli effetti di un archivio. Proprio per motivi materiali, inoltre, la legge sul diritto d'autore non sempre permette di fruire di tutti i vantaggi del digitale rispetto all'analogico⁵⁰, e i concetti stessi di diritto d'autore e di *copyright* negli ultimi anni vengono spesso messi in discussione e criticati⁵¹. E ancora una volta la discrepanza normativa tra materiale analogico e "immateriale" digitale viene ricordata da Valacchi:

«L'articolato castello costruito dal legislatore intorno a documento informatico, per quanto sostanzialmente valido in sé, non ha ponti levatoi (e nemmeno porte) che si aprano in direzione del raccordo giuridico, tecnico, operativo ed economico tra due modelli, quello della produzione e gestione e quello della conservazione, che muovo da presupposti diversi verso altrettanto diversi obiettivi. È pur vero che le più recenti disposizioni in materia di archivi intesi come beni culturali sembrano rivolgersi essenzialmente al mondo analogico e ripropongono modelli organizzativi e conservativi che non recepiscono la peculiarità e le emergenze che derivano dalla necessità di conservazione dei documenti elettronici.»⁵²

⁵⁰ Un esempio su tutti: i cosiddetti applicativi di gestione dei diritti *Digital Rights Management* (DRM); il mercato musicale li ha abbandonati in favore di altri sistemi, e il mercato editoriale – per quanto ancora vengano usati moltissimo – comunque ha solide opposizioni istituzionali e non, e anche voci di dissenso interne al mercato editoriale stesso (l'editore è meno pronò dell'agente dell'autore ad applicare i DRM). Di fatto questi applicativi impongono delle restrizioni molto pesanti al punto che, per alcuni aspetti, la lettura in digitale nel suo complesso, può diventare più ostica di quella analogica, senza beneficiare dei pochi vantaggi offerti dal libro digitale.

⁵¹ Per citare solo uno degli articoli più recenti (Falkvinge 2017), che rimanda a propria volta (Karl Fogel 2005).

⁵² (Valacchi 2015, p. 55)

La normativa archivistica e sul documento digitale

«Volendo schematizzare, [...] possiamo intanto ricondurre la normativa vigente a tre leggi fondamentali: il DPR 445/2000 (Testo unico sul documento amministrativo, con le successive modificazioni), il D.lgs. 22/01/2004 n. 42 (Codice dei beni culturali e del paesaggio e successive modifiche) e il D.lgs. 7 marzo 2005, n. 82 (Codice dell'Amministrazione digitale e relative integrazioni, con particolare riferimento alle regole tecniche per la conservazione pubblicate nel 2014). Mantiene poi vigore, fatta salva l'abrogazione di alcuni articoli, il DPR 1409/1963. A questi strumenti di ordine generale si aggiungono regolamenti, regole tecniche e norme particolari nonché l'insieme delle procedure di gestione che ogni soggetto produttore [normalmente inteso come soggetto pubblico, ma nel nostro caso non necessariamente], sulla base delle proprie peculiarità ed esigenze è tenuto a mettere in pratica.⁵³»

Nel nostro specifico caso la normativa generale sugli archivi di Stato non sarà di particolare utilità, in quanto si riferisce principalmente agli enti statali o pubblici, e alla relativa documentazione amministrativa (così come non saranno di grande rilievo le norme sui documenti amministrativi); tuttavia, il d.p.r. 1409/1963 “Norme relative all’ordinamento ed al personale degli Archivi di Stato” agli articoli 37 e 38 delibera quanto segue:

«Art. 37: I privati proprietari, possessori o detentori, a qualsiasi titolo, di archivi di cui facciano parte documenti di data anteriore all'ultimo settantennio, hanno l'obbligo, [...] di darne notizia per iscritto al sovrintendente archivistico e al prefetto della Provincia.[...]»

Ma, soprattutto, l’art. 38, sancisce l’obbligo – per i privati che non ricorrano al deposito (art. 39) – di espletare tutta una serie di azioni tra cui:

⁵³ *Ibidem*, p. 50

«Art. 38. I privati proprietari, possessori o detentori degli archivi o dei singoli documenti dichiarati di notevole interesse storico hanno l'obbligo di:

a) conservare gli archivi e i singoli documenti, nonché ordinarli e inventariarli, o consentire che all'ordinamento e all'inventariazione provveda il sovrintendente archivistico. [...]

b) permettere agli studiosi, [...] la consultazione dei documenti;

d) procedere al restauro dei documenti deteriorati, o consentire che vi provveda il competente sovrintendente archivistico;

e) non trasferire, a titolo oneroso o gratuito, la proprietà, il possesso o la detenzione degli archivi o dei singoli documenti, senza darne preventiva notizia [...];

f) non esportare dal territorio della Repubblica gli archivi o i singoli documenti senza la preventiva autorizzazione [...];

g) non smembrare gli archivi, [...];

h) non procedere a scarti senza osservare la procedura [...art. 42];

i) consentire al sovrintendente archivistico di procedere, previa intese, a visite per accertare l'adempimento degli obblighi [...].»

In maniera simile anche il Codice dei Beni Culturali e del Paesaggio (artt. 19, 21, 30, 127), però questa volta, facendo provenire la norma dall'ambito dei beni culturali, e non dalla regolamentazione dell'attività amministrativa. Quanto alla normativa sull'amministrazione digitale saranno dati alcuni riferimenti, ricordando però come questa riguardi l'ambito e i documenti della pubblica amministrazione, e che pertanto le esigenze, e anche il lessico utilizzato, non aderiranno sempre e sistematicamente a quelle del contesto in esame.

«Nel caso della descrizione di documentazione informatica, invece, la vigente normativa nazionale individua nei metadati lo strumento necessario per la descrizione del contesto e del contenuto della documentazione archivistica, come espressamente specificato nel glossario allegato alle regole

tecniche in materia di: sistema di conservazione (d.p.c.m. 3 dicembre 2013); protocollo informatico (d.p.c.m. 3 dicembre 2013); formazione, trasmissione, copia, duplicazione, riproduzione e validazione temporale dei documenti informatici nonché formazione e conservazione dei documenti informatici delle pubbliche amministrazioni (d.p.c.m. 13 novembre 2014).⁵⁴»

Nel d.p.c.m. 3 dicembre 2013 tra gli standard riconosciuti e adottati come riferimento per la conservazione vengono citati ISO 14721:2012 (OAIS) e ISO 15836:2009 (*Dublin Core Metadata Element Set*). Tuttavia, è da notare come *Dublin Core* non sia previsto nativamente dal modello OAIS in quanto gli aspetti inerenti le scelte tecniche e tecnologiche vengono «*ritenuti competenza di esperti di settore e oggetto di specifici interventi di normalizzazione*»⁵⁵, e lasciati pertanto a discrezione dei singoli casi che vengono a presentarsi, non decisi a monte (anche perché OAIS nasce per il trattamento e la conservazione di dati spaziali: l'ampia diffusione senz'altro non dipende dalla specificità delle scelte per un determinato ambito, quanto dall'efficacia del modello che propone). In ogni caso, una volta che vengano rispettati questi parametri normativi (già noti e rispettati da chi opera negli archivi), si è conclusa solo la parte meno delicata delle procedure legali e burocratiche a cui il nostro contesto deve necessariamente sottoporsi.

La normativa sul diritto d'autore

Poiché stiamo parlando di autori, e di autori viventi, una parte del materiale da conservare sarà soggetto senza ombra di dubbio alla legge sul diritto d'autore: stesure definitive di romanzi, racconti, bozze contenenti materiali già pubblicati, e altro materiale pubblicato di cui il conferimento conterrà sicuramente "le copie originali". Naturalmente si tratta al contempo di materiale d'archivio, e pertanto soggetto a un

⁵⁴Sisti, *La descrizione archivistica...*, in (Allegrezza e Gorgolini 2016, p. 82).

⁵⁵(Carucci e Guercio 2008) p. 301

accesso piuttosto limitato, per il quale ci sono alcune considerazioni da fare. Fino a non troppo tempo fa, quando il materiale d'archivio era esclusivamente cartaceo – e l'archivio ne conservava pertanto gli originali – l'accesso ai documenti era permesso sulla base della tipologia documentaria (amministrativo o meno) e di conseguenza su base temporale (a distanza di un certo tempo dalla formazione del documento). Nel corso del Novecento, comincia a svilupparsi una certa sensibilità anche per le “carte d'autore”, per quei documenti – privi di un legittimo deposito istituzionale – che testimoniano il lavoro creativo e intellettuale degli autori, e che non hanno precedentemente trovato casa altrove: se è vero che il regio decreto 27 maggio 1875 n. 2552 all'articolo 13 dice che

«Gli atti che hanno carattere puramente storico, letterario o scientifico; le sentenze e i decreti dei magistrati; le decisioni e i decreti delle autorità governative e amministrative; gli atti dello stato civile delle persone; gli atti delle provincie, dei comuni e dei corpi morali occorrenti alla loro amministrazione; gli atti necessari allo esercizio dei diritti elettorali, alla prova dei servizi civili o militari, ed allo svincolo delle cauzioni dei contabili dello Stato; sono pubblici qualunque sia la loro data»

è anche verso che si riferisce pur sempre ad “atti”, e al loro livello di interesse “puro”, cioè al fatto che i relativi contenuti con buona probabilità non abbiano più efficacia amministrativa, senza con questo intendere che non siano effettivamente atti amministrativi. Solo con la legge 2006 del 22 dicembre 1939 è istituito un intero titolo – il titolo VI (artt. 21-28) – dedicato agli archivi privati di interesse storico: la legge però a causa della seconda guerra mondiale, ebbe solo “*parziale e tardiva attuazione*”, come ricorda Valacchi. Il già citato d.p.r. 1409/1963 conferma – negli articoli evidenziati più sopra – questa linea operativa orientata definitivamente all'inclusione anche degli archivi d'interesse storico e culturale, cioè non esclusivamente amministrativi. Poco più tardi cominciano a nascere alcune realtà più specifiche: l'atto notarile della nascita del “Centro di ricerca sulla tradizione manoscritta di autori moderni e contemporanei” di Pavia del

1973, poi definitivamente riconosciuto con decreto rettorale nel 1980; l'istituzione nel 1975 dell'Archivio Contemporaneo (poi Bonsanti) presso il Gabinetto Vieusseux di Firenze, solo per citare due esempi importanti e non dedicati a un unico autore. Ci si è trovati quindi in un periodo piuttosto recente a dover fare i conti con materiali più particolari sotto il profilo giuridico rispetto a quelli normalmente in possesso degli archivi di Stato: posto, però, che in archivio sono conservati i documenti originali, che l'accesso tendenzialmente è limitato a motivi di ricerca e che i documenti particolarmente sensibili sono preclusi all'accesso (o comunque non è possibile citarli, in quanto non accessibili), questi vincoli fisici e giuridici non hanno portato a preoccuparsi particolarmente della legge sul diritto d'autore. Questo però dipende anche dalla tipologia di lavoro (o pubblicazione) a cui si va incontro: un ricercatore che utilizzi personalmente i blocchetti di Montale (conservati presso il Centro di Pavia) per un lavoro di ricerca li utilizzerà esclusivamente *in loco*, e pubblicherà citandoli opportunamente; inoltre è strano anche solo pensare che possa accedere a quei materiali con obiettivi e finalità diverse dalla ricerca; se però lo stesso ricercatore volesse pubblicare un epistolario, la legge sul diritto d'autore, all'articolo 93 dichiara che

«le corrispondenze epistolari, gli epistolari, le memorie familiari e personali e gli altri scritti della medesima natura, allorché abbiano carattere confidenziale o si riferiscano alla intimità della vita privata, non possono essere pubblicati, riprodotti od in qualunque modo portati alla conoscenza del pubblico senza il consenso dell'autore, e, trattandosi di corrispondenze epistolari e di epistolari, anche del destinatario. Dopo la morte dell'autore o del destinatario occorre il consenso del coniuge o dei figli, o, in loro mancanza, dei genitori; mancando il coniuge, i figli e i genitori, dei fratelli e delle sorelle, e, in loro mancanza, degli ascendenti e dei discendenti fino al quarto grado. Quando le persone indicate nel comma precedente siano più e vi sia tra loro dissenso, decide l'autorità giudiziaria, sentito il Pubblico

Ministero. È rispettata, in ogni caso, la volontà del defunto quando risulti da scritto.»

Benché questa tipologia di documenti complichi notevolmente la possibilità di utilizzare determinati materiali, per motivi di studio, su materiale eminentemente “di lavoro”, o comunque a carattere non privato, non dovrebbero esserci grossi problemi per poterlo utilizzare. Potrebbe non essere esattamente la stessa situazione per quanto riguarda l’ambito digitale: conservando solo “copie autenticate”, ma pur sempre copie, e scambiando il materiale digitale attraverso un processo di copia, non avrò più un unico originale consultabile, fisicamente conservato in un luogo geograficamente ben definito, ma potenzialmente avrò tante copie identiche per quanti fanno richiesta e ottengono di poter accedere. Per questo motivo – nel caso in cui ci fosse un qualsiasi tipo di vincolo giuridico – è necessario e indispensabile mettere in atto delle misure tecnologiche volte a permettere l’accesso ma non la copia dei materiali in consultazione: diversamente chiunque potrebbe copiare, riutilizzare, rivendere o disporre in qualsiasi modo di un patrimonio di cui, invece, non potrebbe legalmente disporre. In questo senso la legislazione vigente non ha ancora ben compreso le implicazioni molto differenti dell’impatto del digitale: basti pensare che la legge in materia di “protezione del diritto d’autore e di altri diritti connessi al suo esercizio”⁵⁶, benché aggiornata l’ultima volta con D.lgs.15 gennaio 2016, n. 8, è una legge risalente al 1941, e che giustamente prevedeva delle casistiche molto più stabili (la materialità dei supporti non ne permette questa sorta ubiquità che invece è possibile nel digitale). Il tema nel suo complesso è troppo ampio per presentarne qui una discussione estesa che non potrebbe comunque avere la pretesa di essere né esaustiva né tantomeno risolutiva, tuttavia è importante segnalare questi aspetti come spunti di riflessione: nel momento in cui si “mette qualcosa sul web” non sempre si è consci di come lo si fa, di cosa questo possa implicare o se ci siano o meno i diritti per utilizzare un determinato oggetto digitale; a maggior ragione per chi vuole

⁵⁶(*Protezione del diritto d’autore e di altri diritti connessi al suo esercizio* 1941)

conservare il digitale, è fondamentale riflettere su come questi dati saranno accessibili, da chi, in che forma e in che misura, che cosa potrà farne e come, o se non saranno accessibili esternamente. Questo già è in parte, ma per alcuni aspetti dovrebbe esserlo maggiormente, terreno di confronto tra le istituzioni della conservazione e i politici, per determinare un quadro legislativo atto a garantire il rispetto e la conformità con l'ordinamento, ma anche un'adeguata flessibilità rispetto al mezzo che si usa; auspicando con ciò non già un lassismo, ma una riflessione coordinata e produttiva tra tecnici e politici.

Il diritto d'autore di “terze parti”

Quando ci si confronta con queste tipologie d'archivio non è affatto raro trovare documenti non propriamente dell'autore. Il computer, al giorno d'oggi, è uno strumento estremamente versatile, e l'integrazione possibile con *smartphone* e *tablet* non fa che confermare questo aspetto. Grazie all'informatica e alle connessioni è possibile lavorare da qualsiasi posto e quasi con qualsiasi dispositivo: scrivere e rispondere a mail, prendere appunti su pagine web, comporre articoli e bozze mentre si è in viaggio in aereo, in treno o in quasi qualsiasi altra situazione. La rapidità e soprattutto la facilità con cui è possibile copiare e riutilizzare documenti, immagini, video e altri oggetti digitali, ci rende sempre meno coscienti del fatto che se è possibile farlo materialmente, non è detto che allo stesso tempo sia possibile o opportuno farlo anche legalmente. Un articolo di tre anni fa⁵⁷ mostrava – in un'esperienza nata per gioco – come il web ha cambiato moltissimo il panorama informativo: l'importanza è la rapidità, non per forza l'esattezza (e questo – a volte – a costo di prendere qualche grosso granchio, come mostra l'articolo); mostra anche però quanto sia più semplice, ma soprattutto “facile” copiare i contenuti. Gli scrittori e gli intellettuali, non avranno archivi di lavoro contenenti solo una unica tipologia documentaria ma tante quanti sono i lavori che seguono. Emblematico in questo senso è Beppe Severgnini: nella sezione

⁵⁷ (Virgillito 2014)

autobiografica del suo sito è possibile notare le diverse attività che svolge o ha svolto⁵⁸: editorialista, autore di libri, moderatore di blog, autore e conduttore televisivo, collaboratore di emittenti radiofoniche e corrispondente all'estero. L'archivio di Severgnini, conseguentemente, rispecchierà e conterrà i documenti relativi al suo lavoro, a ciascuno degli ambiti del suo lavoro. Il discorso però si può fare per qualsiasi autore, poiché sono veramente pochissimi gli scrittori o gli intellettuali che riescono a vivere solo ed esclusivamente della scrittura di libri (e anche in quel caso non è detto che i documenti contenuti siano di un solo tipo). Nel caso di Severgnini ci troviamo di fronte a una compagine piuttosto nutrita e diversificata di materiali, e possiamo vedere numerosi casi: esistono le bozze dei libri e degli articoli, esistono i lavori per la televisione; soprattutto, e sono molto numerosi, esistono quelli che potremmo definire “materiali preparatori”. Severgnini, in quanto personaggio celebre, non di rado viene invitato in televisione (in programmi di cui non è autore o conduttore, in riferimento ai materiali d'archivio) per essere intervistato o per parlare di determinati argomenti; inoltre, come giornalista e saggista scrive articoli o saggi su argomenti di cui non è necessariamente esperto prima di cominciare a scrivere: in tutti questi casi, ma anche in moltissimi altri, una qualsiasi persona si documenta. Se uno dovesse lavorare a un saggio o a un articolo per cui alcune delle fonti (o tutte nel caso di temi di estrema attualità) si trovano in rete, è molto più comodo averle tutte insieme copiandone il testo e incollandolo in un unico documento, magari rielaborandole e incrociandole tra loro, e lo stesso vale nel momento in cui ci si prepara per parlare di un tema in radio o in televisione, per un'intervista come anche per un libro. Inoltre, alcuni autori possono conservare – sempre copiando e incollando il testo in uno o più documenti – le recensioni alle proprie pubblicazioni, gli articoli che parlando del proprio lavoro e altri documenti ancora. In tutti questi casi, i materiali (che fanno parte effettivamente dell'archivio), hanno un aspetto in comune: pur se il soggetto produttore autorizzasse nel modo più ampio possibile i diritti

⁵⁸ <http://www.beppevergnini.com/chi-sono/>; nel capitolo relativo a Pavia Archivi Digitali verrà approfondita meglio la presenza di Severgnini.

d'uso del proprio archivio non avrebbe nessun potere nell'autorizzare tutti questi documenti conferiti i cui diritti effettivi sono in capo a terzi: tutti coloro cioè che sono gli autori di questo materiale. Trattandosi però di documenti i cui autori sono, di norma, viventi – e quindi soggetti per lungo tempo ancora al diritto d'autore – ed essendo in un ambiente, quello digitale, in cui i costi e le difficoltà di ottenere una copia sono minimi, occorre sempre avere una particolare attenzione nel capire come si possano gestire questi documenti, senza dover necessariamente impedire un accesso dall'esterno.

Le licenze aperte

Talvolta, tuttavia, alcuni casi possono essere più semplici del previsto, per esempio nel caso in cui i documenti copiati siano pubblicati con una licenza aperta. Da diverso tempo, ormai, si stanno diffondendo – particolarmente in ambito digitale – queste tipologie di licenza: cercheremo, brevemente, di capire che cosa sono, come possono essere utilizzate, e cosa implicano per quanto concerne l'ambito di questa trattazione. Il fenomeno delle licenze “*open*”, o meglio licenze “*open source*” nasce in ambito informatico o, più precisamente, in questo contesto nasce il *software* chiuso. Da quando iniziano a diffondersi i primi *Personal Computer*, con relativo *software*, iniziano a diffondersi anche le versioni proprietarie dei programmi (benché fino a quel momento fosse una pratica comune condividere codice e conoscenze tra gli sviluppatori): senza scendere nei dettagli né storici né tecnici, sia sufficiente sapere che a livello tecnico il codice sorgente (il “*source*” di cui sopra) è il programma scritto con un linguaggio di programmazione nella fase iniziale e che poi, dopo una serie di passaggi, sarà possibile utilizzare: il *software* proprietario non permette di leggere il codice sorgente (che quindi non sarà *open*) né di modificare il programma. A livello storico invece è utile menzionare che nella prima metà degli anni Ottanta Richard Stallman fondò la *Free Software Foundation*, dove “*Free*” è da intendere nel senso di libero, non di gratuito; da quel momento in poi nel corso degli anni (primi anni Novanta) nacque

GNU/Linux e poi ancora, verso la fine degli anni Novanta cominciò a diffondersi il termine *open source*, per designare più chiaramente che non si trattava di un concetto legato al valore commerciale, ma alla trasparenza e al coinvolgimento nello sviluppo di una comunità di persone. Numerose altre esperienze si sono ancora sviluppate nel corso degli anni, sempre legate all'ambito *software*: ma da questa idea ha iniziato a prendere piede anche il cosiddetto movimento per il *copyleft* (gioco di parole in opposizione a *copyright*) per cui l'idea di base è che anziché avere “tutti i diritti riservati” ci siano piuttosto “alcuni diritti riservati”. Gli obiettivi principali sono la diffusione e la condivisione, nell'ottica che se più persone collaborano alla realizzazione di un'idea questa potrebbe dare un risultato migliore qualitativamente, ed essere contemporaneamente più sostenibile per tutti (ma ricordiamoci che aperto/libero non significa gratuito). Le licenze più diffuse (per le opere creative in genere, non tra quelle esclusive per il comparto *software*) sono le licenze Creative Commons (CC)⁵⁹: nate nel 2002 da un'idea di Lawrence Lessig, docente di legge presso la Stanford University. Le licenze CC non sono “contro” il *copyright*, vogliono piuttosto essere una mediazione tra esso e il pubblico dominio; principalmente si basano su due libertà: condividere e riutilizzare (ove previsto); oltre a ciò ci sono quattro possibili condizioni di utilizzo dell'opera: l'attribuzione (BY) (l'unica condizione sempre obbligatoria), l'uso non commerciale (NC: *Non Commercial*; non si possono ricavare profitti dal riuso dell'opera), l'uso non derivativo (ND: *Non Derivative Works*; non è previsto che l'opera possa essere riutilizzata per realizzare nuove opere), e quello “condividi allo stesso modo” (SA: *Share Alike*; incompatibile con l'uso non derivativo: tutte le opere derivate che utilizzino un'opera con licenza SA sono obbligate a utilizzare una licenza SA anche per l'opera derivata). La combinazione delle due libertà e delle quattro condizioni d'uso genera sei possibili scenari che dal più aperto al più ristretto sono i seguenti:

- CC BY : è necessario indicare solo ed esclusivamente l'attribuzione dell'autore (può essere riutilizzata in opere

⁵⁹ Una storia sommaria la si può trovare in (Creative Commons 2017); il sito ufficiale è <https://creativecommons.org/>.

derivate e anche per scopi commerciali, inoltre l'opera derivata potrà essere pubblicata con qualsiasi altra licenza, o anche sotto *copyright*);

- CC BY – SA : oltre all'attribuzione è necessario condividere l'eventuale opera derivata con la stessa licenza (è possibile farne anche un uso commerciale);
- CC BY – ND : non è prevista la possibilità che vengano create opere derivate (e perciò non è nemmeno prevista la condizione SA che sarebbe contraddittoria);
- CC BY – NC : è previsto qualsiasi uso dell'opera, purché non sia per trarne profitto commerciale;
- CC BY – NC – SA : è previsto qualsiasi uso dell'opera purché non sia per trarne profitto e purché l'eventuale opera derivata venga pubblicata con la stessa licenza;
- CC BY – NC – ND : non è possibile creare opere derivate né fare un uso commerciale dell'opera stessa.

C'è, infine una settima possibile condizione d'utilizzo dell'opera dell'ingegno: quella che ricade dentro il pubblico dominio (previsto dalla legge sul diritto d'autore italiana – come in molti altri stati – a settant'anni dalla morte dell'autore); poiché, però, in moltissimi stati non è possibile rinunciare totalmente ai diritti morali sull'opera, vivente l'autore, si configura più spesso come il minimo dei diritti esercitabili da parte dell'autore sull'opera stessa.

La licenza CC è composta da tre “livelli”: c'è il *legal code* ovvero la licenza vera e propria, il documento che ha valore legale e permette che la licenza sia giuridicamente rilevante e si possa applicare; c'è il *commons deed*, che non ha un vero e proprio valore legale ma è il testo che si trova associato di solito alle licenze e serve a spiegarne il funzionamento all'utente medio; infine, l'ultimo livello è il *digital code*, cioè i metadati in RDF⁶⁰ che servono a rendere la licenza rintracciabile in automatico dai motori di ricerca e da altri sistemi di

⁶⁰ *Resource Description Framework*: è uno standard del W3C per il web semantico e ne costituisce lo standard per lo scambio delle informazioni; al proposito cfr. l'illustrazione del “*semantic web stack*”:
http://bnode.org/media/2009/07/08/semantic_web_technology_stack.png

automazione informatica (particolarmente utile per identificare la licenza e le relative attribuzioni). Queste licenze possono essere applicate a qualsiasi opera dell'ingegno e non è assolutamente necessario che l'opera sia distribuita o pubblicata sul web o in rete: certo è che la flessibilità rispetto alla condivisione e al riutilizzo le rendono ottimali per i contenuti che vengono pubblicati quotidianamente online. Tuttavia, cercando di riprendere un po' le fila di questo discorso piuttosto ampio sui termini legali – che non sempre sono decisi secondo competenza del legislatore nel campo che vanno a normare – il funzionamento di queste licenze non è ancora abbastanza diffuso da venire compreso da tutti, soprattutto da chi poi ne regola l'applicazione: Simone Aliprandi – avvocato esperto nel diritto di proprietà intellettuale, che si occupa da anni delle licenze CC – notava informalmente nel suo blog⁶¹ come il sito del ministero dell'interno abbia adottato con formule poco consone le licenze CC per i propri contenuti, ritenendo che l'istituzione statale nel suo complesso dovrebbe essere preparata e competente nelle scelte che intraprende:

«Innanzitutto, c'è da rilevare che trattare come un tutt'uno "dati, informazioni testuali ed elementi multimediali" non risulta la scelta migliore, dato che per effetto della cosiddetta Direttiva Public Sector Information, dati e documenti prodotti dalla PA sono sottoposti a una disciplina diversa rispetto alle altre opere dell'ingegno che vengono prodotte e diffuse dalla PA (fotografie, filmati, musiche, contenuti multimediali). Sarebbe quindi più opportuno tenere separate le due categorie anche nelle note legali dei siti, oppure - scelta consigliabile - applicare ad entrambi il regime di copyright più aperto.

Poi - e questo è il nodo più delicato - dire genericamente "licenza Creative Commons 2.5" non è sufficiente. Quella dicitura potrebbe infatti riferirsi a sei diverse licenze; anzi, a ben vedere, se non si indica anche la versione nazionale della licenza (Italia, Francia, USA...) potrebbe riferirsi a più di un centinaio di

⁶¹ (Aliprandi 2016): l'osservazione delle incongruenze hanno portato a una modifica spontanea e ottimale da parte del sito del Ministero dell'interno, come segnalato dall'articolo.

licenze. È per questo che, a scanso di equivoci, le linee guida di Creative Commons (e anche la buona prassi del settore (che - lasciatemi dire - ormai ogni legale specializzato in diritto della proprietà intellettuale o diritto dell'informatica dovrebbe conoscere) consigliano di indicare sempre l'URL completo in cui poter risalire al testo integrale della licenza. Il tutto sembrerebbe risolto da un link che si trova invece nel footer del sito [...]; questo link però rimanda a una licenza Creative Commons Attribuzione - Non commerciale - Non opere derivate 3.0 Italia (questo sarebbe il nome completo e corretto della licenza) quindi non a una versione 2.5, come invece indicato sopra. Ci si chiede dunque quale sia la licenza realmente prescelta. Possiamo forse risolvere il tutto con il buon senso, evitando di essere eccessivamente rigidi e pensando che il riferimento da considerare sia quello del footer in quanto più aggiornato e completo di link. Ma i problemi non finiscono qui. Se la licenza che il Ministero voleva applicare era davvero la BY-NC-ND, bisogna far presente che, almeno per quanto riguarda dati e documenti, quella licenza è contraria a quanto previsto dalle "Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico (anno 2014)" emesse dall'Agenzia per l'Italia Digitale (AGID) e tutt'ora principale riferimento per il settore. Al Capitolo 8 delle linee guida si legge infatti (in un box ben evidenziato):

Facendo riferimento alla definizione Open Data fornita dall'Open Knowledge Foundation (OKF) per cui un dato è aperto se è “usabile, riutilizzabile e ridistribuibile liberamente da chiunque anche per finalità commerciali, soggetto al massimo alla richiesta di attribuzione e condivisione allo stesso modo”, le sole licenze ammesse per abilitare l'effettivo paradigma dell'Open Data sono classificate come mostrato in Figura 7. Come evidenziato in figura, tutte le altre licenze che non consentono lavori derivati, anche per finalità commerciali (i.e., licenze che riportano chiaramente clausole Non Commercial - NC

e/o Non Derivative – ND e/o ogni altra clausola che limita la possibilità di riutilizzo e redistribuzione dei dati) non possono essere ritenute valide per identificare dati di tipo aperto.»⁶²

Secondo quanto riportato ci sarebbero diverse incongruenze inerenti la comprensione sia del sistema delle licenze, che della loro applicazione che, infine, del rapporto che intercorre tra l'applicazione delle licenze e l'applicazione della normativa nazionale. Sia detto senza spirito polemico che non c'è da meravigliarsi né da indignarsi rispetto a questi punti: è chiaro che le rapide evoluzioni sul fronte della tecnica non permettono un'altrettanto rapida ed efficace presa di posizione da parte del legislatore; si vorrebbe però che chi amministra sia capace di affidare a chi è competente in materia (ove non sia possibile farlo in prima persona) l'incarico di elaborare strategie efficaci e propositive per affrontare questi cambiamenti in modo opportuno.

In conclusione, non sembra sbagliato almeno auspicare la possibilità che gli autori nativi digitali che conferiscono i propri archivi per la conservazione possano, in futuro – se non è ora il caso –, valutare la possibilità di rendere disponibili i propri documenti (naturalmente quelli per i quali non sono previste restrizioni d'accesso) con una licenza aperta, che ne permetta l'uso e la diffusione.

Se questi documenti fossero resi disponibili in modo aperto potrebbero essere sia delle fonti assolutamente interessanti, per vedere come gli autori ricercano, cosa selezionano, come cambiano e interpretano i contenuti che consultano (e come evolvono i paradigmi di ricerca nel tempo), che allo stesso tempo delle risorse autoritative, tanto nel senso dell'autore che dichiara una determinata cosa, quanto in quello dell'istituzione (in questo caso l'istituzione che ospita il conferimento dell'autore) che fornisce descrizioni istituzionali riutilizzabili per creare un sistema di fonti autorevoli nel web.

⁶²*Ibidem.*

IL PROFILO CONCETTUALE

«L'emergere di archivi multimediali ha confuso la netta distinzione tra il passato (conservato) e (l'illusione della) presenza e perciò è ben più che solo un'estensione o un riassetamento di ben note pratiche archivistiche. I fantasmi dell'archivio nel cyberspazio sono una deviazione ideologica dell'immediata cancellazione degli archivi (sia hardware che software) nel mondo digitale. "Il ventesimo secolo, il primo nella storia ad essere esaustivamente documentato da archivi audiovisivi, si trova sotto la maledizione di quello che un filosofo contemporaneo ha definito il 'mal d'archivio'⁶³, un male che, viste le capacità

⁶³ (Derrida 1996, p. 2): «Come parlare di una "comunicazione degli archivi" senza trattare dapprima dell'archivio "dei mezzi di comunicazione"? Niente archivio senza lo spazio istituito di un luogo di impressione. All'esterno, direttamente su qualche supporto, attuale o virtuale.» Certo, la riflessione di Derrida è prevalentemente incentrata sull'approccio psicanalitico, ma non mancano le suggestioni per un approccio filosofico agli archivi (che non mancheranno in generale in questo capitolo, p. 8): «Il concetto di archivio accoglie in sé, naturalmente, questo ricordo del nome archè. [...] In qualche modo, il vocabolo rinvia infatti, come si ha ragione di credere, all'archè nel senso fisico, storico o ontologico, cioè all'originario, al primo, al principiale, al primitivo, al cominciamento insomma. Ma ancor di più, e più-tosto, «archivio» rinvia all'archè nel senso nomologico, archè del comando. Come l'archivium o l'archium latino (parola che si usa al singolare, come si fece in un primo tempo per il francese archive che una volta si diceva al singolare e al maschile: un archive), il senso di "archivio", il suo unico senso, gli viene dall'archeion greco: in primo luogo una casa, un domicilio, un indirizzo, la residenza dei magistrati supremi, gli arconti, coloro che comandavano. Ai cittadini che detenevano e significavano quindi il potere politico, si riconosceva il diritto di fare o di rappresentare la legge. Tenuto conto della loro autorità così pubblicamente riconosciuta, è presso di loro, in questo luogo che è la loro casa che [...] hanno il potere di interpretare gli archivi.»

d'archiviazione del WWW, non è probabile che diminuisca." L'archivio diventa una metafora nel contesto dello spazio multimediale? Questo è un appello per archiviare lo stesso termine "archivio" per la descrizione dei processi di conservazione multimediale. L'archeologia digitale, perciò, non è per le generazioni future, ma deve essere realizzata già nel presente. Nell'era della "digitalizzabilità", che è questa, l'opportunità di conservare ogni sorta di informazione si manifesta attraverso un fenomeno paradossale: il cyberspazio non ha memoria.⁶⁴»

Da tempo questo paradosso, non certo in questi termini, viene espresso attraverso la contrapposizione tra due posizioni banali ed estremiste, due slogan: “Internet conserva tutto” contro “Internet non conserva niente”. Di fatto ci sono diversi errori tipici della semplificazione necessaria agli slogan, primo fra tutti il fatto che Internet non è un luogo o un generico “qualcosa” con la capacità di conservare, ma è semplicemente l’infrastruttura globale sulla quale si spostano i dati attraverso il protocollo TCP/IP (perciò sarebbe vero che “Internet non conserva niente”, ma non si intende certo in questo senso); di solito, inoltre, con “Internet” si intende il web, che è uno spazio informativo accessibile attraverso Internet⁶⁵, ma che è solo uno dei possibili servizi accessibili attraverso la rete, e che presenta dei documenti (risorse), ma non è certo costituito da aggregati di conservazione dei documenti in comunicazione tra loro, anzi. Per quanto negli ultimi anni con casi eclatanti e comparsi sulle cronache nazionali e internazionali⁶⁶ si sia cominciato a parlare di “diritto all’oblio”, in realtà la persistenza del materiale digitale non è così forte: in alcuni casi, come in quelli citati in nota, dipende dal fatto che possono essere fatte molte, o moltissime copie, e che attraverso i canali di distribuzione disponibili (come

⁶⁴ (Chun e Keenan 2006, p. 119); TdA.

⁶⁵ Secondo la definizione del W3C: «Il World Wide Web (WWW, o semplicemente Web), è uno spazio informativo in cui gli oggetti di interesse, a cui ci si riferisce come risorse, sono identificati da identificatori globali chiamati *Uniform Resource Identifiers* (URI).»; il web si basa su HTTP, URL e HTML («Architecture of the World Wide Web, Volume One» 2004)

⁶⁶ Giusto per citarne due molto famosi: Amanda Todd e Tiziana Cantone. In entrambi i casi si tratta di materiale a contenuto sessuale, ma non è l’unico caso possibile; è una possibilità che riguarda ogni genere di dati personali e non solo.

YouTube o Facebook) sia possibile che un contenuto abbia una diffusione molto ampia⁶⁷ per un determinato periodo di tempo (solitamente piuttosto breve). Infatti, se è vero che in questi casi di cronaca è legittimo appellarsi al diritto all'oblio, e che il problema diventa estremamente pervasivo e diffuso, questo è possibile per il semplice fatto che deriva da un'enorme diffusione dei materiali: in genere, non ci si limita a guardare ma – specialmente con materiale video o fotografico – si cerca di ottenerne una copia, e la si diffonde. Questo meccanismo vale in sé per buona parte dei fenomeni che siamo abituati a vedere nel web: condivisione di immagini e video di qualsiasi tipo e copia e condivisione attraverso altre applicazioni o social network. Il risultato è di avere moltissime copie, la cui struttura non è organizzata, ma il cui esito è quello di dare l'impressione che un determinato contenuto sul web sia “impossibile” da perdere o cancellare per sempre: ci si potrebbe facilmente convincere – con questo ragionamento – che una volta che un elemento entra in questo spazio informativo sia impossibile che scompaia. Purtroppo però se questo è possibile, se materialmente è possibile avere questa impressione, dipende solo dal fatto che un contenuto con centinaia o migliaia di copie distribuite ha statisticamente buone probabilità di riemergere periodicamente e di poter essere recuperato attraverso diverse fonti. Questo può capitare casualmente o, meglio, può capitare per quei contenuti di consumo che hanno una grande diffusione attraverso queste piattaforme globali ma, se dovesse essere una strategia applicata da chi possiede un'ingente mole documentaria da conservare

⁶⁷ Uno dei canali YouTube che propongono il video di Amanda Todd, forse quello originale (*My story: Struggling, bullying, suicide, self harm* 2012), ha ad oggi (dopo quattro anni e mezzo) oltre dodici milioni di visualizzazioni, a queste vanno aggiunte le visualizzazioni di video copiati su altri canali, su Facebook, le traduzioni in altre lingue: un gran numero di persone, anche se solo per qualche giorno, si sono interessate alla vicenda, e non possiamo escludere che tra esse, una piccola percentuale possiede, o abbia copiato o scambiato, il materiale incriminato: per quanto piccola possa essere questa parte si tratterà comunque di migliaia di persone, cioè di migliaia di copie. Va specificato comunque che l'esempio preso in considerazione è più esplicativo dei meccanismi della diffusione che non di altre tematiche etiche legate alla condivisione di materiale sessualmente esplicito relativo alla sfera privata delle persone: il contesto specifico in cui Amanda Todd è stata vittima di bullismo non era così ampio prima del suicidio; tuttavia la situazione è precipitata nel momento in cui questo materiale è stato pubblicato in rete.

sarebbe economicamente insostenibile: comporterebbe infatti la moltiplicazione della disponibilità (anche economica) di spazio necessario per conservare un documento almeno decine, se non centinaia o migliaia, di volte⁶⁸. Che quest'idea non sia altro che un'impressione sbagliata (quand'anche non ci fosse uno studio approfondito delle tecnologie e dei sistemi che utilizziamo per produrre, scambiare e conservare questi dati) basterebbe a dimostrarlo l'attivismo di Vinton Cerf nell'ottica della sensibilizzazione a questa tematica. Cerf, infatti – considerato “padre di Internet” insieme a Robert Kahn, in quanto ideatori del protocollo TCP/IP, nonché dal 2005 vice presidente e *Chief Internet Evangelist* di Google – da alcuni anni sta cercando di diffondere al maggior numero di persone possibile il problema della conservazione in ambito digitale, facendo ben presente che è un problema (forse) di domani a cui dobbiamo trovare delle soluzioni oggi⁶⁹. Se aziende che lavorano quotidianamente con i dati (solo per ricordare quelle già citate: Cisco, Google, Seagate) si preoccupano già da tempo della conservazione dei dati digitali e del divario tra dati e supporti, dovrebbe almeno sorgere il dubbio, quando non la consapevolezza, che “l'internet”, o il web, o comunque venga definito dai media lo spazio informativo e informatico in cui siamo immersi quotidianamente, non è minimamente autosufficiente a conservare i dati e i contenuti informativi che vengono prodotti. A questo proposito, proprio perché è possibile che qualcosa venga

⁶⁸ Luciano Floridi, nel suo intervento *The forgetful memory of the digital (or why we need libraries more, not less)* (LUMSA, 21 aprile 2016, Roma: *La memoria labile del digitale* 2016) sottolinea, con dati dell'azienda Cisco e dell'università di Berkeley – ma gli stessi dati e le stesse conclusioni vengono riportate anche da un'intervista a Mark Whitby, Senior Vice President, Branded Products Group presso la Seagate Technology, EMEA (Athow 2015) –, quello che viene definito il “*data capacity gap*”, ovvero il limite fisico tra la produzione di contenuti multimediali sul e per il web, e la produzione materiale di dispositivi per la memorizzazione dei dati: il divario tra i dati prodotti e quelli effettivamente memorizzabili su supporto si sta ampliando sempre di più, implicando necessariamente che sempre più dati, nel futuro, saranno necessariamente effimeri, poiché non esiste lo spazio fisico in cui possano esistere per un periodo di tempo che non sia breve o brevissimo.

⁶⁹ Alcuni degli interventi registrati sono disponibili online: (Cerf 2015; *Bit rot (on digital vellum)* | *Vint Cerf* | TEDxRoma 2014; *CMU Distinguished Lecture Series - Vint Cerf* 2015; *Vint Cerf - Digital Vellum* 2015; *Vint Cerf speaks on Digital Preservation at NASA GSFC* 2016); il titolo con cui Cerf presenta di solito i propri interventi è “Digital Vellum”, ovvero quello che potremmo intendere come il corrispettivo della pergamena, cioè resistente e duraturo nel tempo, in ambito digitale.

conservato ma non possiamo controllare cosa, e non è detto che sia ciò che vogliamo conservare, non possiamo pensare che una volta che un contenuto sia entrato in questo spazio informativo, non sia più necessario fare niente perché si mantenga nel tempo. Inoltre, come riflessione più contestuale e meno generale, è opportuno notare come le aziende che stanno largamente investendo in questo senso siano, appunto, aziende private: non c'è nessuna filantropia da parte di queste multinazionali, ma solo l'interesse a tutelare legittimamente il proprio patrimonio e la propria posizione sul mercato. Tuttavia, proprio perché si tratta di aziende private che operano sul mercato, non c'è nessuna garanzia che le scelte operate da esse siano conformi a un qualsiasi standard, e non è necessario che lo siano. In questo senso, però, come ricordato da Floridi nel suo intervento presso la LUMSA⁷⁰, chi controlla quali domande vengano poste (e permette che siano poste solo in certi termini) e quali risposte possano avere queste domande possiede un enorme potere, il potere – in sostanza – di decidere quali siano le informazioni a cui si può avere accesso e quali no, in che modo debba essere organizzato il sapere, e che cosa possa essere conservato nel tempo: un potere che però poserebbe le sue basi (il gigante dai piedi d'argilla) su dati che non possono essere mantenuti e i cui scopi potrebbero oscillare in base a logiche di mercato. È innegabile che in questo panorama sia necessario un intervento più istituzionale, che non giunga però suggerito o posto dall'ambiente politico, ma piuttosto dall'ambiente professionale (principalmente quello degli archivisti e dei bibliotecari) che si è sempre storicamente occupato di organizzare l'informazione e garantirne l'accesso. La strada per adottare degli standard, in ogni contesto, è sempre lunga – e forse siamo già in ritardo – ma il grado di consapevolezza almeno istituzionale, ad oggi, è consistente: la possibilità che un domani questa consapevolezza diventi coscienza e che gli strumenti per garantire la conservazione del nostro presente digitale siano alla portata di tutti non sembra neanche così utopica.

⁷⁰ Floridi: *The forgetful memory of the digital (or why we need libraries more, not less)* (LUMSA, 21 aprile 2016, Roma: *La memoria labile del digitale* 2016)

OAIS conforme oppure no?

Un quesito a cui è particolarmente difficile rispondere, parlando di standard, è se esista uno standard, al di fuori di quella che è la normativa ufficiale (nazionale), per tutto ciò che non viene direttamente indicato da questa, e che quindi rientri sotto di essa.

«Negli ultimi anni sta diventando sempre più incalzante la necessità di affidare i propri archivi digitali ad aziende ed enti che forniscono servizi di archiviazione e conservazione a lungo termine; tuttavia questa operazione presuppone un'implicita fiducia nell'affidabilità del fornitore e di conseguenza comincia ad essere impellente l'individuazione di criteri sulla base dei quali valutare i depositi digitali. Nel panorama italiano esistono norme ben precise per le pubbliche amministrazioni, che devono necessariamente affidare i loro archivi digitali a conservatori che abbiamo ottenuto l'accreditamento da parte dell'Agenzia per l'Italia Digitale (AgID), ma non vi sono regole altrettanto stringenti per le aziende o per le singole persone che possono decidere di consegnare i propri archivi a qualsiasi conservatore digitale, anche non accreditato. Inoltre, nulla vieta di rivolgersi al mercato europeo od internazionale, dove le legislazioni sono differenti. Ecco, quindi, che diventa importante individuare delle modalità condivise a livello internazionale per valutare l'affidabilità di un deposito digitale al fine di “provare la credibilità del deposito e dei suoi contenuti”»⁷¹

Nel suo articolo, Allegrezza, offre una rapida panoramica sui tentativi operati nel corso degli anni per dare forma a linee guida e procedure condivise⁷², passando poi a trattare questo aspetto in riferimento al

⁷¹ (Allegrezza 2015, p. 39)

⁷² *Ibidem*, p. 40 «Nel contempo sono stati avviati numerosi studi e progetti che hanno portato alla definizione di linee-guida, criteri e strumenti necessari per guidare un processo di valutazione dell'affidabilità di un deposito: tra questi vanno certamente ricordati il *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)*, il *Catalogue of Criteria for Trusted Digital Repositories (NESTOR Catalogue)*, il *Repository Audit Method Based on Risk Assessment (DRAMBORA)* rilasciato nel 2007 dal *Digital Curation Centre (DCC)* e *Digital Preservation Europe*

primo passo per una certificazione di affidabilità, il *Data Seal of Approval* (DSA). Come abbiamo già chiarito, il modello OAIS (ISO 14721:2012) è a tutti gli effetti uno standard, ma questo non implica necessariamente che un qualsiasi deposito digitale sia certificato con OAIS, anche perché se è vero che OAIS è a tutti gli effetti uno standard ISO, stabilire con precisione quali siano effettivamente i criteri particolari (desunti nell'applicazione rispetto alle necessità di ogni singolo contesto), derivati dalle indicazioni a carattere universale del modello, non è un'operazione particolarmente semplice in quanto

«tratti fondamentali di OAIS sono la sua flessibilità e il suo livello di astrazione: non vi sono preclusioni di principio sulle modalità concrete con le quali concetti e modelli debbano essere implementati, e non vengono imposti requisiti specifici circa le tecnologie da usare per supportare le implementazioni.»⁷³

Inoltre, come ulteriormente specificato più avanti

«L'importanza della conformità ad OAIS, in ultima analisi, si misura sulla capacità di produrre un impatto tangibile sulla fiducia che le parti interessate ripongono nell'idoneità di un deposito digitale ad assolvere i propri obiettivi di conservazione. Biblioteche, musei e altre istituzioni di conservazione, ad esempio, si trovano di fronte alla prospettiva di affidare segmenti insostituibili del patrimonio scientifico e culturale di cui sono responsabili a sistemi di archiviazione digitale, la cui effettiva capacità di provvedere ad un'efficace gestione a lungo termine

(DPE). Dall'evoluzione dello studio NESTOR è nato lo standard tedesco DIN 31644: [...] si basa su una serie di 34 criteri e le prime certificazioni di conformità a questo standard sono state rilasciate nel 2014. Analogamente, dall'evoluzione della checklist predisposta dal gruppo di lavoro RLG-NARA (RLG-NARA 2007) e di quella predisposta nell'ambito del progetto NESTOR è nato lo standard ISO 16363: Space data and information transfer systems - Audit and certification of trustworthy digital repositories, che costituisce uno dei principali riferimenti in tema di certificazione dei depositi digitali. Ufficialmente riconosciuto standard nel 2012, si pone "l'ambizioso obiettivo di dar vita ad uno standard internazionale riconosciuto" coerente con il modello ISO 14721:2012(OAIS)»

⁷³ (Lavoie 2016)

non è ancora stata comprovata. La conformità ad OAIS può rendere più semplice questa decisione?⁷⁴»

Rispetto a questo interrogativo ci sono diversi indicatori (non solo riferiti ad OAIS) che portano a rispondere in maniera positiva. Una prima parte di questa risposta (quella prevalentemente non riferita ad OAIS) si può trovare suggerita dal già citato articolo di Allegrezza: il DSA si può ottenere come una sorta di autocertificazione controllata da un organismo esterno: il deposito digitale che intenda “certificarsi” deve inviare una richiesta compilando l'*application form* presente sul sito web del DSA; il DSA *Board* rende disponibile l'accesso ad una procedura *online* di autovalutazione (*DSA online tool*); infine la documentazione prodotta viene valutata dal DSA *Board*, che ha il compito di verificare la correttezza delle risposte fornite e la conformità del deposito ai parametri minimi richiesti (il DSA è composto da sedici linee guida, per ciascuna delle quali è necessario allegare la documentazione e il cui livello – previsto su una scala di parametri da 0 a 4 – deve essere almeno di 3 per la metà dei punti e di 4 per i rimanenti). Il DSA, in questo senso, si pone al livello più basso delle possibili certificazioni (o autocertificazioni) e ha un numero piuttosto esiguo di linee guida (appena sedici); tuttavia, il fatto che sia necessario allegare la documentazione relativa ai vari punti e che ci sia un organismo esterno che ne valuta il contenuto, lo rende a un tempo abbastanza flessibile da permettere anche a istituzioni minori di tentare una valutazione di questo tipo (o in caso di capire quali sono le lacune migliorabili nella propria struttura), ma sufficientemente “imparziale” da valutare sulla base di requisiti minimi accettabili in ambito anche internazionale. Secondo il *Memorandum of Understanding to create a European Framework for Audit and Certification of Digital Repositories*⁷⁵ (firmato da David Giarretta, chair di CCSDSI/ISO Repository Audit and Certification Working Group (RAC), Henk Harmsen, chair del Data Seal of Approval (DSA) Board e Christian Keitel, Chair del DIN Working Group "Trustworthy Archives –

⁷⁴ *Ibidem.*

⁷⁵ (Giarretta, Harmsen, e Keitel 2010)

Certification"), la certificazione ha tre livelli: la *basic certification* è costituita dal conseguimento in autovalutazione del DSA; l'*extended certification* rappresenta il secondo livello di certificazione e viene rilasciato ai depositi che, dopo aver ottenuto la *basic certification*, si sottopongono a una valutazione interna secondo le norme ISO 16363 o DIN 31644; la *formal certification* viene rilasciata a quei depositi che, dopo aver ricevuto l'*extended certification*, si sottopongono a una valutazione esterna completa sempre secondo ISO 16363 o DIN 31644 (è da notare come – ad eccezione del livello di complessità, che è certamente ben distribuito – nei tre livelli siano presenti DSA, ISO 16363 e DIN 31644, rappresentati ciascuno da uno dei tre firmatari). Ma cosa comporta, a questo punto la conformità a OAIS? Intanto, va specificato – a livello di standard – che ISO 16363:2012 è un *Magenta Book* del *Consultative Committee for Space Data Systems*⁷⁶. ISO 16363 “*Audit and Certification of Trustworthy Digital Repositories*” deriva dalla valutazione di una serie di possibilità per applicare e quindi essere conformi a OAIS: TRAC⁷⁷, sviluppato da *Research Libraries Group* (RLG) e *US National Archives and Records Administration* (NARA), pubblicato nel 2007. In seguito alla “standardizzazione” del 2012, come ISO 16363:2012, si è ritenuto opportuno elaborare uno standard anche per i requisiti minimi per l’accreditamento di organizzazioni certificatrici di depositi conformi allo standard e per il processo di certificazione: *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories*, requisiti pubblicati prima come *Magenta Book* (652.1-M-2) e poi come ISO 16919:2014.

«La lista di controllo TRAC affronta le difficoltà esaminate nella sezione precedente riguardo alla conformità ad OAIS: come nota

⁷⁶ Anche OAIS (ISO 14721:2012) è un *Magenta Book* del CCSDS: i *Magenta Book* corrispondono alle pratiche raccomandate, ovvero al modo in cui la comunità internazionale suggerisce di utilizzare al meglio determinati (set di) standard in questo momento, o come è meglio che vengano usati per gli obiettivi futuri.

⁷⁷ *Trustworthy Repositories Audit & Certification: Criteria & Checklist* (TRAC). La lista di controllo TRAC rappresenta la miglior pratica e la riflessione più matura sul tema dei requisiti di una infrastruttura tecnica e organizzativa affidabile e certificabile. Il campo di applicazione della lista di controllo investe tre grandi aree: organizzazione e *governance*, gestione di oggetti digitali, aspetti tecnologici.

la stessa lista «le istituzioni hanno iniziato ad autodichiararsi ‘conformi ad OAIS’ per sottolineare l’affidabilità dei propri depositi digitali, ma non esisteva una interpretazione acclarata di cosa fosse la ‘conformità ad OAIS’, al di là del soddisfare le responsabilità di alto livello definite dallo standard». In quest’ottica, la lista di controllo TRAC può essere vista come un possibile modo per definire in termini concreti come debba essere un archivio OAIS-conforme: termini a loro volta basati su criteri ben definiti e misurabili, che possano essere trasposti e applicati in depositi, organizzazioni e sistemi reali. In altre parole, TRAC spacchetta e interpreta ad un livello di maggior dettaglio i requisiti generali di conformità specificati in OAIS (che abbiamo descritto nella sezione precedente). Va comunque notato che TRAC rappresenta solo una possibile interpretazione di questi requisiti di conformità. Dato il livello molto astratto dei requisiti di conformità OAIS, è senz’altro possibile che un ulteriore sforzo per standardizzare – in modo OAIS-conforme – la certificazione dei depositi possa pervenire ad un insieme molto diverso di criteri specifici, che potrebbe nondimeno trovare posto nel perimetro degli stessi larghi criteri di conformità sui quali TRAC è basato. Tutte queste diverse interpretazioni sarebbero comunque ancorate allo stesso fondamento concettuale (OAIS), e, prese in combinazione, potrebbero anche rivelarsi complementari tra di loro. Tuttavia, esse differirebbero nel modo in cui i concetti originari di OAIS sarebbero declinati nelle caratteristiche verificabili dei sistemi di archiviazione reali e concreti.»⁷⁸

In conclusione, è opportuno evidenziare la tendenza a muoversi in direzione di alcuni standard che possano garantire affidabilità, verificabilità e certificabilità dei processi per la conservazione e delle strutture che la attuano: nonostante lo sforzo collettivo di assicurare un futuro ai nostri dati attuali, non si può ancora ritenere che la garanzia di un esito positivo possa essere rilasciata da alcuna organizzazione.

⁷⁸ (Lavoie 2016)

Quale approccio per gli archivi d'autore?

Nel complesso abbiamo già avuto modo di vedere come questa tipologia di archivi ricada sostanzialmente in una zona grigia, anche a guardarla da più profili: da un punto di vista archivistico puro e semplice, in quanto materiale per così dire “corrente” (ma di privati) non è un compito proprio degli archivi di Stato (e inoltre si tratta di documenti che ricadrebbero all'esterno dalle procedure amministrative digitalizzate), e generalmente le fondazioni hanno i propri archivi, oltre al fatto che nascono per atto pubblico o testamentario (e istituire fondazioni non sembra essere un'abitudine degli scrittori viventi, giovani o meno giovani, così frequente da ritenerla statisticamente rilevante); da un punto di vista del documento informatico, non essendo un documento della pubblica amministrazione, non è tenuto ad entrare – ma non è nemmeno previsto che lo faccia – nella filiera della conservazione normata dal punto di vista legislativo per quanto riguarda la documentazione informatica; non essendo intervenuta la dichiarazione d'interesse storico – e con buona probabilità, in alcuni casi, essendo ben lungi dall'intervenire⁷⁹ – non rientra nemmeno sotto la tutela dei beni culturali. Infine, questo ci dovrebbe far pensare, siamo in una fase di transizione per cui da qui a cinquant'anni è possibile che molti dei sistemi tecnici e delle leggi in materia vengano modificati anche di molto e che, per esempio, un autore che pubblicasse un primo libro, da quel momento dovrebbe per legge utilizzare una piattaforma in *cloud computing* per tutti i propri documenti (con funzioni di *versioning* automatiche come già offre Google Docs, e con molte altre funzionalità che ora mi sarebbe impossibile anche solo immaginare); oppure si troverà un modo per ottimizzare i supporti di memoria e di dati di tutti saranno integralmente conservati (fatto decisamente poco probabile, ma non si sa mai). Ma gli autori

⁷⁹ Basti pensare, come caso recente, a Giacomo Mazzariol (nato il 9 gennaio 1997), autore (con il supporto di Fabio Geda) per Einaudi di *Mio fratello rincorre i dinosauri*: prima che intervenga la dichiarazione d'interesse storico per l'archivio di un ragazzo di vent'anni, ci sono buone probabilità che il ragazzo cambi almeno una decina di supporti; in questo caso, se tra alcuni anni ci fosse interesse a recuperare quei documenti, sarebbe forse più facile recuperare i file di questo primo libro da Fabio Geda.

contemporanei, che hanno abbracciato per comodità o per altri motivi la scrittura in digitale⁸⁰, in questa fase di transizione, sono ancora di più in una zona grigia, perché lo è il periodo che stiamo vivendo, e proprio loro, sotto il profilo della conservazione, sono i più deboli e i più a rischio: quelli che non lavorano più (o lavorano molto poco) in cartaceo, e che ancora non conoscono o non si preoccupano per la conservazione dei propri documenti digitali.

La transizione, però, è per tutti: per il legislatore che con difficoltà – e a volte scarso approfondimento del settore – cerca di normare una situazione che si è imposta *de facto*, ma a cui non era preparato; per l'autore, che produrrà una parte del proprio archivio in cartaceo e una parte invece in digitale (cercando di ottimizzare questa separazione non per forza intenzionale); per il conservatore che deve fare i conti con un ambiente per nulla amichevole: è più lungimirante – forse – Carlo Federici quando dice che

«Si conserva (e conseguentemente si restaura) soltanto la materia dei beni culturali. Da questo assioma si deduce che non è possibile conservare (e quindi restaurare) un "oggetto" (perdonate l'ossimoro) immateriale. Del resto è evidente che nessuno conserva il digitale – così come tradizionalmente si conservano i beni culturali – giacché la sua conservazione consiste, in buona sostanza, nella sua continua copiatura (castamente definita "refreshing"). Diversi anni fa, nel corso di un convegno su questi temi, dissi coram populo a Mariella Guercio che, se davvero voleva conservare i contenuti del digitale, il mio consiglio poteva essere solo quello di stamparli su carta.»⁸¹

Tuttavia, nonostante mi sia trovato personalmente a condividere quest'affermazione di Federici (basti pensare ai numerosi link con

⁸⁰ Solo per fare un esempio a me noto, un autore come Stefano Benni scrive le bozze con carta e penna, correggendosi direttamente sul foglio; la stesura definitiva, o quella che ritiene possa essere tale, la batte a macchina (perché per non buttare via un intero foglio battuto a macchina per errori o ripensamenti è portato a pesare e scegliere al meglio le parole) e soltanto alla fine prepara il testo per la stampa al computer: in questo caso ci sarà meno materiale da destinare alla conservazione rispetto a chi lavora al computer dalle prime fasi fino alla fine.

⁸¹ (Guerrini 2016, slide 11)

errore 404 trovati, sui siti di progetti per la conservazione digitale⁸²), questo è un fatto a cui nessun conservatore può opporsi: può non condividere questo cambiamento di paradigma, ma non può invertire la tendenza; deve invece cercare di conservare al meglio questi documenti, senza necessariamente riporre eccessive o acritiche speranze nel digitale, ma lavorando al meglio per permettere a questi documenti di durare il più possibile, e cercando, se il futuro ne porgerà il destro, di trovare soluzioni alternative e migliori a queste che, per il momento, sono l'approccio più solido – e ancora non danno garanzie – alla conservazione del nostro presente digitale.

Come, cosa e perché conservare?

In base a quello che ci siamo detti finora, ci si può chiedere se sia allora il caso di affrontare le problematiche tecniche (che richiedono competenze di informatica e informatica forense, ambiti per i quali non è generalmente prevista una formazione – se non molto di base – per bibliotecari e archivisti), giuridiche (che richiederebbero un rapporto piuttosto disinvolto con la normativa di settore, ma anche con quella sul diritto d'autore e non solo), e di settore (mancano degli strumenti testati e che diano la garanzia di un risultato certo, e anche gli standard, come abbiamo visto, non hanno ancora un livello di avanzamento tale – ma comunque in linea con l'avanzamento tecnico – da essere prescrittivi a livello di procedura) per ottenere non si sa bene quale risultato. La risposta però non può essere che affermativa: se lasciamo che questi documenti vengano conservati dai servizi privati di *storage* e *backup*, non si avranno (potrebbero esserci, ma non è obbligatorio) standard e buone pratiche condivisi per la gestione di tutte le fasi della conservazione del documento, dall'acquisizione alla migrazione, o all'emulazione e la copia; inoltre non sapremo con certezza – perché

⁸² Non solo: in (Weston 2013) una parte consistente delle note rimandano a documenti digitali, in formato PDF o HTML che ad oggi non sono più reperibili nemmeno cercando solo il titolo del contributo nella speranza di trovarlo, copiato o in un altro formato, altrove; in questo caso solo la bibliografia cartacea sarà sicuramente reperibile e garantirà completezza o la possibilità di approfondire il testo del volume, per le citazioni digitali invece non c'è alcuna garanzia, e per questa tesi sarà lo stesso.

dipende dai contratti di ciascun servizio – per quanto tempo verranno conservati i documenti (senza contare che l’accesso ai documenti è legato agli account personali, che costituiscono una cospicua maggioranza rispetto alle cartelle condivise o ad accesso pubblico). Pertanto il conservatore dovrà farsi carico di tutte queste competenze a livello individuale? In realtà, riprendendo quando già detto nell’introduzione, per quanto sarebbe opportuno che il singolo avesse, a prescindere, il maggior numero di competenze possibili, difficilmente una persona sola potrebbe avere una competenza effettivamente approfondita in tutti questi ambiti contemporaneamente, e non è il caso di mutare così radicalmente gli insegnamenti attuali per adattarli a queste esigenze: non c’è nessuna intenzione di creare il “superconservatore”, che conoscerebbe a livello molto approfondito l’informatica forense, alcuni aspetti del diritto e le risorse tecniche e concettuali per attuare queste procedure: si ritroverebbe a lavorare al fianco di persone con cui non potrebbe condividere un sapere e un saper fare comuni⁸³. L’auspicio, invece, è – come per Conservare il Novecento – di coinvolgere attorno allo stesso tavolo le competenze specifiche di ciascuno degli specialisti del settore per creare (con l’unione delle rispettive competenze) una soluzione che integri al meglio le soluzioni per le problematiche specifiche di ogni ambito (e vedremo nel capitolo seguente che è così che lavora il progetto PAD). Facendosi forza di queste competenze collettive bisognerà considerare che è opportuno intraprendere un percorso verso la conformità agli standard, e che cerchi di integrare (per quanto piuttosto liberamente concesso – come abbiamo avuto modo di notare) le necessità operative e quotidiane con le indicazioni concettuali suggerite dai modelli: sarà pertanto molto opportuno considerare in prima battuta di poter garantire una continuità temporale al progetto (sostenibilità economica nel tempo) e, secondariamente, di guardare al modello OAIS se non

⁸³ Lungi dall’escludere a priori una soluzione di integrazione delle competenze, inoltre nulla esclude, come formazione aggiuntiva, che si possano approfondire questi percorsi separatamente o insieme, come nel caso del DigCCurr (*Digital Curation Curriculum*: <https://ils.unc.edu/digecurr/>) che si propone l’obiettivo di fornire un curriculum di studi adeguato per chi voglia intraprendere il percorso della Digital Curation; questo non esclude affatto che sarebbero sempre opportune delle solide basi in comune con i conservatori più “tradizionali”.

per una immediata o completa certificazione, almeno come un tracciato di linee guida a cui fare riferimento. Ricordandoci che si tratterà di lavorare con gli scrittori viventi, e che i dispositivi utilizzati per il lavoro non saranno necessariamente utilizzati *solo* per il lavoro, bisognerà anche valutare con cura quali siano i documenti da destinare alla conservazione: se in alcuni casi potrà essere ovvio cosa includere, talvolta non sarà sempre altrettanto facile capire cosa invece escludere, poiché è estremamente difficile indovinare su quale base verranno valutati l'interesse o l'utilità di un documento a distanza di anni⁸⁴.

Certamente verranno destinati alla conservazione tutti i documenti preparatori per le opere dell'autore, non sempre così per eventuale corrispondenza, o per i documenti "copiati" da terzi: in linea di massima l'ultima volontà da rispettare è quella dell'autore che, affiancato possibilmente da un archivista, deciderà cosa è opportuno conferire; il contributo di affiancamento dell'archivista è piuttosto rilevante poiché permette di mediare tra le intenzioni soggettive dell'autore e quelle invece più specifiche del conservatore (che ha di norma un quadro generale più chiaro su quali siano i documenti più adatti al conferimento e quali invece siano superflui). Ma, soprattutto, bisognerà anche riflettere attentamente sull'altro versante della mediazione da parte del conservatore: il pubblico a cui si rivolge (in lessico OAI: *comunità di riferimento*). Infatti è importante che i dati acquisiti per la conservazione possano essere accessibili⁸⁵, oltre che conservati, e che lo siano in modo ottimale per chi andrà a utilizzarli: sarà opportuno descrivere in maniera standard l'archivio dell'autore (pur consci che si tratta di un archivio corrente per il quale

⁸⁴ Roncaglia ricorda come in fase di scarto presso la biblioteca del Senato della Repubblica, siano stati eliminati alcuni documenti che a distanza di anni sarebbe invece stato utile avere ancora a disposizione (*LUMSA, 21 aprile 2016, Roma: La memoria labile del digitale* 2016).

⁸⁵ Nel già citato articolo di Allegrezza a pp. 44-45 si ricorda che «*Le DSA Guidelines si basano su cinque principi fondamentali, che, nel loro insieme, determinano se i dati versati nel deposito digitale possono essere considerati come archiviati in maniera "sostenibile": i dati devono essere resi disponibili su Internet [tondo mio]; i dati devono essere accessibili tenendo conto della normativa in materia di tutela della privacy e del diritto d'autore; i dati devono essere disponibili in un formato utilizzabile; i dati devono essere affidabili; deve essere possibile creare riferimenti ai dati (ad esempio, mediante l'utilizzo di persistent identifier)*».

l'accrescimento non è prevedibile in maniera regolare secondo serie già stabilite), e si avrà il tempo – ad ogni acquisizione – per farlo? Oppure non si descriverà affatto (se non con una semplice copia automatica della struttura dell'archivio attraverso *filename* e *directory*)? Oppure ancora bisognerà trovare una soluzione che sia più rapida e richieda meno risorse di una descrizione completa dell'archivio, ma che al contempo offra una strutturazione più ragionata e accurata che il semplice utilizzo del percorso del file?

Metadati tecnici, amministrativi, di conservazione

Fondamentali non solo alla descrizione, ma all'acquisizione e alla gestione stessa dei file sono i metadati, che possono essere di varie tipologie; principalmente i metadati⁸⁶ possono essere:

- **Amministrativi:** metadati usati per la gestione e l'amministrazione delle collezioni e delle risorse informative (ad esempio: informazioni di valutazione e acquisizione; *tracking* dei diritti e delle riproduzioni; documentazione sui requisiti e i protocolli legali, culturali e per l'accesso da parte della comunità; informazioni sulla localizzazione; criteri di selezione per la digitalizzazione);
- per la **Conservazione:** metadati relativi alla gestione della conservazione per le collezioni o altre risorse collegate (ad esempio: documentazione sulle condizioni fisiche delle risorse; documentazione delle azioni intraprese per la conservazione delle versioni analogiche e digitali delle risorse come *refreshing*, migrazione, emulazione, ...; documentazione di ogni mutamento avvenuto durante la digitalizzazione o le operazioni per la conservazione);
- **Descrittivi:** metadati utilizzati per descrivere, autenticare e identificare le collezioni e altre risorse informative affidabili collegate (ad esempio: metadati generati dal creatore e dal sistema originali; SIP: *Submission Information Package*; record

⁸⁶ (Baca 2016)

catalografici; strumenti di ricerca; controllo di versione; relazioni tra le risorse; indici specializzati; informazioni di curatela; descrizioni, annotazioni e correzioni fatte dai creatori o da altri utenti);

- **Tecnici:** metadati relativi a come funziona il sistema o al comportamento dei metadati (ad esempio: documentazione *hardware* e *software*; informazioni procedurali generate dal sistema, come metadati sugli eventi e sui processi; informazioni tecniche sulla digitalizzazione, come formati e livelli di compressione; tracciamento dei tempi di risposta del sistema; dati inerenti l'autenticazione e la sicurezza, come chiavi crittografiche, password);
- **d'Utilizzo:** metadati relativi al livello e al tipo d'uso delle collezioni o delle risorse informative (ad esempio: record di circolazione; record inerenti le mostre fisiche e virtuali; uso e *tracking* dell'uso da parte degli utenti; informazioni sul riuso dei contenuti e sul *multiversioning*; *logs* di ricerca; metadati sui diritti, come IPR, *copyright*, ...)

Principalmente, i metadati che autonomamente un archivio digitale può fornire o, per dirla meglio, i metadati che possono essere estratti dai file contenuti senza che ne siano forniti di espliciti, sono i metadati tecnici: i metadati tecnici sono – al di là dei tracciati definiti come standard⁸⁷ o meno – quelle informazioni che sono generate e quindi memorizzate sul file che riguardano aspetti apparentemente meno utili per l'utente finale (come il *software* e il sistema operativo che hanno generato il file, nonché tutta una serie di altre informazioni tecniche, appunto, sul file in oggetto), ma che sono invece fondamentali per il conservatore in quanto permettono di capire quali *software* utilizzare per accedere al file in questione e quali strumenti utilizzare per eventuali migrazioni o emulazioni, e quindi come presentare e come permettere all'utente di accedere a quei documenti. Altri metadati a cui viene dato un particolare rilievo (particolarmente in OAIS) sono i metadati per la conservazione:

⁸⁷ Cfr: (Dappers e Lenders 2010)

«OAIS descrive un ampio insieme di requisiti per i metadati necessari per supportare le attività di un archivio di tipo OAIS. Questi requisiti hanno avuto una profonda influenza sul successivo sviluppo di un certo numero di schemi di metadati per la conservazione. I metadati per la conservazione sono metadati che supportano il processo di conservazione digitale a lungo termine, e comprendono informazioni sulla provenienza, sui diritti di proprietà intellettuale e sugli ambienti tecnici e interpretativi di un oggetto digitale archiviato. In questo senso, la finalità dei metadati per la conservazione si sovrappone notevolmente con i tipi di metadati definiti nel modello informativo OAIS e quindi OAIS è un punto di partenza naturale e un riferimento obbligato nello sforzo per sviluppare schemi di metadati per la conservazione. Sebbene le iniziative riguardanti i metadati per la conservazione siano state numerose, una – PREMIS – è emersa come lo standard de facto. [...] Sebbene l'ambito del Data Dictionary di PREMIS non si sovrapponga esattamente con i requisiti per i metadati descritti nel modello informativo OAIS (per esempio, il concetto OAIS di Informazioni descrittive non è contenuto in PREMIS), i concetti chiave di OAIS di Informazione sulla rappresentazione e di Informazioni sulla conservazione sono fondamentali per PREMIS, come già lo erano stati per la struttura di metadati per la conservazione su cui PREMIS si è basato. PREMIS può davvero essere visto come una implementazione di questi concetti, anche se non necessariamente copre l'intera gamma di informazioni che si prevede che questi debbano comprendere.

Il punto chiave è che il modello informativo OAIS funge da fondamento per ciò che attualmente è lo standard prevalente per i metadati per la conservazione. Nello stesso modo in cui il modello funzionale OAIS ha esercitato una profonda influenza sullo sviluppo dei sistemi e dei servizi di deposito di conservazione digitale, il modello informativo OAIS ha

influenzato la progettazione degli oggetti informativi che i depositi devono gestire.»⁸⁸

All'interno del manuale del già citato progetto PARADIGM – che, ricordiamo, non ha prodotto una documentazione sugli esiti, ma il cui risultato più documentato per ora, e molto approfonditamente documentato, è il manuale *Workbook on Digital Private Papers* – si dà un riscontro piuttosto ampio alla metadattazione dei documenti acquisiti: il capitolo relativo ai metadati amministrativi e per la conservazione è il più ampio di tutto il manuale, e affronta nel dettaglio sia l'aspetto concettuale (e in riferimento al modello OAIS) sia quello tecnico, offrendo parecchi esempi pratici di come si sono utilizzati determinati formati. Nella fattispecie andremo a vedere alcuni degli esempi prodotti in questo contesto poiché, mancando un vero e proprio standard sulle procedure, si tratta – probabilmente – della serie di esempi e possibilità più ampia con cui possiamo confrontarci. Innanzitutto la riflessione parte dalla necessità di capire quali sono gli elementi che vanno identificati all'interno del conferimento⁸⁹:

<p>The archive The archive, as a collection, must have an identity so that it can be associated with objects belonging to it using their identifiers. <i>archive:1</i> <i>archive:2</i></p>	<p>An accession In order that digital manuscripts can be associated with an accession, an accession id is needed. <i>archive:1</i> <i>accession:1</i> <i>accession:2</i></p>
<p>A digital manuscript (the intellectual entity) This is the conceptual item that might be described in a catalogue, and will be given a traditional identifier, such as a reference code or shelfmark, when it is catalogued. An example is: <i>archive:1</i> <i>accession:1</i> <i>object:1 – Website of Politician X, 20 Jul. 2006 (also known as MS. Eng. 23)</i></p>	<p>A representation A representation is a particular instance of an intellectual entity, so MS. Eng. 23 above could have two representations as follows: <i>archive:1</i> <i>accession:1</i> <i>object:1 [MS. Eng. 23]</i> <i>representation:1 – Website of Politician X, 20 Jul. 2006 (html and jpeg files)</i> <i>representation:2 – Website of Politician X, 20 Jul. 2006 (PDF capture of website)</i></p>
<p>file A representation is composed of one or more files, for example: <i>archive:1</i> <i>accession:1</i> <i>object:1 [MS. Eng. 23]</i> <i>representation:1 – Website of Politician X, 20 Jul. 2006</i> <i>file:2 – html file</i> <i>file:3 – jpeg file</i> <i>representation:2 – Website of Politician X, 20 Jul. 2006</i> <i>file:5 – pdf file</i></p>	

⁸⁸ (Lavoie 2016); per PREMIS: <http://www.loc.gov/standards/premis/>

⁸⁹ Tabella tratta da (Bodleian Library, John Rylands University Library of Manchester, e Joint Information Systems Committee 2007, p. 54)

Un ampio contributo a questa schematizzazione deriva dall'applicazione del PREMIS *Data Dictionary*; ne consegue quindi la necessità di individuare e identificare singolarmente ogni atomo informativo, e ricomporli attraverso i collegamenti possibili tra metadati e oggetti. Non è questa la sede per trattare diffusamente i dettagli propriamente tecnici dei possibili identificatori (come URI, URL, URN, ARK, DOI, PURL, Handle o Fedora Identifier, ciascuno con i propri vantaggi e i propri limiti), né i dettagli di PREMIS o di altri tracciati per i metadati, ma è importante piuttosto soffermarsi brevemente su alcuni punti e su alcuni casi, per capire quali siano gli obiettivi primari dietro questi schemi di metadati, e come – rispetto alla formulazione teorica – vengano impiegati nella pratica.

«Un esempio largamente citato è il Metadata Encoding and Transmission Standard (METS), un formato per documenti basato su XML che supporta la codifica, o l'impacchettamento, di una grande varietà di metadati associati con gli oggetti digitali archiviati. Come fa notare Mc Donough, c'era il desiderio, da parte di METS, di agevolare lo scambio e l'interoperabilità di oggetti presenti in biblioteche digitali tra sistemi di biblioteche digitali e di fornire al tempo stesso da supporto per la conservazione a lungo termine di oggetti presenti in biblioteche digitali, fungendo da potenziale Pacchetto di versamento, Pacchetto di archiviazione e Pacchetto di distribuzione all'interno del modello di riferimento OAIS. METS non è certo l'unico modo di codificare pacchetti di archiviazione OAIS. Ad esempio, Baglt è una specifica di impacchettamento di file sviluppata congiuntamente dalla Library of Congress e dalla California Digital Library. Il sistema di conservazione digitale Archivematica usa Baglt per creare AIP. Altri schemi di codifica sono disponibili, come Fedora Object XML (FOXML), sebbene molti di questi non siano esplicitamente collegati all'implementazione del concetto di AIP di OAIS. [...] Un documento METS per un oggetto digitale comprende:

- una intestazione auto-descrittiva, metadati descrittivi relativi all'oggetto, metadati amministrativi (in particolare, metadati tecnici, metadati relativi ai diritti, metadati sulla fonte analogica dell'oggetto digitale, e metadati sulla provenienza) [METSHeader; Descriptive Metadata Section (dmdSec); Administrative Metadata Section (amdSec)];
- una lista dei file che costituiscono l'oggetto digitale (la mappa strutturale di tutte le componenti dell'oggetto digitale) [file section (fileSec); structural map (structMap)];
- una lista dei collegamenti che stabiliscono relazioni tra le componenti della mappa strutturale [structLink];
- una lista di "comportamenti" che possono essere associati con l'oggetto digitale. [behaviourSec]

I metadati possono essere codificati internamente al documento METS o essere collegati a partire da una fonte esterna, come un registro. Un documento METS serve come struttura per impacchettare insieme gli oggetti digitali archiviati e i metadati a loro associati. In questo senso, può essere visto come una implementazione del concetto di pacchetto informativo OAIS. [...] METS combina un oggetto digitale e i suoi metadati in un singolo pacchetto logico, che può essere versato come una unità nel deposito (cioè, come Pacchetto di versamento), gestito come una unità nel sistema del deposito (cioè come Pacchetto di archiviazione), o estratto e disseminato come una unità verso parti esterne al deposito, come la Comunità di riferimento o altri depositi (cioè come Pacchetto di distribuzione). Svolgendo queste funzioni, un documento METS si comporta come le informazioni sull'impacchettamento definite nel modello OAIS: informazione che serve a combinare, fisicamente o logicamente, tutte le componenti di un pacchetto informativo, permettendo la loro identificazione, localizzazione e gestione come singola unità logica.»^{90 e 91}

⁹⁰ (Lavoie 2016)

<mets>	
<mets Hdr>	<i>metadata about the mets file</i>
<dmdSec>	<i>descriptive metadata</i>
<amdSec>	<i>administrative metadata</i>
<techMD>	<i>technical metadata</i>
<rightsMD>	<i>rights metadata</i>
<sourceMD>	<i>for digitised material: metadata about the original</i>
<digiprovMD>	<i>digital provenance metadata</i>
<fileSec>	<i>a file inventory, which can refer to or embed digital files</i>
<structMap>	<i>a structural map to record a hierarchy of digital files</i>
<structLink>	<i>a mechanism for linking between divisions of the structural map</i>
<behaviourSec>	<i>behaviour metadata; can be used to associate programs with files</i>

Anche in PARADIGM – che stiamo osservando più da vicino – si è optato per utilizzare il formato METS come contenitore per PREMIS e gli altri metadati previsti, a seconda della tipologia di file⁹²; proprio

⁹¹ L'immagine della struttura di un file METS è ripresa da (Bodleian Library, John Rylands University Library of Manchester, e Joint Information Systems Committee 2007, p. 118)

⁹² *Ibidem*, pp. 111-113. Nello specifico i vari standard individuati sono stati per le immagini (*still images*): MIX for Images (Z39.87), Harvard Digital Repository Service Image, National Library of Australia Image, Digital Images Archiving Study, Exchangeable image file format (Exif); per i video (*moving images*): National Library of Australia Video, Library of Congress VideoMD Data Dictionary, MPEG-7; per i file audio: Harvard Digital repository service Audio, Library of Congress Audio (Source) Data Dictionary, National Library of Australia Audio, MPEG-7, ID3; per i file di testo: Schema for Technical Metadata for Text by Jerome McDonough,

per la natura granulare non tanto delle informazioni in sé (in quanto ciascuna informazione, ciascun dato è atomico nel momento in cui viene scorporato dal proprio contesto) quanto piuttosto dell'approccio concettuale adottato, come si è potuto notare dalla tabella precedente, ma come anche consente – e in qualche modo incoraggia – METS. In PARADIGM si è provveduto a ridurre alla più piccola parte ciascun componente complesso e rendere esplicite le relazioni (tra oggetti e metadati e tra gli oggetti tra di loro) attraverso un solido utilizzo dei metadati. Le possibili opzioni valutate per l'integrazione in METS di tutti i metadati sono state l'approccio a struttura composita o a struttura atomizzata, che hanno ciascuna dei vantaggi e degli svantaggi⁹³.

Approccio a struttura composita: è possibile utilizzare un unico file METS per contenere tutti i metadati relativi a un archivio strutturandoli in questo modo (o analogo):

- un record MODS⁹⁴ per ogni oggetto in *dmdSec*, basato sui metadati descrittivi ricavati automaticamente, in modo da permettere all'archivio di stabilire un primo e basilare controllo intellettuale sull'archivio;
- un record PREMIS per ogni oggetto in *amdSec/techMD* e nella stessa sezione un record per ogni oggetto con i metadati specifici rispetto alla tipologia (*MIX*, *textMD*, ...);
- un record PREMIS relativo ai diritti in *amdSec/rightsMD*;
- un record PREMIS per ogni evento in *amdSec/digiprovMD*;
- un inventario in *fileSec* con regole *fileGrp* per le entità intellettuali, la loro rappresentazione e i file;
- in *structMap* il dettaglio dell'ordine originale delle entità intellettuali nell'archivio, e delle sottosezioni che diano la struttura dettagliata della rappresentazione di ogni entità intellettuale.

Il vantaggio di questo approccio è che richiede la creazione di un unico file METS per descrivere l'intero archivio, tuttavia è evidente

National Library of Australia Text; per i database: National Library of Australia Database; per gli eseguibili: National Library of Australia Executables.

⁹³ *Ibidem*

⁹⁴ *Metadata Object Description Schema* <http://www.loc.gov/standards/mods/>

come questo file sia molto grande e complesso, causando grandi difficoltà a lavorarci, e comunque considerando che ogni minima operazione (soprattutto quelle automatiche) va pianificata, controllata e consolidata nella sua efficacia prima di poterla attivare con sicurezza.

Approccio a struttura atomizzata: si divide l'archivio in tanti piccoli file METS e si sfruttano i meccanismi di collegamento presenti in METS e in PREMIS per collegare tra loro i file. Questo approccio, a seconda di come si intenda collegare e rappresentare tra loro i vari elementi, può essere declinato in differenti modalità:

come prima possibilità possiamo considerare di creare dei file METS separati (e quindi poi collegarli) per:

- le entità intellettuali;
- la rappresentazione delle entità intellettuali;
- i file;
- gli agenti;
- gli eventi e
- i diritti.

Come seconda possibilità possiamo invece creare file METS separati per:

- le entità intellettuali, che includono al loro interno i metadati per la rappresentazione delle entità intellettuali e i file necessari;
- gli agenti;
- gli eventi e
- i diritti.

Ancora, si possono creare file METS separati per ciascuna entità intellettuale, le quali conterranno individualmente tutti i metadati necessari per:

- la rappresentazione delle entità intellettuali;
- i file;
- gli agenti;
- gli eventi e
- i diritti.

Infine, come ultima possibilità si possono creare file METS separati per:

- le entità intellettuali;
- la rappresentazione delle entità intellettuali che conterrà al proprio interno i metadati per i file;
- gli agenti;
- gli eventi e
- i diritti.

L'approccio di atomizzare i contenuti e i file METS ha l'indubbio vantaggio di mantenere ridotte le dimensioni dei file; inoltre la struttura interna dei file sarà decisamente meno complessa, agevolando il lavoro, e permetterà, all'occasione, di collegare la stessa unità informativa a più elementi senza dover duplicare informazioni. Al contempo però (considerando che le informazioni relative a un oggetto, per quanto ridotte ai minimi termini e collegate tra loro non diventano meno complesse nel loro insieme: si può spostare di livello la complessità, ma non evitarla) per fornire un AIP completo potrebbe essere necessario richiedere un gran numero di file METS, e in ogni caso le procedure vanno sempre accuratamente pianificate, controllate e valutate nella loro efficacia prima di renderle operative.

Metadati descrittivi?

Finora abbiamo avuto modo di analizzare abbastanza da vicino le varie possibilità per produrre, organizzare e gestire i metadati relativi alla conservazione e agli aspetti tecnici e amministrativi, quelli che in un linguaggio informatico potremmo chiamare i dati di *backend*, quelli necessari al conservatore per tutte le procedure e le operazioni di mantenimento dei file, ma non quelli di *frontend*, visibili e necessari all'utente per comprendere, consultare e avere accesso ai documenti presentati – se ci rifacciamo al primo principio del DSA, per cui *i dati devono essere resi disponibili su Internet*⁹⁵; se invece i dati sono destinati esclusivamente ad uso e consumo interno potrebbe essere

⁹⁵ (Allegrezza 2015, p. 45)

sufficiente acquisire i metadati per la gestione interna: ma questo risulterebbe in un approccio volutamente non conforme con OAIS. Se queste informazioni sono inevitabilmente necessarie, e ci si può anche interrogare – come abbiamo fatto – sulle modalità tecniche con cui possono essere fornite o acquisite, per quanto riguarda i metadati descrittivi sarà più importante attivare una riflessione su cosa si vuole o è opportuno fornire piuttosto che sugli strumenti tecnici per farlo. Relativamente alla descrizione archivistica ci sono una serie di standard internazionali relativi a vari aspetti come la descrizione degli archivi (ISAD(G)), dei record d'autorità (ISAAR(CPF)), dei soggetti conservatori (ISDIAH) e dei soggetti produttori (ISDF); inoltre esistono le applicazioni informatiche di ISAD(G) e ISAAR(CPF) che sono rispettivamente EAD ed EAC-CPF: questi sono gli standard principali in ambito internazionale per la descrizione degli archivi “tradizionali”, destinati agli archivisti che quotidianamente si impegnano nell'ordinamento di un fondo. Il lavoro archivistico però, oltre a essere esercitato principalmente su materiali analogici, è un lavoro impegnativo e che richiede tempi piuttosto lunghi per essere portato a termine: questo potrebbe essere anche parzialmente accettabile o – almeno – gestibile in un contesto analogico ma, se invece lo rapportiamo in un contesto digitale con l'acquisizione di decine o centinaia di migliaia di file per ogni conferimento, formerebbe un “collo di bottiglia” in cui tutte le operazioni seguenti non potrebbero procedere a causa della descrizione in corso. L'alternativa più rapida sarebbe evitare la fase scientifica della descrizione, dare un'indicazione generica del tipo «*Archivio dell'autore Nome Cognome, dal gg/mm/aaaa al gg/mm/aaaa*», dando come indicazione di base la struttura delle *directory* e lasciando a un momento successivo (se ci sarà tempo) la descrizione più scientifica o alla volontà e all'abilità dell'utente la possibilità di muoversi con queste scarse informazioni. Nel nostro caso specifico è forse più opportuno orientarsi a un sistema che sia di mediazione tra la descrizione archivistica vera e propria e nessun'altra descrizione oltre alla struttura logica, poiché non si potrà guardare alla pubblica amministrazione (che è l'unico ambito per cui esistano delle prescrizioni in questo senso) per

la quale il problema non si pone in quanto esistono già procedure e serie predefinite. Trattandosi di scrittori, e certo anche di intellettuali e autori in genere, un modo abbastanza rapido, e al contempo sufficientemente efficace per descrivere l'archivio è quello di collegare i documenti alle opere con cui sono in relazione, sia come documenti preparatori che come bozze e stesure, e ordinare i documenti rimanenti per tema generale (miscellanea, corrispondenza con..., corrispondenze – in generale – idee per nuove pubblicazioni, ...) come se si trattasse di opere a loro volta. Utilizzando un sistema di questo tipo, o comunque analogo a questo, si possono fornire le informazioni necessarie all'utente per valutare e consultare il documento in rapporto al suo contesto, senza dover ricorrere a una impegnativa descrizione scientifica, ma comunque garantendo un ordinamento sia per *directory* che per “titoli” tale da permettere un accesso e una gestione dei contenuti informativi adeguata alle esigenze della comunità di riferimento. Naturalmente è necessario ricordare come questa modalità non costituisca né possa costituire un sostituto alla descrizione scientifica, e che non è da intendersi se non come il suggerimento di una possibilità tra le tante da valutare nel contesto.

...e oltre?

In conclusione rispetto a quanto ci siamo detti finora, prima di trattare in maniera approfondita il progetto PAD – Pavia Archivi Digitali, che è il primo esempio in Italia di un archivio istituito allo scopo di conservare i documenti (nativi digitali) degli scrittori (per così dire, pure loro) nativi digitali⁹⁶, accennerò piuttosto brevemente ad alcune di quelle che – ad oggi – sembrano poter essere delle innovazioni tecnologiche capaci di offrire un servizio migliore nel trattamento dei dati, ovvero alcune applicazioni degli strumenti offerti dal web

⁹⁶ Ed ecco qui spiegata la voluta ambiguità del titolo: se naturalmente ci si occupa dei *born digital archives*, si parla anche di quegli autori che stanno affrontando questa transizione dall'analogico al digitale e che quindi sono i soggetti produttori di archivi ibridi (in un qualsiasi rapporto tra le due forme) e, specialmente, di quegli autori che – in quanto, essi stessi, “nativi digitali” – non lasceranno nessun archivio cartaceo ai posteri, poiché tutto il loro lavoro è svolto in ambiente digitale.

semantico su queste tipologie di progetti. C'è da premettere subito che le evidenze scientifiche di questi strumenti, tra cui i *Linked (Open) Data* (LOD), sono ancora tutte da dimostrare, o per meglio dire: non facciamoci prendere da entusiasmi eccessivi. In realtà ci sono alcune applicazioni che utilizzano LOD e che funzionano anche molto bene⁹⁷, se si sa come utilizzarle. In estrema sintesi i LOD sono uno strumento che permette di inserire dati (ridotti a livello granulare) all'interno dello spazio informativo del web consentendo, non solo collegarli ad altri dati (il meccanismo del *link* è sempre esistito nel web), ma anche di far capire ai computer quale tipo di relazione collega tra loro dei dati o quale pattern di relazioni può ripetersi tra gruppi di dati, in modo che un ampio insieme di dati possa essere processato in modo automatico estraendo i dati richiesti (e talvolta, attraverso la possibilità di inserire pattern di relazione, consentendo anche la scoperta di relazioni fino a quel momento non esplicitate). Concettualmente, si potrebbe pensare che non si tratti di una novità così rilevante rispetto a come potevano essere trattati precedentemente i dati con altri sistemi, tuttavia bisogna notare come questo sistema permetta di esprimere le relazioni in un modo standard per tutti, e di conseguenza estrarre informazioni (potenzialmente) da tutto l'insieme di dati presenti sul web. La questione più rilevante – al di là del fatto che possa essere o meno una moda – è che nell'ambito dei beni culturali, in genere, non si capisce ancora molto bene quali vantaggi darebbero⁹⁸ rispetto ad altri strumenti che sono già disponibili: per fare un esempio molto banale quando una persona effettua una ricerca nel catalogo di una biblioteca, viene restituita una scheda catalografica, che è l'unità informativa minima, in quanto contiene tutti i dati necessari a identificare inequivocabilmente un determinato esemplare di un determinato titolo. Nel caso dei LOD,

⁹⁷ Dbpedia (<http://wiki.dbpedia.org/>), nato come versione semantica di Wikipedia, a cui ha fatto seguito Wikidata (<https://www.wikidata.org/>); Linked Jazz (<https://linkedjazz.org/>), il progetto semantico della Bibliothèque nationale de France (<http://data.bnf.fr/>), solo per citarne qualcuno.

⁹⁸ Anche perché sono relativamente ancora poco diffusi o mediamente non indirizzati all'utente medio: se metto a disposizione i miei dati attraverso un *endpoint* SPARQL, questo potrà essere interrogato solo da chi ha un minimo di dimestichezza con il linguaggio SPARQL e potrà capire la documentazione relativa alla struttura dei dati per poter porre le *query* in modo appropriato.

l'unità informativa minima è ciascuno dei singoli elementi che vengono riaggregati in un secondo momento a comporre l'unità della scheda: ci si potrebbe giustamente domandare quale vantaggio si possa ottenere dall'investire una buona quantità di risorse per convertire il catalogo, se si possono ottenere risultati analoghi con altri strumenti (come i *discovery tool* che non fanno necessariamente uso dei LOD, pur offrendo funzionalità aggiuntive) e senza dover modificare in maniera così radicale il catalogo. Proprio a questo proposito, in una discussione avuta con Sara Di Giorgio (ICCU), si evidenziava come – ad oggi – l'assenza di applicazioni che possano giustificare, con esiti decisivi, la scelta di utilizzare i LOD, rimanga uno dei principali deterrenti alla loro applicazione. Se però ci sono alcuni inevitabili dubbi su queste tecnologie, ci sono anche alcune esperienze positive, o alcune applicazioni, che ne dimostrano alcuni usi stimolanti; al di là del lavoro molto articolato e complesso della Bibliothèque Nationale de France con il progetto *data.bnf.fr*, che non è evidentemente alla portata di tutti, rispetto alle necessità che abbiamo relativamente agli archivi d'autore che stiamo analizzando, possono esserci alcuni strumenti che, se non direttamente applicabili, senz'altro possono dare alcuni spunti almeno di riflessione su come vorremmo poter utilizzare i nostri dati nel futuro e cosa vorremmo che facessero i nostri dati per noi.

Un primo esempio è quello di *Social Networks and Archival Context* (SNAC) dell'Institute for Advanced Technology in the Humanities (University of Virginia; IATH): attraverso i dati forniti da numerose istituzioni⁹⁹ il progetto ha come obiettivi tre nuove forme di accesso alla documentazione: accesso distribuito alle fonti primarie e secondarie riguardo enti, persone e famiglie; accesso alle descrizioni storiche e biografiche; accesso ai contesti sociali in cui le persone sono vissute e hanno operato. Lo scopo perseguito da questo progetto, pertanto, è quello di creare – a partire da descrizioni archivistiche già esistenti –

⁹⁹ Secondo quanto riportato dal sito: «190,000 finding aids contributed by scores of repositories, 2.2 million MARC21 descriptions contributed by OCLC WorldCat, and approximately 500,000 original descriptions of people contributed by the British Library, the U.S. National Archives and Records Administration, the Smithsonian Institution Archives, and others» («SNAC: Social Networks and Archival Context» 2010)

un'interfaccia (parzialmente grafica: sul sito è disponibile un prototipo¹⁰⁰) che consenta di evidenziare le relazioni e i rapporti esistenti sia tra persone e/o personalità che tra i singoli individui e il contesto sociale a loro contemporaneo: l'idea, trasportando il progetto nel nostro contesto, di costruire questi *social network* per gli autori contemporanei è sicuramente accattivante, in quanto, per esempio, potrebbe permettere un'estrema accuratezza; tuttavia bisognerebbe che oltre al consenso e alla disponibilità da parte degli autori a fornire alcune informazioni, si considerasse in che modo creare e fornire agli utenti questi dati: se aggregati al conferimento all'interno del sistema di conservazione (allora sarebbe necessaria una descrizione archivistica scientifica in fase di acquisizione), separatamente (allora si potrebbe considerare di affidare a terzi la gestione di questi dati?), o in quale altra forma e con quali strumenti.

Un altro spunto interessante proviene invece dal progetto di un gruppo di ricerca dell'Hasso Plattner Institut (HPI) di Potsdam in collaborazione con Yovisto¹⁰¹, un motore di ricerca semantico per materiale video a contenuti accademici: il risultato di questo lavoro è *refer*¹⁰², un plugin (ancora in versione beta, per il momento) per il CMS Wordpress, il cui uso è dimostrato in un blog ospitato da Yovisto¹⁰³. Refer, in breve è costituito da tre componenti principali che ne regolano il funzionamento:¹⁰⁴ un'applicazione di *Named Entity Linking* (NEL), un "visualizzatore di *infobox*" e un "*browser* relazionale e di raccomandazione". Con NEL si intende un sistema in grado di riconoscere il carattere delle parole di un testo (se cioè sono sostantivi, nomi propri, articoli, aggettivi, avverbi, ...) e, con il supporto di una base di dati a cui è collegato, di suggerire quali delle entità contenute all'interno della base di dati possano essere collegate a una determinata parola; questa parte di *refer* è quella attiva nel *backend* del sito, e pertanto non è visibile dall'esterno. L'applicazione di NEL

¹⁰⁰ <http://socialarchive.iath.virginia.edu/snac/search>

¹⁰¹ Tabea Tietz, Jörg Waitelonis e Harald Sack dell'HPI e Joscha Jäger di Yovisto (<http://www.yovisto.com/>)

¹⁰² <https://refer.cx/>; (Tietz et al. 2016)

¹⁰³ <http://blog.yovisto.com/>

¹⁰⁴ (Tietz et al. 2016)

analizza (a richiesta) il testo scritto in un post da pubblicare e lo confronta con i dati di DBpedia, suggerendo per quali delle parole riconosciute si sono trovati dei risultati che combaciano o, in alternativa, è possibile selezionare una o più parole del testo e chiedere la ricerca di una corrispondenza con un'entità; una volta confermate le annotazioni desiderate, si pubblicherà l'articolo normalmente. Dal *frontend*, invece, è possibile vedere il comportamento degli altri due componenti: spostando il mouse sopra le parole annotate semanticamente compare un *infobox*, una piccola tabella contenente i dati principali dell'entità (nazionalità, date e luoghi di nascita e morte per le persone; e altre indicazioni rilevanti come definizioni per termini e sostantivi, o i riferimenti geografici per i luoghi, etc; non avendo avuto la possibilità di utilizzarlo – ma guardando come vengono presentati i dati – forse è possibile scegliere quali sono le informazioni che vengono presentate nell'*infobox*). Se invece di passare solamente il mouse sopra la parola si clicca l'entità annotata compare il “*browser* relazionale e di raccomandazione”, un'interfaccia grafica che permette di navigare attraverso le relazioni: la parte superiore divisa in quattro colonne presenta da sinistra a destra persone, luoghi, concetti, date, mentre la parte inferiore presenta nel box a sinistra gli articoli che hanno più elementi in comune con l'entità selezionata (motore di raccomandazione) e nella fascia rimanente la descrizione sommaria dell'entità selezionata; passando il mouse sopra una qualsiasi delle entità presentate vengono evidenziate le relazioni con le altre entità visibili, mentre se si clicca sopra una qualsiasi delle entità, viene ricaricato in browser con persone, luoghi, concetti, date, articoli collegati e breve descrizione della nuova entità selezionata. Avere uno strumento come *refèr*, nel caso degli archivi nativi digitali d'autore, non sarebbe utile tanto per la gestione interna, quanto piuttosto per la comunicazione all'esterno dell'attività del soggetto conservatore, per la possibilità di interessare un numero maggiore di utenti e – di grande importanza –, per la possibilità di fornire fonti autoritative (che non sono così numerose) nella vasta rete dei LOD.

Naturalmente queste rimangono soltanto delle suggestioni: di primaria importanza è la conservazione degli archivi, e questi strumenti non contribuiscono in maniera significativa alla conservazione, ma piuttosto alla fruizione dei dati una volta che siano stati prodotti; se abbiamo aperto questo spiraglio è per offrire un panorama che sia il più completo e aperto possibile, e per mostrare cosa si sta provando a fare con queste tecnologie¹⁰⁵.

¹⁰⁵Bisogna anche ricordare, in aggiunta che gli stessi L(O)D sono dati digitali per cui sussiste il rischio di perdita se non adeguatamente conservati: per questo motivo (e abbastanza per tempo rispetto al resto dei documenti digitali, considerando che i *Linked Data* sono una tecnologia piuttosto recente) sono nati progetti come PRELIDA: *PREserving LInked DAta* <http://www.prelida.eu/>.

PAD - PAVIA ARCHIVI DIGITALI¹⁰⁶

«Pavia Archivi Digitali è nato nel 2009 grazie a un'idea di Beppe Severgnini (allora presidente dell'Associazione Alunni dell'Università di Pavia) che propose la costituzione di un archivio di memorie digitali di scrittori, giornalisti e intellettuali del nostro tempo. PAD è nato dunque per raccogliere, conservare a lungo termine e mettere a disposizione degli studiosi materiali digitali di diverso tipo: stesure di romanzi, appunti, bozze, inediti, prove di copertina, fotografie, stesure di articoli etc., nella consapevolezza tecnica e culturale che nelle elaborazioni delle fase scritte del testo e di conseguenza nella relativa archiviazione dei documenti a uso personale e professionale, l'imponente diffusione degli strumenti di videoscrittura ha portato via via, dalla metà degli anni Ottanta in avanti, e con rapidità sempre crescente, al graduale se non in alcuni casi definitivo abbandono del tradizionale supporto cartaceo. Un primo nucleo di lavoro dell'Università di Pavia nel 2010 ha iniziato a elaborare un progetto sostenibile che potesse durare nel tempo. Beppe Severgnini ha subito conferito oltre

¹⁰⁶ Le informazioni utilizzate per la redazione di questo capitolo hanno prevalentemente fonti non ufficiali o non ufficialmente pubblicate: alcuni dei materiali sono quelli messi a disposizione sul sito di PAD – Pavia Archivi Digitali, nella sezione relativa ai seminari tecnici (<http://pad.unipv.it/content/seminari-tecnici>); nello specifico a questo indirizzo si trovano le *slides* dei contributi di Fabio Ciotti e Salvatore Vassallo, e di Emmanuela Carbé e Primo Baldini). Un grande ringraziamento va al dottor Primo Baldini e alla dottoressa Emmanuela Carbé, per il dialogo instauratosi sul profilo tecnico e strutturale di PAD, per le conversazioni e i confronti, e per tutti i materiali informativi che hanno condiviso con grande disponibilità, e che sono stati ampiamente utilizzati per illustrare le varie fasi. Un altro ringraziamento per i contributi su PAD, infine, va al professor Weston coordinatore del comitato scientifico di PAD.

sedecimila file e sponsorizzato il progetto presso la sua casa editrice, che ha aderito all'iniziativa: da qui il passo alla donazione di altri autori Rizzoli è stato molto breve; PAD ha infatti ottenuto l'adesione di Silvia Avallone e Gianrico Carofiglio. PAD ha potuto contare su importanti finanziamenti della Fondazione Alma Mater, sulla sponsorizzazione della banca UBI (che ha messo a disposizione un caveau per il deposito di DVD con una copia originale dei conferimenti) e, grazie a Beppe Severgnini, di Nokia e Vodafone. Nel 2015 e 2016 altri autori hanno affidato il loro archivio digitale all'ateneo pavese: Francesco Pecoraro, Paolo Di Paolo e Franco Buffoni.»¹⁰⁷

PAD, avviato quindi dal 2009, nel corso di circa sette anni di attività – ad oggi – ha potuto intraprendere e valutare alcuni approcci e strategie per la conservazione digitale, testando le soluzioni più idonee rispetto alle esigenze pianificate e, progressivamente, individuando le problematiche degli strumenti utilizzati, talvolta correggendoli e talvolta cambiando approccio a vantaggio della funzionalità della struttura nel suo complesso.

Dopo la proposta – accolta – fatta da Severgnini nel 2009, si è avviato quindi uno studio di fattibilità, tra il 2010 e il 2011, che ha portato ad evidenziare le necessità specifiche sia per tipologia di progetto (conservazione a lungo termine di documenti digitali, e comunque con un “approccio archivistico”) che per tipologia di soggetti produttori (scrittori, giornalisti e intellettuali).

Le necessità principali emerse da questo studio sono state:

- Architettura semplice e scalabile: è necessario che, una volta creata la struttura del sistema, questa possa adeguarsi a successivi ampliamenti (più relativi alla quantità di dati che alla struttura stessa) e implementazioni senza ulteriori radicali modifiche;

¹⁰⁷ Dallo schema delle procedure del progetto PAD, che era temporaneamente disponibile sul sito (link controllato l'ultima volta il 10 ottobre 2016) all'indirizzo http://pad.unipv.it/PAD_lista_sw.pdf realizzato da Baldini e Carbé: come vedremo più avanti l'evoluzione del sistema e le modifiche che lo implementano costantemente non sempre permettono che la documentazione sia aggiornata “all'ultima versione” del sistema: come verrà ricordato anche più avanti non si dà il progetto come un risultato ma come un processo, e in quanto tale è previsto che possa evolvere nel tempo.

- Archiviazione veloce di grandi quantità di dati: autori particolarmente prolifici o che comunque producono un'ingente mole documentaria per vari motivi professionali (e Severgnini è un caso emblematico in questo senso) riverseranno probabilmente archivi molto ampi, dalle decine alle centinaia di migliaia di file: è importante che la fase di acquisizione del materiale non crei un collo di bottiglia rispetto al flusso di lavoro di tutti gli altri elementi del sistema (e che parimenti non lo creino altri elementi del processo);
- Archiviazione in assenza dei metadati identificativi¹⁰⁸: come già evidenziato in precedenza a più riprese, in genere i produttori di materiale digitale non sono i “soggetti conservatori ideali”, può capitare che i metadati identificativi non siano consistenti (*filename* irrelati con il contenuto, creazione di cartelle o file vuoti il cui nome servirebbe come nota o promemoria, ...) o che non siano affatto presenti; rispetto a questa situazione è opportuno avere degli strumenti che permettano la gestione dei file, occupandosi solo in un secondo momento dei metadati identificativi;
- Conversione di documenti in modalità non interattiva (*batch*);
- Estrazione automatica dei metadati, ripetibile: è necessario avere a disposizione differenti metadati, per i quali può essere necessario ricorrere a librerie o *tool* di estrazione diversi; inoltre, può essere necessario estrarre i metadati in più fasi dell'acquisizione;
- Visualizzazione della struttura del *filesystem* dei conferimenti: in qualche modo è la struttura stessa dell'archivio conferito, o almeno una sua parte sostanziale, è importante che possa essere ricostruita in quanto rispetta l'ordinamento originario (che acquista un senso rispetto al lavoro dell'autore);

¹⁰⁸ I metadati presi in considerazione per ogni file sono solo quattro: *filename*, *path*, dimensione del file e *checksum*: in un qualsiasi altro deposito istituzionale sarebbe necessario inserire numerosi metadati identificativi. Con questi quattro elementi PAD riesce ad identificare in modo univoco ciascun file, in fase di acquisizione, per integrare in seguito le informazioni con altri dati.

- Archivio accessibile in base alle restrizioni del contratto: è fondamentale che il sistema possa garantire le restrizioni previste nel contratto, secondo la volontà dell'autore;
- Gestione delle occorrenze (come nel caso di file duplicati, che assumono valenze diverse in base alla loro posizione): i documenti duplicati (avranno lo stesso *checksum*) in posizioni differenti avranno significato diverso, e perciò è opportuno conservare tanto la struttura originaria dell'archivio quanto il significato che questa esprime;
- Gestione di più conferimenti di uno stesso autore: è auspicabile, che uno stesso autore possa conferire a più riprese nel corso della propria attività (benché sia anche opportuno non confondere il conferimento del proprio archivio con un servizio di *storage, backup* o di *data management*)
- Sostenibilità economica nel tempo del progetto.

Si è valutato quindi, tra le soluzioni commerciali o *open source* disponibili, il sistema più adatto a rispondere a tutte queste esigenze insieme: sono stati presi in considerazione alcuni tra i sistemi più diffusi e utilizzati a livello globale come: Fedora, Islandora, Xena, Dspace, Archivematica, ArchivesSpace, Rosetta, ExLbiris e Phaidra (nell'ipotesi che in quanto diffusi e utilizzati siano mantenuti nel tempo e possibilmente sufficientemente flessibile e implementabili, almeno in un futuro); e sono state contattate alcune aziende (IBM, Alfresco, EMC²) per verificare la possibilità di sviluppare soluzioni per questo progetto e progetti con esigenze analoghe. Rispetto alle necessità espresse i vari sistemi valutati sono risultati non del tutto soddisfacenti o funzionali e il contatto con le aziende non ha portato all'elaborazione di una soluzione concreta; la questione più rilevante, non tanto per la creazione, ma per la sopravvivenza del progetto stesso riguarda infatti la sua sostenibilità economica, aspetto per cui sono state fatte approfondite ricerche in merito sui sistemi disponibili.

La sostenibilità di PAD

Nei vari progetti che sono stati esaminati da PAD (ma è una pratica piuttosto diffusa), viene sempre omesso il profilo economico, e si concentra tutta l'attenzione sugli aspetti tecnici, i metadati, gli standard, i tracciati e gli strumenti tecnologici a cui ci si affida: in questo modo, finché ci sono i finanziamenti, tutto funziona (o almeno così sembrerebbe) poi, una volta che i fondi non siano più stanziati o vengano a mancare per un qualsiasi motivo, questi progetti non riescono più a funzionare e s'interrompono (talvolta senza nemmeno lasciare traccia della documentazione prodotta). Per PAD, il punto cardine per realizzare una struttura tecnologica di questo tipo è stato – prima di qualsiasi altra valutazione, prima ancora di considerare un qualsiasi standard – produrre un sistema capace di funzionare con finanziamenti minimi o quasi nulli in quanto progetti di ricerca estremamente importanti a livello internazionale – come ad esempio CASPAR¹⁰⁹ – sono stati finanziati, istituzionalmente, con cifre importanti¹¹⁰: ad oggi di CASPAR non rimane nemmeno il sito (e la discrepanza tra l'indirizzo del sito e il risultato ottenuto è quantomeno bizzarra: <http://www.casparpreserves.eu/>).

Tra i progetti valutati e confrontati per capire quali fossero i sistemi più sostenibili nel tempo gli strumenti più diffusi sono risultati il sistema di Fedora Commons, Greenstone, Archivemática, LOCKSS, alcune soluzioni di Alfresco, e altri minori. Il sistema di Greenstone, per esempio, rilascia un aggiornamento ogni tre anni circa (2006, 2009, 2013, 2016) senza implementazioni, e il manuale è fermo al 2006; Archivemática è una soluzione commerciale che riesce a sostenersi economicamente per il gran numero di contatti con istituzioni e università e per l'integrazione con i sistemi più diffusi; LOCKSS, grazie

¹⁰⁹ *Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval*: maggiori informazioni sugli obiettivi del progetto sono disponibili all'indirizzo <http://cordis.europa.eu/docs/projects/cnect/2/033572/080/publishing/readmore/CASPAR-booklet.pdf>.

¹¹⁰ Costi totali del progetto dal 2006 al 2009: 15.078.563,00 €; fondi elargiti dal CORDIS (*Community Research and Development Information Service*): 8.799.996,00 €. Proprio a causa della mancanza di un sito mantenuto ancora attivo almeno per la documentazione del progetto, i dati sono stati recuperati direttamente dal sito del CORDIS http://cordis.europa.eu/project/rcn/92920_en.html

a fondi ingenti, per quanto semplice e “ristretto” – non fa altro che creare un ecosistema internamente distribuito in *peer-to-peer* – riesce a mantenersi in maniera piuttosto solida; Alfresco è un’azienda che propone soluzioni tecnologiche a vari livelli, e perciò l’ambito della conservazione del digitale costituisce solo una parte dei ricavi complessivi. La soluzione più promettente – e più diffusa – sembra essere il *repository open source* di Fedora Commons, ma è vero che ha un costo di mantenimento del progetto che lo rende impraticabile (non da adottare, ma da implementare seguendone il modello): secondo il prospetto finanziario pubblicato nel 2015 relativo agli anni 2013 e 2014¹¹¹ le spese di mantenimento del progetto sono state di oltre tre milioni e mezzo di dollari per entrambi gli anni (precisamente 3.592.577 \$; di cui 1.788.790 \$ per il 2013 e 1.885.991 \$ per il 2014), dei quali ben più della metà destinati al pagamento degli stipendi degli oltre 40 dipendenti (trenta sviluppatori e nove tecnici)¹¹². È chiaro che, per quanto il progetto possa essere ben realizzato e diffuso, i costi per mantenerlo attivo e aggiornato nel tempo (sono stati fatti 4 aggiornamenti solo nel 2016) non possono essere sostenuti da una qualsiasi istituzione: non si può cioè guardare al modello di Fedora per realizzare il proprio *repository* e pensare di mantenerlo nel tempo, principalmente perché la necessità di un *team* molto ampio di sviluppo ha dei costi piuttosto impegnativi; il gran numero di progetti (o sottoprogetti interni a Fedora) che ne fanno utilizzo in tutto il mondo possono però giustificare questa spesa con la grande diffusione del prodotto.

Due tra i progetti pratici a cui si è guardato durante lo studio di fattibilità, che presentano due approcci molto diversi sono quello della

¹¹¹ Reperibile all’indirizzo <http://www.duraspace.org/sites/duraspace.org/files/media/FEDORA%20COMMONS%202014.pdf>

¹¹² Le spese per i dipendenti le ho calcolate sommando il salario, le tasse e i benefit per il personale: 1.041.607 \$ nel 2013 pari al 58,23% delle spese che salgono al 65,7% (1.175.122 \$) se contiamo anche 133.515 \$ per i viaggi; mentre per il 2014 sono ben il 67,29% (1.269.016 \$) che se aggiungiamo i 154.022 \$ dei viaggi salgono al 75,45% (1.423.038 \$): i tre quarti della spesa complessiva; rispetto al totale dei due anni sono il 64,32% (2.310.623 \$) o il 72,3% (2.598.160 \$) se vogliamo considerare anche le spese di viaggio.

St. Andrews University, in Scozia, e della *Digital Preservation Software Platform* (DPSP) realizzata dagli archivi nazionali australiani.

Sono entrambi progetti *open source*, il primo realizzato quasi interamente in casa mentre il secondo è una “collezione” di importanti progetti realizzata per gestire le varie problematiche di conservazione.

Per DPSP il progetto è stato definito e sviluppato ma poi tutto si è fermato, perché l’implementazione serviva a risolvere un problema della struttura. Il vantaggio, però, è che una volta implementato, il sistema ha spese molto basse per la manutenzione *software* (e costi molto alti – forse milioni – per l’*hardware*; ma questo presumibilmente è stato previsto).

Il sistema implementato dalla St.Andrews invece ha dei costi enormi di manutenzione: solo tenere aggiornati i vari moduli (per problemi di sicurezza) ha dei costi esorbitanti e probabilmente è necessario l’intervento di aziende esterne.

Tutte queste considerazioni sono servite a PAD per progettare un sistema di preservazione estremamente semplice da utilizzare (a titolo d’esempio, nella fase di *ingest* l’unico dato richiesto è il codice del conferimento), a prova di errore (merito delle aree e della separazione dei database) ma anche orientato a gestire funzionalità innovative (Data Mining, Machine Learning); l’unica controindicazione è che questo progetto serve solo a gestire PAD: poiché è stato realizzato sulle necessità specifiche di questo progetto, non si propone per essere idoneo a servire bene in altri contesti.

Quindi si è deciso di sviluppare “in casa” un sistema di *software* basati su un’architettura semplice, scalabile, implementabile e interoperabile con altri sistemi. Il sistema sviluppato è realizzato interamente in Java, per quanto riguarda l’architettura; le applicazioni per la consultazione invece sono state sviluppate con *FileMaker*. Si è discusso a proposito dei motivi della scelta di un *software* proprietario rispetto allo sviluppo con *software open source* (OS) e se ne sono valutati gli aspetti positivi e negativi: l’implementazione di una soluzione OS richiederebbe lunghi tempi di sviluppo, lo sviluppo dell’interoperabilità tra sistemi e dispositivi differenti, un *team* di sviluppatori che seguano il progetto per lungo tempo, e la possibilità che un piccolo cambiamento richieda

lunghi tempi di aggiustamento; per contro *FileMaker* gestisce in automatico i problemi relativi all'interoperabilità tra sistemi, permette di effettuare modifiche praticamente in tempo reale ed è economicamente molto più sostenibile (si tratta di poche centinaia di euro di licenza); infine, se in futuro fosse disponibile una soluzione OS che offra le stesse possibilità di questo *software*, sarebbe sufficiente una migrazione dei dati per utilizzare la soluzione OS. Perciò il sistema è stato realizzato in modo da essere facilmente modificabile e implementabile:

- prevedendo la separazione dei fondi per una sicurezza maggiore tra i conferimenti dei vari autori (ogni autore ha un fondo proprio, ogni fondo ha credenziali d'accesso differenti – in modo che un operatore non possa avere accesso a fondi riservati – e ciascun fondo è separato individualmente da tutti gli altri in modo che sia garantita la sicurezza – se ci sono problemi in un fondo si può disattivare singolarmente – e che i documenti rimangano separati);
- garantendo un controllo rigoroso dei permessi e delle fasi lavorative, in modo da sapere sempre chi ha fatto cosa;
- garantendo l'interazione con il sistema di controllo qualità;
- permettendo agli applicativi di trasferire dati da un'area all'altra.

Inoltre si è anche guardato, oltre alle architetture tecniche, alle soluzioni adottate da altri progetti di conservazione di documenti digitali per capire quali fossero i risultati raggiunti e quali gli obiettivi da raggiungere; tra le esperienze attivate precedentemente ci sono:

- il già citato progetto PARADIGM (*Personal ARchives Accessible in DIGital Media*), frutto della collaborazione tra le università di Oxford e Manchester, gli esiti del cui lavoro sono stati raccolti all'interno del *Workbook on Digital Private Papers*¹¹³
- gli archivi di Salman Rushdie¹¹⁴ conservati presso la Woodruff Library della Emory University, e il relativo progetto dei born

¹¹³ <http://www.paradigm.ac.uk/workbook/index.html>

¹¹⁴ (Carroll et al. 2011)

digital archives; la soluzione adottata per l'archivio digitale di Rushdie è stata l'emulazione delle macchine conferite attraverso alcuni computer resi accessibili nella sede della biblioteca;

- il progetto *Demystifying Born Digital*¹¹⁵ realizzato da OCLC, volto a supportare l'attività di archivisti e bibliotecari che si trovano a dover affrontare la conservazione del materiale digitale;
- *When Not All Papers are Paper: A Case Study in Digital Archivy*¹¹⁶: un caso di studio sull'acquisizione dell'archivio ibrido di Michael Joyce da parte dell'Harry Ransom Center presso l'Università di Austin, in Texas.

Gli esiti di questi e altri progetti – per il momento – come già accennato, non hanno però saputo fornire le risposte adeguate alle necessità di PAD: consideriamo infatti che PARADIGM non ha prodotto un *report* dei risultati, ma un manuale di procedure e scelte tecniche (molto ben approfondito nelle tematiche che tratta), che l'archivio di Salman Rushdie è consultabile solo in locale e solo in emulazione delle macchine originali, e che parecchi progetti internazionali destinati alla conservazione delle risorse native digitali non hanno più un sito attivo, o il sito è mantenuto ma i collegamenti no, pur essendo terminati in tempi recenti (CASPAR, SCIDIP-ES¹¹⁷). Possiamo certamente andare a vedere in questi progetti quali siano state le strutture tecnologiche utilizzate (e per quali motivi o necessità), ma non potranno ancora offrirci spunti per valutarne l'efficacia nel tempo e nei confronti del pubblico. Ammettere che questi progetti, anche ad altissimo livello, non possano ancora dare risposte certe e risultati sicuri è un punto con cui bisogna necessariamente confrontarsi: per chi oggi cominciasse ad affacciarsi al panorama – anche internazionale – della

¹¹⁵ Il progetto ha prodotto alcuni materiali informativi e *report* che sono disponibili sulla pagina dedicata al progetto sul sito di OCLC: <http://www.oclc.org/research/themes/research-collections/borndigital.html> (Erway e OCLC Research 2012a; Erway e OCLC Research 2012b; Barrera-Gomez, Erway, e OCLC Research 2013; Erway et al. 2014; Erway et al. 2016)

¹¹⁶ (Stollar Peters 2006)

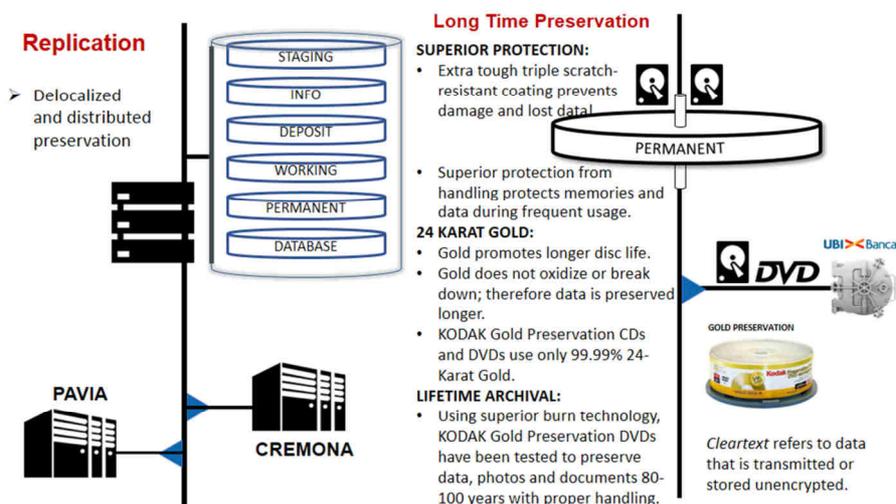
¹¹⁷ <http://www.casparpreserves.eu/> (non più disponibile); <http://www.scidip-es.eu/> (i materiali messi a disposizione hanno link morti); <http://www.planets-project.eu/>

conservazione dei materiali digitali è possibile guardare ad alcuni di questi progetti, la più parte terminati anche da diversi anni, che offrono metodi e strategie elaborate spesso sugli obiettivi da conseguire (comparabili tra loro, ma che necessariamente rispondono a determinate esigenze *locali*) ma non ancora risultati su cui poter fare affidamento.

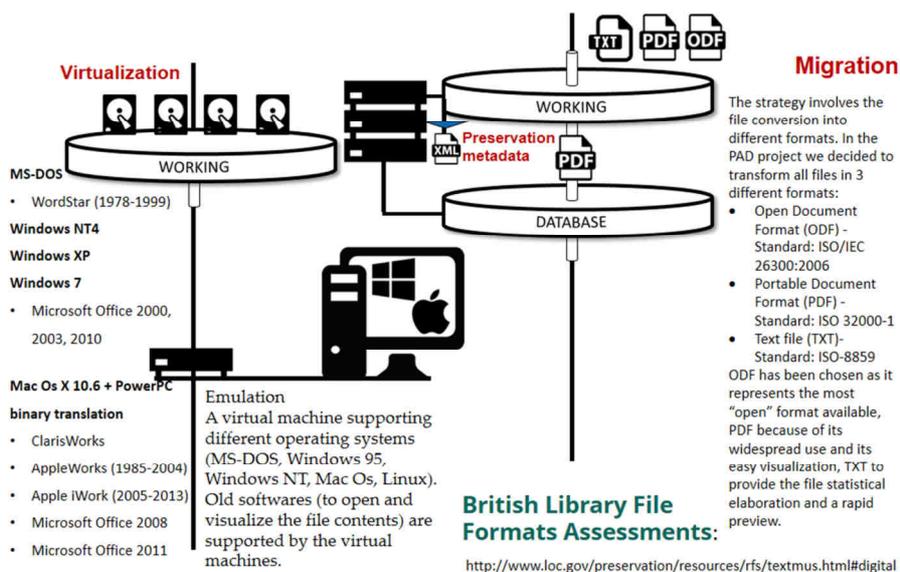
Per ogni archivio c'è la necessità di gestire i file conferiti (ovvero l'archivio vero e proprio dell'autore), le conversioni dei file e tutta la documentazione digitale e digitalizzata (contratto con l'università, questionario informativo, informazioni sul conferimento, set di metadati). A oggi non esiste un metodo univoco e certo per la conservazione a lungo termine di materiali digitali, ed è dunque necessario attuare strategie diversificate come il *refreshing* dei supporti, la duplicazione dei documenti (vedi il progetto LOCKSS di Stanford), l'emulazione virtuale dell'ambiente nativo (vedi il caso di Salman Rushdie), la migrazione a formati aperti, più recenti o mirati alla conservazione. PAD sfrutta queste strategie, talvolta combinate tra loro, per aumentare le probabilità di successo: continua comunque a migliorare e aggiornare le proprie strategie con l'evolversi delle nuove tecnologie e prevede futuri trasferimenti di dati su eventuali nuovi supporti. Per quanto riguarda la sola conservazione una copia dell'archivio è conservata in forma criptata, dunque non accessibile liberamente, su un server primario del centro di calcolo dell'Università di Pavia.

«La struttura dell'hardware utilizzata ricorre alla tecnologia RAID (Redunant Array of Independent Disk), che duplica le attività del disco fisso su più dischi garantendo così la ridondanza dei dati nel caso di errore su uno di essi: il disaster recovery è garantito da un secondo server, localizzato a Cremona (sede distaccata dell'Università di Pavia), che mantiene una copia costantemente aggiornata dei dati conservati sul server primario. Una copia dell'archivio viene masterizzata su un DVD archival grade con

triplo strato metallico in oro, posto in busta sigillata in vuoto e conservato presso il caveau di una banca.»¹¹⁸

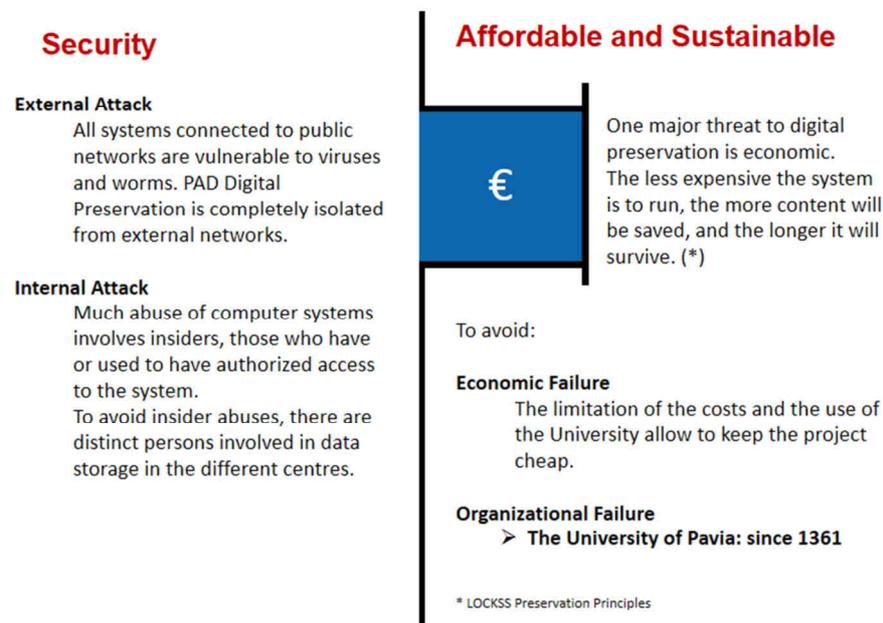


Oltre a ciò, per garantire l'accesso al documento nel suo ambiente originale si ricorre a tecniche di virtualizzazione ed emulazione e alla migrazione, ovvero la conversione, in formati più adatti alla conservazione: l'emulazione supporta un largo numero di sistemi operativi e *software*, mentre la conversione avviene – per ogni documento – in tre formati: ODF (ISO/IEC 26300:2006), PDF/A-1 (ISO 19005-1) e TXT (ISO/IEC 8859).



Infine, la sicurezza dall'esterno e dall'interno è garantita dal fatto che nessun amministratore ha accesso a tutte le aree, ma solo a singole parti.

¹¹⁸ Dalla guida per l'autore; documento in appendice



Le immagini presentate¹¹⁹ sono basate su una rappresentazione della struttura di PAD, che diverrà più chiara ora che andremo ad analizzarla.

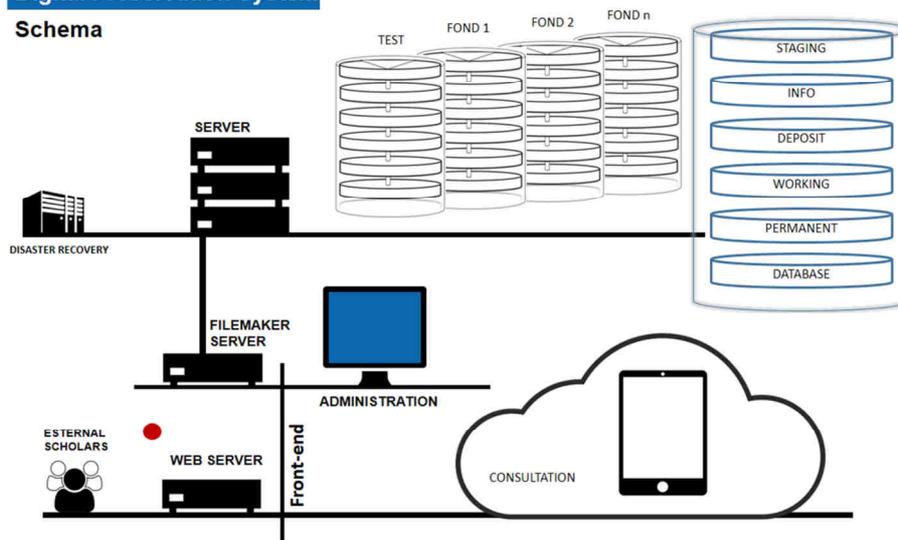
La struttura di PAD e le funzioni del modello OAIS

L'architettura del progetto PAD, è ben articolata rispetto alle funzioni che deve svolgere e cerca di essere il più possibile conforme al modello OAIS, il flusso di lavoro può essere analizzato seguendo il percorso che compie il conferimento, rappresentato dallo schema evidenziato nella figura seguente, oppure rispetto alle funzioni previste dal modello OAIS; utilizzeremo entrambi gli approcci (per aree e per funzioni OAIS) per capirne il funzionamento, la logica, e le intersezioni tra i due “modelli”.

¹¹⁹ Tutte le figure qui presentate sono tratte dalle presentazioni di PAD e delle relative sezioni realizzate da Primo Baldini.

Digital Preservation System

Schema



La logica sottesa alla divisione in aree è quella di poter gestire in maniera sicura i documenti e i dati relativi, isolando tra loro i diversi momenti del “percorso” compiuto dalle fasi di preacquisizione fino alla conservazione e alla successiva possibilità di riutilizzare i dati e i documenti: potendo garantire che in ogni fase del processo la gestione dei contenuti del conferimento sia sempre ottimale e che eventuali errori o problemi rimangano contenuti solo ed esclusivamente nella fase in cui si manifestano e non possano propagarsi agli altri livelli ma rimangano circoscritti, si possono limitare eventuali danni e – nel remoto caso di problematiche effettive (in quanto come vedremo dalla struttura è piuttosto improbabile) – recuperare solo la parte di lavoro andata perduta, e non il conferimento nel suo complesso. La separazione tra la “struttura ad aree” e quella che segue invece le funzioni del modello OAIS è assolutamente arbitraria, in quanto non esiste questa separazione internamente rispetto al sistema, e anche le divisioni sono solo funzionali alla descrizione, e rappresentano di per sé le fasi di un processo piuttosto che dei momenti ben definiti: è importante rilevare che non esiste un ordine obbligatorio prestabilito per le aree: dal momento in cui il conferimento entra nel sistema di PAD gli esiti finali sono la copia definitiva dei documenti elaborati per la conservazione sui vari supporti destinati allo scopo e la possibilità di riutilizzare per motivi di studio questi documenti; i file verranno

spostati da un'area all'altra secondo necessità e non seguendo passaggi obbligati. L'ordine previsto inizialmente era:

- Area di *staging*;
- Area di deposito;
- Area per la conservazione permanente;
- Area di lavoro;
- Area delle informazioni;
- Area del database.

La logica di quest'ordine dipendeva dalla necessità di avere una prima area di “approdo dei dati”, in attesa della disponibilità di un operatore per essere processati e per la quarantena dai virus (spostata ora in area deposito), successivamente di un'area di deposito per fare le operazioni di selezione ed eliminazione, per poi copiare il conferimento così definito nell'area della conservazione permanente, da cui venivano presi i documenti per poterci lavorare nell'area di lavoro; le informazioni sul conferimento venivano così allegate nell'area delle informazioni e tutte le altre informazioni del conferimento erano invece salvate e aggiornate nell'area database. Tuttavia, lavorando con una quantità cospicua di file (il conferimento di oltre cinquantamila file di Francesco Pecoraro nel 2016) ci si è trovati nella necessità pratica di procedere con un ordine differente: frequenti interventi manuali sui file acquisiti hanno portato a trasferire il conferimento prima nell'area di lavoro, e solo in seguito nell'area per la conservazione permanente. Inoltre si è notato che l'area delle informazioni sul conferimento può venire aggiornata dall'operatore con i materiali relativi appena dopo aver importato il conferimento in PAD, e quindi appena “dopo” l'area di *staging*. In ogni caso, l'ordine dipende dalla complessità del conferimento: l'area di lavoro e quella per la conservazione permanente sono uguali ad eccezione dei permessi di accesso e del fatto che l'area permanente è un'area di sola lettura, quindi a seconda di quanto lavoro “manuale” è necessario prima della conferma, questo passerà prima per l'area di lavoro o per quella permanente (ma casi così complessi vengono considerati eccezionali).

Questa separazione tra i due modelli (OAIS e “ad aree”) risulta però strumentale rispetto ad un punto abbastanza preciso: se da un lato possiamo analizzare le fasi operative nel loro susseguirsi, possiamo al contempo confrontarle con quanto previsto dallo standard a livello concettuale e valutarne la conformità.

Nel complesso, le azioni che rimandano alla descrizione secondo le funzioni del modello OAIS (benché la fase preliminare non ne faccia parte), spesso si troveranno accavallate tra due (o più) aree della struttura, in quanto le “fasi funzionali” tengono insieme tutti gli elementi per espletare una funzione, mentre le aree sono divisioni tecniche (e talvolta fisiche) percorse dai documenti che frammentano intenzionalmente il processo per garantire ad ogni livello la sicurezza e il controllo sui documenti e sui dati; pertanto non è previsto che combacino integralmente.

Infine, per ogni fase del processo, è prevista l’integrazione con il sistema di gestione e controllo della qualità QUANDO, *Quality control for Archiving and Networking of Digital Objects*, che tiene traccia di tutte le operazioni svolte e che avremo modo di approfondire nella sezione dedicata (pertanto, ogni volta che si incontrerà l’interazione con QUANDO in questa sezione, ci si accontenti – per il momento – di prenderne nota mentalmente).

Operazioni preliminari

In realtà, ben prima che l’archivio dell’autore entri “materialmente” (o virtualmente) in PAD, c’è un sistema già in piena attività: le primissime fasi sono per così dire “esterne” all’architettura logica di PAD che ne permetterà il trattamento e le procedure per la conservazione: sono affidate al Comitato di Gestione, e quindi sono in capo all’Università di Pavia, o almeno a chi – delegato per essa – è incaricato della gestione di PAD (nel suo complesso, cioè tecnica ma anche burocratica e amministrativa)¹²⁰. Se volessimo guardare molto brevemente quali sono

¹²⁰ Il Comitato di Gestione è così composto:

Presidente: Fabio Ruge (rettore dell’Università di Pavia); Coordinatore del progetto: Paul Gabriele Weston; il Comitato Scientifico composto da Simone Albonico, Umberto Anselmi Tamburini, Gianni Danese, Maria Antonietta Grignani, Clelia Martignoni, Giorgio Panizza, Maria Giovanna Ruberto, Beppe Severgnini, Angelo Stella, Luigi Carlo Ubertazzi; Responsabili Tecnico-scientifici: Primo

le fasi amministrative preliminari che portano all'avvio di tutta la procedura del sistema di PAD, troveremmo questo percorso:

- il Comitato seleziona, ma anche accetta la candidatura di, un autore per proporgli di conservare il proprio archivio digitale con PAD, quindi il verbale di delibera viene registrato;
- un operatore prende contatto con l'autore designato per presentare il progetto e proporre la collaborazione (volta al conferimento) con PAD, il contatto con l'autore e il relativo esito vengono registrati;
- in caso di esito positivo inizia la fase di preparazione amministrativa, di cui si occupa la segreteria: all'autore vengono inviati il questionario e il contratto;
- a questo punto può iniziare la fase di preparazione tecnica: un responsabile tecnico preparerà il materiale (dischi fissi criptati) mentre l'incaricato si accorderà con l'autore per il trasferimento dei dati (compresi i supporti esterni obsoleti e non emersi dal questionario informativo).

Per quanto concerne la fase precedente l'acquisizione vera e propria (quella della copia dei dati), le procedure sono illustrate nella brochure informativa¹²¹ per l'autore (vi sono illustrate anche le fasi successive, ma con minor dettaglio tecnico, in quanto non sono procedure che riguardano l'autore o che debbano essere espletate da lui):

«Un incaricato dell'Università di Pavia fisserà un appuntamento con l'autore e realizzerà una copia su hard disk dei file conferiti. È possibile conservare in PAD qualsiasi tipo di file: documenti di testo, foto, immagini, audio, video, pagine web e social network. Le situazioni degli archivi conferiti possono essere molto varie, e sarà perciò necessario individuare caso per caso le soluzioni più idonee. [...] Quanto più la donazione è ampia e la copia eseguita da PAD fedele alla situazione originaria, tanto più sarà possibile utilizzare e interpretare i materiali raccolti. È prevista la possibilità di conferire parti di collezioni in momenti diversi: nel caso in cui

Baldini, Emanuela Carbé; Relazioni esterne: Grazia Bruttocao; Segreteria: Nicoletta Leone.

¹²¹ La brochure informativa si trova tra i materiali allegati in appendice.

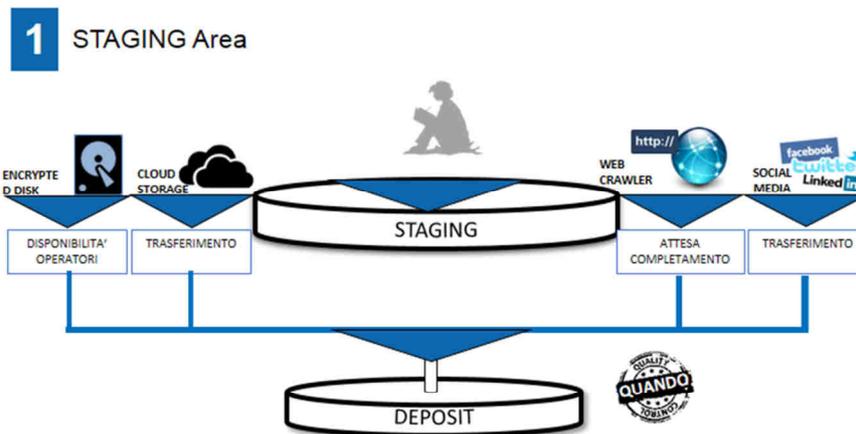
l'autore voglia ampliare il proprio fondo PAD, l'archivio verrà aggiornato con un sistema che mantenga le strutture e i dati precedenti e accolga separatamente, pur con interconnessioni, le directory e i file conferiti in momenti successivi. [Come primo passaggio l'incaricato] dell'Università di Pavia si recherà dall'autore con un hard disk appositamente predisposto, su cui verranno copiati i file che l'autore deciderà di conferire. Si tratterà di un semplice "copia e incolla": i dati presenti sul computer dell'autore non saranno in alcun modo modificati. L'hard disk è dotato di criptazione hardware e ha una password conservata dal responsabile tecnico di PAD. La criptazione è una misura preventiva in caso di furto dell'hard disk durante le fasi di trasporto.»

Una volta che il disco fisso criptato contenente i file copiati dal computer dell'autore sia giunto all'Università di Pavia possono cominciare le procedure preliminari di controllo che porteranno all'acquisizione definitiva dei file e alla relativa conservazione permanente:

Area di *staging* (area temporanea)

Rappresentata nella figura seguente, l'area di *staging* è la prima area temporanea dove tutti i file del conferimento (cioè quelli copiati nel disco fisso criptato, ed eventuali altri supporti conferiti dall'autore come dischi fissi esterni, CD-ROM, DVD, vecchi *floppy disk*, ...) e altri materiali (come possono essere per esempio i file provenienti da servizi di *cloud storage*, *webmail*, piattaforme di *social media*, blog, siti o altri materiali non provenienti dal diretto conferimento da parte dell'autore) vengono mantenuti in attesa della disponibilità degli operatori a procedere con le successive fasi di acquisizione e verifica dei file: nel caso in cui l'operatore di PAD iniziasse subito il lavoro non sarebbe necessario far transitare i dati per l'area di *staging*, ma verrebbero copiati direttamente in area deposito e i dati inerenti il conferimento verrebbero inseriti nell'area delle informazioni. L'area di *staging*, infine, è particolarmente utile come deposito temporaneo (ma sempre al sicuro all'interno di PAD) in attesa di ricevere tutti i dati: in

particolare, i servizi legati al web, come i servizi di cloud, i social networks, i siti web, i blog, necessitano di tempi abbastanza lunghi per la raccolta dei dati.



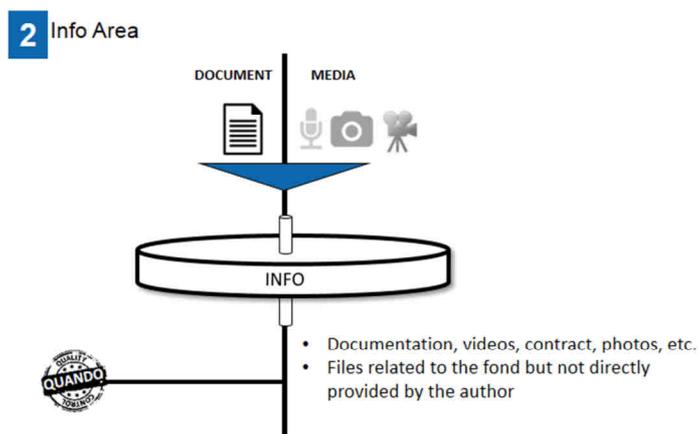
Ufficialmente, la preacquisizione non fa parte del modello OAIS, ma permette di svolgere alcune operazioni preliminari: in questa fase vengono svolte le seguenti azioni:

- Inserimento dei dati nel sistema QUANDO (dati “primari” del conferimento come ad esempio le date del conferimento, dell’invio e della ricezione dei documenti da parte dell’autore, etc);
- Ricezione del conferimento: i dati copiati sul disco fisso criptato, o provenienti da supporti esterni conferiti dall’autore vengono riversati all’interno dell’area di *staging*;
- Messa in sicurezza dei dati: le operazioni di *disaster recovery* prevengono la perdita dei dati ricevuti (dal momento in cui i dati entrano in PAD è sempre attiva la sincronizzazione tra i server primari dell’Università di Pavia e quelli a Cremona);
- Attesa della disponibilità degli operatori per poter procedere al vaglio automatico e manuale dei file, della documentazione completa e del contratto firmato dall’autore;

Area delle informazioni (area statica)

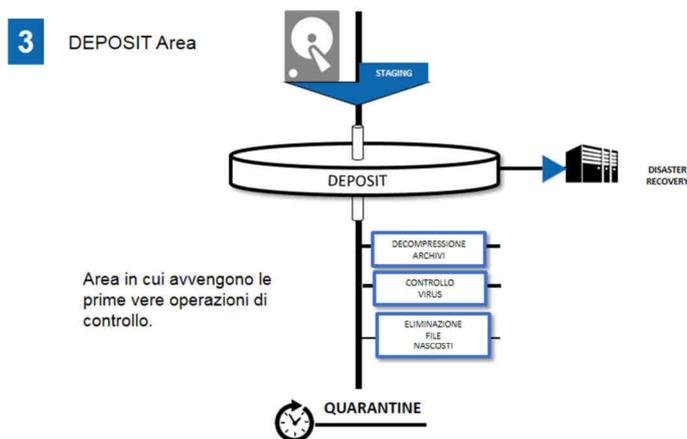
L’area delle informazioni (nella figura successiva) è un’area statica della struttura di PAD destinata alla conservazione di materiali collegati al conferimento ma non prodotti direttamente dall’autore: in quest’area

vengono conservate (e collegate al conferimento) le digitalizzazioni del contratto firmato dall'autore, della lista dei file conferiti, dell'intervista informativa, i documenti collegati a vario titolo al fondo (amministrativi e non), alcune foto, interviste anche filmate dell'autore del fondo; una prima parte di questi documenti (la lista dei file del conferimento, non sarà tra questi) possono venire inseriti contemporaneamente all'ingresso dei file in PAD: ecco perché rispetto all'ordine pensato inizialmente quest'area è in seconda posizione.



Area di deposito (area temporanea)

L'area di deposito (illustrata nella figura successiva), è l'area in cui i file entrano a tutti gli effetti all'interno del processo di conservazione: uno o più operatori si occuperanno da questo momento della corretta acquisizione del conferimento, della stesura definitiva del contratto con l'autore – per quanto concerne i file effettivamente e volontariamente conferiti – e della gestione dei file.

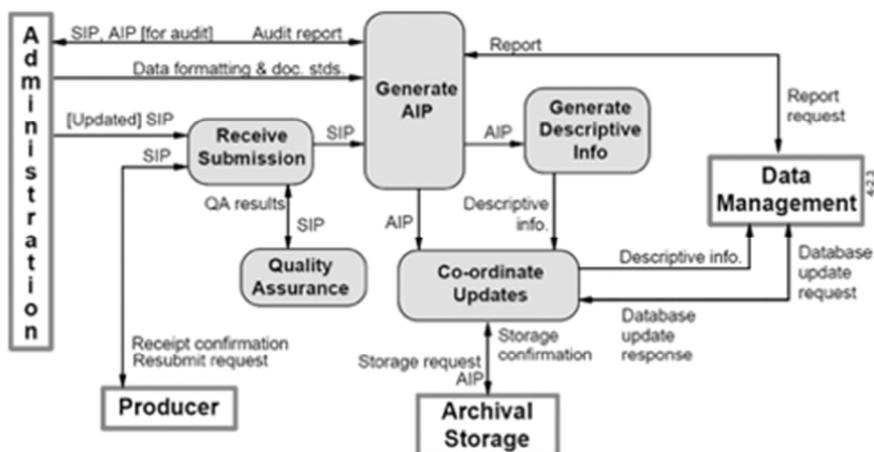


Le varie fasi che possiamo distinguere in quest'area sono:

- a. Controllo visivo: ad un primo controllo tutti i file che non sono evidentemente parte del conferimento (file del sistema, altri file non intenzionalmente copiati come file audio, video, immagini, eseguibili e di altro tipo che non possono essere considerati come parte dell'archivio dell'autore) vengono segnalati per l'eliminazione, come anche altri file da escludere indicati dall'autore: va comunque chiarito che la decisione finale di se e cosa eliminare spetta sempre all'autore, PAD può solo segnalare eventuali file che potrebbero non essere adeguati. Inoltre, se all'interno dell'archivio fossero presenti file o cartelle compressi (come ad esempio .zip, .rar) il programma PadCleaner li decomprime per visualizzare il quadro generale del conferimento;
- b. Eliminazione file nascosti: di solito i file nascosti sono file che vengono generati dal sistema operativo, e pertanto non fanno parte del conferimento. Nel caso in cui fossero nascosti dei file creati dall'autore, non spetta a PAD indagarne i motivi (potrebbero contenere dati sensibili), e vengono perciò eliminati automaticamente: se era nelle intenzioni dell'autore conferire anche alcuni file nascosti dovrà conferirli nuovamente modificandone gli attributi;
- c. Controllo integrità/virus: *«il programma PadChecker rileva la presenza di virus: se individuati, l'autore viene tempestivamente informato, e se necessario gli si offre una prima assistenza. Nell'archivio PAD i 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, "codici" univoci che rimarranno associati per sempre ai singoli file. PadChecker effettua una scansione quotidiana*

basata sull'SHA-1 e segnala all'amministratore eventuali manomissioni.»¹²²

- d. Generazione elenco file: «il programma PadPrint genera una lista di tutti i file conferiti, che verrà inviata all'autore per una convalida. In caso di ripensamenti l'autore 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, l'autore potrebbe infatti aver conferito alcuni materiali inconsapevolmente. La lista inviata mostra in maniera ordinata e chiara tutta la collezione. PAD vuole dare la massima garanzia all'autore instaurando un rapporto di trasparenza e fiducia: solo con la firma del documento, poi allegato al contratto, vengono avviate le procedure di immagazzinamento, conservazione e archiviazione.»¹²³
- e. *Disaster Recovery*: il conferimento e i dati relativi – benché ancora non confermati – vengono comunque assicurati tramite procedure di *disaster recovery* (tutti i dati sono sempre in copia con i server di Cremona, con l'unica eccezione dell'area di lavoro che può sempre essere ricreata).



La fase di acquisizione (come funzione del modello OAIS, illustrata nello schema precedente) opera la presa a carico e la preparazione

¹²² Dalla brochure informativa per gli autori.

¹²³ *Ibidem*

automatizzata dei dati (SIP) per l'archiviazione di lunga durata. Il sistema di acquisizione segue un processo definito e deve integrarsi con il controllo qualità; le azioni svolte in questa fase:

- *Submission Information Package* (SIP): sono i metadati creati dal soggetto produttore, o più probabilmente dall'operatore; il conferimento viene ricevuto in area deposito;
- QUANDO riceve i metadati relativi al conferimento in formato XML, e vengono archiviati (QUANDO gestisce i metadati di tutti i conferimenti, mentre gli applicativi solo di quello in lavorazione: perciò c'è un programma che si occupa della sincronizzazione tra QUANDO e il fondo);
- Viene controllata la presenza di virus tra i file del conferimento;
- Vengono generati i *checksum* dei file con la funzione di HASH SHA1;
- I file vengono normalizzati;
- L'*Archival Information Package* (AIP) viene prodotto come file *manifest* del conferimento in formato XML;
- Vengono utilizzati alcuni *tool* di *file characterization* per assicurare la corretta caratterizzazione del file (ne vengono usati anche più d'uno sullo stesso file per ottenere informazioni diverse: ad esempio Exiftool, DROID, JHOVE, ...);
- Le operazioni vengono costantemente monitorate;
- Vengono aggiornati i metadati relativi presenti in QUANDO, con quelli estratti in questa fase.

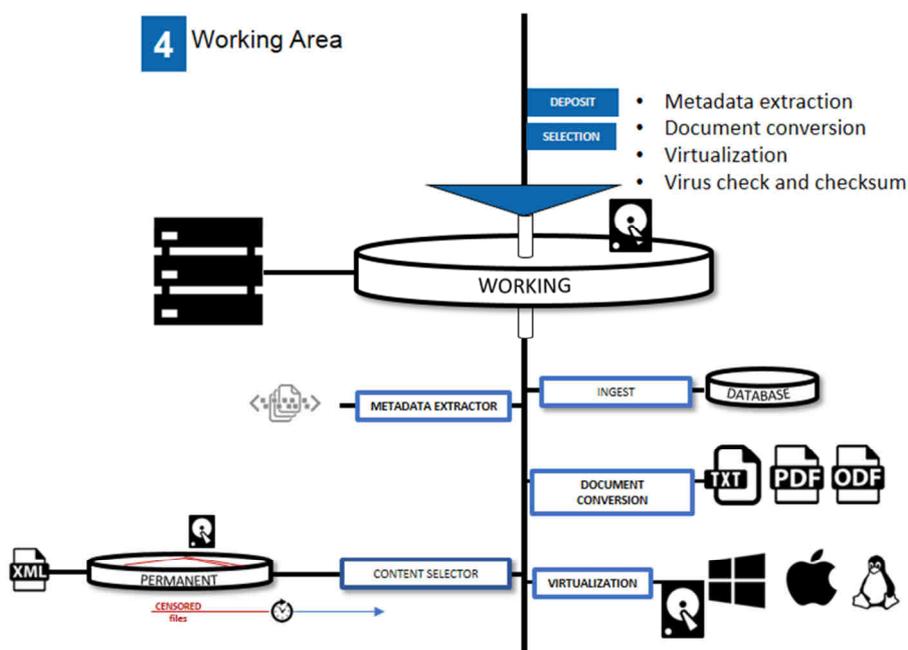
A cavallo tra quest'area e la successiva si concludono anche le procedure amministrative e burocratiche iniziate dal comitato con la scelta dell'autore:

- Il responsabile tecnico riceve i materiali contenenti il conferimento (dischi fissi criptati, *hardware* esterno e/o obsoleto), e attiva le procedure tecniche;
- Il responsabile tecnico attiva le procedure di sicurezza mettendo in area di *staging* il conferimento;
- la segreteria riceve il contratto firmato dall'autore: il conferimento viene trasferito nell'area di deposito;

- Il responsabile tecnico produce la lista dei file ricevuti, la manda in segreteria che la gira all'autore per la convalida;
In caso di conferma il contratto viene chiuso e i file destinati alla conservazione vengono trasferiti nell'area apposita.

Area di lavoro (area dinamica)

L'area di lavoro (mostrata nella figura seguente) è un'area dinamica a cui fanno capo diverse funzionalità di PAD: se l'area per la conservazione permanente è il nucleo centrale da un punto di vista statico, e della conservazione intesa come fissazione stabile dei dati, l'area di lavoro, invece, è il fulcro dinamico del sistema, quello in cui vengono realizzate tutte le funzioni per una conservazione “dinamica” e che permette la fruizione dei documenti conservati. Nel caso in cui ci sia una frequente necessità di intervenire manualmente sul conferimento, i documenti passano nell'area di lavoro prima di venire copiati nell'area della conservazione permanente, nell'ordine in cui lo vediamo ora; è da segnalare però che ultimamente capita sempre più spesso che gli autori conferiscano un *backup* del proprio disco e poi deleghino a qualcuno il controllo del contenuto, rendendo sostanzialmente un passaggio obbligato questa fase di selezione.



Le principali attività che vengono effettuate nell'area di lavoro sono:

- a. Estrazione metadati: con l'utilizzo di alcune e differenti librerie si estraggono i metadati tecnici dai file. L'utilizzo di diversi strumenti è dovuto al fatto che i comportamenti dei vari strumenti sono differenti e permettono di estrarre di volta in volta metadati tecnici diversi (più avanti se ne parlerà in maniera più approfondita).
- b. Per poter conservare l'accesso al contenuto e la rappresentazione del documento così come viene conferito il sistema di PAD provvede alla conversione del documento di testo in tre formati differenti: una versione in formato TXT, utile per l'analisi dei testi elettronici con varie applicazioni (da luglio 2016, per esempio, c'è la possibilità di analizzare in maniera automatica un testo con Voyant Tools¹²⁴) e per l'anteprima del documento negli applicativi di PAD; una versione in Open Document Format, per cercare di mantenere il più possibile l'ambiente simile a quello di un text editor (anche se è stato elaborato con un differente *software* proprietario il cui formato di uscita non è ODF) e perché rappresenta il formato più "aperto" disponibile; e una versione in PDF/A-1, per mantenere l'impostazione della pagina così come prevista dal documento originario (compresi font, formattazione del testo, spaziature, linee e tutto quanto sia fissato sulla pagina).
- c. Se la conversione è fondamentale ai fini della conservazione del contenuto nel tempo, non conserva però la forma originaria: come nella filologia sono rilevanti anche quegli aspetti non esclusivamente testuali, ma piuttosto legati alla materialità del libro e della pagina, così nel nostro caso l'ambiente *software* con cui il documento è stato creato (sia il sistema operativo che il text editor) ne costituisce il contesto informatico d'origine. Il modo migliore per renderlo – e probabilmente anche l'unico – è l'emulazione: all'interno dell'area di lavoro è prevista la

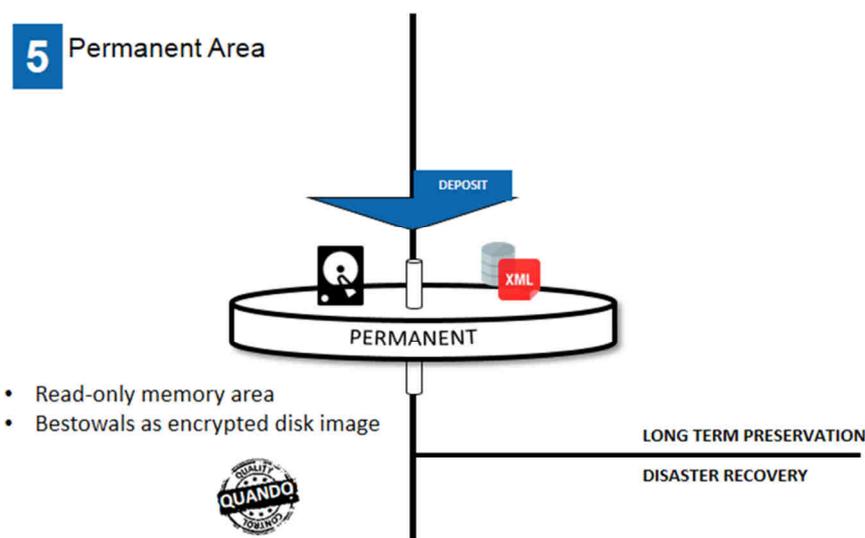
¹²⁴ <https://voyant-tools.org/>

virtualizzazione dei sistemi operativi e dei programmi che hanno creato i documenti, permettendo di poterli esaminare nel loro ambiente originario.

- d. Quotidianamente viene fatto un controllo con *software* antivirus sui documenti e vengono confrontati i *checksum* per valutare che non ci siano errori o danni ai documenti.
- e. Poiché non si può accedere direttamente ai file in sola lettura nell'area per la conservazione a lungo termine, i file vengono consultati dall'area di lavoro attraverso le modalità che abbiamo visto (la conversione e l'emulazione).

Area della conservazione permanente (area statica)

Questa è l'area di lavoro centrale rispetto alla conservazione a lungo termine: in questo punto (illustrato nella figura seguente) vengono ottimizzate tutte le procedure per la conservazione a lungo termine del conferimento digitale. Una volta che l'autore ha confermato la lista dei file conferiti – o che tramite successivi aggiustamenti si sia arrivati a questo punto – i file verranno trasferiti dall'area di deposito a quella per la conservazione.



Il conferimento viene copiato nell'area deputata alla conservazione a lungo termine in forma criptata sui server del centro di calcolo dell'Università di Pavia: una volta che la copia è completata (e che quindi il conferimento sia terminato), l'accesso a quest'area è consentito in modalità di sola lettura, e i documenti non possono essere

modificati. Per garantire la maggior sicurezza possibile, come anche richiesto dai *NDSA Levels of Digital Preservation*¹²⁵, non è certo l'unica copia disponibile dei dati conferiti, né l'unica forma in cui vengono conservati da PAD, che invece utilizza diversi sistemi per garantire che i dati non vadano persi, e allo stesso tempo per poterli recuperare in caso di problemi:

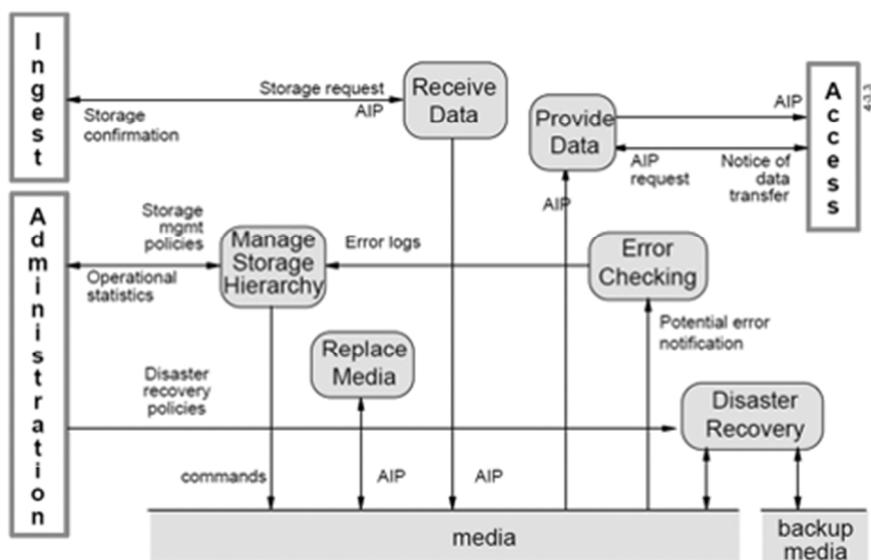
- *Long Term Preservation*. I dati vengono masterizzati su DVD per la conservazione *archival grade* con strato in oro, in forma non criptata, impacchettati sottovuoto (per limitare il più possibile il deterioramento fisico dovuto ad agenti esterni) e conservati in cassetta di sicurezza presso il *caveau* di una banca: questo permette che possano essere riutilizzati in futuro. La masterizzazione dei dati è il procedimento tecnico; è opportuno però tenere in considerazione se sia il caso di copiare solo i dati originali (e quindi trasferire il *bitstream* dei documenti) oppure anche le conversioni dei file in TXT, PDF/A-1, ODF e i dati del conferimento: rispetto a questo punto si stanno valutando alcune implementazioni data l'importanza del contesto;
- *Disaster recovery*: oltre alla sincronizzazione sul secondo server a Cremona vengono fatte delle copie di *backup* su nastro del server primario (in questo modo vengono osservati i parametri suggeriti dall'NDSA di avere “almeno tre copie geograficamente separate con differenti trattamenti di *disaster recovery*”: copie criptate sul server primario di Pavia e su quello di Cremona, *backup* su nastro e DVD non criptata in cassetta di sicurezza).

Una volta trasferito completamente il conferimento nell'area di sola lettura dell'archiviazione, le aree di *staging* e deposito vengono cancellate; saranno riutilizzate in caso di un nuovo conferimento da parte dello stesso autore (perché bisogna ricordare che ogni autore ha una sua specifica area). A questo punto l'unica area che contiene i dati è l'area della conservazione permanente, ed è possibile accedervi solo in modalità di lettura: qualsiasi operazione si debba fare sui documenti

¹²⁵ La tabella già citata si trova tradotta in appendice; (Phillips et al. 2013)

conferiti da questo momento in poi, andrà effettuata nell'area di lavoro in cui verranno copiati i file.

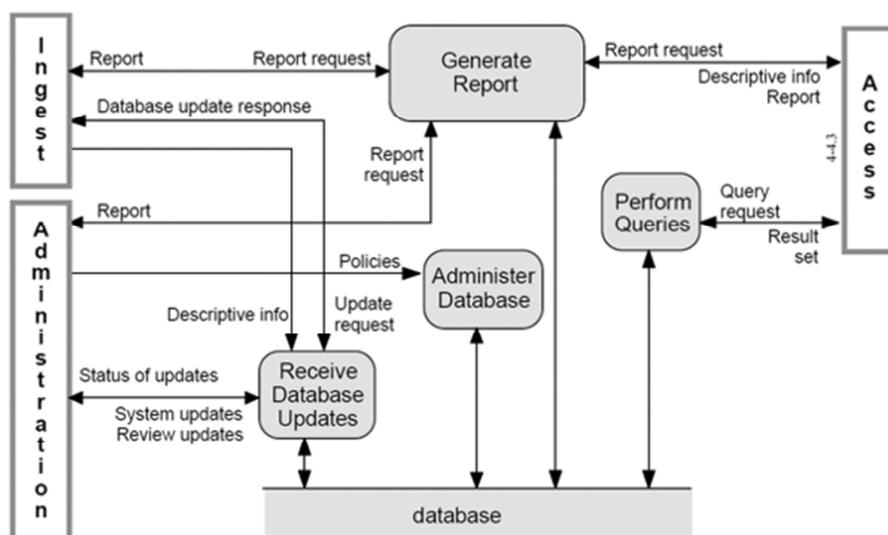
In realtà, terminata questa fase con le copie ridondanti sui server dell'Università, i *backup* su nastro e i DVD in cassetta di sicurezza, le operazioni vere e proprie di conservazione a lungo termine potrebbero in buona parte considerarsi concluse. Se, cioè, l'unico obiettivo fosse garantire la possibilità che gli archivi degli autori sopravvivano autentici nel tempo a livello di *bitstream*, le fasi che abbiamo affrontato finora soddisfano questo requisito poiché garantiscono che questi stessi file copiati nella struttura di PAD possano mantenersi intatti per gli anni a venire, *esattamente come sono*. L'altra sfida che pone la conservazione digitale a lungo termine, però, è quella di mantenere i file intatti, oltre che a livello di *bitstream*, anche nella possibilità di accedere al contenuto e alla rappresentazione del contenuto (e questo non sempre è possibile contemporaneamente): per questo la struttura e gli obiettivi di PAD non si esauriscono con l'acquisizione del conferimento nell'area di deposito permanente: oltre a garantire la conservazione anche a livello di rappresentazione del contenuto, nell'area di lavoro la struttura lavora sui metadati e sul riuso dei dati (e dei contenuti).



In riferimento alla funzione del modello OAIS questa è la fase di archiviazione o *archival storage* (illustrata nell'immagine precedente), in cui i documenti effettivamente selezionati per il conferimento vengono copiati nell'area della conservazione permanente:

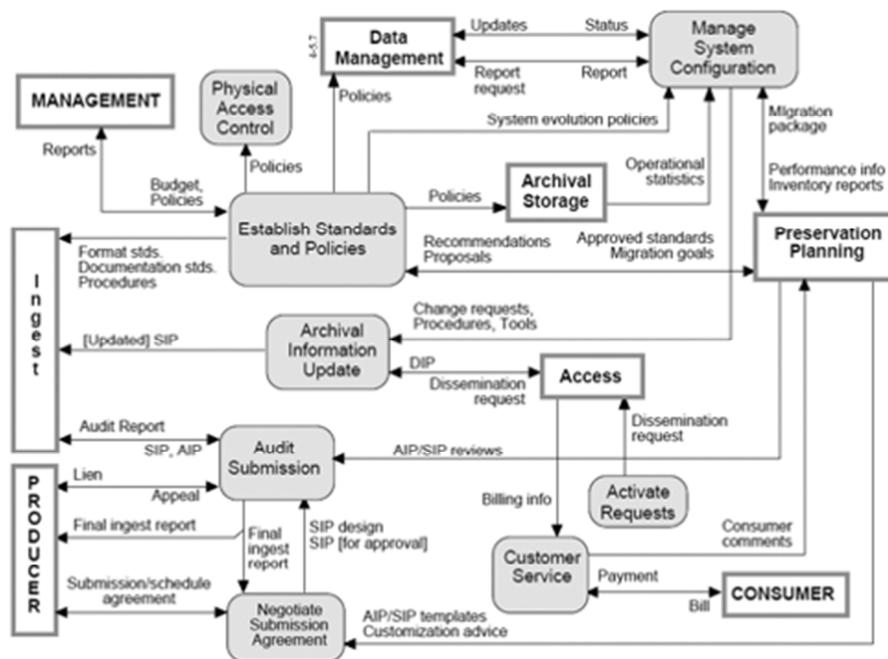
- Viene mantenuta l'immagine disco del materiale conferito;
- Si procede alla masterizzazione del DVD per la conservazione;
- Quotidianamente il sistema provvede al controllo degli errori attraverso la verifica dei *checksum*;
- Quotidianamente viene operato, sempre automaticamente, il controllo dei virus;
- Si controlla che le varie procedure di *disaster recovery* siano attive (*backup* su nastro e server geograficamente distaccato)
- Aggiornamento dei metadati di QUANDO con i risultati dei vari controlli, e la segnalazione di eventuali anomalie.

Per quanto riguarda le funzioni del modello OAIS, le funzioni rimanenti (*Data management,, Administration, Access, Preservation planning* e *Common Services* ovvero Gestione dei dati, Amministrazione, Accesso, Pianificazione della conservazione e Servizi comuni) riguardano l'ordinaria amministrazione della struttura, necessaria a garantire un corretto funzionamento di tutto il sistema e a mantenere i dati sempre aggiornati; queste funzioni attraversano interamente o per una buona parte il sistema di PAD, e le azioni che prevedono non saranno pertanto esclusivamente ascrivibili ad un'area piuttosto che ad un'altra . Rendiamo conto delle rimanenti funzioni simultaneamente in questo punto, prima dell'area del database.



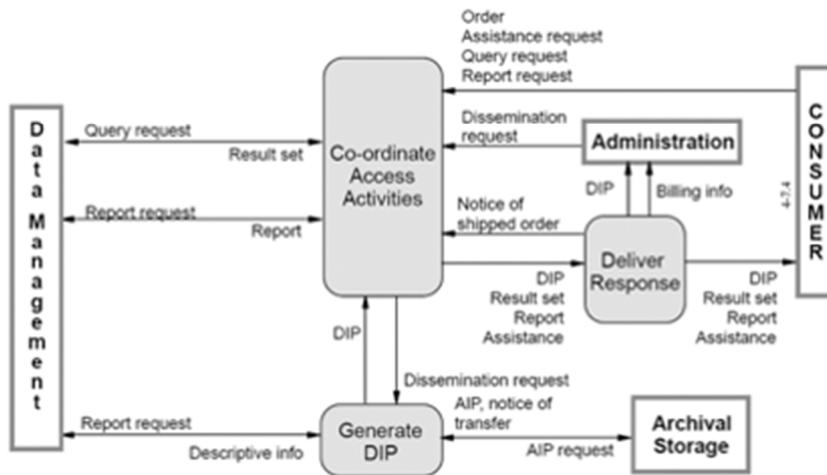
Per la gestione dei dati (illustrata nell'immagine precedente) le azioni previste sono:

- La regolare amministrazione del database;
- L'attivazione fondi e conferimenti (in fase di acquisizione e nei vari passaggi tra una fase e l'altra);
- Creare *query* e statistiche da utilizzare in fase di catalogazione del conferimento;
- *Report* sui file del conferimento da allegare al contratto;
- Aggiornamento metadati di QUANDO ad ogni modifica e operazione effettuata.



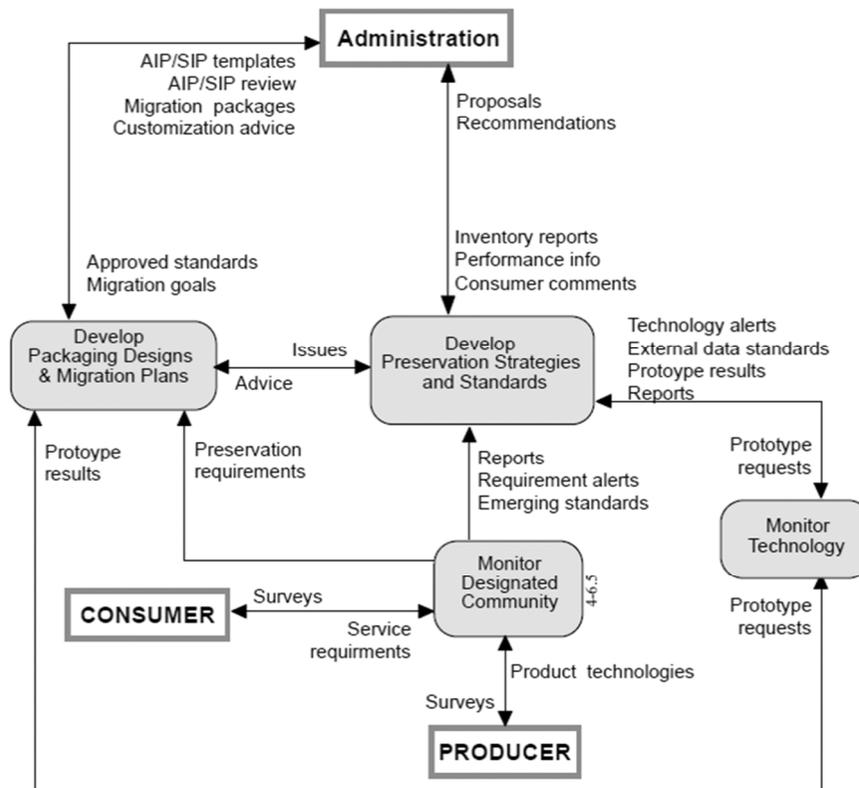
L'amministrazione dei dati e del sistema (nella figura precedente) provvede alle seguenti operazioni:

- Audit delle operazioni;
- Interazione con il controllo qualità QUANDO (per le verifiche);
- Abilitazione utenti e permessi in base alle restrizioni dei documenti o dei conferimenti;
- Registrazione del contratto con l'autore.



Le politiche di accesso (nell'immagine precedente) principalmente provvedono a:

- Garantire il rispetto del diritto d'accesso, secondo la volontà dell'autore;
- Fornire un'interfaccia utente per la navigazione dei documenti;
- Fornire una piattaforma di ricerca open source (Solr/Lucene) che opera sui documenti di un fondo.



Le azioni intraprese per pianificare la conservazione dei documenti (e quindi le operazioni da tenere sotto controllo per non rischiare di

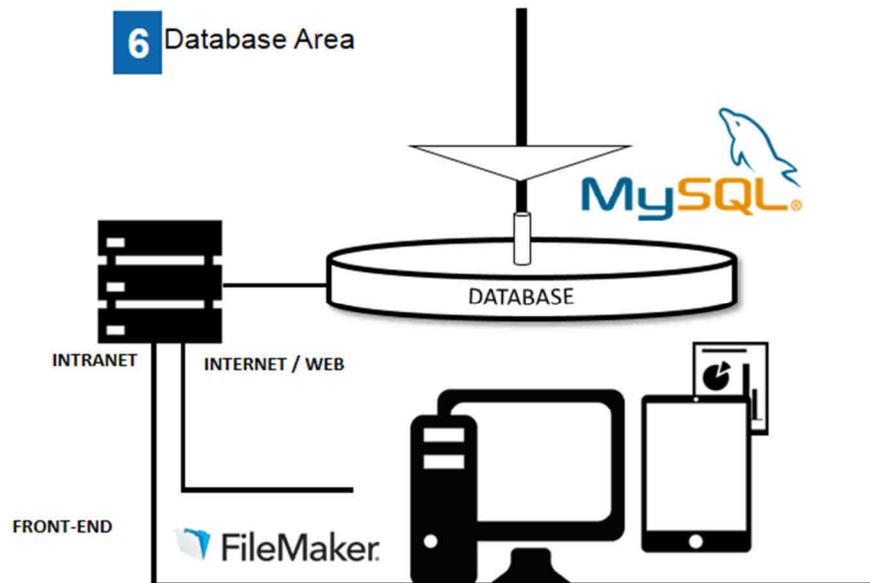
ritrovarsi con errori, virus, perdita di dati e altre problematiche, così come illustrate nell'immagine precedente) sono:

- Audit delle operazioni, in modo da garantire sempre un'adeguata qualità del lavoro svolto (anche in automatico);
- Interazione con QUANDO (per verificare quali possono essere i punti – se ce ne sono – che hanno generato errori);
- Sorveglianza negli aggiornamenti delle procedure di:
 - Estrazione metadati,
 - Conversione documenti,
 - Virtualizzazione;
- Sorveglianza degli sforzi fatti in materia d'archiviazione per avere fondi e dati consistenti anche sotto il profilo archivistico;
- Formulazione di raccomandazioni per il mantenimento della leggibilità dell'informazione stoccata;
- Pianificazione delle migrazioni e dei processi di copiatura.

Infine, i Servizi comuni, non hanno uno schema delle funzioni, poiché si intende un insieme di relazioni con tutti gli altri elementi pervasivo al punto che una schematizzazione grafica genererebbe solo più confusione.

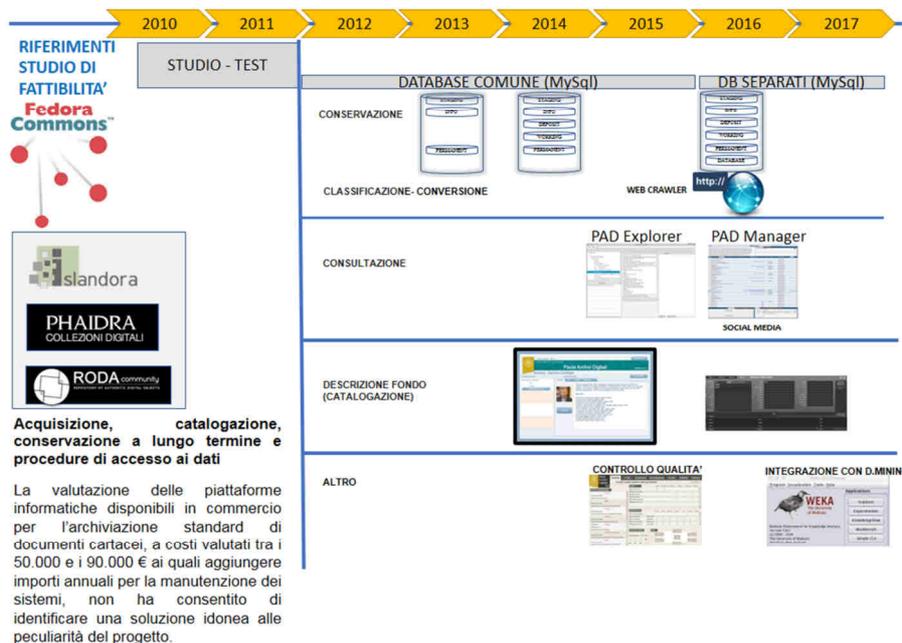
Area del database (area dinamica)

L'area del database (nella figura seguente) è un'area dinamica all'interno del sistema di PAD che serve a conservare e riutilizzare i dati estratti dal conferimento: i parametri che riguardano i fondi e i conferimenti, l'elenco dei file conferiti (non quello firmato dall'autore – presente nell'area delle informazioni – ma quello per la gestione del sistema), i metadati tecnici e quelli estratti dai *tool*, i *checksum* del conferimento: è un'area fondamentale per collegare i dati estratti dai file del conferimento, ai file stessi a cui si riferiscono; vengono utilizzati MySQL per la parte della conservazione e FileMaker per la parte di consultazione.



QUANDO e PAD Manager

Per PAD è stato elaborato un sistema di gestione e controllo qualità, denominato QUANDO, cioè *Quality control for Archiving and Networking of Digital Objects*, che permette di tenere traccia di ogni fase, di quali operazioni siano state compiute e quale sia l'operatore responsabile dell'operazione. Posto che, come già ricordato, si tratta di un processo e non di uno stato, proveremo a dare un breve resoconto del funzionamento di QUANDO, del principale applicativo cioè PAD Manager e dell'evoluzione delle implementazioni tecnologiche operate finora senza pretendere di essere esaurienti sotto questo profilo: non sarà infatti possibile fissare le ultime attuali o possibili future implementazioni, inseguendo "l'ultima versione"; bisogna tenere in considerazione che a livello tecnologico è sempre possibile implementare il sistema (e che se ad oggi non fosse possibile recuperare un determinato documento perché il formato non viene riconosciuto, non è da escludere che in futuro possa tornare ad essere accessibile).



Il sistema di controllo qualità “QUANDO”

Il Sistema *Quality control for Archiving and Networking of Digital Objects* è fondamentale all'interno della struttura, in quanto permette al sistema stesso – attraverso la sua continua e costante integrazione con le operazioni – di avere una certa flessibilità per quanto riguarda le tecnologie implementate e il personale della struttura: in questo modo un avvicendamento tra personale o tecnologie non porterebbe a una perdita irreparabile del lavoro o delle operazioni compiute, in quanto ogni movimento (dal primo contatto informativo con l'autore fino all'aggiornamento automatico dei dati del conferimento) è registrato ed è registrato l'operatore che lo ha svolto. Attraverso le schermate commentate del sistema, andremo ora a vederne il funzionamento e l'integrazione con le varie fasi del processo di conferimento¹²⁶.

Il menu principale, come mostrato nella figura seguente, offre come schermata principale un riepilogo dell'intero sistema, e permette di muoversi tra alcuni menu tematici: quello dei fondi in cui sono tracciate tutte le procedure relative al trattamento tecnico dei fondi una volta ottenuti, quello dei conferimenti, inerente le procedure di conferimento; quello della documentazione relativa ai conferimenti (o

¹²⁶ Le immagini commentate sono prese dal manuale del sistema di controllo qualità QUANDO disponibile nella sezione “Aiuto” interna al programma.

altra attinente ma non direttamente connessa); quello delle persone tutti i soggetti coinvolti nelle operazioni di PAD.

PAGINA INIZIALE

Università di Pavia Archivi Digitali

Controllo Qualità

Riepilogo Fondi Conferimenti Documentazione Contatti Gestione Persone

Controllo qualità e gestione delle fasi operative Vers. 15.3.11

Attività programmate

Fondi

Fondo	Metadati	Numerazione	Questionari	Biografie	Catalogazione	Assistenza
Silvia Avallone	X	X				
Gianrico Carefiglio	X	X				
Paolo Di Paolo						
Dacia Marzini						
Francesca Pecoraro						
Beppe Severgnini	X	X				

Conferimenti

Prep. doc.	Prep. HO	Conferimento	deputato	Test	Cartacce	Archivati
Francesco Pecoraro						
Paolo Di Paolo	13/02/2015	X	X	X	X	X
Silvia Avallone	26/07/2011	X	X	X	X	X
Gianrico Carefiglio	25/05/2011	X	X	X	X	X
Beppe Severgnini	16/10/2010	X	X	X	X	X

Controllo archivi

Integrità Backup Virus Sync

Aiuto

Archiviazione Conferenze Consultazione Quarta Esposizione Archivi

Menù principale

Versione programma

Riepilogo fondi e conferimenti: le caselle con le operazioni principali sono spuntate quando si conclude la compilazione delle relative pagine

Aiuto: guida all'utilizzo del programma

Riepilogo attività programmate: una to-do-list mostra in breve le attività inserite dall'operatore in Gestione > Attività

Riepilogo dello stato degli archivi: i programmi dialogano con il Controllo Qualità e segnalano eventuali problemi tecnici

Dalle schermate successive vedremo in che modo ogni passaggio e ogni fase siano integrati nel sistema QUANDO, e come questo tenti di tenere traccia del maggior numero possibile di informazioni: vedremo principalmente in dettaglio le fasi di gestione dei conferimenti e dei fondi, mentre accenniamo qui brevemente che nel menu “Documentazione” è possibile inserire tutta la documentazione collegata al fondo (contratto, questionario, ...); nel menu “Contatti” c’è la lista dei contatti e le attività a cui sono collegati; nel menu “Gestione” si tiene conto delle attività svolte e pianificate, delle riunioni svolte e pianificate, delle risorse *hardware* e *software* disponibili; nel menu “persone” sono segnati i responsabili e i collaboratori (anche non più in servizio) di PAD.

2. CONFERIMENTI

Contatti:

Di Paolo

Nome autore (collegato a "contatti") ed altre eventuali persone coinvolte (esclusi operatori PAD)

-Numero del conferimento: ordine di arrivo in PAD
- Numero (romano) del conferimento di uno stesso autore

Es.
1. Severgnini (I)
2. Avallone (I)
3. Carofiglio (I)
4. Severgnini (II) → secondo conferimento di Severgnini, quarto conferimento PAD

Riepilogo delle procedure amministrative e tecniche da completare prima del conferimento.

Controlla le procedure tecniche per l'archiviazione dei file e la creazione del "Fondo"

Dettagli del conferimento

Procedure tecniche e amministrative da attivare dopo il conferimento

Campo "note" per una descrizione preliminare dell'archivio conferito.

CONFERIMENTO > Preparazione

Preparazione documenti legali:	12/03/2015	Operatore:	Emmanuela Carbè
Invio guida:	11/02/2015	Operatore:	Emmanuela Carbè
Invio contratto per presa visione:	04/02/2015	Operatore:	Emmanuela Carbè
Invio questionario:	11/03/2015	Operatore:	Emmanuela Carbè
Data ricezione questionario:			

Procedure completate:

Hardware Autore: Mac portatile + HD esterno

Hardware PAD: HD USB con crittatura hardware e formattazione per MAC OS, SSD USB con crittatura software e formattazione per MAC OS

Operatore: Primo Baldini

Procedure completate:

Note:
Il 4 febbraio 2015 è stata inviata a Di Paolo la nuova versione del contratto (l'autore aveva la versione cartacea del contratto precedente).
L'11 febbraio 2015 Carbè ha inviato il questionario e la guida appena ultimati. Dato il poco tempo a disposizione Carbè e Di Paolo hanno deciso che il questionario si sarebbe compilato venerdì 13 alla Casa delle Letterature di Roma, prima del conferimento.
Il 17 febbraio è stato inviato a Paolo Di Paolo il questionario compilato a Roma (l'autore ha chiesto di integrarlo).
La preparazione dei documenti legali è stata molto problematica, Carbè ha consultato più persone dell'ateneo senza ricevere risposte univoche, sollevando a più riprese il problema dell'assoluta urgenza di stabilire nei dettagli le procedure legali del conferimento.
Mercoledì 11 febbraio 2015 Francesco Ottavio Muzzini, dell'Ufficio Legale, ha dato parere favorevole all'utilizzo del nuovo contratto per il conferimento di Di Paolo.
Il 24 febbraio la dott.ssa Giulia Viola ha scritto via e-mail a Carbè che non ha eccezioni da sollevare sul nuovo contratto.
Nicoletta Leone ha avuto ripetuti contatti con l'ufficio legale, in particolare con Giulia Viola, per definire i tempi di preparazione delle procedure legali del conferimento.

Se la procedura è completata il programma segnala la fine del processo nel "Riepilogo" della pagina iniziale.

Se la procedura è completata il programma segnala la fine del processo nel "Riepilogo" della pagina iniziale. Nella preparazione al conferimento la conclusione delle procedure non tecniche è svincolata da quella (due campi).

Procedure tecniche: preparazione degli hard disk di PAD per trasferire i dati (Hardware PAD) in base al computer dell'autore (Hardware autore). Si richiede il nome dell'operatore.

Per ogni procedura non tecnica è richiesto l'inserimento di data e operatore:

- Preparazione documenti legali (contratto/informativa privacy)
- Invio della guida al conferimento
- Invio e ricezione del questionario.

CONFERIMENTO > Conferimento

Preparazione Conferimento Post Conferimento Archiviazione Descrizione

Data: 13/02/2015 Ora: _____
 Luogo: Casa delle Letterature, ROMA

Operatori: Paul Gabriele Weston
 Emmanuela Carbè

Mansione: coordinatore
 custodia e trasporto hd

Hardware: MacBook Pro con sistema operativo Mac OS X 10.8

Problemi: Problemi di alimentazione dell'HD criptato, risolto sostituendo il portatile.

Note:

Data Fine: 13/02/2015 Ora: _____

Conferimento in custodia all'operatore: Emmanuela Carbè

Procedure completate:

Nomi operatori e mansione assegnata

Data, ora dell'inizio operazioni e luogo del conferimento

Data, ora della conclusione delle operazioni

Nome dell'operatore che prende in carico l'hard-disk di PAD e lo trasporta all'Università di Pavia

Se la procedura è completata il programma segnala la fine del processo nel "Riepilogo" della pagina iniziale.

Problemi e note: in caso di criticità tecniche o burocratiche è necessario che l'operatore o gli operatori presenti appuntino in questa sezione tutti i dettagli. Nella sezione "note" si richiede una descrizione delle modalità di conferimento.

Gli hardware da cui provengono il conferimento. Nella sezione Conferimento>Preconferimento è presente l'indicazione dell'Hardware segnalato dall'autore prima del conferimento. È necessario che l'operatore o gli operatori verifichino che non ci siano errori. Il tecnico che prende in carico il materiale si baserà esclusivamente sui dati tecnici indicati in questa schermata.

CONFERIMENTO > Post conferimento

Preparazione Conferimento Post Conferimento Archiviazione Descrizione

CARICO A PAD
 In Data: 16/02/2015 L'operatore: 1
 Depositato a: CASSAFORTE DIP. CHIMICA

procedure completate:

CONTROLLO ARCHIVIO
 Data: 16/02/2015 Operatore: PRIMO BALDINI Programma: PadCleaner
 Note: Rimossi i file nascosti
 Rimosse due cartelle di dati del sistema operativo che inquinavano il conferimento.

MESSA IN SICUREZZA
 Duplicazione archivio temporaneo: SI

procedure completate:

CHIUSURA CONTRATTO
 Invio lista file: 03/03/2015 Programma: PadList
 Convalidazione: _____ Programma: PadPrint
 Protocollo data: _____ Numero: _____ Operatore: _____

procedure completate:

Dettagli sulle prime procedure di trasferimento: data, nome operatore che si occupa della presa in carico e luogo dove viene temporaneamente conservato l'hard disk

Se le procedure sono completate il programma segnala la fine dei processi nel "Riepilogo" della pagina iniziale. In questa fase particolarmente delicata le procedure sono divise in tre parti.

Dettagli sulle procedure di chiusura del contratto. Data invio della lista dei file all'autore (con il nome del software che ha eseguito la stampa), data della convalidazione, data e numero protocollo contratto a cui viene allegata la lista dei file.

Dettagli sulle prime procedure tecniche del post conferimento: data del primo controllo dell'archivio e nome operatore, nome del software utilizzato (inserito automaticamente all'avvio del programma), eventuali note su criticità di file o cartelle. Alla fine della procedura l'operatore crea un archivio temporaneo, in attesa della convalidazione definitiva da parte dell'autore.

CONFERIMENTO > Post conferimento

Preparazione Conferimento Post Conferimento Archiviazione Descrizione

ARCHIVIAZIONE

Programma: **PadArchiver**

Data: Operatore:

Numero file:

Note:

Procedure completate:

Se le procedure sono completate il programma segnala la fine del processo.

In questa sezione viene segnalato automaticamente il software che gestisce l'archiviazione, che comunica al gestionale il numero dei file conferiti. L'operatore, il cui nome va inserito nell'apposita casella, aggiungere la data in cui è stata effettuata l'archiviazione ed eventuali note.

CONFERIMENTO > Post conferimento

Preparazione Conferimento Post Conferimento Archiviazione Descrizione

Prima sommaria descrizione dell'archivio conferito

FONDI

UNIVERSITÀ DI PAVIA Pavia Archivi Digitali Controllo Qualità

Riepilogo **Fondi** Conferimenti Documentazione Contatti Gestione Persone

Nome Beppe Nome Fondo Beppe Severgnini
 Cognome Severgnini Scadenza contratto

1

Contatti & Conferimenti Media Type Normalizzazione Bibliografia Questionario Catalogazione Accessibilità Altri media

Eventuale data di scadenza del contratto

Riepilogo dei contatti e dei conferimenti

Spazio per caricare, visualizzare ed esportare i file con la bibliografia dell'autore. Il programma può visualizzare più bibliografie. Richieste automatiche: data compilazione, nome compilatore.

Campo temporaneamente vuoto

Memorandum sui termini di accessibilità (da remoto/solo presso la sede di PAD; accessibile solo a partire da una certa data)

File con il questionario compilato dall'autore. Si possono caricare più questionari.

Tipologie di file conferiti

Dettagli sulle conversioni dei file

Campo nota: si forniscono notizie su eventuali altri media (interviste video/audio, fotografie) prodotte da PAD nel corso del conferimento.

FONDI > Contatti e conferimenti

Contatti e Conferimenti Media Type Normalizzazione Bibliografia Questionario Catalogazione Accessibilità Altri media

Contatti:

Carofiglio	VAI
Zaninoni	VAI
Canton	VAI

Conferimenti:

Data	Luogo
I 25/05/2011	email

Riepilogo dei conferimenti (data e luogo) associati a un numero progressivo: ogni autore può conferire più fondi. Cliccando la freccia a destra viene visualizzata la pagina del relativo conferimento.

Attenzione: il numero (romano) progressivo non corrisponde al numero del conferimento:
 Es.
 1. Severgnini (I)
 2. Avallone (I)
 3. Carofiglio (I)
 4. Severgnini (II)

Lista dei nomi coinvolti nel conferimento (esclusi gli operatori PAD): autore ed eventuali trami (agenti letterari, operatori di case editrici, etc.). Cliccando il bottone "VAI" si visualizza la scheda dettagliata del contatto.

FONDI > Media Type

Nome: Beppe
Cognome: Severgnini

Nome Fondo: Beppe Severgnini
Scadenza contratto:

Contatti & Conferimenti Media Type Normalizzazione Bibliografia Questionario Catalogazione Accessibilità Altri media

Numero file: 13228
File unici: 7774
Testi Normalizzati: 6900

Media Type	Num. files
BOBO/CWVP	3782
MSWD/WDBN	1884
application/msword	1280
image/jpeg	402
image/gif	67
text/html	64
text/rtf	59
application/pdf	52
SCONOSCIUTI	25
video/quicktime	19
CSOm/TEMP	16
application/vnd.ms-excel	16
txt/TEXT	14
image/tiff	13
audio/x-pn-realaudio	11

Programmi: **PadExtractor**
PadMdsExtractor

Data esecuzione: 05/03/2015

Operatore: Primo Baldini

Procedure completate:

Operatore: **Primo Baldini**

- Annalisa Doneda
- Clelia Martignoni
- Emmanuela Carbè
- Nicoletta Leone
- Paul Gabriele Weston
- Primo Baldini**
- Simone Albonico
- Umberto Anselmi Tamburini

Programmi: **PadExtractor**

- PadArchiver
- PadChecker
- PadCleaner
- PadConverter
- PadExtractor**
- PadList
- PadMdsExtractor
- PadPrint
- PadSync

Riepilogo dei file: numero totale dei file (somma di tutti i conferimenti), numero file unici (potrebbero esserci più copie dello stesso file), numero testi normalizzati

Programmi utilizzati per l'estrazione delle informazioni sui file. Il programma viene selezionato da un menù a tendina. Ogni programma è catalogato nel sistema con le relative informazioni (libreria usata, versione, sistema operativo, data installazione)

Se la procedura è completata il programma segnala la fine del processo nel "Riepilogo" della pagina iniziale.

Riepilogo tipologia dei file (testo, immagine, applicazione, audio, video, etc.) e conteggio dei file che appartengono alle diverse tipologie

FONDI > Normalizzazione

DOCUMENTI

Sono convertiti i file con Media Type:

- application/msword
- MSWD/WDBN
- BOBO/CWVP
- text/rtf

Contatti e Conferimenti Media Type Normalizzazione Bibliografia Questionario Catalogazione Accessibilità Altri media

Programma: **PadConverter**

Data esecuzione: 04/03/2015

Operatore: Primo Baldini

Test dei risultati: ✕

Procedure completate:

Tipo di programma utilizzato (per ora solo PadConverter), data esecuzione del programma e operatore che ha svolto la procedura

Se la procedura è completata il programma segnala la fine del processo nel "Riepilogo" della pagina iniziale.

Elenco tipologie di materiale normalizzato

Le altre tre schede, cioè “catalogazione”, “accessibilità”, “media” offrono rispettivamente la possibilità di “catalogare” o meglio, collegare (come si vedrà più avanti nella descrizione di PAD Manager) i vari file o documenti (e ne vedremo la differenza) al rispettivo livello di descrizione previsto e organizzato dal sistema, segnalare i vincoli temporali o limitazioni all’accesso di qualsiasi tipo, e collegare materiali audio o video realizzati in occasione del conferimento a PAD.

PAD Manager

L’applicativo principale utilizzato nella gestione tecnica dei conferimenti è PAD Manager, che al proprio interno contiene anche il modulo PAD Archiver: i nomi sono sufficientemente autoesplicativi rispetto alle funzioni per cui sono stati creati, ma permettono una interessante serie di possibilità che sono state via via aggiornate nel corso del tempo (e che continuamente vengono implementate), in base ai suggerimenti sull’utilizzo provenienti principalmente da Emmanuela Carbé. Anche per questa sezione è previsto il riutilizzo di schermate commentate, ma andremo a spendere anche qualche parola in più relativamente ad alcune scelte o ad alcune riflessioni che sono nate dall’uso pratico e quotidiano di questi applicativi. Il mercato offrirebbe diversi *software* e soluzioni per svolgere operazioni di DAM (*Digital Asset Management*), ma lo sviluppo interno di questa applicazione è stato un passaggio quasi obbligato: nessuna delle soluzioni commerciali o *open source* offriva gli strumenti, la flessibilità, oppure non rispondeva in buona parte alle esigenze e alle necessità di PAD, che se da un lato deve essere tecnicamente fruibile e flessibile rispetto alle funzioni per un uso quotidiano da parte degli operatori, deve anche rispondere (almeno con una conformità parziale) agli standard di settore.

PAD (Asset) Manager è l’applicazione unica di PAD che permette di integrare diverse funzionalità al proprio interno:

- L’attivazione e la disattivazione dei moduli del sistema;
- La possibilità di inserire e collegare la biografia dell’autore;

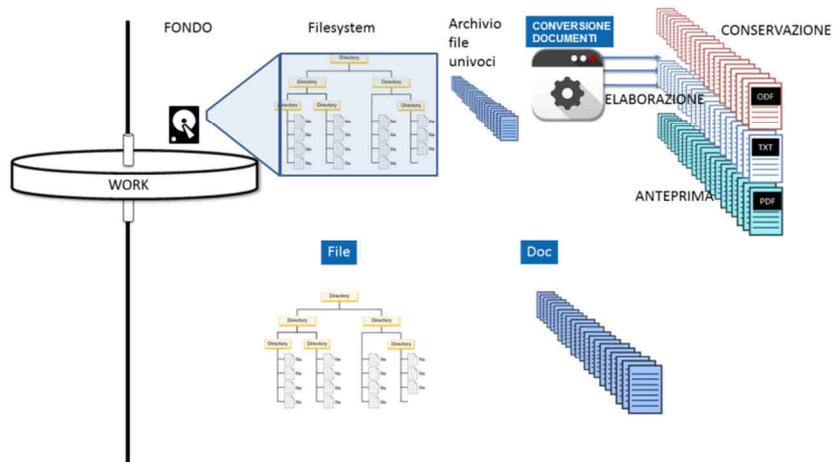
- La possibilità di inserire e collegare la bibliografia dell'autore (operazione importante rispetto all'organizzazione dei documenti come rispettivamente relativi a un'opera);
- La possibilità di descrivere gli oggetti;
- La navigazione attraverso il fondo e attraverso la sua struttura;
- L'anteprima del documento (dalla conversione in formato TXT);
- La ricerca generale su testi e metadati;
- La possibilità di aggiungere ai documenti e ai file TAG, Classi e Bookmark (archiviare, verificare, ...);
- La gestione del *database*;
- La creazione di documenti in formato PDF criptato (la lettura del file integrata – dalla conversione in PDF e TXT offre alcuni vantaggi: il controllo dei permessi sulla lettura, l'impossibilità di copiare il documento da parte dell'utente, e un passaggio rapido al documento successivo);
- L'esportazione dei dati.

PAD Archiver, invece, è lo strumento interno a PAD Manager che permette la gestione potremmo dire “strutturale” dei file presenti nel conferimento, e offre queste funzioni:

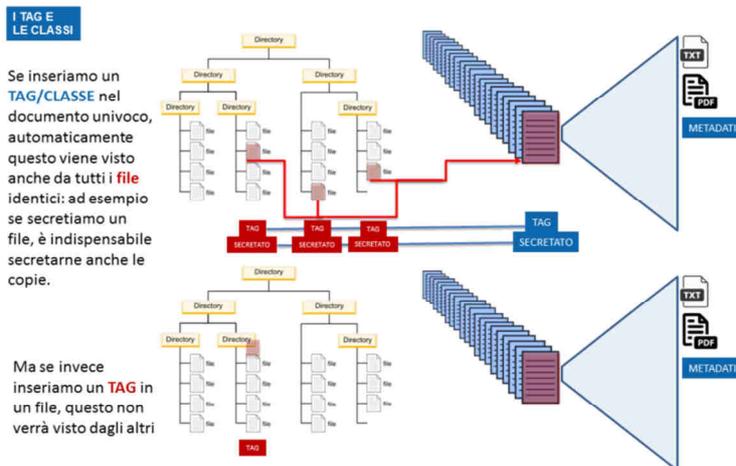
- Visualizzazione file duplicati (poiché si è scelto di mantenere i file duplicati in posizioni differenti, come documenti con funzioni differenti);
- Spostamento rapido tra file duplicati (permesso dal fatto che i file duplicati hanno lo stesso *checksum*);
- Visualizzazione *directory* duplicate;
- Assegnazione multipla in base alle righe selezionate;
- Sostituzione di TAG;
- Visualizzazione dei metadati;
- Statistiche e conteggi vari;
- Assegnazione di un oggetto definito dalla procedura “Project”;
- Utilizzo di *tool* esterni (esempio Voyant Tools);
- Assegnazione di più oggetti relazionati;

- Inserimento rapido degli oggetti senza la procedura “Project”;
- Assegnazione di *bookmark*.

Prima di visualizzare alcune schermate commentate di PAD Manager e PAD Archiver, vale la pena aggiungere alcune spiegazioni rispetto alla distinzione tra file, o *item*, e documento. Come schematizzato nell’immagine seguente, il documento e il file risultano due entità distinte.

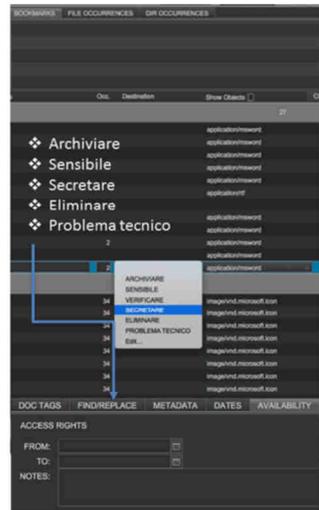
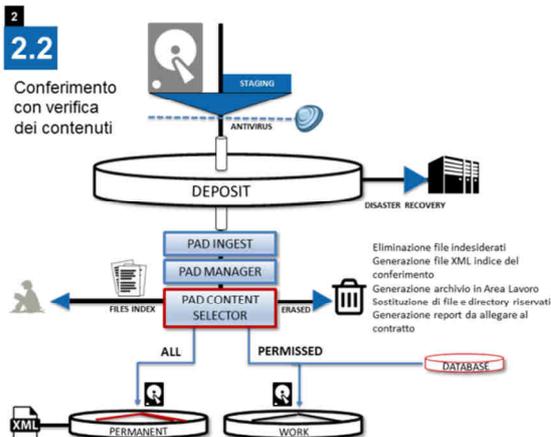
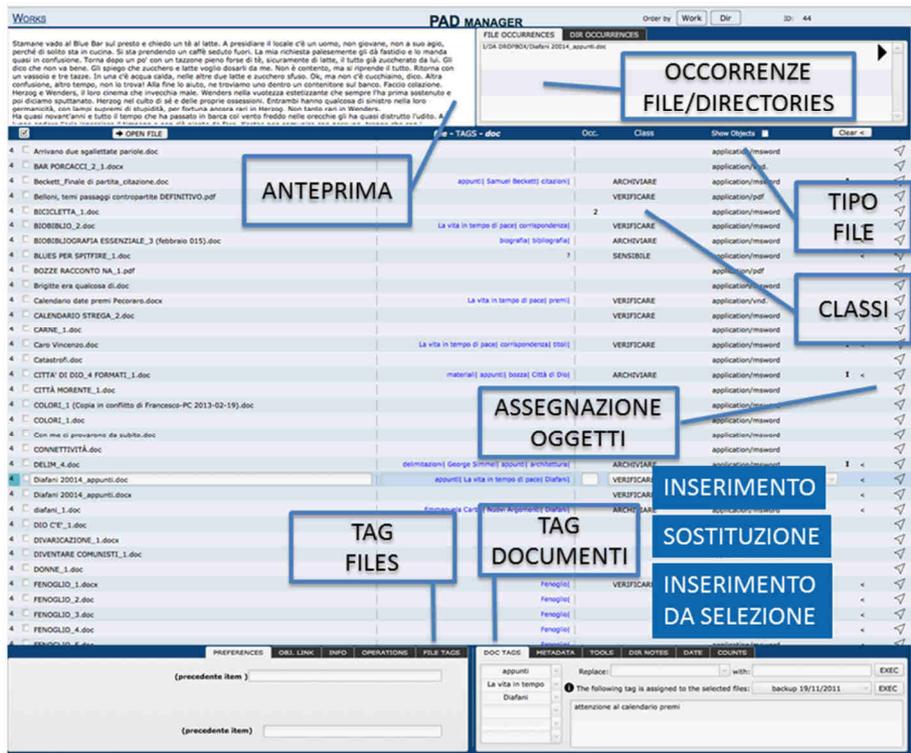


La differenza tra documento e file, o item, viene efficacemente spiegata nell’immagine successiva: nell’area di lavoro vengono copiati i file archiviati nell’area per la conservazione a lungo termine, e vengono conseguentemente convertiti (o vengono convertiti prima, nel caso passino prima per l’area di lavoro e poi in quella della conservazione permanente); le *directory*, e quindi la struttura del conferimento, vengono mantenute intatte; vengono prodotti i *checksum* e quindi vengono evidenziati i file identici.



La scelta di distinguere tra documento e file è piuttosto importante: dal momento che in un archivio di questo tipo (cioè un archivio d'autore), non è infrequente trovare documenti identici (ad esempio: il testo della stessa mail mandata a più editori, la bozza di un testo da mandare a più editor; e in ogni caso ciascuna copia salvata nella cartella relativa alla documentazione che riguarda il destinatario) che però testimoniano un processo o un rapporto lavorativo, e quindi non sono da scartare. Pertanto la distinzione riguarda l'aspetto logico e quello numerico: ciascun file è un file singolo, e rappresenta una unità numerica nel conteggio finale, e ciascun file può essere rispettivamente marcato con un *tag* rispetto ad un altro (tutti i testi preparati per un editore, tutti i documenti per una certa circostanza); in questo senso i *tag* sono utili e versatili in quanto non hanno un vocabolario controllato e sono sufficientemente flessibili – se usati con coscienza – da permettere una marcatura che, per quanto non possa sostituirsi alla descrizione, agevola notevolmente le operazioni più immediate. Nel caso in cui invece un *tag* venga inserito in un documento verrà automaticamente riportato su tutti i file identici, poiché verrà considerato come inerente il contenuto del documento e non quella specifica copia. Analogo il comportamento delle classi (Archiviare, Sensibile, Secretare, Eliminare, Problema tecnico), che si applicano ai documenti, per agevolarne il trattamento dall'area di deposito, e concludere la lista dei file da mandare all'autore per la conferma.

Dall'immagine successiva cominciamo a vedere propriamente PAD Manager e le sue funzionalità: possiamo vedere in alto a sinistra il box dell'anteprima in TXT, utile per poter vedere il contenuto del documento senza doverlo aprire, e alla destra il box con le occorrenze dei file e delle *directory*; nella finestra principale vediamo l'elenco dei documenti contenuti, il tipo di file, le classi assegnate a quel documento e gli oggetti (cioè i file con lo stesso *checksum*) assegnati a quel documento. In basso a sinistra invece troviamo il box per inserire i *tag* sul singolo file (non è l'unica funzione prevista, e sono diverse per file e documento), senza riferirli a tutti i documenti uguali, e a destra il box per assegnare *tag* ai documenti, e altre funzioni generali sul contenuto (e quindi riferibili al documento).



Dall'immagine precedente possiamo invece notare come opera il modulo "PAD Content Selector" nella fase di deposito: è il modulo che permette l'eliminazione dei file indesiderati, genera la lista in XML dei file conferiti, quindi genera l'archivio in area di lavoro, rende invisibili i file e le *directory* riservate e genera i *report* da allegare al contratto.

La descrizione del contenuto, in PAD, è un momento che ha richiesto una riflessione sugli standard di settore, e sull'opportunità di esservi conformi o meno: in linea con l'interoperabilità (che però riguarda più i formati che gli standard, e la comunicazione dei dati che il contenuto), oppure più in linea con la praticità (che in base al contesto può anche essere slegata rispetto agli standard)? Posto che una descrizione archivistica vera e propria è un processo opportunamente lungo, ma che le necessità di PAD sono più celeri, e posto che il materiale "descritto" è in relazione principalmente (quando non esclusivamente) con materiale bibliografico, inizialmente si è guardato all'utilizzo di FRBR_{ER} e alle sue classi: uno standard bibliografico, quindi, per dare una prima e sommaria descrizione dei contenuti dell'archivio, basata sulla bibliografia dell'autore, e sulle relazioni tra i documenti e le opere a cui si riferiscono. La rigidità di FRBR_{ER}, però non è stata ritenuta ottimale per le necessità di PAD, e si è preferito rivolgersi alla struttura più flessibile di FRBR_{OO}. Tuttavia (queste ultime modifiche sono state operate circa tra luglio e agosto 2016), per ovvi motivi di contenuto (non necessariamente o esclusivamente bibliografico), se pure il modello di FRBR_{OO} è più funzionale rispetto alle necessità di PAD, la conformità con lo standard non è mai del tutto completa: perciò si è deciso di mantenere la struttura di FRBR_{OO} come ispirazione per la "descrizione" dei documenti, ma al contempo dichiarando apertamente una non completa conformità con il modello. Così, nella schermata seguente possiamo vedere i quattro livelli presenti anche in FRBR_{ER/OO}, cioè *Work*, *Expression*, *Manifestation*, *Item*, mutati rispettivamente in *Project*, *Object*, *Class*, *Item* che sono al livello concettuale più astratto i corrispettivi analoghi: *Project* è la creazione intellettuale astratta a cui fa capo una determinata opera; *Object* è la realizzazione (o tutte le realizzazioni) di quel *Project*, *Class* è una delle parti che compongono *Object*, mentre *Item* è il file specifico collegato a quella classe per quell'*Object*. Nelle schermate successive vedremo alcuni comportamenti del programma rispetto ai livelli.

1. PROJECT *creazione intellettuale astratta* Italiano in America

2. OBJECT *realizzazione*

- Un Italiano in America
- An italian in America
- Documentazione USA

3. CLASS

- **Bozza 1**
- Bozza 2
- Definitiva
- Stampa
- Corrispondenza
- Recensioni
- Vendite

4. ITEMS

FILES

- ****introduzione**
- **indice**
- **1aprile**
- **2maggio**
- **3giugno**

RELATED FILES

- **Articolo corriere**
- **Articolo corriere**
- **Articolo corriere**

DIR //BPSA/1/loDonna/...

LINK http://.....

GENERAZIONE OGGETTO

Project Italiano in America **1. DIRETTO**

Object Un Italiano in America

Notes

Type

Wiki

2. DALLA BIBLIOGRAFIA

MENU ARCHIVER PROJECT BIBLIO **Wikipedia Citation Styles**

Template: libro

Title: **Aureli fatto la fine di Turing**

Author:

Year: 2015

Language:

Volume:

Opera:

OCOLC:

Id:

Cid:

Export

From selected row

Editor:

ISBN:

URL:

Edizione:

Capitolo:

URL capitolo:

Page:

Page:

Page:

Position:

Citation:

Data Accesso:

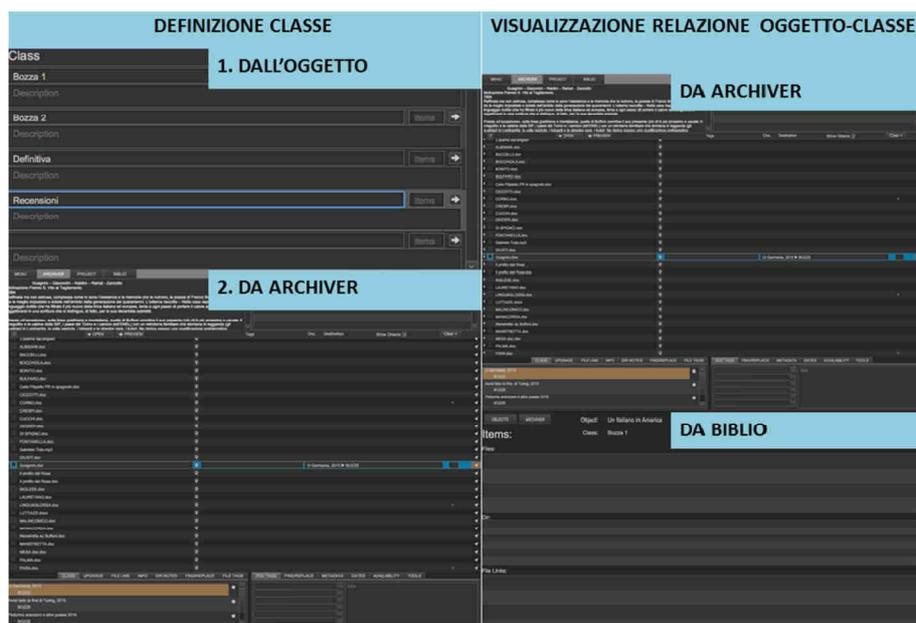
URI archivio:

Data archivio:

URI morto:

Object: Aureli fatto la fine di Turing, 2015

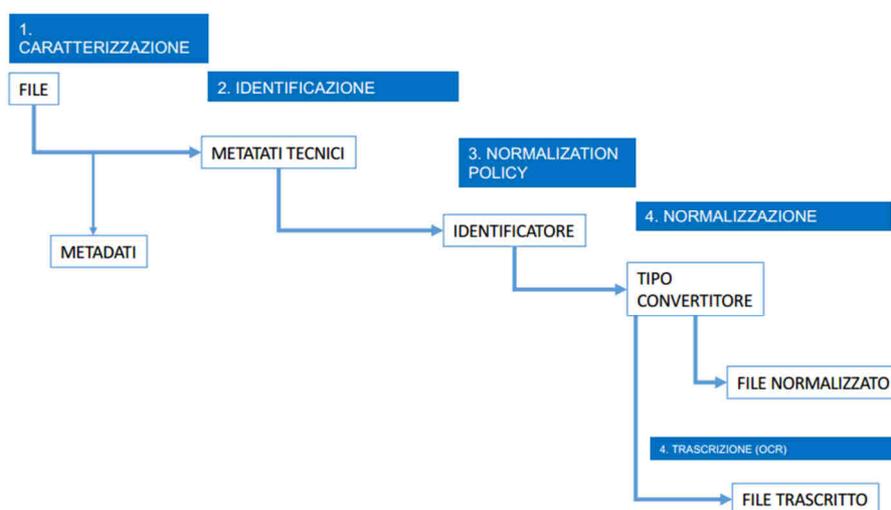
Create Object from this citation



Scendendo sempre più nel dettaglio degli aspetti tecnici, rendiamo conto di alcune implementazioni che sono state fatte nel corso del tempo in PAD, per rendere ancor più evidente quanto questo sia effettivamente un processo, tanto per le componenti che concettualmente, che può sempre evolvere e migliorare. Alcuni degli aspetti tecnici più rilevanti (oltre alle scelte operate su FRBR_{ER} prima, e FRBR_{OO} poi, a cui abbiamo già accennato, e all'implementazione dei vari sistemi) riguardano i passaggi che partono dalla caratterizzazione per arrivare fino alla normalizzazione. Teniamo comunque in considerazione che i conferimenti non sono mai a ritmo costante e sono intervallati da periodi di inattività; le operazioni di archiviazione possono essere attivate solo da un amministratore, che deve avviare la definizione di un fondo e di una sequenza ben precisa di operazioni; non è prevista una gestione multiutente dell'archiviazione su uno stesso fondo, mentre per fondi diversi l'operazione sarebbe possibile; infine, i materiali inseriti non devono essere necessariamente disponibili subito alla consultazione (bisogna compiere prima una serie di operazioni: quarantena, eliminazione dei file nascosti, controllo visivo, normalizzazione).

Prima del 2015, le prime versioni del sistema di normalizzazione erano complesse per la presenza di macchine virtuali che gestivano la

conversione dei file obsoleti e in conseguenza di ciò c'era la necessità di spostare manualmente i risultati delle conversioni negli archivi di conservazione. Inoltre non c'era nemmeno una vera gestione della caratterizzazione, ma il sistema si basava solamente sul metadato *mediatype*. Perciò, nei fondi più vecchi, la percentuale di file convertiti era molto bassa, in quanto mancavano strumenti adeguati per la caratterizzazione dei file (il metadato *mediatype* non è minimamente sufficiente): **a partire dal 2015** la situazione è migliorata a seguito della disponibilità di librerie di caratterizzazione (questo può essere considerato un esempio pratico del fatto che per i formati più obsoleti – se non sono presenti gli strumenti adatti alla lettura e alla corretta interpretazione del formato stesso – può essere che in futuro qualcuno sviluppi gli strumenti adatti). Si tenga presente che progetti come Archivematica o DAITSS – Florida Digital Archive hanno usato schemi per l'estrazione di metadati simili a quelli elaborati da PAD. I passaggi che vedremo in dettaglio in questo modulo sono in ordine: la caratterizzazione, l'identificazione, la *format policy*, la normalizzazione e la trascrizione (OCR).



- La **caratterizzazione** è il processo di estrazione dei metadati tecnici contenuti all'interno di un file. I metadati tecnici sono importanti attributi per la comprensione e la gestione dei fondi archivistici digitali, in particolare per la ricerca di procedure di trasformazione del formato originale del file in formati standard. Per caratterizzare un file si usano dei *tool* specifici: dopo

approfondite sperimentazioni si è arrivati alla conclusione che un solo *tool* non è sufficiente per caratterizzare file molto diversi (testi, immagini, audio, applicativi proprietari). PAD, in questo momento, utilizza le seguenti librerie: Apache Tika, Exiftool, OIS Audio Information e un *tool* sviluppato da PAD per la gestione dei vecchi file Apple Mac.

Nella versione 2017 di PAD sono state risolte alcune criticità che erano presenti in questa fase: questi *tool* di caratterizzazione producono metadati ma non esiste uno standard comune né per chi crea questi metadati, né per chi deve estrarli, in quanto non c'è per forza una corrispondenza tra l'ordine o il tipo di metadati estratti con due o più *tool* differenti (nella pagina seguente si troverà uno schema comparativo ad esemplificare), e non sempre sono certi: potrebbero venire alterati da vari passaggi (ad esempio da un sistema operativo a un altro) oppure da impostazioni non corrette, come il nome dell'autore (registrato dal *software*) che non corrisponde con l'autore effettivo; inoltre, questi strumenti devono essere aggiornati spesso per consentire l'identificazione di un numero maggiore di tipologie.

Per estrarre le informazioni importanti ai fini della caratterizzazione e per di ottenere dei metadati tecnici affidabili la Harvard Library ha sviluppato un sistema chiamato FITS¹²⁷ (*File Information Tool Set*) che esegue quest'operazione e ha come risultato un unico file XML: questo sistema, nato più o meno nello stesso periodo del progetto PAD, ha una impostazione simile, anche se con scopi diversi; infatti, mentre FITS è sviluppato per interagire con altri *repository*, il sistema di PAD è semplicemente un modulo del sistema.

¹²⁷ Il *File Information Tool Set* (<http://projects.iq.harvard.edu/fits/home>) identifica, convalida ed estrae metadati tecnici per una vasta gamma di formati. Utilizza diversi *tool open source* (ADL Tool, Apache Tika, DROID, Exiftool, File Utility, Jhove, MediaInfo, National Library of New Zealand Metadata Extractor, OIS Audio Information, OIS File Information, OIS XML Information) e produce un output in XML. Non è da confondere con il *Flexible Image Transport System* (<https://fits.gsfc.nasa.gov/>), formato per la conservazione delle immagini digitali mantenuto dalla NASA e utilizzato anche dalla biblioteca vaticana (Allegrezza 2011).

TIKA	EXIF TOOL
<ul style="list-style-type: none"> • Application-Name: Microsoft Macintosh Word • Application-Version: 15.0000 • Author: Microsoft Office User • Character Count: 2955 • Character-Count-With-Spaces: 3467 • Content-Length: 26674 • Content-Type: application/vnd.openxmlformats-officedocument.wordprocessingml.document • Creation-Date: 2016-11-13T12:56:00Z • Last-Author: baldini@unipv.it • Last-Modified: 2016-11-15T15:01:00Z • Last-Printed: 2016-11-15T11:47:00Z • Last-Save-Date: 2016-11-15T15:01:00Z • Line-Count: 24 • Page-Count: 3 • Paragraph-Count: 6 • Revision-Number: 5 • Template: Normal.dotm • Total-Time: 110 • Word-Count: 518 • X-Parsed-By: org.apache.tika.parser.DefaultParser • X-Parsed-By: org.apache.tika.parser.microsoft.OOXMLParser • X-TIKA:digest:MD5: b269da367cc7fbd19975f12a9112962d • X-TIKA:digest:SHA256: 4dcd07c77debc798d93552e43696a616e66110d95397b1bcb828232f36fa3ad2 • cp:revision: 5 • creator: Microsoft Office User • date: 2016-11-15T15:01:00Z • dc:creator: Microsoft Office User • dcterms:created: 2016-11-13 T12:56:00Z • dcterms:modified: 2016-11-15 T15:01:00Z • extended-properties:AppVersion: 15.0000 • extended-properties:Application: Microsoft Macintosh Word • extended-properties:Template: Normal.dotm • extended-properties:TotalTime: 110 • meta:author: Microsoft Office User • meta:character-count: 2955 • meta:character-count-with-spaces: 3467 • meta:creation-date: 2016-11-13T12:56:00Z • meta:last-author: baldini@unipv.it • meta:line-count: 24 	<ul style="list-style-type: none"> • ExifTool Version Number: 10.33 • File Name: 10-Workflow.docx • Directory: . • File Size: 26 kB • File Modification Date/Time: 2016:11:15 16:01:53+01:00 • File Access Date/Time: 2016:11:15 17:37:49+01:00 • File Inode Change Date/Time: 2016:11:15 17:25:04+01:00 • File Permissions: rw-r--r-- • File Type: DOCX • File Type Extension: docx • MIME Type: application/vnd.openxmlformats-officedocument.wordprocessingml.document • Zip Required Version: 20 • Zip Bit Flag: 0x0006 • Zip Compression: Deflated • Zip Modify Date: 1980:01:01 00:00:00 • Zip CRC: 0x36e18bfb • Zip Compressed Size: 386 • Zip Uncompressed Size: 1965 • Zip File Name: [Content_Types].xml • Title: • Subject: • Creator: Microsoft Office User • Keywords: • Description: • Last Modified By: baldini@unipv.it • Revision Number: 5 • Last Printed: 2016:11:15 11:47:00Z • Create Date: 2016:11:13 12:56:00Z • Modify Date: 2016:11:15 15:01:00Z • Template: Normal.dotm • Total Edit Time: 1.8 hours • Pages: 3 • Words: 518 • Characters: 2955 • Application: Microsoft Macintosh Word • Doc Security: None • Lines: 24 • Paragraphs: 6 • Scale Crop: No • Heading Pairs: Title, 1

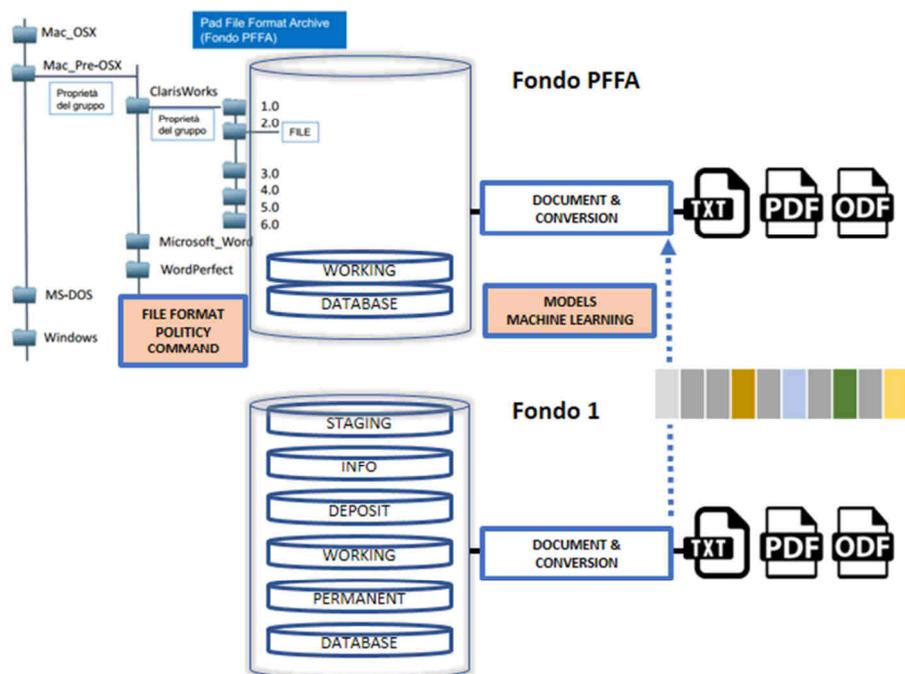
- **L'identificazione** ha il compito di identificare in tipo di file che abbiamo di fronte poiché noi ne conosciamo i metadati e l'identificatore, ma non sappiamo ancora come trattarlo. Principalmente PAD utilizza due strumenti a disposizione dai National Archives: DROID¹²⁸ (*Digital Record Object Identification*) e il registro tecnico di PRONOM¹²⁹.
- **La Normalization Policy**, ovvero quella che potremmo definire la procedura di conversione, è in sostanza una tabella di conversione tra l'identificatore (in PRONOM) del tipo di file e l'applicativo da utilizzare per la conversione, o le eventuali procedure più complesse da attivare: attraverso queste corrispondenze il sistema potrebbe automaticamente valutare in maniera ben più rapida le procedure da attivare per arrivare alla conversione (in TXT, PDF/A-1 e ODF) e alla trascrizione con OCR (il procedimento dall'ID all'applicativo, è uguale a quello per la conversione).

Nel **luglio 2016**, all'avvio del lavoro di descrizione del fondo Pecoraro ci si è resi conto che mancava la normalizzazione per alcuni tipi di file, che ne impediva quindi la visualizzazione in anteprima (in formato TXT). Il problema era stato generato dall'aggiornamento di una libreria che aveva cambiato un metadato, portando conseguentemente alcuni errori nell'identificazione; benché questo errore sia stato riconosciuto e risolto abbastanza rapidamente, è stato causa di un notevole stress, e si è voluto migliorare il controllo sulla fase di normalizzazione: non potendo, ad ogni aggiornamento, simulare la conversione di tutti i fondi, si è pensato di creare il PAD *File Format Archive* o fondo PFFA: il fondo PFFA non è un conferimento d'autore ma un insieme di documenti, immagini e file audio, video, eseguibili, etc... raccolto nel corso degli anni e raggruppato per sistema operativo, programma che li ha generati e versione del programma; in questo modo ogni gruppo può contenere le indicazioni per stabilire le corrette operazioni di normalizzazione per la propria tipologia di file o

¹²⁸ <http://digital-preservation.github.io/droid/>

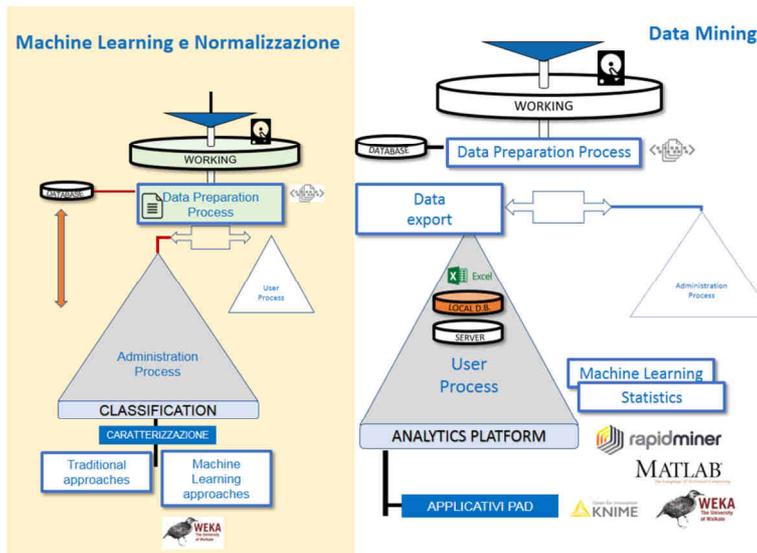
¹²⁹ <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

documento (e la figura seguente schematizza la struttura del fondo PFFA).



L'archivio PFFA è stato parzialmente realizzato con un archivio di test delle prime versioni del progetto, ma successivamente si è scelto di mettere a punto un sistema di generazione automatica (quando si parla di decine di migliaia di file non è più un lavoro che si può fare a mano in un tempo utile): con questo sistema, quando il programma di caratterizzazione incontra una nuova classe di file lo copia nel fondo PFFA, in modo tale che possa essere inserito tra le tipologie note e quindi tra le procedure conosciute. Il fondo può essere elaborato come tutti gli altri e raggruppa in sé tutte le problematiche di conversione e le tipologie di file individuate nel progetto PAD.

Si è deciso quindi di implementare per il 2017 l'utilizzo di algoritmi di *Machine Learning* (ML) e tecniche di *Data Mining* (DM) all'interno di questa specifica procedura, in modo tale da migliorare l'approccio alla classificazione dei file. Per testare questa funzione si è scelto di utilizzare il sistema di ML MatLab con gli algoritmi e i modelli di classificazione forniti già nel sistema, oltre ai sistemi di DM Weka, Rapidminer e Knime.



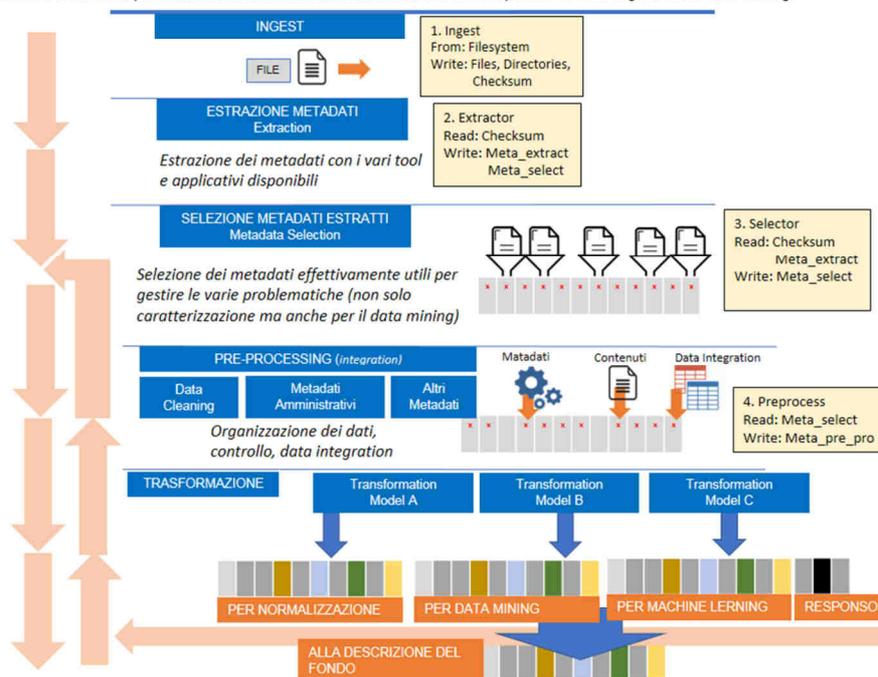
Dopo un primo periodo in cui è necessario “allenare” l’algoritmo con un set di dati noti, il risultato che ci si attenderebbe è quello di fare in modo che in automatico la procedura di ML riesca a classificare i vari file: se il tipo di file è noto ed esiste nel fondo PFFA (combaciano tutti i metadati individuati come significativi a stabilire che sia effettivamente lo stesso tipo di file); se solo alcuni dei metadati individuati come significativi combaciano, e quindi il file si può ritenere simile, il file viene copiato nel vecchio fondo di prova di PAD e si avvia il test automatico di verifica; se invece il file non esiste nel fondo PFFA viene copiato nel fondo di prova e vengono effettuati alcuni controlli manuali.



Dalla figura seguente possiamo vedere schematizzata la procedura per il trattamento dei dati, nelle varie fasi, attraverso questi sistemi automatici: dall'acquisizione iniziale, passando per l'estrazione dei metadati, poi la selezione dei metadati utili per la caratterizzazione e il DM, organizzazione e il controllo dei dati nella fase di *preprocessing* volta all'integrazione dei dati in PAD e, infine, il passaggio dei dati attraverso algoritmi di trasformazione rispetto agli obiettivi (ML, DM, normalizzazione).

1 - Processo di preparazione e trasformazione dei dati

Insieme di tecniche per inizializzare correttamente i dati utilizzati come input ad un certo algoritmo di Data Mining



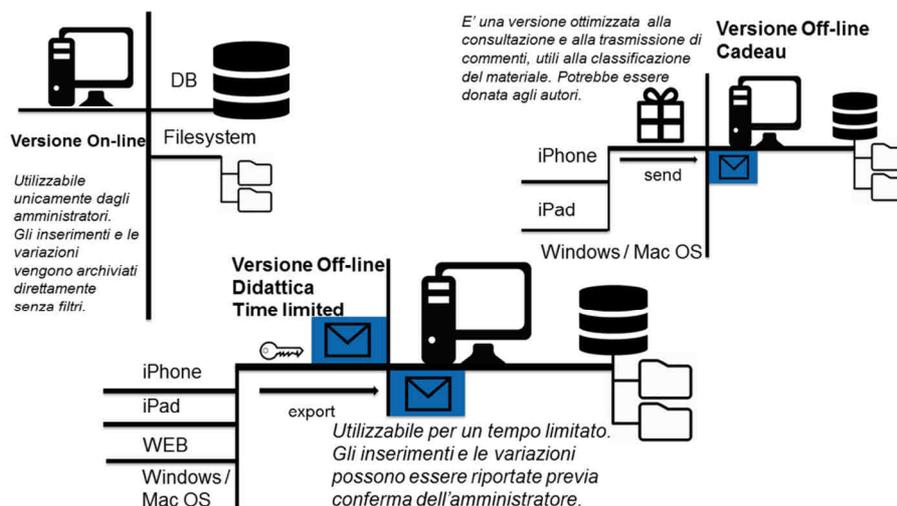
Rispetto ai risultati attestati, tuttavia, le performance ottenute finora non sembrano essere abbastanza soddisfacenti, ma si sta ancora lavorando per migliorarne la resa: uno dei problemi principali riguarda la significatività dei metadati, cioè il fatto che un metadato o un insieme di metadati garantiscano che quel file è proprio in quel formato; naturalmente è da escludere che un vaglio di questo tipo possa essere fatto da un operatore, poiché vengono estratti più di quattrocento diversi metadati dalle librerie di estrazione, e un utente si limiterebbe a guardarne appena qualcuno. Un importante vantaggio dell'integrazione di questi due sistemi, cioè il fondo PFFA e il sistema di ML e DM è che nel momento in cui si cambiassero o si aggiornassero librerie o

convertitori, basterebbe provare a convertire e controllare il fondo di test per verificarne il corretto funzionamento. È preferibile non convertire i documenti in tempo reale e solo quando vengono richiesti, ma già da subito, perché di volta in volta bisognerebbe attendere il termine del processo di conversione e il relativo trasferimento di dati e in alcuni casi, dovendo controllare molti testi quest'operazione appesantirebbe troppo il carico di lavoro; inoltre convertendo i documenti fin da subito si ha tutto il tempo di verificare la qualità della conversione ottenuta. In ogni caso è sempre opportuno ripetere le conversioni a distanza di tempo e ne abbiamo già potuto valutare alcuni benefici (come la disponibilità di strumenti adatti alla conversione che prima non c'erano): i convertitori, infatti, non sono perfetti, e la stessa conversione non sarà mai identica all'originale. PAD cerca continuamente il meglio della tecnologia disponibile, che è in continua evoluzione, e avere una conversione più accurata a distanza di qualche tempo può aiutare l'operatore, oltre al fatto che font, grafici, immagini mettono a dura prova i convertitori e avendo decine di migliaia di file è impensabile un intervento manuale per ogni conversione. Con questi sistemi la possibilità di generare procedure di classificazione e descrizione del fondo con all'interno tutti i testi da analizzare è un sistema che, una volta messo a punto il funzionamento, è completamente automatico. Tra gli sviluppi, almeno ipotetici e presi in considerazione, una volta messo a punto il sistema, si possono facilmente generare altri modelli di classificazione che consentano ulteriori elaborazioni che non riguardano la parte di caratterizzazione, ad esempio: la semplificazione della struttura del conferimento o l'individuazione versioni.

La didattica in PAD

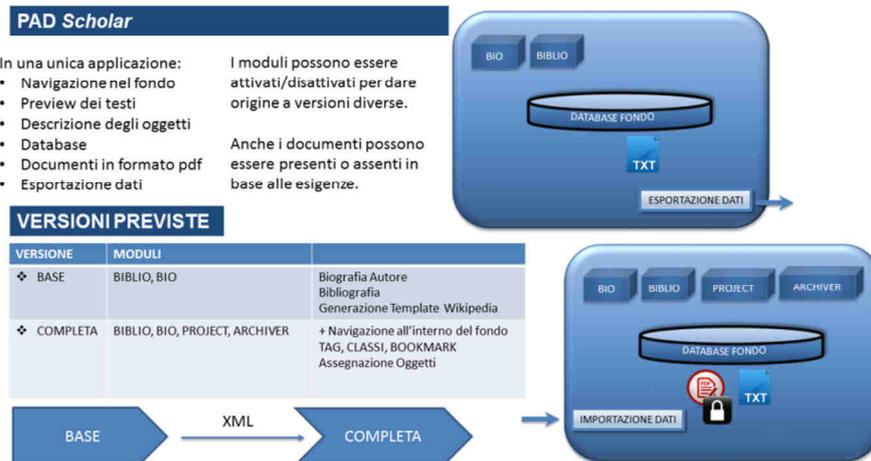
Dall'anno accademico 2016/2017 si è pensato di integrare le attività di PAD anche con la didattica: di fronte a un sistema formativo (quello italiano) in cui è piuttosto presente lo scollamento tra la formazione e il mondo del lavoro – a meno di intraprendere un percorso professionale

nel mondo accademico – PAD non propone certo un sistema risolutivo, ma offre l’opportunità di supportare la didattica con delle ore di laboratorio utilizzando gli applicativi reali e di fatto contribuendo ai dati dell’archivio.

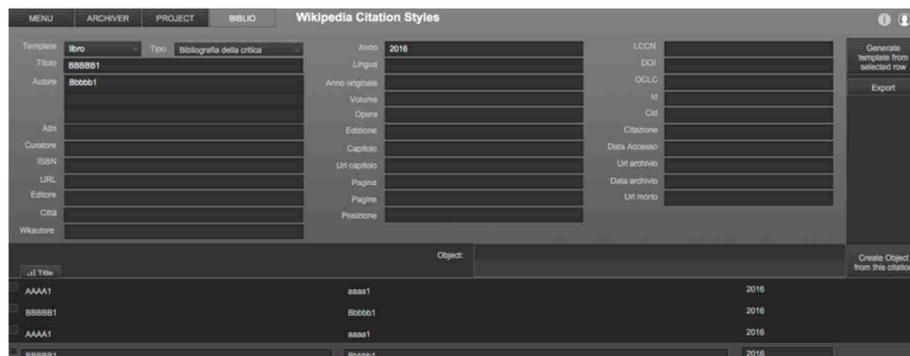


Naturalmente l’accesso agli studenti non viene garantito integralmente e con tutti i permessi (ricordiamo comunque che per motivi di sicurezza in PAD non esiste un unico amministratore con tutte le autorizzazioni, ma solo i permessi di diversi amministratori costituiscono l’accesso completo), ma in modalità controllata: PAD ha – per il momento – tre tipologie di accesso diverso: una versione online, che è quella per così dire “ufficiale” attiva e funzionante, ospitata sui server di ateneo, ed è quella a cui ha accesso il personale interno di PAD; una versione per la didattica che è offline, e a cui è previsto un accesso limitato nel tempo (ci si può connettere solo quando è aperto quel canale, cioè durante gli orari di laboratorio e previa autenticazione) con le dovute restrizioni per gli utenti/studenti e i cui dati – una volta controllati e validati dal tutor/docente – possono essere inseriti nella versione “ufficiale”; infine esiste una terza versione offline, chiamata “cadeau”, destinata agli autori e contenente per ciascuno il proprio conferimento: sarebbe utile poter ricevere commenti o l’ampliamento di alcuni dati direttamente dai creatori dei documenti.

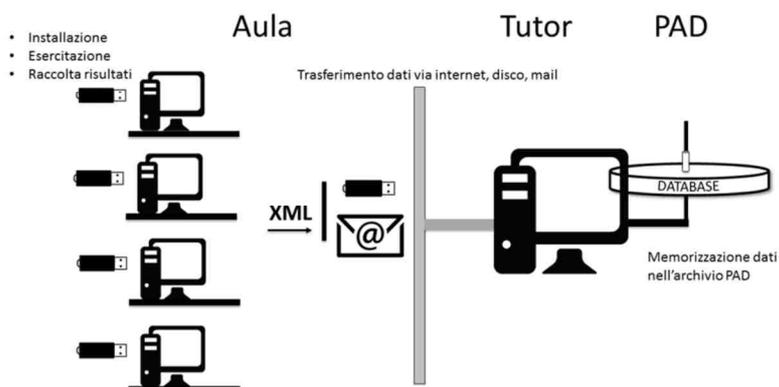
L'applicativo che permette agli studenti di accedere a PAD e imparare a capirne le logiche e il funzionamento è PAD Scholar, di cui esistono due versioni, con differenti limitazioni:



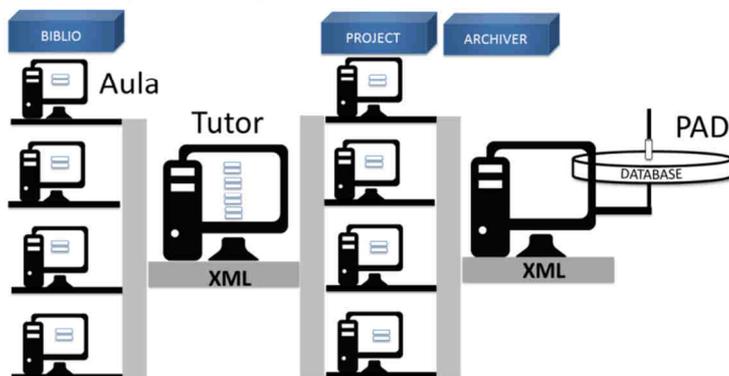
La versione di base permette la possibilità per gli studenti di andare a compilare la biografia e la bibliografia completa di un determinato autore (utilizzando – per ora – un *template* per i libri sul modello di Wikipedia, che potrebbe essere implementato in futuro con una scheda più conforme rispetto agli standard bibliografici):



Le esercitazioni, operabili solo in laboratorio informatico e durante gli orari di lezione con il collegamento permesso da PAD, vengono poi raccolte, controllate e verificate dal docente, e possono infine essere inviate a PAD per essere inserite effettivamente nei dati disponibili:



I dati raccolti nella prima esercitazione, possono essere raccolti elaborati e riutilizzati nella seconda



La possibilità di integrare il lavoro di laboratorio, cioè di operare tecnicamente con determinate tecnologie, con la parte più teorica dell'insegnamento, per gli studenti è un indubbio vantaggio, in quanto questo può garantire a loro la possibilità di approcciare nella pratica problematiche e strumenti che rischierebbero di rimanere solo nozioni teoriche, a cui manca la capacità di incidere nell'esperienza quotidiana.

In ultima battuta, vista la contiguità tematica (una contiguità che è anche territoriale) con il Fondo Manoscritti di Autori moderni e Contemporanei, di cui PAD si potrebbe considerare come il naturale prosieguo digitale e benché attualmente non siano state istituite collaborazioni, non si può escludere in via assoluta che in futuro possano stringersi dei rapporti anche più solidi, in questa fase di transizione tra il cartaceo e il digitale: solo a titolo d'esempio, il poeta Franco Buffoni ha conferito a PAD la parte digitale del proprio archivio, mentre al Fondo Manoscritti quella cartacea.

APPENDICI

- A. Traduzione della tabella “*Version 1 of the Levels of Digital Preservation*” in (Phillips et al. 2013, p. 3);

- B. Brochure informativa di PAD per gli autori;

- C. Contratto di PAD per il conferimento;

- D. Questionario informativo per gli autori.

Version 1 of the Levels of Digital Preservation

in (Phillips et al. 2013, p. 3)

	PROTEGGI I DATI	CONOSCI I DATI	MONITORA I DATI	RIPARA I DATI
STOCCAGGIO E LOCAZIONE GEOGRAFICA	Due copie complete conservate separatamente; Per supporti eterogenei copiare il contenuto all'interno del proprio sistema	Almeno tre copie complete; Almeno una geograficamente separata; Documentare il sistema di stoccaggio e che cosa ti serve per utilizzarli	Almeno una copia geograficamente separata con un differente trattamento di <i>disaster recovery</i> ; Monitorare l'obsolescenza di supporti e sistemi	Almeno tre copie geograficamente separate con differenti trattamenti di <i>disaster recovery</i> ; Attuare un piano che permetta di mantenere documenti e metadati sui supporti correntemente accessibili
FISSITÀ DEI FILE E INTEGRITÀ DEI DATI	Controllare la fissità dei file all'acquisizione se i <i>checksum</i> sono forniti con il contenuto; Se non forniti, creare le informazioni di fissità	Controllare la fissità ad ogni acquisizione; Utilizzare <i>write blocker</i> per accedere ai supporti originali; Controllare con antivirus i contenuti a rischio	Controllare la fissità del contenuto a intervalli regolari; Mantenere registri delle informazioni di fissità; Notare i dati corrotti; Controllare con antivirus tutti i dati	Controllare la fissità del contenuto in seguito a specifiche operazioni; Riparare/cambiare i dati corrotti; Assicurarsi che nessuno abbia accesso in scrittura a tutte le copie
SICUREZZA	Identificare che ha i vari permessi sui file individuali; Limitare il più possibile che ha questi permessi	Documentare le restrizioni di accesso sul contenuto	Mantenere registri di chi ha compiuto ogni operazione all'interno del sistema	Praticare ispezioni dei registri
METADATI	Inventario del contenuto e di dove è conservato; Assicurare il <i>backup</i> e la conservazione separata dell'inventario	Archiviare i metadati amministrativi; Archiviare i metadati "trasformativi" e i registri degli eventi	Archiviare metadati tecnici standard e metadati descrittivi	Archiviare metadati standard di conservazione
FORMATI DEI FILE	Incoraggiare, se possibile, la creazione di file in un limitato numero di formati aperti e noti	Inventario dei formati in uso	Sorvegliare i problemi legati all'obsolescenza dei formati in uso	Mettere in atto strategie di emulazione, migrazione e interventi analoghi all'occorrenza



PAD conserva gli archivi digitali di scrittori, giornalisti, intellettuali rappresentativi della cultura e società odierna, riconducibili a diverse generazioni e formazioni; le collezioni conferite sono messe a disposizione degli studiosi nel rispetto delle norme sulla privacy e dei limiti di consultabilità stabiliti dall'autore. PAD raccoglie un patrimonio ampio e trasversale di documenti in formato digitale e studia le strategie per la conservazione a lungo termine.

Informazioni e contratto

Il rapporto tra l'autore che cede documenti digitali a PAD e l'Ateneo è formalizzato da un contratto. PAD agisce secondo le norme che regolano la privacy e il diritto d'autore. L'autore può decidere:

- quali file non possono essere consultati dagli studiosi ed eventualmente per quanto tempo;
- le modalità di consultazione (esclusivamente nella sede di PAD presso l'Università di Pavia, o anche da remoto);
- di incrementare il suo archivio con successivi conferimenti;
- di chiedere a PAD in qualsiasi momento un backup in caso di perdita dei dati conservati sul proprio computer personale.

Gli autori saranno aggiornati periodicamente dai responsabili tecnico-scientifici sulle novità e lo stato dei lavori. Ogni autore ha un'area riservata sul sito pad.unipv.it, dove può consultare i dati riassuntivi e i dettagli dell'archivio conferito.

Questionario informativo e risorse bio e bibliografiche

Prima del conferimento viene chiesto all'autore di compilare un questionario informativo, che precisa la tipologia dell'archivio. PAD, in relazione alle risorse di cui potrà disporre, si impegna a raccogliere le informazioni biografiche e bibliografiche sull'autore, che entreranno sul sito del progetto.

Preparare il conferimento

Un incaricato dell'Università di Pavia fisserà un appuntamento con l'autore e realizzerà una copia su hard disk dei file conferiti. È possibile conservare in PAD qualsiasi tipo di file: documenti di testo, foto, immagini, audio, video, pagine web e social network. Le situazioni degli archivi conferiti possono essere molto varie, e sarà perciò necessario individuare caso per caso le soluzioni più idonee. Schematizzando, si possono prevedere quattro tipologie di conferimento:

SITUAZIONE	AZIONI
L'archivio conferito coincide con l'archivio dell'autore.	Si trasferisce sull'hard disk la copia dei file e delle directory così come sono distribuite nel computer di provenienza.
L'archivio conferito coincide solo in parte con l'archivio dell'autore: la struttura nella sostanza è immutata, alcuni materiali sono stati eliminati.	Si trasferisce sull'hard disk la copia dei file e delle directory così come sono distribuite nel computer di provenienza. L'autore tramite il questionario può aiutare PAD dando informazioni generiche sui file che non ha conferito (tipologia, posizione, etc.).
L'archivio conferito coincide in minima parte con l'archivio dell'autore: cambia sia nella struttura delle directory che nel numero dei file.	Si trasferisce sull'hard disk la copia delle directory e dei file. L'operatore chiede all'autore, nei limiti del possibile e nel rispetto delle sue volontà, di prendere nota della struttura generale dell'archivio di provenienza.
L'autore ha selezionato i file eliminando la struttura dell'archivio di provenienza. L'autore ha inserito i file in una o più cartelle create appositamente per il conferimento.	Si trasferisce sull'hard disk la copia dei file conferiti. In questo caso i file non rappresentano l'organicità dell'archivio personale, si presentano come materiali autonomi dalla situazione di provenienza di cui conseguentemente PAD non terrà traccia.

Quanto più la donazione è ampia e la copia eseguita da PAD fedele alla situazione originaria, tanto più sarà possibile utilizzare e interpretare i materiali raccolti. È prevista la possibilità di conferire parti di collezioni in momenti diversi: nel caso in cui l'autore voglia ampliare il proprio fondo PAD, l'archivio verrà aggiornato con un sistema che mantenga le strutture e i dati precedenti e accolga separatamente, pur con interconnessioni, le directory e i file conferiti in momenti successivi.

Fasi di conferimento e archiviazione

1. Copia dei file e delle cartelle dal computer dell'autore

Un incaricato dell'Università di Pavia si recherà dall'autore con un hard disk appositamente predisposto, su cui verranno copiati i file che l'autore deciderà di conferire. Si tratterà di un semplice "copia e incolla": i dati presenti sul computer dell'autore non saranno in alcun modo modificati.

L'hard disk è dotato di criptazione hardware e ha una password conservata dal responsabile tecnico di PAD. La criptazione è una misura preventiva in caso di furto dell'hard disk durante le fasi di trasporto.

2. Prime procedure di controllo

All'Università di Pavia i dati vengono provvisoriamente trasferiti su un server protetto per attivare le procedure di controllo: se all'interno dell'archivio fossero presenti file o cartelle compressi (come ad esempio .zip, .rar) il programma PadCleaner li decompime per visualizzare 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 sono stati, almeno in parte, generati dall'autore e potrebbero contenere dati sensibili conferiti inconsapevolmente.

3. Identificazione e sicurezza

Il programma PadChecker rileva la presenza di virus: se individuati, l'autore viene tempestivamente informato, e se necessario gli si offre una prima assistenza. Nell'archivio PAD i 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, "codici" univoci che rimarranno associati per sempre ai singoli file. PadChecker effettua una scansione quotidiana basata sull'SHA-1 e segnala all'amministratore eventuali manomissioni. Ogni archivio è criptato ed è accessibile in modalità *read-only*: non può dunque essere modificato.

4. Convalida del conferimento

Il programma PadPrint genera una lista di tutti i file conferiti, che verrà inviata all'autore per una convalida. In caso di ripensamenti l'autore 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, l'autore potrebbe infatti aver conferito alcuni materiali inconsapevolmente. La lista inviata mostra in maniera ordinata e chiara tutta la collezione. PAD vuole dare la massima garanzia all'autore instaurando un rapporto di trasparenza e fiducia: solo con la firma del documento, poi allegato al contratto, vengono avviate le procedure di immagazzinamento, conservazione e archiviazione.

5. Conservazione a lungo termine

Il programma PadConverter converte automaticamente i file in diversi formati, che garantiscono la conservazione dei dati secondo i principi di *long-term preservation*. Ogni archivio è costituito dai file conferiti, dalle conversioni e da tutta la documentazione digitale e digitalizzata (contratto, questionario, informazioni sul conferimento, set di metadati). A oggi non esiste un metodo univoco e sicuro per la conservazione a lungo termine di materiali digitali, ed è dunque necessario attuare strategie

diversificate (*refreshing*, duplicazione, emulazione, migrazione), talvolta combinate tra loro, per aumentare le probabilità di successo. Il progetto PAD migliora con l'evolversi delle nuove tecnologie e prevede futuri trasferimenti di dati su eventuali nuovi supporti.

Una copia dell'archivio è conservata in forma criptata, dunque non accessibile liberamente, su un server primario del centro di calcolo dell'Università di Pavia. La struttura dell'hardware utilizzata ricorre alla tecnologia RAID (Redunant Array of Indipendent Disk), che duplica le attività del disco fisso su più dischi garantendo così la ridondanza dei dati nel caso di errore su uno di essi: il *disaster recovery* è garantito da un secondo server, localizzato a Cremona (sede distaccata dell'Università di Pavia), che mantiene una copia costantemente aggiornata dei dati conservati sul server primario. Una copia dell'archivio viene masterizzata su un DVD *archival grade* con strato metallico in oro, posto in busta sigillata in vuoto e conservato presso il caveau di una banca.



**OFFERTA AL PUBBLICO E
CONDIZIONI GENERALI DI CONTRATTO DI PAD – PAVIA ARCHIVI DIGITALI**

A tutti gli interessati

L'Università di Pavia attraverso il progetto Pavia Archivi Digitali (di seguito PAD) ha organizzato un archivio di documenti in formato digitale/elettronico (d'ora in avanti "documenti" o "documento"), i cui obiettivi e la cui organizzazione sono descritti ulteriormente nell'allegato n. 1. L'Università degli Studi di Pavia offre al pubblico ex art. 1336 c.c., e con ciò 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, la proposta di un contratto tra l'autore interessato e l'Università degli Studi di Pavia alle seguenti condizioni generali:

1. L'Autore consegna a un incaricato del PAD una copia dei documenti che indicherà al momento della consegna.
2. Al momento della consegna l'Autore indicherà altresì se PAD può rendere accessibili i suoi documenti soltanto in loco presso il PAD o anche da remoto, immediatamente o soltanto a far tempo dalla data successiva indicata dall'Autore. L'Autore potrà indicare una data diversa per ciascun documento. Tutte le date si intendono indicate dall'Autore irrevocabilmente: salvo la sua facoltà di consentire successivamente a PAD per iscritto l'accesso ai suoi documenti anche prima di esse.
3. L'Autore autorizza PAD e ogni possibile terzo che si rivolga a PAD (ma solo nei limiti necessari a PAD e a ogni possibile terzo per lo studio, la ricerca, la citazione, la pubblicazione anche on line dei risultati della ricerca relativa ai documenti dell'Autore) all'utilizzazione dei suoi documenti consegnati a PAD, e ne consente la riproduzione, la comunicazione al pubblico e la messa a disposizione al pubblico.
4. L'Autore consente il trattamento dei dati che comunica al momento della consegna dei documenti e dei dati contenuti in questi documenti, per ogni attività di PAD comunque relativa all'organizzazione e alla promozione di PAD e alle utilizzazioni dei documenti indicate dal precedente art. 3: e garantisce che queste utilizzazioni non violano il diritto alla privacy di alcun terzo.
5. PAD si obbliga a inserire sollecitamente i documenti dell'Autore nell'archivio PAD; a provvedere alla manutenzione dei relativi software e dei documenti; e a rendere disponibili al pubblico i documenti degli autori secondo le disposizioni degli articoli che precedono.
6. Tenuto conto degli investimenti necessari per l'organizzazione e la promozione di PAD e per l'archiviazione e la manutenzione dei documenti dell'Autore, delle finalità universitarie di PAD e dei vantaggi di ogni genere che potranno derivare all'Autore dall'organizzazione di PAD, l'Autore rinuncia a compensi monetari e a ulteriori attività corrispettive a suo favore da parte di PAD.



7. L'Autore si obbliga a non consegnare a terzi, ai fini di studio e ricerca, alcun altro documento corrispondente a quelli consegnati a PAD, per tutto il tempo in cui i suoi documenti consegnati a PAD resteranno oggetto di diritti d'autore o connessi.
8. Il contratto con l'Università potrà essere concluso dall'Autore soltanto sulla base dell'atto scritto di accettazione delle presenti condizioni generali di contratto, allegato al corrente documento con il n. 2, che fornisca a un tempo le indicazioni previste dal precedente art. 2.
9. Per ogni controversia comunque relativa al contratto tra PAD e l'Autore, a ogni atto compiuto in esecuzione dello stesso, e alla sua cessazione per qualsiasi ragione è competente in via esclusiva il foro di Milano.

Pavia, _____

Università degli Studi di Pavia
Il Rettore

(Prof. Fabio Ruge)



Allegato n. 1

PROGETTO PAD – PAVIA ARCHIVI DIGITALI
REPOSITORY DELLA CULTURA ITALIANA DEL XXI SECOLO

La diffusione degli strumenti di videoscrittura ha portato, a partire dalla metà degli anni ottanta, a una graduale scomparsa del supporto cartaceo come strumento di archiviazione, sia nell'uso personale che in quello professionale. La quasi totalità dei documenti elaborati negli ultimi due decenni è ora contenuta in file non strutturati, prodotti con l'ausilio dei più diversi applicativi. Si tratta di un enorme e ricchissimo patrimonio di informazioni relative alla vita e alla produzione culturale recente, conservato, al momento attuale, in modo spesso precario e senza garanzie di mantenimento sul medio e lungo periodo. In assenza di un progetto specifico di salvaguardia, tale patrimonio rischia di andare irrimediabilmente perduto per incuria, anomalie tecniche e obsolescenza del software, lasciando gli storici e i filologi del futuro quasi completamente privi di fonti dirette per la ricostruzione dei percorsi della produzione culturale di questi decenni.

Il progetto PAD si propone di fornire uno strumento per raccogliere documenti in formato digitale/elettronico di personalità della cultura garantendone la conservazione nel tempo e l'eventuale successivo studio, nel rispetto della privacy e del diritto d'autore. PAD prevede l'incrementabilità nel tempo del materiale conferito, perché l'archivio resti aggiornato sugli autori che già detiene.

LA SOLUZIONE TECNICA

Il progetto prevede una struttura tecnica molto semplice, svincolata da prodotti commerciali, che privilegia l'apertura e la semplicità del sistema, rispetto alla quantità e complessità delle sue funzioni.

PAD garantisce una conservazione a lungo termine dei dati in condizioni di massima sicurezza e predispone applicativi specifici per la consultazione.

Nelle sue linee essenziali la soluzione tecnica identificata prevede un'architettura a più livelli: un livello interno è costituito da due server fisicamente e logicamente separati, che sono delegati alla pura conservazione dei file originali, della documentazione e dei file normalizzati secondo i principi di long term preservation.

Un livello più esterno è costituito da un server di lavoro su cui risiedono unicamente i dati accessibili nei termini stabiliti dall'autore nell'allegato n. 2 del presente contratto. Questo livello è utilizzato per l'attività di accesso ed elaborazione dei dati.

Un terzo livello di sicurezza è garantito da una copia di backup realizzata su DVD archival grade con strato metallico in oro, posto in busta sigillata in vuoto e conservato presso il caveau di una banca.

I documenti in formato digitale/elettronico conferiti dall'autore, trasportati mediante un hard disk dotato di crittazione hardware, vengono inseriti, con una procedura controllata, nel livello più interno del sistema, dopo essere stati oggetto di un esame dei contenuti, allo scopo di verificare l'integrità del materiale ricevuto e di identificarlo sulla base di uno schema ridotto di metadati.

A eccezione della copia di backup in DVD tutti i dati vengono memorizzati in forma criptata.



UNIVERSITÀ DEGLI STUDI DI PAVIA

PAD – Pavia Archivi Digitali

Via Luino, 12 – 27100 PAVIA – tel. 0382 50984819 – fax 0382-984889 – e-mail pad@unipv.it – web pad.unipv.it

Allegato n. 1

IL COMITATO DI GESTIONE

PRESIDENTE

FABIO RUGGE

COMITATO SCIENTIFICO

Simone Albonico

Umberto Anselmi Tamburini

Gianni Danese

Maria Antonietta Grignani

Clelia Martignoni

Giorgio Panizza

Giovanna Ruberto

Beppe Severgnini

Angelo Stella

Luigi Carlo Ubertazzi

Paul Gabriele Weston

PROMOZIONE E RAPPORTI ESTERNI

Grazia Bruttocao

RESPONSABILI TECNICO-SCIENTIFICI

Primo Baldini

Emmanuela Carbé

SEGRETERIA

Nicoletta Leone



Allegato n. 2

ACCETTAZIONE DELL'AUTORE

1. L'Autore dichiara di accettare integralmente l'offerta al pubblico dell'Università degli Studi di Pavia e le relative condizioni generali di contratto di PAD indicate agli artt. 1 e seguenti.
2. L'Autore precisa che i propri documenti in formato digitale/elettronico dovranno essere
 - accessibili soltanto in loco presso la sede PAD
 - accessibili presso la sede PAD e solo in parte da remoto
 - accessibili presso la sede PAD e da remoto

 - accessibili immediatamente
 - accessibili solo in parte immediatamente
 - accessibili soltanto a far tempo dalla seguente data: _____ .
3. L'Autore consegna all'incaricato di PAD i suoi documenti in formato digitale/elettronico e riceverà entro 15 giorni l'elenco dei dati conferiti che dovrà sottoscrivere indicando precisamente le eventuali limitazioni d'accesso all'archivio.

Per accettazione

L'Autore

Per accettazione espressa ex art. 1341 c.c. delle clausole indicate dai seguenti articoli delle condizioni generali di contratto di PAD: art. 7 (esclusiva a favore di PAD) e art. 9 (foro competente).

L'Autore

Pavia, 23/01/2016

Gentile Autore,

PAD desidera proporle un questionario informativo da allegare alla documentazione del suo conferimento.

Le saranno chieste informazioni sul computer e sui dispositivi che utilizza, sull'organizzazione dell'archivio che intende conferire, sul rapporto tra il suo lavoro di scrittura e gli strumenti tecnologici. Se non fosse in grado di rispondere ad alcune domande, o se le ritenesse troppo impegnative, potrà decidere di saltarle (lasciando eventualmente un commento).

Il questionario è un passo fondamentale per il conferimento: dalle sue risposte si potrà capire meglio la struttura dell'archivio al fine di una migliore classificazione.

Grazie fin da ora per la sua collaborazione.

Data di compilazione / /

1. Come si configura l'archivio che conferirà a PAD?

<input type="checkbox"/>	È una copia che coincide con il mio archivio personale di lavoro, sia nella struttura (cartelle e sottocartelle) che nei contenuti.
<input type="checkbox"/>	È una copia simile al mio archivio personale di lavoro: ho solo eliminato alcune cartelle o file che non desidero conferire.
<input type="checkbox"/>	È una copia che coincide con il mio archivio personale di lavoro, sia nella struttura (cartelle e sottocartelle) che nei contenuti.
<input type="checkbox"/>	Altro:

eventuali commenti:

2. Archivio personale / archivio conferito

Potrebbe anche solo brevemente descrivere com'è organizzato l'archivio che ha conferito? Nel caso in cui ci fossero importanti differenze con il suo archivio personale, potrebbe dare qualche informazione per capire meglio il "tavolo di lavoro" originario da cui derivano i materiali che PAD conserverà?

3. Struttura e contenuti dell'archivio

Ha fatto o ha intenzione di apportare modifiche al suo archivio prima del conferimento a PAD (es. nella struttura delle directory, ai nomi delle cartelle, ai contenuti o ai nomi dei file)?

<input type="checkbox"/>	Sì
<input type="checkbox"/>	Solo in parte
<input type="checkbox"/>	No

eventuali commenti:

4. Consultazione

Vuole che parte dei contenuti non sia consultabile per un certo numero di anni? Potrà segnalarci i documenti non consultabili nella lista dettagliata dei file che le sarà inviata per convalidare il conferimento.
<input type="checkbox"/> Sì Per quanti anni desidera che non siano consultabili?
<input type="checkbox"/> No

eventuali commenti:

5. Modalità di consultazione

Il suo archivio potrà essere consultabile anche da remoto o vi si potrà accedere solo dalla sede di PAD?
<input type="checkbox"/> Anche da remoto
<input type="checkbox"/> Solo presso la sede PAD

eventuali commenti:

6. Privacy

L'archivio che ha conferito contiene dati non propriamente suoi (ad es. file generati da terze persone, e-mail ricevute, etc.)?
<input type="checkbox"/> Sì Saprebbe indicarne gli autori e se possono essere contattati?
<input type="checkbox"/> No

7. Incrementare l'archivio

Avrebbe intenzione di incrementare il suo archivio con successivi conferimenti?

Sì

Acconsente a che PAD la contatti periodicamente (ogni 12/24 mesi) per conferimenti successivi?

Sì No

No

eventuali commenti:

8. Conferimenti diversi dall'archivio

Riterrebbe utile conferire anche altro tipo di materiale (e-mail, pagine web, social network, newsletter da lei curate, etc.)?

Sì

No

eventuali commenti:

9. Computer principale

Che computer utilizza per lavorare?

Quando ha acquisito il computer?

È lo stesso da cui provengono i file che conferirà a PAD?
<input type="checkbox"/> Sì
<input type="checkbox"/> Solo in parte
<input type="checkbox"/> No

eventuali commenti:

10. Macchine precedenti

Con che frequenza cambia gli strumenti informatici, e in particolare il computer principale?

Cosa utilizzava prima di questo computer?

La macchina o le macchine sostituite sono ancora nella sua disponibilità?
<input type="checkbox"/> Sì
<input type="checkbox"/> No
<input type="checkbox"/> Non lo so

I file che ha intenzione di conferire sono stati generati anche da computer precedenti?
<input type="checkbox"/> Sì
<input type="checkbox"/> No
<input type="checkbox"/> Non lo so

eventuali commenti:

11. Sistema Operativo

Conosce il Sistema Operativo (Windows/Mac/Linux) installato sul suo computer?
<input type="checkbox"/> Sì [specificare:]
<input type="checkbox"/> No

Utilizza o ha utilizzato altri sistemi operativi?
<input type="checkbox"/> Sì [specificare:]
<input type="checkbox"/> No

12. Altri computer e dispositivi

Utilizza più computer (ad. es. fisso/portatile)?
<input type="checkbox"/> Sì [specificare:]
<input type="checkbox"/> No

Utilizza anche altri dispositivi per l'attività di scrittura e lavoro (ad es. tablet, smartphone)?
<input type="checkbox"/> Sì [specificare:]
<input type="checkbox"/> No

Se utilizza altri dispositivi, provvede poi a trasferire sul computer principale i testi elaborati altrove?
<input type="checkbox"/> Sì
<input type="checkbox"/> No

Pensa che alcuni suoi testi siano ora soltanto su dispositivi diversi dal computer principale?
<input type="checkbox"/> Sì
<input type="checkbox"/> No
<input type="checkbox"/> Non lo so

I file che ha intenzione di conferire sono stati generati anche da altri dispositivi (ad es. app di tablet /smartphone, macchine fotografiche)?
<input type="checkbox"/> Sì
<input type="checkbox"/> No
<input type="checkbox"/> Non lo so

Eventuali commenti:

13. Organizzazione dei dati

Conosce, all'incirca, la disposizione all'interno del suo computer dei documenti da lei prodotti?
<input type="checkbox"/> Sì
<input type="checkbox"/> No

Quando salva un file sceglie una collocazione particolare?
<input type="checkbox"/> Sì
<input type="checkbox"/> No

Eventuali commenti:

14. Programmi di scrittura e revisione dei testi

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 del file per ottenere versioni diverse e progressive?

Le capita di eliminare 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 che conferirà ci sono file frutto di un lavoro condiviso con i redattori?

15. Processi di scrittura

Utilizza ancora la scrittura tradizionale su carta?
<input type="checkbox"/> Sì
<input type="checkbox"/> Solo in parte
<input type="checkbox"/> No

Se la utilizza, che rapporto c'è tra la scrittura su carta e la scrittura digitale?

16. Scrittura e web

Che rapporto ha con le nuove tecnologie e con il web? Legge e-book, si documenta via internet mentre scrive? Ritiene che i nuovi strumenti tecnologici a disposizione stiano influenzando il suo lavoro?

Hai mai tenuto un sito internet o un blog?
<input type="checkbox"/> Sì indirizzo: è ancora attivo? <input type="checkbox"/> sì <input type="checkbox"/> no
<input type="checkbox"/> No

Ha mai scritto su siti/blog collettivi?
<input type="checkbox"/> Sì specificare:
<input type="checkbox"/> No

Eventuali commenti:

17. Social Network

Ritiene che i social network siano oggi rilevanti per promuovere al pubblico l'attività di un autore?

È iscritto a social network?

Facebook

Twitter

Altro [specificare:]

No

Se li utilizza abitualmente, ritiene che questa attività influisca sui suoi processi di scrittura?

Ritiene che i materiali da lei elaborati per i social network potrebbero entrare nell'archivio PAD?

Sì

No

Eventuali commenti:

18. Ulteriori note e commenti

[Empty rectangular box]

Note dell'operatore PAD

[Empty rectangular box]

BIBLIOGRAFIA

- Aliprandi, Simone. 2016. «Il Ministero dell'Interno ha problemi con le licenze CC e con l'open data». novembre 6. <http://aliprandi.blogspot.com/2016/11/ministero-interno-problemi-licenzeCC-opendata.html>.
- Allegrezza, Stefano. 2011. «Analisi del formato FITS per la conservazione a lungo termine dei manoscritti. Il caso significativo del progetto della Biblioteca Apostolica Vaticana». *DigItalia 2* (0): 43–72.
- . 2015. «Digital repositories certification: the Data Seal of Approval». *JLIS.it* 6 (3): 39–56. doi:10.4403/jlis.it-11332.
- Allegrezza, Stefano, e Luca Gorgolini, a c. di. 2016. *Gli archivi di persona nell'era digitale: il caso dell'archivio di Massimo Vannucci*. Percorsi. Bologna: Società editrice Il mulino.
- Arbasino, Alberto. 2008. *L'ingegnere in blu*. Piccola biblioteca Adelphi 565. Milano: Adelphi.
- «Architecture of the World Wide Web, Volume One». 2004. dicembre 15. <https://www.w3.org/TR/webarch/#intro>.
- Athow, Desire. 2015. «The data capacity gap: why the world is running out of data storage». *TechRadar*. febbraio 10. <http://www.techradar.com/news/computing-components/storage/the-data-capacity-gap-why-the-world-is-running-out-of-data-storage-1284024/2>.
- Baca, Murtha. 2016. «Introduction to Metadata». InteractiveResource. luglio 2016. <http://www.getty.edu/publications/intrometadata>.

- Barrera-Gomez, Julianna, Ricky Erway, e OCLC Research. 2013. *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read in-House*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf>.
- Bit rot (on digital vellum) | Vint Cerf | TEDxRoma*. 2014. <https://www.youtube.com/watch?v=GV0A82TCrf0&feature=youtu.be>.
- Bloch, Arthur. 1990. *Il terzo libro di Murphy: altre pessime ragioni per cui il mondo va comunque avanti!* Milano: Longanesi.
- Bodleian Library, John Rylands University Library of Manchester, e Joint Information Systems Committee. 2007. *Paradigm: Workbook on Personal Digital Archives*. Oxford: Bodleian Library. <http://www.paradigm.ac.uk/workbook/>.
- Carroll, Laura, Erika Farr, Peter Hornsby, e Ben Ranker. 2011. «A Comprehensive Approach to Born-Digital Archives». *Archivaria* 72: 61–92.
- Carucci, P., e Maria Rosaria Guercio. 2008. *Manuale di archivistica*. 1. ed. Beni culturali 32. Roma: Carocci.
- Cerf, Vinton. 2015. «Digital Vellum». In . aaas. <https://aaas.confex.com/aaas/2015/webprogram/Paper14064.html>.
- «Che lingua fa?» 2016. *Nuovi argomenti* 5.
- Chun, Wendy Hui Kyong, e Thomas Keenan, a c. di. 2006. *New media, old media: a history and theory reader*. New York: Routledge.
- CMU Distinguished Lecture Series - Vint Cerf*. 2015. https://www.youtube.com/watch?v=K_DIwiSDaT8.
- Corte di Giustizia dell'Unione Europea. 2012. «L'autore di un software non può opporsi alla rivendita di proprie licenze "usate" che consentono di utilizzare i suoi programmi scaricati via Internet - European Commission - PRESS RELEASES - Press release». luglio 3. http://europa.eu/rapid/press-release_CJE-12-94_it.htm.
- Corti, Maria. 1997. *Ombre dal fondo*. I coralli 61. Torino: Einaudi.
- Creative Commons. 2017. «Chi siamo». Consultato gennaio 30. <http://creativecommons.it/About>.

Dappers, Angela, e Markus Lenders. 2010. «Digital Preservation Metadata Standards». *Information Standards Quarterly* 22 (2): 4. doi:10.3789/isqv22n2.2010.01.

«Dario Fo lascia tutto il suo archivio alla Svezia: il sospetto sul Nobel». 2015. agosto 4. <http://www.liberoquotidiano.it/news/personaggi/11817076/Dario-Fo-lascia-tutto-il-suo.html>.

«Dario Fo vill skänka sitt teaterarv till Sverige». 2015. *DN.SE*. luglio 24. <http://www.dn.se/kultur-noje/scen/dario-fo-vill-skanka-sitt-teaterarv-till-sverige/>.

DECRETO DEL PRESIDENTE DEL CONSIGLIO DEI MINISTRI 3 dicembre 2013: Regole tecniche in materia di sistema di conservazione ai sensi degli articoli 20, commi 3 e 5-bis, 23-ter, comma 4, 43, commi 1 e 3, 44, 44-bis e 71, comma 1, del Codice dell'amministrazione digitale di cui al decreto legislativo n. 82 del 2005. (14A02098) (GU Serie Generale n.59 del 12-3-2014 - Suppl. Ordinario n. 20). 2013. <http://www.gazzettaufficiale.it/eli/id/2014/03/12/14A02098/sg>.

DECRETO LEGISLATIVO 22 gennaio 2004, n. 42 Codice dei beni culturali e del paesaggio, ai sensi dell'articolo 10 della legge 6 luglio 2002, n. 137. (GU n.45 del 24-2-2004 - Suppl. Ordinario n. 28). 2004. <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2004-01-22;42!vig=>.

Deleuze, Gilles, Claire Parnet, e Pierre Boutang. 2005. «Abecedario di Gilles Deleuze video-intervista in tre DVD». DeriveApprodi.

Derrida, Jacques. 1996. *Mal d'archivio: un'impressione freudiana*. Napoli: Filema.

Desideri, Laura, e Giuliana Zagra, a c. di. 2010. *Conservare il Novecento: gli archivi culturali: Ferrara, Salone internazionale dell'arte del restauro, 27 marzo 2009*. Roma: Associazione italiana biblioteche.

Digital Preservation Coalition. 2015. «Digital Preservation Handbook». <http://www.dpconline.org/handbook/>.

EBLIDA. 2014. «The Right to E-read Position Paper and Statement (archived) - European Bureau of Library Information and Documentation Associations (EBLIDA)». maggio. <http://www.eblida.org/e-read/the-right-to-e%E2%80%90read-position-paper-and-statement.html>.

- «EBLIDA briefing on the e-lending judgement of the CJEU - European Bureau of Library Information and Documentation Associations (EBLIDA)». 2017. gennaio 19.
<http://www.eblida.org/news/eblida-briefing-on-the-e-lending-judgement-of-the-cjeu.html>.
- «EBLIDA Position Papers and Statements - European Bureau of Library Information and Documentation Associations (EBLIDA)». 2017. Consultato gennaio 29.
<http://www.eblida.org/activities/position-papers.html>.
- Erway, Ricky, Ben Goldman, Matthew McKinley, e OCLC Research. 2014. *Agreement Elements for Outsourcing Transfer of Born Digital Content*. Dublin, Ohio: OCLC Research.
<http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-born-digital-content-transfer-2014.pdf>.
- Erway, Ricky, Laurence Horton, Amy Nurnberger, Reid Otsuji, Amy Rushing, e OCLC Research. 2016. *Building Blocks: Laying the Foundation for a Research Data Management Program*.
<http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-data-management-building-blocks-2016.pdf>.
- Erway, Ricky, e OCLC Research. 2012a. *Swatting the Long Tail of Digital Media: A Call for Collaboration*. Dublin, Ohio: OCLC Research.
<http://www.oclc.org/research/publications/library/2012/2012-08.pdf>.
- . 2012b. *You've Got to Walk before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media*. Dublin, Ohio: OCLC Research.
<http://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf>.
- Falkvinge, Rickard. 2017. «The Internet has reversed the assumption that justified copyright». *Privacy Online News*. gennaio 10.
<https://www.privateinternetaccess.com/blog/2017/01/modern-copyright-was-built-fundamental-assumption-internet-reversed/>.
- Ghersetti, Francesca, Loretta Paro, e Fondazione Giuseppe Mazzotti per la civiltà veneta, a c. di. 2012. «Archivi di persona del Novecento: guida alla sopravvivenza di autori, documenti e addetti ai lavori». In . Treviso: Fondazione Bettinetton studi ricerche : Fondazione Giuseppe Mazzotti per la civiltà veneta : Antiga.

- Giaretta, David, Henk Harmsen, e Christian Keitel. 2010. «Memorandum of Understanding». http://www.dataealofapproval.org/media/filer_public/2014/08/28/20100709_020_signedmoutocreateaeuropeanframeworkforauditandcertificationofdigitalrepositories.pdf.
- Guerrini, Mauro. 2016. «Biblioteca digitale.pdf». https://e-l.unifi.it/pluginfile.php/227893/mod_resource/content/1/Biblioteca%20digitale.pdf.
- (ICAR), Antonella Mulè. 2016. «Inaugurato l'Archivio Fo - Rame». aprile 15. <http://www.ilmondodegliarchivi.org/rubriche/gli-archivi-si-raccontano/318-inaugurato-l-archivio-fo-rame>.
- «Individuazione delle modalità semplificate per l'informativa e... - Garante Privacy». 2014. maggio 8. <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3118884>.
- International Council of Archives, a c. di. 2000. *ISAD(G): General International Standard Archival Description ; Adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999*. 2. ed. Ottawa: International Council of Archives.
- Karl Fogel. 2005. *The Surprising History of Copyright and The Promise of a Post-Copyright World*. <http://archive.org/details/TheSurprisingHistoryOfCopyrightAndThePromiseOfAPost-copyrightWorld>.
- Lavoie, Brian. 2016. *Il modello di riferimento per un Sistema informativo aperto per l'archiviazione*. A cura di Paul Gabriele Weston. Tradotto da Flavia Cancedda, Chiara Consonni, Agnese Galeffi, Andrea Marchitelli, e Lucia Sardo. Roma: Associazione Italiana Biblioteche.
- «Lettura digitale in biblioteca: legalizziamola!» 2014. *AIB-WEB*. aprile 16. <http://www.aib.it/attivita/2014/42592-lettura-digitale-biblioteca/>.
- LUMSA, 21 aprile 2016, Roma: La memoria labile del digitale*. 2016. https://www.youtube.com/watch?v=f_R73ljxoQc.
- Mascheroni, Luigi. 2015. «Dario Fo lascia tutto il suo archivio alla Svezia». *ilGiornale.it*. agosto 4. <http://www.ilgiornale.it/news/cultura/dario-fo-lascia-tutto-suo-archivio-svezia-1158056.html>.

Messina, Maurizio, Giuliana Zagra, Associazione italiana biblioteche, Istituto centrale per la patologia del libro Alfonso Gallo, e Istituto per i beni artistici, culturali, naturali della Regione Emilia-Romagna, a c. di. 2003. *Conservare il Novecento: oltre le carte: Convegno nazionale, Ferrara, Salone internazionale dell'arte del restauro e della conservazione dei beni culturali e ambientali, 5 aprile 2002: atti*. Roma: Associazione italiana biblioteche.

Messina, Maurizio, Giuliana Zagra, Istituto per i beni artistici, culturali, naturali della Regione Emilia-Romagna, Associazione italiana biblioteche, e Istituto centrale per la patologia del libro Alfonso Gallo, a c. di. 2001. *Conservare il Novecento: Convegno nazionale, Ferrara, Salone internazionale dell'arte del restauro e della conservazione dei beni culturali e ambientali, 25-26 marzo 2000: atti*. Roma: Associazione italiana biblioteche.

My story: Struggling, bullying, suicide, self harm. 2012.
<https://www.youtube.com/watch?v=vOHXGNx-E7E>.

NDSA, Paula De Stefano, Carl Fleischhauer, Andrea Goethals, Michael Kjörling, Nick Krabbenhoft, Chris Lacinak, et al. 2014. «Checking Your Digital Content: What is Fixity, and When Should I be Checking It?»
<http://nds.org/documents/NDSA-Fixity-Guidance-Report-final100214.pdf>.

«paradigm | workbook on digital private papers». 2017. Consultato gennaio 7. <http://www.paradigm.ac.uk/workbook/>.

Pavone, Claudio, Italy, e Dipartimento per i beni archivistici e librari. 2006. *Storia d'Italia nel secolo ventesimo: strumenti e fonti*. Roma: Ministero per i beni e le attività culturali, Dipartimento per i beni archivistici e librari, Direzione generale per gli archivi.

Phillips, Megan, Jefferson Bailey, Andrea Goethals, e Trevor Owens. 2013. «The NDSA Levels of Digital Preservation: An Explanation and Uses». *National Digital Stewardship Alliance - Digital Library Federation*.
http://nds.org/documents/NDSA_Levels_Archiving_2013.pdf.

Protezione del diritto d'autore e di altri diritti connessi al suo esercizio. 1941. <https://dirittodautore.it/leggi/legge-22-aprile-1941-n-633/>.

«SNAC: Social Networks and Archival Context». 2010. 2015.
http://socialarchive.iath.virginia.edu/home_research.html.

- Stollar Peters, Catherine. 2006. «When Not All Papers are Paper: A Case Study in Digital Archivy». *Provenance, Journal of the Society of Georgia Archivists* 24 (1).
<http://digitalcommons.kennesaw.edu/provenance/vol24/iss1/3>.
- Tietz, Tabea, Joscha Jäger, Jörg Waitelonis, e Harald Sack. 2016. «Semantic Annotation and Information Visualization for Blogposts with refer». *VOILA 2016 Visualization and Interaction for Ontologies and Linked Data*, ottobre 17.
- «Usa, Corte Suprema dà torto agli autori: Google Books libera di scansionare milioni di libri». 2016. *Repubblica.it*. aprile 19.
http://www.repubblica.it/tecnologia/2016/04/19/news/copyright_google_books_corte_suprema_usa-137949980/.
- Valacchi, Federico. 2015. *Diventare archivisti: competenze tecniche di un mestiere sul confine*. Biblioteconomia e scienza dell'informazione 7. Milano: Editrice bibliografica.
- Vint Cerf - Digital Vellum*. 2015.
<https://www.youtube.com/watch?v=STeLOogWqWk&t=3s>.
- Vint Cerf speaks on Digital Preservation at NASA GSFC*. 2016.
<https://www.youtube.com/watch?v=H0l4GLjboDw&t=6s>.
- Virgillito, Daniele. 2014. «Come ho fregato TG, politici e giornali con Wikipedia». *Wired*. gennaio 15.
<https://www.wired.it/internet/2014/01/15/come-ho-fregato-tg-politici-e-giornali-con-wikipedia/>.
- Weston, Paul Gabriele. 2013. *Il libro, gli archivi e la memoria digitale*. Milano: UNICOPLI.

Tutti i *link* sono stati controllati l'ultima volta in data 17 febbraio 2017.

RINGRAZIAMENTI

Questo lavoro non sarebbe stato possibile senza il supporto e il contributo del professor Paul Gabriele Weston, di Primo Baldini ed Emmanuela Carbé, che ringrazio principalmente; devo inoltre un ringraziamento per i contributi e i confronti su vari aspetti e tematiche anche a Sara Zucchi e Camilla Fusetti.

A tutta la mia (numerosa) famiglia che mi ha s(u/o)pportato nel corso dei miei anni di studio, e agli amici grazie a quali sono quello che sono. In particolar modo agli Amici della Biblioteca di Cadoneghe, e ai miei fratelli e sorelle del Vespertilio.

A Valentina.

Grazie.