

AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti le Scienze dell'Informazione, la Documentazione, la Gestione Documentale e l'Organizzazione della Conoscenza. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi possono riguardare topics quali Documentazione, Scienze dell'informazione e della comunicazione, Scienze del testo e del documento, Organizzazione e Gestione della conoscenza, Terminologia, Statistica testuale e Linguistica computazionale e possono illustrare studi sperimentali in domini specialistici, casi di studio, aspetti e risultati metodologici conseguiti in attività di ricerca applicata, presentazioni dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Settore 11/A4 e censita per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali, così come dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) che la annovera tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE – Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; Letteratura Professionale Italiana – Associazione Italiana Biblioteche; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Summon™ – by Serialsolutions; Ulrich's; UniCat – Union Catalogue of Belgian Libraries; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

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

AIDAinformazioni

Anno 40

N. 1-2 – gennaio-giugno 2022

CACUCCI  EDITORE
BARI

PROPRIETÀ LETTERARIA RISERVATA

© 2022 Cacucci Editore – Bari
Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220
<http://www.cacuccieditore.it> e-mail: info@cacucci.it

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

Sommario

Contributi

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

Note e rubriche

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

Contributi

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

STEFANO ALLEGREZZA*

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

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

1. Introduzione

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

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

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

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

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

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

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

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

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

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

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

È importante riflettere anche sul concetto di email come documento in senso archivistico («email as a record»). È evidente che non tutte le email inviate, ricevute o comunque scambiate all'interno di una organizzazione (o da una persona fisica nel caso delle email personali) sono da considerare documenti in senso archivistico; tuttavia alcune di queste lo sono e come tali sono meritevoli di particolare attenzione e devono essere conservate per periodi più o meno lunghi (in alcuni casi per sempre). Uno dei progetti che ha affrontato con maggiore rigore metodologico questo aspetto è stato InterPares3⁴. Dal punto di vista della diplomazia del documento contemporaneo esso ha individuato cinque caratteristiche che un oggetto digitale deve possedere affinché possa essere considerato un documento in senso archivistico (InterPARES3 Project 2011, 8-10): un contenuto stabile ed una forma fissa; un legame con un atto a cui il documento fornisce supporto o proceduralmente o come parte di un processo decisionale; la presenza di un vincolo archivistico; il coinvolgimento di alcuni soggetti (un autore, uno scrittore, un destinatario); la collocazione in uno dei seguenti cinque tipi di contesto: giuridico-amministrativo, finanziario, procedurale, documentale, tecnologico. Non è certamente questa la sede opportuna per approfondire ulteriormente l'argomento e si rimanda per una sua trattazione esaustiva ai numerosi studi che sono stati pubblicati come risultati del progetto⁵; preme, comunque, sottolineare come la questione del riconoscimento del valore archivistico dei messaggi di posta elettronica sia di primaria importanza, perché da esso discendono tutta una serie di adempimenti ed obblighi – anche di legge – in ordine alla loro tenuta e conservazione da parte dei soggetti che trasmettono o ricevono tali messaggi e, in prospettiva,

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

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

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

3. Le sfide della conservazione degli archivi di posta elettronica

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

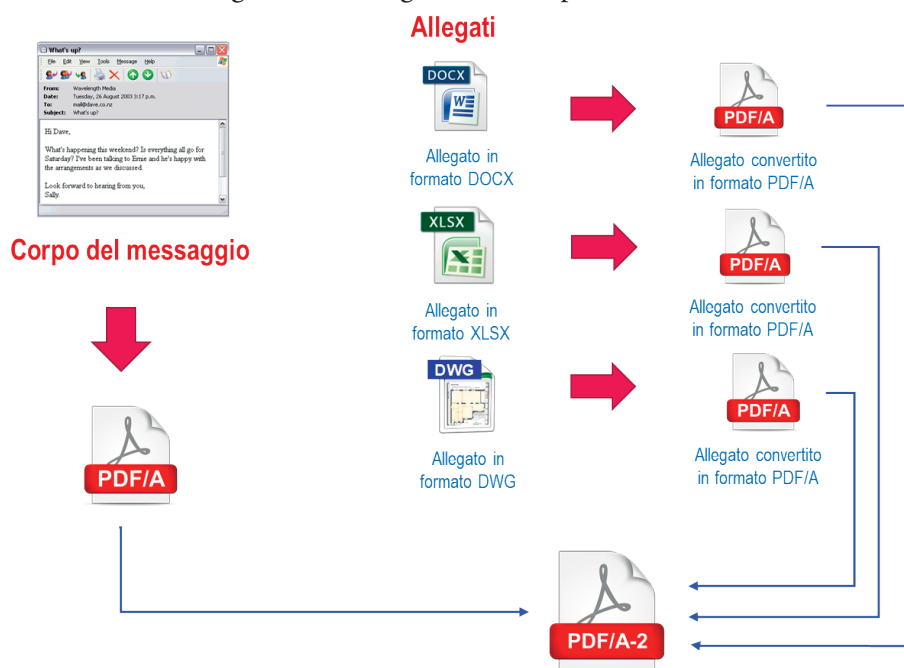


Figura 1: Esempio di utilizzo del formato PDF/A-2 per la conservazione di un messaggio di posta elettronica.

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

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

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

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

¹⁴ La famiglia di standard ISO che definiscono i vari profili PDF/A è composta come segue: PDF/A-1 (ISO 19005-1:2005 – Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4); PDF/A-2 (ISO 19005-2:2011 – Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1); PDF/A-3 (ISO 19005-3:2012 – Document management – Electronic document file format for long-term preservation – Part 3: Use of ISO 32000-1 with support for embedded files); PDF/A-4 (ISO 19005-4:2020 – Document management – Electronic document file format for long-term preservation – Part 4: Use of ISO 32000-2).

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

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

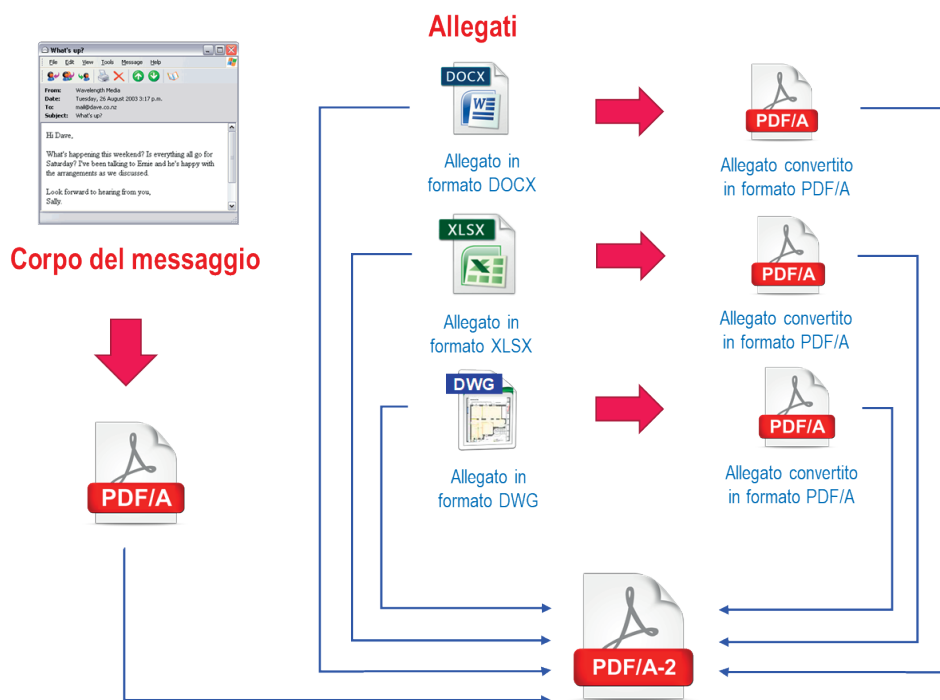


Figura 2: Esempio di utilizzo del formato PDF/A-3 per la conservazione di un messaggio di posta elettronica.

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

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

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

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

5. Recenti sviluppi nella conservazione della posta elettronica

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

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

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

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

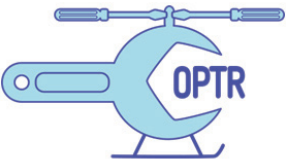
Un altro progetto che merita una segnalazione è RATOM (Review, Appraisal, and Triage of Mail) dell'Università della Carolina¹⁶. RATOM ha l'obiettivo di sviluppare un'applicazione software per assistere le istituzioni della memoria nelle attività di valutazione, selezione e scarto degli archivi di posta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

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

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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

6. Conclusioni

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

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

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

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

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

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

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

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

Riferimenti bibliografici

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

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

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

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

Le Risorse Linguistiche nell'era del Web Semantico

Un insieme di servizi informatici per la gestione di lessici e terminologie

ANDREA BELLANDI*

ABSTRACT: The application of digital technologies in lexicography and terminology is becoming increasingly pervasive. The Semantic Web offers a great potential for representing data in order to process it, understand its meaning, and share it across different scientific communities. In this area scientific communities defined computational models for the representation of linguistic data together with best practices for their use. However, applications for managing this data in the context of the Semantic Web, are a few. This paper fills this gap by presenting a set of web services that are general enough to be used as a basis for the development of task-oriented lexicography and terminology applications, such as editing, data visualization, querying, and using for linguistic/conceptual full-text searches. The final part of the contribution will describe some use cases of the presented services.

Keywords: Semantic Web, Lexicography, Terminology, OntoLex-Lemon.

1. Introduzione

Le risorse linguistiche e le tecnologie digitali, hanno recentemente stretto un rapporto di simbiosi che, per molti aspetti, sembra ormai indissolubile. Le potenzialità della rete e delle tecnologie del Web Semantico possono aiutare a studiare la storia delle lingue, aumentando le nostre conoscenze e fornendo nuove modalità di accesso alle informazioni, capaci di interconnettere vocabolari e risorse digitali in funzione di nuovi orizzonti di ricerca (Arcidiacono 2019). La lessicografia diventa così lessicografia elettronica, o computazionale, e prevede la ristrutturazione e lo sfruttamento dei dizionari tradizionali ai fini computazionali; le risorse tradizionali, come il dizionario cartaceo, vengono riorganizzate, ristrutturate e rese trattabili da strumenti informatici. Diventa sempre più importante anche la rappresentazione digitale delle terminologie; dato il rapido sviluppo dell'economia, della cultura, della scienza e della tecnologia in tutto il mondo, il valore della terminologia come risorsa di comunicazione è indiscu-

* Istituto di Linguistica Computazionale "A. Zampolli", Consiglio Nazionale delle Ricerche - Pisa, Italy. andrea.bellandi@ilc.cnr.it.

tibilmente in aumento. Diventa quindi imprescindibile una rappresentazione delle risorse linguistiche in modo tale che le macchine possano: i) elaborarle, ii) comprenderne la semantica, iii) interrogarle, iv) condividerle tra le comunità scientifiche. Questi aspetti pongono delle sfide da affrontare attraverso la collaborazione e la standardizzazione su diversi fronti: linguaggi formali di rappresentazione, interoperabilità e modelli linguistici. Le tendenze attuali nelle risorse linguistiche e nella scienza del dato, consentono oggi di immaginare un ecosistema interconnesso, su larga scala, di un insieme di tecnologie e dati linguistici aperti, interrogabili e standardizzati (Monachini e Khan 2018).

Con l'avvento del Web Semantico, le ontologie sono diventate un metodo sempre più importante per modellare formalmente la semantica del dato e condividere la conoscenza di un dominio sul Web. In questo contesto, è stato sviluppato il modello *lemon* (McCrae et al. 2012; McCrae et al. 2017), (anche chiamato *OntoLex-Lemon*), basato sul linguaggio formale Ontology Web Language (OWL), con lo scopo di fornire una ricca base linguistica per le ontologie concettuali. Esso comprende la rappresentazione delle proprietà morfologiche e sintattiche delle voci lessicali, così come l'interfaccia sintassi-semantica, ovvero la rappresentazione del loro significato rispetto ai concetti di cui esse sono le lessicalizzazioni. Negli ultimi anni, anche i terminologi hanno iniziato ad adottare modelli sviluppati nell'ambito della lessicologia, per descrivere in modo più ricco le relazioni tra i termini, spostandosi dalla metodologia tradizionalmente usata per la descrizione dei termini incentrata più sulle relazioni tassonomiche e meronimiche¹. Molti dei modelli lessicali e terminologici proposti in letteratura, hanno trovato una loro corrispondenza in *OntoLex-Lemon*, dimostrando che il modello è in grado di catturare sia aspetti più lessicografici² (Bosque-Gil, Gracia, e Montiel-Ponsoda 2017) che terminologici³ (Cimiano et al. 2017). Vedremo nella sezione 2, come sia possibile riferire definizioni concettuali sia da un senso, che direttamente da una voce lessicale.

Nel presente articolo si offrono dei servizi informatici, chiamati LexO-server, per gestire risorse lessicografiche e terminologiche secondo il modello *OntoLex-Lemon*. I servizi proposti sono offerti come una Web API, che implementa funzionalità di scrittura, lettura e modifica di repertori lessicali o terminologici, sopra la quale è possibile costruire diverse interfacce utente. Questa separazione netta tra le funzioni sul dato e l'interfaccia utente, è alla

¹ Diversamente, i modelli lessicografici, basati su un approccio orientato alla parola, ed in particolare al senso, tengono conto di un insieme più ricco di relazioni rilevanti, ad esempio quelle lessico-semantiche.

² Vedere il modulo dedicato per la lessicografia al seguente indirizzo <https://www.w3.org/2019/09/lexicog/> (ultima consultazione: 10/03/2022).

³ Strumento per convertire database terminologici nel formato standard per le terminologie Term Base eXchange (TBX) in *OntoLex-Lemon*, <http://tbx2rdf.lider-project.eu/converter/index.html> (ultima consultazione: 10/03/2022).

base delle architetture software REST. I servizi qui presentati si rivolgono agli sviluppatori del dominio linguistico, che possono utilizzare LexO-server come backend delle loro applicazioni lessicali e terminologiche. Il presente lavoro si basa sull'esperienza maturata nello sviluppo dell'applicazione LexO⁴ (Bellandi 2021), un'interfaccia utente web e collaborativa, che rende fruibile il modello *OntoLex-Lemon* ad un vasto target di utenza, astruendo l'utente dalle conoscenze delle tecnologie del Web semantico.

L'articolo è strutturato come segue: nella sezione 2 si introduce il Web Semantico descrivendone caratteristiche e vantaggi; nell'ultima parte è presentato il modello linguistico alla base dei servizi. Nella sezione 3 viene descritto LexO-server, il backend di servizi lessicali, e vengono discussi i lavori simili. Nella sezione 4 si presentano tre casi d'uso mostrando altrettante applicazioni che si basano su LexO-server: un editor di lessici, un sistema di navigazione grafica delle relazioni semantiche di un lessico e un sistema di ricerca full-text basato su informazioni linguistiche.

2. Rappresentazione del dato linguistico nel Web Semantico

La presente sezione descrive i principi base del Web semantico e riassume i vantaggi derivanti dall'utilizzo delle tecnologie su cui esso si basa. In questo contesto viene successivamente presentato il modello linguistico *OntoLex-Lemon*, lo *standard de facto* per la rappresentazione formale di lessici e terminologie, creato nell'ambito del World Wide Web Consortium (W3C).

2.1. Caratteristiche e vantaggi del Web Semantico

Riprendendo quanto affermato in (Berners-Lee, Hendler, e Lassila 2001), «il Web semantico è un tentativo di *rappresentare e collegare* i dati (risorse linguistiche, nel nostro caso) in un modo che sia comprensibile per le macchine». Il compito di *rappresentazione* viene svolto mediante l'utilizzo di linguaggi formali basati su sottoinsiemi della logica del primo ordine, in particolare la Logica Descrittiva (Baader et al. 2017). I dati vengono resi *collegabili* attraverso il paradigma dei Linked Data (LD), consentendo a ciascuna entità di un dataset (concetti, relazioni, attributi e così via) di essere identificata in modo univoco da un Internationalized Resource Identifier (IRI) e di essere disponibile sul Web tramite il Protocollo di trasferimento ipertestuale (HTTP).

L'uso di un linguaggio di rappresentazione formale, come l'OWL, fornisce un modo comune per rappresentare e codificare la semantica dei dati garantendo al contempo l'interoperabilità; consente inoltre di eseguire ragionamenti automa-

⁴ Il codice di LexO è open source e disponibile al seguente indirizzo <https://github.com/andreabellandi/LexO-lite> (ultima consultazione: 10/03/2022).

tizzati sui dati, ad esempio per garantire coerenza logica, calcolare chiusure di relazioni (ad esempio la transitività dell'iperonimia) e dedurre nuove conoscenze sulla base di tassonomie di classi e gerarchie di proprietà. Il collegamento dei dati consente di federarsi con contenuti provenienti da risorse esterne, consentendo di dedurre nuovi fatti attraverso il web, per la scoperta di nuova conoscenza. Inoltre facilita il riutilizzo dei dati stessi, in virtù della natura intrinseca dei LD.

I vantaggi della rappresentazione del contenuto lessicografico e terminologico secondo gli standard del Web Semantico sono quindi la reperibilità, la riutilizzabilità, l'accessibilità e la visibilità su scala web, secondo i principi FAIR (Wilkinson et al. 2016). Negli ultimi anni, ciò ha portato a una serie di attività guidate da comunità che hanno promosso l'adozione del paradigma dei LD per la pubblicazione di dati linguistici, in particolare l'Open Linguistics Working Group (Chiarcos, Hellmann, e Nordhoff 2012) e l'Open Knowledge Foundation (Chiarcos, Hellmann, e Nordhoff 2011). L'approccio del paradigma dei LD facilita sia lo sviluppo distribuito della rete di risorse, sia la collaborazione tra i ricercatori che forniscono e utilizzano questi dati e che impiegano un insieme condiviso di tecnologie⁵.

Il paradigma dei LD richiede l'utilizzo del modello di dati Resource Description Framework (RDF). I dati devono essere rappresentati in una struttura a tripla, costituita da un soggetto, un predicato e un oggetto, dove il soggetto è una risorsa, il predicato è una proprietà e l'oggetto è una risorsa, ad esempio `<lemon, meronym, lemon_tree>` o un *datatype* (una stringa, un intero, un decimale, ecc.), ad esempio `<lexicon, hasPhonetic, "lɛksikən">`. Insieme a RDF, RDF Schema (RDFS) aiuta a descrivere istanze, gruppi di risorse correlate e proprietà che collegano queste risorse. Ad esempio, `<synonym, hasDomain, LexicalSense>` e `<synonym, hasRange, LexicalSense>`, significa che la relazione `synonym` collega i sensi lessicali tra loro; `<MultiWordExpression, subclassOf, LexicalEntry>`, significa che ogni espressione multiword è un tipo di voce lessicale; `<hypernym, inverseOf, hyponym>`, significa che la relazione di iperonimia è l'inversa della relazione di iponimia.

Le ontologie aggiungono una rappresentazione semantica degli elementi definiti con RDF, introducendo, per esempio, operatori logici come *and* e *or*, operatori insiemistici come *unione*, *intersezione*, *disgiunzione*, il quantificatore *esistenziale* e quello *universale*⁶. Una voce lessicale, ad esempio, può essere formalmente definita in OWL come una parola, un composto o un affisso che può essere caratterizzata da una specifica parte del discorso, un modello morfologico, un'etimologia, un insieme di sensi, avente almeno una forma ortografica e al massimo una

⁵ Vedi il Linguistic Linked Open Data Cloud, <https://linguistic-lod.org/> (ultima consultazione: 10/03/2022).

⁶ Per una completa definizione di OWL si rimanda al seguente link <https://www.w3.org/TR/owl2-overview/> (ultima consultazione: 10/03/2022).

forma canonica. Il modello linguistico formale, scritto in OWL, che andremo a presentare nella sezione successiva, contiene appunto questo tipo di definizioni.

2.2. Il Modello linguistico

Il modello LExicon Model for ONtologies (*lemon*) (McCrae et al. 2017), sviluppato dal gruppo della comunità W3C OntoLex⁷, è diventato uno standard di fatto per rappresentare e pubblicare risorse lessicali nel Web Semantico. Il modello, spesso chiamato anche *OntoLex-Lemon*, nasce accogliendo un'ampia serie di casi d'uso rilevanti, e astruendo i requisiti generali da modellare. Tutti i problemi e le decisioni del modello sono pubblicamente disponibili e accessibili sul web⁸. L'architettura di *OntoLex-Lemon* è suddivisa in sei moduli, come illustrato nella Figura 1. Ciascun modulo tiene conto di aspetti specifici nella modellazione delle informazioni lessicali. Di seguito daremo una breve spiegazione di ogni modulo fornendo, laddove opportuno, dei piccoli esempi⁹.

Il modulo principale indicato come *CORE*, Fig. 1(a), si basa sul principio noto come “semantic by reference” (Buitelaar 2010), ovvero il modello permette di considerare la descrizione del comportamento morfologico e sintattico di una parola, come separata dalla descrizione ontologica dei concetti a cui la parola si riferisce. Gli elementi principali del *CORE* sono quattro: la classe `LexicalEntry` che rappresenta un insieme di voci lessicali classificate in base al loro tipo, cioè, parola singola (`Word`), parola composta (`MultiWordExpression`) o parte di una parola (`Affix`); la classe `Form` che istanzia tutte le forme flesse di una voce lessicale, incluso il lemma; la classe `LexicalSense` che rappresenta una reificazione tra la voce lessicale e il concetto; la classe `LexicalConcept` che permette di esprimere il fatto che una certa voce lessicale evochi un certo concetto mentale o un'unità di pensiero (ad esempio il verbo “*morire*” può evocare il concetto di “*morte*”). I concetti lessicali possono anche raggruppare insiemi distinti di sensi, come ad esempio i *synsets* in WordNet (Miller 1995). Per quanto riguarda la dimensione concettuale invece, in un'ottica semasiologica, in cui il senso è al centro della rappresentazione, il concetto può essere riferito dal singolo senso lessicale, tramite la proprietà *reference*. Nell'ambito di un approccio onomasiologico, in cui il

⁷ <https://www.w3.org/community/ontolex> (ultima consultazione: 10/03/2022).

⁸ I casi d'uso sono disponibili al seguente link https://www.w3.org/community/ontolex/wiki/Specification_of_Use_Cases (ultima consultazione: 10/03/2022); i requisiti generali sono disponibili al seguente link https://www.w3.org/community/ontolex/wiki/Specification_of_Requirements (ultima consultazione: 10/03/2022); problemi e decisioni del modello sono disponibili al seguente link <https://lists.w3.org/Archives/Public/public-ontolex/> (ultima consultazione: 10/03/2022).

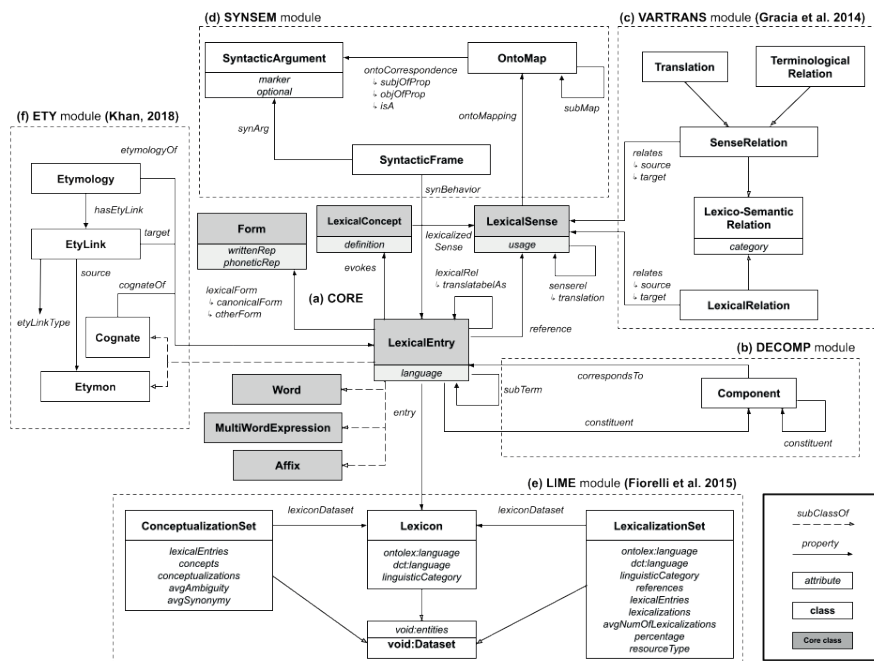
⁹ Per una completa guida al modello riferirsi al seguente link <https://www.w3.org/2016/05/ontolex/> (ultima consultazione: 10/03/2022).

concetto è al centro della rappresentazione, quest'ultimo può essere denotato direttamente dall'entrata lessicale, tramite la proprietà *denote*.

Il modulo in Fig. 1(b), denominato *DECOMP*, è dedicato a rappresentare quali sono i componenti di un composto. È possibile modellare sia l'ordine che i tratti morfologici di ogni componente.

Tutte le relazioni tra le parole, sono modellate dal modulo Variazione e Traduzione nella Fig. 1(c), denominato *VARTRANS*. Vengono presi in considerazione due tipi di relazioni: le relazioni lessicali, che collegano direttamente tra loro voci lessicali, ad esempio `<LexO, acronymOf, Lexicon_and_Ontology>`¹⁰, e le relazioni tra sensi, che collegano direttamente tra loro i sensi lessicali, ad esempio la sinonimia. Se è necessario rappresentare informazioni aggiuntive sulla relazione, è possibile reificare la relazione stessa e modellarla come classe. In questo modo è possibile specificare alcune proprietà della relazione, ad esempio un grado di confidenza. Consideriamo il caso della traduzione (Gracia et al. 2014). Supponiamo che un algoritmo di allineamento automatico proponga che “casseruola” sia la traduzione italiana di “pot” in inglese con un grado di confidenza pari a 0,8. La classe *Translation* avrà un elemento *t* con un *source* `<t source casseruola>`, un *target* `<t target pot>` e un grado di confidenza `<t confidence 0,8>`.

¹⁰ *OntoLex-Lemon* è agnostico ad uno specifico vocabolario linguistico. Le categorie grammaticali, le relazioni semantiche, e tutti gli oggetti linguistici in genere, sono definiti in vocabolari esterni. Nel nostro caso, come da raccomandazione della comunità, LexO-server utilizza il vocabolario LexInfo <https://lexinfo.net/> (ultima consultazione: 10/03/2022).

Figura 1: Architettura del modello *OntoLex-Lemon*.

Il modulo Sintassi e Semantica illustrato nella Fig. 1(d), denominato *SYNSEM*, consente di descrivere i frames sintattici di un elemento lessicale, specificando il tipo (frame verbale, frame transitivo, frame sostantivo preposizionale e così via), gli argomenti sintattici, le loro funzioni (soggetto, oggetto, ecc.), la loro posizione e l'eventuale obbligatorietà di presenza nel frame. Il modulo permette infine di far corrispondere ogni argomento del frame a un'entità di un'ontologia concettuale. Consideriamo ad esempio il frame transitivo “X possiede Y”. Il predicato può essere corrisposto su una proprietà ontologica “*possiede*”, dove il soggetto grammaticale X e il complemento oggetto Y si riferiscono, rispettivamente, al dominio e al codominio della proprietà, che saranno classi ontologiche. Il modulo è in grado di rappresentare anche strutture multi-argomentali più complesse.

Il modulo Metadata (Fig. 1(e)), denominato *LIME* (Fiorelli, Pazienza, e Stellato 2013), permette di rappresentare informazioni quantitative e qualitative a grana grossa sul dataset linguistico, ad esempio il numero di entrate, di sensi lessicali, di concettualizzazioni, i metadati descrittivi su un sottoinsieme di un insieme di lessicalizzazioni (per il sottoinsieme che rappresenta tutte le lessicalizzazioni per un certo tipo di entità concettuale riferita), o l'endpoint per accedere alla risorsa. Lo scopo principale è quello di consentire all'uomo e

alle macchine di conoscere quale materiale lessicale è disponibile e quindi di capire meglio come utilizzarlo per scopi specifici.

L'ultimo modulo che presentiamo (Fig. 1(f)), è denominato *ETY* e non fa ancora parte dell'architettura ufficiale di *OntoLex-Lemon* e gestisce l'etimologia. Il modulo prevede una prima semplice gestione dell'aspetto etimologico delle parole (Khan 2018). Le etimologie (classe *Etymology*) rappresentano la storia di una parola tracciando una sorta di albero genealogico linguistico o genealogia. Per questo la classe *LexicalEntry* del modulo *core* è estesa con due sottoclassi: la classe *Etymon* e la classe *Cognate*. Una parola può essere quindi associata ai suoi etimi e affini direttamente o indirettamente prendendo in esplicita considerazione altri etimi/affini attraverso relazioni che rappresentano dei processi storico-linguistici, rappresentati dalla classe *EtymLink*. Questi processi sono: il prestito, che si riferisce al processo mediante il quale gli elementi linguistici sono trasferiti da una lingua all'altra tramite il contatto linguistico (valore *borrowing* della proprietà *etymLinkType* in Fig. 1(f)), e l'eredità che si riferisce all'eredità di parole da una lingua madre, o a fasi precedenti della stessa lingua (valore *inheritance* della proprietà *etymLinkType* in Fig. 1(f)).

3. LexO-server

In questa sezione viene proposta una breve analisi dello stato dell'arte relativo alle applicazioni e ai servizi web disponibili per la gestione di risorse linguistiche, in particolare facendo riferimento a soluzioni che operano nel contesto del Web Semantico e col modello *OntoLex-Lemon*. Successivamente verrà fornita una descrizione dettagliata dell'architettura e del funzionamento dei servizi proposti.

3.1. Relazione con lavori simili

In generale, esistono diversi sistemi o servizi per il trattamento di risorse linguistiche, ognuno basato su modelli differenti e con scopi diversi, ad esempio per la costruzione di lessici di lingua, per la retro digitalizzazione di dizionari, o per la gestione di terminologie. Nel contesto del Web Semantico, a conoscenza dell'autore, il lavoro qui proposto è unico nel suo genere. Per quanto riguarda il modello *OntoLex-Lemon* infatti, esistono pochi strumenti dedicati alla sua fruizione, ma sono tutti applicazioni *full stack* che si rivolgono direttamente all'utente finale, al lessicografo o al terminologo e non servizi generici come quelli qui proposti, rivolti a chi costruisce applicazioni finali. Daremo di seguito una breve descrizione di tali strumenti. *Lemon source*¹¹ è uno strumento in stile Wiki per la manipolazione e la pubblicazione di risorse

¹¹ <https://lemon-model.net/download/source.php> (ultima consultazione: 10/03/2022).

lemon. Consente di caricare un lessico e condividerlo con altri utenti. È un progetto open source, ed è disponibile gratuitamente online per l'uso. Tuttavia, gestisce versioni vecchie del modello *OntoLex-Lemon* e sembra essere uno strumento non più mantenuto. Lemonade (Rico e Unger 2015), è un editor di *lemon* basato su un insieme di *lemon patterns* (McCrae e Unger 2014), una grammatica che consente di costruire entrate lessicali esprimendole in un semplice linguaggio formale orientato all'utente, senza occuparsi della loro codifica in OWL. Il linguaggio però non permette di specificare le relazioni tra voci lessicali o tra sensi lessicali. VocBench (Stellato et al. 2017) è uno strumento web e collaborativo per costruire ontologie OWL, thesauri Simple Knowledge Organisation System (SKOS) e dataset RDF generici multilingue. Tra le altre caratteristiche offre anche la funzionalità di editing per risorse *OntoLex-Lemon*. Infine LexO (Bellandi 2021) è un sistema dedicato all'editing di risorse *OntoLex-Lemon* con un'interfaccia utente che astrae le complessità del modello allo studioso. Una comparazione tra LexO e VocBench è discussa dagli autori in (Fiorelli et al. 2020).

Esistono invece lavori che propongono servizi Representational State Transfer (REST) simili a quelli qui descritti¹², ma che non si basano sul modello *OntoLex-Lemon*. Ad esempio, un set di API¹³ è stato sviluppato da K Dictionaries¹⁴ (ex Kernerman Dictionaries), per la fruizione di molti dizionari per lo più bilingui in oltre 50 lingue, inclusa una serie di dizionari per studenti apprendenti di varie lingue. L'Università di Oxford ha invece sviluppato una API¹⁵ che contiene diversi dizionari di inglese monolingue (destinati a madrelingua inglese) e dizionari bilingue in oltre 30 lingue abbinate all'inglese. Infine l'Università di Cambridge ha sviluppato una API¹⁶ per l'accesso al loro dizionario monolingua Inglese per studenti. Tutti questi servizi sono comunque orientati esclusivamente all'accesso e alla fruizione di dizionari mono e pluri lingue già esistenti, mentre la proposta dell'autore è focalizzata anche sull'offerta di servizi per la costruzione, in modo da servire un ampio set di possibili tasks per lo sviluppo di applicazioni dedicate, per la lessicografia e la terminologia, da parte di terzi. Un altro aspetto che caratterizza e distingue LexO-server dai servizi sopra citati, è la possibilità di riferire ontologie extra-linguistiche per assegnare una descrizione formale al senso o direttamente all'entrata, permettendo quindi anche un accesso concettuale alla risorsa terminologica e lessicale.

¹² Un'interessante rassegna può essere consultata al link <https://www.lexiconista.com/dictionary-apis/> (ultima consultazione: 10/03/2022).

¹³ <https://api.lexicala.com> (ultima consultazione: 10/03/2022).

¹⁴ <https://lexicala.com/> (ultima consultazione: 10/03/2022).

¹⁵ <https://developer.oxforddictionaries.com> (ultima consultazione: 10/03/2022).

¹⁶ <https://dictionary-api.cambridge.org> (ultima consultazione: 10/03/2022).

3.2. Architettura e struttura dei servizi

LexO-server è una API REST che espone un insieme di funzionalità (o servizi) che permettono di essere invocate da programmi esterni via rete e, a seguito di un'elaborazione, restituiscono un risultato. Nello specifico LexO-server si basa sul protocollo HTTP e scambia i dati in formato JavaScript Object Notation (JSON)¹⁷. I servizi sono conformi a OpenAPI¹⁸, una specifica per file di interfaccia leggibili dalle macchine per descrivere, produrre, consumare e visualizzare servizi REST¹⁹.

I servizi sono raggruppati nelle seguenti categorie, in base alla tipologia di funzionalità che realizzano:

- *Linguistic Vocabulary*. Questo gruppo di servizi è volto a fornire la lista dei valori del vocabolario linguistico utilizzato da LexO-server (*LexInfo* come descritto in sezione 2.2), ad esempio le categorie grammaticali, le relazioni semantiche, i tratti morfologici, i tipi di frames sintattici, ecc. Lo scopo di questi servizi è quello di permettere alle interfacce delle applicazioni che li useranno di mostrare tali valori come possibili scelte, in opportuni menù in fase di editing della risorsa.
- *Lexicon Creation*. I servizi di questo gruppo permettono di creare tutti gli elementi del modello presentato in sezione 2.2. Per ognuno di essi, i servizi registrano l'utente creatore e la data di creazione (si veda in seguito la gestione degli utenti). Le entrate lessicali vengono gestite con uno stato a tre valori: i) *working*, per indicare che l'utente ha creato un'entrata e che questa è in fase di lavorazione (inserimento forme, sensi, ecc.); ii) *completed*, per indicare che l'utente ha terminato la compilazione dell'entrata; iii) *revised*, per indicare che l'entrata è definitiva e non è più modificabile. Questa gestione può essere utile nel contesto dello sviluppo di editors collaborativi in cui, ad esempio, è prevista una figura di utente revisore che ha il compito di validare le entrate redatte da utenti lessicografi o terminologi.
- *Lexicon Data*. Questo gruppo colleziona tutti i servizi per accedere ai dati della risorsa. È previsto un recupero dei dati di ogni elemento sia di sintesi, cosicché ad esempio un'interfaccia possa presentarli come liste o come alberi, sia di dettaglio, ad esempio per la loro visualizzazione in tabelle o forms per l'editing. I servizi di ricerca del dato permettono la specifica di parametri di input, che implementano ricerche avanzate.
- *Lexicon Update*. I servizi di questo gruppo permettono di modificare gli elementi del modello, aggiungendo o modificando proprietà e valori degli stessi.

¹⁷ <https://www.json.org/json-en.html> (ultima consultazione: 10/03/2022).

¹⁸ <https://www.openapis.org/> (ultima consultazione: 10/03/2022).

¹⁹ <https://licodemo.ilc.cnr.it/LexO-backend-beta/> (ultima consultazione: 10/03/2022).

- *Lexicon Deletion*. I servizi di questo gruppo permettono di eliminare elementi o singole proprietà. Al momento i servizi adottano una politica di cancellazione *lazy*, nel senso che se un elemento da eliminare è collegato ad altri elementi della risorsa, il servizio non procede con la cancellazione ma ritorna un codice di errore.
- *Lexicon Statistics*. Questo gruppo di servizi può essere di supporto a differenti tasks, fornendo alcuni elementi quantitativi della risorsa (numero di lingue, numero di voci per lingua, numero di parole per tipo, ecc.)
- *Graph Visualization Support*. I servizi implementano funzionalità tipiche della navigazione di un grafo. Al momento viene ipotizzato che i nodi siano i sensi e gli archi tutte le relazioni possibili tra sensi²⁰. Alcune relazioni possono avere delle caratteristiche specifiche, ad esempio la sinonimia può essere simmetrica, l'iperonimia transitiva, l'olonimia l'inversa della meronimia, ecc. Di conseguenza, i servizi specificano anche se l'informazione relativa ad una relazione è inferita o esplicita. Questo può essere utile in un task di visualizzazione perché, ad esempio, un arco potrebbe essere colorato in maniera diversa a seconda della natura della relazione che rappresenta.
- *Query Expansion Support*. I servizi di questo gruppo sono funzionali alla ricerca full-text e quindi prevedono l'espansione di: un'entrata con le sue forme flesse e i relativi tratti morfologici; un concetto con le forme dei sensi che lo riferiscono; un senso con le sue relative forme.

Per quanto riguarda la parte concettuale, LexO-server offre la possibilità di costruire un'ontologia SKOS²¹. SKOS consiste in una serie di classi RDFS e proprietà di RDF che sono usate per rappresentare il contenuto e la struttura di base dei cosiddetti concept-schema cioè thesaurus, schemi di classificazione, elenchi di titoli di argomento, tassonomie, glossari e altri tipi di vocabolario. La classe `LexicalConcept` del modulo core di *OntoLex-Lemon* è definita come una classe di SKOS e quindi i concetti lessicali possono essere strutturati tra di loro, ad esempio, tramite relazioni di gerarchia. LexO-server comunque gestisce il vocabolario SKOS anche indipendentemente dalla formalizzazione del concetto lessicale e permette quindi di costruire ontologie concettuali in SKOS, che possono rappresentare i concetti le cui lessicalizzazioni sono descritte tramite il modello linguistico.

Come riportato in Fig. 2, i componenti principali dell'architettura sono, i) la parte di logica dei servizi, ii) il repository semantico e iii) un insieme di

²⁰ Al momento i servizi ritornano oggetti secondo il formato della libreria Cytoscape <https://cytoscape.org/>. È una delle librerie free e open source più utilizzate per la costruzione di sistemi di visualizzazione (ultima consultazione: 10/03/2022).

²¹ Simple Knowledge Organisation System (SKOS) <https://www.w3.org/2004/02/skos/> (ultima consultazione: 10/03/2022).

servizi REST esterni e integrabili in LexO-server, tramite dei dati di configurazione.

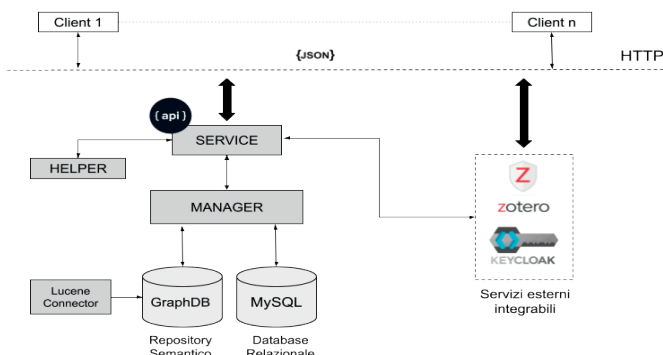


Figura 2: Architettura di LexO-server.

Riguardo la parte di logica, i servizi sono scritti in Java 15²². Ogni servizio effettua la verifica sintattica della propria invocazione, validando gli eventuali parametri in input con l'aiuto del modulo di Helper. Il servizio poi demanda l'elaborazione della funzionalità che esso implementa, al modulo del Manager che si interfaccia con il repository. Quest'ultimo risponde al Manager e, in base al tipo di risposta, il manager chiede al modulo di Helper di preparare il dato di risposta nella struttura dati opportuna. L'Helper trasmette il JSON al servizio, che poi lo invia al chiamante. In caso di problemi o errori, il servizio risponde con il relativo codice HTTP di stato.

GraphDB è la tecnologia scelta per la realizzazione del repository semantico. È un'implementazione dell'interfaccia di storage e reasoning del framework RDF4J. Gestisce basi di conoscenza in OWL 2, e ha delle ottime performance di scalabilità su dataset contenenti bilioni di triple. Si integra con Apache Lucene²³, un sistema di indicizzazione del testo, grazie al quale si ottengono alte prestazioni in fase di interrogazione e recupero di informazioni. I servizi non fanno riferimento ad una particolare istanza del repository. Chi utilizza i servizi dovrà installare un'istanza di GraphDB, anche su una macchina fisica o virtuale diversa da quella dove sono installati i servizi, e configurare il proprio repository, vuoto o importando una risorsa *OntoLex-Lemon* esistente.

LexO-server è attualmente integrabile con Zotero²⁴, un gestore di record bibliografici e con KeyCloack²⁵, un manager per la gestione degli utenti. Per quanto riguarda la bibliografia, ogni entità lessicale (entrata, forma, senso ed etimologia) può essere associata ad elementi bibliografici presenti in una li-

²² <https://openjdk.java.net/projects/jdk/15/> (ultima consultazione: 10/03/2022)

²³ <https://lucene.apache.org> (ultima consultazione: 10/03/2022).

²⁴ <https://www.zotero.org> (ultima consultazione: 10/03/2022).

²⁵ <https://www.keycloak.org/> (ultima consultazione: 10/03/2022).

breria di Zotero, tramite la sua API REST²⁶. Per quanto riguarda gli utenti, la cui anagrafica, i diritti e i permessi sono gestiti da KeyCloack, i servizi di LexO-server possono essere configurati in modo tale che richiedano un'autenticazione basata sulla base dati di KeyCloack, interfacciandosi con la sua API REST²⁷. La parte di autorizzazione invece (diritti e permessi), richiede un po' di lavoro in più rispetto alla semplice configurazione, ovvero la definizione dei ruoli utente e dei diritti sulle risorse, che dipendono dall'applicazione finale che andrà ad essere sviluppata. Il codice di ogni servizio dovrà essere modificato di conseguenza.

Alla luce di quanto detto, in un'apposita tabella di configurazione gestita in MySQL di LexO-server, è possibile specificare, tra gli altri:

- l'endpoint del repository semantico (GraphDB);
- l'endpoint della libreria *Zotero* all quale ci si vuole connettere;
- l'endpoint di *Keycloak* per la gestione degli utenti;
- un parametro per stabilire se i servizi sono aperti o il loro utilizzo deve richiedere un'autenticazione;
- il tipo di lessicalizzazione dei concetti SKOS (utilizzo della proprietà *label* del vocabolario RDFS o utilizzo delle proprietà *prefLabel*, *altLabel* e *hiddenLabel* del vocabolario SKOS);
- l'insieme delle lingue ammissibili per le lessicalizzazioni degli elementi di SKOS e la lingua di default per la loro visualizzazione.

I servizi funzionano con un web server come Apache Tomcat o Jetty. Nei casi d'uso che saranno presentati nella sezione successiva, i servizi utilizzano Tomcat 9.03.

4. Casi d'uso

In questa sezione vengono presentati tre strumenti informatici che utilizzano LexO-server. Ognuno di essi è dedicato a un caso d'uso diverso: il primo, sviluppato nel contesto del P.R.I.N. "Lingue e culture dell'Italia antica: linguistica storica e modelli digitali"²⁸, è dedicato alla creazione di lessici multilingua; il secondo permette una visualizzazione sofisticata delle relazioni tra sensi di un lessico; il terzo, sviluppato all'interno del "Progetto di Traduzione del Talmud Babilonese"²⁹, permette la ricerca su base linguistica all'interno di un

²⁶ https://www.zotero.org/support/dev/web_api/v3/basics (ultima consultazione: 10/03/2022).

²⁷ <https://www.keycloak.org/docs-api/15.0/rest-api/index.html> (ultima consultazione: 10/03/2022).

²⁸ <https://www.prin-italia-antica.unifi.it/> (ultima consultazione: 10/03/2022).

²⁹ <https://www.talmud.it/> (ultima consultazione: 10/03/2022).

testo. Le tre sottosezioni seguenti sono dedicate ad una breve descrizione dei sopra elencati sistemi.

4.1. Editing

Lo strumento chiamato EpiLexO (Quochi et al. 2022), mostrato in Fig. 3, è un editor web collaborativo di lessici storici multilingua. Nello specifico rappresenta un esempio di utilizzo di LexO-server per la costruzione di un lessico a partire da epigrafi. Come mostrato in Fig. 3, la colonna alla sinistra dell'interfaccia mostra la struttura del lessico con alcuni dati di sintesi come lo stato di editing della voce lessicale e l'utente che l'ha creata. In alto è possibile effettuare delle ricerche avanzate in base ad alcuni parametri. Questa è la parte di interfaccia che utilizza i servizi del gruppo Lexicon Data. Al centro ci sono i dati di dettaglio delle entità selezionate nella colonna di sinistra. Questi possono essere creati, modificati ed eliminati tramite i servizi dei gruppi *Lexicon Create*, *Lexicon Update*, e *Lexicon Deletion*. Nella colonna più a destra vengono editate le informazioni di contesto di ciò che è stato selezionato nella colonna di sinistra come la bibliografia, i links esterni, le note e alcuni metadati. Anche questa parte dell'interfaccia utilizza gli stessi gruppi di servizi usati dalle altre.

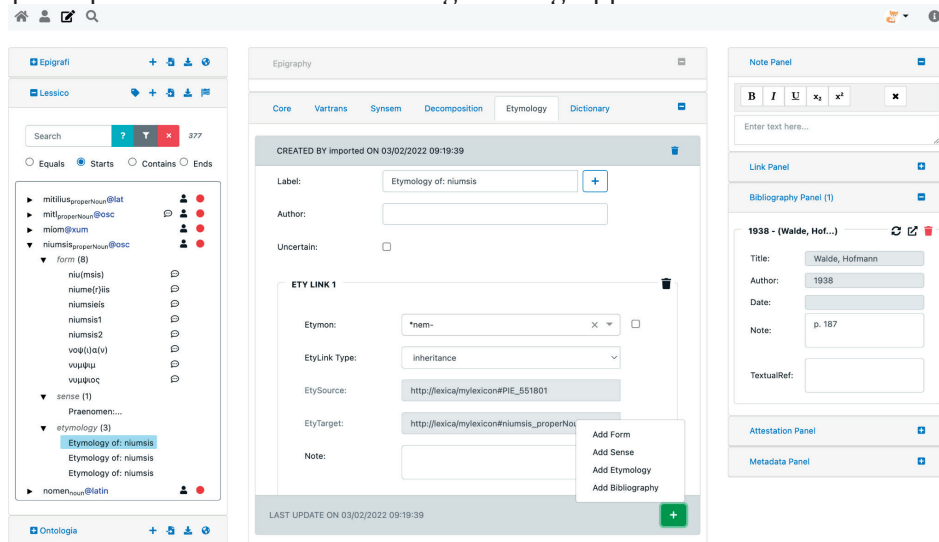


Figura 3: EpiLexO, un editor di lessici.

In questo caso i servizi sono configurati anche col gestore degli utenti (*Keycloak*); l'utilizzatore del servizio dovrà quindi autenticarsi per poter invocare LexO-server.

4.2. Visualizzazione

Il lavoro presentato in (Colombo e Giovannetti 2022), è un primo esperimento di strumento per permettere la navigazione di “PAROLE-SIMPLE-CLIPS” (PSC), un lessico computazionale dell’italiano, sviluppato dal 1996 al 2003 dell’Istituto di Linguistica Computazionale “A. Zampolli” (Rumy et al. 2002). Lo strumento mostra i sensi lessicali e le loro relazioni tramite un grafo. Come riportato in Fig. 4, l’interfaccia si divide in due parti. Nella colonna a sinistra viene mostrata la lista dei sensi disponibili nel lessico con la possibilità di filtrarli in base a vari parametri. Questa parte di interfaccia utilizza i servizi del gruppo Lexicon Data. Al centro è possibile navigare graficamente i sensi in base alle loro relazioni. Nell’esempio di Fig. 4 è mostrato un esempio di relazioni di meronimia e olonimia del senso di “nocca” e “unghia”. La relazione che i due sensi sono meronimi di “dito” è esplicita. In questo caso viene chiesto al sistema di ricostruire, tramite le regole di inferenza, la relazione di meronimia fino al senso di braccio (visualizzata in un colore diverso). I servizi relativi a questo tipo di navigazione sono quelli del gruppo *Graph Visualization Support*.

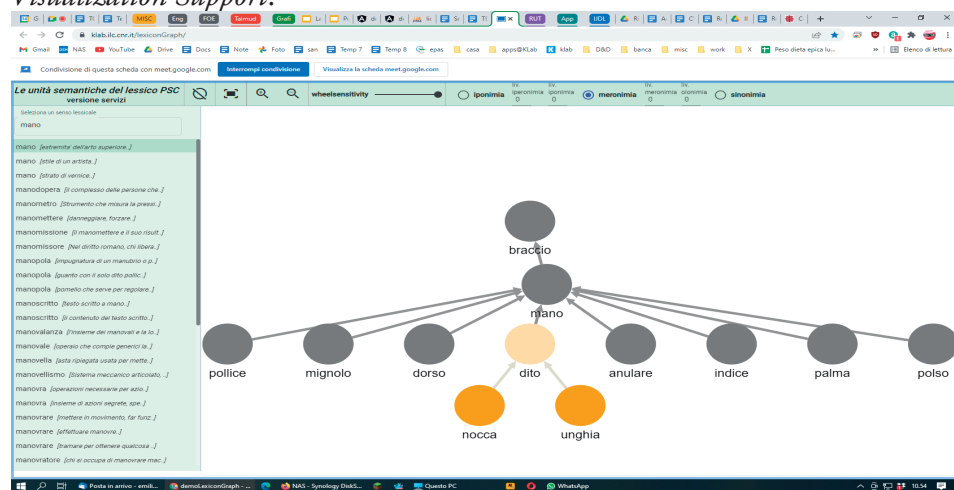


Figura 4: Visualizzatore di relazione semantiche.

4.3. Supporto alla ricerca full-text

In (Giovannetti et al. 2021), gli autori hanno sviluppato uno strumento per accedere ad un testo su base linguistica e concettuale. In questo caso vengono utilizzati i servizi del gruppo Lexicon Data e quelli del gruppo Query Expansion Support. In Fig. 5 viene mostrato un esempio di ricerca linguistica sul testo. Nell’ambito del “Progetto di Traduzione del Talmud Babilonese” i testi utilizzati sono le traduzioni italiane di alcuni trattati del testo sacro

della cultura ebraica³⁰. In questo caso, l'utente chiede i contesti della parola "recipiente". Come mostrato in Fig. 5, l'interfaccia propone le voci lessicali ambigue ritornate da LexO-server. L'utente seleziona i sensi desiderati della categoria grammaticale "nome" e specifica che le forme desiderate devono essere al plurale. LexO-server espande l'interrogazione con gli iponimi diretti delle forme scelte (Fig. 5 in basso) e invia l'elenco delle forme risultanti al modulo di ricerca nel testo. Nella parte destra dell'interfaccia è possibile vedere alcuni contesti restituiti al sistema. I servizi del gruppo *Query Expansion Support*, partendo dal lemma "recipiente" hanno permesso di ottenere risultati come "le anfore ...", "... proprietari di barili ...", "... bicchieri colmi di vino ...". La risorsa linguistica usata è PSC. Siccome, ogni senso riferisce uno dei tipi rappresentati nell'ontologia SIMPLE (Lenci et. al. 2001), è possibile anche accedere ai contesti desiderati del testo partendo da caratteristiche concettuali (quindi extra-linguistiche). Componendo un'interrogazione relativa ai sensi riferiti al campo semantico di "Animale del cielo", si otterranno ad esempio i contesti relativi alle forme (solo lemma o anche forme flesse) dei sensi di "uccello", "mosca", "cavalletta" e così via.

The screenshot displays the LexO-server interface. On the left, a sidebar titled 'Limit the research context' lists various resources with checkboxes. The main search area has tabs for 'Keyword', 'Form/Lemma', and 'Semantic traits'. The 'Keyword' tab is active, showing the search term 'recipiente' and its grammatical category 'noun'. Below this, the 'Computational Lexicon' section shows a list of senses for 'recipiente', including 'recipiente - adjective' and 'recipiente - noun - [number: plural]'. The 'Form/Lemma' tab is also visible, showing a list of forms corresponding to the parameters entered. The 'Semantic traits' tab is also visible, showing a list of semantic traits for the search term. The bottom section shows a list of occurrences of the search term in the text, with examples like 'E Ghidón e cento uomini che erano con lui vennero al bordo del campo all'inizio della veglia di mezzo, quando avevano appena stabilito le guardie, e suonarono lo shofar e ruppero le anfore che erano nelle loro mani (Giud. 7:19);' and 'Le anfore intatte vanno al fiume,'.

Figura 5: Interfaccia di ricerca full-text basata su conoscenza linguistica.

5. Conclusioni

In questo articolo è stato presentato LexO-server, un insieme di servizi informatici, sotto forma di API REST, per la gestione di lessici e terminologie, nel contesto del Web Semantico. I servizi sono di libero utilizzo e a codice

³⁰ Il testo si intende linguisticamente annotato.

aperto. Sono rivolti a chi vuole sviluppare applicazioni lessicali e/o terminologiche che si basano sugli standard del Web Semantico. Sono stati presentati tre casi d'uso che mostrano la generalità e la flessibilità di LexO-server, mostrando il loro utilizzo in altrettanti strumenti.

I lavori futuri proseguiranno su due fronti, ovvero il miglioramento dei servizi attuali e il concepimento di nuovi servizi per la gestione di ontologie concettuali in OWL.

Per quanto riguarda il primo punto, è prevista un'attività di revisione, risoluzione di bugs, eventuale aggiornamento e ampliamento dei servizi, sulla base anche dei feedback ricevuti da chi sviluppa applicazioni che utilizzano i servizi stessi.

Riguardo la seconda linea, si intende fornire dei servizi di supporto sia alla costruzione e fruizione di ontologie OWL, sia alle interrogazioni avanzate che sfruttano la potenza espressiva di OWL per accedere in maniera più sofisticata dal concetto al dato linguistico.

Riferimenti bibliografici

- Arcidiacono, Salvatore. 2019. *Per una Lessicografia Elettronica degli Antichi Volgari Italiani*. Quaderni di Artesia. Centro Studi Filologici e Linguistici Siciliani.
- Baader, Franz, Ian Horrocks, Carsten Lutz, and Uli Sattler. 2017. *An Introduction to Description Logic*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781139025355>.
- Bellandi, Andrea. 2021. "LexO: an Open-source System for Managing OntoLex-Lemon Resources." *Language Resources and Evaluation* 55, no. 4: 1093-126.
- Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284, no. 5: 34-43.
- Bosque-Gil, Julia, Jorge Gracia, and Elena Montiel-Ponsoda. 2017. "Towards a Module for Lexicography in OntoLex." In *Proceedings of the 1st Workshop on the OntoLex Model 18 June 2017 Galway, Ireland*, 74-84.
- Buitelaar, Paul. 2010. "Ontology-based Semantic Lexicons: Mapping between Terms and Object Descriptions." In *Ontology and the Lexicon. A Natural Language Processing Perspective*, edited by Chu-ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, Laurent Prevot, 212-23. Cambridge: Cambridge University Press.
- Chiarcos, Christian, Sebastian Hellmann, and Sebastian Nordhoff. 2011. "Towards a Linguistic Linked Open Data cloud: The Open Linguistics Working Group." *TAL* 52, no. 3: 245-75.

- Chiarcos, Christian, Sebastian Hellmann, and Sebastian Nordhoff. 2012. "Linking linguistic resources: Examples from the open linguistics working group." In *Linked Data in Linguistics*, 201-16. Berlin Heidelberg: Springer.
- Cimiano, Philipp, John Philipp McCrae, Víctor Rodríguez-Doncel, Tatiana Gornostay, Asunción Gómez-Pérez, Benjamin Siemoneit, and Andis Lagzdins. 2017. "Linked Terminologies: Applying Linked Data Principles to Terminological Resources." In *Proceedings of the Electronic Lexicography in the 21st Century 19-21 September 2017 Leiden, the Netherlands*, 504-17.
- Colombo, Marianna, e Emiliano Giovannetti. 2022. "La Visualizzazione Grafica di Sensi e Relazioni Semantiche di un Lessico Computazionale della Lingua Italiana." In *Proceedings of AIUCD 2022 Digital Cultures. Intersections: Philosophy, Arts, Media 01-03 June 2022 Lecce*, 155-60.
- Fiorelli, Manuel, Armando Stellato, Tiziano Lorenzetti, Andrea Turbati, Peter Schmitz, Enrico Francesconi, Najeh Hajlaoui, and Brahim Batouche. 2020. "Editing OntoLex-Lemon in VocBench 3." In *Proceedings of the 12th Conference on Language Resources and Evaluation, Marseille, 11-16 May 2020*, 7196-205.
- Fiorelli, Manuel, Maria Teresa Pazienza, and Armando Stellato. 2013. "LIME: Towards a Metadata Module for Ontolex." In *Proceedings of the 2nd Workshop on Linked Data in Linguistics: Representing and Linking Lexicons, Terminologies and Other Language Data September 2013 Pisa, Italy*.
- Giovannetti, Emiliano, Davide Albanesi, Andrea Bellandi, Simone Marchi, Mafalda Papini, and Flavia Sciolette. 2021. "The Role of a Computational Lexicon for Query Expansion in FullText Search." In *Proceedings of the 8th Italian Conference on Computational Linguistics 26-28 June 2022 Milan*.
- Gracia, Jorge, Elena Montiel-Ponsoda, Daniel Vila-Suero, and Guadalupe Aguado-de-Cea. 2014. "Enabling Language Resources to Expose Translations as Linked Data on the Web." In *Proceedings of 9th Language Resources and Evaluation Conference May 2014 Reykjavik, Iceland*, 409-13.
- Khan, Anas Fahad. 2018. "Towards the Representation of Etymological Data on the Semantic Web." *Information* 9, no. 12: 304.
- Lenci, Alessandro, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, et al. 2000. "SIMPLE: A General Framework for the Development of Multilingual Lexicons." *International Journal of Lexicography* 13, no. 4: 249-63.

- McCrae, John Philipp, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. "The Ontolex-Lemon model: development and applications." In *Proceedings of the Electronic Lexicography of the 21st Century September 2017 Leiden, the Netherlands*, 19-21.
- McCrae, John Philipp, and Christina Unger. 2014. "Design Patterns for Engineering the Ontology-Lexicon Interface." In *Proceedings of the Towards the Multilingual Semantic Web*, edited by Paul Buitelaar, Philipp Cimiano, Berlin Heidelberg: Springer, 15–30.
- McCrae, John Philipp, Guadalupe Aguado de Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, et al. 2012. "Interchanging Lexical Resources on the Semantic Web." *Language Resources and Evaluation* 46, 701-19.
- Miller, George A. 1995. "WordNet: a lexical database for English." *Communications of the ACM* 38, no. 11: 39-41. <https://doi.org/10.1145/219717.219748>.
- Monachini, Monica, and Anas Fahad Khan. 2018. "Towards the Construction of a Lexical Data and Technology Ecosystem: The Experience of ILC-CNR." In *Proceedings of the Language Resources and Evaluation Conference, Workshop Globalex 7-12 May 2018 Miyazaki, Japan*, 52-4.
- Quochi, Valeria, Andrea Bellandi, Fahad Khan, Michele Mallia, Francesca Murano, Silvia Piccini, Luca Rigobianco, Alessandro Tommasi, and Cesare Zavattari. 2022. "From Inscriptions to Lexica and back: A Platform for Editing and Linking the Languages of Ancient Italy." In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022) 20-25 June 2022 Marseille*.
- Rico, Mariano, and Christina Unger. 2015. "Lemonade: A Web Assistant for Creating and Debugging Ontology Lexica." *Natural Language Processing and Information Systems NLDB 2015 Lecture Notes in Computer Science* 9103, edited by Chris Biemann, Siegfried Handschuh, Andre Freitas, Farid Meziane, Elisabeth Méitas, Cham: Springer, 448-52. https://doi.org/10.1007/978-3-319-19581-0_45.
- Ruimy, Nilda, Monica Monachini, Raffaella Distante, Elisabetta Guazzini, Stefano Molino, Marisa Ulivieri, Nicoletta Calzolari, and Antonio Zampolli. 2002. "Clips, a Multi-level Italian Computational Lexicon: A Glimpse to Data." In *Proceedings of the 3rd International Conference on Language Resources and Evaluation May 2002 Las Palmas, Canary Islands, Spain*.

- Stellato, Armando, Andrea Turbati, Manuel Fiorelli, Tiziano Lorenzetti, Eugeniu Costetchi, Christine Laaboudi, Willem Van Gemert, and Johannes Keizer. 2017. "Towards VocBench 3: Pushing Collaborative Development of Thesauri and Ontologies Further Beyond." In *Proceedings of the 17th European Networked Knowledge Organization Systems Workshop 21 September 2017 Thessaloniki, Greece*, CEUR-WS, 39-52.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3.

E se l'archivio non rispecchia l'istituto?

Pavone e il rispecchiamento: analisi di una bozza preliminare

GIORGIA DI MARCANTONIO*

ABSTRACT: The paper reports the transcription and analysis of the first version of Claudio Pavone's article *Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?*, published in 1970 in the *Rassegna degli Archivi di Stato*. The draft has recently been discovered during the description and arrangement of the scholar's archive deposited in the Central State Archives. Although the contents of the two versions are not very different, the preparatory study, on which Pavone began working as early as 1968, is full of examples and unpublished considerations that have only partially emerged in the published article. In comparing the two versions, further elements of reflection can be gathered regarding the influence that Pavone's thought has had, and continues to have, on contemporary archival science.

Keywords: Claudio Pavone, Inventariazione, Metodo storico, Ordinamento, Teoria archivistica.

1. Introduzione

Nel novembre del 2008 Claudio Pavone comunica al Sovrintendente dell'Archivio centrale dello Stato¹ di voler donare parte della propria documentazione all'Istituto. Qualche mese più tardi vengono depositate in ACS 149 buste corredate da un elenco di versamento di mano dello stesso studioso². Il complesso donato comprende un arco cronologico che va dal 1943 al 2006 e raccoglie corrispondenza, materiale relativo alla preparazione delle sue pubblicazioni scientifiche e documentazione riguardante l'intensa attività di studio e di ricerca sulla storia contemporanea.

Claudio Pavone, prima della donazione, seleziona e prepara le carte, scegliendo di concentrare nella casa di famiglia a Torchiara, nel Cilento, la documentazione ritenuta di carattere più personale. Il fondo custodito in ACS è in

* Università degli studi di Macerata, Macerata, Italia. g.dimarcantonio@unimc.it.

¹ Da adesso ACS.

² In realtà le buste elencate nell'elenco di versamento sono 166, alcune delle quali collocate all'interno di un unico faldone. Per questa ragione i pezzi indicati al momento della donazione sono 149.

larga misura la testimonianza del profilo di storico dello studioso considerando che l'intensa attività svolta in qualità di funzionario archivista è limitata a pochi fascicoli.

Questi assetti documentari sono confermati anche dal profilo del complesso emerso dal progetto di dottorato di ricerca che chi scrive ha concluso nel novembre del 2021. La struttura si articola infatti nelle serie che seguono:

- Formazione giovanile, 1943-1957;
- Corrispondenza, 1954-2004;
- Attività didattica e scientifica (Convegni, 1959-2007; Pubblicazioni, 1954-2006; Seminari e laboratori, 1961-2006; Studi preparatori, 1944-2005; Università degli studi di Pisa, 1974-1992);
- Amministrazione archivistica, 1975-2007;
- Incarichi, premi e impegno civile, 1954-2009;
- Interviste e programmi televisivi, 1958-2003;
- Miscellanea.

Considerando la struttura e il contenuto del complesso sembra davvero “pacifico” affermare che proprio le carte di Claudio Pavone sono la testimonianza più evidente del deficit di rispecchiamento cui vanno soggetti gli archivi. La tesi centrale del famigerato articolo *Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?* (Pavone 1970), per ragioni che meritano di essere analizzate in altra sede con più attenzione, è confermata dalla natura del versamento che rispecchia solo in parte le attività da lui svolte e obbedisce a un ordine formale frutto di un'attenta selezione e organizzazione da parte del soggetto produttore. L'articolo pubblicato è però la versione riveduta di una prima stesura che Pavone, per nostra fortuna, ha scelto di inserire nel complesso donato all'ACS³ e che fornisce ulteriori spunti di riflessione in merito alle sue teorie. La bozza di cui qui si dà conto è a tutti gli effetti un lavoro preparatorio che sarebbe poi stato raffinato in vista della versione finale, ma proprio questo processo di “revisione ragionata” può risultare molto utile per meglio comprendere e contestualizzare le posizioni di Pavone, poi cristallizzate nella versione pubblicata.

2. E se l'archivio non rispecchia l'istituto?

Nel 1968 Pavone inizia a lavorare ad un articolo destinato a diventare nel tempo un punto fermo della disciplina archivistica della seconda metà del Novecento. Come noto, la rilevanza del contributo, pubblicato due anni più tardi sulla Rassegna degli Archivi di Stato, sta principalmente nel mettere in discussione la teoria del rispecchiamento e di conseguenza la meccanica appli-

³ *Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?*, 1965-1971. Busta 134, fasc. 1-2. Fondo di Claudio Pavone. Archivio centrale dello Stato.

cazione del metodo storico. A distanza di quarant'anni, le riflessioni dello studioso sono ancora attuali e continuano ad alimentare un dibattito importante (Valacchi 2019b).

Claudio Pavone, prima di pubblicare l'articolo, ne inviò una bozza a Filippo Valenti e Isabella Zanni Rosiello, suoi amici e interlocutori privilegiati⁴. Questo modo di procedere era usuale per Pavone che spesso condivideva i suoi lavori preliminari con persone fidate per raccogliere le loro suggestioni o eventuali correzioni. Ne sono altra prova, ad esempio, i molti suggerimenti ricevuti alle bozze del volume *Prima lezione di storia contemporanea* (Pavone 2007) che lo studioso inserisce in un fascicolo esplicitamente intitolato 'Commenti dei colleghi'⁵.

Tornando al nostro articolo, dall'analisi comparata delle due versioni emerge una sostanziale coerenza dei contenuti essenziali. Nella bozza alcuni concetti vengono però espressi in maniera più ampia e articolata. Un'importante differenza, probabilmente non solo formale, si riscontra nel titolo, che ha un tono più forte dell'originale, *E se l'archivio non rispecchia l'istituto?*, tant'è che Pavone, fin dalle prime righe del contributo preparatorio, sente l'esigenza di giustificare questa scelta:

Lo scopo di queste brevi note, stimulate anche da conversazioni con alcuni colleghi, è soltanto quello di riaprire la discussione su alcuni principi correttamente accettati dall'archivistica italiana. Di qui il titolo "provocatorio", che pone il punto interrogativo dopo un canone ritenuto fondamentale: quello appunto della corrispondenza tra archivio e istituto.

Lo studioso poi continua:

È noto che tale corrispondenza è scaturita come corollario del "metodo storico", contrapposto a quel metodo "per materia" che ha assicurato, fra gli archivisti, eterna ignominia al milanese Peroni. Quando infatti ci si è posti la domanda quale sia la storia che il riordinatore di archivi deve rispettare, in quanto iscritta negli archivi stessi, la risposta è stata: la storia dell'istituto che ha prodotto l'archivio; donde poi la tesi della conversione della archivistica speciale nella storia delle istituzioni.

È anche noto, tuttavia, che l'applicazione rigorosa di questo criterio all'opera di riordinamento archivistico e di stesura di inventari ha incontrato e incontra spesso gravi difficoltà. Fra gli scopi delle nostre osservazioni c'è anche questo: non attribuire senz'altro tali difficoltà alla piccolezza degli uomini di fronte alla maestà dei principi, ma cercare di intravedere cosa possa nascondersi sotto le constatabili incongruenze.

⁴ I commenti alla bozza dell'articolo inviata da Claudio Pavone a Filippo Valenti e Isabella Zanni Rosiello sono inseriti all'interno del fascicolo 1 della busta 134 del fondo, già citata.

⁵ *Commenti dei colleghi*, 2001. Busta 1, fasc.3. Fondo di Claudio Pavone. Archivio centrale dello Stato.

Da qui Pavone prosegue la sua trattazione segnalando il problematico rapporto tra introduzione e inventario che si ritrova, anche se in una versione più sintetica, nell'articolo edito:

Cominciamo da un esempio che dovrebbe essere familiare a molti: il rapporto fra introduzione all'inventario e inventario. Nella introduzione l'archivista diligente si sforza di ricostruire la storia dell'istituto: lo fa però il più delle volte soprattutto sulle fonti normative, integrate da nozioni di storia generale. In via esemplificativa, ne risulterà il quadro dell'istituto come avrebbe dovuto funzionare e dell'archivio come avrebbe dovuto essere organizzato per rispecchiare quella ideale vita dell'istituto che lo produceva. Quando tuttavia si viene a riferire sui criteri usati nell'inventario spesso si racconta che le carte sono state in realtà disposte in modi che non corrispondono a quelli del dover essere prima descritto. Le soluzioni che vengono adottate di fronte a questa frattura sono varie e di diverso impegno; ma sotto molte è sottesa l'idea che occorre rassegnarsi a rispettare, con qualche eventuale ritocco, lo stato di fatto, anche se dottrinarmente abnorme. Così si deprecano i guasti prodotti da quegli infelici capri espiatori che sono gli "eruditi settecenteschi", ma si finisce con concludere che, poiché gli studiosi hanno già utilizzato e citato i documenti secondo quel cattivo ordine, non rimane che prenderne malinconicamente atto.

Ci sembra che questa così frequente esperienza ponga almeno tre ordini di problemi.

La sostanza di queste prime righe non segnala particolari scostamenti tra la prima versione e quella poi edita. L'unica grande differenza, con molta probabilità stimolata dai commenti di Flippo Valenti, si riscontra nella parte che segue, ossia nelle considerazioni sulla traduzione del manuale di Adolf Brenneke che non figurano nella bozza (Brenneke 1968). Pavone, infatti, riprendendo alcuni passaggi dell'amico e collega, torna sulla differenza tra le operazioni di registrazione e l'istituto, ossia tra l'attività attraverso la quale il corpus documentario dovrebbe assumere un determinato ordine organico e il reale stato nel quale versa l'archivio dell'istituto stesso. Secondo Brenneke, l'archivista, per effetto di quella che Pavone chiama «funzione creativa», dovrebbe rimodellare la documentazione «secondo gli schemi ideali che nascono dalla sintesi provenienza-contenuto», ossia ricondurre le carte a «quella realtà che gli uomini registratori hanno disposto solo empiricamente» (Pavone 1970, 146). E proprio utilizzando le parole di Valenti, Pavone introduce la questione sul problematico rispecchiamento tra archivio e istituto, richiamando alcuni punti a favore della sua tesi. Questa parte nello studio preparatorio è molto più ampia e dettagliata di quanto non la sarà nella versione finale.

Il primo problema è quello di una migliore definizione di "istituto" e di un più approfondito esame delle relazioni che intercorrono fra il complesso di norme che lo regolano, la prassi amministrativa, i rapporti giuridici che si svolgono nell'ambito di essi e i rapporti sociali nei quali essi cercano la loro forma gi-

uridica. [...]. È chiaro, in realtà, che discorsi di questo tipo non possono che stare a monte del lavoro archivistico: ma per remoti che siano, permettono di portare in seno stesso all'istituto il problema dei vari suoi piani di funzionamento e quindi quello dello scarto fra il suo dover essere e l'essere effettuale. Le conseguenze per l'archivista sono evidenti: a quale livello di vita dell'istituto [...] dobbiamo cercare di rintracciare la corrispondenza fra esso e l'archivio? Seconda esigenza è una migliore definizione di quell' "ordinamento originario" che si afferma essere compito dell'archivista ricostruirlo. Il problema sta quindi nell'individuazione del momento cui riferirsi.

Nella bozza è evidente che la domanda che l'archivista dovrebbe porsi in fase di ordinamento è più netta e in qualche modo ha una connotazione più pratica. Qual è l'ordine originario al quale l'archivista deve ricondurre il fondo? Le vicende della conservazione a che livello devono influenzare il lavoro di ordinamento? Ma soprattutto in che cosa consistono quei rapporti sociali che sostanziano la concretezza della forma giuridica? In una sintesi che da sola ci restituisce l'originalità e la profondità del pensiero di Pavone. Nei passaggi successivi, gli esempi proposti dallo studioso, più minuziosi rispetto alla versione edita, chiariscono meglio tale criticità:

È un problema analogo a quello del restauro, ben noto a storici dell'arte, archeologi e architetti, nessuno dei quali proporrebbe la demolizione dei palazzi di Michelangelo per ricondurre il Campidoglio alla fisionomia "originaria" (vero è che molti degli inconsulti sventramenti fatti in periodo fascista perseguivano proprio mete assurde di tal fatta; né può escludersi che tentazioni germoglianti di analoga radice abbiano talvolta attraversato la mente di qualche archivista). Dati gli scopi pratici che presiedono alla formazione degli archivi, sembra di poter affermare che finché l'archivio è "vivo" l'istituto che lo crea ha pieno diritto di rimaneggiarlo per renderlo più consono ai propri fini. Fra tutte le fasi "storiche" che l'archivio ha attraversato, il riordinatore dovrebbe dunque assumere come modello quella conclusiva perché, dal suo punto di vista, ha ragione chi ha parlato per ultimo. Se ne deduce che l'istituto più immediatamente "rispecchiato" dall'archivio è l'istituto nella fase finale della sua vita, o almeno l'istituto com'era l'ultima volta che ritenne conveniente riorganizzare il proprio archivio.

Gli ordinamenti precedenti – tutti a loro modo "originari" rispetto alla fase a cui si riferiscono – non è che non debbano interessare il riordinatore. Dovrebbe essere anzi sua cura renderne conto criticamente nell'introduzione che in tal modo sarebbe aiutata a perdere quel ruolo di "prologo in cielo" che spesso le viene riservato.

Pavone, quindi, individua chiaramente quale possa essere la soluzione nei casi in cui il riordinatore si trovi a dover trattare fondi che presentano diversi e successivi rimaneggiamenti operati dall'istituto nel corso della sua attività o dovuti ai diversi processi conservativi. «Finché l'archivio è "vivo" l'istituto che lo crea ha pieno diritto di rimaneggiarlo per renderlo più consono ai propri

fini» e «ha ragione chi ha parlato per ultimo». E tra queste due asserzioni che si gioca la partita di una riorganizzazione consapevole e fedele ai ritmi reali dei complessi documentari. L'archivista – secondo questa chiave di lettura – dovrebbe assumere l'ultima organizzazione formale come parametro di riordinamento, utilizzando l'introduzione non più come “prologo in cielo” ma quale strumento atto a specificare le scelte operate sul fondo e, soprattutto, le ragioni che le hanno determinate.

L'analisi prosegue poi richiamando le coordinate che dovrebbero orientare l'ordinatore rispetto ad eventuali «errori archivistici» che, non di rado, si possono rintracciare nella sedimentazione delle carte:

Emerge infine il problema dell'“errore” archivistico consumato nel passato. Ci sembra che – una volta chiarito il corretto rapporto intercorrente fra norme, quando esistono, e prassi archivistica – per quelli che sarebbe conveniente continuare a definire “errori” il criterio dovrebbe essere questo: rispettare gli errori significativi, correggere gli altri, dando naturalmente conto di entrambi.

Ma quali errori devono essere considerati “significativi”? È necessario, anche in questo caso, aver presente un sistema di riferimento. Ad esempio, per chi volesse sottoporre a trattamento psicanalitico i nostri colleghi del passato prossimo e remoto potrebbero essere riccamente significativi errori che, in un altro contesto, non presenterebbero alcun interesse.

È questo, dunque, è un caso in cui il metodo storico e la cultura storica dell'archivista, devono davvero entrare in campo. Se Crispi, ad esempio, ha ecceduto nella pratica, corrente fra gli uomini di governo, di portare a casa atti d'ufficio, questo non interessa per denunciare che egli violò, più di altri suoi colleghi, i principi della dottrina archivistica, ma per portare una nuova, e sia pur secondaria, tessera alla ricostruzione della personalità autoritaria e accentratrice di quello statista. E se gli archivi di gabinetto si sono venuti gonfiando, nei ministeri, di documenti sempre più numerosi, questa non è, da un punto di vista storico, violazione del celebre regio decreto del 1900, ma il risultato di un evolversi dell'organismo stesso ministeriale.

Il concetto di “errore significativo”, da rintracciare anche nelle pieghe psicanalitiche della descrizione e dei descrittori, segna un altro punto fermo nel definire il profondo rispetto che Pavone ha per l'archivio in sé, indipendentemente dal mito aurorale delle origini. Anche gli errori, quando non siano meramente materiali, fanno parte dell'archivio. Sono infatti frutto di un sentimento, se non di un preciso bisogno operativo, che a prima vista può sfuggire e che ha invece il suo significato informativo.

Si arriva così all'ultimo punto della trattazione di Claudio Pavone ossia il problema della periodizzazione. Nella versione edita le questioni legate alle cesure cronologiche da adottare in fase di ordinamento vengono solo accenna-

te⁶. Nello studio preparatorio, invece, questa riflessione è più estesa e presenta alcuni spunti degni di nota:

Quale periodizzazione occorre adottare nell'ordinamento degli archivi? La domanda è strettamente connessa al problema del rapporto archivio-istituto e può spezzarsi in queste altre: le periodizzazioni della storia degli istituti pubblici coincidono sempre con quelle degli ordinamenti sovrani (quando individuabili) di cui essi fanno parte? Queste ultime coincidono a loro volta con le periodizzazioni usate nella storia generale? E quanto, agli archivi, vanno essi agganciati a una delle periodizzazioni sopra accennate, o bisogna andare in cerca di una loro periodizzazione autonoma?

Per avviare un tentativo di risposta giova ricordare che le periodizzazioni convenienti ai singoli settori della ricerca storiografica possono non coincidere. Basterà fare a riguardo l'esempio classico del Rinascimento: periodizzazione sufficientemente chiara nella storia della cultura sul cui terreno è narrata, ma che si è rivelata inadatta nel campo della storia economica. In linea di principio, dunque, non esistono pregiudiziali contrarie a una eventuale periodizzazione degli archivi diversa da altre periodizzazioni in uso nelle discipline storiografiche. Si tratta soltanto di esaminare se e quando essa possa essere scientificamente produttiva.

Come si accennava in precedenza, la bozza del contributo venne inviata da Pavone ad alcuni colleghi e quasi tutti espressero orientamenti diversi dai suoi sulla questione della periodizzazione. In sostanza, mentre lo studioso propendeva per mantenere anche nell'ordinamento del fondo una periodizzazione storica canonica, fatte salve alcune eccezioni particolari alle quali si accennerà in seguito, i colleghi concordavano sul fatto che l'inventario dovesse restituire le cesure realmente esistenti nel fondo. Pavone però non la pensava così, convinto che per gli studiosi fosse più agevole muoversi nel fondo alla luce di una periodizzazione canonica. In risposta ad una lettera inviata da Isabella Zanni Rosiello, il 9 giugno del 1966⁷, lo studioso torna sulla questione affermando:

Interessante il rapporto periodizzazione-generalizzazione. Certo andrebbe approfondito, per indagare fino in fondo se si tratti davvero, sotto ogni profilo, della stessa operazione logica. Quanto poi all'unicità o irripetibilità del singolo evento, cronologicamente ben individuato, mi sembra [...] che il problema non venga esaurito da quella della periodizzazione o generalizzazione. Queste sono operazioni di sistemazione scientifica e quindi convenzionale [...] della realtà. Ma l'esigenza di tale operazione nasce proprio dalla presenza di tante

⁶ Su questi temi, al centro del dibattito progettuale per la costruzione della *Guida generale degli Archivi di Stato*, Claudio Pavone, insieme a Piero D'Angiolini, tornerà diffusamente proprio nell'introduzione all'imponente opera, in particolare si veda: Claudio Pavone e Piero D'Angiolini, «La Guida generale degli Archivi di Stato italiani: un'esperienza in corso», *Rassegna degli Archivi di Stato* XXXII, n. 2 (1972): 13-14.

⁷ *Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?*, 1965-1971. Busta 134, fasc. 2. Fondo di Claudio Pavone. Archivio centrale dello Stato.

individualità. Anche il singolo gatto è irripetibile [...]. Altro problema interessante è quello della non coincidenza delle diverse periodizzazioni. Me ne sono dovuto occupare proprio in rapporto agli archivi; [...] Cioè: la periodizzazione degli istituti pubblici può non coincidere con quella degli ordinamenti sovrani di cui fanno parte? E quella degli archivi può non coincidere né con l'una né con l'altra? A me ripugna sempre concedere troppa autonomia all'archivistica. Gli archivi mi interessano assai di più come fonti che come fenomeno meritevole di specifica attenzione. [...] Ritengo che in linea di massima la periodizzazione da seguire per gli archivi sia quella delle grandi partizioni che la storiografia politica ha trattato da cambiamenti di regimi, ordinamenti sovrani, ecc. E questo non perché non mi renda conto del carattere problematico e convenzionale di tale periodizzazione; ma perché ritengo che, dovendosi anche per gli archivi adottare una periodizzazione e dando una pretesa autonoma alle loro periodizzazioni, i risultati per gli utenti degli archivi stessi risulterebbero poco comprensibili[...].

Sulla periodizzazione, quindi, la posizione di Claudio Pavone pare abbastanza chiara e, nella bozza, per avvalorare la sua tesi, lo studioso si affida alla sua esperienza di archivista sul campo:

L'esperienza dei lavori archivistici pone in evidenza il problema soprattutto in due casi:

- a. Nell'inventario di un singolo archivio, quando questo trapassa dall'uno all'altro ordinamento sovrano e dall'uno all'altro periodo storico (secondo le partizioni in uso nella storia generale);
- b. Nella guida di un intero archivio di Stato (o di altro archivio generale), quando si tratta di scegliere il criterio della successione e dell'articolazione dei vari fondi.

Si possono al riguardo individuare alcuni nodi di discussione, raggruppabili attorno a queste tre parole: continuità, vischiosità, frattura. Il problema cioè è quello di definire cosa occorre guardare per dire che un archivio è concluso e ne comincia un altro.

Continuità, vischiosità, frattura. In linea con le riflessioni che attraversano anche l'ambizioso e per certi versi rivoluzionario progetto della Guida Generale (D'Angiolini et al. 1981), con questi tre concetti, Claudio Pavone, definisce efficacemente i parametri con cui la disciplina archivistica deve confrontarsi in cerca di soluzioni che permettano di governare la complessità documentaria. La memoria archivistica, se è lecito definirla così, è per sua natura annidata e per certi versi circonvoluta. Il tempo degli archivi non è necessariamente lineare, non si muove lungo l'asse di un'astratta evoluzione. Gli archivi sanno resistere tenacemente al cambiamento istituzionale e politico, le carte possono intrecciarsi tra loro senza soluzione di continuità e le fratture periodizzanti sono spesso modelli astratti. Quando si specchia nella storia istituzionale, l'archivio non si riconosce perché in fondo è un manufatto che obbedisce a regole molto più basiche e concrete dei principi del diritto. L'archivio, per così dire,

si fa e non si pensa a tavolino, almeno fino a un certo momento. Per inciso, il fenomeno della vischiosità documentaria in rapporto alla storia delle istituzioni è da sempre stato oggetto di attenzione nei lavori di Claudio Pavone, tant'è che per il primo centenario dell'Unità pubblicherà un'opera in tre volumi sugli Archivi dei governi straordinari e provvisori tra il 1859 e il 1861, i cui esiti sono richiamati successivamente anche nella bozza (Pavone 1961). Ma tornando alle riflessioni esposte dallo studioso, quindi, quando è che un fondo può dirsi concluso? Come e quanto influiscono le vicissitudini storiche, politiche e istituzionali nell'organizzazione della documentazione? A quali elementi il riordinatore deve appoggiarsi per fornire una risposta a queste problematiche? A tal riguardo Pavone tenta di fornire delle soluzioni:

Una prima risposta porta a dar rilievo decisivo alla contiguità materiale delle carte, collocate una dietro l'altra secondo gli stessi moduli – razionali o empirici – usati in precedenza, quale che sia l'evento storico-politico o storico-istituzionale nel frattempo intervenuto.

L'esperienza insegna che si hanno nella realtà casi di tal genere prolungatisi per un numero di anni più o meno lunghi. Di fronte ad essi non si può tuttavia, a nostro avviso, assumere un atteggiamento univoco, ma occorre guardare da vicino il significato che ciascuno di essi volta volta ha, non dimenticando mai il principio secondo cui, in linea generale, l'amministrazione è vischiosa rispetto alla politica e gli archivi sono vischiosi rispetto all'amministrazione. Entro questo ambito si possono -nel caso in esame -formulare alcune delle seguenti ipotesi:

tutto è dovuto a mera pigrizia e neghittosità degli impiegati addetti agli archivi; il nuovo regime o l'istituto successore sono inefficienti e stentano a organizzarsi secondo propri schemi originali;

il nuovo regime o il nuovo istituto attribuiscono scarsa importanza a documenti ed archivi, e abbandonano deliberatamente questo settore ad impiegati in sottordine:

il nuovo regime o il nuovo istituto pur attribuendo importanza agli archivi, accettano deliberatamente di continuare a gestirli secondo le vecchie regole; lo sconvasso è stato così profondo che non sono rimasti tempo e fiato per pensare, almeno in un primo momento, agli archivi.

Sono tutte ipotesi plausibili, fra le quali il riordinatore è chiamato a compiere una scelta motivata e a comportarsi di conseguenza, in un arco interpretativo che va da un massimo di frattura a un massimo di continuità.

Una seconda risposta più che alla materiale disposizione degli atti guarda alla continuità della funzione esercitata dal vecchio e dal nuovo istituto. Una visione di tal tipo è ad esempio sottintesa da coloro che, per indicare due dei più tipici fondi archivistici meridionali, scrivono "intendenza, poi prefettura" e "giudicate regio, poi pretura". Ma se la risposta esaminata nel punto precedente, non criticamente rielaborata, rischiava di ridurre la "storia" che devono rispettare i riordinatori di archivi a quella vissuta dai nostri più umili colleghi del passato, questa seconda contiene in sé il pericolo di assumere come criterio

storico un'astrazione giuridica (al limite, quella della divisione dei poteri), con in più la possibilità di ricadere, con solo apparente paradosso, in un criterio per materia. Ad esempio: se la continuità è data dalla funzione, a prescindere dall'organo in cui quella si incarna, non sono da condannare quelle raccolte di "atti demaniali", che si trovano presso alcuni archivi di Stato meridionali, e che raccolgono tutti gli atti relativi alla liquidazione degli usi civici e dei demani comunali, a partire da Gioacchino Murat fino ai tuttora operanti commissariati per la liquidazione degli usi civici.

Il discorso critico sulle risposte richiamate nel punto precedente ci conduce infine ad accennare alla terza possibile risposta: quella secondo cui occorre guardare soprattutto il succedersi degli ordinamenti statuali sovrani. Diciamo subito che è la risposta per la quale, in prima approssimazione, maggiormente propendiamo, sia perché per gli archivi pubblici il fatto "stato" ha, in ultima analisi, un valore predominante, sia perché la chiave di ricerca imperniata su di esso risulta nella maggior parte dei casi la più limpida per gli studiosi, e non soltanto per quelli di storia politica.

Claudio Pavone, quindi, propone delle possibili soluzioni, affidandosi in definitiva alla sensibilità e alla soggettiva competenza del riordinatore. Quest'ultimo, infatti, dovrà tener conto delle vicissitudini conservative che possono avere influenzato l'ordinamento e quindi la periodizzazione di quell'archivio. In altre parole, dovrà saper riconoscere la vischiosità delle prassi influenzate da pigrizia, inefficienze o disattenzioni. E, allo stesso tempo, il riordinatore dovrà tener conto della continuità delle funzioni che i nuovi e vecchi istituti hanno assunto tenendo a mente che il fatto "Stato" prevale nell'ordinamento perché l'organizzazione dell'informazione che deriva da una struttura nota è più intellegibile per gli studiosi.

Ma, andando oltre a queste condivisibili soluzioni, Pavone, che all'epoca aveva già accumulato molta esperienza negli archivi, sente l'esigenza di puntualizzare alcuni aspetti:

Occorrono tuttavia alcune precisazioni.

La prima è la più ovvia, è che lo Stato, inteso come sovranità, non è per nulla un fatto sempre pacificamente constatabile, specie in una storia complessa, spezzettata e pluralistica come quella italiana. Così il grado di autonomia del singolo istituto rispetto all'ordinamento statale non è affatto uguale per tutti gli istituti, tempi e Stati, e nel valutare questa differenza bisogna far ricorso a una distinzione della quale finora abbiamo prescinduto, avendo chiamato "istituto" sia quello fornito di personalità giuridica sia quello che ne è privo. Un comune – è appena il caso di ricordarlo – si è sempre collocato nei confronti dello Stato che lo ha inglobato in una posizione diversa da quella degli organi proprio dello Stato stesso. Perciò un archivio comunale ha diritto di vedersi riconosciuto un coefficiente di continuità superiore a quello di un archivio amministrativo statale, sebbene anche per i comuni le cose non siano andate sempre nello stesso modo. Infatti se è lecito partire, ad esempio, dall'ipotesi

che il passaggio dalla dominazione spagnola a quella austriaca interferisca poco nella continuità dell'istituto-comune, discorso diverso dovrebbe farsi quando, da Napoleone in poi, i comuni vengono rigidamente inquadrati in una struttura amministrativa uniforme. In altre parole, non sarebbe, a nostro avviso, un'eresia porre in rilievo, in un inventario di un archivio comunale, la frattura data dall'inquadramento del comune nel regno italico. Analoga frattura, invece, non sarebbe probabilmente giustificata nell'archivio, ad esempio, di una Università degli studi come quella di Bologna.

Come la continuità della funzione non legittima la deduzione della continuità dell'archivio così tanto meno legittima la continuità della funzione sovrana propria dello Stato. Così, quando un ordinamento statale crolla, come avvenne per i vecchi Stati italiani nel biennio 1859-1860 o per l'impero austro-ungarico nel 1918, parrebbe strano che negli ordinamenti e negli inventari degli archivi dei suoi organi il fatto passasse inosservato. Posto ad esempio che a Trieste gli uffici dell'amministrazione austro-ungarica siano sopravvissuti nei nomi e nell'organizzazione interna fino al 1924, quando iniziarono a funzionare regolarmente i nuovi organi periferici italiani, non ci sembrerebbe giusto tirarne la conseguenza che la data del 4 novembre 1918 non abbia avuto rilevanza istituzionale e archivistica.

Un altro esempio: quando si dovette procedere all'inventario degli archivi dei governi provvisori che ressero le varie parti d'Italia fra il 1859 e il 1861, ci si trovò, in Toscana, di fronte soprattutto alle "code" degli archivi dei vari organi centrali granducali. Si decise di considerare l'insieme di queste "code" come l'archivio del Governo provvisorio toscano, articolato secondo i suoi organi centrali. Questa soluzione – benché realizzata solo sulla carta, senza cioè materiale spostamento di documenti – è stata da qualcuno criticata come violentatrice del fatto archivistico in favore di una periodizzazione storica generale. L'archivio del ministero dell'interno, si è ad esempio detto, è lo stesso prima e dopo il 27 aprile 1859, e quindi non era lecito spezzarlo in conseguenza di un fatto non archivistico, come la fuga di un granduca. Questa tesi non disdegnava nemmeno di estremizzarsi nell'affermazione che non solo l'archivio, ma il ministero in quanto tale era rimasto lo stesso.

Anche in questo caso ci sembra tuttavia che non si possano formulare giudizi di continuità o di rottura senza un preciso sistema di riferimento. Non c'è dubbio che il ministero era "lo stesso" se si guarda al fatto che esercitava una funzione – l'amministrazione interna – uguale o almeno analoga, a quella svolta sotto il granduca, ed era anche lo stesso se si considera l'identità fisica della maggioranza degli impiegati che lo facevano funzionare, cominciando dagli archivisti che continuarono ad applicare le medesime regole. Ma il ministero non era più lo stesso se si dà rilievo al fatto, di importanza storica essenziale, che non era più organo del governo granducale, bensì di un governo provvisorio che stava addirittura liquidando lo Stato toscano; e che la mutata situazione politica non poteva non ripercuotersi nel contenuto degli affari trattati, cosicché di fatto il ministero svolgeva un'azione in vario grado diversa. Cosa doveva fare allora il riordinatore o inventariatore? Scegliere fra gli opposti possibili modi di vedere

le cose. E la scelta a favore della novità contro la continuità apparve legittima perché è basata sugli elementi che sembrò giusto porre in preminente rilievo, e come l'unica, per di più, che non disviava verso una risposta falsamente negativa lo studioso che avesse posto la domanda, per lui primordiale, sull'esistenza di un archivio del governo provvisorio toscano.

Lo studioso, nel passaggio appena riportato, fa riferimento al già citato e magistrale lavoro che avrebbe poi preso corpo nei volumi pubblicati nel 1961 (Pavone 1961, 2004b) e attinge ad alcune riflessioni precedentemente espresse nel 1964, dopo la pubblicazione del DPR del 30 settembre del 1963 n. 1409 (Pavone 1964). Nei passaggi successivi, Pavone torna sul concetto nevralgico del suo contributo, ossia sulla questione del metodo storico, mettendo in guardia gli studiosi dai possibili pericoli di quella che lui chiama "fantarchivistica". Con tale espressione, Pavone intende sintetizzare l'applicazione dogmatica del metodo che stenta spesso a restituire quelle vischiosità, continuità e fratture delle quali si accennava in precedenza:

Ai fini di tutto il discorso che stiamo conducendo, ci sembra utile ricordare che una cosa è la "vita dell'istituto" in quanto tale, considerata ai molteplici livelli sopra richiamati, altra cosa è l'attività dell'istituto in quanto volta ad organizzare la propria capacità di autodocumentarsi. È questa una distinzione che potrebbe probabilmente rivelarsi assai feconda; ed è una distinzione che è venuta acquistando un rilievo sempre maggiore, e che ancora di più ne acquisterà nel prevedibile futuro. Riusciva relativamente facile al notaio-cancelliere di un comune medievale conservare gli atti via via prodotti secondo schemi abbastanza vicini al concreto modo di funzionare dell'istituto comunale. L'empiria stessa del metodo era in quel caso garanzia di aderenza alla prassi giuridica e amministrativa. La mole crescente degli affari e delle carte è da ritenere sia stato il primo elemento che ha posto in crisi tale seducente correlazione, che è poi quella sulla cui esperienza nacque il "metodo storico" in archivistica. L'introduzione del titolario ad opera di Napoleone spezzò programmaticamente questo rapporto ingenuo ed aurorale fra archivio ed istituto. Il titolario non è infatti che il metodo razionalistico -tanto deprecato dagli idealisti- applicato agli archivi in fieri. Questa razionalità, il cui criterio risiede non tanto nella mera conservazione degli atti quanto nel rendere facile e pronto il reperimento di un singolo atto in mezzo alla mole sempre crescente di tutti gli altri, si è sviluppata secondo una sua linea, collegata sempre meno direttamente alla vita dell'istituto considerato in tutte le altre sue dimensioni. Facciamo un esempio a noi molto vicino: il titolario in uso presso la Direzione Generale degli Archivi di Stato, che si ferma al livello divisionale, rispecchia ben poco il reale modo di funzionare di tale organo; e lo studioso del futuro che volesse prenderlo a base per tracciare la storia della Direzione verrebbe condotto notevolmente fuori strada, anche volendo ammettere - in via molto ipotetica - che quel titolario sia fedelmente applicato.

Ma, a prescindere dai titolari e dalle discussioni in corso sul loro necessario rinnovamento, oggi appare evidente che la tecnica della documentazione si è

venuta costituendo in disciplina sempre più autonoma, aperta all'acquisizione di strumenti lontani da quelli che un tempo erano manovrati dagli autori o dai destinatari stessi dei documenti.

Possiamo fermare qui le nostre osservazioni, non senza tuttavia aver messo in guardia dai pericoli della fantarchivistica, sempre in agguato quando si parla di macchine applicate agli archivi, e non senza aver ricordato che compito dell'archivista di Stato rimarrà pur sempre, a nostro avviso, quello di un ausiliario della ricerca storiografica, che arriva a cose, cioè ad archivi, fatti, e traduce in termini istituzionali e documentari la domanda postagli dallo storico.

A margine di queste riflessioni si potrà osservare oggi come la "fantarchivistica" di Pavone sia alla fine sbarcata sulla Terra. Si potrebbe cioè riflettere sul valore relativo delle periodizzazioni che servono a incasellare l'informazione ma non hanno, né potrebbero avere, scansioni rigorose e immutabili e soprattutto non danno conto dell'articolazione archivistica reale. A ben guardare nell'individuazione della fluidità utile di queste cesure e nel bisogno di datare le diverse relazioni in ossequio a periodizzazioni appropriate emergono i prodromi di un antidoto alla vischiosità e, in prospettiva, logiche proprie dei sistemi informativi archivistici e del loro bisogno di datare le relazioni.

L'ultimo punto analizzato, prima di avviarsi alla conclusione di questo studio preparatorio, parte ancora da una domanda, ossia «[...]cos'è propriamente significante in un archivio, l'ordine reciproco dei pezzi che lo compongono o il loro contenuto documentario?». Il riferimento, esplicitato nella trattazione, è al concetto di *universitars rerum* di Giorgio Cencetti (Cencetti 1937).

Sotto quanto detto finora serpeggia un ultimo interrogativo che vorremmo cercare di rendere esplicito: cosa è propriamente significante in un archivio, l'ordine reciproco dei pezzi che lo compongono o il loro contenuto documentario? È un modo questo per riproporre il problema dell'autonomia dell'archivio, vista ora come possibilità di costruire un rigoroso e formalizzato sistema (o struttura) di segni, a prescindere dal loro contenuto semantico. (E sarebbe forse anche la strada per rivestire di qualche insperata dignità una vecchia massima, ripetuta anche nella relazione ministeriale a stampa del 1952, e che ha sempre urtato le persone di buon senso: la massima cioè che solo colui che non legge le carte che ordina è perfetto archivista).

La nota tesi, formulata dal Cencetti, dell'archivio come *universitas* è solo parzialmente analoga a quella sopra prospettata. La teoria dell'*universitas* pone sì l'accento sul vincolo che lega un documento all'altro, ma questo vincolo vede soprattutto sotto un profilo giuridico e, ci sembra, non fa completa astrazione delle finalità pratiche, extrarchivistiche, dei produttori d'archivi, non considera del tutto indifferenti le cose testimonianti. Si tratterebbe ora invece di vedere se sia possibile costruire una teoria dell'archivio sulla base di una totale non presa in considerazione della materia documentata.

La nostra risposta è negativa per un duplice ordine di considerazioni. Innanzitutto, non crediamo che l'archivistica possieda un sufficiente grado di au-

tonomia metodologica. L'archivistica è circoscritta nel campo tecnico, e non può quindi costruirsi in scienza iuxta propria principia. Può darsi che si possa ricondurla sotto la teoria dell'informazione ma si tratterebbe, per l'appunto, di un limitato campo applicativo di una teoria di larghissimo significato.

In secondo luogo, anche se si volesse ammettere che operando miracoli di formalizzazione, l'archivistica possa in futuro essere presentata come scienza autonoma (e l'archivio quindi come un sistema chiuso), ci sembra che un simile risultato sarebbe di ben scarso interesse, almeno per un archivista di Stato. Si tratterebbe infatti di una costruzione così flebile ed esangue che, appena la si cominciasse ad usare, se ne sfonderebbe la forma per riprendere contatto con il contenuto ed essere rinviiati ad altri tipi di discorso. Solo così l'archivio tornerebbe a giovare alla storia, che poi l'unica finalità che tutti diciamo di attribuire agli archivi quand'essi abbiano perduto quell'amministrativa. Siamo del resto convinti che anche il più formalista è il più puro degli archivisti di Stato quando dice che è un archivio è "interessante" non può fare a meno di pensare anche al contenuto delle carte.

Che oggi si possa parlare di archivistica come disciplina autonoma sembra assodato, non solo considerati gli sviluppi metodologici e applicativi registrati dagli anni Settanta ad oggi, ma anche perché la disciplina ha assunto una propria connotazione dentro a un dibattito scientifico ampio e articolato. Il "miracolo di formalizzazione" c'è stato e si è rivelato non tanto negli standard di descrizione quanto nel dibattito che da lì si è sviluppato. La formalizzazione, poi, ha messo a disposizione anche strumenti tecnologici duttili e potenti in misura adeguata a governare, in buona parte, la contraddizione tra forma e sostanza che attraversa l'"archivistica debole" di Pavone. Fuori dai vincoli della bidimensionalità descrittiva molte questioni trovano infatti soluzioni possibili nella capacità comparativa che noi oggi abbiamo per tenere a bada le riluttanze al metodo di ogni archivio. Detto questo, però, lo stesso Pavone sembra rifuggire da un'archivistica attenta solo ai meccanismi teorici della produzione e della sedimentazione delle carte e non al contenuto. In questo caso, è necessario sforzarsi di leggere queste sue ultime battute collocandole nel contesto a lui coevo. Da lì a poco sarebbe stato istituito il Ministero per i beni culturali e ambientali e gli archivi, simbolo del potere statale, sarebbero stati ancora più inglobati nel dominio del patrimonio culturale (Pavone 1975). L'archivistica, quindi, da disciplina volta a governare la memoria del potere, si connoterà diversamente come memoria della memoria. Forse, grazie anche a questo cambio di passo, l'archivistica ha assunto negli anni delle sfumature più ampie ed è a tutt'oggi indubbia la sua autonomia scientifica.

È però interessante notare come Pavone ponga la questione dell'avalutatività, questa invece ancora non propriamente risolta e anzi rilanciata dalle selezioni digitali e dalla nascita dei cosiddetti archivi inventati. Si torna infatti sulla percezione dell'archivio e sulle responsabilità del riordinatore che deve farsi mediatore tra quello che l'archivio "è", e quello che l'archivio "dovrebbe esse-

re". In altre parole, l'archivista dovrebbe da un lato astenersi da personali considerazioni sul fatto archivio, e dall'altro governarlo. Per far ciò deve scegliere, per certi versi arbitrariamente, in che modo restituire all'utente l'ordinamento affinché abbia una percezione del fondo quanto più reale e realistica possibile. Gli standard di descrizione archivistica, discussi e diffusi dagli anni Novanta, hanno sicuramente aiutato i riordinatori a non "perdersi nelle carte" utilizzando un approccio più omogeneo e armonico. Ma, almeno in prima battuta, sul piano concreto i modelli di normalizzazione si fermano a una lettura meccanica, alla formalizzazione di una struttura e soprattutto delle relazioni che intercorrono tra le diverse entità informative lasciando però aperto il problema della individuazione delle entità stesse e delle relazioni (Vitali 2014; Valacchi 2016). Sono comunque dirimenti e imprescindibili nell'impostazione del lavoro di ordinamento e nella progettazione di strumenti che Pavone non poteva immaginare per quanto in certi suoi passaggi sembri preconizzarli, perché aiutano a fare chiarezza, almeno sul piano formale. I modelli proposti dall'*International Council on Archives*, infatti, dettano, e non è poco, delle linee di indirizzo che consentono comunque all'archivista di scegliere le soluzioni più idonee ed efficaci al fine di rendere intellegibile il fondo ai futuri utenti. E proprio pensando a questi ultimi, Pavone conclude:

Quale conclusione trarre per ora da un discorso che ci auguriamo venga da altri ripreso e approfondito? Soprattutto queste: che la rappresentatività dell'archivio rispetto all'istituto non è un dogma, ma un problema o se si preferisce un'ipotesi di lavoro. Non esistono parallelismi metafisici e armonie prestabilite; e nemmeno esiste la possibilità di evasione caricando l'archivio di un valore pieno e autonomo che esso non può avere. Esistono soltanto situazioni da vagliare criticamente. Solo così il rapporto archivio-istituto diventa davvero "storico"; solo così la sua ricostruzione e chiarificazione impegna la cultura dell'archivista, costringendolo a fare appello a ogni informazione che egli possa trarre, da qualsiasi fonte, sui diversi livelli di funzionamento dell'istituto e a confrontarla con lo scheletro archivistico, dando un senso anche alle lacune e alle discordanze. Soltanto così, infine, l'inventario, redatto secondo chiavi lucide e rigorosamente applicate, può svolgere quella che è la sua funzione specifica: giovare ai ricercatori.

In ultima analisi è proprio questo i *leitmotiv* che ha guidato la complessa stesura dell'inventario delle carte di Claudio Pavone, durante la quale, chi scrive, ha cercato di tenere sempre presenti le possibili esigenze dei ricercatori o, per meglio dire, in chiave più contemporanea, degli utenti.

3. Conclusioni

Si potrebbe accettare la ripugnanza di Pavone per l'eccessiva autonomia dell'archivistica e chiudere senza bisogno di altre parole. Sembra invece che

proprio l'analisi comparativa che si è sviluppata dimostri una volta di più il bisogno di quel pensiero archivistico che lo studioso, quasi fosse un vezzo, colloca sullo sfondo, nel momento stesso in cui lo rigenera e lo prepara alla modernità.

In conclusione, allora, le riflessioni di Claudio Pavone restano di sicuro interesse anche nella dimensione archivistica contemporanea che anzi in alcuni passaggi sembrano quasi saper immaginare. Le criticità che lo studioso rileva nell'applicazione del metodo storico e la fatica, se non l'impossibilità, di cogliere un reale rispecchiamento tra archivio e soggetto produttore si ripropongono del resto anche nei complessi documentari ibridi o nativi digitali contemporanei (Valacchi 2019a). Solo per fare un esempio, che meriterebbe ben altro approfondimento, l'idea stessa di interoperabilità, intesa come capacità di diversi soggetti di partecipare alla definizione di un progetto o di una pratica, può mettere in forte difficoltà l'idea "monolitica" che abbiamo a lungo avuto di soggetto produttore. Sempre più spesso, nell'ordine naturale delle cose, gli archivi sono il risultato piuttosto fluido di vere e proprie cooperative di produttori. Lo specchio, insomma, anche se esistesse, andrebbe comunque in frantumi. Se nel contesto analogico la disciplina ha assorbito, ed accettato, le indicazioni di Pavone, nella dimensione digitale la riflessione è ancora aperta e decisamente complicata. Quali criteri e strategie di restituzione dovranno adottare gli archivisti di un futuro ormai prossimo? Che ne sarà del rapporto univoco e fin qui dato per scontato tra produttore e produzione? I dati e i metadati sopravvissuti all'obsolescenza con quale forza riusciranno a rispecchiare le attività svolte dai soggetti che li hanno prodotti? Avrà ancora senso parlare di ordinamento e inventariazione? Anche in questa direzione Pavone scrisse: «Quanto più sottile e sofisticata si farà la registrazione delle informazioni necessarie alla vita di un organismo tanto più l'inventario del futuro non potrà non prendere atto anche di questi mutamenti» (Pavone 2004a, 93). Gli archivi digitali richiederanno di ripensare o di studiare metodi e modalità di restituzione delle informazioni diverse da quelle attuali ed è per questa ragione che assume un senso rileggere quanto Pavone ed altri hanno scritto. Perché anche se le problematiche da risolvere potrebbero essere differenti da quelle del passato, le domande da porsi e le modalità per arrivare a soluzioni sostenibili ed efficaci dovranno seguire quel rigore scientifico e metodologico che Pavone ed altri hanno perseguito. In un mondo dove la rincorsa digitale sembra orientare qualsiasi processo ma si fa ancora fatica ad avere una coscienza critica sui bit che troppo meccanicamente vengono prodotti, la frammentazione gestionale e conservativa della memoria del futuro è dietro l'angolo: *e se l'archivio (digitale) non rispecchi(erà) l'istituto?* La partita, quindi, è ancora aperta e la consapevole curiosità di Pavone, insieme al rigore di un metodo che sia confronto concreto e non astratta speculazione, restano strumenti indispensabili per giocarla. Magari con la consapevolezza che, proprio come ha tentato di insegnare Pavone,

questa partita non finisce mai perché le dinamiche di generazione e gestione degli archivi sono inesauribili e in continua evoluzione.

Riferimenti bibliografici

- Brenneke, Adolf. 1968. *Archivistica: contributo alla teoria ed alla storia archivistica europea. Traduzione italiana di Renato Perrela*. Milano: Giuffrè.
- Cencetti, Giorgio. 1937. "Sull'archivio come 'Universitas rerum'." *Archivi* IV: 7-13.
- D'Angiolini, Piero, Claudio Pavone, Paola Carucci, Antonio Dentoni Litta, e Vilma Piccioni Sparvoli. 1981. *Guida generale degli archivi di Stato italiani*. Roma: Ministero per i beni culturali e ambientali, Ufficio centrale per i beni archivistici.
- Pavone, Claudio. 1961. *Gli archivi dei governi provvisori e straordinari (1859-1861)*. Roma: Pubblicazioni degli archivi di Stato.
- Pavone, Claudio. 1964. "Archivi fatti e archivi in fieri." *Rassegna degli Archivi di Stato* XXIV, no. 3: 359-60.
- Pavone, Claudio. 1970. "Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?" *Rassegna degli Archivi di Stato* XXX, no. 1: 145-49.
- Pavone, Claudio. 1975. «Gli archivi nel lungo e contraddittorio cammino della riforma dei beni culturali». *Rassegna degli Archivi di Stato* xxxvi, no. 1:2-3: 143-60.
- Pavone, Claudio. 2004a. "Inventariazione e problemi di metodo." In *Intorno agli archivi e alle istituzioni. Scritti di Claudio Pavone*, a cura di Isabella Zanni Rosiello, 89-94. Roma: Ministero per i Beni e le Attività culturali, Dipartimento per i beni archivistici e librari, Direzione generale per gli archivi.
- Pavone, Claudio. 2004b. "Mutamenti istituzionali e storie di carte." In *Intorno agli archivi e alle istituzioni. Scritti di Claudio Pavone*, a cura di Isabella Zanni Rosiello, 43-70. Roma: Ministero per i Beni e le Attività culturali, Dipartimento per i beni archivistici e librari, Direzione generale per gli archivi.
- Pavone, Claudio. 2007. *Prima lezione di storia contemporanea*. Bari: LaTerza.
- Valacchi, Federico. 2016. "Pezzi di cose di cose nel mondo. Il processo di integrazione delle descrizioni archivistiche nei sistemi interculturali." *JLIS.it* 7, no. 2: 331-67. <https://doi.org/10.4403/jlis.it-11529>.
- Valacchi, Federico. 2019a. "Quiddam divinum. Riflessioni sul metodo storico." *Archivi* XV, no. 1: 69-87.

- Valacchi, Federico. 2019b. "Ripartire da Pavone. Spunti di archivistica." *Parole Chiave. Claudio Pavone* 1, no. 2: 161-76.
- Vitali, Stefano. 2014. "La descrizione degli archivi nell'epoca degli standard e dei sistemi informatici." In *Archivistica. Teorie, metodi, pratiche*, a cura di Linda Giuva e Maria Guercio, 179-210. Roma: Carocci.

Strumenti per il reperimento di informazioni rilevanti in Rete

Applicazioni nel dominio della comunicazione pubblica in
materia di immigrazione in Italia

GERARDO GIARDIELLO, CHIARA FIORAVANTI,
FRANCESCO ROMANO, MARIASOLE RINALDI*

ABSTRACT: Access to legal information is a fundamental means of guaranteeing people the protection of their rights and the knowledge of their obligations under the law. This access often takes place via thematic public web portals. A legal informatics tool aimed to facilitate the daily monitoring of institutional online information about immigration procedures and legislation has been designed and developed. In this paper we will show the information resources to be monitored, the monitoring tool features and the emerging results during the first months of use of the tool. The emerging difficulties and the ways to fix them will also be highlighted, offering ideas for future developments.

Keywords: Public platforms, Public communication, Immigration, Digital forensics, Web crawler.

1. Il contesto e i bisogni informativi

L'accesso alle risorse informative nell'odierna società dell'informazione e della conoscenza è sempre più da considerarsi un bene primario, che infatti come tale è promosso da accordi e convenzioni internazionali. L'Agenda ONU 2030, ad esempio, prevede tra i propri obiettivi anche quello di «garantire un pubblico accesso all'informazione e proteggere le libertà fondamentali, in conformità con la legislazione nazionale e con gli accordi internazionali»¹.

In un contesto, come quello attuale, in cui le informazioni sono molte e in cui si rischia che il mondo possa «diventare illeggibile per overdose di informazioni» (Galimberti 2020, 117), diviene fondamentale in ambito pubblico disporre di informazioni rilevanti, aggiornate e derivanti da fonti attendibili².

* Istituto di Informatica Giuridica e Sistemi Giudiziari del CNR. gerardo.giardiello@igsg.cnr.it, chiara.fioravanti@igsg.cnr.it, francesco.romano@igsg.cnr.it, mariasole.rinaldi@igsg.cnr.it.

¹ Sui rischi di società in cui i cittadini non siano informati e sul formarsi di élite in grado di dirigere lo Stato e la società, si veda anche (Nichols 2017, 216).

² L'articolo 53 (Siti Internet delle pubbliche amministrazioni) del Codice dell'amministrazione digitale (D.Lgs 7 marzo 2005, n. 82, ss.mm.ii.) prevede, al comma 1, che le «pubbliche

L'accesso all'informazione giuridica, in particolare, diviene poi un irrinunciabile strumento per garantire la possibilità di tutelare i propri diritti e conoscere i propri doveri, specie in società sempre più aperte e come tali soggette all'arrivo di nuovi soggetti, che hanno la necessità primaria di conoscere le norme del Paese di accoglienza³.

In particolare, la reperibilità delle norme è riconosciuta come diritto dei cittadini, i dati giuridici sono collocati tra i dati pubblici essenziali ed è ormai condivisa la necessità di prevedere azioni finalizzate a consentirne l'accessibilità telematica. Si comprende, quindi, la necessità di creare raccolte normative costantemente aggiornate al fine di fornire ai soggetti interessati uno strumento immediato di supporto al lavoro e di accesso all'informazione desiderata⁴.

Dall'esigenza di fornire un accesso unico e attendibile all'informazione giuridica in Rete nel dominio dell'immigrazione, nasce nel 2004 il Progetto PA-eSI (Pubblica Amministrazione e Stranieri Immigrati), frutto della collaborazione tra Regione Toscana, Prefettura di Firenze e Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG) del Consiglio Nazionale delle Ricerche (CNR), nell'ambito del Piano di azione regionale e Toscana, programma di sviluppo dell'e-government della Regione Toscana (Fioravanti e Rinaldi 2010, 93–131) (Fig. 1).

Il progetto si poneva l'obiettivo di contribuire alla diffusione e standardizzazione delle informazioni online su procedure e norme in materia di immigrazione, così da supportare il lavoro degli operatori della pubblica amministrazione e del privato sociale che assistono i cittadini stranieri nelle pratiche burocratiche necessarie per vivere e lavorare in Italia. Tale compito risulta,

amministrazioni realizzano siti istituzionali su reti telematiche che rispettano i principi di accessibilità, nonché di elevata usabilità e reperibilità, anche da parte delle persone disabili, completezza di informazione, chiarezza di linguaggio, affidabilità, semplicità di consultazione, qualità, omogeneità ed interoperabilità (...)». Inoltre l'articolo 6 decreto legislativo n. 33 del 2013 prevede che «Le pubbliche amministrazioni garantiscono la qualità delle informazioni riportate nei siti istituzionali nel rispetto degli obblighi di pubblicazione previsti dalla legge, assicurandone l'integrità, il costante aggiornamento, la completezza, la tempestività, la semplicità di consultazione, la comprensibilità, l'omogeneità, la facile accessibilità (...)».

³ Sulla necessità che la comunicazione istituzionale adempia alla sua «funzione di collante sociale» si veda (Bertolo 2005, 49).

⁴ La necessità che i migranti siano in grado di conoscere le norme dei Paesi di arrivo è un'esigenza sancita in apposite convenzioni internazionali ormai abbastanza datate, come la Convenzione OIL n. 143 del 24 giugno 1975 (sulle migrazioni in condizioni abusive e sulla promozione della parità di opportunità e di trattamento dei lavoratori migranti), che all'articolo 12 invitava i Paesi ratificanti a fare in modo che i lavoratori migranti possano conoscere i loro diritti, i loro obblighi e le attività destinate a dar loro effettiva assistenza per garantire la loro protezione e permettere loro di esercitare i propri diritti. Tale diritto è stato anche ribadito nella Convenzione internazionale sulla protezione dei diritti di tutti i lavoratori migranti e dei membri delle loro famiglie del 1990 all'articolo 33. Sui diritti degli stranieri in Italia si veda anche (Bonetti 2011).

infatti, piuttosto gravoso, sia per la grande complessità e variabilità della materia, sia per il fatto che le competenze sulle diverse procedure sono suddivise fra numerose amministrazioni, nazionali e territoriali, con la conseguenza che le informazioni sono comunicate in modo frammentato nei rispettivi diversi siti⁵.

Il progetto si è concretizzato nella predisposizione di un portale web⁶, tutt'ora aggiornato e sviluppato, che fornisce informazioni di livello nazionale e territoriale (per la Regione Toscana) su procedure, norme e servizi in materia di immigrazione attraverso due distinti sotto-siti, uno indirizzato agli operatori e l'altro agli stranieri, con contenuti progettati *ad hoc* per le due diverse utenze⁷.

La realizzazione di questo punto di accesso unico ha comportato e comporta tuttora, una collaborazione continua con i funzionari che operano negli uffici territoriali degli enti responsabili dei diversi procedimenti (Questura, Prefettura, Camera di Commercio, Comuni, etc.), nonché un'attività giornaliera di monitoraggio dell'informazione istituzionale su procedure e norme resa disponibile nei siti di diversi enti.

Per quanto riguarda le informazioni sulle procedure amministrative rivolte agli operatori di settore, il Portale fornisce un database di schede informative che descrivono in dettaglio le diverse procedure che interessano i cittadini stranieri, indicando tempi del procedimento, dati del responsabile e dell'ufficio, requisiti necessari per l'attivazione, documentazione richiesta, moduli occorrenti, modalità di presentazione e ricezione del provvedimento, riferimenti normativi e domande e risposte frequenti⁸. Tali contenuti richiedono un aggiornamento costante alla luce delle nuove norme o di modifiche organizzative territoriali. L'aggiornamento e la creazione di nuovi contenuti interessano, inoltre, tutte le sezioni tematiche del Portale, a partire da quelle dedicate alle novità normative o procedurali, ai nuovi servizi, ai progetti e alle opportunità nazionali e territoriali⁹.

⁵ Sulla comunicazione amministrativa via web, la comunicazione a cittadini stranieri nei siti istituzionali, nonché sulla pubblicazione e la trasparenza delle informazioni nei siti istituzionali comunali, si vedano vari saggi in (Cattani 2018).

⁶ www.immigrazione.regione.toscana.it/ (ultima consultazione: 30/05/2022).

⁷ Sulla necessità di selezionare solo le informazioni per il proprio destinatario si veda, ad esempio, (Mason, Zilio 2019, 61) e anche (Barker 2005).

⁸ Si tratta di informazioni utili per districarsi nei meandri della burocrazia e dei servizi pubblici italiani, un tipo di informazione definita come «comunicazione di servizio» (Arena 1997).

⁹ La fruibilità e l'accessibilità dei contenuti di siti che si occupano di fornire informazioni a questo target di utenti è tema da tempo dibattuto, così come la necessità che la pubblica amministrazione acquisisca competenze comunicative di tipo interculturale. Si veda su questi temi specifici (Ducci 2012, 40–42) e anche (Ciancio 2014, 136).

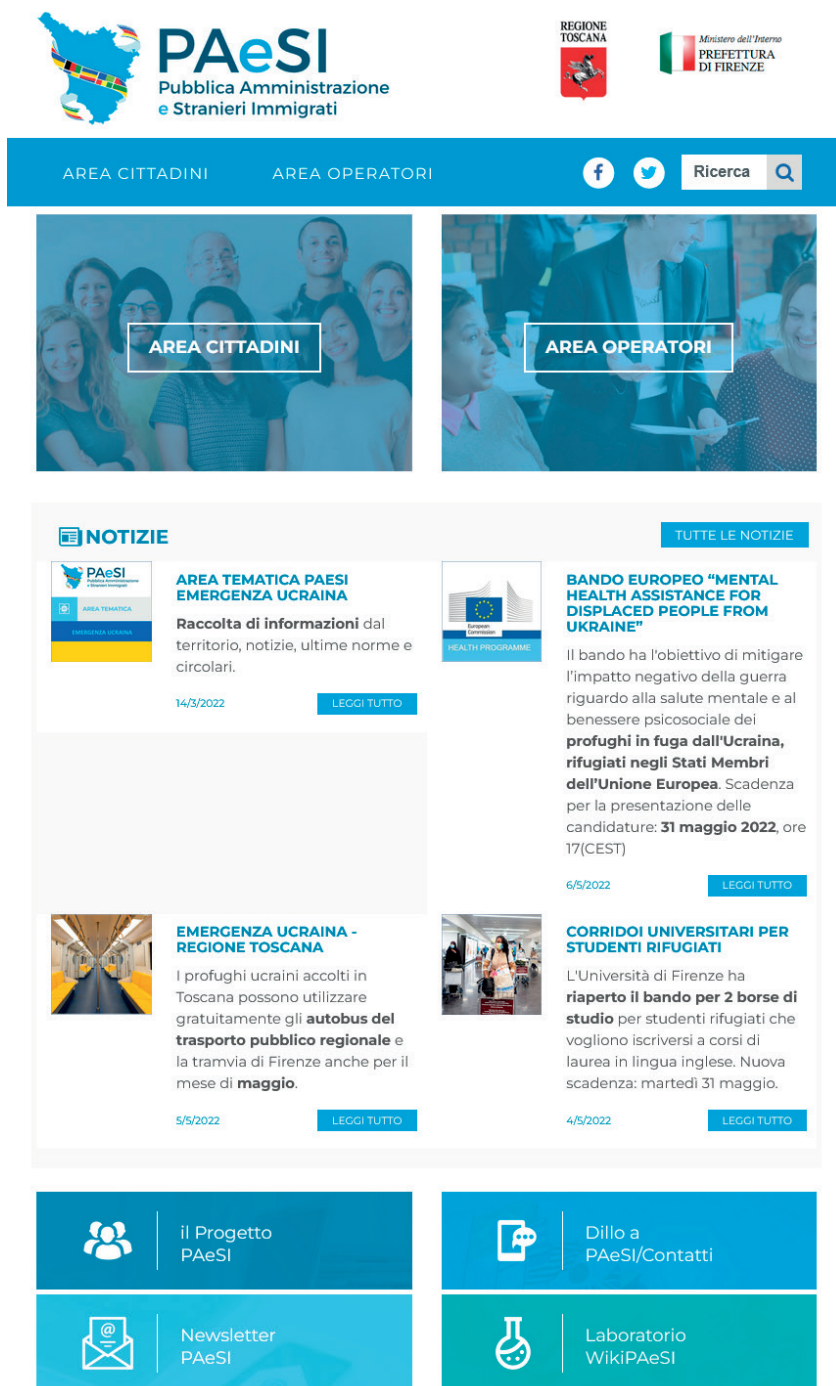


Figura 1: Immagine della home del Portale PAeSI.

Per quanto riguarda invece il reperimento delle norme che riguardano il dominio dell'immigrazione, il Portale, sin dalla sua nascita, ha dedicato una rilevante attività all'implementazione di un database di normativa nazionale, europea/internazionale, regionale e più recentemente anche comunale. Gestire l'informazione normativa nel settore dell'immigrazione presenta numerose problematiche e criticità, prima fra tutte, la necessità di rimanere aggiornati con la costante introduzione di nuove norme nell'ordinamento italiano. La tendenza della Pubblica Amministrazione a trattare, alle volte con eccessiva discrezionalità, le procedure, tanto da creare disomogeneità operativa tra gli stessi Enti presenti in province diverse, è un ulteriore elemento che complica il quadro di riferimento. Da qui deriva il sentito bisogno di offrire una raccolta normativa nazionale di riferimento il più possibile chiara e facilmente consultabile. Esiste infine una terza criticità, legata all'interpretazione normativa, che rende decisamente necessaria la pubblicazione in banca dati di circolari ministeriali esplicative, prevalentemente del Ministero dell'Interno, al fine di poter avere una chiave di lettura e un'interpretazione meno equivocabile della normativa in questione.

Il lavoro incessante di incremento e mantenimento della banca dati consiste dunque nell'introduzione, preceduta dal processo di reperimento e selezione, di nuova normativa e nell'aggiornamento di quella già presente, permettendone la visualizzazione, nelle differenti versioni del testo nel tempo¹⁰.

Tutto ciò rende pertanto fondamentale la consultazione del sito della Gazzetta Ufficiale e di molti altri siti istituzionali in cui reperire informazioni aggiornate.

Le norme in oggetto vengono selezionate nel testo originale e poi convertite in modo automatico o semi automatico in formato eXtensible Markup Language (XML) (Sartor et al. 2011). Tale linguaggio costituisce uno standard per la rappresentazione delle norme che fornisce il vantaggio di poter strutturare il contenuto di un documento utilizzando marcatori personalizzati attraverso un editore specializzato¹¹. Successivamente il testo viene analizzato con un *parser*, che individuando nel testo di legge i riferimenti testuali espliciti, crea in modo automatico link ipertestuali in base ad una sintassi predefinita¹². Inoltre, tale

¹⁰ La Direttiva (UE) 2016/2102 ha introdotto un concetto di accessibilità che va inteso non solo come «l'insieme di principi e tecniche da rispettare nella progettazione, costruzione» di siti web ma anche nella loro «manutenzione e aggiornamento». Si veda sul punto (Caporale 2019, 360).

¹¹ Si veda Circolare 22 aprile 2002 n. AIPA/CR/40 *Formato per la rappresentazione elettronica dei provvedimenti normativi tramite il linguaggio di marcatura XML*. Vedi (Ciampi et al. 2001).

¹² Si tratta della sintassi stabilita dallo standard Uniform Resource Name (URN). Lo standard serve a identificare le risorse normative in modo univoco e persistente indipendentemente dalla loro collocazione fisica in rete, si veda Circolare del 6 novembre 2001, n. AIPA/CR/35 *Assegnazione dei nomi uniformi ai documenti giuridici*. Vedi (Francesconi et al. 2010, 233-52).

standard consente anche di individuare le disposizioni di modifica presenti nel testo di una legge, procedendo al relativo trattamento. Nella banca dati dunque, i provvedimenti sono consultabili nel testo originale, in quello vigente ad una data intermedia, ovvero in una forma redazionale cosiddetta *multivigente* in cui le modifiche intervenute sul testo, fin dalla sua emanazione, sono tutte riportate e visibili nella loro interezza¹³.

Proprio per facilitare l'attività giornaliera di monitoraggio dell'informazione istituzionale online su procedure e norme, che per molti anni è stata svolta manualmente, è stato progettato e realizzato uno strumento di informatica-giuridica, che ha reso l'operazione più veloce ed efficiente ed è oggetto del presente contributo.

Nel presente articolo illustreremo le risorse informative che costituiscono il flusso da monitorare, lo strumento di monitoraggio, le sue funzionalità e i risultati emersi nei primi mesi d'uso dello strumento.

In conclusione, saranno evidenziate le criticità riscontrate e le modalità per poterle gestire, offrendo spunti per possibili sviluppi futuri.

2. Le risorse informative per il Portale PAeSI

Come detto, per mettere a disposizione dell'utenza le informazioni provenienti da diverse fonti istituzionali attraverso i contenuti del Portale PAeSI, risulta necessario svolgere giornalmente un'attività di monitoraggio dei siti degli enti e degli altri soggetti pubblici che hanno competenza in materia.

Per quanto riguarda l'informazione online utile allo sviluppo e aggiornamento dei contenuti informativi su procedure, servizi e opportunità, è necessario innanzitutto monitorare l'informazione di fonte statale. Vengono in particolare controllati i siti del Ministero dell'Interno nella sezione "Immigrazione e Asilo", del Ministero del Lavoro e delle Politiche Sociali (tema immigrazione) e del Ministero dell'Istruzione per le informazioni sull'accesso allo studio in Italia dei cittadini stranieri.

I tre Ministeri hanno inoltre realizzato nel 2012 un portale tematico (integrazionemigranti.gov.it) per favorire l'integrazione dei migranti attraverso la diffusione di informazioni sui servizi pubblici nazionali e territoriali.

D'interesse anche i siti del Ministero della Salute per quanto riguarda la tutela della salute dei migranti, del Ministero degli Esteri per comunicazioni sui servizi consolari e le opportunità per lo studio in Italia e del Ministero della Giustizia per le informazioni sul riconoscimento di alcuni titoli di studio e il tema dei detenuti stranieri.

Il sito del Dipartimento delle Pari Opportunità della Presidenza del Consiglio dei Ministri viene controllato per verificare le iniziative in materia di

¹³ Per le modalità di marcatura di tali documenti si veda (Francesconi 2012).

contrasto alla tratta e alle mutilazioni genitali femminili ed il sito di UNAR, l'Ufficio Nazionale Antidiscriminazione Razziale, per quelle relative al contrasto alle discriminazioni.

Il sito del SAI (Sistema Accoglienza Integrazione), il sistema nazionale di protezione per titolari di protezione internazionale e per minori stranieri non accompagnati, deve essere monitorato per i report, le risorse formative, i progetti e gli avvisi su accoglienza e protezione internazionale.

Informazioni rilevanti vengono ricavate anche dal sito dell'Istituto Nazionale della Previdenza Sociale (INPS) nella specifica sezione dedicata ai lavoratori migranti e dal sito Istituto nazionale Assicurazione Infortuni sul Lavoro (INAIL) per quanto riguarda il tema della sicurezza sul lavoro.

A livello territoriale, avendo il progetto PAeSI come riferimento informativo la Toscana, sono monitorati i siti degli enti che sul territorio sono responsabili delle procedure nazionali, come le Prefetture e le Questure delle province toscane.

In ambito regionale, poi, oltre al sito della Regione Toscana, sono sottoposti a monitoraggio il sito del Centro regionale di documentazione per l'infanzia e per l'adolescenza, per le informazioni relative ai minori stranieri non accompagnati in Toscana e quello Centro Salute Globale della Regione Toscana, in tema di diritto alla salute dei migranti.

Per quanto riguarda l'accesso dei cittadini stranieri al servizio sanitario regionale, risultano d'interesse i siti delle tre Aziende usl toscane (Toscana centro, Toscana nord-ovest e Toscana sud-est). Per informazioni territoriali in materia di lavoro autonomo, vengono controllati i siti delle Camere di Commercio delle diverse province. In materia di lavoro d'interesse anche i siti delle associazioni di categoria, come Confartigianato e Confagricoltura toscana.

Il monitoraggio territoriale comprende poi i siti dei Comuni per le informazioni sulle procedure di anagrafe, stato civile e cittadinanza di loro competenza, sui servizi informativi e di assistenza alla compilazione delle pratiche per i cittadini stranieri e sulle iniziative comunali in materia di integrazione.

Le risorse informative da monitorare, dal punto di vista normativo, legate al tema dell'immigrazione sono molteplici in quanto nel sito PAeSI sono presenti risorse nazionali, internazionali, europee e locali (Toscana).

Le norme, infatti, devono essere costantemente monitorate sia per inserirne di nuove, sia per aggiornare quelle già presenti nel database.

Per questo tipo di lavoro il principale punto di riferimento è sicuramente rappresentato dal sito web della Gazzetta Ufficiale. Questa rappresenta la fonte ufficiale di conoscenza delle norme in vigore in Italia e lo strumento di diffusione, informazione e ufficializzazione di testi legislativi. Al fine di reperire norme d'interesse viene consultata la "Serie generale"¹⁴ con l'elenco delle gazzette pubblicate negli ultimi 30 giorni.

¹⁴ Questa viene pubblicata tutti i giorni feriali.

Nel sito della Gazzetta Ufficiale si trova anche la sezione relativa alla “Unione Europea” (2^a Serie Speciale) che riporta gli atti comunitari come regolamenti, decisioni e direttive.

Questi atti normativi sono reperibili anche dal sito “Eurlex”, sia nella sezione “Gazzetta Ufficiale dell’UE”, sia nella sezione “Diritto dell’UE”- “Cerca negli atti giuridici”.

Qui, attraverso diverse chiavi di ricerca, troviamo atti come i trattati dell’UE, le direttive, i regolamenti, le decisioni e la legislazione consolidata¹⁵.

Nel sito della Gazzetta Ufficiale della Repubblica Italiana si trova, inoltre, la sezione “Regioni” (3^a Serie Speciale)¹⁶; in questa pubblicazione trovano spazio tutti gli atti normativi e amministrativi di interesse nazionale emanati dalle singole Regioni¹⁷, come leggi regionali, decreti presidenziali e regolamenti.

Questi ultimi possono essere reperiti chiaramente anche nel sito della Regione Toscana nella sezione “Leggi, Atti e normative”, in cui, con modalità di ricerca avanzata, è possibile circoscrivere le risorse esistenti per materia o per parole nel testo.

Per avere un quadro completo e al fine di monitorare la produzione normativa nazionale in materia di immigrazione è opportuno consultare anche il sito dell’INPS. Nella sezione “Inps Comunica”- “Atti” sono raccolti tutti gli atti ufficiali relativi alla materia previdenziale e pubblicati nella Gazzetta Ufficiale (Atti ufficiali) nonché quelli emanati direttamente dall’INPS (Circolari, Messaggi e Normativa).

Infine, importanti risorse normative, utili a incrementare il vasto dominio dell’immigrazione, sono sicuramente rappresentate dagli atti emanati dal Ministero dell’Interno.

Nel sito web di questo ministero, infatti, nella sezione “Atti amministrativi generali” l’amministrazione pubblica decreti, direttive, circolari, programmi e ogni altro atto che dispone in generale sull’organizzazione, sulle funzioni, sugli obiettivi, sui procedimenti, o che determina l’interpretazione di norme giuridiche che riguardano o dettano disposizioni per l’applicazione.

In particolare, nella sezione “Circolari”, è possibile visualizzare le circolari di maggior interesse per gli utenti adottate dai dipartimenti del Ministero dell’Interno, dal Gabinetto del ministro e dai commissari governativi e straordinari. Altre circolari del Ministero disponibili online sono consultabili nelle pagine web delle singole direzioni o dei vari dipartimenti.

Il Dipartimento per gli Affari interni e territoriali, ad esempio, svolge funzioni di supporto alle attività di governo locale, di garanzia della regolare costituzione degli organi elettivi, del loro funzionamento e attività di collaborazio-

¹⁵ La consolidazione consiste nel proporre un documento che raggruppi l’atto giuridico di base e tutte le modifiche o rettifiche apportate successivamente.

¹⁶ Questa viene pubblicata il sabato.

¹⁷ Nel nostro caso sono d’interesse quelli emanati da Regione Toscana.

ne con gli enti locali, e vigilanza sullo stato civile e sull'anagrafe. Tra gli Uffici di diretta collaborazione per questo dipartimento, di particolare importanza per il dominio di interesse è la Direzione centrale per i Servizi demografici che svolge compiti di indirizzo, coordinamento e vigilanza in materia di anagrafe e stato civile.

Di interesse anche le circolari emesse dal Dipartimento della Pubblica sicurezza, in quanto titolare di tutte le attività connesse alla gestione dell'ordine e della sicurezza pubblica, al coordinamento tecnico-operativo delle Forze di polizia. Ed infine sono fondamentali quelle emesse dal Dipartimento per le Libertà civili e l'Immigrazione.

Le circolari di questo ufficio, pur rientrando negli atti di interpretazione, istruzione ed organizzazione destinati ai vari uffici, centrali e periferici del Ministero dell'Interno, per una corretta e puntuale applicazione delle norme, si caratterizzano, in gran parte dei casi, per la diretta incidenza sul tessuto socio-economico del Paese. In tal senso sono significative le tante circolari emanate in tema di regolarizzazione dei lavoratori extracomunitari, procedure riconducibili alle competenze dello Sportello Unico per l'Immigrazione e sulle tematiche riguardanti la concessione della cittadinanza. Tra i provvedimenti presenti in questa sezione possiamo trovare anche gli avvisi emanati dalle Autorità Responsabili dei Fondi dell'Unione europea, i bandi di gara, l'approvazione di graduatorie e gli atti di natura organizzativa tipici delle amministrazioni pubbliche.

In occasione del censimento e continuo monitoraggio dei siti web dei Comuni toscani relativamente a informazioni su servizi¹⁸ e procedimenti si è deciso di realizzare e incrementare una raccolta di atti amministrativi e normativi (in formato XML) di competenza comunale in materia di immigrazione.

Le tipologie di atti individuati (accordi, circolari, deliberazioni, determinazioni, protocolli d'intesa) relative a precise tematiche, contribuiscono a realizzare e perfezionare una sezione dedicata in cui mettere in evidenza le informazioni prodotte dagli Enti locali del territorio su servizi e procedimenti dei Comuni che spesso si trovano collocati in posizioni di non immediata visibilità nelle rispettive reti civiche e di facilitarne la fruizione.

3. Descrizione dello strumento

Al fine di supportare l'attività di reperimento dell'informazione rilevante per lo sviluppo e l'aggiornamento del Portale PAeSI, è stato progettato e realizzato un applicativo informatico in grado di effettuare le seguenti azioni:

¹⁸ <https://www2.immigrazione.regione.toscana.it/?q=procedimenticomunali> (ultima consultazione: 30/05/2022).

- monitorare regolarmente alcune fonti online di informazione e recuperare in automatico i testi pubblicati;
- sottoporre all'esperto di dominio le risorse selezionate come rilevanti;
- trarre dalle valutazioni dell'esperto, informazioni utili per correggere, quindi migliorare, le ulteriori ricerche dello strumento.

Le varie azioni saranno descritte nei paragrafi che seguono¹⁹.

3.1. *Monitoraggio e recupero dell'informazione*

Per quanto concerne la prima azione e cioè il monitoraggio delle fonti informative e il recupero delle risorse, lo strumento effettua una ricerca giornaliera (programmata in orario notturno) su tutte le fonti che l'esperto di dominio ha individuato e inserito come da monitorare. Operativamente l'applicativo effettua le stesse azioni che gli operatori del Portale PAeSI compiono con regolarità, ovvero la consultazione di siti istituzionali alla ricerca di norme e informazioni in materia di immigrazione.

La capacità di valutare se questi documenti pubblicati sono interessanti, ovvero trattano temi legati al dominio del Portale, è simulata andando a verificare la presenza di un determinato numero di argomenti citati nei loro testi. In pratica, lo strumento ricerca all'interno di tali testi certe espressioni che l'esperto di dominio ha precedentemente fornito e che, in base alla sua esperienza, sono quelle che più di frequente ricorrono nelle risorse scelte come rilevanti. Il metodo che abbiamo definito, seppur semplice, copre la casistica delle espressioni presenti nei testi di interesse, che è di fatto limitata, visto che andiamo ad analizzare pagine web che utilizzano un linguaggio istituzionale, meno soggetto a variazioni linguistiche rispetto ad altri tipi di testo. La compilazione di tale elenco di espressioni è avvenuta in vari momenti.

All'esperto di dominio sono state inizialmente fornite analisi statistiche sulle parole utilizzate nei testi estratti dalla banca dati normativa nazionale in tema di immigrazione presente sul Portale PAeSI. Tale analisi non ha evidenziato un set di parole che potesse identificare in modo univoco il dominio dell'immigrazione, ma ha permesso di individuare le parole riguardanti i temi trattati dalla normativa attuale e, con una integrazione di termini, ha consentito di stilare un primo elenco di parole da ricercare.

Successivamente è stato utilizzato tale elenco per verificare se con esso lo strumento sarebbe stato in grado di selezionare (e quindi proporre all'operatore per la sua valutazione) tutti i testi della banca dati normativa precedentemente analizzata; l'analisi dei testi non individuati dallo strumento ha

¹⁹ Lo strumento è accessibile, con le apposite credenziali, a questa URL www2.immigrazione.regione.toscana.it/spider (ultima consultazione: 30/05/2022).

permesso all'esperto di dominio di selezionare ulteriori parole significative da inserire nel suddetto elenco.

Infine, è stato utilizzato l'elenco di parole per analizzare le ultime 400 decisioni, 400 direttive e 400 regolamenti pubblicati, in lingua italiana, su EUR-lex, per individuare i testi in cui tali parole erano presenti. Gli esperti del dominio hanno poi valutato la pertinenza dei testi individuati e hanno evidenziato le norme europee in materia di immigrazione che non erano state segnalate dallo strumento. Dopo tale verifica si è potuto incrementare l'elenco delle parole precedentemente individuate per la ricerca automatica delle informazioni d'interesse. Questo processo manuale semi guidato ha permesso di affinare e convergere verso un numero limitato di espressioni che sono state, quindi, generalizzate per coprire le eventuali varianti d'uso, tradotte in espressioni regolari e fornite allo strumento (Fig. 2)²⁰.

Espressioni da monitorare: 121		
La lista è redatta manualmente in base ai precedenti documenti selezionati. Si possono segnare eventuali pesi per evitar falsi positivi dello spider.		
(crisi processo contesto fluss.) (im)?migratori(,)?	Esempi individuali - crisi migratoria - flussi migratori - contesto migratori - processo immigratorio -	100
tratta(di degli)*(esseri)* umani	Esempi individuali - tratta (degli esseri umani) -	100
(nullaosta nulla osta)(all per)? (lavoro matrimonio)	Esempi individuali - nulla-osta al lavoro - nulla-osta al lavoro -	100
vist.(di per)?(corto lungo)? (soggiorno lavoro)	Esempi individuali - visti per lavoro - visti corto soggiorno -	100
(regolar(, mente) status permesso carta)(di)? soggiorno(olante)	Esempi individuali - carta (di) soggiorno - status di soggiornante - regolarmente soggiornante - permesso (di) soggiorno -	100
xenofobi(c)?alrazzisi(ta mo)	Esempi individuali - razzismo - razzista - xenofobia - xenofobia -	100
protezione (internazionale (dei)?minori sussidiaria temporanea)	Esempi individuali - protezione temporanea - protezione sussidiaria - protezione (dei) minori - (status di) protezione internazionale -	100
(discriminazione motivi intolleranza integrazione origine) (raz(z)?ia ,etnic.)	Esempi individuali - origine etnica - intolleranza razziale - integrazione etnica - motivi etnici - motivi razziali - motivi etnici - discriminazione (etnica) razziale -	100
lavoro ((in)?nero irregolare)	Esempi individuali - lavoro (in) nero - lavoro irregolare -	55

Figura 2: Elenco delle espressioni regolari fornite allo strumento.

Lo strumento dispone di una ulteriore chiave di ricerca: è in grado di estrarre automaticamente la lista delle norme presenti nella banca dati normativa italiana del Portale PAeSI, e i relativi identificatori URN. Tramite un *parser* linguistico, lo strumento è in grado di ricercare nei testi analizzati gli estremi dei riferimenti ad atti normativi/amministrativi e calcolare i corrispondenti

²⁰ In precedenti progetti di informatica giuridica sono state usate espressioni regolari al fine di identificare in testi giuridici e 'marcare' in linguaggio XML porzioni di testo significative per il riconoscimento di nessi normativi quali riferimenti testuali espliciti e modifiche testuali esplicite. Vedi (Mercatali, Romano, Bolioli 2004).

identificatori URN. Testando questi identificatori è in grado quindi di ricercare nei testi le citazioni ad atti presenti nel database delle norme sull'immigrazione, trattando di fatto questi riferimenti come espressioni di selezione in Rete di testi utili per gli operatori del Portale.

L'esperto di dominio fornisce allo strumento anche una lista di indirizzi di pagine online dove sono pubblicate notizie (pagine principali o sezioni 'notizie' dei portali istituzionali) e norme (banche dati o gazzette online). Lo strumento visita regolarmente tali pagine e scarica tutti i testi pubblicati che non aveva precedentemente scaricato. Tali testi sono analizzati con l'ausilio delle espressioni regolari e la lista dei riferimenti normativi e valutati come rilevanti o non rilevanti in materia di immigrazione in base all'individuazione o meno di parole e riferimenti delle suddette liste²¹.

3.2. *Intervento dell'esperto del dominio*

Lo strumento compie il suo lavoro senza necessità di interventi da parte dell'operatore. Tuttavia, l'operatore ha a disposizione una serie di informazioni dettagliate e aggregate su pagine visitate, documenti scaricati, tempi impiegati, documenti valutati, errori che inevitabilmente possono verificarsi durante l'attività.

È possibile visualizzare un report (Fig. 3) di eventi aggregati occorsi nell'ultima esecuzione dell'attività di ricerca e analisi dello strumento.



Figura. 3: Report eventi ultima esecuzione dello strumento.

Tale report evidenzia il numero documenti (cioè i link) verificati e presenti sui siti monitorati, il numero di documenti nuovi (rispetto alla visita precedente) aggiunti al database interno dello strumento di ricerca, il numero di errori e delle segnalazioni occorse.

²¹ In un progetto precedente è stato sperimentato un approccio basato unicamente sulla ricerca dei riferimenti normativi nazionali ed europei (in materia di disciplina del trattamento dei dati personali) al fine di mantenere aggiornato il DPO (*Data protection officer*) del progetto sul modificarsi della normativa di riferimento. Vedi (Arezzini et al. 2018, 221-32).

Nel rapporto (Fig. 4) si mostrano, invece, le attività aggregate occorse in tutte le esecuzioni dello strumento: tempo impiegato nella selezione dei link dei documenti, tempo impiegato per l'analisi degli stessi, dettagliato in tempo impiegato nel prelevamento dei documenti e analisi dei testi, altro tempo impiegato dallo strumento.

L'utente operatore guardando il report con i valori aggregati, può verificare che lo strumento si sta comportando correttamente, verificando che il numero di documenti, il tipo e i tempi impiegati rimangano in trend noti/limitati.

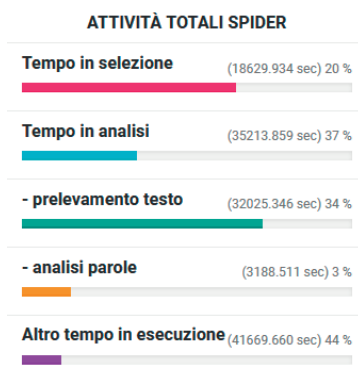


Fig. 4: Report attività strumento.

Un numero di segnalazioni particolarmente elevato o la segnalazione di tempi eccessivi di esecuzione, potrebbe essere indagata andando a consultare l'elenco completo delle attività eseguite dallo strumento di ricerca (Fig. 5). Lo strumento classifica tali attività in eventi normali (informazioni), in attività che hanno generato un errore (errori) e altre attività che, pur non essendo errori, potrebbero nascondere una qualche problematica (segnalazioni): ad esempio nel caso in cui lo strumento di ricerca abbia analizzato un testo molto corto.

L'utente può, quindi, approfondire la sua analisi andando a visionare sia l'elenco completo delle attività dello strumento, sia ogni singolo documento scaricato e analizzato, andando a leggere il testo scaricato e poi analizzato e le parole individuate (evidenziate nel testo) che hanno permesso di selezionare il documento come pertinente.



Fig. 5: Report log.

Il sistema mostra (Fig. 6) all'esperto di dominio anche l'attività svolta (documenti che lo strumento ha individuato e proposto all'utente e da esso valutati come pertinenti o non pertinenti) e da svolgere (documenti che lo strumento di ricerca ha proposto alla valutazione dell'esperto, ma ancora non valutati da quest'ultimo), tramite un cruscotto informativo.

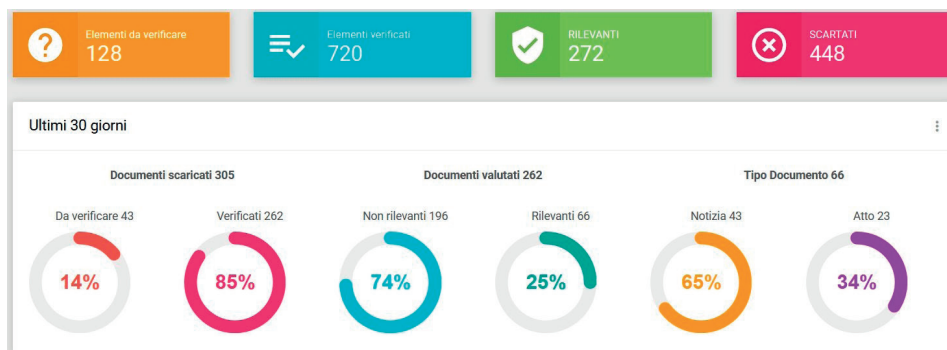


Fig. 6: Cruscotto informativo.

Nel secondo riquadro è possibile avere una visualizzazione degli ultimi 30 giorni di attività (sono selezionabili anche gli intervalli di 7 e 15 giorni) in cui sono mostrati anche la ripartizione dei documenti rilevanti in notizie e norme.

Infine (Fig. 7), lo strumento indica all'utente i documenti che ha automaticamente selezionato e ritiene rilevanti come un elenco di documenti caratterizzata da una data (di reperimento del testo online), da un titolo del documento (recuperato dal titolo della pagina o dalla parte testuale del link ipertestuale che ha individuato lo strumento nella sua ricerca), dall'indirizzo della risorsa online (per visionare il documento basta cliccare sul titolo e si apre una nuova finestra all'indirizzo della risorsa) e dalla lista di parole che hanno permesso di selezionare il documento.

Comunicato Stampa https://www.ispettorato.gov.it/it-it/notizie/Pagine/Comunicato-Stampa-20042021.aspx Parole segnalate: caporalato, sfruttamento lavorativo	21/04/2021	☐ Rilevante ☐ Non rilevante
Lavoratori italiani all'estero: retribuzioni convenzionali 2021 http://www.inps.it/nuovoportaleinps/default.aspx?itemdir=55261 Parole segnalate: paesi extra UE, lavoratori extracomunitari, lavoro irregolare	21/04/2021	☐ Rilevante ☐ Non rilevante
Decreto Sostegni: semplificazioni e requisiti per indennità a lavoratori http://www.inps.it/nuovoportaleinps/default.aspx?itemdir=55262 Parole segnalate: lavoratori domestici, lavoro irregolare, lavoratori stagionali	21/04/2021	☐ Rilevante ☐ Non rilevante
Cina: nella biblioteca di Canton nasce la sezione italiana https://www.esteri.it/mae/it/sala_stampa/archivionotizie/retediplomatica/cina-nella-biblioteca-di-canton-nasce-la-sezione-italiana.html Parole segnalate: Console Generale	21/04/2021	☐ Rilevante ☐ Non rilevante
Accordo Regione-sindacati per i diritti dei rider https://www.toscana-notizie.it/archivio/-/asset_publisher/lyd2ls2gGDzu/content/id/56724114 Parole segnalate: caporalato, sfruttamento lavorativo	21/04/2021	☐ Rilevante ☐ Non rilevante
MINISTERO DEGLI AFFARI ESTERI E DELLA COOPERAZIONE INTERNAZIONALE - COMUNICATO http://www.gazzettaufficiale.it/eli/id/2021/04/20/21A02340/SG Parole segnalate: Consolato generale	21/04/2021	☐ Rilevante ☐ Non rilevante
Gazz. uff. dell'UE L 134/2021 https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=OJ:L:2021:134:FULL Parole segnalate: cle	21/04/2021	☐ Rilevante ☐ Non rilevante
ITL Ravenna Forlì Cesena: lavoro nero e omessi riposi https://www.ispettorato.gov.it/it-it/notizie/Pagine/ITL-Forlì-Ravenna-Cesena-lavoro-nero-e-omessi-riposi.aspx Parole segnalate: Lavoro nero, cittadini extracomunitari, lavoro nero	20/04/2021	☐ Rilevante ☐ Non rilevante

Fig. 7: Lista documenti valutati come rilevanti.

All'utente è sufficiente selezionare il riquadro rilevante, o non rilevante, per confermare allo strumento se il documento è effettivamente pertinente.

3.3. Informazioni utili per migliorare le ulteriori ricerche

L'aggiunta di molte fonti informative da monitorare ha, come previsto, portato rumore alla ricerca. Il sistema ha iniziato a proporre, all'esperto di dominio, documenti che risultavano completamente estranei al tema dell'immigrazione.

Per ovviare a questo problema, è stata introdotta una modalità di lavoro in base alla quale l'operatore può agire sul sistema andando a pesare le espressioni di ricerca in modo tale che lo strumento di ricerca non selezioni un documento se trova unicamente o un numero limitato di occorrenze di espressioni a peso ridotto.

L'operazione di pesatura di queste parole può essere fatta *ex post*, trovando materialmente documenti proposti dallo strumento mal valutati, o *ex ante*, cercando espressioni nella banca dati interna dello strumento, che lo strumento di ricerca incrementa con i documenti automaticamente scaricati e manualmente valutati dall'esperto di dominio.

Interrogato con una parola/espressione, lo strumento visualizza i risultati raggruppando i documenti risultanti in: documenti non segnalati, documenti rilevanti, documenti scartati.

Risulta evidente che l'aggiunta di una parola/espressione che seleziona molti documenti, già valutati come da scartare, porterebbe rumore alla ricerca, andrebbe quindi inserita con un peso basso o non inserita.

Differentemente, aggiungere una parola, non presente fra quelle ricercate, che individua un numero consistente di documenti "rilevanti", potrebbe migliorare le prestazioni della ricerca dello strumento.

4. Primi risultati

Guardando i dati degli ultimi 30 giorni, un documento su quattro, tra quelli proposti dal sistema, risulta valutato come rilevante dall'esperto di dominio.

Tra i documenti valutati come rilevanti, circa un terzo riguarda documenti provenienti da siti istituzionali di produzione normativa, mentre i restanti due terzi sono selezionati da siti che pubblicano informazione qualificata (tra questi ultimi circa un terzo riguarda atti normativi e circa due terzi notizie relative al dominio).

Se si valutano le prestazioni totali, avvenute in circa quattro mesi di attività, il rapporto fra documenti valutati come rilevanti, rispetto a quelli segnalati, si attesta sul 33% (2 su 3 non rilevanti).

Inizialmente le percentuali di riconoscimento erano migliori ma il sistema monitorava quasi esclusivamente le gazzette dei siti istituzionali a livello europeo, italiano e toscano e pochi altri siti istituzionali, come quello del Governo italiano o della Regione Toscana.

Attualmente, invece, le fonti monitorate sono oltre 100 e comprendono tutte le risorse informative indicate nel paragrafo 2; il rapporto fra documenti rilevanti e segnalati si attesta sul 25% (3 su 4 non risultano rilevanti).

Ovviamente, monitorando molte più fonti informative il 'rumore', come detto già, è aumentato, ma il sistema continua a funzionare in maniera molto utile a sostegno dell'attività di aggiornamento del Portale.

5. Conclusioni

Come abbiamo sopra riferito, sono già state operate alcune modifiche alla lista delle parole e alle espressioni regolari da ricercare al fine di evitare ricerche troppo "rumorose".

Valutazioni analoghe potrebbero essere fatte sulle ricerche effettuate sulla base delle norme. Attualmente non è stato ritenuto necessario introdurre la pesatura dei riferimenti normativi, ma trovare citato un riferimento normativo, piuttosto che un altro, è un forte indicatore di rilevanza.

Ad esempio, se il testo che lo strumento di ricerca sta analizzando contiene un riferimento al Testo unico sull'immigrazione, è molto probabile che sia

più rilevante rispetto ad un testo che cita la legge di bilancio dello Stato, che conteneva al suo interno un solo comma in materia di immigrazione e che per questo motivo fa parte del nostro database.

Analogamente potrà essere perfezionato il sistema di monitoraggio per alcune fonti informative, andando a differenziare il set di parole da usare nella ricerca o la pesatura delle stesse.

Ad esempio, sul sito delle Prefetture e delle Questure ci sono moltissime notizie, di cronaca giornalistica che, usando un linguaggio molto meno controllato e trattando il tema dell'immigrazione, possono comportare la selezione di notizie che, pur contenendo le parole del nostro set di ricerca, non sono tuttavia pertinenti per l'aggiornamento del Portale PAeSI.

Allo stesso modo, la selezione di informazioni da siti del medesimo tipo (ad esempio da siti di ministeri diversi) possono comportare criticità connesse all'uso di una data parola in domini diversi.

È il caso di ad esempio CIE (Centri di Identificazione ed Espulsione) e CPT (Centri di Permanenza Temporanea) che sono utilizzate anche in altri domini con altri significati.

Lo strumento attualmente è già in grado di mostrare all'utente (Fig. 8) valutazioni prettamente statistiche sulle parole individuate e quindi di aggregare, a vario livello, indici sui termini trovati o sui documenti valutati.



Figura 8: Statistiche sulle parole rilevate nei siti analizzati.

Tali informazioni potrebbero essere sufficienti a innescare il processo di ri-valutazione sui pesi dati alle varie parole, sui siti che attualmente sono mo-

nitorati e quindi potrebbero rivelarsi utili per progettare analisi più accurate, differenziando le stesse a seconda del contesto²².

Riferimenti bibliografici

- Arena, Gregorio. 1997. "Introduzione all'amministrazione condivisa". *Studi parlamentari e di politica costituzionale*, nn. 117-118.
- Arezzini, Silvia, Paolo Bosco, Sara Calderoni, Alberto Ciampa, Simone Coscetti, Stefano Cuomo, Luca De Santis et al. 2016. "ARIANNA: un Ambiente di Ricerca Interdisciplinare per l'Analisi di Neuroimmagini nell'autismo", in *Informatica e diritto*, no. 2: 221-32.
- Barker, Iain. 2005. *Information scent helping people find the content they want* https://www.steptwo.com.au/papers/kmc_informationscent/.
- Bertolo, Carla. 2005. *L'interfaccia e il cittadino. Comunicazione pubblica tra tecniche e riflessività*. Milano: Edizioni Angelo Guerini e associati.
- Bonetti, Paolo. 2011. *Diritti fondamentali degli stranieri*. <http://www.adir.uni-fi.it/odv/documenti/frontiere/bonetti.htm>.
- Caporale, Marina. 2019. "L'accessibilità ai siti web e alle applicazioni mobili delle pubbliche amministrazioni". *Giornale di diritto amministrativo* no. 3: 357-67.
- Cattani, Paola, e Giuseppe Sergio, cur. 2018. *Comunicare cittadinanza nell'era digitale. Saggi sul linguaggio burocratico 2.0*. Milano: Franco Angeli.
- Ciampi, Costantino, Paolo Carlo Guidotti, Mario Ragona, Lorian Serrotti, Fiorenza Socci, Pierluigi Spinosa, Andrea Marchetti et al. 2001. "Progetto NIR – Fase 2 Accesso alle norme in rete. Recupero della normativa pregressa in formato XML e standards del progetto (DTD e URNs)". *Informatica e diritto*, no. 1: 1-266.
- Ciancio, Bruno. 2014. *Sviluppare la competenza interculturale. Il valore della diversità nell'Italia multietnica. Un modello operativo*. Milano: Franco Angeli.
- Ducci, Gea. 2012. *La comunicazione pubblica e la sfida dell'interculturalità. Lo "sguardo" dei mediatori culturali nella Regione Marche*. Milano: Franco Angeli.

²² Lo strumento descritto è stato sviluppato nell'ambito del progetto Tuscany Empowerment Actions for Migrant System (TEAMS), finanziato dal Fondo Asilo, Migrazione e Integrazione (FAMI) 2014-2020 – Obiettivo Specifico 2. Integrazione/Migrazione legale. Obiettivo nazionale (ON) 2. Integrazione. Piani d'intervento regionali per l'integrazione dei cittadini di paesi terzi.

- Fioravanti, Chiara e Mariasole Rinaldi. 2010. "Il sistema informativo PAeSI: un accesso telematico unico a informazioni, norme e procedimenti in materia di immigrazione". *Informatica e diritto*, no. 1–2: 93–131.
- Francesconi, Enrico. 2012. "Parser e Editor xmLeges". In *Linee guida per la marcatura dei documenti normativi secondo gli standard*, a cura di Maria Pia Giovannini, Monica Palmirani, Enrico Francesconi. Firenze: European Press Academic Publishing, 2012.
- Francesconi, Enrico, Carlo Marchetti, Remigio Pietramala, and Pierluigi Spinosa. 2010. "URN-based Identification of Legal Acts: The Case of the Italian Senate". *Informatica e diritto*, no. 1–2: 233–52.
- Galimberti, Umberto. 2020. "Se le nuove tecnologie rendono inutile comunicare". In *Il primato delle tecnologie. Guida per una nuova iperumanità*, a cura di Carlo Bordini. Milano: Mimesis.
- Mason, Giacomo, e Piero Zilio. 2019. *La comunicazione digitale per la PA*. Santarcangelo di Romagna: Maggioli.
- Mercatali, Pietro, Francesco Romano, and Andrea Bolioli. 2004. "Formal Models for a Legislative Grammar. Explicit Text Amendment". In *Knowledge Management in Electronic Government (KMGov2004)*, a cura di Maria Wimmer, 210–27. Berlin: Springer.
- Nichols, Tom. 2017. *La conoscenza e i suoi nemici. L'era dell'incompetenza e i rischi per la democrazia*. Tradotto da Chiara Veltri. Roma: LUISS University Press.
- Sartor, Giovanni, Monica Palmirani, Enrico Francesconi, and Maria Angela Biasiotti, cur. 2011. *Legislative XML for the Semantic Web. Principles, Models, Standards for Document Management*. Berlin: Springer.

Termini e testi nella comunicazione dei gruppi No-vax

CLAUDIA LANZA*

ABSTRACT: The study presents an analysis of the recursive linguistic structures characterizing the No-vax movement's messages on the social media environment, in particular on Twitter, where the collection of specific time-framed public posts is feasible to be achieved and they represent a solid documental dataset in input from which to start the semantic analysis tasks. The goal of this paper is to present a case-study overview of the influence the information flow within web documentary typologies exerts on new forms of linguistic standardized models. The results will display some of the main recurrent semantic chains used by outmembers on Twitter to serve as references to demonstrate how prefixed has become the language proper to specific anti-pandemic vaccination and regulations groups.

Keywords: COVID-19, No-vax, Semantic chains, Terminology, Twitter.

1. Introduzione

Nell'attuale contesto pandemico da COVID-19 la mole di dati e l'immediatezza con la quale questi vengono resi fruibili al largo pubblico sono caratterizzate da una forte ciclicità e livellamento informativo che hanno determinato un alto grado di disinformazione (Furini and Menegoni 2018). Nello specifico, in questo particolare contesto, si parla di "infodemia", termine derivante dalla fusione del concetto di distribuzione su larga scala delle informazioni e del concetto di epidemia come elemento di viralità in un ecosistema (Rovetta 2022). L'Organizzazione Mondiale della Sanità (OMS) definisce il fenomeno dell'infodemia come:

An infodemic is too much information including false or misleading information in digital and physical environments during a disease outbreak. It causes confusion and risk-taking behaviours that can harm health. It also leads to mistrust in health authorities and undermines the public health response. An

* Università della Calabria, Dipartimento di Culture, Eduzione e Società (DiCES), Arcavacata di Rende (CS), Italia. claudia.lanza@unical.it.

infodemic can intensify or lengthen outbreaks when people are unsure about what they need to do to protect their health and the health of people around them. With growing digitization – an expansion of social media and internet use – information can spread more rapidly. This can help to more quickly fill information voids but can also amplify harmful messages.

Infodemic management is the systematic use of risk- and evidence-based analysis and approaches to manage the infodemic and reduce its impact on health behaviours during health emergencies¹.

Si tratta di un fenomeno che provoca la disseminazione di un elevatissimo numero di dati spesso discordanti tra loro che disorienta i processi di analisi critica da parte dei cittadini, i quali perdono di volta in volta fiducia nelle istituzioni garanti della pubblicazione di notizie potenzialmente risolutive. Questa babele di «narrazioni giornalistiche contraddittorie e spettacolarizzanti» (Pedroni 2020) rende difficile la realizzazione di un sistema di credibilità omogeneo che appoggi le scelte governative nazionali e ne sia promulgatore per un convincimento massivo. Al contrario, i messaggi vengono trasmessi attraverso il web in modo ripetitivo e allo stesso tempo contrastante in quanto riflesso del cambiamento legato alla trasformazione dell'infettività del virus COVID-19 e, di conseguenza, delle azioni di contrasto adottate per fronteggiare i processi evolutivi pandemici. Tutto ciò genera un contesto che facilita la creazione delle cosiddette *echo chambers*, ovvero,

“camere dell’eco” (*echo chambers*), «ambienti chiusi nei quali le parole e le frasi “risuonano” sempre uguali a sé stesse fino a cancellare il dissenso e a non essere più messe in discussione», ciò che «comporta anche una ulteriore assuefazione a parole e concetti che possono assumere la forma di slogan, “parole d’ordine” pressoché desmantellate» (Negri e Tagliani 2021).

Pertanto, in una tale cornice di viralità informativa e creazione di miscredenze basate su un senso comune di frustrazione proveniente dall'incongruenza tra la diffusione di notizie legate alle misure contrastive e la difficile risoluzione concreta e definitiva dello stato pandemico (Valensise 2021), la disposizione delle opinioni pubbliche non risulta univoca e diverse voci basate su un numero elevato di notizie non attendibili circolate come *fake news* prendono forma in gruppi oppositivi sui social media.

Il movimento dell'antivaccinismo è una posizione radicata nel tempo. Se ne tiene traccia a partire dai gruppi oppositivi all'introduzione dell'obbligo vaccinale contro il vaiolo – *Anti Vaccination League* e *Anti-Compulsory Vaccination League*² – considerato come strumento mirato a ledere i diritti e le libertà individuali, nonché causa di malattie collaterali. Tra i sostenitori della formazione

¹ World Health Organization (WHO) – Infodemic, https://www.who.int/health-topics/infodemic#tab=tab_3 (ultima consultazione: 26/05/2022).

² National Library of Medicine, Digital Collections, <https://collections.nlm.nih.gov/catalog.nlm.nlmuid-101235983-bks> (ultima consultazione: 26/05/2022).

e crescita di gruppi di contrasto alle cure vaccinali spiccano anche figure legate al mondo della medicina, come ad esempio Alexander Milton Ross, medico canadese, editore di "The Anti-Vaccinator"³ e autore del pamphlet (Fig. 1) che nel 1885 ebbe larga diffusione per la sua contestazione ai benefici apportati dal vaccino scoperto da Edward Jenner e per il suo incitamento alla convinzione dell'inutilità di sottoporre la gente all'obbligo della vaccinazione.

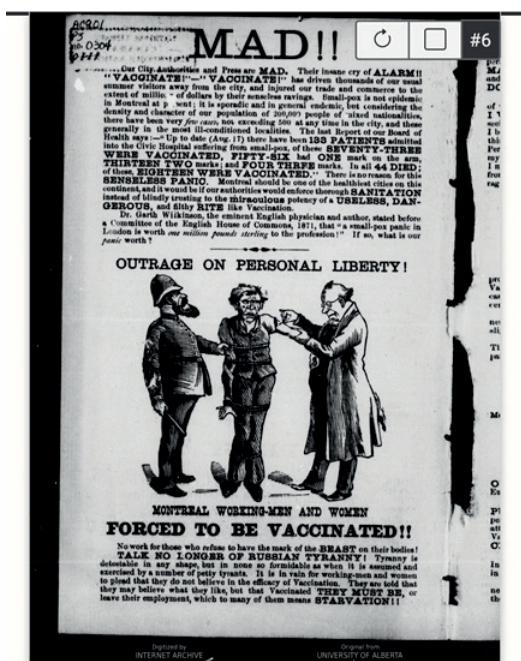


Figura 1: Pamphlet⁴.

*No-vax*⁵, *anti-vaxxers*, *antivaccinators*, *pro-plague*, *conspirazionisti* sono coloro che, come in passato accadde per le cure contro il vaiolo o morbillo, rimangono scettici nei confronti di scoperte scientifiche che hanno permesso di debellare epidemie e pandemie, basando il loro spirito critico sull'adesione irremovibile a messaggi diffusi dai gruppi oppositivi di appartenenza (Knippenberg, Lossie and Wilke 1994, 289-300).

L'esitazione del singolo e della comunità a sottoporsi alla somministrazione di cure vaccinali – per vari fattori, come ad esempio la mancanza di consapevolezza sui rischi della malattia, il timore dell'ago, l'incertezza sulla disponibilità

³ The anti-vaccinator, and advocate of cleanliness. Ross, Alexander Milton, 1832-1897.

⁴ University of Oxford, Faculty of History, "COVID-19 anti-vaxxers use the same arguments from 135 years ago", <https://www.history.ox.ac.uk/article/covid-19-anti-vaxxers-use-the-same-arguments-from-135-years-ago> (ultima consultazione: 26/05/2022).

⁵ "Chi o che si oppone alla vaccinazione obbligatoria o di massa." (Treccani, il portale del sapere).

dei vaccini – viene considerata come uno dei dieci rischi globali per la salute pubblica secondo l'OMS (Whitehead et al. 2019).

L'obiettivo del presente articolo consiste in una prima individuazione di un livello di ricorsività nella terminologia e conseguente formulazione scritta delle specifiche asserzioni da parte dei movimenti No-vax in riferimento all'attuale contesto di pandemia da COVID-19. Nel dettaglio, si presenterà un'analisi introduttiva sulle modalità espressive che caratterizzano i messaggi diffusi nei mesi di settembre e ottobre 2021 da un set di profili identificabili come contrari alla campagna vaccinale e alle conseguenti ordinanze nazionali ministeriali come l'obbligo del Green pass. L'indagine ha previsto uno studio analitico dei messaggi divulgati tramite social network (nello specifico, Twitter) attraverso la verifica delle strutture linguistiche più ricorrenti all'interno degli stessi al fine di isolare delle batterie semantiche, sotto una prospettiva di analisi delle nuove forme linguistiche sui social network (Caleffi 2015), ripetute all'interno di gruppi social che guidano alla formazione di messaggi "manifesto" identificativi dell'appartenenza a un'ideologia oppositiva. In questo modo è stato possibile avviare un'indagine preliminare per la rilevazione di sequenze di espressioni che possano rappresentare un livello di prototipicalità semantica (Knippenberg 2011) basata sull'osservazione della concatenazione di elementi linguistici.

In particolare, l'analisi che verrà presentata si concentra sulle nuove modalità comunicative impiegate dagli utenti tramite la diffusione di messaggi su social network e, nello specifico, mediante la pratica di costruzione di contenuti informativi basati su hashtag. L'hashtag, come definito dall'Accademia della Crusca, è una

parola (o sequenza continua di parole) preceduta dal simbolo #, usata nell'ambito dei social network per categorizzare e rendere ricercabili contenuti correlati.

È dal luglio 2009 che il social network Twitter ha introdotto ufficialmente gli hashtag come chiavi di ricerca per catalogare e rintracciare commenti attinenti a uno stesso argomento. Un commento con un determinato hashtag viene categorizzato insieme a tutti gli altri, scritti da utenti diversi, a cui sia stato associato il medesimo hashtag. La funzione è stata incorporata negli ultimissimi anni da altri popolari social network come Instagram, Google+ e Facebook. [...]

è una parola inglese composta da hash, uno dei tanti termini che indicano il simbolo # (in italiano detto cancelletto), e tag, 'etichetta', 'marcatore'. Coniata per la prima volta nel 2007 in un blog statunitense e poi proposta su Twitter dagli stessi utenti, la parola all'inizio indicava solamente il simbolo del cancelletto impiegato per creare chiavi di ricerca, non la chiave di ricerca stessa, e questo significato si è mantenuto nella terminologia ufficiale di Twitter: "il

simbolo #, denominato hashtag, viene utilizzato per contrassegnare parole o argomenti chiave in un Tweet. È stato concepito dagli utenti Twitter come metodo per categorizzare i messaggi”. Nell’uso comune, tuttavia, il significato si è esteso a tutta la sequenza, che nella terminologia ufficiale è chiamata *hashtagged word*.

In letteratura sono presenti molteplici studi sui processi di adattamento e nuove formazioni di significato nelle pratiche linguistiche all’interno dei nuovi contesti virtuali (Budnik 2019; Caleffi 2015; Zappavigna 2015), orientando le indagini ai meccanismi di associazione delle intenzioni comunicative agli hashtag. Ad esempio, (Caleffi 2015) individua tra gli usi e le funzioni più comuni degli hashtag i seguenti: denominazioni, espressioni di sentimenti ed emozioni personali, hashtag come mezzi comunicativi per sostenere iniziative o eventi, hashtag come veicoli comunicativi di autoironia o di promozione di brand. L’autrice analizza ancora più precisamente la combinazione dei significati a seconda della composizione delle parole contenute all’interno degli hashtag: se si tratta di una sola parola che segue il simbolo ‘#’ le possibili valutazioni sono relative o a un sostantivo singolare che non presenta un valore pragmatico o semantico rilevante, o a un aggettivo che normalmente esprime un sentimento o un commento agendo da meta-commento sull’intero messaggio (Caleffi 2015, 54); quando un hashtag è costituito dalla combinazione di due parole, generalmente questa tipologia è formata da una frase nominale di tipo attributivo (aggettivo + sostantivo) (Caleffi 2015, 55) o sostantivo (aggettivo+sostantivo) e queste espressioni vengono utilizzate per esprimere un sentimento di protesta, incoraggiamento, emozione o per condividere un commento; la struttura di un hashtag formato da tre parole, nel contesto della lingua italiana, ha un significato equivalente a espressioni esclamative, sarcastiche, di avvertimenti o emozioni, mentre gli hashtag composti da quattro parole hanno una funzione di commento o suggerimento.

(Zappavigna 2015) ha esplorato le nuove funzioni linguistiche messe a disposizione dall’uso degli hashtag: indicatori dell’*aboutness* dei testi presenti nei messaggi pubblicati sui social media, veicoli di formazione di tribù virtuali, mezzi comunicativi di costruzione di tematiche emergenti e, non meno importante, tecniche di formazione di ‘meta-discorsi’. È su quest’ultimo aspetto che l’autrice si sofferma in modo più dettagliato indicando come questa combinazione di parole sui social network precedute dal simbolo ‘#’ «support communication that reflexively points back at itself (and is ‘meta’ in this sense). Like traditional metadata, hashtags afford abstraction, enabling information at higher and more complex orders of experiences to be appended to the main content of a post». Infatti, l’hashtag ha, secondo l’autrice, una funzione sincretica di istanze conversazionali tra gli utenti: in questo senso parla di “metadati sociali” creati nel tempo dalle comunità di utenti, le quali restituiscono così un sistema di annotazione descrittiva ai propri messaggi agevolando un processo

meta-discorsivo nella condivisione delle informazioni nonché di aggregazione di contenuti per tematiche.

(Budnik, Gaputina and Boguslavskaya 2019) suddividono le funzioni degli hashtag nelle seguenti tipologie: mezzi di formazione di nuove tematiche e strumenti di navigazione, strumenti di marketing costituendosi come strutture ipertestuali che riescono a focalizzare gli elementi da promuovere ricorrendo a un vocabolario caratterizzato da espressioni superlative e divenendo, pertanto, degli slogan vincenti; strumenti di condivisione dei giudizi e delle emozioni condensando in brevi parole una visione del mondo da diffondere; veicoli di significati concettuali, indicando con ciò la possibilità di attribuire un serie di significati concettuali agli hashtag relativi a determinati eventi, persone o situazioni.

È nella direzione delle suddette funzioni associate agli hashtag che l'analisi del presente studio si orienterà. Nello specifico, verrà verificato l'uso linguistico-veicolare degli hashtag nei tweet estrapolati considerandoli come strutture ipertestuali di condivisione di stati emotivi e giudizi durante i mesi successivi alla prima campagna vaccinale di massa nel corso della pandemia da COVID-19 e i risultati di questa analisi saranno integrati dall'identificazione di espressioni tipizzate di movimenti No-vax.

2. Infodemia da COVID-19

La continua trasmissione delle notizie riferite alle misure contrastive per ridurre il livello di infettività della malattia da COVID-19 appare inarginabile, e gli spazi del web in cui vengono condensati questi dati contraddittori e incessantemente pervasivi offrono una dimensione virtuale a chi contrasta con teorie inventive le evidenze scientifiche sull'efficacia dei vaccini. Come riportano (Negri e Tagliani 2021, 45) discutendo della nuova infodemia da COVID-19:

L'effetto è evidente: non sapendo interpretare i dati, ciascuno di noi può giungere a conclusioni errate, e generare fake news. Gli epidemiologi hanno modelli matematici che, se applicati ai dati grezzi, restituiscono una proiezione dell'andamento dell'epidemia e del tasso di letalità; sono modelli che noi non possediamo e non possiamo applicare. L'interpretazione dei dati "fai da te" non è dunque attendibile, e si somma all'alea d'incertezza della proiezione elaborata dai modelli scientifici, che non può restituire un'interpretazione univoca ma solo stimata, essendo il processo ancora in atto e i dati non definitivi. In questo contesto la "infodemia" a cui, nella società postmoderna, siamo esposti, ci porta alla generazione di fake news molto verosimili, non solo perché non sono pensate per ingannare, ma perché sorgono per sovra-interpretazione dei dati e, cosa ancor più diabolica, si basano sui dati veri.

In questo contesto pandemico, seguendo le parole dell'autore, le motivazioni che spingono i gruppi No-vax a rimanere riluttanti nell'accettazione della

campagna vaccinale per combattere la circolazione del COVID-19 si riferiscono principalmente all'incapacità di interpretare correttamente una fitta serie di informazioni che circolano in modo frenetico sul web. Accanto a questo senso di disorientamento informativo, il meccanismo di disillusione innescato nei confronti degli strumenti di difesa vaccinali è altrettanto legato al fatto che questi stessi vengano considerati, come in passato, strumenti che ledono le libertà di scelta individuali (Jolley 2014). A cambiare sono gli spazi pubblici in cui chi propaganda sommosse intellettuali contro l'obbligatorietà vaccinale trova confronti perpetui, ora confinati alle piattaforme web dei social network e dei blog. Infatti, i social network rappresentano il luogo della concretizzazione comunicativa dei movimenti No-vax e di teorie cospirazioniste su *Big Pharma* e sulla dittatura sanitaria (Cossard 2020; Jolley 2014).

2.1. *Messaggi tipizzati dei gruppi anti-vaccino*

La gestione e l'acquisizione dei dati utili per gli scopi di ricerca sulla rilevazione di un livello di ricorsività linguistica dei gruppi No-vax sono state impostate su una metodologia di recupero dell'informazione basata sulla consultazione di contenuti provenienti da fonti del web. La documentazione disponibile all'interno delle piattaforme virtuali permette, infatti, una maggiore espandibilità in tempo reale del tessuto informativo oggetto di studio. Il web, seguendo le parole di (Montagnani 2007, 74) che riflette sul lavoro di Baldazzi, è connotato da un forte tratto di transdisciplinarietà e ipertestualità in cui di volta in volta «Si è andati verso una dissoluzione dei supporti, dei quadri scientifici e tassonomici di classificazione, di possibilità di contenimento in un solo luogo dell'informazione, del carattere enciclopedico del sapere, dirigendosi verso una pluralità e equivocità di quadri ontologici instabili». Ed è in questo contesto virtuale altamente multiforme che il ruolo del documentalista si è trasformato in quello che Baldazzi chiama *cybrarian* a supporto dell'estrazione di contenuti informativi da piattaforme web e della loro giusta interpretazione attraverso un processo di "metacognizione".

Per avviare l'indagine oggetto di studio è stato preso in considerazione il social network Twitter. Una maggiore fluidità del linguaggio, intendendo con ciò l'assenza di una struttura nella disposizione degli elementi lessicali nei messaggi, e la sempre crescente rapidità negli aggiornamenti dei messaggi da parte degli utenti sono alcuni dei tratti specifici di questa piattaforma (Bernard et al. 2009; Sankaranarayanan et al. 2009) che la rendono un portale in cui sono presenti contenuti eterogenei di supporto al recupero dell'informazione per poter avviare analisi diacroniche su nuovi usi terminologici. Considerando l'obiettivo di ricerca, ovvero uno studio preliminare dei processi di condivisione di espressioni comunicative dei gruppi No-vax per mezzo dei messaggi pubblicati tramite i tweet, l'estrazione di dati utili è stata effettuata per mezzo

di parole chiave di riferimento per la tematica in oggetto, hashtag. In primo luogo è stato definito il periodo per l'estrazione dei tweet. Sono stati isolati alcuni messaggi pubblicati online tra settembre e ottobre 2021, arco temporale che coincide con i mesi successivi alla prima campagna di vaccinazione, al fine di mostrare quanto il contenuto dei post sia diventato, in seguito alle varie ondate della pandemia da COVID-19 e con il passare dei mesi, ripetitivo e standardizzato, quasi a costituire sia nel mondo virtuale che reale un vocabolario di riferimento contenente delle espressioni fisse riutilizzabili per il rafforzamento dell'identificazione comunitaria No-vax nel web (Furini 2021).

Una volta delimitato il range temporale, il secondo passaggio ha previsto l'identificazione di un set di hashtag per il recupero dei tweet inerenti ai suddetti scopi di ricerca al fine di effettuare un'indagine conoscitiva dei principali usi e convenzioni semantiche legati al contenuto dei messaggi estratti. Nel dettaglio, gli hashtag utilizzati per gli scopi di questa ricerca sono stati estratti dallo studio su lavori presenti in letteratura per l'identificazione delle principali parole chiave associate ai movimenti anti-vaccino relativi alla pandemia da COVID-19 (Di Giovanni et al. 2021; Monaci 2021; Quintana et al. 2022). Attraverso ognuno di questi hashtag è stato possibile ottenere un'estrazione di un elenco di tweet corrispondenti.

3. Metodologia

La metodologia con cui si è avviata l'analisi semantica dei messaggi estrapolati dalla ricerca per hashtag sulla piattaforma Twitter, identificativi dell'appartenenza di utenti a gruppi anti-vaccino, è stata articolata in quattro fasi, così come mostrato dal seguente schema (Fig. 2):

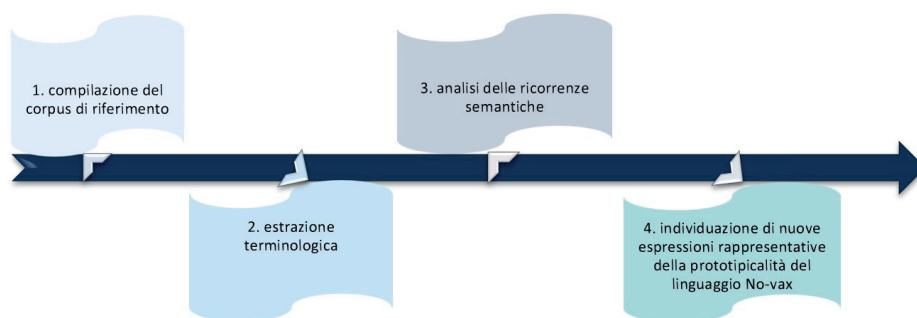


Figura 2: Fasi metodologiche.

Compilazione del corpus

Una volta stabiliti i parametri di ricerca nella prospettiva di avviare uno studio conoscitivo delle modalità espressive di determinati tipi di utenti, la prima fase dello studio si è incentrata sulla realizzazione del corpus di partenza.

Nel processo di estrazione della conoscenza, per ottenere i migliori risultati è importante comprendere al meglio il dominio applicativo di riferimento ed individuare chiaramente gli obiettivi che si intendono raggiungere. [...] Il secondo passo del processo consiste nella selezione dei dati su cui sarà attivato il processo di scoperta. (Lancia e Lapicciarella 2008, 281)

La definizione di un dominio applicativo è, infatti, una delle precondizioni necessarie per una corretta individuazione della popolazione di riferimento con cui far combaciare gli obiettivi di ricerca sulla rappresentazione della conoscenza di settore (Biber 1993).

Un corpus è una raccolta di testi o, in alcuni casi, di parti di testi che vengono scelti in base a criteri prestabiliti perché siano rappresentativi, per quanto possibile, di una lingua, di una varietà linguistica, di un genere, di un dominio di discorso, di un autore o di un argomento (Bowker e Pearson 2002) che possono essere analizzati in un particolare momento o in un certo periodo di tempo (analisi sincronica o diacronica di un corpus) (Torsello 2009, 26).

Le tecniche di compilazione dei corpora seguono alcuni punti nodali che portano alla collazione di testi di riferimento inerenti alla tematica di studio. (Pearson 1998) identifica una serie di requisiti che i testi contenuti nei corpora dovrebbero possedere per configurarsi come strutture di conoscenza aderenti ai campi di studio di cui si cerca di investigare la terminologia:

- attualità: bisognerebbe privilegiare testi che possano riflettere l'informazione corrente del settore di studio in modo da impostare un asset documentale in grado di fornire in output informazioni concettuali aggiornate;
- eterogeneità autoriale: è preferibile inserire nei corpora di partenza testi scritti da esperti provenienti da diversi settori ai fini di un potenziamento della varietà terminologica;
- originalità: i testi non dovrebbero essere traduzioni in modo da garantire l'originalità della terminologia;
- completezza: i testi non devono essere frammentari in quanto si rischierebbe di perdere un certo numero di informazioni utili ai fini di un'analisi terminologica completa.

Per (Barrière 2006) la costruzione di un corpus rappresenta il primo passaggio per la realizzazione di una base di conoscenza terminologica, i.e., *termino-*

logical knowledge bases (TKBs), da cui verranno sviluppate risorse semantiche orientate all'organizzazione e rappresentazione della conoscenza.

Nel caso specifico, il recupero dell'informazione relativa alle modalità espressive di gruppi No-vax è stato impostato sull'estrazione dei contenuti testuali presenti nella piattaforma di microblogging Twitter che ha rappresentato la fonte principale per la configurazione del corpus di partenza. Nel dettaglio, i testi presenti nel corpus sono tweet pubblicati su questa piattaforma, estratti utilizzando un accesso tramite API, inserendo come chiavi di ricerca nelle queries un gruppo di hashtag, ognuno dei quali ha generato una lista di messaggi temporalmente circoscritti ai mesi di settembre e ottobre del 2021 divenendo i documenti costitutivi del corpus di tweet da analizzare per avviare un'indagine sulla terminologia dei gruppi No-vax. In questo modo il set documentale finale è risultato composto da otto documenti che racchiudono il raggruppamento dei post relativi agli hashtag ricercati così come mostrato dal seguente schema (Tab. 1).

hashtag	n. tweet	documenti	Tot
#DittaturaSanitaria	4.625	Doc1	21604 tweet e 8 documenti
#nomask	250	Doc2	
#nogreenpassobbligatorio	11.871	Doc3	
#NoVaccinoObbligatorio	575	Doc4	
#novaccinoCOVID	34	Doc5	
#novax	3196	Doc6	
#nocavie	50	Doc7	
#noobbligovaccinale	1.013	Doc8	

Tabella 1: Numero di tweet per hashtag.

In totale, il corpus compilato sulla base di un raggruppamento di documenti per hashtag, concettualmente correlati ai principi guida dei movimenti anti-vaccino, è composto da 21.604 tweet per un totale di 81.023 token⁶.

I risultati ottenuti sono parte integrante di un'analisi terminologica realizzata attraverso il processamento automatico dei testi, in questo caso i tweet estrapolati tramite API, eseguito da software di estrazione terminologica pre-addestrati come Text To Knowledge (T2K) (Dell'Orletta et al. 2014), SketchEngine (Kilgariff 2014) e VoyantTools (Miller 2018), al fine di poter rilevare le informazioni tipizzate sugli usi terminologici più frequenti, in questo caso di contrasto alle misure di prevenzione alla propagazione dell'infezione da COVID-19. Questi tool consentono di disporre al termine della fase di

⁶ Totale delle occorrenze nel corpus di partenza calcolato con il tool di analisi dei corpora T2K.

acquisizione dei documenti e della loro analisi linguistica di una serie di misure statistiche che forniscono informazioni sul grado di varietà terminologica. In particolare, T2K riporta le seguenti informazioni in relazione al corpus in input, in questo caso costituito da otto documenti, ciascuno dei quali corrispondente al gruppo dei tweet estratti in base all'hashtag:

Documenti	Totale token	Totale di lemmi diversi	Densità Lessicale	Rapporto tipo/unità
8	81.023	7.802	0,621	0,096

Sulla base della definizione della misura Type Token Ratio (TTR):

il valore percentuale che indica la varietà di parole diverse contenute in un testo. Più alto è il valore percentuale maggiore è la varietà del vocabolario usato in un testo. Il TTR è calcolato dividendo il numero delle parole diverse usate in un testo (types) per il numero delle parole complessive di quel testo (tokens) moltiplicato per 100⁷.

si evince come il risultato ottenuto attraverso l'analisi del profilo lessicale mediante il tool T2K non raggiunga un valore percentuale alto, e questo è un dato che può confermare la ripetitività dal punto di vista di usi del linguaggio all'interno dei documenti del corpus in input, ovvero i tweet dei gruppi anti-vaccino.

Estrazione terminologica

La seconda fase è mirata all'isolamento dei termini di riferimento nel corpus di partenza ai fini di un'analisi semantica che possa far emergere delle osservazioni circa le modalità espressive dei suddetti gruppi sociali. Questa fase è successiva alla definizione del corpus per l'elaborazione testuale e si riferisce alla configurazione di un sistema di processamento terminologico all'interno dei testi presi in esame. In particolare, si è proceduto con l'estrazione dei termini rappresentativi dei testi contenuti nel corpus di partenza ai fini di portare avanti uno studio esplorativo delle sequenze espressive più ricorrenti nei messaggi No-vax. Per l'esecuzione di processi analitici terminologici ci si avvale generalmente del supporto di software semi-automatici (Loginova et al. 2012) sviluppati per l'implementazione di tecniche di analisi linguistico-statistiche che mirano a rilevare quali siano gli usi terminologici più frequenti ed emergenti e, al contempo, fare inferenza su quali siano le informazioni tipizzate entro contesti linguistici specifici.

⁷ Paisà, <https://www.corpusitaliano.it/it/help/readability.html> (ultima consultazione: 26/05/2022).

Dopo una prima valutazione di sistemi di analisi terminologica esistenti⁸ il processo di estrazione dei dati è stato eseguito attraverso i software T2K e SketchEngine, tool di analisi dei corpora utilizzati nel campo della lessicografia, il secondo dei quali più flessibile nelle fasi di inclusione di *stopwords* esterne e per la possibilità di ottenere una rappresentazione grafica delle co-occorrenze morfosintattiche dei lemmi utile ai fini delle successive valutazioni semantiche. Questo passaggio ha previsto l'applicazione di alcuni parametri statistici di partenza al fine di ottenere in output una lista di termini rappresentativi che potessero garantire un buon livello quantitativo di copertura semantica rispetto alle informazioni contenute nei post oggetto di studio. Ai fini di una corretta esecuzione di analisi testuale è stata effettuata una pulitura preliminare dei documenti attraverso la rimozione di alcuni dati non essenziali come simboli (più precisamente, emoticon) e riferimenti ipertestuali (Gupta and Rattikorn 2018, 341) e un elenco di *stopwords* non funzionali ai processi di recupero delle informazioni di contesto. Il numero di unità linguistiche singole estratte è stato pari a 11.988, mentre il numero di unità linguistiche composte pari a 15.840.

Nella Tabella 2 si riportano i primi risultati ottenuti in seguito all'analisi linguistico-statistica del corpus di partenza dal punto di vista dell'individuazione dei termini più rappresentativi, eseguita dal software T2K applicando una misurazione di contrasto con un secondo corpus al fine di individuare il livello di specificità terminologica. Il contrasto è stato effettuato attraverso l'attivazione di un confronto tra il corpus di partenza costituito dai tweet estratti e un corpus in lingua italiana giornalistico inserito come documentazione direttamente all'interno del tool T2K (Bonin et al. 2010). L'approccio contrastivo descritto da (Bonin et al. 2012, 209) prevede l'identificazione di unità terminologiche rilevanti ai fini di un'indagine conoscitiva dell'informazione specialistica in corpora di dominio attraverso un confronto tra raggruppamenti documentali che presentano caratteristiche differenti nella distribuzione delle unità terminologiche rappresentative:

L'estrazione di unità terminologiche monorematiche e polirematiche è cioè condotta a partire dal confronto della distribuzione delle unità terminologiche monorematiche e polirematiche nel corpus di acquisizione rispetto a un corpus di riferimento (detto anche 'corpus di contrasto'). In questo modo, la lista finale di unità terminologiche estratte conterrà quelle unità che sono maggiormente rilevanti nel corpus di acquisizione rispetto (ovvero 'per contrasto') al corpus di riferimento. A questo scopo sono state sviluppate una serie di metodologie in grado di computare la misura della diversa rilevanza di unità terminologiche all'interno dei due corpora che vengono confrontati. La possibilità di discriminare termini e non-termini è così empiricamente realizzata sulla base di un'analisi 'contrastiva' della loro distribuzione in un corpus di

⁸ Si vedano i tool Sketch Engine, Text to Knowledge (T2K), Voyant Tools, Wordsmith.

dominio (il corpus di acquisizione) rispetto a un corpus rappresentativo della lingua comune (usato come 'corpus di contrasto').

estrazione terminologica		applicazione corpus di contrasto	
termini	frequenza	termini	frequenza
#NoGreenPassObbligatorio	308	#NoGreenPassObbligatorio	308
cose	272	#quartarepubblica	176
vaccino	213	#greenpass	129
massa	184	conto corrente bancario	174
conto	182	#novax	63
soldi	181	#novaccinoobbligatorio	55
#quartarepubblica	176	#vaccino	51
conto corrente bancario	174	#dittatura	45
#greenpass	129	#vaccini	43
persone	110	Vaccinati	40
libertà	102	RT	37
obbligo	89	Covid	36
dose	78	#nogreenpass	28
anni	76	#noobbligovaccinale	27
governo	75	#nogreenpassobbligatorio	25
diritto	74	#BASTADITTATURA	25
pass	64	#libertà	22
#novax	63	Pass	64
emergenza	57	#BASTADITTATURA	19
scuola	57	#DITTATURASANITARIA	19
parte	55	Pandemia	19
#novaccinoobbligatorio	55	#NoGreenPassObbligatorio	18
#vaccino	51	#NoVaccinoObbligatorio	18
vita	48	Grazie	30
volta	48	Cure	17
tampone	47	reazioni avverse	15
giorno	47	#nessunacorrelazione	14
#dittatura	45	Greenpass	31
paese	45	Mascherina	13
#vaccini	43	dittatura sanitaria	13
vaccinati	40	#Dittatura	13
lavoro	40	#Draghi	13
		#draghi	13
		@borghi_claudio	13

figli	40	#greenpass.	13
diritto costituzionale	24	#Fuoridalcoro	13
RT	37	Vaccino	213
via	37	#NoGreenPassObbligatorio	12
		#noobbligovaccinale	
virus	37	#NoGreenPassObbligatorio	12
		#NoObbligoVaccinale	
legge	36	#drittoeroveschio	12
covid	36	#governodeipeggiori	12
rischio	36	#libertadiscegliere	12
casa	35	#dimartedi	12
decessi	34	#DittaturaSanitaria	12
tempo	34	#mattarelladimettiti	12
lavoratori	34	@matteosalvinimi	12
problema	33	Contagi	25
vaccinazione	32	immunità di gregge	11
verità	32	#staseraitalia	11
paura	32	#terzadose	11
		#novax	
mascherina	31	#NoGreenPassObbligatorio	10
punto	31	#nogreepass	10
caso	31	#vaccinoobbligatorio	10
#BASTADITTATURA	19	Mail	10
#DITTATURASANITARIA			

Tabella 2. Estrazione terminologica dai tweet elaborata con T2K e contrasto con un corpus generale.

Da questa tabella si possono evincere alcune osservazioni. In primo luogo, dal momento che il corpus di partenza risulta costituito da otto documenti contenenti i tweet estratti tramite ricerca per hashtag, i termini che sono considerati più rappresentativi applicando la misura di contrasto sono gli hashtag utilizzati dagli utenti per esprimere le proprie posizioni su Twitter. Come già osservato nei precedenti paragrafi, le funzioni che possiedono gli hashtag nei contesti comunicativi dei social network sono riferite soprattutto alla possibilità di veicolare tramite meta-discorsi pratiche discorsive degli utenti. Infatti, in rosso si può notare il cambiamento di posizione dei termini a seconda della lista di appartenenza e del livello di specificità acquisito nel momento in cui si confronta la terminologia raccolta nel corpus di riferimento con uno generale. e.g., come *greenpass*, *RT*, *novaccinoobbligatorio*, viceversa, termini come *mascherine* e *reazioni avverse* risultano rappresentativi in un corpus italiano

generale rispetto alla frequenza osservata nel raggruppamento documentale specifico della terminologia No-vax su Twitter, ma sono termini che vengono riportati in posizioni superiori per la rilevanza che assumono nel corpus di partenza.

Questa prima fase di estrazione terminologica risulta utile ai fini di un recupero di espressioni che presentano un alto grado di specificità rispetto alle tematiche trattate in un contesto d'uso oggetto di studio e che possono essere prese in considerazione per ulteriori indagini semantiche mirate alla scoperta di strutture linguistiche prefissate ricorsive di questi gruppi oppositivi.

Analisi delle ricorrenze semantiche

Gli elementi informativi del linguaggio sono inseriti in

sistemi di relazioni di ordine lineare, di incassamento gerarchico tra sintagmi e di dipendenze a lungo raggio che nel loro insieme vengono a comporre la struttura sintattica delle frasi. Queste ultime si articolano infine in complesse strutture retoriche che formano la tessitura del discorso. La struttura linguistica è la porta di accesso al contenuto del testo (Lenci, Montemagni e Pirrelli 2005, 241).

I testi che hanno costituito il gruppo documentale di partenza presentavano al loro interno molteplici informazioni linguistiche che hanno fatto emergere dati interessanti per l'avvio della terza fase metodologica. L'analisi dei termini rappresentativi contenuti all'interno del complesso documentale dei tweet ha permesso di mettere in evidenza una struttura di concatenazioni semantiche prototipiche per la rappresentazione concettuale dell'ideologia dissidente dei gruppi No-vax. Infatti, all'interno dei messaggi analizzati come testo non-strutturato, ad esempio, la maggior parte delle volte il termine *vaccino* compare associato a verbi e aggettivi negativi, quali 'inutile', 'pseudo', 'mortifero', 'morire', 'funzionicchiano', 'alterare', come si può osservare dalle seguenti figure (Fig. 3, Fig. 4, Fig. 5).

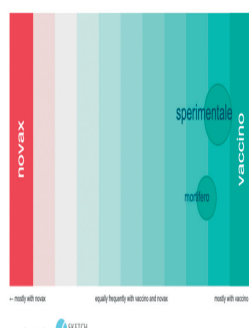


Figura 3: Aggettivi correlati al termine "vaccino".

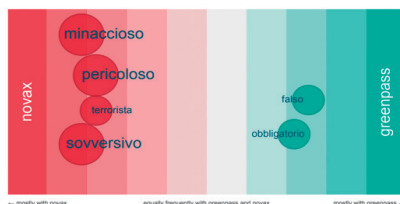


Figura 6: Confronto tra le co-occorrenze dei termini “novax” e “greenpass”.

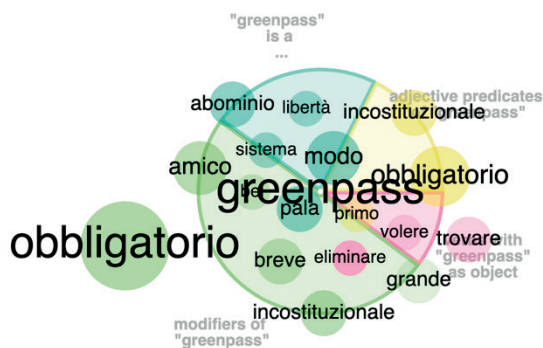


Figura 7: Concatenazioni semantiche riferite al termine “green pass”.

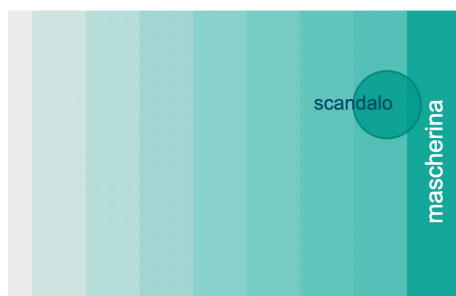


Figura 8: Esempio di relazione associativa per il termine “mascherina”.



Figura 9: Rete semantica per il termine “dittatura sanitaria”.

Identificazione di nuove catene semantiche del linguaggio No-vax

Twitter è una piattaforma di microblogging (Qui et al. 2012, 710-18) grazie alla quale si riescono a seguire in modo costante processi evolutivi dell'uso del linguaggio da parte degli utenti attraverso l'osservazione di nuove formazioni terminologiche in hashtag (Mehmood 2013). Questi ultimi sono caratterizzati dal fatto di non essere dipendenti da un contesto specifico, di contenere una serie di slang e abbreviazioni e espressioni non standardizzate (Gutba 2018). Per la rilevazione di ulteriori sequenze semantiche attribuibili a un uso ricorsivo e stereotipato da parte di utenti Twitter No-vax l'ultima fase del presente studio ha riguardato l'isolamento di tutti i termini preceduti dal simbolo '#', che tipicamente precede la presenza di hashtag sui social network, all'interno della lista terminologica ottenuta in seguito al processamento dei testi contenuti nel corpus di partenza. Da questi è stato possibile identificare nuovi trend semantici (Lau, Collier and Baldwin 2012) associabili alle modalità espressive dei gruppi No-vax.

Pertanto, attraverso l'individuazione delle entità hashtag all'interno dei risultati dell'estrazione terminologica, rappresentativa del corpus documentale dei tweet presi in esame, è stato possibile verificare la presenza di associazioni di elementi semantici ricorsivi all'interno dei messaggi pubblicati da profili Twitter contrari alle cure vaccinali nel contesto della pandemia da COVID-19.

Di seguito viene presentata una lista esemplificativa delle co-occorrenze più frequenti rintracciate nella lista dei termini riferiti al contenuto dei tweet dei gruppi No-vax oggetto di studio:

#fascisti

- #fascisti #nazisti

#guerrieri

- #guerrieri #novax

#nessunacorrelazione

- #nessunacorrelazione #vaccinibrodaglia

#provax

- #provax contro #novax

#terroristi

- #terroristi #vaccino

#terzadose

- #terzadose di #Terrorismo

#vaccino

- #vaccino #giocopubblico
- #vaccino #novax #complotto

#vaccinoobbligatorio

- #depopolazione

#nogreenpassobbligatorio

- #governodeipeggiori
- #NoVaccinePassport

#pandemia

- #falsapandemia #neonazisti
- #variantisenzafine

Si può evincere un uso terminologico identificativo dell'appartenenza ideologica di profili No-vax in cui si ricorre spesso a formule stereotipate negative nei confronti del controllo della situazione pandemica attuale considerando le soluzioni adottate come 'complotto' o forme di 'terrorismo'.

4. Conclusioni

Dai risultati ottenuti attraverso l'analisi del corpus di partenza costituito dal raggruppamento di tweet estratti grazie all'uso di determinati hashtag in riferimento ai movimenti antivaccinismo, è stato possibile osservare l'uso di nuove modalità espressive adottate dai profili iscritti a piattaforme sociali per esprimere proprie posizioni. Il processo di costruzione fraseologica, caratterizzata dall'uso di termini specifici al contesto pandemico di riferimento, entro i confini virtuali trova nell'utilizzo massivo degli hashtag, come strutture meta discorsive, la configurazione di nuovi modelli espressivi. In particolare, in questo studio preliminare sono state analizzate una serie di concatenazioni espressive estrapolate dal corpus di partenza e isolate tramite l'applicazione di misure statistiche in seguito al processamento terminologico, molte volte caratterizzate dalla presenza di hashtag come strumenti di supporto alla rappresentazione sintetica di concetti più complessi diffusi da gruppi No-vax all'interno della piattaforma di microblogging Twitter. L'obiettivo è stato basato sull'intenzione di isolare un set di messaggi ricorrenti utilizzati per manifestare una forma di appartenenza ideologica e atti a diventare un vocabolario di espressioni linguistiche tipizzate dei suddetti gruppi sociali. Infatti, automatizzando dei processi di identificazione di associazioni semantiche su un gruppo di documenti rappresentativi di tematiche specifiche è possibile osservare e isolare delle costruzioni linguistiche ricorsive. Lo studio della ricorsività di strutture morfo-sintattiche presenti all'interno dei tweet oggetto di questo studio ha permesso di verificare come l'uso congiunto di alcuni termini che rientrano nell'uso del linguaggio proprio di un gruppo di utenti specifico abbia supportato la definizione di espressioni regolarmente usate nel linguaggio comune delle comunità virtuali No-vax per esprimere determinati concetti e opinioni. Queste ultime divengono nel corso del tempo delle espressioni manifesto prefabbricate da utilizzare in modo da rispecchiare un'ideologia di appartenenza. Nello

specifico, costituiscono strutture di mediazione e interpretazione semantica orientate all'individuazione di informazioni emblematiche e alla loro integrazione con schemi di pensiero omogenei. Il linguaggio dei social network offre la possibilità di seguire a ritmi costanti l'inserimento di nuove parole che nel tempo formano veicoli espressivi di contenuti semantici per la descrizione di nuove realtà come quella pandemica, diventando altamente emblematiche di determinati contesti descrittivi. Appare pertanto utile disporre di analisi e metodologie terminologiche come quelle qui presentate che possano guidare alla scoperta di espressioni linguistiche contrassegnate da valori iterativi per l'assimilazione di trend costitutivi di fenomeni storici.

Riferimenti bibliografici

- Baldazzi, Anna. 2004. "Documentazione: saperi e professioni in evoluzione." *Rassegna bibliografica infanzia e Adolescenza* 5, no. 2: 5–25.
- Barrière, Caroline. 2006. "Semi-automatic corpus construction from informative texts." In *Lexicography, Terminology and Translation: Text-Based Studies in honour of Ingrid Meyer*, edited by Lynne Bowker. University of Ottawa Press.
- Biber, Douglas. 1993. "Representativeness in Corpus Design." In *Literary and Linguistic Computing* 8, no. 4: 243–57.
- Bonin, Francesca, Felice Dell'Orletta, Giulia Venturi, and Simonetta Montemagni et al. 2010. "A contrastive approach to multi-word term extraction from domain corpora." In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*.
- Bonin, Francesca, Felice Dell'Orletta, Giulia Venturi, and Simonetta Montemagni. 2012. "Lessico settoriale e lessico comune dell'estrazione di terminologia specialistica da corpora di dominio." In *XLIV congresso internazionale di studi della società di linguistica italiana*, 207–20, Viterbo, 27–29 settembre 2010.
- Budnik Ekaterina, Violetta Gaputina, and Vera Boguslavskaya. 2019. "Dynamic of hashtag functions development in new media: Hashtag as an identificational mark of digital communication in social networks." In *Proceedings of the XI International Scientific Conference Communicative Strategies of the Information Society (CSIS'2019)*. Association for Computing Machinery, New York, NY, USA, Article 28, 1–5. <https://doi.org/10.1145/3373722.3373795>.
- Caleffi, Paola-Maria. 2015. "The 'hashtag': A new word or a new rule?" *SKASE journal of theoretical linguistics* 12, no. 2.

- Cossard, Alessandro, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. 2020. "Falling into the Echo Chamber: The Italian Vaccination Debate on Twitter". In *Proceedings of the International AAAI Conference on Web and Social Media* 14, no. 1: 130–40. <https://ojs.aaai.org/index.php/ICWSM/article/view/7285>.
- Dell'Orletta, Felice, Giulia Venturi, Andrea Cimino, and Simonetta Montemagni. 2014. "T2K: A system for automatically extracting and organizing knowledge from texts." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Di Giovanni, Marco, Lorenzo Corti, Silvio Pavanetto, Francesco Pierri, Andrea Tocchetti, and Marco Brambilla. 2021. "A Content-based Approach for the Analysis and Classification of Vaccine-related Stances on Twitter: The Italian Scenario." In *Information Credibility and Alternative Realities in Troubled Democracies @ ICWSM 2021*, 1–6.
- Furini, Marco and Gabriele Menegoni. 2018. "Public Health and Social Media: Language Analysis of Vaccine Conversations." In *2018 International Workshop on Social Sensing (SocialSens)*, 50–55. 10.1109/SocialSens.2018.00022.
- Furini, Marco. 2021. "Identifying the features of ProVax and NoVax groups from social media conversations", *Computers in Human Behavior*, 120, 106751. <https://doi.org/10.1016/j.chb.2021.106751>.
- Gupta, Vibhuti, and Hewett Rattikorn. 2020. "Real-Time Tweet Analytics Using Hybrid Hashtags on Twitter Big Data Streams." *Information*, 11(7), 341. <https://doi.org/10.3390/info11070341>.
- Jansen, Bernard J., Mimi Zhang, Kate Sobel, and Abdur Chowdury. 2009. "Twitter power: Tweets as electronic word of mouth." *Journal of the American Society for Information Science and Technology* 60, no. 11: 2169–88.
- Kalima, Juha, and Mikael Matti. 2013. "The effect of topic on word formation and the frequency of neologism use on Internet forums." Master's Thesis, University of Jyväskylä.
- Kilgarrieff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. "The Sketch Engine: ten years on." In *Lexicography ASIALEX* 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- Knippenberg, Daan van. 2011. "Embodying who we are: Leader group prototypicality and leadership effectiveness." *The Leadership Quarterly* 22, no. 6: 1078–91. <https://doi.org/10.1016/j.leaqua.2011.09.004>.

- Knippenberg, Daan van, Nathalie Lossie and Henk A. M. Wilke. 1994. "In-group prototypicality and persuasion: Determinants of heuristic and systematic message processing." *British Journal of Social Psychology* 33, n. 3 (settembre): 289-300.
- Lancia, Maurizio, e Andrea Lapicciarella. 2008. "Data Mining e Text Mining". In *Dal Documento all'Informazione*, a cura di Roberto Guarasci, Milano: ITER Srl.
- Lau Jey Han, Nigel Collier, and Timothy Baldwin. 2012. "On-line Trend Analysis with Topic Models: #twitter Trends Detection Topic Model On-line." In *Proceedings of COLING 2012*, 1519-34, Mumbai, India. The COLING 2012 Organizing Committee.
- Lenci, Alessandro, Simonetta Montemagni, e Vito Pirrelli. 2005. *Testo e computer. Elementi di linguistica computazionale*. Roma: Carrocci Editore.
- Loginova Clouet E., Elizaveta, Anita Gojun, Helena Blancafort, Marie Guegan, Tatiana Gornostay, and Ulrich Heid. 2012. "Reference Lists for the Evaluation of Term Extraction Tools." In *Terminology and Knowledge Engineering Conference (TKE)*, Madrid, Spain.
- Mehmood, Rizwan Hermann Maurer, and Muhammad Tanveer Afzal. 2013. "Knowledge discovery in hashtags#." In 2013 *IEEE 9th International Conference on Emerging Technologies (ICET)*, 1-6, doi: 10.1109/ICET.2013.6743538.
- Miller, Alissa. 2018. "Text mining digital humanities projects: Assessing content analysis capabilities of voyant tools." *Journal of Web Librarianship* 12(3):169-197.
- Monaci, Sara. 2021. "The Pandemic of Conspiracies in the COVID-19 Age: How Twitter Reinforces Online Infodemic." *Online Journal of Communication and Media Technologies*, 11(4): e202120.
- Montagnani, Daniele. 2007. "Anna Baldazzi Documentazione: Saperi e professioni in evoluzione." *Biblioteche oggi* 3: 74-7.
- Negri, Antonella, e Roberto Tagliani, a cura di. 2021. *Fictio, falso, fake: sul buon uso della filologia*. Milano: Ledizioni Ledipublishing.
- Pearson, Jennifer. 1998. "Terms in context." *Studies in Corpus Linguistics* 1, no. 12: 246. Dopo Pearson inserire:
- Pedroni, Marco. 2020. "Narrazioni virali. Decostruire (e ricostruire) il racconto dell'emergenza coronavirus." *Mediascapes Journal* 15: 24-43.

Mettere il nome della rivista in corsivo

- Pedroni, Marco. 2020. "Narrazioni virali. Decostruire (e ricostruire) il racconto dell'emergenza coronavirus." *Mediascapes Journal* 15: 24-43.
- Quintana, Ignacio Ojea, Ignacio Ojea, Marc Cheong, Mark Alfano, Ritsart Reimann, Colin Klein. 2022. "Automated clustering of COVID-19 anti-vaccine discourse on Twitter." arXiv <https://doi.org/10.48550/arXiv.2203.01549>.
- Qiu, Lin, Han Lin, Jonathan E. Ramsay, and Fang Yang. 2012. "You are what you tweet: Personality expression and perception on Twitter." *Journal of Research in Personality* 46, n. 6 (december): 710-18. <https://doi.org/10.1016/j.jrp.2012.08.008>.
- Rovetta, Alessandro, e Lucia Castaldo. 2022. *Siamo sicuri di aver compreso appieno cosa sia un'infodemia? Una prospettiva globale sui problemi infodemiologici*. 10.6084/m9.figshare.16863439.
- Sankaranarayanan, Jagan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. 2009. "Twitterstand: News in tweets." In the *17th ACM SIGSPATIAL International Conference*. New York, New York, USA: Association for Computing Machinery Press. <https://doi.org/10.1145/1653771.1653781>.
- Torsello, Carol Taylor, Katherine Ackerley, Erik Castello, Fiona Dalziel, Sara Gesuato, Francesca Helm, e Lisa Lena Opas-Hänninen. 2009. "Lavorare con un corpus, in EXPLICS – (EXPL)ointing (I)nternet (C)ase Studies and (S)imulation Templates for Language Teaching and Learning." In *L'uso di studi di un caso e di simulazioni su internet per l'insegnamento e l'apprendimento delle lingue*, a cura di Johann Fischer Maria, Teresa Musacchio, Alison Standring, 25-34.
- Valensise, Carlo M., et al. 2021. "Lack of evidence for correlation between COVID-19 infodemic and vaccine acceptance." arXiv preprint arXiv:2107.07946.
- Whitehead, Martin, Niall Taylor, Alex Gough, Danny Chambers, Mike Jessop, and Phil Hyde. 2019. "The anti-vax phenomenon." *The Veterinary record*. 184, no. 24 (jun): 744. <https://doi.org/10.1136/vr.l4027>.
- Zappavigna, Michele. 2015. "Searchable talk: The linguistic functions of hashtags." *Social Semiotics* 25, no. 3: 274-91.

La posta elettronica negli archivi di persona: conservazione e accesso^{*}

ANNA ROVELLA^{**}

ABSTRACT: The increasingly widespread use of email both as tool for personal communication and work, requires answers about the access to and the digital preservation of these kind of digital objects. This paper aims to analyze some critical aspects related to the long-term digital preservation of email archives in personal papers.

Keywords: Email, Digital preservation, Personal records, Lifecycle emails, Knowledge extraction.

1. Introduzione

Gli archivi di persona, la loro gestione e conservazione rappresentano un tema articolato e complesso che da diversi anni anima la discussione negli ambienti archivistici. La caratterizzazione che tali complessi documentari assumono nel corso delle loro formazione sedimentazione e trasmissione unitamente all'elevata possibilità di smembramento, di eccesso di selezione e di perdita di materiale incoraggiano la ricerca incessante di metodologie archivistiche adeguate¹. Il quadro di precarietà degli archivi di persona, si complica ancor più per gli archivi prodotti nel Ventunesimo secolo a causa delle criticità connesse

^{*} Il presente lavoro è la rielaborazione di un intervento tenuto nel "Convegno Personal Digital Memories: i fondi di persona dall'analogico al digitale", Ravenna 14 e 15 dicembre 2021.

^{**} Dipartimento di Culture, Educazione e Società, Università della Calabria, Rende (CS), Italia. anna.rovella@unical.it.

¹ Antonio Romiti (2011, 16) afferma «nel periodo postunitario, si intensificarono le vicende disgregative, senza dubbio sempre più rapide, che passarono dalle episodicità a procedure di sistematicità» (Guarasci e Pasceri 2011). Secondo Romiti (2011, 14) lo smembramento della documentazione privata «ha rappresentato una preoccupazione da parte del legislatore italiano il quale, così come aveva già fatto nel testo unico dei Beni Culturali del 1999, si è preoccupato di proteggere il materiale anche nel Codice dei Beni culturali del 2004» (Guarasci e Pasceri 2011). Per una rassegna di archivi di persona e degli istituti di conservazione si consiglia la lettura di Barrera (2006, 617-57).

al crescente utilizzo di documenti nativi informatici. La posta elettronica, ad esempio, testimonia come la transizione dall'analogico al digitale possa realizzarsi rapidamente allorquando entrino in gioco tecnologie ubique e pervasive. Le lettere, i carteggi, forse più che altre forme di documenti consentono di rinvenire le relazioni che il soggetto produttore ha nel tempo intrattenuto con il mondo esterno. Esse costituiscono uno straordinario strumento per collocare il personaggio nel suo contesto, per comprenderne il rilievo politico, sociale, istituzionale e culturale, per delineare una sorta di via d'accesso privilegiata al pensiero e all'indole del produttore, utile a svelare lati della personalità e del carattere ma anche a fornire una chiave di lettura di aspetti organizzativi e/o contenutistici dell'archivio. Se la corrispondenza ha un forte impatto sulla memoria del soggetto produttore e presenta un notevole potenziale informativo giova ricordare che essa è anche la più fragile delle componenti di un archivio di persona. Basti pensare alle diverse implicazioni che spingono, i produttori o gli eredi (biologici, culturali ecc.) ad operare, ancor prima del versamento, forme di selezione, di controllo e di frammentazione della documentazione da sottoporre a conservazione (Navarrini 2005). A volte tali scelte sono dettate dalla volontà di tutela delle vicende personali del soggetto produttore, non di rado strettamente intrecciate anche con quelle di altri membri della famiglia o a soggetti terzi, oppure vi è l'esigenza di una trasmissione selettiva della memoria finalizzata, ad esempio, alla diffusione di una determinata immagine del produttore. In questi casi la scelta deliberata di non lasciare traccia di documenti connessi con quanto viene a vario titolo considerato molto riservato, comporta un depauperamento dell'archivio e della trasmissione di conoscenza. Si tratta di interventi che possono giungere alla distruzione parziale o totale del complesso documentario a causa di una scarsa sensibilità verso la conservazione del patrimonio o di una mancata consapevolezza della fragilità dei documenti, come avviene, ad esempio, nel caso dei messaggi di posta elettronica che in assenza di idonei processi preservativi possono andare inesorabilmente perduti.

Su un piano molto generale si può affermare che la trasmissione del corpus delle email di un archivio di persona è una procedura complessa e problematica per la presenza di molteplici fattori che la letteratura ha, in parte, evidenziato a più riprese. Così malgrado la conservazione della posta elettronica rappresenti un aspetto molto dibattuto si può affermare che soluzioni definitive e sistematiche sono ancora certamente lontane da realizzarsi.

L'obiettivo del presente contributo è quello di apporre alcune sintetiche note a margine di un tema vasto e certamente complesso con la finalità di focalizzare elementi significativi per il processo di conservazione.

2. Considerazioni sulla natura delle email

A poco meno di un decennio dall'avvio dei primi processi organici di conservazione dei documenti informatici è possibile constatare che pratiche consapevoli e coerenti di creazione e di gestione dei documenti operate durante la prima parte del loro ciclo di vita condizionano positivamente la qualità del materiale conservato e la possibilità di utilizzarlo nel tempo, salvaguardando quelle garanzie di completezza e di evidenza probatoria che normalmente si richiedono alle fonti archivistiche. Se le politiche, le regole e i sistemi di gestione documentale rappresentano un assunto fondamentale per gli archivi correnti delle pubbliche amministrazioni e delle imprese – pur nella consapevolezza delle numerose difficoltà di un'applicazione compiuta e capillare in tal senso – nel caso della documentazione prodotta da persone fisiche le possibilità di intervento *ex ante* si riducono a sporadici esempi non esaustivamente rappresentativi di un universo molto composito. Una siffatta situazione di partenza pone un primo importante pregiudizio in tema di conservazione dei messaggi di posta elettronica negli archivi di persona. In molti casi si arriva addirittura a non aver percezione alcuna dell'esistenza delle email e della necessità di tutela di tale tipo di documentazione. Non capita poi tanto raramente che gli archivi di persona contemporanei versati a istituzioni culturali di vario genere, non presentino corpus di email al loro interno².

Stessa sorte tocca alla messaggistica istantanea e, non di rado, anche alle fotografie e ai video presenti nei dispositivi mobile (Allegrezza 2022).. Altre volte il deposito di tali documenti è reso impossibile dalla perdita accidentale dei dati che il produttore o gli eredi hanno subito per cause legate alla tecnologia, alla sicurezza, o più semplicemente ad un fattore culturale. Così una parte importante dei documenti di personalità che hanno avuto un ruolo nella società del Ventunesimo secolo viene perduta definitivamente senza che produttori, donatori o archivisti ne accusino neppure il danno.

Nel nostro Paese la letteratura archivistica in materia di conservazione di email nei fondi di persona è ancora molto esigua a testimonianza della scarsa percezione generale del problema (Allegrezza e Gorgolini 2016; Carassi 2018; Weston et al. 2019; AIB 2019). Per meglio comprendere le ragioni della poca attenzione archivistica verso le email può essere utile, in primis, formulare alcune considerazioni sulla natura e sul ciclo di vita di questo particolare oggetto digitale.

In inglese come in italiano l'uso corrente metonimico del termine email può indicare il vettore o il messaggio e talvolta anche l'account. La polisemia di significato viene evidenziata da Richard Pearce-Moses, in *A Glossary of Archival & Records Terminology*, che definisce le email (electronic mail) come «1. An asynchronous message, especially one following the RFC 2822 or MIME

² Stessa sorte tocca alla messaggistica istantanea e, non di rado, anche alle fotografie e ai video presenti nei dispositivi mobile (Allegrezza 2022).

standards, including a header, body, and - optionally - attachments, sent via a computer network held in online accounts to be read or downloaded by the recipients. 2. A system for transmitting email» (Pearce-Moses 2005). Anche per Tullio De Mauro la e-mail è «posta elettronica o messaggio inviato tramite posta elettronica» (De Mauro 2016). Solo apparentemente, sotto il profilo archivistico, il messaggio rappresenta l'oggetto da gestire, conservare e rendere fruibile nel tempo. Le implicazioni probatorie³ e informative dell'account, della relazione tra account e messaggio, dell'eventuale documento allegato, delle informazioni di processo di trasmissione, ecc. costituiscono elementi fondamentali la cui indubbia ricaduta sull'affidabilità del messaggio obbliga anche alla scelta di specifiche strategie conservative. La posta elettronica al pari di altri oggetti digitali complessi (ad esempio i database o le piattaforme di gestione documentale (Landino e Marzotti 2018)) rimanda a sfide di digital preservation metodologicamente e tecnologicamente più articolate rispetto alla semplice conservazione di file PDF e metadati descrittivi⁴. In tali sistemi è necessario conservare gli oggetti, le relazioni, le informazioni relative ai flussi e ai processi⁵. La completezza informativa che il digitale pone come condizione necessaria al recupero e all'uso successivo dei documenti richiede un approccio di conservazione di archivi di posta elettronica, di account, piuttosto che di soli messaggi. Ecco allora una prima importante criticità ma anche una sfida verso una nuova pratica documentale che obbliga archivisti e conservatori di

³ Sul valore probatorio delle email l'art. 2712 Codice Civile recita «Le riproduzioni fotografiche, informatiche o cinematografiche, le registrazioni fonografiche e, in genere, ogni altra rappresentazione meccanica di fatti e di cose formano piena prova dei fatti e delle cose rappresentate, se colui contro il quale sono prodotte non ne disconosce la conformità ai fatti o alle cose medesime». A precisazione degli elementi stabiliti nel citato articolo del Codice Civile la I sez. civile della Corte di Cassazione con l'ordinanza n. 19155 del 17 luglio 2019 afferma: «il messaggio di posta elettronica (cd. e-mail) costituisce un documento elettronico che contiene la rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti che, seppure privo di firma, rientra tra le riproduzioni informatiche e le rappresentazioni meccaniche di cui all'art. 2712 c.c. e, pertanto, forma piena prova dei fatti e delle cose rappresentate se colui contro il quale viene prodotto non ne disconosce la conformità ai fatti o alle cose medesime. ... 1) l'e-mail rappresenta senza dubbio un documento informatico, nell'accezione fornita dall'articolo, 1, comma 1, lettera b), del D.P.R. 445/2000, a mente del quale per documento informatico si intende la rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti; 2) in particolare, l'e-mail costituisce documento informatico sottoscritto con firma elettronica "leggera", "in quanto il mittente, per poter creare ed inviare detta mail, deve eseguire un'operazione di validazione inserendo il proprio username e la propria password"; 3) tale processo di validazione è da considerare equivalente alla firma elettronica leggera, così come definita in precedenza».

⁴ «Therefore, any programmatic attempt to preserve email must begin not only with an understanding of the specific technologies used in an email ecosystem, but a detailed knowledge of how server administrators and end users have configured software and hardware» (Prom 2011, 10).

⁵ Sulla natura delle email e la complessità ad esse associata (Pontevolpe e Salza 2009).

gitali italiani ad una rilettura del modello OAIS (ISO 2012), dei pacchetti informativi (di versamento e di archiviazione e di distribuzione), dello standard UNI SInCRO (UNI 2020), degli schemi nazionali di metadati e soprattutto delle strategie di interoperabilità degli oggetti della conservazione⁶.

Un altro elemento di difficoltà nella conservazione di archivi di posta elettronica è rappresentato dai diversi account che un soggetto produttore può possedere ed utilizzare. Si va dalla posta ordinaria a quella certificata⁷, quest'ultima largamente utilizzata, ad esempio, dai professionisti. L'urgenza della conservazione della PEC è certamente più sentita visto il carattere di recapito certificato e di opponibilità a terzi che la contraddistingue. Non sempre gli eredi che versano i documenti sono al corrente dei diversi account o hanno possibilità di accedervi o decidono di versare tutti gli account di posta del produttore ad un unico soggetto conservatore. Se lo scenario di frammentazione degli account di posta elettronica di un soggetto produttore costituisce, ai fini della conservazione, un potenziale ulteriore rischio di perdita di documenti e di difficoltà di visione di insieme è anche vero che gli archivi di email di personalità stanno assumendo progressivamente una loro dimensione convergente con un ruolo informativo anche all'interno di raccolte di altri oggetti culturali. Gli archivi, i documenti non sono avulsi dal tempo in cui si producono, dal loro contesto e inevitabilmente gli archivi contemporanei vivono una forte sollecitazione verso l'interconnessione, anelano ad esprimersi in una rete della conoscenza che per esplicitarsi deve necessariamente trovare il suo fondamento nella rappresentazione coerente degli oggetti e nell'interoperabilità di metadati e informazioni. In tale ottica il digitale manifesta le sue potenzialità, la sua capacità di restituire una nuova ricchezza conoscitiva multidimensionale, che i sistemi interculturali (musei, archivi e biblioteche) provano a teorizzare con risultati applicativi ancora non del tutto incoraggianti (Valacchi 2015, 54-7, 237-41; Carassi 2018, 11). Fin quando la conservazione digitale resterà, in Ita-

⁶ Ad oggi la conservazione di interi archivi di email pone criticità metodologiche e tecnologiche. Le sfide aperte sono principalmente riconducibili alla formazione di pacchetti di versamento esaustivi anche relativamente a metadati di processo, o all'utilizzo di tecnologie di emulazione o di migrazione sicure e performanti in termini di quantità e qualità dei dati. Per una panoramica dei diversi progetti, tool e scenari di workflow dalla donazione alla conservazione si consiglia la lettura di CLIR (2018).

⁷ In Italia i dati AgID (Agenzia per l'Italia Digitale) della Posta elettronica certificata nell'anno 2021 indicano circa 14 milioni di caselle PEC attivate e 2 miliardi e mezzo di messaggi scambiati. <https://www.agid.gov.it/it/piattaforme/posta-elettronica-certificata/statistiche-utilizzo-pec> (ultima consultazione: 26/05/2022). Ovviamente i dati AgID sulle PEC sono in parte rilevante riferibili alla pubblica amministrazione ma, certamente, vi è un numero non trascurabile di professionisti che fa largo uso di questo strumento per lo svolgimento delle attività lavorative. Al trend di crescita della produzione di documenti digitali si contrappone un dato preoccupante di documenti conservati digitalmente per i quali si registra una battuta di arresto nel 2021 con numeri stazionari rispetto ai due anni precedenti (Dati AssoConservatori 2019–2021, Stati generali della Conservazione, Università della Calabria).

lia, confinata ad una prospettiva a medio termine, senza la modellazione di un quadro evoluto, complessivo e condiviso di *long term digital preservation*, ogni tentativo di sistematizzazione rischia di rimanere confinato a definiti domini e specifiche casistiche.

Quanto evidenziato mette in luce anche l'urgenza di una rilettura, secondo una logica più inclusiva, della conservazione digitale che è chiamata a superare il mero concetto di adempimento riferito quasi esclusivamente a tipologie documentali ben determinate.

3. Considerazioni sul ciclo di vita delle email in un archivio di persona

La posta elettronica è un servizio a diffusione molto ampia a causa del continuo utilizzo nella sfera privata e nell'ambito lavorativo. L'essenzialità di tale servizio unitamente alle caratteristiche di ubiquità e pervasività della tecnologia di base rimandano ad una produzione estremamente copiosa di messaggi⁸. I numeri incrementali che caratterizzano il fenomeno delle email⁹ sollecitano un approccio sistematico e consapevole nella scelta di strategie di sicurezza, di tecnologia, di selezione, di recupero, di accesso, di rappresentazione delle informazioni, di mantenimento di evidenze documentali e giuridiche. Ciò anche in ragione del ciclo di vita delle email le cui fasi e relative attività sono condizionate anche dalla quantità e dalla struttura dei messaggi presenti in ogni account. Definire e tracciare il ciclo di vita delle email, in presenza di così tante variabili, non è un'operazione semplice soprattutto negli archivi di persona, nei quali il processo di conservazione subentra in una fase in cui i messaggi sono stati formati, gestiti e consolidati con modalità che non neces-

⁸ Malgrado il dilagare dei social e dei sistemi di messaggistica l'uso della posta elettronica si conferma quale importante sistema di comunicazione con numeri importanti: 306 miliardi di email vengono scambiate ogni giorno con un totale annuo di oltre 101.000 miliardi di messaggi di posta elettronica. Un settore in crescita che vede stime per il 2025 con più di 376 miliardi di email al giorno e che coinvolge oltre 4 miliardi di persone in tutto il mondo utilizzano email, di cui la metà circa tramite smartphone. <https://www.key4biz.it/50-anni-falla-prima-email-oggi-se-ne-inviano-piu-di-300-miliardi-al-giorno-mercato-in-mano-alle-big-tech/352180/> (ultima consultazione: 26/05/2022).

⁹ «Recent trends show that email usage is growing and becoming more deeply embedded in our work and personal lives, blurring the line between business and personal communication[...] Taken as a whole, these trends present an archival paradox. Email and other electronic messages are both ubiquitous and ephemeral, documenting people's professional and personal lives in a chaotic stream of messages and relationships, calling for a new documentary practice[...] But in comparison, email preservation has taken a backseat, with few grant-funded projects and relatively little institutional support[...] Each of these works place attention on the entire range of cultural, legal, ethical, professional and technical considerations that must be addressed if an organization wishes to identify email of permanent value, preserve it in an authentic form, and render it for future use» (Prom 2011, 4).

sariamente permettono una successiva ricostruzione fedele della formazione e sedimentazione. Tuttavia, la definizione, sia pur generica, delle fasi del ciclo di vita delle email aiuta ad individuare attività, politiche, responsabilità e processi che si rivelano fondamentali per la conservazione. Provando a schematizzare il ciclo di vita delle email si possono individuare le seguenti fasi: formazione e uso, trattamento preconservativo, conservazione, recupero e accesso (Fig. 1).



Figura 1: Fasi del ciclo di vita delle email.

Formazione e uso

Negli archivi di persona è difficile e talvolta impossibile esercitare un controllo della posta elettronica fin dal momento della formazione¹⁰. Tuttavia, come accade per tutti i documenti informatici, anche nel caso delle email, la fase di formazione rappresenta un processo delicato e fondamentale nel quale si determinano le condizioni essenziali e le informazioni necessarie per le successive attività di gestione e di conservazione. La scelta delle tecnologie, le azioni in chiave di sicurezza, la gestione dell'account anche mediante applicativi dedicati installati in locale, l'organizzazione dei messaggi e la loro tutela predeterminano, quindi, il corso del ciclo di vita delle email inclusa la loro conservazione e le garanzie di recupero e validità nel tempo. Gli archivisti nord americani e inglesi, consapevoli di tale difficoltà, hanno da tempo proposto una forma di salvaguardia più ampia della posta elettronica negli archivi di persona, che rimanda anche all'instaurazione di un rapporto tra archivista e soggetto produttore da stabilire fin dalle fasi di sedimentazione e selezione del materiale¹¹. La possibilità di raccogliere direttamente dal soggetto produttore informazioni utili alla scelta e alla conservazione dei messaggi di posta è certamente un'opportunità rilevante che consente di sviluppare un progetto condi-

¹⁰ Il controllo della formazione degli account, dei messaggi email e in generale dell'uso della posta elettronica nel caso di pubbliche amministrazioni e o di grandi aziende può essere risolto mediante la definizione di policy preventive e regolamenti specifici che, ad esempio, possono ricondurre la gestione della posta all'interno della gestione documentale con la possibilità di tracciare, di recuperare e di sottoporre a conservazione almeno le comunicazioni connesse a processi lavorativi.

¹¹ Tale pratica va affermandosi anche in Italia. "Ecco allora la prima riflessione, ovvia ma non scontata: quando è possibile, è meglio collaborare con gli esperti della persona di cui si riordina l'archivio fin dall'inizio dell'intervento" (Tealdi 2018, 7-9).

viso di conservazione¹². Tuttavia, sebbene non sia infrequente che i produttori decidano di menzionare nel proprio testamento l'archivio è assai più raro che essi si rendano disponibili, in vita, ad una collaborazione con gli archivisti. Probabilmente i tempi non sono maturi per una sensibilità ed una cultura di tal genere. L'approccio culturale ha un riflesso anche sulle procedure di selezione dei messaggi da sottoporre a conservazione. Un passaggio che si rende necessario in ragione del proliferare delle comunicazioni pubblicitarie, delle copie generate dall'inoltro di messaggi da un account di posta ad un altro di un medesimo produttore o dalla ricezione di materiale di nessuna utilità ai fini della conservazione. Tali messaggi dovrebbero essere periodicamente eliminati manualmente o attraverso il settaggio di regole di selezione automatica o funzioni analoghe normalmente presenti negli applicativi di posta elettronica. Tali tecniche possono supportare anche una gestione più sicura della posta filtrando, ai fini dell'eliminazione, messaggi di spam, phishing, malware ecc. eventualmente non intercettati dal provider.

Il soggetto produttore dovrebbe, dunque, operare uno sfortimento periodico della posta al fine di eliminare gli elementi superflui o addirittura pericolosi. Un'attività di selezione, più analitica e metodica dovrebbe essere poi praticata almeno alla fine di ogni anno, ad esempio, prima della rituale ma necessaria archiviazione dei messaggi di posta prodotti e ricevuti. Il processo dell'archiviazione della posta è certamente una forma di tutela dei messaggi e diventa ancora più efficace se, nel corso dell'anno i messaggi vengono via via classificati in apposite cartelle secondo un particolare ordine ad esempio per mittente, o argomento. Tale classificazione ulteriore, in genere, facilita il successivo recupero della posta ricercata. In sintesi una buona classificazione della posta elettronica potrebbe prevedere un'ordine cronologico annuale e un ordine per materia (mittente, argomento, ecc.). Un modello che richiama metodologie consolidate di classificazione della corrispondenza analogica in cui unità archivistiche cronologicamente ordinate presentano sottodivisioni funzionali organizzate in sottofascicoli per mittente. In ambiente digitale, anche la sola archiviazione per argomenti o corrispondenti potrebbe risultare efficace evitando eventualmente la frammentazione degli affari o degli argomenti su base annuale. La preventiva classificazione e archiviazione della posta elettronica ad opera del soggetto produttore produce un vantaggio nella successiva fase di conservazione e valorizzazione della stessa soprattutto, quando il soggetto produttore, ha anche la lungimiranza di predisporre strumenti atti a documentare la metodologia di classificazione definita ed applicata e a raccordare la posta elettronica con le altre componenti dell'archivio digitale e/o analogico¹³.

¹² <http://www.ariadne.ac.uk/issue/55/williams-et-al/> (ultima consultazione: 26/05/2022).

¹³ Antonio Romiti commentando il saggio di Stefano Moscadelli "L'Archivio di Fabrizio De Andr : osservazioni a volo d'uccello", afferma «In sostanza si nota come non sia stato possibile reperire tra le carte un vincolo naturale, evidenziando cos  una situazione che mostrava

Anche nel caso della tutela dei dati personali un approccio *risk based* mediante classificazione, metadatazione, consapevole valutazione e selezione del materiale da sottoporre a conservazione evita situazioni di *data breach*.

Trattamento preconservativo

La natura di documento informatico dei messaggi di posta elettronica richiama l'urgenza della conservazione al fine di evitare situazioni di perdita, di mancato recupero o di impossibilità di lettura di tali oggetti. Da un punto di vista teorico il processo di conservazione digitale della posta elettronica deve, quindi, essere reso quanto più possibile vicino al momento della formazione dei messaggi, allegati inclusi. Negli archivi di persona tali procedure incontrano ostacoli di varia natura (culturale, tecnologica, economica ecc.) e il rinvio della conservazione diventa una prassi normale se non addirittura scontata. Conseguentemente la fase di trattamento preconservativo, avviene spesso a distanza di molti anni dalla produzione dei messaggi generando scenari molto diversificati. In questa fase del ciclo di vita delle email è necessario verificare l'organicità e la congruenza del materiale da conservare, così come è fondamentale definire metodologie e implementare strumenti che possano supportare una conservazione coerente e consistente. Come già evidenziato le situazioni che possono presentarsi all'archivista a seguito di un processo di donazione o di versamento di un archivio di posta elettronica possono essere molto eterogenee e, in un elevato numero di casi, gli attori coinvolti sono diversi dal soggetto produttore (eredi, archivista ecc.). L'archivista è chiamato, così, a conferire sistematizzazione al materiale da conservare, in situazioni in cui la mancanza di evidenze documentali o informative sufficienti difficilmente permette, malgrado l'impegno meticoloso e responsabile, di restituire archivi di posta completi, documentati e affidabili. La letteratura scientifica di riferimento ritorna più volte su questo aspetto sottolineando la solitudine dell'archivista di fronte a scelte strategiche o di responsabilità quali ad esempio la selezione del materiale, la tutela dei dati personali e delle informazioni riservate, l'autenticità di quanto conservato, in un contesto interdisciplinare in cui mancano modelli stabili di riferimento. Si può affermare che per le criticità citate la fase di trattamento preconservativo è strettamente connessa con processi di qualità in cui la valutazione del rischio, la definizione delle responsabilità e la gestione dei dati assumono particolare rilievo. L'esito positivo di tale fase presuppone un lavoro di squadra in cui sono chiamate in causa diverse

il ricorso ad antecedenti operazioni mirate a compiere consistenti operazioni di scarto. Questi comportamenti sono frequenti nella gestione delle carte degli archivi personali e spesso sono effettuati da persone che non hanno la minima idea del significato dell'operazione che stanno conducendo» (Guarasci, e Pasceri 2011).

professionalità e competenze (documentali, giuridiche, informatiche, storiche, di dominio ecc.).

La complessa e intricata vicenda della posta elettronica di Hilary Clinton durante il suo mandato di Segretario di Stato degli Stati Uniti d'America rappresenta un esempio delle problematiche legate alla gestione e conservazione della posta elettronica anche in ragione del complicato intreccio tra le vicende private e professionali di un soggetto produttore (Salahub and Oldford 2018; Salahub and Oldford 2017; Vasundhara and Mueller 2016).

La presenza di messaggi privati negli account istituzionali o aziendali è, ad esempio, un tema che ritorna frequentemente negli archivi di posta elettronica. La commistione di documentazione pubblico/privata è abbastanza usuale negli archivi di persona specie se il soggetto produttore ha ricoperto incarichi pubblici e istituzionali. Non di rado la documentazione pubblica confluisce nel corpus dell'archivio privato o viceversa i documenti privati presenti negli account istituzionali vengono eliminati senza che il soggetto produttore o gli eredi ne siano informati (Rovella et al. 2020). Tali fenomeni hanno assunto una dimensione ancor più significativa nel periodo della pandemia, durante il quale a seguito del lockdown e della diffusione dello smartworking i dispositivi personali hanno ospitato sempre più frequentemente anche documentazione istituzionale o aziendale.

La costruzione di un buon team con esperienze e competenze diverse produce ricadute positive in termini di organicità, affidabilità e futuro recupero di quanto potrà essere conservato anche a vantaggio della salvaguardia della documentazione e del rispetto della tutela dei dati riservati (personali, sensibili ecc.) presenti nella posta elettronica. Significativi risultati, in tale contesto, arrivano, ad esempio, dall'applicazione di tecniche di Natural Languages Processing (NLP)¹⁴ utilizzate per la modellazione degli oggetti, all'interno delle collezioni, con la finalità di identificare documenti che richiedono restrizioni circa i dati personali o sensibili (Hutchinson 2018, 2251–55).

Dal punto di vista della cybersecurity le email sono particolarmente vulnerabili e anche questo è un aspetto rilevante ai fini della loro conservazione. Una risposta interessante arriva dall'applicazione di tecniche di classificazione automatica delle email (Rovella et al. 2020). I provider di posta elettronica, utilizzano largamente tali strumenti per identificare, ad esempio, i messaggi di spam, phishing ecc. La classificazione automatica della posta può essere utilizzata per dare ordine ai messaggi fin dalla loro formazione permettendo una sedimentazione organizzata per tipologie di messaggi o per contenuto con un impatto positivo sulla consultazione dei contenuti dei messaggi. L'ingente e costante investimento nella costruzione di strumenti per la sicurezza delle mail ha permesso di sviluppare tool e algoritmi sempre più sofisticati in grado

¹⁴ «Natural Language Processing (NLP) is the application of computational methods to model and extract information from human language» (Kamath, Liu and Whitaker 2019).

di lavorare sul piano funzionale e su quello semantico. Tali sistemi partendo da un corpus di email possono associare automaticamente i messaggi a classi pre-determinate o generate dinamicamente. Probabilmente su tale tema potrebbe essere aperta una riflessione con il coinvolgimento degli attori istituzionali non solo italiani ma anche dell'Unione Europea. È evidente che Google, Microsoft e Apple, proprietari dei primi 10 clienti di posta e, di fatto, capaci di controllare il 90% del mercato delle email, non avvertano nessuna pressione della domanda di strumenti a supporto della gestione dei messaggi di posta elettronica. Una concertazione europea, sui ruoli, sulle regole e sulle metodologie di tutela delle mail nel dialogo tra istituzioni e grandi player potrebbe rappresentare un tema di interesse anche in una logica di cittadinanza digitale.

Gli elementi finora considerati sono rilevanti anche ai fini della creazione del pacchetto di versamento (SIP) (ISO 2012) da inviare al sistema di conservazione (AIP). Tutte le scelte operate in merito alle strategie di conservazione che vanno dagli accordi di donazione, alla definizione di schemi di metadati e dei formati tecnologici fino alla redazione del contratto di conservazione, avranno un ruolo molto importante nella composizione del pacchetto che dovrà presentare caratteristiche di coerenza e di interoperabilità, proprietà fondamentali anche per la formazione del successivo e conseguente Pacchetto di Archiviazione. Durante la fase di preconservazione sarà allora necessario predisporre un accurato piano di conservazione, un piano di gestione dei dati corredato da adeguati schemi di metadati, utili a descrivere e documentare le email anche in relazione ai processi di trasmissione, provenienza e agli elementi di autenticità (Dinnen and Kartalić 2020, 413–24). Il modello organizzativo/gestionale così elaborato costituirà una solida base per l'accordo di conservazione.

Conservazione, recupero e accesso ai messaggi

Da quanto finora affermato è facilmente desumibile che anche la fase di conservazione riserva una serie di criticità che richiamano una programmazione attenta di tutti i processi. Per tale ragione è fondamentale che nel sistema di conservazione risultino chiare le responsabilità dei diversi attori che contribuiscono alla creazione dei pacchetti informativi e alla gestione degli oggetti digitali. La formazione dell'AIP richiede al conservatore soluzioni che avranno una ricaduta decisiva sulla possibilità di estrazione dei messaggi e sull'autoconsistenza informativa espressa tramite i metadati.

I possibili scenari di conservazione delle email possono contemplare situazioni diversificate in cui il SIP può essere composto da singoli messaggi completi di metadati o dall'intero account di posta. Il focus si sposta allora sull'AIP che, a prescindere dallo scenario che ha guidato la formazione del SIP, dovrebbe ospitare le email sia in formato originale che in un formato normalizzato (generalmente eXtensible Markup Language – XML). I metadati, i log, l'oggetto

del messaggio, il mittente, gli allegati, le informazioni relative alla stabilità e l'integrità dovrebbero essere salvate in un database XML, o in una struttura XML flessibile da archiviare (come file) all'interno del pacchetto. La struttura dell'indice del pacchetto di archiviazione indicata nello standard UNI SInCRO (2020) permetterebbe di ospitare tutti gli oggetti, i metadati e le informazioni da conservare. I processi di *digital preservation*, gli agenti, gli eventi e le eventuali informazioni aggiuntive potranno essere documentate da metadati specifici in accordo con il modello PREMIS completando il set di metadati anche negli elementi che documentano la provenienza e l'autenticità degli oggetti digitali. Il risultato sarà un nuovo AIP, in cui sono presenti email e metadati ampliati. I pacchetti di archiviazione così creati, con la caratteristica di autoconsistenza potrebbero essere collocati in un repository certificato del tipo Trustworthy.

L'attenzione verso la completezza e l'autoconsistenza dell'AIP facilita la formazione di pacchetti di distribuzione coerenti favorendo l'accesso sicuro ai contenuti conservati nel sistema. Le architetture per l'interoperabilità e la cooperazione applicativa¹⁵ tra i sistemi di conservazione previste nel modello OAIS hanno effetto anche sull'accesso ai documenti conservati. Nel caso degli archivi a lunga conservazione sarebbe auspicabile, come propone Stefano Pigliapoco, «realizzare architetture miste, ad esempio, combinando la *Federated archives* con la *Shared resources archives* per conseguire il duplice obiettivo di ridurre i costi dell'impianto tecnologico e attivare forme avanzate di accesso ai contenuti digitali» (Pigliapoco 2016). In quest'ultimo aspetto assumono rilevanza i diversi sistemi di classificazione e estrazione automatica dei contenuti quali strumenti qualificanti nel processo di accesso alle email. Un discorso complesso e articolato che può chiamare in gioco tecnologie di Machine Learning e Deep Learning, in grado di fornire risultati su un piano generale di organizzazione ed estrazione di contenuti ma anche ad un livello molto dettagliato di analisi dei concetti, di estrazione di dati e metadati anche nel caso di testi che con grafici, tabelle (Rovella et al. 2022), immagini e audio. All'interno del messaggio è possibile trovare diversi elementi, talvolta anche sotto forma di link verso altre risorse, in tal caso molta attenzione deve essere prestata al formato digitale e alla sua capacità di autodocontenimento. Negli ultimi anni è stata considerata la possibilità di utilizzare il formato FITS per la conservazione delle email e una sperimentazione condotta in tal senso potrebbe verificare l'eventuale efficacia di tale scelta (UNI 2022).

¹⁵ Le tre diverse architetture definite nello standard OAIS sono: *Independent archives*; *Cooperating archives*; *Federated archives*; *Shared resources archives* (ISO 2012).

4. Alcune considerazioni conclusive

Le poche note richiamate in questo lavoro rimandano ad una letteratura certamente ampia ma non definitiva sul tema della conservazione delle email negli archivi di persona. Malgrado i numerosi progetti internazionali¹⁶ al momento è comunque difficile tracciare un quadro di insieme (Standford Libraries. n.d; Schneider et al. 2017; Schneider et al. 2019). Manca uno sforzo ulteriore di sistematizzazione verso un approccio inclusivo delle diverse sfaccettature che connotano un tema complesso ed esteso che, a diversi livelli, evidenzia la necessità di un raccordo tra i molteplici attori del sistema in una logica di cooperazione e di accountability necessaria per la buona riuscita del processo di conservazione. Se tali considerazioni sono valide in un contesto internazionale, in Italia, vista anche la scollatura che si è andata generando tra la conservazione digitale e la conservazione digitale a lungo termine, si coglie la traccia di un allontanamento progressivo da soluzioni integrate e definitive.

Riferimenti bibliografici

- Allegrezza, Stefano, e Luca Gorgolini, a cura di, 2016. *Gli archivi di persona nell'era digitale. Il caso dell'archivio di Massimo Vannucci*. Bologna: il Mulino.
- Allegrezza, Stefano. 2022. *Recent developments on email preservation: towards the ultimate solution?* IRCd. http://ircdl2022.dei.unipd.it/downloads/papers/IRCdL_2022_paper_24.pdf.
- Associazione Italiana Biblioteche (AIB). 2019. *Linee guida sul trattamento dei fondi personali*. Milano. <https://www.aib.it/struttura/commissioni-e-gruppi/gbaut/strumenti-di-lavoro/linee-guida-sul-trattamento-dei-fondi-personali/>.
- Barrera, Giulia. 2006. *Archivi di Persona*. Vol. 2 di *Storia d'Italia nel secolo ventesimo. Strumenti e fonti*, a cura di Claudio Pavone. Roma: Ministero per i beni e le attività culturali, Dipartimento per i beni archivistici e librari, Direzione generale per gli archivi.
- Carassi, Marco, a cura di. 2018. "Sugli Archivi Di Persona: Esperienze a Confronto". *Il mondo degli archivi*. I Quaderni 5. http://www.ilmondodegliarchivi.org/images/Quaderni/MdA_Quaderni_n5.pdf.
- Council on Library and Information Resources (CLIR). 2018. *The future of email Archives. A Report from the Task Force on Technical Approaches for Email Archives*. Washington DC.

¹⁶ Per una panoramica sui principali progetti internazionali si consiglia la lettura di CLIR (2018).

- De Mauro, Tullio. 2016. *Nuovo vocabolario di base della lingua italiana*.
- Dinnen, Jesse David, and Maja Kartalić. 2020. "E-Mail as Legacy: Managing and Preserving E-Mail as a Collection." *Libraries and the Academy* 20 (3): 413-24. Doi:10.1353/pla.2020.0022.
- Guarasci, Roberto, e Erika Pasceri, a cura di. 2011. *Archivi privati*: Studi in onore di Giorgetta Bonfiglio-Dosio. Roma: Consiglio Nazionale delle Ricerche, SEGID.
- Kamath, Uday, John Liu, and James Whitaker. 2019. *Deep Learning for NLP and Speech Recognition*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-14596-5>.
- Hutchinson Tim. 2018. "Protecting privacy in the archives: Preliminary explorations of topic modeling for born-digital collections." In 2017 IEEE International Conference on Big Data (Big Data) 2251-55. <https://doi.org/10.1109/BigData.2017.8258177>.
- International Standard Organization (ISO) 14721:2012. *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*.
- Landino, Costantino, e Pasqualina A. Marzotti. 2018. *Memorie dinamiche. La conservazione dei data base e il web archiving*. Roma: Edizioni ANAI.
- Navarrini, Roberto. 2005. *Gli Archivi Privati*, Lucca: Civita Editoriale.
- Pearce-Moses, Richard. 2005. *A Glossary of Archival & Records Terminology*, Chicago: SAA.
- Pigliapoco, Stefano. 2016. "Il modulo ACCESS del modello OAIS. Architetture e funzionalità avanzate." *AIDAinformazioni*, 34 (1-2): 49-59.
- Pontevolpe, Gianfranco, and Silvio Salza. 2009. "Keeping and Preserving Email report." In *Progetto InterPARES*. http://www.interpares.org/ip3/ip3_products.cfm?cat=10.
- Prom, Christopher J. 2011. *Preserving Email*. Digital Preservation Coalition. <http://dx.doi.org/10.7207/twr11-01>.
- Romiti, Antonio. 2011. "Gli archivi privati visti da più prospettive". In *Archivi privati: Studi in onore di Giorgetta Bonfiglio-Dosio*, a cura di Roberto Guarasci e Erika Pasceri, 7-28. Roma: Consiglio Nazionale delle Ricerche, SEGID.
- Rovella, Anna, Roberto Guarasci, e Francesca Parisi. 2020. "Gestione e conservazione dei messaggi di posta elettronica." *AIDAinformazioni*, (1-2):149-75.

- Rovella, Anna, Alexander Murzaku, Eugenio Cesario, Martin Critelli, Armando Bartucci, and Francesca M.C. Messiniti. 2022. "Analysis, evaluation and comparison of knowledge extraction tools in the Environmental and Health domain. A holistic approach." In *Knowledge Organization and Management in the domain of Environment and Earth Observation (KOMEEEO)*, *Proceedings of the International KOMEEEO Conference 2021*, edited by Antonietta Folino, and Roberto Guarasci. ERGON VERLAG.
- Salahub, Christopher, and Wayne Oldford. 2017. *Interactive Filter and Display of Hillary Clinton's Emails: A Cautionary Tale of Metadata*. <https://www.researchgate.net/publication/315876309>.
- Salahub, Christopher, and Wayne Oldford. 2018. "About 'Her Emails.'" *Significance* 15, no. 3: 34–37. <https://doi.org/> <https://doi.org/10.1111/j.1740-9713.2018.01148.x>.
- Schneider, Josh, Peter Chan, Glynn Edwards e Sudheendra Hangal. 2017. "ePADD: Computational Analysis Software Facilitating Screening, Browsing, and Access for Historically and Culturally Valuable Email Collections". *D-Lib Magazine* 23, n. 5/6 (maggio). <https://doi.org/10.1045/may2017-schneider>.
- Schneider, John, Chance Adams, Sally De Bauche, Reid Echols, Callum Mckean, Jessica Moran, and Dorothy Waugh. 2019. "Appraising, Processing, and Providing Access to Email in Contemporary Literary Archives". *Archives and Manuscripts* 47(3):305-326 DOI:10.1080/01576895.2019.1622138.
- Tealdi, Elena. 2018. "Piero Martinetti, filosofo. (Pont Canavese 21 agosto 1872 – Cuorgnè, Torino 22 marzo 1943)". In *Sugli Archivi Di Persona: Esperienze a Confronto*, a cura di Marco Carassi, 7-9. *Il Mondo degli archivi*. I Quaderni 5. http://www.ilmondodegliarchivi.org/images/Quaderni/MdA_Quaderni_n5.pdf.
- UNI 11386:2020. *Supporto all'Interoperabilità nella Conservazione e nel Recupero degli Oggetti digitali (SInCRO)*.
- UNI 11845:2022. *Processi di gestione della conservazione a lungo termine di immagini digitali con l'uso del formato FITS*.
- Valacchi, Federico. 2015. *Diventare archivisti. Competenze tecniche di un mestiere sul confine*. Milano: Editrice bibliografica.
- Vasundharaan, Dehiya, and Mueller Klaus. 2016. "Analyzing Hillary Clinton's Emails." In *Poster Abstracts of IEEE VIS*, October 2016. <https://www3.cs.stonybrook.edu/~mueller/papers/Analyzing%20Hillary%20Clinton%20Emails%20VIS%202016.pdf>.
- Weston, Paul Gabriele, Primo Baldini, Emmanuela Carbé, e Laura Pusterla. 2019. "PAD - Pavia Archivi Digitali e gli archivi degli scrittori." In *Digitalia*, XIV(1).

Il Blog di qualità

Proposta di un nuovo format di pubblicazione nell'editoria scientifica

GIOVANNI SALUCCI*

ABSTRACT: This paper aims at presenting the Quality Blog as a new inspiring model to reduce the critical issues typical of almost all scientific journals. In continuity with more traditional formats, the Quality Blog aims at visibility and it is characterized by five essential properties: i) a more communicative cut of the contributions, organized in sections/themes and often by invitation; ii) an internal organization consisting of a small group of experts supporting the Director; iii) a substantial focus on editorial and metadata management quality; iv) the lack of a peer-review process in favor of an internal evaluation; v) the coexistence of scientific and informational contributions having a more fluid structure and the presence of multimedia materials. Just like the blog or a scientific journal in *Diamond Open Access*, the Quality Blog adopts membership, sponsorship, and/or funding as the main economic models; it is free of charge both for authors and readers.

Keywords: Quality Blog, Journal, Academic publishing, Scholarly communication, Blog.

1. Introduzione

L'editoria scientifica è sin dagli inizi del Duemila costretta ad affrontare numerose sfide e difficoltà dettate principalmente da fattori economici e logiche di sostenibilità (Kyrillidou 2012)¹. Si tratta di criticità di ampia portata, che neanche la rivoluzione digitale e i programmi di diffusione in *Open Access*² sono stati in grado, pur nella loro capillarità, di arginare: lungi dal creare un

* Università degli Studi di Firenze, Firenze, Italia. giovanni.salucci@unifi.it.

¹ L'autore raccoglie e presenta in particolare i dati relativi al costo di acquisto di monografie e abbonamenti a riviste scientifiche da parte delle biblioteche universitarie americane nel periodo 1986-2011, evidenziandone il tasso di crescita nel tempo.

² In ordine di importanza, PLAN-S rappresenta l'iniziativa più strutturata e ampia nel settore, <https://www.coalition-s.org/why-plan-s/> (ultima consultazione: 26/05/2022): muovendo dall'intento dei promotori di Coalition-S di rendere disponibile in *Open Access*, entro il 2020, la letteratura scientifica finanziata con contributi pubblici e di fondazioni no-profit, l'iniziativa raccoglie l'esperienza maturata da soggetti pubblici e privati nel perseguimento di tale scopo (per citarne alcuni, *Wellcome Trust*, *Gates Foundation*, *UKRI*, *OpenAire*).

sistema più democratico, inclusivo e sostenibile, questi hanno anzi contribuito ad accentrare il mercato editoriale accademico nelle mani di pochi editori globali (Larivière, Haustein, e Mongeon 2015; Guatelli 2016).

In tempi recenti si è assistito a una vivace discussione sulle possibili cause all'origine di questa situazione (Brembs et al. 2021)³. Inutile omettere che lo scopo principale del dibattito è stato (ed è tuttora) quello di verificare se il mondo delle università e della ricerca possa gestire autonomamente i propri risultati scientifici senza necessariamente coinvolgere gli editori, con l'effetto immediato di ridurre così i costi globali di produzione e diffusione dei risultati stessi.

Senza addentrarsi in analisi approfondite e marginali rispetto agli obiettivi del presente contributo, possiamo così riassumere i tratti salienti del suddetto dibattito:

- i. critiche al sistema della *peer-review* (Hames 2014; Eve et al. 2021) anche per motivazioni di sostenibilità economica (Aczel, Szaszi, e Holcombe 2021);
- ii. innovazione di ruoli e responsabilità nella *peer review* (Guerrini 2022)⁴, rispetto al processo standard⁵;
- iii. dubbi sull'utilizzo o meno dell'*Impact Factor* come criterio di misurazione della qualità scientifica delle riviste⁶;
- iv. limiti di sostenibilità dei principali modelli economici di diffusione dell'*Open Access* (Salucci e Cinotti 2021);
- v. ricorso ai *preprint server*⁷ o *repository* (Urbano et al. 2021)⁸, oppure agli *overlay journals* (Rousi e Laakso 2022) come modalità di deposito e

³ Si tratta di una delle posizioni più dure espresse contro gli editori accademici globali, ritenuti tra i principali responsabili della attuale crisi, a causa del loro sistema di potere economico-monopolistico.

⁴ Si vedano le considerazioni sul processo e sul ruolo del revisore.

⁵ Per la definizione di *standard peer review* si fa riferimento a Coalition-S «A 'standard peer review process' is defined as involving at least two expert reviewers who observe COPE guidelines and do not have a conflict of interest with the author(s). An implicit validation has occurred when the reviewers state the conditions that need to be fulfilled for the article to be validated. An explicit validation is made by an editor, an editorial committee, or community overseeing the review process». <https://www.coalition-s.org/statement-on-peer-reviewed-publications> (ultima consultazione: 26/05/2022).

⁶ L'insieme delle critiche è presente nel manifesto DORA (*San Francisco Declaration on Research Assessment*: <https://sfdora.org/read/>) varato da The American Society for Cell Biology (ASCB) a San Francisco, CA, il 16 dicembre 2012. A tale manifesto hanno aderito numerose istituzioni di ricerca e oltre duecento università in tutto il mondo.

⁷ <https://scity.org/> (ultima consultazione: 26/05/2022).

⁸ Gli autori si focalizzano in particolare sul *case study* della conferenza EDICIC 2019 che, incoraggiando i relatori a pubblicare i risultati della loro ricerca in appositi *repository* prima del debutto dei lavori, non ha prodotto atti, facilitando così la diffusione pre-conferenza della ricerca e utilizzando la conferenza stessa come luogo di dibattito per aggiornare in

diffusione dei contributi in grado di ridefinire il concetto stesso di pubblicazione.

Tutti questi elementi creano spazio per la comparsa di nuovi modelli di pubblicazione scientifica che prevedano una riduzione dei costi e la semplificazione dei processi.

Esemplare in questo senso è l'iniziativa della *Open Research Europe*⁹, recente piattaforma destinata alla pubblicazione di contributi (articoli, saggi, recensioni) prodotti nel quadro di progetti finanziati tramite il piano Horizon della Commissione Europea. Vale la pena, ai fini della nostra argomentazione, addentrarsi pur brevemente nelle specifiche, con relative conseguenze, di questa iniziativa. Essa propone di fatto il modello ibrido di *preprint* e *post-publication peer-review* con pubblicazione in *Open Access* e sotto la licenza Creative Commons-BY (CC-BY); il finanziamento avviene mediante il contributo attivo dell'autore (APC) rimborsabile da fondi di ricerca debitamente allocati¹⁰. La *peer-review* prevista è ben lontana dalla tradizionale *double blind peer review*, poiché essa viene in questo caso effettuata dopo la pubblicazione dell'articolo (*post-publication*); i revisori di norma sono segnalati dagli stessi autori dell'articolo in valutazione; l'intera documentazione di ciascun referaggio viene pubblicata in chiaro insieme all'articolo di riferimento (*open*). Alla luce di ciò, è evidente che la piattaforma europea propone un nuovo processo, articolato su più fasi, per selezionare ciò che va pubblicato (e che, *sul piano formale*, ha dunque valore scientifico).

Dati questi presupposti, proponiamo un nuovo format progettato per adattarsi a contesti di diffusione differenti da quelli tradizionali cui ricercatori e studiosi sono generalmente abituati. Anche se numerosi sono gli esempi di contenitori già esistenti per la comunicazione scientifica, nessuno risulta ad oggi raccogliere tutte le specifiche previste in simultanea nella nostra proposta; i paragrafi a venire ne detaglieranno gli aspetti.

2. Metodo: modelli, posizionamento e contaminazione

Per proporre e promuovere un nuovo modello di pubblicazione che possa risolvere o almeno contribuire a ridurre i problemi dell'attuale sistema dell'editoria accademica, non si può che partire analizzando nel dettaglio il modello principale e universalmente diffuso di pubblicazione: l'articolo scientifico.

tempo reale i risultati scientifici esposti.

⁹ All'insegna del motto *Rapid and Transparent Publishing*, la piattaforma è ad oggi accessibile sia per le *submission* che per la consultazione degli articoli: <https://open-research-europe.ec.europa.eu/> (ultima consultazione: 26/05/2022).

¹⁰ Per la terminologia utilizzata si veda "Glossario dell'Open Access (1.0)". Zenodo: 1-21. <https://zenodo.org/record/1208418> (ultima consultazione: 26/05/2022).

Normalmente accolto all'interno di una rivista scientifica, l'articolo presenta una serie di caratteristiche ed elementi peculiari che ne hanno fatto, nel tempo, lo strumento più adatto per la diffusione dei risultati della ricerca scientifica soprattutto nell'ambito disciplinare STM (*Science, Technology, Medicine*). Lo stesso può dirsi anche per quello HSS (*Humanities and Social Sciences*), dove comunque la monografia, in certe aree disciplinari, continua a rappresentare ancora oggi il modello prevalente di pubblicazione. D'altro lato, volgendo l'attenzione dalla comunicazione accademica a quella più divulgativa e generalista, si osserva che il blog è il contenitore più diffuso di pubblicazione e, al suo interno, il post rappresenta l'elemento unitario di comunicazione al pari dell'articolo in rivista scientifica.

Questa breve digressione preannuncia il contesto di adozione e applicabilità del format che si presenta qui per la prima volta. In fase di ideazione, si è deliberatamente optato per un ibrido che presentasse elementi tipici dei due modelli appena citati; anche nella scelta del nome abbiamo ugualmente optato per una nomenclatura che testimoniassse questa contaminazione.

2.1. *La scelta del nome e il posizionamento*

Blog di qualità è il nome che è stato scelto per il nuovo format, con il chiaro intento di evidenziare i due aspetti fondamentali che ne denotano lo scopo: a) puntare alla comunicazione e diffusione dei risultati, b) con un particolare focus sulla qualità dei contenuti.

Nell'individuazione del nome abbiamo prima di tutto verificato che esso fosse *nuovo* oltre che *innovativo*, in modo da gettare le basi di riconoscibilità e possibilmente di successo per il formato stesso. Effettuando ricerche in banche dati scientifiche, aggregatori accademici e finanche in Google emerge che sia per "Blog di qualità" che per la versione inglese "quality blog" (altrimenti digitata "quality-blog") nessun risultato attestato sembri pertinente rispetto all'editoria accademica. Le occorrenze trovate sul web si riferiscono semmai a indicazioni di *web writing* o a linee guida di ottimizzazione dei contenuti dei *post* del blog in vista di un incremento di visibilità sul web; in nessun caso l'espressione "Blog di qualità" è attualmente legata al mondo accademico o della ricerca scientifica o a un possibile formato o piattaforma di pubblicazione scientifica.

Esistono esempi strutturati di blog istituzionali che pubblicano contenuti scientifici di qualità, quali in ambito economico il network di blog LSE¹¹ della *London School of Economics* oppure, negli ambiti di cultura, politica e scienze, il blog della Cambridge University Press¹². Questi esempi tuttavia, così come altri pubblicati in lingue e contesti differenti, pur presentandosi come blog a tema

¹¹ <https://blogs.lse.ac.uk/> (ultima consultazione: 26/05/2022).

¹² <https://www.cambridgeblog.org/about/> (ultima consultazione: 26/05/2022).

scientifico, non hanno però molte delle specifiche del formato che proponiamo, e proprio per questo sono stati da noi considerati come l'altra linea di demarcazione, al pari delle riviste scientifiche, entro cui collocare il nuovo formato.

In una fase iniziale abbiamo analizzato nel dettaglio le caratteristiche peculiari delle riviste scientifiche e dei blog scientifici, mettendoli a confronto in modo da dedurre tratti di analogia e discontinuità. Dopo questa fase sinottica, si è passati a decidere quali delle proprietà desunte dovessero considerarsi come irrinunciabili e caratterizzanti per il Blog di qualità. Come criterio generale, si è stabilito che un qualunque contributo pubblicato nel Blog di qualità debba per definizione avvicinarsi il più possibile ad un articolo in rivista, bilanciando due condizioni essenziali: che esso si discosti dal suo modello soprattutto negli aspetti che incidono sui costi di realizzazione, garantendo però il giusto compromesso per assicurare la qualità complessiva del prodotto. Le scelte fatte sono state anche confrontate con le recenti linee programmatiche per la diffusione del modello *Diamond Open Access* (Ancion et al. 2022).

Nello schema seguente illustriamo il posizionamento del "Blog di qualità" rispetto agli altri due principali formati standard (Fig. 1).



Figura 1: Rappresentazione del posizionamento del Blog di qualità rispetto ai due formati tradizionali di riferimento.

Chiariti gli orientamenti principali del nostro metodo, si presentano di seguito le proprietà caratterizzanti il Blog di qualità, con un focus particolare sulle scelte necessarie da intraprendere a livello di testata; rimandiamo ad altra sede la descrizione puntuale e approfondita degli elementi che invece caratterizzano il singolo contributo.

2.2. Le proprietà del Blog di qualità a livello di testata

2.2.1. Promotori e proprietà

Il soggetto promotore del Blog di qualità ha come tratto distintivo quello di gravitare attorno al mondo della produzione scientifica; rientrano in questa descrizione numerose categorie di soggetti anche differenti sul piano giuridico, quali ad esempio un'associazione professionale o scientifica, una società scientifica o accademia, un dipartimento universitario, o ancora più semplicemente un gruppo di esperti coinvolti in un progetto collegiale su un tema scientifico. Tutti questi possibili soggetti muovono dall'intento di realizzare un'iniziativa editoriale atta in primo luogo a valorizzare e diffondere contributi scientifici auto-prodotti. Di norma, quindi, manterranno la proprietà dell'iniziativa editoriale e ne cureranno anche il progetto scientifico e il posizionamento, il che equivale a dire che non vige, nella fattispecie, l'esigenza di affidarsi a un editore esterno, come solitamente accade per le riviste scientifiche.

Inoltre, se la rivista scientifica nasce generalmente per accogliere articoli firmati da autori esterni, il Blog di qualità pubblica in prima istanza contributi prodotti direttamente da chi è implicato nella sua gestione, o in alternativa da autori esterni invitati a dare il loro contributo su un dato tema di cui sono ritenuti esperti. Questo aspetto è ciò che differenzia maggiormente il Blog di qualità da una rivista accademica, che per natura si rivolge ad autori non già implicati nella gestione della rivista stessa.

2.2.2. Scopo e finalità

L'obiettivo di questo assetto è di far nascere e consolidare il Blog di qualità come fonte autorevole di informazione e discussione su uno o più temi oggetto di indagine scientifica. La vastità delle tematiche affrontate dal format dipenderà da una molteplicità di fattori: dall'ampiezza delle risorse umane disponibili alla collaborazione, dai mezzi finanziari e dalle competenze coinvolte, per arrivare al taglio comunicativo e al livello di approfondimento scelto. In aggiunta, il Blog di qualità può porsi come uno strumento di comunicazione che, pur potendo assorbire attività svolte in precedenza su altri canali o strumenti (ad esempio, bollettini, newsletter, dossier sul sito internet o attività

sui social network), aiuti a creare un nuovo modo di raccontare e diffondere la ricerca. Infine, uno degli scopi precipui del progetto potrebbe essere quello di appoggiare e valorizzare alternativamente iniziative formative, scientifiche, professionali e/o commerciali (si pensi ad esempio a Master universitari oppure organizzazione di eventi o corsi di formazione).

2.2.3. *Target di riferimento*

La scelta del target di riferimento sarà una delle più impegnative da compiere. Due istanze ne denotano il processo:

- 1) definire le tematiche di cui si occuperà il progetto in modo che siano sostenibili (nel tempo) con le risorse a disposizione;
- 2) delimitare il pubblico di riferimento, identificando il target in termini di livello di studio, età, lingua e interesse. Ciò può essere particolarmente vero se si considera il Blog di qualità come iniziativa di appoggio e pubblicità di iniziative formative, professionali e/o commerciali.

2.2.4. *Finanziamento e modelli economici*

Il Blog di qualità rientra senza dubbio nella macro-categoria dei progetti auto-finanziati, la stessa che contempla e accetta la coesistenza di diversi modelli economici (ad esempio, *membership*, *funding*, *cross-subsidiary model*, ecc.), seguendo i quali il progetto verrebbe rispettivamente sostenuto nel tempo da una molteplicità di garanzie (quote sociali, finanziamenti esterni o interni) o utilizzando il margine operativo di altri progetti. Quale modello usare o se ricorrere a una combinazione ragionata di più metodi dipenderà dal singolo caso, in funzione – evidentemente – del promotore. Non è in nessun caso contemplata l'applicazione del modello APC (*Article Processing Charge*) che prevederebbe il versamento di un contributo alla pubblicazione direttamente dall'autore o il finanziamento dello stesso grazie a fondi di ricerca stanziati dall'istituzione o ente di provenienza.

2.2.5. *Presenza in rete*

Il Blog di qualità viene pubblicato in rete con un proprio dominio internet. Nel caso di un progetto che parta da zero, occorrerà registrare un nuovo dominio; si potranno usare estensioni generali (.net, .com, .org, .info, ecc.) oppure geografiche (.it, .de, .ch, ecc.), indirizzando la scelta in funzione delle specifiche di ciascun progetto.

Inoltre, nella scelta del nome della testata e nella registrazione del nome di dominio si terranno in considerazione i criteri di semplicità, riconoscibilità, imme-

diatezza al fine di favorire la diffusione. Più nel dettaglio, nella predisposizione del sito internet due ipotesi sono prevedibili: limitarsi alla sola presenza del Blog di qualità oppure prevedere alternativamente di estendere questo a delle componenti preesistenti legate all'istituzione, ente e/o associazione che lo gestisce. In tal caso, si auspica l'impiego di due sottodomini differenti, proprio per mantenere il Blog di qualità come entità a sé stante, con un proprio posizionamento.

2.2.6. Adempimenti burocratici e periodicità

Il Blog di qualità si pone come una pubblicazione periodica esclusivamente digitale; se edita da un soggetto italiano è dunque sottoposta alla normativa nazionale sulla stampa. In merito alla necessità di registrazione al tribunale della testata e del relativo sito internet di pubblicazione, il D. Lgs. 18/5/2012 n. 63¹³ ne esclude esplicitamente l'obbligo e, analogamente, in quanto pubblicazione scientifica digitale, non è richiesta la presenza di un direttore responsabile iscritto all'Albo dei giornalisti (o all'elenco speciale), come invece è necessaria per i periodici a stampa.

Per attestare e garantire che si tratti di un'iniziativa scientifica è comunque indispensabile che si adempia ad alcune incombenze amministrative. Di queste, la più importante: fare richiesta al Consiglio Nazionale delle Ricerche (CNR) del codice International Standard Serial Number (ISSN) nella versione di periodico digitale¹⁴, stabilendone a monte la periodicità. Nel fare questa scelta si consideri che, secondo le attuali regole di assegnazione, per fare domanda per ottenere l'identificatore ISSN occorre che la pubblicazione sia già disponibile online con almeno il primo numero pubblicato.

2.2.7. Management e organizzazione

Il Blog di qualità presenta una struttura più snella rispetto a quella che caratterizza attualmente le riviste scientifiche. Essa si articola in un direttore scientifico (che è anche il responsabile del progetto) e nel comitato scientifico, un gruppo (anche ridotto) di esperti che affianca il direttore sia nella produzione dei contenuti che nella valutazione scientifica. Per progetti limitati, o nella fase iniziale di progetti destinati a crescere, non serve una redazione strutturata ma solo un coordinatore; questo ruolo viene svolto direttamente dal direttore o assegnato a un altro membro del comitato scientifico. Con questa configurazione interna, il Blog di qualità si posiziona a un livello intermedio tra una rivista (provvista di un ampio numero di referee e di un comitato scientifico

¹³ <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:2012-05-18;63> (ultima consultazione: 26/05/2022).

¹⁴ Così come indicato nelle raccomandazioni del CNR: v. <https://www.bice.cnr.it/seriali-correnti> (ultima consultazione: 26/05/2022).

internazionale, i cui membri devono essere “incardinati” nel mondo universitario) e un blog (che, non prevedendo un processo formale di valutazione scientifica, non presenta di norma alcun comitato). Questa organizzazione dipende poco dalla disciplina oggetto della testata.

2.2.8. *Tipologia di contributi e lingua*

Il Blog di qualità si doterà di norme di gestione chiare e trasparenti, rese note sul proprio sito internet. Altresì, esso accetterà contributi scritti in prevalenza direttamente da autori coinvolti nell’organizzazione del Blog stesso o da membri dell’istituzione di riferimento. È possibile da qui dedurre i generi di contributi più comunemente accettati e pubblicati: editoriali e *position paper* (Iuliano 2021), rassegne tematiche, report e linee guida, articoli di approfondimento.

Si prevede che siano pubblicati articoli nella lingua originale dell’autore con la predisposizione aggiuntiva, a cura della redazione, di un abstract e una lista di parole chiave in lingua inglese, da utilizzare sia per i metadati che ai fini di una più efficace diffusione.

2.2.9. *Organizzazione interna dei contenuti*

La natura ibrida del Blog di qualità si manifesta anche nella coesistenza di contributi sia scientifici che divulgativi; pur comparando con formati editoriali e *pattern* comunicativi differenziati tra loro, essi saranno organizzati in rubriche o sezioni in base alle tematiche comuni affrontate.

Tra la rigida organizzazione dei contenuti prevista per un articolo di rivista e la libera esposizione tipica del post del blog, il contributo nel Blog di qualità rappresenta una via intermedia; è prevista certo una struttura, ma dinamica e adattabile alla tipologia di contributo e ai temi in esso trattati. Potranno inoltre prevedersi al suo interno materiali multimediali non di semplice approfondimento, ma integrati al testo stesso.

2.2.10. *Gestione, controllo e identità tecnologica*

Per la gestione del Blog di qualità occorre che sia presente una piattaforma in grado di tracciare e gestire tutte le fasi previste dal *workflow* editoriale tramite i moduli fondamentali di valutazione, impaginazione, metadattazione e, infine, di pubblicazione. Per garantire un sistema completo e qualitativamente bilanciato, sarebbe preferibile orientarsi verso piattaforme informatiche integrate¹⁵, oppure adottare più piattaforme specializzate collegate tra loro e

¹⁵ <https://www.academic-publishing-services.it/rassegna-software-piattaforme-leditoria/> (ultima consultazione: 26/05/2022) raccoglie una lista aggiornata di piattaforme e servizi per riviste e monografie.

in grado di coprire l'intero flusso. Se ne deduce, quindi, che l'utilizzo di un solo Content Management System (CMS) standard, quale che sia (Wordpress, Drupal, Plone, Joomla, TYPO3, ecc.), non è in grado di soddisfare i requisiti del progetto nella loro integralità.

2.2.11. *Valutazione scientifica*

La messa a punto con successiva applicazione dei criteri di valutazione scientifica incideranno sul livello di qualità finale. Oltre ad appurare la pertinenza del contributo rispetto alle linee guida di cui il progetto del Blog di qualità si è dotato, il processo di valutazione stimerà l'importanza scientifica, l'originalità del contributo, la coerenza metodologica e l'organizzazione del materiale, nonché la chiarezza e lo stile dell'esposizione. Tuttavia, a differenza delle riviste scientifiche, che applicano di prassi la *double blind peer review* o una differente ma simile tipologia di *peer-review*, il Blog di qualità adotta una condotta diversa: la certificazione del valore scientifico ed editoriale dei contributi sarà svolta direttamente dal direttore e dal gruppo di esperti del comitato scientifico.

2.2.12. *Piattaforma, codice etico e policy di valutazione scientifica*

Il processo di valutazione sarà di norma gestito attraverso una piattaforma informatica; configurando tuttavia un flusso più agile, con procedure meno vincolanti per la *peer review*, si ottiene un notevole risparmio di tempo rispetto a quanto necessario per le riviste scientifiche. Utilizzare una piattaforma informatica garantisce anche la tracciabilità e favorisce la gestione dell'intero processo.

Il direttore, unitamente ai suoi collaboratori e agli autori, saranno soggetti a un codice etico che li impegna ad adottare delle buone pratiche in linea con gli standard internazionali (ad esempio, con quelli della Committee on Publication Ethics - COPE)¹⁶. L'adozione di policy in linea con gli standard internazionali della editoria accademica è uno degli aspetti che differenzia maggiormente il Blog di qualità dai blog tradizionali, anche quelli di tipologia scientifica segnalati in precedenza.

2.2.13. *Standard di pubblicazione e scelta di un service editoriale*

Per pubblicare un'iniziativa digitale di qualità occorre prevedere lo svolgimento di numerosi processi nella fase produttiva: controllo antiplagio, con-

¹⁶ Si tratta di linee guida redatte con lo scopo di consigliare gli editori sulle pratiche etiche di pubblicazione e gestire casi di condotta impropria nella diffusione dei risultati della ricerca scientifica: <https://publicationethics.org/> (ultima consultazione: 26/05/2022).

trollo linguistico, *editing*, controllo delle citazioni e delle altre norme editoriali, verifica e assolvimento dei diritti di riproduzione per eventuali materiali coperti da copyright di terzi, realizzazione di grafici, tabelle e immagini, impaginazione, scambio bozze con gli autori, creazione di indici e sintesi, creazione dei metadati, promozione e lancio sui social network, ecc.

Il fatto che il Blog di qualità sia un formato di pubblicazione auto-prodotto non significa necessariamente che tutti i servizi necessari per la sua realizzazione siano svolti internamente, ma solo che la proprietà è il responsabile dei contenuti e della valutazione vengono a coincidere.

Per garantire un livello alto di qualità costante nel tempo, è da preferirsi che i servizi editoriali, tecnici e a base informatica, che non rientrano quindi usualmente nelle competenze specifiche del gruppo responsabile dell'iniziativa, siano svolti da un *service* editoriale esterno selezionato dalla proprietà. La scelta del fornitore è un aspetto determinante per il posizionamento di qualità e la sostenibilità del progetto editoriale nel tempo.

Il raggiungimento della qualità stabilita passa senz'altro dalla sinergia tra la redazione interna (che sia strutturata o rappresentata dal solo coordinatore) e il *service* editoriale scelto; affidarsi a fornitori esterni non è sinonimo di perdita di controllo sulla lavorazione, ma piuttosto di smistamento delle attività ai giusti interlocutori e referenti.

2.2.14. Norme editoriali e formati di pubblicazione

Sarà predisposto un documento da diffondere e consegnare agli autori per la messa a conoscenza di norme editoriali specifiche da seguire che definiranno le linee guida editoriali della testata; previa accettazione, i materiali da pubblicarsi saranno comunque sottoposti a revisione editoriale esterna, secondo le stesse linee guida divulgate ai potenziali autori.

Il formato di pubblicazione nativo del Blog di qualità è HTML, anche se è da prevedersi una versione PDF generata possibilmente in automatico dalla versione HTML stessa e che contenga la numerazione progressiva delle pagine – requisito, quest'ultimo, essenziale non solo per la citabilità secondo le norme dell'Agenzia nazionale di valutazione del sistema universitario e della ricerca (ANVUR), ma anche per l'assegnazione del codice ISSN. Laddove ci fossero risorse, si potrebbe inoltre prevedere un flusso di produzione a partire da XML (JATS) da cui generare automaticamente gli altri formati digitali, come già accade d'altronde per le riviste scientifiche più prestigiose. In alternativa, e qualora ciò non fosse possibile per carenza di risorse, l'impaginazione avverrà a cura del *service* editoriale esterno, direttamente in HTML.

2.2.15. Licenze e diritti

Agli autori dei contributi sarà lasciato il copyright e si chiederà loro che siano concessi i diritti per la pubblicazione, utilizzando una licenza CC. In particolare, si favorirà l'adozione della licenza CC-BY che favorisce la diffusione e il riutilizzo, secondo le più recenti direttive della *Open Science* (Margoni e Tsiavos 2018).

2.2.16. Diffusione e valorizzazione

Per garantire il massimo della visibilità e della diffusione in rete, il Blog di qualità adotterà i principali identificativi persistenti standard della editoria scientifica – Open Researcher and Contributor ID (ORCID), Digital Object Identifier (DOI), International Standard Name Identifier (ISNI), ISSN – e gestirà un ampio spettro di profili di metadati (Smith-Yoshimura 2020): grazie ad essi, infatti, unitamente alla contemporanea assegnazione sistematica del DOI a tutti i contributi, si potrà consentire la presenza di questi ultimi all'interno dei principali aggregatori e banche dati della ricerca scientifica. Accanto ai profili dei metadati adottati nella editoria scientifica (*Dublin Core* e *Highwire Press*), ai fini della diffusione sui social e i principali motori di ricerca generalisti, si invita all'adozione anche dei profili *Open Graph*, *Twitter* e *Schema.org*.

2.2.17. Accesso gratuito

Il Blog di qualità, tranne rare eccezioni, si configura come un'operazione gratuita nel cui ambito i contributi sono accessibili liberamente a tutti e soggetti alle policy di riutilizzo in base alle licenze adottate. In questo contesto applicare o meno la dicitura *Open Access* è discutibile, in quanto, come detto in precedenza, la valutazione scientifica nel Blog di qualità è svolta non attraverso un processo di *peer-review* standard.

2.2.18 Archiviazione e preservazione a lungo termine

Grazie all'adozione di standard internazionali (metadati in formato XML, formato PDF e HTML per il *full-text*, utilizzo di licenze CC) ci sono le condizioni perché i materiali e contributi pubblicati sul Blog di qualità possano essere archiviati facilmente e gratuitamente su repository istituzionali; può essere implementato anche un sistema di preservazione a lungo termine, utilizzando servizi specializzati (ad es. Clockss)¹⁷.

¹⁷ <https://clockss.org/> (ultima consultazione: 26/05/2022).

3. Risultati, in pratica e in prospettiva

Il Blog di qualità si propone come un nuovo format per la comunicazione e diffusione di contributi scientifici che, senza differenziarsi troppo da un tradizionale articolo pubblicato all'interno di una rivista scientifica, può garantire analoghi livelli di qualità e di impatto a costi inferiori di realizzazione, grazie alla semplificazione delle procedure messe in atto nell'ambito del *workflow* di pubblicazione appositamente configurato.

Questo format può essere scelto come strumento privilegiato di comunicazione in diversi contesti.

Il primo che si può ipotizzare guarderebbe al Blog di qualità come progetto pilota o primo *step* strategico in vista della realizzazione di una rivista scientifica. Per fondare una rivista occorrono non solo molti prerequisiti formali, ma anche un arco temporale di almeno un paio di anni che permetta la definitiva messa a regime dell'iter di pubblicazione. Le difficoltà iniziali insite nel riconoscimento della qualità scientifica della rivista e i complessi percorsi di accreditamento (ANVUR¹⁸ e/o assegnazione dell'*Impact Factor*)¹⁹ richiedono alcuni anni di pubblicazioni prima del riconoscimento ufficiale. Potrebbe quindi essere una valida alternativa partire con un Blog di qualità da convertire, dopo un anno o due, in rivista scientifica. Ma come può il Blog di qualità dimostrare la *compliance* con delle procedure di funzionamento e gestione di una rivista scientifica pur nascendo esso come prodotto alternativo, sia nei propositi che nella forma?

I requisiti tecnici ed editoriali ANVUR sono soggetti a modifiche ed è probabile che nei prossimi anni le regole di valutazione si facciano meno stringenti, se non altro per assecondare i cambiamenti di un panorama scientifico internazionale in rapida evoluzione. Se anche i requisiti non cambiassero, due possibili opzioni sono disponibili:

1. si potrebbe procedere a un'attività esplicita e tracciata di *peer review*, sempre secondo requisiti ANVUR, anche successivamente alla pubblicazione (*post-publication-peer-review*) ma precedente alla domanda di accreditamento; assegnando e registrando un DOI a ciascun referaggio²⁰, è inoltre possibile ag-

¹⁸ In un contesto scientifico che valorizza sempre di più articoli pubblicati in riviste scientifiche che adottino sistemi accertati di selezione e valutazione dei contributi, è facile immaginare quanto cruciale sia la fase di accreditamento del periodico in termini di posizionamento e soddisfacimento di standard editoriali di qualità, a livello non solo nazionale ma anche internazionale: <https://richiestarevisioneriviste.cineca.it/anvur/> (ultima consultazione: 26/05/2022).

¹⁹ Originariamente pubblicato nel 1994, l'approfondimento di *Clarivate Analytics*, allora conosciuto come *The Institute for Scientific Information*, pone le basi (anche teoriche) dell'applicabilità dell'*Impact Factor* nella reputazione della ricerca scientifica: <https://clarivate.com/webofsciencegroup/essays/impact-factor/> (ultima consultazione: 26/05/2022).

²⁰ Possibilità gestibile secondo il più recente schema 5.3.1. di Crossref: <https://www.>

ganciare la valutazione all'articolo, certificandone anche la data di svolgimento (grazie ai metadati previsti dallo schema per la registrazione del DOI);

2. si potrebbe in alternativa percorrere la strada della richiesta di inserimento nelle Banche dati WoS (*Web of Science*) o Scopus che garantiscono, ai sensi dell'art 13.2 del regolamento²¹, l'ammissibilità d'ufficio negli elenchi delle riviste accreditate.

Un secondo caso di applicazione del Blog di qualità può veder coinvolta una società scientifica interessata a pubblicare le proprie riflessioni, conferenze o atti di giornate di studio. Laddove l'istituzione non voglia disperdere la conoscenza e il patrimonio di riflessioni sulle tematiche di cui si occupa, ecco che diviene opportuno – se non altamente raccomandato – costruire un ambiente idoneo a tale conservazione e diffusione: più che un sito internet istituzionale o un semplice *repository*, il Blog di qualità può costituire un contenitore appropriato a un simile contesto di applicazione.

In terzo e ultimo luogo, il Blog di qualità si presta particolarmente come format di pubblicazione adatto a iniziative di comunicazione ibride tra il mondo universitario e quello professionale: si pensi ad esempio agli albi professionali, alle professioni sanitarie e alle associazioni di categoria, che non hanno la forza accademica e la struttura interna per fondare una propria rivista ma che, allo stesso tempo, hanno necessità di dare voce alla propria posizione culturale e scientifica nel mondo della ricerca. La natura ibrida del Blog di qualità aiuta anche alla pubblicazione di contributi provenienti più da una esperienza pratica sul campo che da riflessioni teorico-speculative, come spesso accade proprio nel mondo associativo professionale.

4. Conclusioni

Le specifiche passate in rassegna nel corso dell'articolo, partendo dai promotori per finire con i fattori di valorizzazione e diffusione, fanno del Blog di qualità un progetto editoriale intermedio tra il blog e la rivista scientifica: come si è visto, da quest'ultima esso eredita la gran parte delle caratteristiche ma non il tipico *workflow* editoriale, che nel caso del format in questione risulta semplificato e tale da consentire un risparmio dei costi di gestione. Proprio alla luce dei vantaggi offerti dal suo iter gestionale, il Blog di qualità emerge come una valida via percorribile anche per realtà che, per quanto piccole e meno strutturate, sono pur sempre interessate a promuovere un progetto di

crossref.org/documentation/content-registration/content-types-intro/peer-reviews/ (ultima consultazione: 26/05/2022).

²¹ https://www.anvur.it/wp-content/uploads/2019/02/REGOLAMENTO-PER-LA-CLASSIFICAZIONE-DELLE-RIVISTE_20022019.pdf (ultima consultazione: 26/05/2022).

editoria accademica e comunicazione scientifica tanto nel settore delle STM che in quello delle HSS. In prospettiva futura, particolare attenzione andrà posta nella messa a punto del *workflow* e delle procedure operative e organizzative proprie del Blog di qualità: in attesa di raccogliere un numero sufficiente di *case studies*, possiamo per ora affermare che il bilanciamento tra la semplificazione e la qualità attesa sarà un elemento decisivo per il successo di questo format nel tempo.

Ringraziamenti

Si ringraziano Eleonora Colangelo e Francesco Sensoli per l'aiuto nella revisione del testo.

Riferimenti bibliografici

- Aczel, Balazs, Barnabas Szaszi, and Alex O. Holcombe. 2021. "A billion-dollar donation: estimating the cost of researchers' time spent on peer review." *Res Integr Peer Rev* 6, no. 14: 1-8. <https://doi.org/10.1186/s41073-021-00118-2>.
- Ancion, Zoé, Lidia Borrell-Damián, Pierre Mounier, Johan Rooryck, and Bregt Saenen. 2022. "Action Plan for Diamond Open Acces." <https://doi.org/10.5281/zenodo.6282402>.
- Brembs, Björn, Philippe Huneman, Felix Schönbrodt, Gustav Nilsson, Toma Susi, Renke Siems, Pandelis Perakakis, Varvara Trachana, Lai Ma, and Sara Rodriguez-Cuadrado. 2021. "Replacing academic journals." <https://doi.org/10.5281/zenodo.5526635>.
- Eve, Martin Paul, Cameron Neylon, Daniel Paul O'Donnell, Samuel Moore, Robert Gadie, Victoria Odeniyi, and Shahina Parvin. 2021. *Reading Peer Review: PLOS ONE and Institutional Change in Academia*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108783521>.
- Guatelli, Fulvio. 2016. "Lo spazio economico di una *open access journal platform*." In *Open Access e scienze umane. Note su diffusione e percezione delle riviste in area umanistica*, edited by Luca Scalco, 43-53. Milano: Ledizioni.
- Guerrini, Mauro. 2022. "Sua Maestà il revisore: alcune considerazioni sul processo di peer-review all'interno della LIS." *AIB Studi* 61 (3): 585-92. <https://doi.org/10.2426/aibstudi-13328>.
- Hames, Irene. 2014. "The Changing Face of Peer Review." *Science Editing* 1 (1): 9-12. <https://doi.org/10.6087/kcse.2014.1.9>.

- Kyrillidou, Martha. 2012. "Research Library Trends: an historical picture of services resources, and spending." In *Research Library Issues: A Quarterly Report from ARL, CNI, and SPARC*, Washington, edited by Association of Research Libraries 280, 20-27. <https://doi.org/10.29242/rli.280.4>.
- Larivière, Vincent, Stefanie Haustein, and Philippe Mongeon. 2015. "The Oligopoly of Academic Publishers in the Digital Era." *PLOS ONE* 10 (6). <https://doi.org/10.1371/journal.pone.0127502>.
- Margoni, Thomas, and Prodromos Tsiavos. 2018. "Toolkit for Researchers on Legal Issues."
- Rousi, Amtti Mikal, and Mikael Laakso. 2022. "Overlay journals: a study of the current landscape." arXiv:2204.03383. <https://doi.org/10.48550/arXiv.2204.03383>.
- Salucci, Giovanni, e Roberto Cinotti. 2021. "Open Access e studi umanistici." *Studi Italiani* 65 (1): 159-84.
- Smith-Yoshimura, Karen. 2020. "Transitioning to the Next Generation of Metadata." Dublin, OH: OCLC Research. <https://doi.org/10.25333/rqgd-b343>.
- Urbano, Cristóbal, Sara Tafalla, Ángel Borrego, and Ernest Abadal. 2021. "Preprints as an alternative to conference proceedings: A hands-on experience at EDICIC Iberian Meeting 2019." *Learned Publishing* 34: 558-67. <https://doi.org/10.1002/leap.1402>.

Note e rubriche

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

ROBERTO GUARASCI*

La lettura del libro di Paola Castellucci e Sara Mori (Castellucci e Mori 2022) su Suzanne Briet oltre a offrire ai lettori una traduzione italiana del testo *Qu'est-ce que la Documentation* (Briet 1951) presenta una serie di considerazioni sul contesto nel quale il testo e il pensiero della Briet maturarono e si formarono e lascia di sfondo, pur accennandone a più riprese, per gli ovvi motivi derivanti dalle dimensioni del testo, gli interrogativi di più generale portata relativi sia all'evoluzione della documentazione ma, ancor più, all'evoluzione del termine documento la cui corretta definizione è – ancora oggi – non compiutamente esplicitata all'interno del generale e composito panorama delle scienze del documento e dell'informazione. Nel 1997 Michael Buckland pubblica un breve testo dal titolo *What is a Document?* nel quale comparando le definizioni di documento di Otlet e della Briet con quelle successive dei teorici delle scienze dell'informazione rileva come, pur con qualche diversa accentuazione, esse fossero molto vicine al concetto antropologico di cultura materiale o a quello di oggetto come segno della semiotica. Più avanti puntualizza che un documento convenzionale, come un messaggio di posta o un rapporto tecnico, esiste fisicamente nella tecnologia digitale come una stringa di bit, ma lo stesso vale per tutto il resto in un ambiente digitale. In questo senso, qualsiasi caratteristica distintiva di un documento come forma fisica è ulteriormente ridotta e la discussione su “cos'è un documento digitale?” diventa ancora più problematico se non ricordiamo il percorso del ragionamento alla base delle discussioni in gran parte dimenticate sugli oggetti di Otlet e sull'antilope di Briet¹.

* Laboratorio di Documentazione, Dipartimento di Culture, Educazione e Società, Università della Calabria, Rende (CS), Italia. roberto.guarasci@unical.it.

¹ «The evolving notion of “document” among Otlet, Briet, Schürmeyer, and the other documentalists increasingly emphasized whatever functioned as a document rather than traditional physical forms of documents. The shift to digital technology would seem to make this

Vent'anni dopo, nel 2017, Luciana Corts Mendes riprende lo stesso filo del discorso provando a leggere sincronicamente le definizioni di documento di Otlet e Briet ad esse aggiungendo quelle di Jean Meyrat e Roger T. Pédauque² ed esaminando anche i contesti nei quali esse si formarono e come dopo un lungo periodo di oblio le nazioni non francofone «turn their attention to document theory and Documentation Science in order to deal with KO and LIS's issues» (Corts Mendes 2017). Questa possibilità di una rilettura dell'opera della Briet in un contesto – come quello attuale – nel quale la definizione di documento è centrale nella costruzione di un fondamento epistemologico delle varie scienze del documento e dell'informazione è il filo conduttore del libro e forse l'elemento di maggior pregio anche in virtù delle integrazioni e divagazioni colte – forse a volte sovrabbondanti – che contribuiscono a definirne le complesse intersezioni culturali. La notazione che la documentazione può essere configurata come una tecnica culturale offre uno spunto interessante per cercare di superare il problema posto nel 1995 dalla mappa epistemologica di Ingwersen che nel definire le interconnessioni tra le scienze dell'informazione collocava al livello dell'interdisciplinarietà le scienze della comunicazione, della cognizione e dei sistemi mentre collocava al livello applicativo la documentazione e la biblioteconomia che venivano sostanzialmente configurate come delle mere tecniche. Se, però, ad ogni tecnica è associata una scienza che ne determina i fondamenti teorici e concettuali che la prima si limita a rendere operativi quale potrebbe essere la scienza alla quale documentazione e biblioteconomia e più in generale le scienze del testo e del documento si riferiscono? Pur rimarcando a più riprese la differenza tra documentalisti e bibliotecari, fino ad arrivare all'affermazione che il documentalista non è un bibliotecario specializzato ma bensì è il bibliotecario ad essere un documentalista che si occupa di una particolare tipologia di documenti, la scienza di riferimento per Briet è quella nascente delle informazioni della quale in più occasioni tratteggia delle immagini futuribili e della quale la documentazione è un *minus inter pares*, appunto una tecnica culturale. In questo anche lo scostamento da Otlet per il quale – nella visione teleologica finale – la documentazione diventerà una scienza universale nella quale ogni cosa sarà codificata. La visione di Briet

distinction even more important. Levy's thoughtful analyses have shown that an emphasis on the technology of digital documents has impeded our understanding of digital documents as documents (e.g. Levy 1994). A conventional document, such as a mail message or a technical report, exists physically in digital technology as a string of bits, but so does everything else in a digital environment. In this sense, any distinctiveness of a document as a physical form is further diminished and discussion of "What is a digital document?" becomes even more problematic unless we remember the path of reasoning underlying the largely forgotten discussions of Otlet's objects and Briet's antelope» (Buckland 1997).

² Roger T. Pédauque è lo pseudonimo usato dai partecipanti a RTP-doc, gruppo di ricerca del Centre National de la Recherche Scientifique (CNRS) sui documenti digitali (Sultan e Salaün 2010).

è laica, concreta, per obbiettivi raggiungibili e non utopistici, anche in considerazione del lasso di tempo che la separa dal bibliografo belga. Per il resto i due hanno molto in comune: la fiducia nella standardizzazione, la fiducia nella cooperazione internazionale e la convinzione del valore etico e culturale delle nuove tecnologie in termini di capacità di acquisire conoscenza. Ma tra l'omo documentor della Briet e il Documentalista superuomo di Otlet la differenza è netta e puntuale.

Il 1951, anno di pubblicazione del testo della Briet, è anche l'anno nel quale in Italia si tiene, presso il Consiglio Nazionale delle Ricerche (CNR), il XVIII congresso mondiale di Documentazione (La Documentazione in Italia 1952) al quale la Briet partecipa come relatrice nella sezione sull'Insegnamento professionale per documentalisti. In netto contrasto con quanto poi ella dirà il presidente della sessione, Bruno Balbis, nel delineare lo scenario nazionale e internazionale ripropone una visione estremamente restrittiva della disciplina, una bibliografia specializzata, specificando che «allo stato attuale delle cose sarebbe opportuno che l'insegnamento della Documentazione in Italia penetrasse per gradi, evitando che i suddetti programmi, a vasto respiro avessero effetti controproducenti» (La Documentazione in Italia 1952, 72). Balbis è il direttore del Centro Nazionale di Documentazione Scientifica Tecnica del Consiglio Nazionale delle Ricerche³ che aveva avuto un ruolo importante nella diffusione dell'informazione tecnico-scientifica nel Paese in organico collegamento con l'Istituto Internazionale di Bibliografia dal quale mutua metodi e spesso acquisisce le schede ma, ciononostante, non riesce ad affrancarsi dalle critiche mosse da decenni a Paul Otlet⁴ tanto che, per cercare di trovare una qualche possibilità di discostarsi da quel filone, propone di cambiare nome alla disciplina da documentazione in documentologia «per giungere ad una maggiore chiarezza dei limiti e delle finalità scientifiche della nuova scienza della documentazione» (La Documentazione in Italia 1952, 72).

Per Suzanne Briet, come correttamente notano le autrici, la formazione è invece un momento centrale, propedeutico alla definizione del ruolo che il documentalista deve svolgere nel panorama culturale e, partendo dalla constatazione dell'inadeguatezza dei percorsi all'epoca in essere, ella ne definisce e postula uno specifico anche in virtù di una situazione sostanzialmente privilegiata come quella francese (Fayet Scribe 2001, 123 e sgg).

³ Sulla Storia del Centro Nazionale di Documentazione Scientifica del CNR si veda (Guarasci 2011).

⁴ Sull'argomento si veda (Fumagalli 1906). Anche in occasione della redazione da parte di Otlet del sistema di classificazione dell'Istituto Internazionale di Agricoltura il segretario generale, Pasquale Iannaccone, specificherà «che se è pur vero che l'Istituto ha tratto giovamento dalla visita di Otlet alcune delle migliori suggerite erano già preventivate, altre risultano inapplicabili e, solo una parte, è stata prodotta su specifica sollecitazione» (Guarasci e Taver-niti 2013, 55).

In chiusura del volume il richiamo a Ronald Day e alla sua valutazione di Briet come momento di cerniera tra la le scienze del testo e del documento e le scienze dell'informazione tocca un punto nodale del dibattito sul rapporto tra le scienze che meriterebbe una più approfondita disamina su base nazionale non potendosi riscontrare, se non in grandi linee, una comune direttrice di evoluzione a livello sovranazionale.

Riferimenti bibliografici

- Briet, Suzanne. 1951. *Qu'est-ce que la Documentation ?*. Paris : Editions documentaires, industrielles et techniques.
- Buckland, Michael Keeble. 1997. "What is a "document"?". *Journal of the American Society for Information Science* 48, 804-9.
- Castellucci, Paola, e Sara Mori. 2022. *Suzanne Briet nostra contemporanea*. Milano: Mimesis.
- Corts Mendes, Luciana. 2017. "The francophone development of the concept of document: the works of Paul Otlet, Suzanne Briet, Jean Meyriat and Roger T. Pédauque." In *Fondements épistémologiques et théoriques de la science de l'information-documentation : Actes du 11e colloque ISKO France 2017 11-12th July 2017 Paris*, sous la direction de Widad Mustafa El Hadi. Londres : ISTE éditions, 180-90.
- Fayet Scribe, Sylvie. 2001. *Histoire de la Documentation en France*. Paris : CNRS Editions.
- Fumagalli, Giuseppe. 1906. "La Conferenza internazionale bibliografica di Bruxelles e il repertorio bibliografico universale." *Rivista delle Biblioteche e degli Archivi*, anno VI, no. 9-10, 129-50.
- Guarasci, Roberto. 2011. "La Memoria della Scienza: L'Archivio Tecnico Italiano e il Centro Nazionale di Documentazione Scientifica." In *Archivi Privati*, a cura di Roberto Guarasci, Erika Pasceri, 195-218. Roma: CNR.
- Guarasci, Roberto, e Maria Taverniti. 2013. *Paul Otlet, il Rapport per l'Istituto Internazionale di Agricoltura*, Roma: Aracne.
- La Documentazione in Italia. Atti del Congresso Mondiale di Documentazione*. 1952. Roma: CNR.
- Sultan, Frederic, et Jean Michel Salaün. 2010. "Roger T. Pédauque, l'aventure d'une écriture collective." *dialogues, propositions, histoires pour une citoyenneté mondiale*, no. 3, <http://base.d-p-h.info/fr/fiches/dph/fiche-dph-8220.html>.

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

Erika Pasceri*

La pubblicazione del *Nuovo Glossario di biblioteconomia e scienza dell'informazione* di Ferruccio Diozzi (Diozzi 2021) giunge in un periodo storico in cui le Scienze dell'informazione acquisiscono sempre più forza nella loro trasversalità di utilizzo e applicazione. L'autore in questa edizione aggiornata del *Glossario di biblioteconomia e scienza dell'informazione* (Diozzi 2003) pubblicato per la prima volta nel 2003 integra 850 nuovi lemmi con l'intento di guidare gli utenti nell'interpretazione dell'evoluzione culturale che ha condizionato – per alcuni versi – ma aggiornato per altri - le discipline cui fa riferimento.

Alcune tematiche - come evidenziato dallo stesso autore – sono di particolare rilievo ed estremamente attuali per la loro complessità e rilevanza nello scenario normativo. Basti pensare come in tempi di *smart working*, sia aumentato il *digital divide* (o divario digitale) e siano stati incrementati i processi di *dematerializzazione*, ma soprattutto di *digitalizzazione* e si sia reso necessario l'utilizzo di strumenti di validazione quale la *firma digitale*.

Un altro aspetto cruciale in quest'ultimo millennio è l'avvento dei social media quali strumenti di informazione collettiva, che stanno diventando uno strumento di diffusione delle notizie, delle quali non sempre viene verificata l'autenticità e la veridicità (*fact checking*).

Se da una parte il legislatore si sta muovendo velocemente per delineare un quadro normativo chiaro e completo, dall'altra la rapida evoluzione delle tecnologie informatiche - come è noto - sta in qualche modo stravolgendo tutti i settori e tutti i relativi contesti applicativi, quindi anche in questo senso le precisazioni terminologiche rispetto a cosa “non esiste più” o rispetto alle “nuove funzioni” risultano – ancora una volta – fondamentali. È il caso specifico del DPO (acronimo di *Data Protection Officer*) figura introdotta dal *Regolamento Generale sulla Protezione dei Dati* n. 679 del 2016 che ha come responsabilità principale

* Laboratorio di Documentazione, Dipartimento di Culture, Educazione e Società, Università della Calabria, Rende (CS), Italia. erika.pasceri@unical.it.

quella di assicurare che la gestione del trattamento di dati personali (e dunque la loro protezione) all'interno di un'azienda (pubblica che privata), sia trattata nel rispetto delle normative sulla privacy europee e nazionali. Per quanto riguarda la precisazione effettuata rispetto ad alcuni termini relativi a delle attività specifiche, non si può non evidenziare quella della *gestione documentale*, in precedenza *gestione dei documenti*, che non è più una semplice «attività di back office che si sviluppa dalla scelta del documento da acquisire e/o da accedere e termina con il suo scarto, attraversando tutti i passaggi di trattamento dell'informazione» (Diozzi 2021), ma ora vero e proprio processo di *gestione documentale*, normato dal Decreto Legislativo 7 marzo 2005 n. 82, altrimenti conosciuto come il *Codice dell'Amministrazione Digitale* (CAD), “finalizzato al controllo efficiente e sistematico della produzione, ricezione, tenuta, uso, selezione e conservazione dei documenti”. Forse in questo caso ci si poteva aspettare il riferimento normativo anche nel termine *documento* e che – tra le tante accezioni proposte dall'autore – ci fosse anche il termine “documento informatico”, che nel CAD riveste il ruolo di protagonista assoluto in ambiente digitale.

La lingua come è noto segue le normali e fisiologiche evoluzioni culturali e pertanto è destinata a crescere e ad alimentarsi continuamente con nuovi termini individuati dalle necessità di utilizzo e dal contesto in cui ci si muove. Magari in un futuro aggiornamento, per alcuni termini si potrebbe pensare di citare altre risorse che trattano le stesse tematiche e che sono di riferimento per alcuni ambiti della *gestione documentale* come ad esempio le *Linee Guida sulla formazione, gestione e conservazione dei documenti informatici*, dove il *Nuovo Glossario* potrebbe sicuramente trovare una giusta integrazione ed un ampliamento circa i contenuti specifici soprattutto nelle parti relative alle precisazioni terminologiche in esso contenute.

In conclusione, l'ambizione dell'autore sul voler creare uno strumento di supporto per gli studi nei domini disciplinari di riferimento con il *Nuovo Glossario di Biblioteconomia e Scienza dell'informazione* ad oggi ha sicuramente raggiunto il suo obiettivo grazie ad un lavoro di dettaglio che rispecchia la multidisciplinarietà, ma anche la tradizione citando come giusti riferimenti le opere che hanno segnato e da sempre continuano ad essere un modello assoluto per tutti gli addetti del settore, senza snaturare le principali discipline direttrici nelle loro peculiarità e specificità.

Riferimenti bibliografici

- Diozzi, Ferruccio. 2003. *Glossario di Biblioteconomia e Scienze dell'informazione*. Milano: Editrice Bibliografica.
- Diozzi, Ferruccio. 2021. *Nuovo Glossario di Biblioteconomia e Scienze dell'informazione*. Milano: Editrice Bibliografica.

Come mi vuoi, canonica o rivoluzionaria?

Claudio Gnoli*

Il Collegio Borromeo di Pavia ha recentemente organizzato e pubblicato sul suo canale YouTube un ciclo di quattro seminari che, facendo seguito ad altri svolti da Federico Laudisa, muovono “Dalla filosofia della scienza ai fondamenti della fisica”. Li tiene il professor Gianluca Introzzi, che abbiamo già avuto modo di citare in questa rubrica per la sua attenzione al ruolo dei concetti su cui basiamo le nostre teorie sul mondo. Nozioni come quelle di spazio-tempo, di etere o di onda-particella, discusse ciascuna in uno dei seminari, svolgono infatti dei ruoli chiave nell’evoluzione delle rispettive teorie scientifiche.¹

L’*etere*, per esempio, era un concetto necessario per far funzionare il modello aristotelico delle sfere celesti; una volta passati al sistema copernicano, il suo nome è rimasto con significati aggiornati per indicare il mezzo in cui si muovevano gli astri, o almeno la luce. Con l’avvento della teoria della relatività è prevalso un modello che non ha più bisogno dell’etere: Einstein infatti ne fa a meno, considerandola un’ipotesi non necessaria secondo il principio del rasoio di Occam. Nella fisica contemporanea dunque il concetto di etere non c’è più.

Lo stesso destino hanno avuto in passato altri concetti, come quello del flogisto in chimica: essi supponevano l’esistenza di entità che, con l’avvento di nuove teorie, si sono rivelate non necessarie. Questa constatazione ci mostra come i concetti possano essere legati a specifiche teorie, che nel tempo li hanno introdotti o che a un certo punto li aboliscono. La fisica delle particelle ha portato con sé nuovi concetti come quelli di bosone di Higgs e di onde gravitazionali, messi peraltro in dubbio da modelli alternativi. In questo caso, gli esiti di alcuni esperimenti degli ultimissimi anni sono risultati essere in accordo con l’esistenza di entrambe le entità: Popper direbbe che quelle ipotesi sono state *corroborate*.

La sostituzione di una teoria con un’altra è stata discussa in modo particolare da alcuni epistemologi novecenteschi. Norwood Hanson ha scritto che «gli ele-

* Biblioteca della scienza e della tecnica, Università degli Studi di Pavia, Pavia, Italia.
claudio.gnoli@unipv.it.

menti delle loro esperienze sono identici, ma la loro organizzazione concettuale è enormemente diversa. C'è dunque un senso in cui il semplice fatto di vedere è in realtà un'impresa carica di teoria [*theory laden*]» (Hanson 1958). Hanson morì solo 42enne sul suo aereo militare, ma influenzò il classico lavoro di Thomas Kuhn *La struttura delle rivoluzioni scientifiche* (Kuhn 1962), spesso citato nei lavori sull'organizzazione della conoscenza per sottolineare come i concetti siano sempre legati a determinati *paradigmi* e suscettibili di revisione. Alcuni spingono queste considerazioni nella direzione di un forte relativismo, lasciando intendere che nessun sistema di organizzazione della conoscenza (KOS) sia migliore degli altri e tutto dipenda dalle specifiche culture. In realtà Kuhn non si è espresso in modo così relativistico, in particolare negli ultimi anni prima della scomparsa, nei quali stava spostando la sua attenzione dai paradigmi alle *tassonomie* dei concetti che stanno alla base delle teorie nuove. Ce n'è traccia nel libricino pubblicato in italiano col titolo *Le rivoluzioni scientifiche* (Kuhn 2008).

Quest'ultimo tema viene sviluppato anche dallo psicologo canadese Paul Thagard. Il suo testo *Rivoluzioni concettuali* (Thagard 1994), sebbene non troppo scorrevole nella traduzione italiana, analizza anche mediante strumenti informatici le gerarchie dei concetti che stanno alla base di alcune teorie rivoluzionarie per la loro epoca: quelle copernicana, darwiniana e einsteiniana, ma anche la teoria della combustione dell'ossigeno di Lavoisier che ha soppiantato il concetto di flogisto, o la tettonica a zolle che ha sostituito la teoria della deriva dei continenti. Thagard mostra come ciò che cambia da una teoria alla successiva non sia tanto un singolo concetto, quanto la topologia delle relazioni fra un concetto e l'altro: concetti che erano alla radice degli alberi gerarchici diventano, nella nuova prospettiva, solo particolari tipi, o parti, di un concetto più generale. Il nuovo assetto finisce per essere preferito al vecchio perché offre una migliore *coerenza esplicativa*. In altre parole, l'avvento di una teoria nuova è un cambiamento nell'organizzazione delle conoscenze.

Pensiamo a quelle che erano le classi somme nei sistemi del sapere antico: acqua, aria, terra e fuoco; ogni cosa era spiegata come derivata dalle loro combinazioni. Questi quattro concetti sono ancora presenti nei sistemi attuali, ma non più al vertice delle loro gerarchie. L'acqua appartiene alla classe dei composti chimici, l'aria e la terra a quella degli aggregati, il fuoco non è che una proprietà delle reazioni di combustione: le classi di riferimento sono ora gli elementi quali idrogeno e ossigeno, inoltre esistono classi più fondamentali come le particelle elementari. Sono dunque cambiate le relazioni fra concetti e gli alberi gerarchici che li rappresentano.

Come sappiamo, i KOS sono appunto formati da concetti e da relazioni fra concetti, a cominciare appunto dalle relazioni gerarchiche genere-specie e tutto-parte. Il campo dell'organizzazione della conoscenza avrebbe dunque un ruolo chiave nello studiare l'evoluzione della scienza stessa, anche se non è ancora riconosciuto esplicitamente in questi termini – gli stessi Kuhn e Thagard

parlano di tassonomie senza particolari riferimenti al nostro settore. Si può invece considerare le modifiche di concetti e loro relazioni che punteggiano la storia della scienza come una storia dell'organizzazione della conoscenza.

I concetti sono espressi da *termini*, ma come ci avverte Thagard essi vanno considerati non in modo isolato bensì nel quadro di una certa tassonomia. Questo significa che, con il mutare della tassonomia, cambia anche il significato del termine. L'*etere* di Aristotele non è la stessa cosa dell'*etere* dell'astronomia di Tycho Brahe, che a sua volta è diverso da quello dell'elettromagnetismo di Maxwell (per non parlare della relatività di Einstein che ne fa a meno del tutto). Ciò pone un problema cruciale alle tecniche di recupero dell'informazione, in particolare quelle contemporanee digitali che si basano esclusivamente sulla presenza o assenza di singole parole: infatti un testo cinquecentesco sull'*"etere"* non parla della stessa cosa di un testo novecentesco che usa lo stesso termine! I KOS evolvono, e lo fanno di conseguenza anche i loro termini, per cui è necessario occuparsi anche di quella che Joe Tennis ha chiamato "subject ontogeny". Ne ha fornito altri esempi il grande storico Marc Bloch nel suo classico *Apologia della storia*, notando che termini come *patrizi* o *feudi* vanno interpretati in modi diversi a seconda dei contesti dell'epoca.

Queste considerazioni ci possono portare a mettere in discussione anche le stesse categorie su cui basiamo i nostri KOS. Pensiamo ai concetti di materia, spazio o tempo che compaiono anche fra le categorie fondamentali nell'analisi a faccette di Ranganathan e del Classification Research Group. Ogni soggetto, siamo soliti pensare, si può analizzare per i suoi aspetti materiali, per quelli spaziali, per quelli temporali e così via. Ma anche materia, spazio e tempo sono concetti legati a certe teorie fisiche.

Nella fisica relativista, spazio e tempo si fondono nel concetto di *spaziotempo*, che è anch'esso un'entità e non più una sorta di sfondo rispetto al quale ogni cosa può essere misurata (lo sfondo che un tempo veniva immaginato come riempito del fantomatico etere) tanto è vero che esso può deformarsi per velocità vicine a quella della luce. Esisteranno quindi documenti che trattano dello spaziotempo, ma anche documenti che trattano di entità distinte da esso. Le faccette di spazio e di tempo non sono allora categorie assolute, ma si possono applicare solo a certi fenomeni alle scale consuete – ai pianeti o ai telefoni, ma non ai fotoni... Per questo motivo, l'Integrative Levels Classification sostituisce le tradizionali categorie di tempo e spazio con quelle più fondamentali di posizione sequenziale (1) e situazione (2), che poi per la maggior parte dei fenomeni alle scale consuete si traducono in faccette temporali e spaziali.

Altrettanto scontata è per noi l'idea di materia, la cui quantità è misurata dalla grandezza fisica chiamata *massa*, su cui ci basiamo quando diciamo che un certo oggetto è fatto di pezzi di certi costituenti materiali: un martello è fatto di ferro, un cristallo è fatto di atomi eccetera. Ma le teorie di inizio Novecento hanno riferito la materia a sistemi più generali (prima l'elettromagneti-

smo e poi la relatività), portando alcuni studiosi a descrivere la massa come un epifenomeno. La celebre formula di Einstein mostra come la massa equivalga ad energia, e le teorie contemporanee comprendono anche onde-particelle prive di massa, come i fotoni e i gluoni. Non ha allora senso chiedere quali siano i costituenti materiali di un fotone, dal momento che esso non è materia.

Le nostre lingue, da cui derivano anche i termini che compongono le tassonomie e i tesauri, sono in fin dei conti nate dalla percezione comune alla scala della vita umana, nella quale spazio, tempo e materia hanno perfettamente senso. Molte preposizioni e avverbi che usiamo anche con significati astratti sono metafore spaziali: qualcosa va “avanti”, un elemento sta “sopra” un altro e così via. Questo però non significa che possiamo generalizzare tali categorie in modo assoluto: quando arriviamo a trattare di fenomeni più fondamentali le dovremmo abbandonare, per non commettere violazioni ontologiche (chiedere qual è il materiale di un fotone sarebbe come chiedere qual è il gruppo sanguigno di un ghiacciaio). Anche in questo caso, la categoria di materiale andrà generalizzata con una più astratta, come quella di componente: un'equazione o un poliedro possono avere dei componenti ma non dei materiali.

Le ricerche di filosofia della scienza e i nostri esempi ci hanno mostrato come i KOS siano strettamente correlati alle teorie scientifiche, alla competizione fra loro e alla sostituzione di una teoria con un'altra. Non stiamo necessariamente parlando dei KOS che sono stati formalizzati e pubblicati come tali per servire al recupero di informazioni nelle basi-dati o nelle biblioteche; ma anche dei KOS impliciti che si possono trovare in forma diffusa nella letteratura e nella terminologia delle scienze, e che i filosofi della scienza fanno emergere.

Come abbiamo scritto nel n. 3-4 del 2016, l'organizzazione della conoscenza non si limita quindi a riflettere passivamente delle fonti esterne in “ossequiose” tavole basate solo sul principio di garanzia bibliografica (presentando cioè solo i concetti che appaiono nella letteratura già pubblicata), ma può partecipare essa stessa alla natura delle teorie, anche fornendo nuovi modelli alternativi, ed è quindi in fin dei conti al cuore degli sviluppi del sapere.

Riferimenti bibliografici

- Hanson, Norwood R. 1958. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.
- Kuhn, Thomas S. 2008. *Le rivoluzioni scientifiche*. Bologna: il Mulino.
- Kuhn, Thomas S. 1962. *La struttura delle rivoluzioni scientifiche*. Ed. italiana 1979. Torino: Einaudi.
- Thagard Paul. 1994. *Rivoluzioni concettuali. Le teorie scientifiche alla prova dell'intelligenza artificiale*. Milano: Guerini e Associati.

L'universo affascinante del colore: le sociocromie

Claudio Grimaldi*

Lo svolgimento delle *Premières Rencontres OTALF – Orientations Théoriques et Appliquées dans les études de Linguistique Française synchronique et diachronique*, organizzate dall'Osservatorio di Terminologie e Politiche Linguistiche (OTLP) dell'Università Cattolica del Sacro Cuore di Milano¹ nei giorni del 19 e 20 maggio 2022, ha rappresentato l'occasione di visitare il Museo Nazionale della Scienza e della Tecnologia Leonardo da Vinci e l'interessante mostra allestita presso il Museo dal titolo "Sociocromie. 100 anni in 25 colori", curata da Giulio Ceppi, architetto e designer, il cui scopo è quello di ripercorrere attraverso 25 cromotipi eventi di natura politica, sociale, culturale e sportiva dal Novecento fino ai giorni nostri. Analizzare alcune denominazioni ricorrenti ad aggettivi di colore approfondite dalla mostra offre lo spunto per riflettere in questa sede sulla ricchezza degli studi sul colore, che hanno contraddistinto gli ultimi decenni della ricerca scientifica e che ancora oggi si interrogano su tale affascinante concetto, sotto distinte prospettive.

Dal punto di vista scientifico, il colore costituisce un ambito di studi molto ampio e dinamico, che coinvolge distinte discipline quali la fisica, la psicologia, l'antropologia, la fisiologia e la linguistica, in particolare, per quest'ultima, il campo della semantica, della pragmatica, della linguistica cognitiva, della psicolinguistica, della traduzione e della lessicografia, dove i colori sono stati indagati ricorrendo a distinti approcci. Nel corso del Novecento gran parte dei lavori scientifici hanno perseguito l'obiettivo, da un lato, di categorizzare il concetto di colore a livello cognitivo e antropologico, e, dall'altro, di

* Dipartimento di Studi Economici e Giuridici, Università degli Studi di Napoli "Parthenope", Napoli, Italia/Presidente dell'Associazione Italiana per la Terminologia (Ass.I.Term), Roma, Italia. claudio.grimaldi@uniparthenope.it/info@assiterm91.it.

¹ Le *Premières Rencontres OTALF* sono state realizzate in collaborazione con l'Università degli Studi di Bologna, l'Università degli Studi di Napoli "Parthenope", l'Università degli Studi di Perugia e l'Università degli Studi di Verona, con il patrocinio della *Délégation Générale à la Langue Française et aux Langues de France (DGLFLF)*.

comprendere il meccanismo di funzionamento delle componenti linguistiche (principalmente aggettivi e nomi), implicate nelle denominazioni del lessico cromatico. Vari studi significativi sono degni di nota e, tra questi, nell'ambito della psicolinguistica e dell'antropologia, i (Berlin and Kay 1969) sono riconosciuti come i pionieri di una suddivisione dei termini di colore in due distinte categorie: i *basic color terms* (nel numero di 11 e comuni a tutte le lingue) e i *focal color terms*, ovvero le denominazioni linguistiche che designano le sfumature cromatiche. Successivi studi hanno parzialmente messo in discussione i risultati ai quali sono giunti Berlin e Kay, con l'inclusione o l'esclusione delle denominazioni di colore dalla lista degli universali e dalla teoria dei prototipi (Rosch Heider 1972, 1973; Wierzbicka 1996, 2005).

Nella linguistica cognitiva contemporanea, altri studi hanno indagato il rapporto tra le teorie finora citate e la pluralità delle denominazioni linguistiche che ricorrono ad aggettivi e nomi di colore, esplorando soprattutto il legame che intercorre tra tali denominazioni e le pratiche artistiche di utilizzo del colore (Dubois 2006a, 2006b). Da tali indagini è emerso che per studiare le denominazioni cromatiche da un punto di vista linguistico non è possibile fare astrazione dal contesto pratico all'interno del quale queste sono create e utilizzate dai parlanti nel discorso orale e scritto. Tale tipo di indagini ha permesso, inoltre, di esplorare la relazione che esiste tra categorie cognitive e categorie lessicali del colore, nonché i modi di rappresentazione e di realizzazione linguistica del colore stesso (Molinier 2001, 2006; Golka 2014; Krysolova 2005; Rodríguez Pedreira 2002; Grimaldi 2017, 2020). In questa ottica, non è, infine, da tralasciare la dimensione culturale e simbolica implicata nella categorizzazione cognitiva del colore (Mollard-Desfour 1998, 2000, 2002, 2005, 2008a, 2011, 2012, 2015, 2017; Pastoureau 2002, 2008, 2013, 2016) e nella creazione delle denominazioni del lessico cromatico, anche a seconda della tipologia testuale nella quale tali denominazioni sono inserite (Ripoll 2018, 2019). Tale dimensione culturale e simbolica riveste un ruolo essenziale nel campo della traduttologia dove numerose difficoltà traduttive emergono per rendere nella lingua d'arrivo le implicazioni culturali e simboliche che sono alla base della nascita di espressioni e denominazioni cromatiche nella lingua di partenza (Kristol 1978; Silvestre, Cardeira, e Villalva 2006; Mollard-Desfour 2008b).

Le "sociocromie" di Ceppi rappresentano dunque un altro esempio del fascino e dell'interesse del concetto di colore nell'ambito della ricerca scientifica, accademica o extra-accademica, che si presta ad essere indagato in quanto fenomeno culturale, sociale, linguistico e cognitivo particolarmente complesso. Ceppi precisa come le "sociocromie" siano manifestazioni attraverso il colore di un fatto a tutti noto, per quanto espresso sempre con una natura metaforica o a volte metonimica, fatto celebre che diventa riconoscibile e irripetibile proprio attraverso la connotazione offerta dal colore. I sintagmi oggetto della

mostra di Ceppi sono ormai diventati o delle categorie per indicare un gruppo di persone caratterizzato da finalità e scopi comuni – ottimi esempi sono offerti, ad esempio, dai sintagmi “colletti bianchi” e “mano nera”, risalenti ai primi dieci anni del Ventesimo secolo, che sono ormai lessicalizzati in quanto denominazioni che servono per indicare specifiche tipologie di lavoratori che non svolgono mansioni fisiche e associazioni segrete con finalità politiche o di carattere delinquenziale, o ancora dal sintagma “tute blu”, che designa un gruppo di lavoratori dipendenti che svolgono un’attività manuale – o eventi storici ben collocati dal punto di vista temporale, ormai noti a tutti e nel quale il colore svolge una funzione simbolica che rinvia alla natura positiva o negativa dell’evento stesso, come ad esempio “black friday” o “martedì nero”. L’interesse di tali sintagmi, oltre al loro potere evocativo e alla loro valenza storico-sociale, si situa anche sul piano linguistico ovvero nella maniera di combinarsi del nome e dell’aggettivo di colore, dove quest’ultimo perde talvolta la sua funzione qualificativa, facilmente interpretabile da parte del parlante in quanto in possesso di un valore meramente descrittivo, e svolge, come si è detto, un ruolo categorizzante, dove il colore serve da elemento centrale per indicare uno specifico referente. È in questa chiave di lettura che devono essere interpretati sintagmi quali “green economy”, “silver economy” o ancora “rivoluzione arancione”, dove l’aggettivo di colore non ha nessun tipo di valore descrittivo, ma si fa portatore di un significato che deve essere interpretato dal parlante tramite il ricorso a conoscenze extra-linguistiche e culturali: nel caso di “green economy”, l’aggettivo “verde” è da interpretare semanticamente come “ecologico”, poiché denota una scelta di vita consapevole, sostenibile e vicina all’ambiente, in un pianeta dove la maggior parte delle risorse stanno per terminare e dove ognuno si deve impegnare per dare alle generazioni future un pianeta più abitabile; in “silver economy”, l’aggettivo di colore designa la fascia più anziana della popolazione, da prendere in considerazione nelle sue esigenze e nei suoi bisogni che sono diversi da quelli delle persone più giovani; in “rivoluzione arancione”, l’aggettivo di colore è, invece, da leggere in quanto simbolo dell’opposizione ucraina che si è sviluppata in seguito alle elezioni presidenziali del 2004 e dove il colore, utilizzato dalle persone in protesta, evoca le foglie autunnali degli ippocastani che bordano la strada principale di Kiev. Tali esempi dimostrano, quindi, quanto gli aggettivi di colore siano implicati nelle creazioni di nuove denominazioni e in processi neologici, anche in terminologia – basti pensare, in tal senso, alle denominazioni “zona rossa”, “zona gialla” e “zona verde”, create durante la pandemia da coronavirus 2019 –, la cui interpretazione richiede da parte del parlante delle conoscenze culturali, enciclopediche e pragmatiche.

Molteplici sono dunque gli spunti di ricerca offerti dal concetto affascinante di colore che si presta a riflessioni di distinta natura in cui le denominazioni ricorrenti agli aggettivi di colore possono essere indagate con approcci afferenti

a discipline distinte: le “sociocromie” di Ceppi non sono che un ultimo esempio di come poter declinare e interpretare l’universo complesso del colore, che merita di sicuro ulteriori approfondimenti anche dal punto di vista linguistico e nel campo delle ricerche di natura principalmente terminologica e terminografica.

Riferimenti bibliografici

- Berlin, Brent, and Paul Kay. 1969. *Basic color terms: their universality and evolution*. Berkeley: University of California Press.
- Dubois, Danièle. 2006a. “Les “mots” et les catégories cognitives du sensible : des rapports problématiques.” *Cahiers du LCPE* 7: 23-47.
- Dubois, Danièle. 2006b. “De l’expérience subjective des catégories de couleurs à l’objectivité de la couleur : approches cognitives.” *Cahiers du LCPE* 7: 67-78.
- Golka, Maria H. 2014. “La catégorisation linguistique des couleurs : niveaux d’élémentarité des noms de couleurs français.” *Cognitive Studies/Études cognitives* 14: 131-47.
- Grimaldi, Claudio. 2017. “Processus de dénomination des couleurs en français contemporain : réflexion sur l’agencement des composantes dans les structures polylexicales à deux lexèmes.” *Annali dell’Università degli Studi di Napoli “L’Orientale”-sezione romanza* LIX, no. 2: 127-41.
- Grimaldi, Claudio. 2020. “Mots, couleurs et néologie : analyse des dénominations de couleur en français contemporain.” In *Desafíos de la neología en las lenguas románicas en el siglo XXI/Défis de la néologie en langues romanes au XXI^e siècle*, sous la direction de José Carlos de Hoyos, Cécile Poix, Adam Renwick, Corina Veleanu, 103-13. Murcia: Ediciones Universidad de Murcia.
- Kristol, Andres M. 1978. *Color. Les langues romanes devant le phénomène de la couleur*. Berne: Francke.
- Krysolova, Svetlana. 2005. *Contribution à l’étude lexico-sémantique des dénominations chromatiques en russe et en français*. Tesi di dottorato diretta da Antoine Nivière (Université Nancy 2), discussa il 6 dicembre 2005 presso l’Université Nancy 2, Institut de russe et de serbo-croate, 789 p.
- Molinier, Christian. 2001. “Les adjectifs de couleur en français. Éléments pour une classification.” *Revue Romane* 36, no. 2: 193-206.
- Molinier, Christian. 2006. “Les termes de couleur en français. Essai de classification sémantico-syntaxique.” *Cahiers de Grammaire* 30: 259-75.

- Mollard-Desfour, Annie. 1998. *Dictionnaire des mots et expressions de couleur. Le Bleu*. Paris: CNRS Éditions.
- Mollard-Desfour, Annie. 2000. *Dictionnaire des mots et expressions de couleur. Le Rouge*. Paris: CNRS Éditions.
- Mollard-Desfour, Annie. 2002. *Dictionnaire des mots et expressions de couleur. Le Rose*. Paris: CNRS Éditions.
- Mollard-Desfour, Annie. 2005. *Dictionnaire des mots et expressions de couleur. Le Noir*. Paris: CNRS Éditions.
- Mollard-Desfour, Annie. 2008. *Dictionnaire de la couleur. Mots et expressions d'aujourd'hui. Le Blanc*. Paris: CNRS Éditions.
- Mollard-Desfour, Annie. 2008b. "Les mots de couleur : des passages entre langues et cultures." <<https://gerflint.fr/Base/Italie4/mollarddesfour.pdf>>.
- Mollard-Desfour, Annie. 2011. "Le lexique de la couleur : de la langue à la culture... et aux dictionnaires." *Revue d'études françaises* 16: 89-109.
- Mollard-Desfour, Annie. 2012. *Dictionnaire de la couleur. Mots et expressions d'aujourd'hui. Le Vert*. Paris: CNRS Éditions.
- Mollard-Desfour, Annie. 2015. *Dictionnaire de la couleur. Mots et expressions d'aujourd'hui. Le Gris*. Paris: CNRS Éditions.
- Mollard-Desfour, Annie. 2017. "Les couleurs dans la mode. Phénomènes lexicaux et données sociologiques." In *Le français : des mots de chacun, une langue pour tous. Des français parlés à la langue des poètes*, sous la direction de Françoise Argod-Dutard, 113-21. Rennes: Presses Universitaires de Rennes.
- Pastoureau, Michel. 2002. *Bleu. Histoire d'une couleur*. Paris: Le Seuil.
- Pastoureau, Michel. 2008. *Noir. Histoire d'une couleur*. Paris: Le Seuil.
- Pastoureau, Michel. 2013. *Vert. Histoire d'une couleur*. Paris: Le Seuil.
- Pastoureau, Michel. 2016. *Rouge. Histoire d'une couleur*. Paris: Le Seuil.
- Ripoll, Élodie. 2018. *Penser la couleur en littérature. Explorations romanesques des Lumières au réalisme*. Paris: Classiques Garnier.
- Ripoll, Élodie. 2019. "La couleur dans le roman des Lumières. Enjeux, emplois et évolutions." *Dix-huitième siècle* 51: 77-92.
- Rodríguez Pedreira, Nuria. 2002. "Recherches sur les adjectifs de couleur. Description et référenciation." *L'information grammaticale* 95: 27-30.
- Rosch Heider, Eleanor. 1972. "Universals in color naming and memory." *Journal of Experimental Psychology* 93, no. 1: 10-20.

- Rosch Heider, Eleanor. 1973. "On the internal structure of perceptual and semantic categories." In *Cognitive development and the acquisition of language*, edited by Timothy E. Moore, 111-44. New York-San Francisco-London: Academic Press.
- Silvestre, Joao Paulo, Esperança Cardeira, and Alina Villalva. 2006. *Colour and coulour naming: crosslinguistic approaches*. Lisboa: Centro de Linguística de Universidade de Lisboa-Universidade de Aveiro.
- Wierzbicka, Anna. 1996. *Semantics: primes and universals*. Oxford: Oxford University Press.
- Wierzbicka, Anna. 2005. "There are no "Color Universals" but there are universals of visual semantics." *Anthropological Linguistics* 47, no. 2: 217-44.