

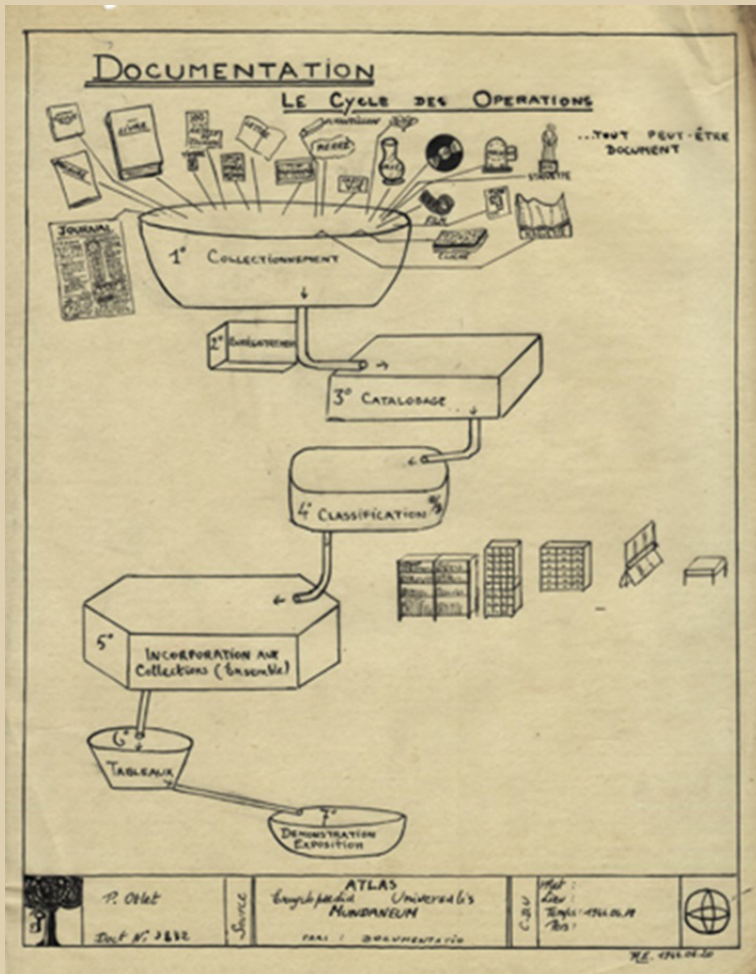
# AIDa informazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 1-2

ANNO 41

GENNAIO-GIUGNO 2023



# **AIDA**informazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

Fondata nel 1983 da Paolo Bisogno

**Proprietario della rivista:**

Università della Calabria

**Direttore Scientifico:**

Roberto Guarasci, *Università della Calabria*

**Direttore Responsabile:**

Fabrizia Flavia Sernia

**Comitato scientifico:**

Anna Rovella, *Università della Calabria*;

Maria Guercio, *Sapienza Università di Roma*;

Giovanni Adamo, *Consiglio Nazionale delle Ricerche* †;

Claudio Gnoli, *Università degli Studi di Pavia*;

Ferruccio Diozzi, *Centro Italiano Ricerche Aerospaziali*;

Gino Roncaglia, *Università della Toscana*;

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

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

Maria Mirabelli, *Università della Calabria*;

Agustín Vivas Moreno, *Universidad de Extremadura*;

Douglas Tudhope, *University of South Wales*;

Christian Galinski, *International Information Centre for Terminology*;

Béatrice Daille, *Université de Nantes*;

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

Federico Valacchi, *Università di Macerata*.

**Comitato di redazione:**

Antonietta Folino, *Università della Calabria*;

Erika Pasceri, *Università della Calabria*;

Maria Taverniti, *Consiglio Nazionale delle Ricerche*;

Maria Teresa Chiaravallotti, *Consiglio Nazionale delle Ricerche*;

Assunta Caruso, *Università della Calabria*;

Claudia Lanza, *Università della Calabria*.

**Segreteria di Redazione:**

Valeria Rovella, *Università della Calabria*

**Editrice:** Cacucci Editore S.a.s.

Via D. Nicolai, 39 – 70122 Bari (BA)

[www.cacuccieditore.it](http://www.cacuccieditore.it)

e-mail: [riviste@cacuccieditore.it](mailto:riviste@cacuccieditore.it)

Telefono 080/5214220



# AIDAinformazioni

## RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

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

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

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

# AIDAinformazioni

Anno 41

N. 1-2 – gennaio-giugno 2023

CACUCCI  EDITORE  
BARI

---

PROPRIETÀ LETTERARIA RISERVATA

---

© 2023 Cacucci Editore – Bari

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

<http://www.cacuccieditore.it> e-mail: [info@cacucci.it](mailto: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, “Effimeri ma non troppo”.	9
CARLOS ALBERTO ÁVILA ARAÚJO, A veritistic turn in information science?	33
DIMITRI BRUNETTI, Le raccomandazioni di Alba per la descrizione dei corrispondenti e la redazione dell’oggetto nel protocollo informatico	55
MANOLA CHERUBINI, FRANCESCO ROMANO, ANDREA BOLIOLI, Intelligenza artificiale e informazione giuridica	71
NATASCIA RALLI, ISABELLA STANIZZI, MARLIES ALBER, COVID-19 e lavoro terminologico: riflessioni a posteriori	91
SALVATORE SPINA, Handwritten Text Recognition as a digital perspective of Archival Science	115
MARC TANTI, Quelles sont les Fake News qui ont circulé lors de la pandémie de Covid-19 ?	133
FEDERICO VALACCHI, Se l’archivio è artificiale. Verso uno ius archivi partecipativo?	153

### Note e rubriche

CLAUDIA LANZA, Recensione “L’innovazione in biblioteca” di Ferruccio Diozzi	173
CLAUDIO GNOLI, TeKaMoLo ovvero le lingue come KOS, e viceversa	177





# Contributi



## “Effimeri ma non troppo”.

La conservazione dei messaggi istantanei:  
è ora di cominciare ad occuparcene?

Stefano Allegrezza\*

**Abstract:** Over the past ten to twelve years, instant messaging systems have become increasingly popular: WhatsApp, Telegram, Signal, WeChat, SnapChat – just to name the most popular apps – but also the instant messaging systems built into Facebook, LinkedIn, or Microsoft Teams are used by anyone and on any occasion, whether formal or informal. This extremely fast and immediate means of interaction increasingly tends to replace those traditionally used so far – such as e-mail, which is considered a more “official” but also less “immediate” means of communication. Consequently, it is necessary to begin to question whether these instant messages can be considered as records and as such should be “captured” in document management systems and somehow preserved for posterity. This paper aims to offer a reflection on these issues and the debate currently underway in the scholarly and professional community of records managers and archivists.

*Keywords:* Instant messaging, Text messaging, Preservation, WhatsApp, Telegram.

### 1. Introduzione

Negli ultimi dieci-dodici anni l'utilizzo dei sistemi di messaggistica istantanea è diventato sempre più diffuso: WhatsApp<sup>1</sup>, Telegram, Signal, WeChat,

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

<sup>1</sup> WhatsApp è stata creata nel 2009 da due ex dipendenti di Yahoo, Jan Koum e Brian Acton, per dare la possibilità agli utenti di scambiarsi messaggi gratuitamente utilizzando il proprio numero di telefono e la rete Internet. Il suo nome deriva dalla fusione tra le parole inglesi “what’s up?” (come va?) e “app” (da “application”). All’inizio l’app era disponibile solamente per gli utenti iPhone, ma in breve tempo venne rilasciata anche la versione per Android, BlackBerry OS, Windows Phone e Symbian (sistemi operativi che oggi hanno una quota di mercato infinitesimale, ma che all’epoca erano utilizzati da una buona percentuale di utenti). In poco tempo WhatsApp è diventato un vero e proprio fenomeno di massa con milioni di utenti entusiasti del servizio offerto. Con una diffusione a macchia d’olio, l’app supera ben presto il traguardo dei cento milioni di utenti attivi ogni mese. Per sostenere le spese di sviluppo, WhatsApp aveva inizialmente introdotto l’obbligo di pagare un abbonamento annuale di 0,99 dollari, una sorta di “tassa” che è stata tolta solamente a gennaio del 2016, da quando

SnapChat – solo per citare le app più diffuse –, ma anche i sistemi di messaggistica istantanea integrati in Facebook, in LinkedIn o in Microsoft Teams sono utilizzati quotidianamente da milioni di persone per scambiare messaggi con amici, familiari, conoscenti. Sempre più spesso ciò avviene anche nel contesto lavorativo, complice anche la pandemia di Covid-19 che ha decisamente influito sul modo di relazionarsi e comunicare. L'uso di questi strumenti di interazione, estremamente veloci e immediati, tende sempre più a sostituirsi a quelli tradizionalmente utilizzati fino ad oggi – come la posta elettronica, considerata un mezzo di comunicazione più “ufficiale” ma anche meno “immediato”<sup>2</sup>. Anche le informazioni istituzionali vengono in molti casi veicolate attraverso sistemi di messaggistica istantanea, con conseguenze, dal punto di vista archivistico, che devono essere valutate attentamente. Occorre, infatti, cominciare a domandarsi se questi messaggi istantanei possano avere un valore documentale e come tali debbano essere “catturati” nei sistemi di gestione documentale e in qualche maniera conservati. Nel seguito si cercherà di rispondere a questa e altre domande collegate e di proporre alcune riflessioni sul tema della conservazione dei messaggi istantanei, a partire da una disamina di alcune tra le più interessanti iniziative presenti nel contesto internazionale.

Prima di entrare nel vivo dell'argomento, è bene specificare esattamente cosa si intende per messaggistica istantanea e quali sono le differenze con gli altri mezzi di comunicazione. Secondo la definizione riportata dai *National Archives and Records Administration* (NARA) degli Stati Uniti, per messaggistica istantanea (*instant messaging*, IM, in inglese) si intende «an electronic messaging service that allows users to determine whether a certain party is connected to the messaging system at the same time. IM allows them to exchange text messages with connected parties in real time» (NARA 2016). Si tratta, quindi, di uno strumento di comunicazione che consente lo scambio di messaggi in maniera istantanea tra le parti connesse allo stesso sistema.

---

l'utilizzo dell'app è diventato gratuito. L'acquisto da parte di Facebook non ha cambiato i progetti di sviluppo di WhatsApp che ha continuato nella sua corsa verso il raggiungimento del miliardo di utenti, avvenuto nel febbraio del 2016 (WhatsApp è diventata la prima applicazione a raggiungere questo traguardo). In base alle statistiche ufficiali, nel 2023 WhatsApp ha raggiunto i 2 miliardi di utenti attivi mensili a livello mondiale, rendendolo l'app di messaggistica istantanea più utilizzata al mondo. In Italia è utilizzata da 33 milioni di utenti (Statista 2023; Cosenza 2023; RAI News 2022).

<sup>2</sup> È curioso osservare come solamente fino a pochi anni fa la posta elettronica venisse considerata uno strumento di comunicazione estremamente veloce se raffrontato alla posta cartacea tradizionale che veniva chiamata anche “snail mail” (posta lumaca). Oggi viene considerata “lenta” la posta elettronica in confronto con altri strumenti come quelli della messaggistica istantanea e ciò non tanto per la velocità della consegna del messaggio (che è pressoché la stessa) ma per il fatto che la risposta da parte del corrispondente risulta in genere meno veloce rispetto a quella che si otterrebbe inviandogli un messaggio istantaneo.

Una distinzione che è opportuno tenere a mente è quella tra messaggistica testuale (*text messaging*) e messaggistica istantanea. In entrambi i casi i messaggi vengono recapitati istantaneamente, ma nel primo vengono inviati attraverso la rete di telefonia mobile (rientrano in questa categoria i ben noti e ormai “obsoleti” messaggi SMS (Short Message Service) e MMS (Multimedia Messaging Service)), nel secondo attraverso la rete Internet, solitamente mediante una connessione wi-fi oppure dati. Inoltre, mentre nel passato attraverso gli SMS era possibile inviare solo messaggi testuali (con una lunghezza limitata a 160 caratteri), oggi questo limite è stato superato ed è possibile inviare anche contenuti multimediali – come registrazioni audio, video e immagini – attraverso gli MMS, con un risultato analogo, anche se non del tutto comparabile, a quello dell’*instant messaging*. A sua volta l’*instant messaging* può essere classificato a seconda della modalità attraverso la quale i messaggi vengono inviati: utilizzando delle app, tipicamente quelle presenti sul proprio smartphone (è questo il caso di WhatsApp e Telegram); mediante delle piattaforme di social media (è il caso del Messenger di Facebook); attraverso delle piattaforme di videoconferenza (come Microsoft Teams) o sfruttando piattaforme collaborative (Tab. 1).

Categoria	Esempi
Text messaging	Short Message Service (SMS), Multimedia Messaging Service (MMS)
Instant messaging attraverso App	WhatsApp, Telegram, WeChat, SnapChat, Signal, Viber, Line, Discord, iMessage, Google Messages
Instant messaging attraverso piattaforme social media	Twitter Direct Message, Facebook Messenger, LinkedIn Messenger, Youtube Messenger, Instagram, TikTok, Pigeon, Yammer, Jive
Instant messaging attraverso piattaforme di videoconferenza	Google Meet, Microsoft Teams, Zoom, Cisco Webex, Jitsi, Streamyard, GoToMeeting, Skype
Instant messaging attraverso piattaforme collaborative	Google Chat (ex Google Hangouts), Microsoft Teams, IBM Sametime, Novell Groupwise Messenger, SharePoint, Slack

Tabella 1: Le varie tipologie di *text* e *instant messaging*.

Tuttavia, occorre rilevare come questa distinzione si vada affievolendo sempre più, anche per l’integrazione che ormai sussiste tra le varie modalità: ad esempio, oggi è possibile inviare messaggi WhatsApp non solo tramite il proprio smartphone ma anche attraverso un browser web, semplicemente raggiungendo il sito “WhatsApp Web”; tutte le chat risulteranno costantemente sincronizzate e i messaggi saranno accessibili sia utilizzando l’una o l’altra modalità. Si noti anche che queste distinzioni non sono sempre mutuamente

esclusive: ad esempio la messaggistica su Microsoft Teams ricade sia nella categoria dell'*instant messaging* attraverso piattaforme di videoconferenza che in quella dell'*instant messaging* attraverso piattaforme collaborative, potendo tale applicazione essere considerata sia come piattaforma di videoconferenza che come piattaforma collaborativa.

## 2. Stato dell'arte

Sebbene il tema della conservazione dei messaggi istantanei abbia sollevato l'interesse della comunità archivistica solo recentemente, a livello internazionale – e in modo particolare nel mondo anglosassone – ferve già uno stimolante dibattito, soprattutto con riferimento all'utilizzo della messaggistica istantanea nei contesti governativi. In questo ambito vanno senz'altro segnalate le direttive riguardo la gestione e la conservazione di *text messaging* e *instant messaging*<sup>3</sup> emanate dai diversi archivi nazionali.

Molto attivi su questo fronte sono gli Stati Uniti: i *National Archives and Records Administration* diffondono, a partire dal 1995, una serie di *Bullettin*, tra cui risulta di particolare interesse quello del 29 luglio 2015 intitolato *Guidance on Managing Electronic Messages* (NARA 2015), che fornisce indicazioni sull'utilizzo e la gestione dei messaggi elettronici e ricomprende tra questi anche i messaggi istantanei. Con il recente bollettino del 5 febbraio del 2023 i *National Archives* ribadiscono che i messaggi elettronici «created or received in the course of agency business are likely federal records. This includes electronic messages sent or received on personal devices that meet the definition of a record. These messages must be forwarded or copied to an official account within 20 days» (NARA 2023a). Quindi, i messaggi elettronici, di qualunque tipo essi siano (compresi quelli istantanei) sono potenzialmente dei documenti in senso archivistico e non ne può essere escluso tale *status* per il semplice fatto che si tratta di messaggi istantanei inviati con i dispositivi personali.

Nel Regno Unito, nel marzo 2022, l'*Institute for Government*, a conclusione di una indagine sull'utilizzo dell'*instant messaging* da parte dei politici e dei funzionari delle agenzie governative, ha pubblicato un rapporto estremamente interessante dal titolo *WhatsApp in Government. How ministers and officials should use messaging apps – and how they shouldn't* (Dourrant, Lilly, e Tingay 2022). L'indagine ha messo in evidenza come l'uso di WhatsApp – e di altre app simili come Signal e Telegram – stia trasformando il modo in

<sup>3</sup> Anche se WhatsApp è certamente l'app di messaggistica istantanea più conosciuta e diffusa, non è l'unica: esistono tante altre opzioni, come Telegram, WeChat, Snapchat, Viber, Kik, Facebook Messenger, Skype, Line, QQ, ecc. Vale la pena notare che la popolarità delle piattaforme di messaggistica istantanea varia in modo significativo a seconda della regione. Ad esempio, WeChat è molto popolare a livello globale, ma la maggior parte di questa popolarità si concentra in Cina e tra le persone con legami con la Cina che vivono altrove.

cui i politici e i membri del Parlamento discutono e prendono decisioni nel contesto istituzionale. I messaggi testuali e quelli istantanei sono ormai entrati nell'uso comune negli ambienti di lavoro del governo britannico, dove sono ampiamente utilizzati per lo stesso motivo per cui lo sono in tutto il mondo: consentono di comunicare in modo rapido ed efficace, in coppia o in gruppo, senza l'impegno organizzativo e logistico di una riunione – si pensi, ad esempio, alla difficoltà di trovare una data e un orario che vada bene a tutti i membri di un gruppo di lavoro – o l'onere di un lungo scambio di email che coinvolge più persone (Dourrant, Lilly, e Tingay 2022). Secondo il rapporto «WhatsApp is changing the way politicians and others debate, take decisions and build (or dismantle) relationships with each other. This is already having an impact on UK politics, but there is little attention paid to how it affects the way government works» (Dourrant, Lilly, e Tingay 2022). In particolare, sempre secondo il rapporto, le *app* di messaggistica istantanea sono utili, ma devono essere usate con attenzione per garantire che taluni aspetti negativi – come la mancanza di trasparenza – non superino quelli positivi. Inoltre, il governo deve affrontare con urgenza la questione della carenza di indicazioni sull'utilizzo della messaggistica istantanea da parte di ministri, consiglieri politici e funzionari pubblici emanando delle direttive e delle policy in materia. Il rapporto suggerisce come possibile raccomandazione quella di stabilire che i ministri, i consiglieri e i funzionari non utilizzino i telefoni personali per le attività di governo (anche se tale obiettivo non è facilmente raggiungibile); dal momento che, tuttavia, la messaggistica istantanea viene comunque utilizzata anche per le attività ufficiali, è necessario stabilire delle regole per individuare i messaggi che devono essere conservati; è, poi, importante prevedere dei meccanismi per garantire che i messaggi rilevanti vengano archiviati e conservati, sia per finalità giuridico amministrative che per finalità di ricerca storica. Per tutte queste ragioni, a conclusione del rapporto l'*Institute for Government* sottolinea quanto sia importante che il governo rifletta su questi temi e metta in atto le strategie opportune.

Sempre nel Regno Unito, la *Digital Preservation Coalition* (DPC) si è occupata di recente della conservazione della messaggistica istantanea pubblicando nel suo apprezzato blog (Mitcham 2022) una articolata discussione sull'utilizzo di WhatsApp, nella quale è stata messa in evidenza la sua enorme diffusione come mezzo rapido ed efficace per comunicare con i propri contatti, passando velocemente dalla sfera privata – di amici e familiari – al contesto lavorativo. Dal blog emerge come sia proprio quest'ultimo ambito quello in cui si rende necessario normare al più presto l'utilizzo della messaggistica istantanea per lo scambio di informazioni e documenti tra colleghi, fornendo anche esempi e buone pratiche.

A livello europeo delle indicazioni importanti sono giunte dall'indagine svolta dall'European Ombudsman, il Mediatore Europeo, che con la comuni-

cazione del 30 giugno 2021 (European Ombudsman 2021) aveva avviato una indagine sul modo in cui le istituzioni, gli organi, gli uffici e le agenzie dell'unione europea<sup>4</sup> registrano i messaggi testuali e i messaggi istantanei inviati o ricevuti dal personale nello svolgimento delle proprie attività lavorative. L'iniziativa era volta a fare il punto della situazione e ad individuare le *best practice* tra quelle in uso nelle amministrazioni coinvolte. Sulla base delle risposte, che dovevano pervenire entro il 15 novembre 2021, il Mediatore ha formulato delle raccomandazioni pratiche rivolte alle amministrazioni dell'Unione Europea<sup>5</sup>. Nella nota di chiusura dell'indagine (European Ombudsman 2022a) si riconosce che tali amministrazioni, come qualsiasi altra amministrazione pubblica, utilizzano i moderni mezzi di comunicazione elettronica – compresi i messaggi testuali e quelli istantanei che, in taluni casi, sono documenti ai sensi del regolamento 1049/2001 sull'accesso pubblico – per lo scambio di informazioni nel lavoro quotidiano. In tale nota il Mediatore osserva che tutte le otto istituzioni dell'Unione Europea coinvolte hanno dichiarato che i messaggi testuali e istantanei rientrano, in linea di principio, nella definizione generale di documento stabilita nelle rispettive direttive sulla gestione documentale. Ad esempio, la Commissione Europea, una delle otto istituzioni destinatarie dell'indagine, dichiara che per quanto riguarda la registrazione di tali messaggi si attiene alla *Commission Decision (C/2020/4482) on records management and archives* che si applica ai documenti in possesso della Commissione «regardless of format and the technological environment in which they are collected, created or generated» (European Commission 2020). Essa definisce come *record* «any information, received and created in the form of a document, a collection of data or other form in a digital or analogue medium that is captured in an official repository and managed and maintained as evidence and as an asset» (European Commission 2020). Nello specifico, «documents shall be registered if they contain important information which is not short-lived<sup>6</sup> or if they may involve action or follow-up by the Commission or one of its departments» (European Commission 2020), un'affermazione che viene accompagnata da una serie di esempi (non esaustivi) per guidare il personale nel processo di selezione dei messaggi che vanno considerati come *record*, e quindi suscettibili di registrazione, e quali no. Tuttavia, alla domanda sulle regole da applicare in merito alla messaggistica testuale e istantanea, la Commissione Europea

<sup>4</sup> Sono state coinvolte le seguenti amministrazioni: European Medicines Agency, Council of the European Union, European Commission, European Central Bank, European Chemicals Agency, European Food Safety Authority, European Parliament, European Border e Coast Guard Agency.

<sup>5</sup> Si veda il paragrafo “Raccomandazioni pratiche”, *infra*.

<sup>6</sup> Short-lived, ovvero di breve durata, effimero, è il termine che individua i messaggi con un'utilità temporanea e che possono essere cancellati una volta esaurito il loro scopo; al contrario, sono rilevanti i messaggi da cui può derivare una decisione, un procedimento o comunque un'azione da parte dell'istituzione.



risponde che: «due to their short-lived and ephemeral nature, text and instant messages are not, in general, considered to contain important information relating to policies, activities and decisions of the Commission, nor are they considered to be drawn up or received in the meaning explained in the criteria set out above. They do therefore, as a general rule, not meet the registration criteria set out above» (European Commission 2021).

Come regola generale, quindi, la Commissione Europea non prevede un sistema di gestione e registrazione dei messaggi, in quanto il personale tende – o dovrebbe tendere – a non usarli per le comunicazioni importanti, scegliendo invece forme considerate istituzionali come l’email. Tuttavia, qualora un messaggio testuale o istantaneo dovesse contenere informazioni importanti e non effimere, i responsabili della gestione documentale hanno il compito di garantire che tali informazioni siano correttamente registrate, archiviate e conservate. Inoltre, se la messaggistica istantanea dovesse diventare un mezzo di comunicazione correntemente utilizzato, e se la pratica di scambiare informazioni importanti e documenti ufficiali attraverso questa modalità dovesse diventare più comune, la Commissione potrebbe valutare l’individuazione o la realizzazione di strumenti per l’importazione diretta dei messaggi testuali/istantanei dai sistemi in cui risiedono (ad esempio, app come WhatsApp e Telegram) nei sistemi di gestione dei documenti.

Quello che emerge dal quadro appena delineato è che le istituzioni, soprattutto in ambito anglosassone, dimostrano di avere una certa attenzione verso il tema della conservazione della messaggistica istantanea e stanno cominciando a fornire indicazioni sotto forma di direttive e raccomandazioni. Per ora queste si soffermano sull’uso che deve essere fatto della messaggistica istantanea nei contesti istituzionali (specificando quando il personale deve o non deve utilizzare la messaggistica e per quali comunicazioni). In linea generale tutte le direttive suggeriscono una serie di domande (organizzare come una sorta di *check-list*) rispondendo alle quali si possono trarre elementi utili per individuare quei messaggi istantanei che debbono essere considerati come *record*. Sono solitamente esclusi i messaggi *short-lived*, anche se si tratta, tuttavia, di una decisione non priva di criticità perché a volte una conversazione può iniziare tramite un messaggio istantaneo apparentemente senza valore ed evolvere poi in una discussione rilevante che porta a decisioni importanti e quindi è meritevole di essere conservata.

Un elemento molto interessante che emerge dalla disamina delle direttive è che la scelta di archiviare o meno un determinato messaggio nel sistema di gestione documentale non deve dipendere dal mezzo – che si tratti di una lettera ufficiale, di una email o di un messaggio istantaneo – ma dal suo contenuto (European Ombudsman 2022a, 2022b)<sup>7</sup>. La scelta su cosa considerare

<sup>7</sup> Infatti, il Mediatore Europeo afferma che è necessario «take into account, in document management rules and practices, the ever-increasing use of text and instant messaging. Con-

importante e cosa no viene generalmente affrontata in maniera piuttosto pragmatica. Ad esempio, la Commissione Europea nel rispondere alla domanda «Is the information contained in the document important and not short-lived?» individua i seguenti criteri:

This question requires subtle judgement taking into account the content and context of the document concerned. A document that requires action or follow-up by the European Commission or one of its departments, or involves the responsibility of the institution, is important and not short-lived. Likewise, a document that may later be needed as proof in accordance with the rules and regulations applicable to the underlying “business process” is considered important and not short-lived; in contrast, information is considered to be unimportant and short-lived if not keeping it would have no negative administrative or legal effect for the European Commission (European Commission 2021).

Nella risposta si sottolinea come la distinzione tra le informazioni importanti e quelle non rilevanti richieda un giudizio “sottile” che tiene conto del contenuto e del contesto del documento in questione. Si tratta, a tutti gli effetti, di un’operazione di selezione che richiede tutte le avvertenze e le cautele del caso per poter essere condotta in maniera corretta.

A conclusione di questa sintetica analisi dello stato dell’arte non si può non rilevare come in Italia, purtroppo, il tema dell’archiviazione e conservazione della messaggistica istantanea non è ancora stato preso in considerazione dalla comunità scientifica e professionale degli archivisti, sintomo questo di una ancora insufficiente attenzione. Tuttavia, si tratta di un tema che non potrà essere ignorato a lungo, stante l’uso crescente della messaggistica istantanea che in taluni ambiti, sia nel settore pubblico che in quello privato, è giunta addirittura a sostituirsi alla posta elettronica come mezzo privilegiato di comunicazione e di trasmissione di documenti.

### 3. Alcune riflessioni dal punto di vista archivistico

Sotto il profilo archivistico l’utilizzo dei messaggi istantanei solleva diverse questioni che meritano di essere approfondite. La prima è relativa al valore dei messaggi istantanei come documenti in senso archivistico: a questo proposito l’analisi delle direttive e raccomandazioni emanate da diverse istituzioni mostra come si stia sviluppando una riflessione attorno al concetto di *instant messaging as a record*, del tutto simile a quella che si era sviluppata nei primi anni del XXI secolo intorno al concetto di *email as a record*. Oggi come allora ci si chiede quali sono le caratteristiche che un messaggio elettronico (nel caso

---

cretely, this implies that the decision to record a certain piece of information in the administration’s document management system should not be dependent on the medium - be it a letter, an email, a text or instant message - but on its content» (European Ombudsman 2022a).

in questione, un messaggio istantaneo) deve possedere affinché possa essere considerato come *record* e come tale “meritevole” di trasferimento prima nel sistema di gestione documentale e poi nel sistema di conservazione? Questa sicuramente è la domanda più complessa tant’è vero che diverse organizzazioni hanno sentito l’esigenza di redigere delle direttive in materia di formazione, gestione e conservazione della messaggistica istantanea.

A questo proposito è interessante leggere la risposta alla domanda «Does IM Content Qualify as a Federal Record?» contenuta nelle *Frequently Asked Questions about Instant Messaging* (NARA 2016): «Agencies that allow IM traffic on their networks must recognize that such content may be a Federal record under that definition and must manage the records accordingly». Al fine di facilitare il riconoscimento dei messaggi istantanei come *federal records* i *National Archives* rendono disponibile un elenco, non esaustivo, di domande che possono aiutare le agenzie federali in questa analisi (NARA 2023b). Ad esempio, occorre domandarsi se il messaggio elettronico contiene prove delle policy, delle attività o della missione di un’agenzia; se le informazioni sono disponibili solamente in quel messaggio e non anche attraverso altri mezzi più “formali”; se l’agenzia utilizza lo strumento di comunicazione elettronica per trasmettere informazioni ufficiali dell’agenzia; se l’agenzia ha bisogno di quell’informazione per le sue attività istituzionali, ecc.<sup>8</sup> Anche il *Department for Levelling Up, Housing and Communities* (DLUHC) fornisce «a list of the types of messages that should be recorded, including those involving “decisions to start or end a project” and “approvals to spend”» (Dourrant, Lilly, e Tingay 2022). Anche per i *New Zealand Archives* i messaggi testuali e ogni altra tipologia di messaggi istantanei, se utilizzati per finalità di tipo amministrativo o nella conduzione di affari, sono considerati *record*: «If an organisation uses text messaging or any other instantaneous, non-sequential electronic communication mechanism to conduct business, e.g. social media, these communications are considered records under the Public Records Act 2005. As such, they must be managed accordingly» (New Zealand Archives 2020). Dello stesso avviso è il *National Archive of Australia* secondo il quale «Social media and instant messaging posts, media, comments, messages and analytics that are created or received as part of Australian Government business are Commonwealth records. To be accountable, your agency must also manage these types of records» (NAA 2023).

In generale, le indicazioni sono concordi nel ritenere che non tutti i messaggi, ricevuti o scambiati all’interno di un’istituzione, tra una o più persone,

---

<sup>8</sup> Il bollettino (NARA 2023b) riporta una «non-exhaustive list of questions may help agencies determine record status of electronic message content: Does it contain evidence of an agency’s policies, business, or mission? Is the information only available in the electronic message? Does the agency use the tool to convey official agency information? Is there a business need for the information?».

sono da ritenere documenti in senso archivistico. Anzi, la maggior parte dei messaggi istantanei sono privi di un valore giuridico-amministrativo e non avrebbe senso conservarli neanche in prospettiva storica. Tener conto di questo aspetto è estremamente importante, anche perché ha un impatto significativo sul processo di valutazione e selezione dei messaggi che hanno valore documentale. Si tratta di un'attività che richiede grande attenzione e competenza – insostituibili se si vuole individuare quella piccola percentuale di messaggi che merita di essere selezionata – e la cui sostenibilità, anche sul piano operativo, è messa in crisi dall'enorme numero di messaggi istantanei che vengono prodotti e che rischia di diventare rapidamente incontrollabile. Questa “iperproduzione” – le cui cause possono essere rinvenute da una parte nella facilità con cui è possibile creare ed inviare messaggi istantanei, dall'altra nella grande capacità di archiviazione dei dispositivi oggi disponibili, che sembra quasi non porre limiti al numero di messaggi che è possibile “salvare” – fa sì che nei dispositivi personali si accumulino quantità ormai ingestibili di messaggi istantanei (oltre che delle varie e «nuove (o parzialmente nuove) tipologie di memorie personali digitali» (Roncaglia 2023) che si formano continuamente nell'arco della giornata, dal momento del risveglio fino a quando non ci si addormenta).

Riguardo l'aspetto della selezione, occorre tener presente che esiste una certa percentuale di messaggi che contengono le prove di approvazioni, permessi o concessioni (si pensi, ad esempio, ad una autorizzazione relativa ad una determinata situazione inviata dal dirigente ad un funzionario mediante un messaggio WhatsApp, che può essere costituito anche semplicemente da un “Ok” o da un emoji con il “pollice su”) oppure le motivazioni alla base di decisioni e azioni (come nel caso dei messaggi istantanei scambiati tra un ministro e il suo staff, ad esempio relativamente ad una determinata questione in discussione in Parlamento). In casi del genere i messaggi istantanei dovrebbero essere considerati alla stregua di documenti ufficiali e quindi acquisiti e conservati.

In altri casi, i messaggi istantanei devono essere conservati per periodi di tempo più o meno lunghi per soddisfare esigenze fiscali o finanziarie. Ad esempio, nel settore privato ormai è sempre più diffusa la consuetudine di effettuare ordini di acquisto tramite messaggi istantanei (sia attraverso un messaggio testuale che un messaggio vocale di WhatsApp) così come attraverso gli stessi mezzi vengono inviate le conferme degli ordini, le copie di cortesia delle fatture, le conferme delle prenotazioni, le autorizzazioni al ritiro della merce, ecc. In tutti questi casi il messaggio deve essere conservato, almeno per il periodo di tempo previsto dalla legge.

Per quanto riguarda l'assimilazione dei messaggi istantanei a veri e propri documenti vale la pena ricordare che il Tribunale di Catania, con l'ordinanza del 27 giugno 2017, ha ritenuto legittimo, sotto il profilo della sussistenza della forma scritta e della validità della sua comunicazione, un licenziamento intimato a mezzo “WhatsApp” (Battaglia 2022). A tal fine, il Giudice ha

condotto un ragionamento estremamente rigoroso sotto il profilo giuridico: innanzitutto, per quanto attiene alla verifica sulla necessaria sussistenza della forma scritta della comunicazione di recesso, vi sono pochi dubbi: il messaggio inviato a mezzo “WhatsApp”, infatti, assolve pienamente l’onere della forma scritta, trattandosi di un «documento informatico dattiloscritto, che parte ricorrente ha con certezza imputato al datore di lavoro, tanto da provvedere a formulare tempestiva impugnazione stragiudiziale» (AvveComm n.d.). Quanto alla certezza temporale della ricezione della comunicazione di recesso da parte del lavoratore il Giudice ha osservato che il sistema di WhatsApp garantisce al mittente e al destinatario del messaggio valida evidenza in ordine alla data, all’ora e alla prova dell’avvenuta ricezione (grazie alla “doppia spunta grigia”). In aggiunta, garantisce al mittente anche la prova di avvenuta lettura da parte del destinatario (grazie alla “doppia spunta blu”), cosa che, per esempio, evidentemente non è permessa con l’utilizzo di altri mezzi, come la “classica” raccomandata con avviso di ricevimento: è possibile, infatti, in questo secondo caso, avere contezza solo dell’avvenuta ricezione della raccomandata ma evidentemente non del giorno e l’ora in cui questi abbia effettivamente deciso di aprire la busta e leggere il contenuto.

Nel mondo degli archivi di persona la situazione è ancora più complessa perché sono meritevoli di attenzione non solo i messaggi aventi una rilevanza giuridica o amministrativa ma anche quelli con un valore affettivo o sentimentale (Allegrezza 2023). Chi avrebbe il coraggio di eliminare un messaggio ricevuto da un personaggio illustre, una persona famosa o anche solo particolarmente importante per la propria vita? E infatti si tende a non eliminare i messaggi presenti nelle chat, lasciandoli fino a quando la memoria dello smartphone non si esaurisce, e anche in quel caso si è disposti a cambiare telefono, aumentando la memoria, pur di non perderli. Purtroppo, molti messaggi importanti sono già stati cancellati automaticamente dai sistemi, che li preservano solo per determinati periodi di tempo (si pensi alle chat che racchiudono i messaggi scambiati durante le sessioni di lavoro su Zoom), o si sono persi perduti durante le varie sostituzioni dei dispositivi che l’obsolescenza tecnologica periodicamente impone, con la conseguenza che, con buona probabilità, molti messaggi istantanei scambiati tra due o più persone, sono già andati perduti.

Un’altra questione che andrebbe affrontata al più presto è quella relativa alla difficoltà di formare delle aggregazioni documentali che siano complete e corrette dal punto di vista archivistico. La straordinaria diffusione della messaggistica istantanea, in tutte le sue varie forme, fa sì che sia estremamente difficile ricondurre all’interno della stessa aggregazione documentale i messaggi scambiati, ad esempio, tra due o più corrispondenti. Infatti, non di rado un “carteggio elettronico” inizia con un messaggio di posta elettronica, prosegue con una risposta attraverso un’app di *instant messaging* (ad es. WhatsApp), continua con una successiva risposta, magari attraverso un’altra app di *instant*

*messaging* (ad es. Telegram), poi prosegue spostandosi sulla messaggistica collegata ad una piattaforma di social media (ad es. LinkedIn) e così via. Casi di questo genere sono oggi sempre più frequenti, e in questo modo i vari messaggi si sedimentano sulle piattaforme più svariate: il sistema di posta elettronica, WhatsApp, Telegram, LinkedIn, Microsoft Teams, Facebook Messenger, ecc., con la conseguente difficoltà – se non addirittura l'impossibilità – di riunire all'interno di una unica aggregazione documentale tutti i messaggi scambiati con un determinato corrispondente o relativi ad un determinato affare. Spesso poi non si tratta più solo di messaggi testuali, ma anche di messaggi di altra natura (ad esempio, di messaggi vocali) o contenenti anche immagini, video, emoticons ed emoji<sup>9</sup>, ecc. Come si può, allora, tentare di ricostruire le aggregazioni documentali, e con esse il vincolo archivistico, dal momento che i vari messaggi non si trovano all'interno della stessa unità archivistica ma sono sparsi sui vari dispositivi, sulle varie piattaforme, sui vari *loci* virtuali? Come si può cercare di ricondurre tutto all'interno dello stesso "fascicolo"? Nel tentativo di trovare una soluzione, alcuni hanno proposto di considerare la posta elettronica come sistema di archiviazione principale e di individuare dei metodi o degli strumenti per ricondurre i messaggi ricevuti o inviati attraverso altre piattaforme in tale sistema. Ad esempio, è stato proposto di effettuare degli *screenshot* dei messaggi presenti nei sistemi di messaggistica istantanea e di importarli nel sistema di posta elettronica copiando quello *screenshot* in una email e inviandola al proprio indirizzo di posta elettronica o ad un indirizzo appositamente dedicato a questo scopo. Si tratta, evidentemente, di una soluzione estemporanea che può andare bene nell'immediato ma che deve necessariamente evolvere in qualcosa di più strutturato. Inoltre, anche laddove ciò fosse tecnicamente possibile, occorre considerare che in una siffatta aggregazione documentale potrebbero astrattamente confluire tipologie documentali assai diverse tra loro (messaggi di posta elettronica, messaggi testuali, messaggi istantanei, messaggi vocali, immagini, video, ecc.); si tratterebbe, cioè, di un contesto multitipologico ed integrato dove i concetti di "documento" e di "fascicolo" – non più definibile solamente come «complesso di carte, memoria di un'attività» (Romiti 2020) – ma anche il concetto stesso di "vincolo archivistico" e, in ultima analisi, quello di "archivio", devono necessariamente modificarsi al fine di ricomprendere queste nuove tipologie documentarie. Infatti, il riconoscimento dello *status* documentale ai messaggi istantanei – oltre che ad altre "nuove" tipologie di documenti: si pensi ai post sui social media, ai blog,

---

<sup>9</sup> Gli "emoji" sono simboli pittografici divenuti popolari in Giappone alla fine degli anni 1990 e utilizzati attualmente sia nella scrittura di messaggi testuali ed istantanei, che nella scrittura di email. Si tratta di una evoluzione commerciale delle emoticon, da cui differiscono «per standardizzazione e tipologia di raffigurazione, oltretutto per le modalità di espressione dei contenuti non verbali, attingendo direttamente dalle forme di rappresentazione tipiche di manga e anime» (Treccani, n.d.).



ai siti web – produce delle conseguenze su questi concetti e in particolare su quello di “archivio”: se l’archivio è il «complesso dei documenti acquisiti da un ente (magistrature, organi e uffici centrali e periferici dello Stato; enti pubblici; istituzioni private, famiglie o persone) durante lo svolgimento della propria attività» (Carucci 2019), i messaggi istantanei inviati o ricevuti «durante lo svolgimento della propria attività» devono essere anch’essi considerati come documenti a tutti gli effetti e, in quanto tali, una volta sottoposti ad un accurato processo di selezione, non possono non entrare a far parte dell’archivio del soggetto produttore.

Un’altra questione che deve essere urgentemente affrontata, soprattutto con riferimento al contesto italiano, dove, come si è visto, manca del tutto una riflessione su questi temi, è la quasi totale assenza di indicazioni – sotto forma di linee guida, raccomandazioni, policy e best practice – sia sull’utilizzo corretto della messaggistica istantanea che, soprattutto, sui criteri di valutazione e selezione di quei messaggi che si ritengono importanti e meritevoli di essere consegnati alla futura memoria. Soprattutto nelle pubbliche amministrazioni servono delle direttive unitarie e chiare che diano indicazioni su come utilizzare i vari strumenti di messaggistica, come selezionare i messaggi che hanno valore documentale, quali azioni porre in essere per gestirli e conservarli correttamente. Impedire, come qualcuno propone, l’impiego della messaggistica e di altri sistemi di comunicazione per lo svolgimento di attività ufficiali sarebbe un provvedimento sproporzionato e di difficile applicazione, non fosse altro perché occorrerebbe rinunciare ai vantaggi della messaggistica istantanea, come la rapidità e la velocità, ormai insostituibili. Allo stesso modo affermare semplicisticamente che tutto ciò che viene inviato o ricevuto tramite messaggi istantanei è privo di rilevanza e può essere eliminato non è aderente alla realtà e potrebbe portare a perdite incontrollate di materiale potenzialmente molto interessante anche nella prospettiva della futura ricerca storica.

Infine, non si può prescindere da una attenta riflessione sul modello conservativo da adottare per la conservazione della messaggistica istantanea. Così come per altre “nuove” tipologie di documenti (si pensi, ad esempio al dibattito in corso sulla conservazione dei siti web e dei social media)<sup>10</sup> occorre chiedersi a chi spetti il compito di conservare i messaggi istantanei ritenuti rilevanti: alla singola persona – ad esempio il dipendente di un ente o una

<sup>10</sup> A questo proposito si segnala la Summer School in *Web and Social media archiving and preservation*, organizzata dal Dipartimento di Beni culturali dell’Università degli Studi di Bologna, che è giunta nel 2023 alla terza edizione e che intende offrire «una formazione di alto livello sui temi emergenti dell’archiviazione e conservazione dei siti web e dei social media, che rappresentano una nuova e diversificata tipologia di materiale la cui conservazione è imprescindibile per tutta una serie di ambiti scientifici (si pensi alla ricerca storica, sociologica, politica, artistica, musicale, geografica, filologica, letteraria, demo-etno-antropologica, religiosa...) ai fini della futura ricostruzione dell’attuale civiltà» (Alma Mater Studiorum Università di Bologna 2023).

azienda – che invia o riceve tali messaggi ed è materialmente in possesso del dispositivo? All’organizzazione presso la quale lavora e che deve essere in grado di avere accesso alla messaggistica?<sup>11</sup> Alle piattaforme che consentono di inviare tali messaggi e attraverso le quali questi transitano? Alle istituzioni archivistiche che hanno per scopo statutario quello della salvaguardia della memoria e che, quindi, potrebbero essere individuate come soggetti conservatori anche per questa nuova categoria di documenti? Si tratta di domande che meriterebbero un’approfondita analisi e alle quali non è certamente possibile fornire una risposta compiuta in questa sede. Tuttavia è importante sottolineare la necessità di avviare al più presto una riflessione che porti a definire le indicazioni di massima finalizzate alla progettazione di un modello conservativo efficace, sostenibile e in grado di rispondere all’esigenza di mantenimento della memoria.

#### 4. Raccomandazioni pratiche

Tra le raccomandazioni che si possono fornire in materia di archiviazione e conservazione dei messaggi istantanei, risultano interessanti innanzitutto quelle fornite dal Mediatore europeo, il quale sulla base delle informazioni raccolte durante l’indagine precedentemente citata (European Ombudsman 2021) ha dato delle indicazioni rivolte alle amministrazioni dell’Unione Europea ma valide in generale anche per altre organizzazioni. Queste raccomandazioni, riferite non solo ai messaggi istantanei ma anche ai messaggi testuali, sono state rese note il 13 luglio 2022 attraverso la lettera intitolata *Closing note on the strategic initiative on how EU institutions, bodies, offices and agencies record text and instant messages sent/received by staff members in their professional capacity* (European Ombudsman 2022a) e possono essere così riepilogate:

– per prima cosa, nelle policy e nei manuali di gestione documentale, occorre tenere conto dell’uso sempre più frequente dei messaggi istantanei e del fatto che la decisione di registrare nel sistema di gestione documentale una determinata informazione non dovrebbe dipendere dal mezzo con cui è stata inviata – sia esso una lettera, un’email ufficiale, un messaggio testuale o un messaggio istantaneo – ma dal suo contenuto;

---

<sup>11</sup> In alcuni enti pubblici italiani sono stati adottati dei regolamenti interni in base ai quali i dispositivi mobili assegnati ai dipendenti e utilizzati da questi per finalità di servizio vengono periodicamente consegnati al Servizio informatico dell’ente che provvede ad effettuare l’acquisizione della messaggistica istantanea e il trasferimento nei sistemi di gestione documentale. Si tratta di un’operazione che si basa sul principio che tutto ciò che è “memorizzato” nel dispositivo assegnato per lavoro al dipendente non è di proprietà di quest’ultimo ma dell’ente per cui questi lavora. Il processo è piuttosto veloce (in genere bastano pochi minuti) e al termine il dispositivo viene restituito al dipendente che può subito riprenderne l’utilizzo.



- occorre poi individuare delle soluzioni tecnologiche che consentano di trasferire con facilità nei sistemi di gestione documentale quei messaggi istantanei che hanno un valore documentale;
- occorre fornire indicazioni chiare al personale che per lavoro utilizza dispositivi elettronici su come estrarre dalle app o dalle piattaforme di messaggistica i messaggi istantanei con valore documentale e come trasferirli al sistema di gestione dei documenti;
- è necessario sensibilizzare il personale sulla necessità di garantire che i messaggi istantanei con valore documentale vengano estratti e trasferiti nel sistema di gestione dei documenti con regolarità e di routine; ciò può essere fatto, ad esempio, offrendo una specifica formazione sia al personale già in servizio che a quello neo assunto;
- è importante assicurarsi che i tempi di conservazione dei messaggi istantanei contenuti nei dispositivi elettronici utilizzati per lavoro siano in linea con i piani di conservazione (ad esempio, evitando l’uso dei messaggi cd. “effimeri”, che si auto-cancellano trascorsi determinati periodi di tempo rendendo impossibile il trasferimento nei sistemi di gestione documentale);
- occorre prestare particolare attenzione a cosa fare nel caso in cui un dipendente o un collaboratore lasci il servizio, stabilendo delle precise istruzioni operative per assicurare che tutti i documenti di lavoro in suo possesso, compresi i messaggi istantanei con valore documentale, vengano trasferiti nel sistema di gestione documentale;
- infine, nel rispondere alle richieste di accesso documentale che potrebbero riguardare messaggi testuali e istantanei, occorre prendere in considerazione tutti i “luoghi” in cui tali messaggi potrebbero trovarsi, compresi i dispositivi elettronici utilizzati dal personale, che va eventualmente assistito nelle operazioni necessarie per rendere disponibili tali messaggi.

Anche i *National Archives and Records Administration* hanno fornito delle indicazioni interessanti sulla gestione della messaggistica istantanea già fin dal 2016, quando era stata affrontata la questione del riconoscimento dei messaggi istantanei come documenti federali (*federal records*) ed erano state fornite alcune raccomandazioni attraverso le *Frequently Asked Questions about Instant Messaging* (NARA 2016). Secondo queste FAQ, le agenzie federali nella redazione delle policy per la gestione della messaggistica istantanea devono garantire che i messaggi siano «managed consistently across the agency in its component offices. An effective policy addresses the authorized use of the IM technology and provides guidelines for the management of the records generated during an IM session. This is especially important because IM content may be subject to various types of access requests, including under the Freedom of Information Act (FOIA) or as part of a discovery process in a litigation context».

Con il recente Bollettino del 5 gennaio 2023 (NARA 2023a), i *National Archives* suggeriscono di rivedere politiche e approcci delle agenzie governative

per garantire il rispetto dei requisiti di gestione dei documenti, e di conseguenza predisporre procedure e strumenti per la conservazione a lungo termine dei messaggi scambiati attraverso sistemi di messaggistica elettronica, ricomprendendo tra questi non solo i messaggi istantanei ma anche i messaggi testuali e i messaggi inviati attraverso i sistemi di videoconferenza (Microsoft Teams, Zoom, Google Meet, ecc.), sia su reti e dispositivi delle agenzie, sia su dispositivi personali o di fornitori terzi. Il bollettino prevede di estendere l'approccio Capstone<sup>12</sup> – già largamente adottato da un vasto numero di agenzie federali per la selezione dei messaggi di posta elettronica da trasferire nel sistema di gestione documentale – anche alla messaggistica istantanea e, più in generale, a tutti i messaggi elettronici. Tale metodo presenta certamente grandi benefici ma anche delle difficoltà. Tra i primi vi è la possibilità di gestire la selezione delle email in maniera uniforme, con un approccio semplificato e automatizzato e senza dover richiedere al personale di effettuare la scelta dei messaggi da trasferire nel sistema di gestione documentale analizzando singolarmente il contenuto di ciascun messaggio. Di contro, come per la gestione delle email, la difficoltà nell'applicare questo approccio risiede nel riuscire ad individuare correttamente tutte le "figure Capstone", operazione delicata ma fondamentale per la sua buona riuscita. Analogamente al caso dell'email, anche per la messaggistica istantanea la responsabilità di assicurare la corretta gestione e conservazione risulta in capo alle agenzie federali.

---

<sup>12</sup> L'approccio Capstone è «a successful approach for managing email records to text messages and other digital communications as part of a bid to help agencies with an ever-increasing deluge of electronic records» (NARA 2023b). Si tratta di una metodologia per la valutazione della posta elettronica da destinare alla conservazione permanente e alla consultazione per finalità storico-culturali sviluppato dai National Archives and Record Administration (NARA) degli Stati Uniti per aiutare le agenzie federali statunitensi a conformarsi ai dettami del memorandum presidenziale *Managing Government Records* firmato il 28 novembre 2011 dal Presidente Obama ed entrato in vigore nel 2016, a cui ha fatto seguito la direttiva M-12-18 *Managing Government Records* del 24 agosto 2012. In sostanza, si tratta di un approccio automatizzato alla gestione della posta elettronica, semplificato rispetto all'utilizzo di applicazioni di gestione documentale che richiederebbero al personale di selezionare e archiviare i messaggi di posta elettronica individualmente. L'idea alla base di questo approccio è che si possano selezionare le email in base al lavoro e/o alla posizione del titolare dell'account di posta. In termini pratici, ciò significa che solo le caselle di posta elettronica di funzionari/direttori/dirigenti o altre figure apicali, all'interno di una organizzazione dovrebbero essere prese in considerazione ai fini della conservazione. Questo tipo di approccio *role-based* era stato introdotto per prima volta con il bollettino *Guidance on a New Approach to Managing Email Records* del 29 agosto 2013: «Under a role-based approach, final disposition is determined by the role or position of the record creator or receiver. Specifically, records of certain roles or positions may be scheduled as permanent for transfer to the National Archives. These are usually at the top of an organization. The records of all other roles and positions may be scheduled for disposal as temporary records. In the context of this bulletin, electronic messages of high-level roles and positions could be scheduled as permanent and electronic messages of all others could be scheduled as temporary» (NARA 2013).

Al bollettino ha fatto seguito l'aggiornamento del *General Records Schedule 6.1: Email and Other Electronic Messages Managed under a Capstone Approach* (NARA 2023b) – e delle relative *Frequently Asked Questions* (NARA 2023c) – che possono essere utilizzate per alcuni tipi di messaggi istantanei, i messaggi testuali e le chat. In realtà sono previste delle estensioni che vale la pena evidenziare. Sono, infatti, inclusi, oltre ai messaggi provenienti da servizi di messaggistica forniti su dispositivi mobili (WhatsApp, Signal, Telegram, WeChat, SnapChat, ecc.), anche i messaggi collegati a funzioni di chat o messaggistica del sistema di posta elettronica (Google Chat, Microsoft Outlook Chat, ecc.). Sono, invece, esclusi i messaggi collegati a servizi di messaggistica diretta presenti sui social media (Facebook Messenger, Twitter, Instagram, TikTok, YouTube, ecc.); i messaggi di chat riconducibili alle piattaforme di collaborazione (Microsoft Teams, SharePoint, ecc.), che perdono il loro valore quando vengono separati dai contenuti condivisi nella piattaforma di collaborazione (ad esempio, un documento di progetto su cui si sta lavorando insieme) e a cui normalmente fanno riferimento; allo stesso modo sono esclusi i messaggi di chat scambiati all'interno di una piattaforma di videoconferenza (Zoom, Skype, Google Meet, GoToMeeting, Cisco WebEx, ecc.) che, anche in questo caso, perdono il loro valore se separati dai contenuti a cui si riferiscono (ad esempio, una registrazione video di un incontro o un convegno). Senza queste esclusioni, le agenzie federali sarebbero costrette ad acquisire tutte le chat dalle piattaforme di collaborazione o di videoconferenza a cui partecipa anche una sola “figura Capstone”.

Il *General Records Schedule* non detta un insieme di regole e procedure pre-stabilite da osservare, ma propone piuttosto delle considerazioni sulle quali le singole agenzie possono basarsi per definire le proprie policy di conservazione dei messaggi, lasciando tuttavia un certo margine di intervento. Alle agenzie federali è richiesto di predisporre sistemi che permettano di conservare in maniera permanente i contenuti scambiati attraverso i sistemi di messaggistica istantanea, per periodi che vadano dai 15 ai 30 anni, o comunque dopo la loro declassificazione, nel caso possano essere soggetti a tale procedura (PAR-ER 2023). In conclusione, il Bollettino del 5 gennaio 2023 dei *National Archives* rappresenta un importante passo avanti, in quanto estende l'approccio *role-based* del metodo Capstone alla gestione dei messaggi elettronici di qualunque natura, ricomprendendo tra questi anche i messaggi istantanei. Come risulta evidente, il nodo cruciale è l'individuazione di quei messaggi che, per il loro contenuto, per la rilevanza della persona a cui sono stati inviati o da cui sono stati ricevuti, o per altre motivazioni, sono meritevoli di essere acquisiti e conservati. Questo è il vero punto cruciale e su questo servono delle regole chiare e basate su principi archivisticamente solidi.

Delle indicazioni interessanti contenute anche nella *Policy on instant Electronic massaging Records* del *New York City Department of Records and Informa-*

*tion Services* (The New York City Department of Records and Information Services 2022). Dopo aver precisato che tutti i dati e le informazioni in formato digitale, prodotti o ricevuti in base a leggi o ordinanze, o in relazione alle attività istituzionali sono considerati documenti comunali, la policy scoraggia i dipendenti comunali dall'utilizzare i sistemi di messaggistica istantanea per scopi diversi dall'invio di brevi comunicazioni legate al lavoro, sempre che non siano disponibili o risultino poco tempestivi altri metodi, come la posta elettronica. Inoltre, viene effettuata una distinzione tra messaggi "sostanziali" e "transitori": i primi sono quelli che documentano o giustificano decisioni e azioni di un ufficio dell'ente, al contrario dei messaggi transitori che sono relativi a questioni di limitata rilevanza o puramente logistiche (ad esempio, i messaggi scambiati tra colleghi per l'individuazione dell'orario o di un luogo per una riunione). I dipendenti sono fortemente scoraggiati dal creare messaggi elettronici istantanei di natura sostanziale. Tuttavia, nel caso vengano creati, devono essere importati su un sistema che ne consenta la corretta gestione secondo il piano di conservazione dei documenti dell'ente. Riguardo la modalità di importazione dei messaggi sostanziali, la policy suggerisce di farne uno *screenshot* sul proprio dispositivo mobile e di inoltrarlo alla propria email di lavoro, oppure di salvarlo sul servizio cloud comunale (come Microsoft OneDrive). I messaggi transitori, invece, possono essere eliminati non appena perdono il loro valore sia per il mittente che per il destinatario e possono anche essere programmati per la cancellazione automatica.

In ultimo, meritano una segnalazione le raccomandazioni fornite dal rapporto *WhatsApp in government* (Dourrant, Lilly, e Tingay 2022) che si basano su due indicazioni fondamentali. La prima riguarda l'importanza di individuare i messaggi che devono essere conservati: molti messaggi istantanei sono puramente operativi e non devono essere acquisiti, ma altri sono importanti e devono essere conservati. Allo scopo di rendere questa operazione di selezione meno soggettiva possibile, il rapporto cita il *Department for Levelling Up, Housing and Communities* (DLUHC) che fornisce una lista delle tipologie di messaggi da conservare, come quelli che riguardano le "decisioni di iniziare o terminare un progetto" e le "approvazioni di spesa": si tratta, in sostanza, di quei messaggi che sono utili per giustificare una decisione assunta o motivare un'azione intrapresa. In caso di dubbio sul valore documentale di un messaggio, occorre rivolgersi al responsabile della gestione documentale del dipartimento, che ha le competenze per dirimere i casi dubbi. Inoltre, per quanto la policy stabilisca che i messaggi selezionati debbano essere trasferiti sul sistema di gestione documentale entro poche settimane dall'invio o dalla ricezione, c'è pur sempre un periodo di tempo in cui messaggi potrebbero essere cancellati, accidentalmente o sulla base di policy preimpostate (come quella che stabilisce la cancellazione automatica in WhatsApp); pertanto il rapporto raccomanda che i messaggi non vengano eliminati se non dopo essere stati verificati relati-

vamente alla loro importanza come documenti; inoltre, suggerisce di disattivare tutte le funzioni che consentono l'auto-eliminazione. La seconda indicazione è relativa alla necessità di individuare metodi e strumenti per trasferire i messaggi nel sistema di gestione documentale in uso. Su questo punto alcuni dipartimenti hanno già fornito le prime indicazioni: ad esempio, il *Department for Business, Energy and Industrial Strategy* (BEIS) ha stabilito che i messaggi debbano essere «saved in SharePoint as BEIS' official records management system to ensure accountability and transparency» mentre il DLUHC fornisce delle istruzioni che guidano passo passo il personale nell'operazione di trasferimento delle chat di WhatsApp sui sistemi governativi. Tuttavia, non sembra esserci un approccio uniforme tra i vari dipartimenti dell'amministrazione britannica, probabilmente in considerazione della diversità di sistemi di gestione documentale in uso e anche della novità della materia. Negli Stati Uniti, il *Presidential and Federal Records Act Amendments* del 2014 (United States Congress 2014) richiede che «an officer or employee of an executive agency [...] forwards a complete copy of the record [from a non-official electronic messaging account] to an official electronic messaging account [...] no later than 20 days after the original creation or transmission of the record». Assicurarsi che i messaggi di WhatsApp siano trasferiti dal sistema di messaggistica elettronica non ufficiale (ad esempio WhatsApp) al sistema ufficiale (ad esempio, quello di posta elettronica) consentirà poi il loro versamento ai *National Archives* una volta trascorso il periodo previsto dalla normativa statunitense. Nel caso in cui un funzionario avesse la necessità di cancellare i messaggi (ad esempio, per liberare “spazio di memorizzazione”) o ci sia la possibilità di perderli (ad esempio, a causa di un guasto al proprio dispositivo mobile), è opportuno che venga contattato il responsabile della gestione documentale del dipartimento al fine di mettere in atto le strategie opportune.

## 5. Conclusioni

Come si è potuto dedurre dalle considerazioni esposte nei paragrafi precedenti, la messaggistica istantanea è ormai ampiamente utilizzata nelle pubbliche amministrazioni e nelle organizzazioni private e in molti casi si sta velocemente sostituendo ad altri strumenti considerati più “istituzionali”, come la posta elettronica. Sebbene nella maggior parte dei casi venga utilizzata per inviare e ricevere informazioni effimere e transeunti, che non necessitano di essere conservate, vi sono, tuttavia, diverse situazioni in cui i messaggi istantanei sono rilevanti e devono essere conservati, sia per finalità di tipo giuridico-amministrativo che per future finalità di ricerca storica. Personaggi politici, ministri, membri del parlamento, dirigenti e funzionari di pubbliche amministrazioni, manager di aziende e organizzazioni utilizzano quotidianamente i messaggi istantanei e sarebbe del tutto irrealistico e controproducente cercare

di porre un freno a questo fenomeno, che non è più arginabile e che risulta ancor più complicato per l'assottigliarsi del confine tra l'uso professionale e quello privato di queste nuove forme di comunicazione. Piuttosto, occorre cercare di governarlo ed è, quindi, urgente avviare una riflessione su questi temi, partendo dalla delicata questione della selezione dei messaggi istantanei aventi un valore documentale – alla quale fino ad oggi è stata data una scarsa importanza, forse anche in virtù del fatto che, generalmente, si associa il messaggio istantaneo ad una forma di comunicazione informale e dal contenuto “effimero” – per giungere all'individuazione di metodologie e strumenti per il trasferimento dei messaggi nei sistemi di gestione documentale e per la loro conservazione, che tenga conto anche delle questioni relative alla tutela dei dati personali e alla privacy, fino ad arrivare alla condivisione di buone pratiche e alla redazione di linee guida e raccomandazioni, che andranno a beneficio di tutti coloro che si troveranno, a breve, a dover gestire questi nuovi sistemi di comunicazione.

Alla luce di quanto esposto, si comprende come il problema della conservazione della messaggistica istantanea costituisca per la comunità degli archivisti una nuova sfida che deve essere affrontata con urgenza (Mitcham 2022) dal momento che qualsiasi ritardo nell'avviare la ricerca di idonee strategie e nell'acquisizione della consapevolezza che i messaggi istantanei possono rappresentare una fonte fondamentale e insostituibile per la ricostruzione del periodo storico che stiamo vivendo, porteranno inevitabilmente a gravi perdite nella memoria istituzionale degli enti e delle organizzazioni e, più in generale, nella comune eredità culturale. È davvero giunta l'ora che gli archivisti comincino ad occuparsene.

## Riferimenti bibliografici

- Allegrezza, Stefano. a cura di. 2023. *Personal Digital Memories: i fondi di persona dall'analogico al digitale*. Torre del Lago Puccini: Civita Editoriale.
- Alma Mater Studiorum Università di Bologna. 2023. “Summer School Web and Social Media Archiving and Preservation.” <https://site.unibo.it/web-and-social-media-archiving-and-preservation>.
- AvveComm. n.d. “Legittimo il licenziamento via whatsapp.” Consultato il 30 maggio 2023. <https://www.avvecomm.it/legittimo-licenziamento-via-whatsapp>.
- Battaglia, Elio. 2022. “Messaggi WhatsApp, modalità di acquisizione sulla utilizzabilità nell'ambito di un procedimento penale. Nota a sentenza n. 39529/2022.” Il Sole24ore, 4 novembre 2022. [https://ntplusdiritto.ilsole24ore.com/art/messaggi-WhatsApp-modalita-acquisizione-utilizzabilitaambito-un-procedimento-penale-AENeIEEC?refresh\\_ce=1](https://ntplusdiritto.ilsole24ore.com/art/messaggi-WhatsApp-modalita-acquisizione-utilizzabilitaambito-un-procedimento-penale-AENeIEEC?refresh_ce=1).



- Carucci, Paola. 2019. *Le fonti archivistiche: ordinamento e conservazione*. Roma: Carocci editore.
- Cosenza Vincenzo. 2023. “I social e gli instant messenger più usati al mondo: le mappe del 2023.” Viscos Blog, 4 febbraio 2023. <https://vincos.it/2023/02/04/i-social-e-gli-instant-messenger-piu-usati-al-mondo-le-mappe-del-2023>.
- Dourrant, Tim, Alice Lilly, and Paeony Tingay. 2022. *WhatsApp in government. How ministers and officials should use messaging apps – and how they shouldn't*. London: Institute for Government.
- European Commission. 2020. “Commission Decision (EU) 2021/2121 of 6 July 2020 on records management and archives (C/2020/4482).” <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-%3A32021D2121>.
- European Commission. 2021. “Reply from EC to the Ombudsman’s Strategic Initiative on the recording of text and instant messages (SI/4/2021/MIG).” 25 November 2021. <https://www.ombudsman.europa.eu/en/doc/correspondence/en/158616>.
- European Ombudsman. 2021. “Letter from the European Ombudsman to the European Commission on how EU institutions, bodies, offices and agencies record text and instant messages sent/received by staff members in their professional capacity.” Strasbourg, 30 June 2021. <https://www.ombudsman.europa.eu/en/doc/correspondence/en/143787>.
- European Ombudsman. 2022a. “Closing note on the strategic initiative on how EU institutions, bodies, offices and agencies record text and instant messages sent/received by staff members in their professional capacity (SI/4/2021/MIG).” Strasbourg, 13 July 2022. <https://www.ombudsman.europa.eu/en/doc/correspondence/en/158383>.
- European Ombudsman. 2022b. “The recording of text and instant messages sent/received by staff members in their professional capacity - Practical recommendations for the EU administration.” Strasbourg, 14 July 2022. <https://www.ombudsman.europa.eu/en/doc/correspondence/en/158302>.
- Mitcham, Jenny. 2022. “What’s up with using WhatsApp?” DPC (Digital Preservation Coalition), 31 March 2022. <https://www.dpconline.org/blog/what-s-up-with-using-whatsapp>.
- NAA (National Archive Australia). 2023. “Managing social media and instant messaging (IM).” Consultato il 23 maggio 2023. <https://www.naa.gov.au/information-management/manage-information-assets/types-information/managing-social-media-and-instant-messaging-im>.

- NARA (National Archives and Records Administration). 2013. "Guidance on a New Approach to Managing Email Records." *Bullettin* no. 2013-02, 29 August 2013.
- NARA (National Archives and Records Administration). 2015. "Guidance on Managing Electronic Messages". *Bulletin* no. 2015-02, 29 July 2015. <https://www.archives.gov/records-mgmt/bulletins/2015/2015-02.html>.
- NARA (National Archives and Records Administration). 2016. *Frequently Asked Questions about Instant Messaging*. <https://www.archives.gov/records-mgmt/initiatives/im-faq.html>.
- NARA (National Archives and Records Administration). 2023a. "Expanding the Use of a Role-Based Approach (Capstone) for Electronic Messages." *Bulletin* 2023-02, 5 January 2023. <https://www.archives.gov/records-mgmt/bulletins/2023/2023-02>.
- NARA (National Archives and Records Administration). 2023b. "General Records Schedule 6.1: Email and Other Electronic Messages Managed under a Capstone Approach." Transmittal no. 33, January 2023. <https://www.archives.gov/files/records-mgmt/grs/grs06-1.pdf>.
- NARA (National Archives and Records Administration). 2023c. "Frequently Asked Questions (FAQs) about GRS 6.1, Email and Other Electronic Messages Managed under a Capstone Approach." <https://www.archives.gov/files/records-mgmt/grs/grs06-1-faqs.pdf>.
- New Zealand Archives. 2020. "Text messages and other communications." 25 November 2020. <https://www.archives.govt.nz/manage-information/how-to-manageyour-information/digital/text-messages-and-other-communications>.
- PAR-ER (Polo Archivistico dell'Emilia Romagna). 2023. "Stati Uniti: conservazione digitale anche per la messaggistica chat e mobile." <https://poloarchivistico.regione.emilia-romagna.it/notizie/america-conservazione-digitale-anche-per-la-messaggistica-chat-e-mobile-1>.
- RAI News. 2022. "Ci sentiamo su WhatsApp: 33 milioni di italiani utilizzano la chat. Seguono Messenger e Telegram." 16 agosto 2022. <https://www.rainews.it/articoli/2022/08/whatsapp--la-chat-preferita-da-33-milioni-italiani-seguono-messenger-e-telegram-3d8f11ba-4fe2-40ac-b944-49fd-06d3036b.html>.
- Romiti, Antonio. 2020. *Archivistica generale. Primi elementi*. Torre del Lago Puccini: Civita Editoriale.
- Roncaglia, Gino. 2023. "Memorie personali ed ecosistema digitale." In *Personal Digital Memories: i fondi di persona dall'analogico al digitale*, a cura di Stefano Allegrezza. Torre del Lago Puccini: Civita Editoriale.



Statista. 2023. “Most popular global mobile messenger apps as of January 2023, based on number of monthly active users.” <https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps>.

The New York City Department of Records and Information Services. 2022. “Policy on Instant Electronic Messaging Records.” 5 January 2022. <https://www.nyc.gov/assets/records/pdf/Records-Management/RMD-P-2022-01%20-%20Instant%20Electronic%20Messages.pdf>.

Treccani. n.d. “Emoji.” Consultato il 30 maggio 2023. <https://www.treccani.it/enciclopedia/emoji>.

United States Congress. 2014. “H.R.1233 – Presidential and Federal Records Act Amendments of 2014.” <https://www.govinfo.gov/content/pkg/COMPS-11304/pdf/COMPS-11304.pdf>.



# A veritistic turn in information science?

An answer in an infodemic scenario

Carlos Alberto Ávila Araújo\*

**Abstract:** This article deals with an epistemic problematization for information science based on Jonathan Furner’s proposal of a veritistic turn for the field. Furner calls for the consideration of “truth” as a central concept to the field instead of “relevance”. In this text, his arguments – based on social epistemology and epistemic justice – are confronted with a set of questions relating to the contemporary phenomenon of post-truth. Elements raised by both discussions are analysed in the light of paradigms and “turns” already undergone by information science, to assess the relevance of a possible veritistic science in the area.

*Keywords:* Information science epistemology, Veritistic turn, Post-truth, Conceptual turns in information science.

## 1. Introduction

The aim of this text, whose title poses a question, is to reflect on the pertinence of proposing a possible veritistic turn in the field of information science. To do this, first, a discussion of “turns” in information science is presented. Studies in the epistemology of information science highlight that this area was born within a physicist perspective in the 1960s and underwent its first “turn” in the late 1970s, when a cognitivist perspective made its contribution. This was followed by a second turn in the mid-1990s, when a pragmatist sociocultural perspective was constructed.

The following are some questions about the truth and the current post-truth and infodemic scenario in information science. A number of aspects are used to characterize the phenomenon of post-truth – from scientific denialism, human confirmation biases, the bubble effect, to the mass proliferation of fake news, among others – which are generating a profound change in the

---

\* School of Information Science, Federal University of Minas Gerais, Belo Horizonte, Brazil. carlosaraujoufmg@gmail.com.

ways in which information is produced, received and reproduced (Santaella 2019) thus invoking a conceptual change in the field of information science.

In a third moment, the proposal of a “veritistic turn” is presented, from Fallis and Budd, arriving at Furner’s most recent proposal. Furner’s proposal is based on social epistemology, epistemic injustice and human rights, and articulates the concepts of truth, relevance and justice. This discussion is confronted with a second one, not found in the author’s work, which is an understanding of contemporary times as an era of post-truth.

While, admittedly, a “turn” cannot cancel the importance of previous knowledge, but raises new problems and ways of approaching them, I ask how far the conditions are given for the field of information science to experience a new “turn” from the third decade of the twentieth century. Such a movement would prepare the field more fully for the study of contemporary realities.

The article is thus structured in four topics: the epistemological discussion of information science revolves around the idea of “turns”, the question of truth, the question of post-truth and the proposal of a veritistic turn to information science.

## 2. “Turns” in information science

The idea of a “turn” in information science first came about in the early 1980s, when the works of Brookes (1980) and Belkin (1980) exposed the need for information science studies to include cognitive dimensions. Since the 1960s information science has mainly focused on studying fluxes or processes of information transfers, on identifying the forces and agents behind such fluxes (Debons, Horne, and Cronenweth 1998; Davis and Shaw 2001; Gilchrist 2009). In the late 1970s, concerns about meeting the needs of users triggered reflections about information not as something contained in data, in documents. Information could only be considered as such in relation to subjects’ state of knowledge, and its value could be determined only by the extent to which it altered people’s knowledge. This major conceptual shift in the field was given different names: Dervin and Nilan (1986) saw it as a change from the “traditional paradigm” to the “alternative paradigm”; Ellis (1992), from a “physical to a cognitive paradigm”.

In the 1990 new conceptual changes in the field brought about reflections which did not restrict information to subjects’ cognitive dimension but highlighted its connections to all of subjects’ actions in the world and to the context of such actions; information was not an individual phenomenon between persons and data, but existed on an intersubjective, social level. This change was called the “socio-cognitive turn” (Hjørland and Albrechtsen 1995) or “sociological turn” (Cronin 2008).

Several authors have put forth epistemological frameworks for information science that consider both trends. Capurro (1992; 2003), pioneeringly presented what he called the area's three paradigms: the physical, the cognitive and the social paradigms. The same organization is espoused by many others, like Ørom (2000), Molina and Moya-Anegón (2002), Silva and Ribeiro (2002), Linares Columbié (2005), Salaün and Arsenault (2009) and Bawden and Robinson (2012). While Saracevic (1999) does not refer to "turns" or "breaks", the author sees the history of information science as an increasing expansion in the concept of information: from a restricted perspective, a narrow sense (tantamount to a sign or a datum) to a wider or broader sense (more knowledge-related) one and an even wider or broadest sense (integrated into human action and a certain context).

According to this division between three major ways of studying information. The first of them, which began in the 1960s, essentially focuses on a physicalist (hence objective) conception of information, on the idea that information consists of the fastest, cheapest, and most efficient transport of data. Information science must be involved with the development of services, systems and products to ensure the success of transfer processes, acting directly on the flows.

The second way of studying information focuses on a conception focused on users, taken in their cognitive dimension. It is, therefore, a subjective approach and seeks to analyze how people miss information, seek, and use information. Information science must focus on the development of services and products that replicate these human mental activities.

Finally, in the third major way of studying information, the focus is on the intersubjective dimension, that is, on the social constitution of needs, processes of search and use of information. The links between the human action of producing and using information and the very constitution of culture or collective memory are highlighted. In a recent and ample systematic approach to the area's history, covering its almost 60 years of existence, Hjørland (2018a; 2018b) identifies what he calls the six paradigms or traditions in the study of information science: eminently practical studies (without theory); information theory (Shannon and Weaver's mathematical theory of computing); Cranfield's tradition (of the physical paradigm); the cognitive vision; the philosophy of information and the sociocultural vision.

Another recent discussion, and one which quite specifically addresses "turns" in information science is developed by Hartel (2019). Hartel identifies the starting point of information science in the physical paradigm and argues that, since 1986, the area has experienced seven turns: the cognitive turn, the affective turn, the turn of neo-documentation, the socio-cognitive turn, the everyday-life turn, the socio-constructivist turn and the bodily turn.

None of the mappings presented here – not even the most recent ones from the last two years – mention the question of truth or post-truth in any way. Naturally, the main reason is that systematic epistemological approaches are inevitably done *a posteriori*, that is, after publishing. More time would therefore be necessary for works on the field's epistemology to be incorporated.

Thus, it does not mean that the question of truth was not present in information science, in its other models of study. In physicalist perspectives, for example, expert assessments, peer review processes, quality criteria for scientific journals, consistent classification systems, among others, have always been elements with a “veritistic” dimension insofar as they sought objective criteria for information quality. In the same way, from the cognitivist perspective, user judgment criteria also imply a veritistic dimension, insofar as there was a whole work of scientific analysis of the criteria of relevance and preference of users for the design of information retrieval systems. And in the sociocultural approach, the idea of truth of a collective or domain is present.

The difference from the present moment is in the importance and volume of false information – in what has been called an “infodemic”. This concept means an association of the terms information and pandemic characterizes a pathological characterization of the informational dimension: the gigantic scope and speed of dissemination of false information has produced a situation in which false information is more present in people's lives than true and quality information and end up having much more influence in decision-making and in defining courses of action (Zarocostas 2020; Zielinski 2021). This constitutes a “pandemic” nature of informational phenomena, taken from the perspective of their adverse effects or dysfunctions. The term was created in 2020 and promoted by the World Health Organization to precisely designate the role of information in the pandemic scenario caused by Covid-19 (World Health Organization 2020; Pan American Health Organization 2020). In this sense, the term designates the new general conditions through which information is produced, circulated, disseminated, received, used, and appropriated by people at the contemporary moment. The exercise of identifying such conditions also implies considering the technological means of such production and circulation, the economic and legal forces that act on these processes, and the ways people behave in relation to them.

Exactly for that reason, a series of works produced in the past few years, presented at conferences, or published in journals, have raised the issues of post-truth and fake news in the area. Such works have focused on various topics and their impacts on the work of librarians, archivists, and other information professionals (Schlesselman-Tarango 2017; Agosto 2018; Naeem and Bhatti 2020; Revez and Corujo 2021). Studies are still grappling with assessment, implications, and the possible ways of professional engagement with this very recent topic to combat or mitigate negative effects. A more profound

reflection on the meaning of this phenomenon for the very concept of information is yet to be developed.

It is nonetheless worth pointing out that, even within past theoretical models, significant contributions still provide insights into aspects of post-truth. From a physicist perspective some aspects of the logic of algorithms can be identified and the “success” of certain contents, from the mechanisms favoring the popularity of sites and information sources as a measure of relevance and retrieval. From a cognitive perspective, the impact of cognitive biases of confirmation and dissonance can be assessed in the perceptions of lacunae of knowledge, in the identification of search strategies and the search for and use of information. Also, from a sociocultural approach, post-truth can be analyzed as a culture, as a disregard for truth as an attribute of information and a social construction. All three dimensions of the phenomenon of post-truth could therefore be studied by information science. The field is far from ill-prepared for this study. However, the specific problematization of the attribute of information’s “truth” has never been central in it, and this is exactly the focus of Furner’s claim before contemporary phenomena, and the informational realities they raise.

### 3. The question of truth

Naturally, the question of truth, that is, whether the information being stored, organized, and disseminated in information systems and services is true, has always been present in information science. But not a discussion of what the truth is.

The question of the meaning and conditions of existence of “truth” is a problem that has always crossed philosophy. Several authors have even dedicated themselves to systematizing theories and understandings about truth existing in philosophy. Kirkham (1992), for example, presents the theories of correspondence, coherence, pragmatic, semantic, performative, redundancy, appraisal, and truth-as-justification. Roark (1982) presents the following tests or dimensions of truth: correspondence, coherence, pragmatic, verification, and performative. Raatikainen (2021) presents the following theories of truth: classical correspondence, coherence, pragmatist, Epistemic, Formal Approaches, Deflationist and Minimalist.

But it is not exactly in philosophy that the necessary foundations for information science to deal with the question of truth lie. After all, information science does not deal with the philosophical aspects of truth. Information science deals with information sources, with knowledge production authorities, with institutions that create, certify, or reproduce knowledge.

In this sense, it is within the scope of theories that study the social and institutional modes of truth creation that information science is related.

This is the case, for example, of Burke (2000), who studied how different social institutions (universities, academies, craft corporations, markets, states) acted to promote certain knowledge and interfere in the processes of collection, classification, dissemination and sometimes withholding information. In a later work, the same author (Burke 2012) studied how apparently timeless activities – the acquisition and gathering of knowledge, its analysis and organization, its dissemination, and its effective use – are in fact limited and conditioned by the performance of knowledge institutions and professions and take different forms at different times and places.

A similar analysis is made by Blatt (2018), who traces an evolution of the concept of truth, as a socially lived issue, since the 17th century. The author points out the transition from a moment when the truth came from the religious authorities, passing through its inscription in the scope of the performance of institutions such as the university, science, and journalism, and then with the performance of advertising, social networks, and other mediations.

In this sense, a close perspective is the notion of a “regime of truth” introduced by Foucault (1975) in his work *Discipline and Punish*. With this concept, the author sought to see how knowledge and truth were produced by the power structures of a given society. In his study of the penal system of the 18th and 19th centuries, Foucault identified a corpus of knowledge, techniques and discourses that were entangled with the practice of the power to punish, giving rise to a new “regime of truth” (Weir 2008).

What there is in common in the analyzes of authors such as Burke, Blatt and Foucault is the perception that what is accepted as true at a given time depends on the interaction of different social forces. The authors study different phenomena and processes such as the invention of the press, the action of libraries, the constitution of the university, the production of the encyclopedia, among others. A complementary analysis to this is that of Giddens (1991), who studies the mechanisms through which modernity was constituted. In his analysis of how the transition from pre-modern societies to modernity took place, the author identifies the occurrence of several phenomena that profoundly changed the various dimensions of human life (politics, economy, culture, regulation, work). These phenomena produced a new way of life and social organization that emerged in Europe from the 17th century onwards, and which later became worldwide in their influence. Among these changes is the so-called disengaging of social systems, which altered spatial and temporal relationships and inserted a rationalized organization into human life. Giddens is dedicated to studying these mechanisms, which are of two types: symbolic tokens and expert systems. Both fundamentally depend on trust: it is essential for the constitution of modernity’s institutions.

Expert systems are defined by Giddens as structures of technical excellence or professional competence that organize large areas of the material and social



systems in which we live. The author brings, as an example, a ladder, which we use with the certainty that we will not fall, that it will not break – that is, we accept the risk, because we believe in the expertise of those who produced it. Expert systems work in all spaces and environments. Each person, throughout their lives, is faced with situations and problems in which their own knowledge is null or rudimentary (for example, the need to undergo surgery, or the repair of microelectronic equipment) and those situations attribute protagonism in solving problems to another professional actor, endowed with recognized knowledge in that area. The activities start to take place, therefore, despite the knowledge of each of the people involved.

Expert systems allow a vast range of human activities to be performed with greater effectiveness, efficiency, accuracy, and productivity, precisely because they are performed by people with specific training, prior knowledge, and skills. In caring for the body, housing, food, human relationship, in all spheres of human life, it is possible to have a more rational and productive action based on guidance by a functionally instructed and specialized professional. This has happened, therefore, in medicine, engineering, nutrition, gastronomy, and in several other fields. It also took place in the informational field.

Before proceeding, it is important to point out that the performance of expert knowledge is not a creation of modernity. Before modern societies, there were craftsmen's guilds, artisans, specialized knowledge, even universities. The novelty brought by modernity was a complex structure for validating and certifying these expert systems, through professional training courses, supervisory boards, regulatory legislation, among others. And, above all, a broad promotion of the trust to be placed in such expert systems, precisely because of the entire previous certification.

Libraries, archives, museums, and other institutions that deal with human registered knowledge have existed for centuries and conduct, using the current categories of thoughts, actions of "information mediation," in the sense of acting together with human knowledge by selecting, preserving, organizing, disseminating. In modernity, such institutions are supported by scientific knowledge (archival science, library science, museum studies) that provide institutional, professional, legal, and technical support for their interventions with the societies in which they operate. Such institutions and knowledge, throughout their existence in modernity, dealt with different issues: the universalization of access to their contents (democratization); the search for diversity in their collections and actions (epistemic justice); the sophistication of instruments for organizing the knowledge (efficiency for preservation and recovery), among others.

As pointed out in the previous topic, there is a new scenario of production, circulation, and use of information, which has been called infodemic or post-truth. In this scenario, the question of whether the information is true or false

acquires new relevance. In the next topic, this issue will be addressed, based on the unfolding of the post-truth concept.

#### 4. The question of post-truth

As Peters et al. (2018) had noted, the term “post-truth” was first used by Steve Tesich in 1992 in his study of the Gulf War and appeared in a book title for the first time in the work by Ralph Keyes, published in 2004. But it was in 2016 that the expression became widespread, to the point of being considered the word of the year by the Oxford Dictionary, to designate the circumstances under which objective facts become less influential in making public opinion than appeals to emotion and personal belief (D’Ancona 2017). In 2016 the term also became closely associated with two fundamental facts of international politics: Donald Trump’s election for the US presidency and the victory of plans to withdraw Britain from the European Union, known as the acronym Brexit (and abbreviated form of ‘British exit’). The phenomenon of post-truth came to be seen as involving a series of aspects and levels of problems, and various researchers from several areas and countries have engaged in studying and correlating those aspects and levels. From such discussions, the phenomenon can be seen as showing three major dimensions.

The first is connected to technological dynamics and the logic of “personalized” information reinforced by the algorithms structuring search engines and social networks – currently the primary environment from which people receive world news and global information (Kakutani 2018; Noble 2018). These are built from algorithms that select what people probably want or that which confirms their points of view, an effect known as the “bubble effect” (Magallón Rosa 2019). In social networks like WhatsApp, messages are sent massively to people’s devices, without any monitoring or contrast, in an “underground” logic of information dissemination. With the formation of “bubbles” or “echo chambers” that seal users off from new ideas, topics and important information”, and especially when it comes to politics, people are exposed almost exclusively to unilateral visions within the broader political spectrum (O’Connor and Weatherall 2019).

Alongside this phenomenon there is also the mass dissemination of fake news. While this is not exactly a new development it operates in a new logic: fake news travels apocryphally, forwarded by common people, exposing the lack of regulations for them as opposed to the controls of journalistic or educational institutions. The situation suggests that all information would have the same weight or value regardless of quality, of checks and of institutional commitments behind their production. The phenomenon is further compounded by the action of clickbait, that is, the dissemination of false content or injection of sensational headlines to entice users into accessing content to

profit from publicity (Aparici and García-Martín 2019). In the dissemination dynamics of this mass output of fake news, lies become active in shaping people's decision making in different spheres (politics, the economy, education, health, religion) at previously unseen speed.

The second dimension is related to the human level or, more specifically, connected to human cognitive dimensions: the so-called cognitive bias, or confirmation bias, or even cognitive dissonance. This is a tendency in human beings to shape their beliefs and worldviews without basing themselves on reason and evidence, that is, on facts, to avoid psychic discontent. McIntyre (2018) bases his point on three classical studies in social psychology conducted in the United States during the 1950s and 1960s. The first is Festinger's theory of cognitive dissonance, according to which we seek harmony between our beliefs and actions. The second is Asch's theory of social conformity, that states our tendency to yield to social pressure is rooted in our urge to be in harmony with others. The author also puts forth more recent studies about this question, expressed in two concepts: the counterproductive effect (a phenomenon where presenting someone with true information which conflicts with their beliefs in fake data only makes them believe such facts even more fervently) and the Dunning-Kruger effect (by which our lack of capacity to act on something causes us to overestimate our true abilities). Such elements of cognitive bias incline people to shape their beliefs regardless of reason and evidence. This phenomenon is further compounded by the context described above as the bubble effect (Greifeneder et al. 2021).

There is still a third, cultural dimension – to the point that some authors refer to a “culture of post-truth” (Wilber 2017). At present, most people (except, of course, for a sector of global population below a certain economic threshold) have easy and instant access to technology and the possibility of checking the veracity of any given piece of information, through smartphones, notebooks, desktops, and other devices. Unlike other periods in history, when checking whether certain facts about, say, another country's way of life were true or false, today this can be easily verified from our homes. But this is not how people act. People take for real, forward, share and appropriate information they have not bothered to check. This disdain, this disregard for truth, in a context of such privileged access to information, is the novel fact that the idea of “post-truth” as culture seeks to describe.

In this sense, post-truth designates a condition, a context where attitudes of disinterest and even disdain towards the truth are naturalized, disseminated, turned widespread, normalized, and even encouraged. There is a process of accepting and replicating concepts that normalize disdain for the truth: it is an idea, an imaginary, a set of social representations or meanings incorporated by audiences that renders possible the existence of fake news referring to a cer-

tain idea and reaffirm or develop it (Murolo 2019). These dynamics empower clickbait and other mechanisms for the dissemination of false information.

The researcher Wilber (2017) has examined this phenomenon in a book stimulatingly entitled *Trump and post-truth*. He discusses Donald Trump's election as president of the USA, and the UK's exit from the European Union – two phenomena directly associated with the triumph of mass-produced, disseminated and consumed fake news, which oriented people's electoral choices. The author links these choices to other phenomena like the diminishing value placed on democracy, the increase of hatred, of racism, of xenophobia and of bad taste, among others. He thus frames post-truth within a broader worldwide process of change in cultural values – especially in western societies.

Wilber paints a picture of world-dominating values and ideas increasingly accepted currently (what he calls *avant garde*). He identifies that, in the first half of the twentieth century, a wide range of political, cultural, and intellectual movements drove the world according to values associated to the rational, the operational, the conscious, and notions of merit, profit, progress – that is, values directly related to the ideals of modernity. In his analysis, the author considers that since the 1960s, ideas associated to postmodern values gained momentum, such as the defense of plurality, of relativism, self-realization, inclusion, multiculturalism, civil rights, sustainability, the defense of minorities, etc. Continuing his analysis Wilber contends that the second decade of the twentieth century is witnessing a crisis of such a project, a failure of avant-garde progressiveness. This idea is also developed by other authors (Eatwell and Goodwin 2019; Broncano 2019; Casara 2019).

Wilber points out several factors as causes for such a failure. Among them are the relativization of the idea of truth, the notion that local, particular truths would exist, which bringing about a form of generalized narcissism. This causes the inability to communicate with others' perspective, a loss in sense of empathy and hatred towards minority points of view. In turn, this leads to essentialist visions, to racist tendencies, patriarchalism and misogyny. As a result, we are experiencing a crisis in the legitimacy of modern institutions, human rights, reason, science, and democracy (Fukuyama 2018; Cosentino 2020).

## 5. Jonathan Furner's proposal

There are works that discuss the question of truth in information science from a philosophical perspective, especially from the debate between realism and idealism (Hjørland 2004; 2021; Dobson 2001; Spasser 2002). However, another type of discussion, closer to the discussions of Burke, Blatt, Foucault and Giddens, that is, about the social and institutional conditions of construction of truth, is the one in which information science approaches social

epistemology. And it was in this type of theoretical proposal that the idea of a veritistic turn to information science emerged.

The first proposal of a possible veritistic turn in information science was made by Fallis (2000) from the proposal of a veritistic social epistemology made by Alvin Goldman. Fallis reports that this proposal sought to study the social practices that contributed to the production of errors and lies in the social creation of knowledge. The fields of action initially thought of by Goldman were science, law and education. Fallis proposed to bring this theory closer to information science through the incorporation of libraries and other information services and systems in the scope of the studies. In a later work, Fallis (2002) seeks to bring this idea of proposal closer to the social epistemology made by Jesse Shera and Margaret Egan in 1952, thus seeking a foundation for information science by bringing together the two discussions: that of social epistemology and that of the vertical turn.

The defense of Shera and Egan's social epistemology as a general foundation of information science is also carried out by other authors. Among them, Budd (1995) stands out, who proposes that this theory can be useful to conceive information science as an area dedicated to the study of all the ways in which a society deals with the knowledge that it produces and consumes. In a previous work, by the way, Budd (2001) had already analyzed the issue of error and false information in the scope of the construction of scientific knowledge and its impacts on information science.

Taking up points made by these authors, Jonathan Furner elaborated a new proposal. Furner is a professor at the Graduate School of Education and Information Studies at the University of California in the United States. For several years, he has been questioning the epistemic bases of information science, by approaching Shera's (Furner 2002) social epistemology, discussing broader philosophical questions (Furner 2010; 2015) and developing a problematization of the concept of information in all five subareas of information science (informational behavior, information retrieval, metric studies of information, information organization and information ethics) (Furner 2014).

In July of 2018 Furner presented a paper at the "XV International Conference of the ISKO (Information Society for Knowledge Organization)" – taking place in Portugal –, which was later published as the chapter of a book with the events proceedings (Furner 2018). In this work the author seeks an epistemic grounding for knowledge organization (KO) based on contributions from epistemology and ethics – and, more specifically, from social epistemology and epistemic justice. By articulating three concepts (truth, relevance, and justice), the author proposes a veritistic turn so that the area can provide a critical knowledge organization (CKO).

To make his point, Furner (2018) turns to philosophy, more specifically a branch-ontology which studies the philosophy of being, of existing things, of

types of things and how things can be classified. He proposes conceiving KO as an ontology where “facts” can replace “things” and thus become itself a “philosophy of data”. As such, Knowledge Organization would comprise elements from three of philosophy’s traditional branches: the philosophy of the mind, the philosophy of language and the philosophy of beliefs. The latter is taken by the author as a synonym for epistemology or of the philosophy of knowledge. Furner identifies the existence of two types of theories: truth-oriented theories, which can be defined as theories of belief distinguishing between true and false beliefs; and relevance-oriented theories, which can be defined as those theories of belief that distinguish relevant beliefs from irrelevant ones. Based on this categorization the author detects the existence of a historical break-up between epistemology as a subfield of philosophy and library science-information science: in the former, theories of belief are oriented towards the truth; in the latter, towards relevance – relevance becoming the main parameter in determining informational retrieval.

Still within the field of epistemology, Furner puts forth that epistemology can be divided into types following three criteria. The first distinguishes pure epistemology (made up of theories for the description of the nature of concepts and of epistemic practices) from applied epistemology (consisting of normative theories seeking to orient the most appropriate practices for obtaining true or relevant propositions). The second parameter is connected to methodology, and distinguishes rationalist from naturalist epistemologies, depending on the subject’s promptness to admit different types of evidence in favor of conclusions. Finally, according to the main factor in establishing beliefs, epistemologies can be individualistic (when the subject’s interests are considered) or social (when the focus is on social interaction). By combining such criteria, Furner proposes an applied social epistemology (the study of normative questions about those social practices which are more prone to generate true or relevant beliefs) as the adequate theoretical framework for his proposed CKO.

Furner then turns to the question of justice. He presents justice as a value (like truth, relevance, beauty, and liberty) related to a desirable characteristic in people’s decisions and actions. Justice is made when people are treated according to their merits or needs, without prejudice or discrimination, without violating their human rights or limiting their freedom, and without exercising any form of oppression resulting from asymmetric power relations. Furner identifies several types of rights (natural, human, civil, group and individual rights) related to equitable access to certain goods or opportunities. He then charts six types of rights within the field of information: the right to think (to conceptualize, to categorize and classify, believe and have an opinion); the right to express oneself (to voice one’s thoughts in speech, in writing and other ways); the right to access (the possibility to seek, research, find, hear and discover the thought and expression of others); the right to be listened to (to



publish and broadcast, to reach out to an audience without being censored, silenced, hidden or ignored); the right to be “left alone” (to keep one’s privacy); the right to credibility (to be treated as someone reputable).

By discussing theories of justice, Furner first identifies what he calls theories of social or distributive justice, those focused on the results of actions taken to distribute amounts of resources among the members of certain populations, according to fair criteria. These theories have the objective of achieving: the reduction of divides, disparities, and inequalities between rich and poor, or between the powerful and the powerless; fairer distributions of social, cultural, economic, and political opportunities where human rights and liberties are respected. Conversely, the author raises the need for theories of injustice or oppression to be contemplated. Theories that expose processes of exploitation, marginalization, cultural imperialism, and violence. Furner concludes that working for social justice involves the basic reformulation of oppressive and discriminatory social practices and institutions, as well as the redistribution of resources. Among such practices and institutions, he places those involved in the production and consumption of knowledge – among which are libraries, information services, knowledge organization systems (KOSs) and systems of bibliographical classification, topic heading lists and thesauri.

Continuing his argument, Furner introduces the theory of epistemic justice developed by Miranda Fricker (2017), which focuses on the equity that people are treated with in their ability to know and to have beliefs. Furner mentions the distinction made by Fricker between two types of epistemic injustice: the distributive type (which occurs whenever such epistemic resources as education or information are unfairly distributed); and the discriminatory type (which takes place whenever failings are attributed to an individual or group); the testimonial type (when preconceptions or deficits of authority are attributed to those producing certain discourses); the hermeneutic type (when subjects are hermeneutically marginalized, that is, they belong to groups with no access to an egalitarian participation in generating social meanings). In Furner’s assessment, social justice has become the aim of professionals working in libraries and information services. Still, while these professionals have appropriated applied social epistemology, their grasp of the theory of epistemic justice is still shallow.

Based on the categories and concepts developed through his argument, Furner lays out four arguments for the construction of a CKO: to base it on applied social epistemology (by identifying the conditions under which testimonies should be assessed as true or relevant); to be inspired by values of epistemic justice (not just social justice as the primary end of libraries and information services, but also justice in the dissemination and acquisition of true beliefs); respect of human rights (the right to testimonial justice, to have credibility); and finally, to privilege truth instead of relevance. Furner makes

his proclamation of a veritistic turn for information professionals based on this last point. A relevance-oriented organization of knowledge is that which seeks to assess information practices, institutions and products based on satisfying the wants and needs of users; a truth-oriented organization of knowledge is assessed on the grounds that the beliefs acquired by users are true. Once again, according to the previous discussion, in line with the institutions that certify the character of truth of the statements and discourses in circulation in each society.

One month after verbally presenting this work, in August of 2017, Furner took part in a symposium entitled “Social Epistemology as Theoretical Foundation for Information Science: Supporting a Cultural Turn”, at the University of Copenhagen, Denmark. There he gave a conference keynote entitled *Society, Epistemology, and Justice: Prospects for a Critical LIS?* Once again, this lecture set out from the intersection between epistemology and ethics to propose a critical library and information science. To this end he first approached the symposium’s focus, social epistemology, as a possibility for CLIS to apply values of truth and relevance in the construction of systems of access to information. Then the author defended the idea that CLIS’ mission goes beyond social justice and touches upon epistemic justice, that is, equal access for all people to world-recorded knowledge. As a third point he defended the relevance of a “veritistic turn” in information science, from which truth would substitute relevance as the main prerequisite for providing users with information. In his final proposal, he challenged the area’s attachment to the idea of relevance and problematized its adoption of ethical codes claiming neutrality. As a conclusion, he defended the need for a veritistic turn in the face of the “Trump era”, one dominated by the circulation of fake news and “alternative facts” (Hartel 2017). The link that this author made with the contemporary questions related to Donald Trump, the president of the United States elected in 2016, and the intense dissemination of fake news triggered the present problematization of his proposal from the standpoint of aspects, traits and dimensions in the phenomenon of post-truth.

A question that arises as central to the field of information science, in view of these discussions, concerns the expected impacts with a possible adoption of a veritistic turn for the field. Naturally, effecting a veritistic turn does not mean abandoning everything that was built over the previous decades. In this way, the challenges of building effective means of dissemination and promotion of access, more sophisticated mechanisms for retrieving information in databases, repositories, and other systems, as well as strategies for meeting the search processes and satisfying information needs, continue being extremely important for the field.

But the idea of a veritistic turn puts the need to verify and verify whether the information is true at the center of the question. This means that much



of the so-called “information processing” work, historically linked to subject identification, indexing, and retrieval, is now centered on the aspect of truth. Therefore, the question of the quality of information itself needs to stop having as its main criterion its “popularity”, the number of accesses, views, for its reliability. Greater emphasis needs to be given, therefore, to actions and services that carry out the check, that confirm the veracity. This includes promoting both reputable sources, that is, with a credible history, as well as checking agencies, services dedicated exclusively to carrying out research to identify and report false, distorted, denialist and hateful information.

Likewise, one of the expected results of adopting such a perspective in the field of information science is the collaboration with justice in establishing specific laws related to crimes that are committed through information. Although, in many countries, there are already laws for the punishment of crimes such as slander, libel and defamation, the informational reality has generated both new types of crimes as well as new conditions for their execution in terms of scope and speed. Many countries around the world are currently in the process of drawing up procedures so that social media platforms and search engines can be notified or held accountable for false information being circulated. The typification of false information itself needs to be carried out. A pioneering effort was made by Wardle and Derakhshan (2017), who distinguished between misinformation (dissemination of false information without the intent to cause harm), disinformation (disclosure of false information with the intent to cause harm or mislead) and malinformation (disclosure of information not necessarily false but taken out of context or distorted with a clear intent to mislead or confuse). But new categorizations still need to be made, including categorizations of what fake news, fake science, denialism, false testimonials, hate speech and other forms of disinformation are, and their different impacts on society (Araújo 2021).

In another line, it is expected that the adoption of a veritistic turn can help to increase critical information literacy (Downey 2016) actions, especially in what has been called critical competence in information. Information literacy, which began to be theorized and practiced from the reflection of Zurkowski (1974), has developed in recent decades centered on promoting the capacity of individuals to recognize their information needs, adopt the best search strategies, and use information in a productive, ethical, and responsible manner. In a post-truth scenario, it is necessary to add skills to identify the veracity of the information, the suitability of the sources, and it is this issue that has been worked on within a critical perspective. An example is the ability that people need to identify the existence of the bubble effect and think of strategies to break the bubbles in which they may be inserted (Ferrari 2018; Noble 2018). Or knowing when they are incurring cognitive biases. In both cases, people leaving their comfort zones in search of the contradictory, of information with

other types of points of view and worldview, is fundamental to overcome the challenges brought by the current post-truth scenario (Dalkir and Katz 2020).

## 6. Final considerations

To return to my initial argument, it is important to highlight that a “turn” in a scientific discipline does not mean that everything done before it should be abandoned. This is true for other scientific areas and information science is no exception. The cognitive turn of the 1980s did not put an end to studies shaped by a physicist perspective, centered on information transport and retrieval. Neither did the sociocultural turn bring about the extinction of studies centered on the triad: data-information-knowledge or of the centrality of users’ cognitive experience. Proposing a veritistic turn for information science in no way means abandoning either previous perspectives or the approach constructed over the past two decades around the social construction of information and its links to the social, political, economic, cultural, and technological dimensions shaping information regimes.

In any case it is worth assessing to what extent privileging the notion of “truth” in informational studies could imply or demand a “turn”. Some authors have criticized the recurrent “turns” or the emergence of yet another new “paradigm” every twenty or thirty years in information science – a very short span, that could be a symptom of mere fashions. Hartel’s own notion of a turn described above could be considered an exaggeration, as it refers to almost concomitant turns – and which therefore would not necessarily be in fact turns but theories or parallel trends within the same area.

As pointed out in the discussion laid out in this article, the past few years have very effectively put forth a new informational reality which has challenged technologies, the experiences of subjects and even the stability of democracy, science, and peace. It is in this sense – of providing the adequate instruments for new realities – that I put forth the present proposal, of granting a centrality to the notion of truth within contemporary information studies. The different conceptual models (or paradigms) developed by information science throughout its existence have always been directly linked to pressing problems at any given time. Thus, the area is again challenged to show its dual traits: respect for accumulated research findings and theorizations on the one hand and the versatility to produce new models in tune with the specific and empirical field of informational phenomena that must be understood in their full complexity.

In any case, regardless of whether there is a need for a veritistic turn, the most fundamental thing is for information science to be in tune with the informational challenges of its time.

The current moment has been designated by different authors as a society of ignorance (Serrano Oceja 2019), an era of resentment (Fukuyama 2018), an Orwell world (Gómez de Águeda 2019), a post-democratic era (Casara 2019), era of national populism (Eatwell and Goodwin 2019), the great setback (Geiselberger 2017). All of them point, in some way, to the failure of the promises of wisdom and peace made in the 1960s and 1970s around the idea of an “information society”. All point to the centrality of false information in the erosion of democracy, in the increase in hate speech, in the growth of prejudice, among other dimensions.

In addition to diagnosing the problem, it is also necessary for information science to develop intervention strategies and to combat its perverse effects. As pointed out at the end of the previous topic, the main actions have been pointed out are the creation of mechanisms to certify the veracity and quality of information, the construction of accountability mechanisms for crimes committed through false information, increasing the visibility and circulation of checking services, and the promotion of critical information literacy. The effective implementation of such actions is fundamental, above all, for the maintenance of certain values built in recent centuries, such as democracy, inclusion, the defense of diversity and the encouragement of a culture of peace.

## Bibliographic References

- Agosto, Denise E. 2018. *Information literacy and libraries in the age of fake news*. Chicago: American Library Association.
- Aparici, Roberto, and David García-Marín. 2019. *La posverdad: una cartografía de los medios, las redes y la política*. Barcelona: Gedisa.
- Araújo, Carlos Alberto Ávila. 2021. “Infodemia, desinformação, pós-verdade: o desafio de conceituar os fenômenos envolvidos com os novos regimes de informação.” *The International Review of Information Ethics* 30 (1).
- Bawden, David, and Lyn Robinson. 2012. *Introduction to information science*. London: Facet.
- Belkin, Nicholas. 1980. “Anomalous states of knowledge as a basis for information retrieval.” *Canadian Journal of Information Science* 5: 133–43.
- Blatt, Roberto. 2018. *Historia reciente de la verdad*. Madrid: Turner.
- Broncano, Fernando. 2019. *Puntos ciegos: ignorancia pública y conocimiento privado*. Madrid: Lengua de Trapo.
- Brookes, Bertram C. 1980. “The foundations of information science: part I: philosophical aspects.” *Journal of Information Science* 2 (3-4): 125–33.
- Budd, John. 1995. “An Epistemological Foundation for Library and Information Science.” *The Library Quarterly* 65 (3).

- Budd, John. 2001. *Knowledge and knowing in library and information science: a philosophical framework*. Lanham: Scarecrow.
- Burke, Peter. 2000. *A social history of knowledge: from Gutenberg to Diderot*. Cambridge: Polity Press.
- Burke, Peter. 2012. *A social history of knowledge II: From the Encyclopaedia to Wikipedia*. Cambridge: Polity Press.
- Capurro, Rafael. 1992. "What is information science for? A philosophical reflection." In *Conceptions of Library and Information Science*, edited by Pertti Vakkari and Blaise Cronin, 82-96. London: Taylor Graham.
- Capurro, Rafael. 2003. "Epistemologia e ciência da informação." In *Encontro Nacional de Pesquisa em Ciência da Informação. Belo Horizonte, Brazil, 10 de Novembro de 2003*.
- Casara, Rubens. 2019. *A era pós-democrática*. Porto: Exclamação.
- Cosentino, Gabriele. 2020. *Social media and the post-truth world order: the global dynamics of disinformation*. Cham: Palgrave.
- Cronin, Blaise. 2008. "The sociological turn in information science." *Journal of Information Science* 34 (4): 465-75.
- D'Ancona, Matthew. 2017. *Post Truth: The New War on Truth and how to Fight Back*. London: Ebury Press.
- Dalkir, Kimiz, and Rachel Katz, edited by. 2020. *Navigating fake news, alternative facts, and misinformation in a post-truth world*. Hershey: IGI Global.
- Davis, Charles, and Debora Shaw. 2001. *Introduction to information science and technology*. Medford: Information Today.
- Debons, Anthony, Esther Horne, and Scott Cronenweth. 1988. *Information science: an integrated view*. Boston: G. K. Hall.
- Dervin, Brenda, and Michael Nilan. 1986. "Information needs and uses." *Annual Review of Information Science and Technology* 21: 3-33.
- Dobson, Paul J. 2001. "Critical realism and information systems research: Why bother with philosophy?" *Information Research* 7 (2).
- Downey, Annie. 2016. *Critical information literacy: foundations, inspiration, and ideas*. Sacramento: Library Juice Press.
- Eatwell, Roger, and Matthew Goodwin. 2019. *National Populism: the revolt against liberal democracy*. New Orleans: Pelican Publishing Company.
- Ellis, David. 1992. "Paradigms and proto-paradigms in information retrieval research." In *Conceptions of library and information science: historical, empirical and theoretical perspectives*, edited by Pertti Vakkari and Blaise Cronin, 165-86. London: Taylor Graham.

- Fallis, Don. 2000. "Veritistic social epistemology and information science." *Social Epistemology* 14 (4): 305–16.
- Fallis, Don. 2002. "Introduction: social epistemology and information science." *Social Epistemology* 16 (1): 1–4.
- Ferrari, Pollyana. 2018. *Como sair das bolhas*. São Paulo: Armazém da Cultura.
- Foucault, Michel. 1975. *Discipline and punish: the birth of the prison*. New York: Random House.
- Fricker, Miranda. 2017. *Injusticia epistémica*. Barcelona: Herder.
- Fukuyama, Francis. 2018. *Identity: The Demand for Dignity and the Politics of Resentment*. New York: Farrar, Straus and Giroux.
- Furner, Jonathan. 2002. "Shera's social epistemology recast as psychological bibliology." *Social Epistemology* 16 (1): 5–22.
- Furner, Jonathan. 2010. "Philosophy and information studies." *Annual Review of Information Science and Technology* 44: 161–200.
- Furner, Jonathan. 2014. "Information without information studies." In *Theories of information, communication, and knowledge*, edited by Fidelia Ibe-kwe-Sanjuan and Thomas Dousa. London: Springer.
- Furner, Jonathan. 2015. "Information science is neither." *Library Trends* 63 (3): 362–77.
- Furner, Jonathan. 2018. "Truth, relevance, and justice: towards a veritistic turn for KO." In *Challenges and Opportunities for Knowledge Organization in the Digital Age*, edited by Fernanda Ribeiro and Maria Elisa Cerveira, 468–74. Porto: Faculdade de Letras da Universidade do Porto.
- Geiselberger, Heinrich. editado por. 2017. *O grande retrocesso*. Lisboa: Objectiva.
- Giddens, Anthony. 1991. *The consequences of modernity*. Stanford: Stanford University Press.
- Gilchrist, Alan. edited by. 2009. *Information science in transition*. London: Facet.
- Gómez de Águeda, Angel. 2019. *Mundo Orwell: manual de supervivencia para un mundo hiperconectado*. Madrid: Ariel.
- Greifeneder, Rainer, Mariela Jaffé, Eryn Newman, and Norbert Schwarz. edited by. 2021. *The psychology of fake news: accepting, sharing, and correcting misinformation*. London: Routledge.
- Hartel, Jenna. 2017. "Social Epistemology as Theoretical Foundation for Information Science: Supporting a Cultural Turn, Copenhagen, August 16-17, 2017." *Knowledge Organization* 45 (1): 79–84.
- Hartel, Jenna. 2019. "Turn, turn, turn." *Information Research* 24 (4).

- Hjørland, Birger. 2004. "Arguments for philosophical realism in Library and Information Science." *Library Trends* 52 (3): 488–506.
- Hjørland, Birger. 2018a. "Library and information science (LIS), Part I." *Knowledge Organization* 45 (3): 232–54.
- Hjørland, Birger. 2018b. "Library and information science (LIS), Part 2." *Knowledge Organization* 45 (4): 319–38.
- Hjørland, Birger. 2021. "Information retrieval and knowledge organization: a perspective from the philosophy of science." *Information* 12 (3): 135.
- Hjørland, Birger, and Hanne Albrechtsen. 1995. "Toward a new horizon in information science: domain analysis." *Journal of the American Society for Information Science* 46 (6): 400–25.
- Kakutani, Michiko. 2018. *The death of truth: notes on falsehood in the age of Trump*. New York: Tim Duggan.
- Kirkham, Richard. 1992. *Theories of Truth: A Critical Introduction*. Cambridge: MIT Press.
- Linares Columbié, Radamés. 2005. *Ciencia de la información: su historia y epistemología*. Santa Fé: Rojas Eberhard.
- Magallón Rosa, Raúl. 2019. *Unfaking news: cómo combatir la desinformación*. Madrid: Pirámide.
- McIntyre, Lee. 2018. *Post-truth*. Cambridge: MIT Press.
- Molina Fernández, y Félix Moya Anegón. 2002. "Perspectivas epistemológicas 'humanas' en la documentación." *Revista Española de Documentación Científica* 25 (3): 241–53.
- Murolo, Leonardo. 2019. "La posverdad es mentira. Un aporte conceptual sobre fake news y periodismo." In *La posverdad: una cartografía de los medios, las redes y la política*, editado por Ramon Aparici y David García-Marín, 65–80. Barcelona: Gedisa.
- Naeem, Salman, and Rubina Bhatti. 2020. "The Covid-19 'infodemic': a new front for information professionals." *Health Information and Libraries Journal* 37 (3): 233–39.
- Noble, Safiya Umoja. 2018. *Algorithms of oppression: how search engines reinforce racism*. New York: New York University Press.
- O'Connor, Cailin, and James Weatherall. 2019. *The misinformation age: how false beliefs spread*. New Haven: Yale University Press.
- Ørom, Anders. 2000. "Information science, historical changes and social aspects: a Nordic Outlook." *Journal of Documentation* 56 (1): 12–26.



- Pan American Health Organization. 2020. *Understanding the infodemic and misinformation in the fight against COVID-19*. Washington: Regional Office for the Americas of the World Health Organization.
- Peters, Michael, Sharon Rider, Mikko Hyvönen, and Tina Besley. edited by. 2018. *Post-truth, fake news, viral modernity and higher education*. Singapore: Springer.
- Raatikainen, Panu. 2021. "Truth and theories of truth." In *The Cambridge Handbook of the Philosophy of Language*, edited by Piotr Stalmaszczyk, 217–32. Cambridge University Press.
- Revez, Jorge, and Luis Corujo. 2021. "Librarians against fake news: A systematic literature review of library practices (Jan. 2018–Sept. 2020)." *The Journal of Academic Librarianship* 47 (2).
- Roark, David. 1982. *Introduction to philosophy*. Dalmor Publishing.
- Salaün, Jean-Michel, et Claude Arsenault. 2009. *Introduction aux sciences de l'information*. Montreal: Presses de l'Université de Montréal.
- Santaella, Lucia. 2019. *A pós-verdade é verdadeira ou falsa?* Barueri: Estação das Letras e Cores.
- Saracevic, Tefko. 1999. "Information science." *Journal of the American Society for Information Science* 50 (12): 1051-63.
- Schlesselman-Tarango, Gina. 2017. "Critical Librarianship in a Post-Truth Age." *Library Faculty Publications* 36.
- Serrano Oceja, y José Francisco. 2019. *La sociedad del desconocimiento: comunicación posmoderna y transformación cultural*. Madrid: Encuentro.
- Silva, Armando Malheiro, y Fernanda Ribeiro. 2002. *Das 'ciências' documentais à ciência da informação: ensaio epistemológico para um novo modelo curricular*. Porto: Afrontamento.
- Spasser, Mark A. 2002. "Realist activity theory for digital library evaluation: Conceptual framework and case study." *Computer Supported Cooperative Work* 11 (1/2): 81–110.
- Wardle, Claire, and Hossein Derakshan. 2017. "Information disorder: toward an interdisciplinary framework for research and policy making." *Council of Europe*.
- Weir, Lorna. 2008. "The concept of truth regime." *Canadian Journal of Sociology / Cahiers canadiens de sociologie*, 33 (2).
- Wilber, Ken. 2017. *Trump and a post-truth world*. Boulder: Shambhala.
- World Health Organization. 2020. *1st WHO infodemiology conference How infodemics affect the world and how they can be managed*. Geneva: World Health Organization.

- Zarocostas, John. 2020. "How to fight an infodemic." *The Lancet* 395 (10225): 676.
- Zielinski, Christoph. 2021. "Infodemics and infodemiology: a short history, a long future." *Rev Panam Salud Publica* 45: 1–8.
- Zurkowski, Paul G. 1974. *Information services environment relationships and priorities*. Washington D.C.: National Commission on Libraries.



# Le raccomandazioni di Alba per la descrizione dei corrispondenti e la redazione dell'oggetto nel protocollo informatico

Dimitri Brunetti\*

**Abstract:** The purpose of this contribution is to present the Aurora and Alba projects that led to the realisation of some recommendations for the standardised drafting of the Correspondent and Object fields in the IT protocol. The intensive work which was carried out in two phases, the first one in 2008-2009 and the second one as an update for 2022-2023, also contributed to the national reflection on description models in order to avoid inhomogeneity and inconsistent descriptions and to favour the correct formation of archives and interoperability between document systems.

*Keywords:* Aurora, Alba, Object, Correspondent, Records management.

## 1. Introduzione

Giovedì 6 luglio all'Università degli Studi di Bari "Aldo Moro", in occasione della IX edizione di Procedamus, saranno presentati in anteprima i risultati del gruppo di lavoro Alba che ha aggiornato le raccomandazioni di Aurora, per la redazione dei campi corrispondente e oggetto nel protocollo informatico, unitamente alle raccomandazioni di carattere generale e per la scelta di un software in regola con i principi archivistici. Contestualmente il volume che lo contiene sarà distribuito a stampa e messo a disposizione in *open access*<sup>1</sup>.

Si è trattato del momento conclusivo di un progetto durato quasi un anno che ha coinvolto molti addetti ai lavori e che ha prodotto uno strumento indispensabile per una corretta gestione dei documenti. Il percorso di transizione digitale della Pubblica Amministrazione ha preso avvio, com'è noto, con le disposizioni del 1998/2000 sui sistemi di gestione informatica dei documenti che hanno riproposto la centralità del registro di protocollo come strumento

---

\* Dipartimento di Studi Umanistici e del Patrimonio Culturale, Università di Udine, Udine, Italia. dimitri.brunetti@uniud.it.

<sup>1</sup> Il pdf del libro è scaricabile dal sito web di Procedamus (Procedamus, n.d.a) e dal sito web della casa editrice Edizioni Università di Trieste - EUT (Edizioni Università di Trieste, n.d.).

imprescindibile per attribuire valore giuridico-probatorio ai documenti e per monitorarne la collocazione.

L'articolo 53 del DPR 28 dicembre 2000, n. 445 prevede la registrazione di protocollo per ogni documento ricevuto o spedito dalle pubbliche amministrazioni mediante la memorizzazione di alcune informazioni essenziali fra cui il corrispondente (il mittente per i documenti ricevuti, il destinatario per quelli spediti) e l'oggetto del documento. Sembra facile, ma nei fatti si tratta di un'operazione piuttosto complessa perché generare anagrafiche normalizzate e oggetti sintetici ed esaustivi non è affatto una cosa semplice. Eppure è necessario registrare i documenti nel protocollo nel modo più efficace possibile, evitando disomogeneità, descrizioni incoerenti, estrema sintesi e descrizioni bizzarre sia per formare un archivio dove i documenti e i fascicoli siano ben organizzati e facilmente reperibili, sia per garantire una reale interoperabilità fra i sistemi. Per risolvere questo problema nel 2008 si è formato il gruppo di lavoro Aurora, che ha realizzato la prima versione di una serie di raccomandazioni dal tenore agile e operativo. Negli anni quelle indicazioni si sono dovute confrontare con i nuovi strumenti dell'amministrazione digitale e con una normativa in costante evoluzione che ne hanno reso necessario l'aggiornamento. Così "sorge" il progetto Alba (*Aggiornamento dei Lavori e delle Buone pratiche di Aurora*) che propone a tutti coloro che si occupano di gestione documentale raccomandazioni semplici ed efficaci per la produzione del principale mezzo di corredo dell'archivio corrente e per favorire il dialogo tra i sistemi protocollo della Pubblica Amministrazione.

## 2. Il progetto Aurora

Il progetto Aurora nasce a Padova e a Bologna da un'idea di Gianni Penzo Doria. Il gruppo di lavoro viene istituito con decreto del direttore amministrativo dell'Università degli Studi di Padova dell'11 gennaio 2008, n. 95 e poco tempo dopo si svolge a Bologna la prima riunione operativa. Tuttavia l'origine di Aurora ha radici più lontane ed è un elemento del percorso di riflessione teorica e disciplinare sulle politiche di conservazione e gestione degli archivi delle università italiane avviata dal Centro per la storia dell'Università di Padova in occasione del convegno del 1994 "La storia delle università italiane. Archivi, fonti, indirizzi di ricerca". Una ricerca che si richiama agli obiettivi della Section for University and Research Institutions dell'ICA (SUV), riunitasi per la prima volta a Montreal nel 1992, e che è stata sostenuta da numerosi progetti e si è consolidata in occasione della "1ª Conferenza organizzativa degli archivi delle università italiane" del 1998. Nel 1996 viene istituito il Centro interuniversitario per la storia delle università italiane, poi prendono avvio il progetto Titulus 97 per la creazione di un sistema archivistico nazionale, il progetto Thesis 99 per la gestione e la salvaguardia delle tesi di laurea e il pro-

getto Studium 2000 per la tutela e la valorizzazione degli archivi storici delle università che, nel 2002, censisce 58 atenei rilevandone la storia istituzionale e gli archivi. Intanto, nel 1999 l'Università degli Studi di Padova emana le prime *Regole per la redazione delle anagrafiche*. Successivamente, nel 2005 viene avviato il progetto denominato "I calzini del principe Carlo", per la revisione del titolario delle università, e nel 2008 Aurora, cui faranno seguito Cartesio, UNIDOC e altri fino all'istituzione del progetto permanente Procedamus (Penzo Doria 2021). Il progetto Aurora trae spunto anche da alcuni approfondimenti pubblicati in quegli anni (Baldacci e Gelli 2000; Penzo Doria 2005) e dal vivace dibattito sviluppatosi prima e dopo il gennaio 2004, data di applicazione degli obblighi contenuti nel capo IV del DPR 445/2000.

Il progetto Aurora, realizzato da un folto gruppo di lavoro<sup>2</sup> e in collaborazione con la Direzione Generale per gli Archivi e l'ANAI (Associazione Nazionale Archivistica Italiana), si poneva l'obiettivo di definire alcune regole redazionali e descrittive per la normalizzazione delle registrazioni del corrispondente e dell'oggetto nel protocollo, in un periodo di riforma della Pubblica Amministrazione sul fronte della digitalizzazione. Un lessico e un metodo condivisi per garantire una maggiore efficienza ed efficacia alla gestione documentale e, soprattutto, informazioni certe sulla presenza di un documento e sulla sua collocazione in archivio. Aurora (*Amministrazioni Unite per la Redazione degli Oggetti e la Registrazione delle Anagrafiche nel protocollo informatico*) affrontava un problema molto serio, nel contesto documentario, perché fino a quel mo-

---

<sup>2</sup> Del Gruppo di lavoro Aurora facevano parte: Angela Angeli (Università degli Studi di Urbino "Carlo Bo"), Alessandro Boretti (collaboratore "telematico"), Valeria Bosco (Università degli Studi di Padova), Dimitri Brunetti (Università degli Studi di Torino – Coordinatore del sottogruppo Corrispondente), Gennaro Capasso (Università Ca' Foscari di Venezia), Franco Cardin (Azienda Ulss 16 Padova), Fiorindo Carducci (Università degli Studi de L'Aquila), Antonella Casula (Comune di Oristano), Paola Ciandrini (Politecnico di Milano), Teresa Lucia Colombo (Università degli Studi di Pavia), Gianluigi Contini (Regione autonoma della Sardegna), Donatella Corchia (Comune di Parma), Concetta Damiani (Anai (Associazione nazionale archivistica italiana), Antonella Finatti (Comune di Rovigo), Laura Flora (Istituto nazionale di astrofisica (Inaf) (Osservatorio astronomico di Trieste), Paolo Franzese (Ministero per i beni e le attività culturali), Laura Gilardi (Comune di Lecco), Anna Guastalla (Azienda provinciale per i servizi sanitari di Trento), Salvatore Guastella (Azienda Usl 7 Ragusa), Raffaella Gustapane (Soprintendenza archivistica per il Veneto), Donatella Mazzeo (Università degli Studi di Padova), Katia Milanese (Università del Piemonte Orientale "A. Avogadro"), Luca Milani (Agenzia veneta per i pagamenti in agricoltura), Leonardo Mineo (Università degli Studi di Trento), Nicoletta Molfetta (Provincia di Novara), Gianni Penzo Doria (Università degli Studi di Padova), Barbara Pistorozzi – Alma Mater Studiorum (Università di Bologna), Rossella Pozzi (Federazione nazionale dei collegi IPASVI), Fiorella Pregarz (Comune di Pordenone), Micaela Procaccia (Ministero per i beni e le attività culturali – Coordinatrice del sottogruppo Oggetto), Antonella Sannino (Università degli Studi di Salerno), Federica Scaravelli (Comune di Mantova), Elisabetta Scarpa (Regione del Veneto), Simonetta Simoni (Fondazione IRCCS Policlinico San Matteo di Pavia), Sebastiano Solidoro (Università del Salento), Cristina Tognon (Provincia di Rovigo), Nerella Zorzetto (Università IUAV di Venezia).

mento il modo di scrivere le anagrafiche dei corrispondenti o gli oggetti nel registro di protocollo rimaneva perlopiù un fatto personale, privo di regole e di metodo (Penzo Doria 2010).

Il gruppo di lavoro si è riunito diverse volte, alcune in sessione plenaria e altre nei sottogruppi Corrispondente e Oggetto, per concludere i lavori nell'estate del 2009 in un resort alla periferia di Ragusa, che tutti noi ricordiamo sia per le accese discussioni che per i momenti conviviali. Tanto lavoro ha portato ad un risultato importante rivolto principalmente a chi registra un documento nel sistema di gestione informatica, ma anche ai dirigenti, funzionari e ai responsabili di procedimenti amministrativi, con l'idea di offrire spunti di miglioramento delle attività.

*Le raccomandazioni di Aurora* si articolavano in quattro gruppi omogenei: le raccomandazioni di carattere generale, quelle per il corrispondente, quelle per l'oggetto e quelle per il software. Il nucleo principale del lavoro era completato da un capitolo introduttivo sulla normalizzazione delle descrizioni nel protocollo informatico e affiancato da una fitta serie di esemplificazioni, da una raccolta di abbreviazioni, sigle e acronimi, da un atlante diplomatico e da una sorta di "bestiario", chiamato *Protochrorror*, che segnalava le mostruosità e la corretta dicitura che avrebbe dovuto essere inserita a protocollo. Per l'occasione Micaela Procaccia ha scritto la fiaba *I regali della principessa Aurora*, incentrata sull'attenta gestione dei molti regali che Aurora, la principessa de *La bella addormentata nel bosco*, ha ricevuto in occasione delle nozze con il principe Filippo (Università degli Studi di Padova – Archivio generale di Ateneo 2007; Procedamus, n.d.c).

*Le raccomandazioni di Aurora* sono pubblicate nella collana «Instrumenta archivi Studii Patavini» con la presentazione del magnifico rettore dell'Università degli Studi di Padova e le introduzioni di Maria Grazia Pastura per la DGA (Direzione generale archivi), di Isabella Orefice per l'ANAI e di Mariella Guercio (Gruppo di lavoro interistituzionale Aurora 2009).

La presentazione ufficiale del volume e delle raccomandazioni è avvenuta in occasione della "6ª Conferenza organizzativa degli archivi delle università italiane", che si è svolta all'Università degli Studi di Padova il 28 ottobre 2009 con interventi di Alberto Mirandola, Stefano Vitali, Donatella Mazzetto, Mariella Guercio, Dimitri Brunetti, Micaela Procaccia, Giorgetta Bonfiglio-Dossio (secondo l'ordine di intervento della prima sessione del programma della giornata).

### 3. Disseminazione e adozioni

Nella primavera del 2009 alcuni componenti del gruppo di lavoro Aurora hanno presentato le raccomandazioni in una prima campagna di disseminazione a Catania, Lecce, Mantova, Napoli, Torino e Trieste. Si è trattato di

un'operazione significativa perché, in questo modo, si è ottenuto sia di condividere con un vasto pubblico quanto era stato realizzato, sia di avviare una fase di confronto con gli operatori e i responsabili dei servizi di protocollo che negli anni successivi non si è più interrotta.

Nel corso del tempo *Le raccomandazioni di Aurora* sono state richiamate nei testi di archivistica (ad es. Pigliapoco 2020, 107) e nei Manuali di gestione documentale di un gran numero di amministrazioni pubbliche, e comunque hanno rappresentato un riferimento rilevante per i protocollisti, benché non si abbiano dati quantitativi sulla loro effettiva adozione<sup>3</sup>. Alcuni enti hanno anche provato ad adattare alle proprie esigenze, semplificandole o facendone estratti, così come sul finire del 2016 ha fatto l'Università degli Studi di Messina con le Istruzioni Incopia (Istruzioni per la Normalizzazioni del Corrispondente e dell'Oggetto nel Protocollo Informatico di Ateneo).

Però, il segnale più forte in favore della diffusione delle *raccomandazioni di Aurora* è venuto dalla Direzione Generale per gli Archivi, che con sua Circolare del 29 gennaio 2010, n. 4, ha invitato i direttori degli istituti archivistici a favorirne la conoscenza affinché potessero poi essere diffuse anche presso gli enti sorvegliati e vigilati e adottate in base alle modalità opportunamente stabilite dai soggetti interessati.

#### 4. Il progetto Alba

Negli anni successivi alla loro pubblicazione *Le raccomandazioni di Aurora* hanno rappresentato l'unico riferimento per gli operatori di protocollo e per chi, nelle pubbliche amministrazioni, riconosce il valore della normalizzazione delle descrizioni e di una corretta registrazione del corrispondente e dell'oggetto per la formazione di metadati efficaci. Nei fatti le indicazioni del 2009 sono risultate uno strumento utile anche per la redazione dell'oggetto dei fascicoli.

Le ragioni della longevità di questo lavoro sono da ricercarsi nel fatto che il percorso di transizione digitale della Pubblica Amministrazione nel corso del tempo si è certamente consolidato e sviluppato, ma che gli elementi relativi all'identificazione dei documenti, così come enumerati dal Testo unico delle disposizioni legislative e regolamentari in materia di documentazione amministrativa del 2000, sono rimasti inalterati nel contesto della gestione documentale. Il protocollo rimane ancora oggi uno strumento insostituibile per utilità e valore e la sua corretta compilazione una reale necessità.

Però, *Le raccomandazioni di Aurora* necessitavano di una verifica in relazione al mutato contesto normativo e ai nuovi applicativi di protocollo e di registrazione. Era indispensabile confrontarsi con un'organizzazione del lavoro

---

<sup>3</sup> Un primo elenco dei provvedimenti e delle circolari di approvazione e di adozione è disponibile in (Università degli studi di Padova – Archivio Generale di Ateneo, n.d.).

che oggi tratta un numero assai minore di documenti analogici rispetto a quasi quindici anni fa e ad una pubblica amministrazione che sottoscrive i suoi atti con firma digitale, che nelle comunicazioni utilizza abitualmente la posta elettronica certificata e sta guardando a sistemi alternativi di recapito, che ha sviluppato indici nazionali dei domicili digitali per la PA e le imprese, che si avvale di banche dati interne ed esterne di persone fisiche e giuridiche, che ha perfezionato i sistemi di segnatura, metadattazione e interoperabilità. Inoltre, alcuni esempi andavano aggiornati e talune raccomandazioni si erano rivelate non di immediata comprensione. Insomma, *Le raccomandazioni di Aurora* si confermavano indispensabili per il corretto trattamento dei documenti, ma bisognose di un robusto aggiornamento.

Nell'ottobre del 2022 Gianni Penzo Doria ha inviato a tutti gli iscritti alla lista «Procedamus informa», che raggiunge oltre duemila cinquecento persone distribuite sull'intero territorio nazionale, l'avviso della prossima costituzione di un gruppo di lavoro interistituzionale cui ci si sarebbe potuto candidare. La chiamata ha portato alla formazione di un gruppo di persone motivate e preparate<sup>4</sup>. Il progetto *Procedamus* è ormai molto noto, ma vale la pena ricordare che si configura con la più importante iniziativa nazionale di partecipazione e collaborazione fra enti e professionisti sui temi riferiti alla gestione documentale e dei procedimenti amministrativi. *Procedamus* collabora con DGA, AIDUSA (Associazione Italiana Docenti Universitari Scienze Archivistiche), ANAI, ForumPA, ANDIG (Associazione Nazionale Docenti Informatica Giuridica e Diritto dell'Informatica), ANORC (Associazione Nazionale Ope-

---

<sup>4</sup> Del Gruppo di lavoro Alba facevano parte: Angela Angeli (Università degli Studi di Urbino "Carlo Bo"), Laura Berti Ceroni (Comune di Ravenna), Alessandro Boretti (Comune di Sonà), Maela Bortoluzzi (Università Iuav di Venezia), Dimitri Brunetti (Università degli Studi di Udine), Chiara Cabbia (Università Iuav di Venezia), Gennaro Capasso (Università Ca' Foscari di Venezia), Elisabetta Caputo (Guarnerio soc. coop.), Chiara Carlucci (Istituto nazionale per l'analisi delle politiche pubbliche – Inapp), Roberto Carriero (Comune di Lendinara), Marta Castellan (InfoCert spa), Teresa Cincavalli (Provincia di Lodi), Giulia Colombo (Eustema spa), Gianluigi Contini (Regione Sardegna), Valentina Corona (Comune di Reggio Emilia), Claudia Costa (Università degli Studi della Valle d'Aosta), Luca Dalvit (Soprintendenza archivistica e bibliografica del Veneto e del Trentino Alto Adige), Monica del Rio (Archivio di Stato di Venezia), Roberto Donato (Università degli Studi di Firenze), Giovanni Lamborghini (Comune di Ferrara), Delfina Majer (Università Ca' Foscari di Venezia), Luca Milani (Comune di Vicenza), Gianni Penzo Doria (Università degli Studi dell'Insubria), Cesare Petricca (Corte dei conti), Maria Piatto (Azienda Socio Sanitaria Territoriale Rhodense), Simone Pinna (Asl 2 di Savona), Sara Pizzi (Università degli Studi di Pavia), Laura Pochecci (Dipartimento per le politiche giovanili e il servizio civile universale), Micaela Procaccia (ANAI), Beatrice Rossi (Provincia di Pesaro e Urbino), Elisabetta Rossini (Cineca), Federica Scaravelli (Comune di Mantova), Nadia Selizato (Regione Veneto – Veneto agricoltura), Sebastiano Solidoro (Università degli Studi del Salento), Maddalena Taglioli (Scuola Normale Pisa), Cristina Tognon (Comune di Rovigo), Cristina Toso (Università degli Studi di Padova), Damiano Venturini (Azienda territoriale per l'edilizia residenziale della Provincia di Padova).



ratori e Responsabili della Custodia di contenuti digitali) e UNI e organizza incontri di formazione e gruppi di lavoro su problemi comuni basati sulla collegialità e il confronto serrato per raggiungere risultati concreti destinati a divenire standard di fatto.

Nel novembre dello scorso anno si è svolto online il primo incontro del progetto Alba, coordinato da Chiara Cabbia e da Gianni Penzo Doria, in cui ci si è confrontati sui problemi da risolvere e su come organizzare il lavoro e poi sono stati anche formati i sottogruppi: quello sulla descrizione dell'oggetto, coordinato da Angela Angeli, e quello riferito all'identificazione dell'anagrafica del corrispondente, coordinato da Dimitri Brunetti. Nell'incontro è stato anche deciso il nome ufficiale del progetto Alba: *Aggiornamento dei Lavori e delle Buone pratiche di Aurora*. Nei mesi successivi i sottogruppi si sono incontrati diverse volte fino alla stesura di una prima bozza condivisa. Nei primi giorni del marzo scorso è stata consolidata la versione da rendere disponibile sul sito di Procedamus che è rimasta in consultazione pubblica dal 13 al 31 marzo (Procedamus, n.d.b). Nel mese di aprile ci si è confrontati con le sollecitazioni e le proposte di modifica pervenute e a fine maggio il documento è stato consolidato prima dell'ultima lettura preliminare all'invio in stampa.

## 5. Le Raccomandazioni di Alba

Le raccomandazioni propongono principi generali e regole di comportamento, coerenti con le possibili funzioni di un sistema di gestione dei documenti e con le esigenze della comunicazione pubblica, a prescindere dalle soluzioni tecnologiche adottate. Anche per questo le raccomandazioni possono essere utili a tutti, indipendentemente dalla tipologia di ente, dalla dimensione e dai modelli organizzativi. Vengono offerte agli operatori di protocollo e a chi opera nella gestione informatica dei documenti della pubblica amministrazione, sia come riferimento per l'attività ordinaria, sia come ausilio per affrontare situazioni complesse.

Il volume *Le raccomandazioni di Alba*, edito da EUT (Edizioni Università di Trieste), sostituisce quello pubblicato nel 2009 e si apre con la prefazione di Federico Valacchi, che conduce il lettore ad un'ampia riflessione sulla funzione delle regole e delle raccomandazione per lo sviluppo armonioso di un archivio e l'uniformazione delle descrizioni, sulla figura del protocollista e sul ruolo del protocollo per la registrazione dei documenti, sul valore del modello organizzativo e sullo sviluppo di sistemi informativi evoluti basati sull'interoperabilità con viste documentali dinamiche. Valacchi chiude il suo intervento con queste parole: «l'elemento davvero innovativo di queste regole sta nel ricercare tutta la semplicità possibile per affrontare casistiche complicate e controverse, tentando di normalizzare l'antropologia degli approcci ancora prima degli effettivi comportamenti descrittivi». Seguono i saggi introduttivi di Chiara

Cabbia e di Gianni Penzo Doria. Cabbia, a nome di tutti i colleghi che hanno contribuito alla realizzazione del lavoro, illustra il percorso di digitalizzazione della Pubblica Amministrazione, riprende gli elementi di cambiamento tecnologico e gestionale più recenti che hanno determinato la necessità di porre mano alle raccomandazioni di Aurora e di aggiornarle, presenta il lavoro del gruppo di lavoro Alba e offre indicazioni sulle principali modifiche apportate ad Aurora. Penzo Doria sottolinea l'importanza del confronto costruttivo e della collaborazione istituzionale che hanno portato all'aggiornamento delle raccomandazioni di Aurora e contribuiscono a far sentire chi lavora in archivio e al protocollo come parte di un'ampia comunità professionale.

Il primo capitolo affronta il tema della normalizzazione delle descrizioni nel protocollo informatico, a cui seguono i capitoli contenenti le raccomandazioni di carattere generale, per la descrizione del corrispondente, per la redazione dell'oggetto e per la scelta di un software archivisticamente funzionante. Ogni raccomandazione è numerata ed è formata dall'indicazione di un principio e dal suo commento ed è quasi sempre illustrata con esempi di carattere descrittivo, ma non prescrittivo. Completano il libro tre appendici, la prima che ripropone la fiaba di Aurora scritta nel 2009 da Micaela Procaccia, la seconda con un elenco di abbreviazioni, sigle e acronimi e la terza con il cosiddetto "Protochrorror", ovvero un campionario di registrazioni stravaganti o che contengono errori causati da imperizia, distrazione, fretta e inerzia. Completa il volume l'indice delle parole notevoli (Gruppo di lavoro interistituzionale Alba 2023).

### 5.1. Raccomandazioni per la descrizione del corrispondente – RC

La descrizione del corrispondente, cioè il mittente (o i mittenti) per i documenti ricevuti oppure il destinatario (o i destinatari) per i documenti spediti, è un elemento obbligatorio della registrazione di protocollo e costituisce un'attività imprescindibile per la gestione archivistica in senso complessivo. Tuttavia, può risultare equivoca o insufficiente se non si forniscono informazioni adeguate e conformi a criteri condivisi e riconosciuti. In questo contesto la normalizzazione delle informazioni assume un ruolo indispensabile per evitare l'inserimento di anagrafiche in eccedenza, doppie o sbagliate. *Le raccomandazioni di Alba*, che aggiornano quelle precedenti, si propongono di rappresentare una guida utile per la compilazione, all'interno del sistema di protocollo delle amministrazioni pubbliche, di intestazioni anagrafiche dei corrispondenti formalmente omogenee, tali da consentire l'individuazione univoca del soggetto.

Dapprima sono indicati i criteri di intestazione di carattere generale comuni a tutti i corrispondenti, richiamando l'attenzione alla scelta di elementi descrittivi significativi che identifichino il soggetto in modo univoco, alla ne-



cessità di normalizzare le descrizioni e ad un uso consapevole delle abbreviazioni, della punteggiatura e delle lettere maiuscole. Di seguito le RC forniscono indicazioni relative alle persone fisiche, con puntuali riferimenti ai nomi stranieri, ai titoli di deferenza e alle altre qualificazioni dei nomi, e alle strutture pubbliche o private, ponendo l'accento sulle sigle, sui contesti gerarchici e funzionali, alle denominazioni multiple e all'indicazione delle sedi. Poi sono affrontate le questioni dei destinatari plurimi, degli atti anonimi, degli indirizzi postali, delle aggregazioni temporanee, delle sigle sindacali e dei documenti in materia di contenzioso. Infine, ci si sofferma su alcune particolarità riferite all'identificazione dei corrispondenti che utilizzano la posta elettronica.

RC1 – Le intestazioni anagrafiche devono essere formate da elementi che identifichino il corrispondente in modo univoco

RC2 – Le descrizioni dei corrispondenti devono risultare il più possibile normalizzate

RC3 – Le denominazioni dei corrispondenti si descrivono per esteso, senza preposizioni, abbreviazioni o segni di punteggiatura e riducendo per quanto possibile l'uso delle lettere maiuscole

RC4 – I nomi stranieri si riportano nella lingua originale e, di norma, non si traducono

RC5 – I nomi di persona fisica si indicano scrivendo prima il cognome e poi il nome, senza segni di separazione

RC6 – I nomi di persona fisica si indicano senza titoli di deferenza o qualificazione del nome

RC7 – I nomi di struttura si descrivono con l'aggiunta del relativo acronimo, delle virgolette e dei numeri eventualmente compresi

RC8 – L'anagrafica degli uffici e degli organi per l'ente o l'istituto di appartenenza è costituita dalla denominazione del contesto gerarchico di cui fanno parte, seguita da quella dell'articolazione funzionale

RC9 – Alle amministrazioni centrali dello Stato, alle regioni e alle città metropolitane possono essere assegnate anagrafiche multiple. Ai comuni e alle province di norma è assegnata una sola anagrafica

RC10 – Se sono presenti più denominazioni relative alla medesima struttura, si descrive quella prevalente

RC11 – Se sono presenti più intestazioni o sigilli o ragioni sociali relativi a strutture diverse, si descrive quello prevalente

RC12 – Se un ente o associazione ha più sedi o più sezioni, si descrive il nome della struttura principale completo e per esteso con l'aggiunta della denominazione della sede o della sezione

RC13 – Se un documento è spedito con un sistema elettronico di recapito certificato si descrivono tutti i destinatari, indipendentemente dal numero. Se un documento è spedito con altro sistema a quattro o più destinatari si

- descrive l'anagrafica completa del primo con l'aggiunta delle parole "Altri - Elenco allegato alla minuta" oppure "Altri - Vedi minuta"
- RC14 – Nei documenti a firma plurima si indica, di norma, solo il primo firmatario leggibile e non più di tre firmatari
- RC15 – I documenti pervenuti con firma illeggibile o privi di firma, a seconda dei casi si descrivono con le parole "Firma illeggibile" o "Firma mancante"
- RC16 – L'anagrafica di un documento che non comprende alcun elemento utile alla identificazione del mittente si descrive con la parola "Anonimo"
- RC17 – L'indirizzo postale si scrive in forma italiana, inserendo i dati inerenti alla città, alla provincia e al codice di avviamento postale
- RC18 – Le aggregazioni fra strutture si descrivono riportandone per esteso le denominazioni
- RC19 – Nei raggruppamenti temporanei di imprese si registra la società capogruppo
- RC20 – Le sigle sindacali riunite si considerano come un nuovo corrispondente rispetto agli stessi soggetti non aggregati
- RC21 – Le strutture dipendenti da altre si descrivono solo se significative
- RC22 – Nei documenti in materia di contenzioso bisogna distinguere il vettore dall'autore del documento
- RC23 – Si devono evidenziare i riferimenti ai corrispondenti che utilizzano la posta elettronica, anche desumendo i dati dalle informazioni di contesto o da quelli precedentemente dichiarati dal corrispondente
- RC24 – Si deve individuare il vero mittente e verificare l'uso improprio della carta intestata

## 5.2. Raccomandazioni per la redazione dell'oggetto – RO

La redazione dell'oggetto prevede un'operazione intellettuale di analisi e sintesi di un documento diretta ad individuare le azioni, le istanze e il contenuto giuridico rappresentato. L'oggetto costituisce il principale ambito di ricerca, identifica il documento e ne definisce la relazione con il fascicolo. Naturalmente, per interpretare correttamente un documento e per elaborare un oggetto significativo è indispensabile definire criteri rigorosi di normalizzazione, evitando frasi generiche o la copiatura acritica dell'oggetto del documento stesso, testimoniandone invece il carattere giuridico, amministrativo o narrativo.

*Le raccomandazioni di Alba* sull'oggetto sono rivolte in particolare ai soggetti che partecipano alla formazione di un archivio: chi redige l'oggetto del documento, chi effettua la registrazione di protocollo, chi predispone la modulistica, chi produce documenti, chi si occupa del sistema di gestione documentale.

Inizialmente si richiama l'attenzione sulla necessità di cercare un equilibrio fra analisi e sintesi, di scegliere termini e forme verbali comuni, di ricercare un approccio redazionale uniforme in seno alla stessa AOO individuando le parole chiave del documento anche ai fini della successiva ricerca. Poi, dopo aver precisato che la redazione dell'oggetto non è mai la semplice trascrizione di quello presente nel documento, si segnala di fare attenzione alla forma grafica dei nomi composti, dei termini tecnici e dei neologismi, di riportare precisamente date, luoghi, appuntamenti, di non introdurre riferimenti normativi muti o richiami non espliciti. Infine, la RC12 ricorda che i documenti usualmente contengono dati personali e che quindi la redazione dell'oggetto deve essere svolta valutando la pertinenza dei dati e l'indispensabilità degli stessi per identificare e ricercare correttamente il documento, senza eccedere e quindi evitando di inserire nell'oggetto dati personali sensibili e giudiziari, così come previsto dalla normativa in materia.

RO1 – L'oggetto si redige ricercando un equilibrio tra analisi e sintesi

RO2 – L'oggetto si redige scegliendo forme verbali e lessici condivisi

RO3 – L'oggetto si descrive secondo i principi di univocità e uniformità

RO4 – L'oggetto si redige individuando le parole chiave che esprimono l'azione (o le azioni), facendo riferimento a una lista di termini utilizzabili

RO5 – L'oggetto di un documento non si replica nella risposta, né nei documenti successivi

RO6 – La redazione dell'oggetto non si limita alla mera trascrizione dell'oggetto presente nel documento

RO7 – L'oggetto del documento non replica l'oggetto del fascicolo

RO8 – L'oggetto si redige facendo attenzione alla forma grafica dei nomi composti

RO9 – L'oggetto si redige riportando in modo esatto date, scadenze, luoghi e orari

RO10 – I termini tecnici si utilizzano solo se necessari, evitando neologismi

RO11 – L'oggetto di un documento in lingua straniera si riporta per esteso nella lingua originale seguito dalla descrizione in lingua italiana

RO12 – L'oggetto di un documento si redige tenendo conto delle disposizioni vigenti in materia di protezione dei dati personali

RO13 – L'oggetto non deve contenere riferimenti normativi muti o richiami non espliciti

RO14 – I moduli istituzionali si descrivono per esteso

RO15 – Gli allegati non fanno parte del campo oggetto, ma si contano e si descrive la loro tipologia in campi separati

RO16 – Il punto fermo a fine frase risulta pleonastico

### 5.3. Raccomandazioni di carattere generale – RG

Accanto alle raccomandazioni per la redazione delle descrizioni di anagrafiche e di oggetti nell'ambito delle registrazioni di protocollo, il progetto Alba fornisce anche principi generali, criteri e procedure operative riferiti alla gerarchia delle informazioni da inserire (dal generale al particolare), sull'utilizzo limitato delle lettere maiuscole, sull'uso consapevole delle abbreviazioni, dei numeri, delle date, della punteggiatura, degli accenti e dei caratteri speciali. Inoltre, richiama il protocollista a non ripetere la denominazione del campo che si sta compilando e a non riportare nel campo oggetto il mezzo di invio del documento. Infine, suggerisce l'utilizzo di un linguaggio "neutro", privo di riferimenti di genere.

RG1 – La descrizione dei corrispondenti e la redazione degli oggetti si effettuano dal generale al particolare

RG2 – L'uso delle lettere maiuscole deve essere limitato allo stretto necessario

RG3 – Le abbreviazioni si utilizzano solo se approvate, le sigle non si puntano e l'acronimo si riporta anche sciolto

RG4 – Nell'indicazione degli aggettivi numerali ordinali si utilizzano le cifre arabe e per gli intervalli di numeri si usa il trattino

RG5 – Nelle date si privilegia la leggibilità

RG6 – La punteggiatura e gli accenti si utilizzano in modo normalizzato

RG7 – Attenzione all'utilizzo dei caratteri speciali

RG8 – Non si ripete la denominazione del campo che si sta compilando

### 5.4. Raccomandazioni per la scelta di un software archivistamente funzionale – RS

*Le raccomandazioni di Alba* sono indipendenti dagli applicativi utilizzati e si basano principalmente sulla preparazione e sulla buona volontà dei protocollisti, però un sistema informatico inadeguato può influire negativamente sulla possibilità di chi le applica nell'azione di protocollazione del documento. Per questo motivo il gruppo di lavoro ha confermato la scelta di Aurora di proporre anche dei consigli per l'adozione di un software definendone alcuni requisiti pur se riferiti solo alle registrazioni del corrispondente e dell'oggetto, e quindi vincendo la tentazione di affrontare in termini più ampi la questione.

Ci si è soffermati, in particolare, nel suggerire l'adozione di software interoperabili, allineati al quadro normativo e regolamentare, che offrano soluzioni avanzate per la compilazione e la gestione delle informazioni, che garantiscano l'integrità dei dati e l'immodificabilità dei campi.

RS1 – Preferire software che garantiscano l'interoperabilità tra sistemi

RS2 – Preferire software che assicurino identificazione certa, immodificabilità e integrità del documento informatico

- RS3 – Preferire software che garantiscano il rispetto della normativa in materia di metadati e formati obbligatori
- RS4 – Preferire software che offrano thesaurus controllati per la compilazione del campo oggetto
- RS5 – Preferire software che permettano la creazione di liste di autorità (*authority file*) estese
- RS6 – Preferire software che comprendano campi distinti per ciascuna informazione complementare del corrispondente
- RS7 – Preferire software che consentano di visualizzare tutti i documenti collegati a una determinata anagrafica
- RS8 – Preferire software che nella funzione di copia di un record di protocollo prevedano controlli automatici
- RS9 – Preferire software che prevedano per le anagrafiche dei dipendenti l'indirizzo di residenza, di domicilio e della sede di lavoro
- RS10 – Preferire software che trattino il corrispondente e l'oggetto come campi immutabili e non rinnovabili
- RS11 – Preferire software che prevedano l'inserimento di un numero minimo e massimo di caratteri nel campo oggetto
- RS12 – Preferire software che permettano di associare parole chiave al documento.
- RS13 – Preferire software che comprendano campi per annotazioni immutabili e campi per note modificabili
- RS14 – Preferire software che comprendano l'indicazione del numero degli allegati e la loro descrizione come campi obbligatori, distinti e non modificabili, gestibili con diciture controllate
- RS15 – Preferire software che controllino la presenza di documenti già registrati
- RS16 – Preferire software che gestiscano i caratteri accentati e i caratteri speciali

## 6. Conclusioni

*Alba di Aurora* è stato un progetto utile e concreto e le sue 64 raccomandazioni si candidano a diventare uno standard di fatto per la registrazione dei documenti nel protocollo informatico, ma anche come strumento di lavoro quotidiano per la gestione documentale corrente, così come un riferimento per tutte le attività di formazione di anagrafiche controllate e di descrizione di aggregazioni in ogni ambito di intervento. *Le raccomandazioni di Alba* sono aggiornate ed efficaci e dovranno essere fatte conoscere con una disseminazione analoga a quella del 2009, ma anche in altri modi, che certamente la nostra comunità professionale saprà identificare.

Le raccomandazioni, però, non sono perfette e andranno verificate e magari riviste e poi aggiornate. Qualcuno le considera complicate, e forse ha ragione. In effetti, a tratti la lettura non sembra agevole e non sempre la singola raccomandazione è chiara se non posta in relazione ad altre (ma i riferimenti sono sempre ben indicati in nota). La complessità, però, oltre che ad essere la cifra distintiva dell'archivistica contemporanea, è anche la naturale conseguenza dell'incredibile varietà dell'azione amministrativa e della straordinaria e multiforme produzione documentaria che richiedono sempre attenzioni specifiche. Le raccomandazioni affrontano una grandissima quantità di casi e propongono soluzioni e quasi trecento esempi positivi e negativi, sovente approfondiscono le questioni sollecitando la riflessione.

Mi si permetta, in conclusione, avendo partecipato alla redazione sia delle raccomandazioni di Aurora che delle raccomandazioni di Alba, di testimoniare il piacere di aver condiviso con tanti colleghi e amici questo percorso e il valore di questa esperienza fatta di confronto, ricerca e crescita professionale.

## Riferimenti bibliografici

- Baldacci, Sandra e Simona Gelli. a cura di. 2000. *Il protocollista questo sconosciuto!* San Miniato: Archilab.
- Edizioni Università di Trieste – EUT. n.d. Consultato il 30 maggio 2023. <https://eut.units.it>.
- Gruppo di lavoro interistituzionale Alba. a cura di. 2023. *Le raccomandazioni di Alba. Aggiornamento dei lavori e buone pratiche di Aurora*. Trieste: EUT.
- Gruppo di lavoro interistituzionale Aurora. a cura di. 2009. *Le raccomandazioni di Aurora*. 2009. Padova: Cleup.
- Penzo Doria, Gianni. 2005. "Profili archivistici del protocollo informatico." *Archivi & Computer* XV (1): 91-109.
- Penzo Doria, Gianni. 2010. "La solitudine del protocollista e il progetto Aurora." *La gazzetta degli enti locali*, 16 settembre 2010. <https://www.procedamus.it/8-eventi/383-aurora-2009.html>.
- Penzo Doria, Gianni. 2021. *Il progetto Archivi del 1996 dell'Università degli Studi di Padova*. Padova: Cleup.
- Pigliapoco, Stefano. 2020. *Guida alla gestione informatica dei documenti*. Torre del Lago: Civita.
- Procedamus. n.d.a. Consultato il 30 maggio 2023. <https://www.procedamus.it/>.
- Procedamus. n.d.b. "ALBA - Aggiornamento dei Lavori e delle Buone pratiche di AURORA in consultazione pubblica fino al 31 marzo 2023." Consultato il 30 maggio 2023. [www.procedamus.it/8-eventi/419-albainconsultazione31marzo2023.html](http://www.procedamus.it/8-eventi/419-albainconsultazione31marzo2023.html).

Procedamus. n.d.c. “AURORA - Amministrazioni unite per la redazione degli oggetti e la registrazione delle anagrafiche nel protocollo informatico.” Consultato il 30 maggio 2023. <https://www.procedamus.it/8-eventi/383-aurora-2009.html>.

Università degli Studi di Padova - Archivio Generale di Ateneo. n.d. “AURORA Provvedimenti e circolari di approvazione e di adozione.” Consultato il 30 maggio 2023. <https://www.unipd.it/archivio/progetti/aurora/approvazioni/>.

Università degli Studi di Padova - Archivio Generale di Ateneo. 2007. “AURORA Amministrazioni Unite per la Redazione degli Oggetti e la Registrazione delle Anagrafiche nel protocollo informatico”. <https://www.unipd.it/archivio/progetti/aurora/>.





# Intelligenza artificiale e informazione giuridica

Una sperimentazione con GPT-3  
per il testo coordinato delle norme

Manola Cherubini, Francesco Romano\*, Andrea Bolioli\*\*

**Abstract:** The purpose of this paper is to provide a brief overview of the GPT-3 linguistic model, and to explore the potential of using artificial intelligence to reconstruct the current legislative text. The paper presents a detailed description of the use case that was designed, the system outputs, and the evaluations conducted.

*Keywords:* Computational Linguistics, Artificial Intelligence, Legal informatics, Dissemination of legal information, Current legislative texts.

## 1. Introduzione e stato dell'arte

L'accesso all'informazione giuridica costituisce un mezzo irrinunciabile per garantire la possibilità di tutelare i propri diritti e conoscere i propri doveri, specie in società sempre più aperte e interconnesse. Il medium telematico è la modalità principale di accesso a tali risorse informative per cittadini e addetti ai lavori. Per questo motivo il trattamento automatico di questo tipo di informazione – ad esempio la legislazione statale, ma anche la regolamentazione degli enti locali (Aydin e Hakan 2020) – è molto utile. Tramite le tecnologie linguistico computazionali si possono infatti produrre applicazioni per archiviare automaticamente i dati estratti in database, prendere decisioni o elaborare analisi, usare i dati estratti come input in altri processi (Garcia-Constantino et al. 2017, 79), ma anche per produrre riassunti (*summarization*), semplificare l'accesso o comunque aumentare la conoscibilità di informazioni così rilevanti. Per questo motivo ancora oggi la ricerca informatico-giuridica si occupa di mettere a punto metodi e strumenti per estrarre particolari strutture presenti

---

\* Istituto di Informatica Giuridica e Sistemi Giudiziari (IGSG) del CNR. manola.cherubini@igsg.cnr.it, francesco.romano@igsg.cnr.it.

\*\* Ricercatore indipendente. andrea.bolioli@gmail.com.

Il lavoro è frutto della progettazione e della riflessione comune degli autori ma il par. 1 è da attribuire a Romano, il 2 a Bolioli, il 3 a Cherubini e il 4 a tutti i coautori.

nei testi di legge. Modellizzare, per identificare, le singole disposizioni normative è infatti alla base delle applicazioni del web semantico in ambito giuridico (Biagioli 2009; Sartor et al. 2011) ed identificare riferimenti normativi e modifiche testuali esplicite per il riconoscimento delle relazioni tra i documenti legislativi o fra parti di essi è uno dei campi di indagine di chi da anni progetta e mette a punto sistemi di *parsing* per riconoscere ed estrarre questo tipo di nesso (Maat e van Engers 2003; Lesmo, Mazzei, e Radicioni 2008). Ma tale obiettivo rimane ancora oggi una delle priorità di chi ancora si occupa di trattamento automatico dell'informazione legislativa (Bach et al. 2019). Infatti la ricostruzione del testo normativo vigente e anche di un testo multivigente, che mostri cioè le diverse versioni dell'atto normativo nel corso del suo ciclo di vita (Fioravanti e Rinaldi 2010), costituisce un valido aiuto per addetti ai lavori e cittadini per la comprensione di quel naturale ipertesto costituito dall'ordinamento giuridico. Ad oggi la creazione, all'interno di database del testo normativo vigente e multivigente richiede l'applicazione di strumenti automatici, ma anche un rilevante lavoro umano. In particolare, relativamente all'esperienza italiana, la marcatura dei testi normativi in formato XML (*eXtensible Mark-up Language*) e la successiva annotazione della necessaria meta-informazione tramite apposite DTD (*Document Type Definition*) in modo semi-automatico, ha permesso di creare basi dati normative strutturate per la consultazione del testo vigente e multivigente delle norme (Francesconi 2012). Tali metodologie ormai da anni, grazie a progetti istituzionali quali *Normattiva*, consentono di visualizzare le norme statali e regionali oltre alla Costituzione e ai principali codici (Corasaniti 2022, 158-59). La letteratura più risalente in materia (Bolioli et al. 2002), ma anche quella più recente fa riferimento a sistemi basati sulla corrispondenza dei modelli di questo tipo di nesso con espressioni regolari (Garofalakis, Plessas, and Plessas 2016), ma le potenzialità degli attuali strumenti di *Artificial Intelligence* (AI) (tra i quali i *Large Language Models o LLM*), da ricondurre ad una capacità di calcolo aumentata in modo esponenziale, ad "algoritmi sempre più evoluti" e alla disponibilità di «dati che non ha precedenti nella storia» (Vincenzi 2019, 397), pongono interessanti sfide a chi voglia sperimentare queste tecnologie anche nel dominio giuridico. Tale accelerazione induce a iniziare nuove sperimentazioni anche nel campo della comunicazione delle informazioni giuridiche (Cherubini e Romano 2022, 47), ma anche altre ipotesi di utilizzo di queste tecnologie in campo giuridico sono oggetto di sperimentazione scientifica (Macey-Dare 2023).

Nel presente contributo sarà analizzata la possibilità di usare i modelli linguistici di intelligenza artificiale per l'identificazione di varie strutture di modifica testuale esplicita (MTE) e la conseguente ricostruzione del testo coordinato delle norme.

Dopo aver brevemente illustrato le funzionalità dei modelli utilizzabili per questo tipo di sperimentazione, approfondendo, in particolare, le caratteri-

stiche dello strumento usato per i test e cioè GPT-3 (*Generative Pre-trained Transformer 3*), saranno presentate le sperimentazioni per la ricostruzione del testo legislativo vigente, applicando a vari testi normativi alcune disposizioni di modifica testuale esplicita.

Nella fase conclusiva sarà riportata l'analisi dei test effettuati, valutando vantaggi e criticità nell'uso di tali strumenti.

## 2. Il modello linguistico GPT-3

Le tecnologie di Natural Language Processing (NLP) hanno compiuto negli ultimi anni passi in avanti notevoli e quasi imprevisi, al punto che potremmo dire che ha preso avvio una nuova era a partire dai modelli linguistici pre-addestrati (*Large Language Models*) basati su reti neurali profonde. Le innovazioni riguardano il Natural Language Understanding (NLU), ovvero la capacità di “comprendere” un testo, individuare gli argomenti, classificare il documento, riconoscere le entità citate, estrarre informazioni utili, identificare il *sentiment* e le emozioni espresse ecc., ma soprattutto la Natural Language Generation (NLG), ovvero la capacità di generare nuovo testo, rispondere a domande, scrivere riassunti, dialogare (*chatting*) ecc. La generazione automatica non riguarda solo i testi ma anche le immagini, i video e altri tipi di contenuti, e viene definita Intelligenza Artificiale generativa (*generative AI*).

Una svolta importante è avvenuta con il rilascio del modello BERT (*Bi-directional Encoder Representations from Transformers*) (Devlin et al. 2018) nel 2018, e delle successive estensioni del modello e varie applicazioni in diversi domini e task NLP.

Dal 2021 sono entrati in scena i modelli GPT (*Generative Pre-trained Transformers*) dell'azienda statunitense OpenAI; in particolare, il sistema ChatGPT reso disponibile al pubblico a partire da novembre 2022.

La sperimentazione presentata in questo articolo è stata effettuata tra novembre 2022 e febbraio 2023 utilizzando GPT-3, un sistema basato su un modello linguistico autoregressivo con 175 miliardi di parametri, descritto inizialmente in un articolo scientifico nel maggio 2020 (Brown et al. 2020), rilasciato al pubblico alla fine del 2021 e successivamente modificato nel corso dei mesi.

GPT-3 è il modello di terza generazione della serie GPT-n. L'architettura si basa su una rete di *Transformer*, con le dimensioni del contesto di 2048 *token*. Il task del metodo di pre-addestramento utilizzato è prevedere qual è il prossimo token in una sequenza (“pre-allenamento generativo”). Per quanto riguarda il set di dati testuali su cui è stato addestrato, complessivamente è composto da circa 499 miliardi di *token*. Il 60% proviene da una versione filtrata dell'intero web dal 2011 al 2021, detta “Common Crawl”; le altre porzioni sono invece dataset di qualità maggiore e includono una versione espansa del WebText da-

taset (Radford et al. 2019, 9), due corpora di libri digitali (Books1 e Books2) e la Wikipedia in inglese.

Il set di training utilizzato è parzialmente multilingua: circa il 93% delle parole del corpus sono in inglese, mentre il restante 7% sono in altre lingue, tra le quali il francese, il tedesco, lo spagnolo. La lingua italiana è rappresentata nello 0,6% del corpus, con più di 1 miliardo di parole (GitHub, n.d).

GPT-3 ha raggiunto buoni risultati in diversi task NLP tra i quali la traduzione, il *question-answering*, i cosiddetti *cloze task* (inserire la parola mancante in una porzione di testo), così come in compiti che richiedono del ragionamento o l'adattamento a un dominio, senza la necessità di effettuare dell'addestramento o del fine-tuning per lo specifico task o dominio.

A differenza dei sistemi precedenti, GPT-3 mostra risultati sorprendenti nei casi di *few-shot*, *one-shot* e *zero-shot learning*, ovvero:

(a) *few-shot learning*, in cui sono consentite tutte le dimostrazioni (esempi) che si adattano alla finestra di contesto del modello;

(b) *one-shot learning*, in cui è consentita solo una dimostrazione;

(c) *zero-shot learning*, in cui non sono consentite dimostrazioni e al modello viene data solo un'istruzione in linguaggio naturale (Brown et al. 2020).

Nel periodo della sperimentazione, GPT-3 era utilizzabile come servizio a pagamento tramite un'interfaccia web o tramite API, per gli utenti registrati al servizio. Nel sito web di OpenAI si trova la documentazione che spiega come utilizzare il servizio e alcuni esempi di task specifici (rispondere a domande, riassumere testi, scrivere review di ristoranti, generare codice software ecc.) (OpenAI, n.d.a.). Non era possibile accedere al codice sorgente né ai modelli linguistici sottostanti.

Le modalità di utilizzo previste da GPT-3 erano 3:

- *Complete*: è la modalità utilizzata in molti task diversi, nella quale «you input some text as a prompt, and the model will generate a text completion that attempts to match whatever context or pattern you gave it» (OpenAI, n.d.b.);

- *Insert*: per inserire un testo all'interno di un altro testo («inserting text within text by providing a suffix prompt in addition to the prefix prompt»);

- *Edit*: l'utente fornisce al sistema un testo e un'istruzione in linguaggio naturale su come modificarlo («the edits endpoint can be used to edit text, rather than just completing it»).

Per la sperimentazione di ricostruzione del testo normativo abbiamo utilizzato la modalità *Edit*, che si è rivelata la più adatta a questo scopo.

A scopo esemplificativo riportiamo, in Figura 1, una schermata dell'interfaccia grafica (il cosiddetto *Playground*), in cui abbiamo fornito al sistema in modalità *Edit* (beta) una porzione di documento giuridico nella sezione *Input* e l'istruzione di modifica nella sezione *Instructions*. Nella sezione di destra si vede il risultato in *Output*.

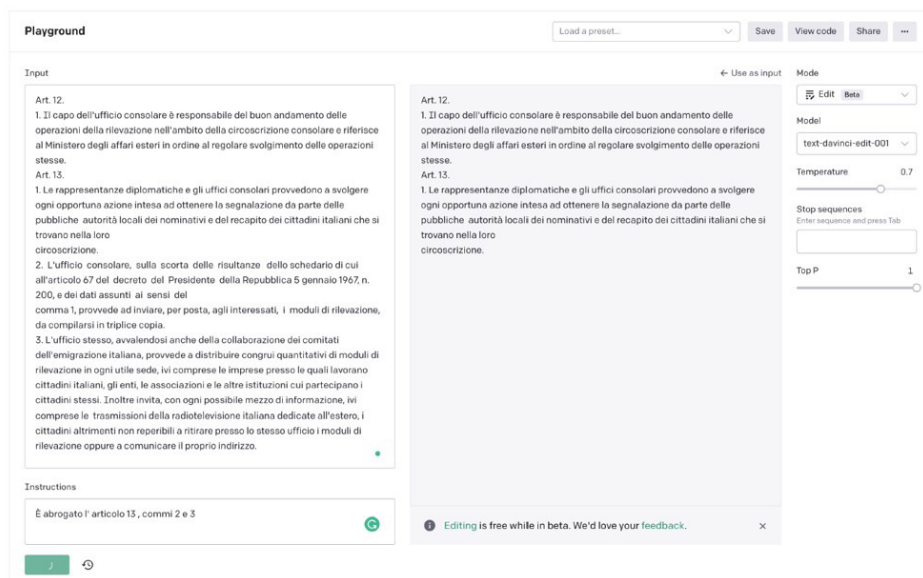


Figura 1: Schermata del Playground di GPT-3 in modalità Edit.

### 3. Sperimentazione di GPT-3 per la ricostruzione del testo normativo vigente

Le tipologie di modifiche presenti nei testi normativi possono determinare la definitiva eliminazione di un testo normativo previgente e, dunque, per così dire, la sua “uscita di scena” dal quadro normativo in vigore, oppure la sua permanenza in vigore con parziali modifiche al testo (modifiche testuali) o, comunque, alla sua portata normativa (modifiche non testuali). La permanenza in vigore del testo modificato è la situazione che si verifica nella stragrande maggioranza dei casi e che può creare anche serie difficoltà ai destinatari del contenuto.

Si distinguono, poi, le modifiche esplicite, che riportano chiaramente le porzioni del testo da modificare e la modifica da effettuare, dalle implicite, che, al contrario, non indicano la modifica da effettuare (modifica tacita) o la disposizione da modificare (modifica innominata).

L'attuale sperimentazione con GPT-3 del task ricostruzione del testo legislativo vigente ha preso in considerazione esclusivamente le modifiche testuali esplicite, ovvero ha inteso verificare la capacità del modello di riconoscere le partizioni dei testi normativi e di apportare le modifiche richieste.

Per modifiche esplicite testuali (anche dette novelle) si intendono, infatti, tutte quelle disposizioni che intervengono in modo esplicito in modifica di un testo normativo previgente, ovvero quando il legislatore prescrive che un testo normativo deve essere modificato secondo opportune formule testuali, indicata in una nuova disposizione di modifica, appositamente emanata.

### 3.1. Individuazione del campione

Per sperimentare con GPT-3 la possibilità di ricostruire il testo legislativo vigente è stato selezionato un campione di modifiche che potesse essere, se non esaustivo, il più possibile rappresentativo delle diverse casistiche utilizzate dal legislatore.

Le modifiche considerate sono, infatti, riferibili a: abrogazione (la nuova disposizione elimina singole disposizioni o parti di esse, eliminando parole); integrazione (la nuova disposizione integra singole disposizioni, aggiungendo parole); sostituzione (la nuova disposizione modifica il contenuto di singole disposizioni, eliminando parole e sostituendole con nuove parole); modifiche multiple, in cui la nuova disposizione contiene nella stessa dicitura più modifiche alla disposizione precedente<sup>1</sup>.

In particolare, le disposizioni di modifica sperimentate hanno preso in considerazione: l'abrogazione di articoli, lettere, commi, numeri o parole, includendo anche casi in cui il legislatore usa il termine "soppressione"; l'integrazione di articoli, commi, lettere e parole; la sostituzione di commi e parole; modifiche multiple; i diversi tipi di numerazione utilizzata, cardinale (compresi -bis, -ter ecc.) e ordinale; alcune perifrasi per indicare la posizione nel testo (ad esempio "in fine"); parti particolari del testo come alinea, periodo e capoverso.

I testi normativi da cui sono state ricavate le porzioni di testo da modificare (testi storici) e le relative novelle per la sperimentazione sono molto eterogenei per tipologia<sup>2</sup>.

<sup>1</sup> Per consultare tutte le formule di modifica si veda l'allegato B delle *Regole e suggerimenti per la redazione dei testi normativi* (manuale per le Regioni promosso dalla Conferenza dei Presidenti delle Assemblee legislative delle Regioni e delle Province autonome con il supporto scientifico dell'Osservatorio legislativo interregionale). Terza edizione del dicembre 2007.

<sup>2</sup> Decreto-legge 27 gennaio 2022, no. 4. "Misure urgenti in materia di sostegno alle imprese e agli operatori economici, di lavoro, salute e servizi territoriali, connesse all'emergenza da COVID-19, nonché per il contenimento degli effetti degli aumenti dei prezzi nel settore elettrico.", come modificato dalla Legge 28 marzo 2022, n. 25. *Gazzetta ufficiale*, 27 gennaio 2022, no. 21; Decreto Legislativo 20 novembre 2008, no. 188. "Attuazione della direttiva 2006/66/CE concernente pile, accumulatori e relativi rifiuti e che abroga la direttiva 91/157/CEE.", come modificato dal Decreto Legislativo 15 febbraio 2016, n. 27. *Gazzetta ufficiale*, 3 dicembre 2008, no. 283 – Suppl. Ordinario no. 268; Costituzione della Repubblica Italiana, come modificata dalla Legge Costituzionale 11 febbraio 2022, no. 1; Decreto 5 marzo 2015, no. 30. "Regolamento attuativo dell'articolo 39 del decreto legislativo 24 febbraio 1998, n. 58 (TUF) concernente la determinazione dei criteri generali cui devono uniformarsi gli Organismi di investimento collettivo del risparmio (OICR) italiani.", come modificato dal Decreto 13 gennaio 2022, no. 19. *Gazzetta ufficiale*, 19 marzo 2015, no. 65; Legge 27 ottobre 1988, no. 470. "Anagrafe e censimento degli italiani all'estero.", come modificata dalla Legge 27 maggio 2002, no. 104. *Gazzetta ufficiale*, 7 novembre 1988, no. 261; Regio Decreto 19 ottobre 1930, no. 1398. "Approvazione del testo definitivo del Codice Penale.", come modificato dal Decreto Legge 4 agosto 2000, no. 220, convertito con modificazioni dalla Legge 6 ottobre 2000, no. 275 e dalla Legge 21 novembre 2000, no. 353. *Gazzetta ufficiale*, 26 ottobre 1930,



Occorre evidenziare che i testi normativi sottoposti a GPT-3, dato il vincolo del limite di parole, sono stati, in alcuni casi, adattati per rendere più agevole la sperimentazione. Le disposizioni di modifica applicate a tali testi rispecchiano però le espressioni utilizzate dal legislatore.

In dettaglio, si riportano di seguito le disposizioni di modifica sottoposte a GPT-3:

1. all'articolo 1: al comma 1, primo periodo, dopo le parole: "n. 221" sono aggiunte le seguenti: «convertito, con modificazioni, dalla legge 18 febbraio 2022, n. 11»;

2. all'articolo 3, il comma 3, è soppresso;

3. all'articolo 1: al comma 2, le parole: "fino al 31 gennaio 2022" sono sostituite dalle seguenti: "fino al 31 marzo 2022" e dopo le parole: "n. 221" sono inserite le seguenti: «convertito, con modificazioni, dalla legge 18 febbraio 2022, n. 11»;

4. all'articolo 3, il comma 2-bis, è abrogato;

5. all'articolo 3, comma 2, lettera b) dopo le parole: "una di quelle" la parola: "attività" è soppressa e le parole: "attività economiche ATECO" sono sostituite dalle seguenti: "attività economiche ATECO 2007";

6. all'articolo 3: dopo le parole: "una di quelle" la parola: "attività" è soppressa;

---

no. 251; Legge 20 febbraio 1958, no. 75. "Abolizione della regolamentazione della prostituzione e lotta contro lo sfruttamento della prostituzione altrui.", come modificata dalla Legge 3 agosto 1998, no. 269. *Gazzetta ufficiale*, 4 marzo 1958, no. 55; Decreto del Presidente della Repubblica 8 giugno 2001, no. 327. "Testo unico delle disposizioni legislative e regolamentari in materia di espropriazione per pubblica utilità. (Testo A).", come modificato dalla Legge 1 agosto 2002, no.166. *Gazzetta ufficiale*, 16 agosto 2001, no. 189 – Suppl. Ordinario no. 211; Regio Decreto 28 ottobre 1940, no. 1443. "Codice di procedura civile.", come modificato dalla Legge 27 maggio 1997, no. 141, poi successivamente modificato dalla Legge 18 giugno 2009, no. 69. *Gazzetta ufficiale*, 28 ottobre 1940, no. 253; Regio Decreto 7 giugno 1923, no. 1252. "Che dispone il passaggio della «Gazzetta ufficiale» del Regno dalla dipendenza del Ministero dell'interno a quella del Ministero della giustizia e degli affari di culto e detta le norme per la compilazione e la pubblicazione di essa.", come modificato dalla Legge 11 dicembre 1984, no. 839. *Gazzetta ufficiale*, 16 giugno 1923, no. 141; Decreto Legislativo 30 luglio 1999, no. 300. "Riforma dell'organizzazione del Governo, a norma dell'articolo 11 della legge 15 marzo 1997, n. 59.", come modificato dal Decreto-legge 12 giugno 2001, no. 217, convertito con modificazioni dalla Legge 3 agosto 2001, no. 317. *Gazzetta ufficiale*, 30 agosto 1999, no. 203 – Suppl. Ordinario no. 163; Legge 30 dicembre 2010, no. 240. "Norme in materia di organizzazione delle università, di personale accademico e reclutamento, nonché delega al Governo per incentivare la qualità e l'efficienza del sistema universitario.", come modificata dal Decreto Legge 9 febbraio 2012, no. 5. *Gazzetta Ufficiale*, 14 gennaio 2011, no. 10 – Suppl. Ordinario no. 11; Decreto del Presidente della Repubblica 29 marzo 1973, no. 156. "Approvazione del testo unico delle disposizioni legislative in materia postale, di bancoposta e di telecomunicazioni.", come modificata dal Decreto legislativo del 1 agosto 2003, no. 259. *Gazzetta ufficiale*, 3 maggio 1973, no. 113 – Suppl. Ordinario.

7. all'articolo 1, comma 3, le parole: «e di cui al decreto legislativo 25 luglio 2005, n. 151» sono sostituite dalle seguenti: «e di cui al decreto legislativo 14 marzo 2014, n. 49»;

8. all'articolo 3: il comma 2 è soppresso;

9. al comma 3 la lettera c) è soppressa;

10. all'articolo 8, comma 1, secondo periodo, le parole: «del decreto legislativo 25 luglio 2005, n. 151» sono sostituite dalle seguenti: «del decreto legislativo 14 marzo 2014, n. 49»;

11. all'articolo 9, comma 1, il secondo periodo è sostituito dal seguente:

Qualora tali rifiuti non possano essere prontamente rimossi dall'utilizzatore finale, i suddetti apparecchi sono progettati in modo tale che i rifiuti di pile e accumulatori siano prontamente rimovibili da professionisti qualificati indipendenti dai produttori. Gli apparecchi in cui sono incorporati pile o accumulatori sono altresì corredati di istruzioni che indicano come l'utilizzatore finale o i professionisti qualificati indipendenti possano rimuoverli senza pericolo. Se del caso, le istruzioni informano altresì l'utilizzatore finale sui tipi di pila o di accumulatore incorporato nell'apparecchio;

12. all'articolo 10, comma 1, lettera b), le parole: «o del decreto 25 luglio 2005, n. 151,» sono sostituite dalle seguenti: «o del decreto 14 marzo 2014, n. 49»;

13. m) all'articolo 27 il comma 5 è sostituito dal seguente: «5. Con decreto del Ministro dell'ambiente e della tutela del territorio e del mare, di concerto con il Ministro dell'economia e delle finanze sono stabilite le tariffe per la copertura degli oneri di cui al comma 4, nonché le relative modalità di versamento»;

14. all'articolo 41 della Costituzione sono apportate le seguenti modificazioni: a) al secondo comma, dopo la parola: "danno" sono inserite le seguenti: "alla salute, all'ambiente," b) al terzo comma sono aggiunte, in fine, le seguenti parole: "e ambientali";

15. all'articolo 1, comma 1, del decreto del Ministro dell'economia e delle finanze 5 marzo 2015, n. 30, dopo la lettera z) sono aggiunte le seguenti: «aa) "portafoglio finanziario": il valore complessivo del portafoglio costituito da depositi bancari, prodotti di investimento assicurativi e strumenti finanziari disponibili anche presso altri intermediari o gestori bb) "prodotti di investimento assicurativi": i prodotti di cui all'articolo 1, comma 1, lettera w-bis.3), del TUF»;

16. all'articolo 3, dopo il comma 2, è inserito il seguente: «2-bis. In considerazione della necessità di inquadrare, anche a livello statistico, le imprese operanti nel settore dei matrimoni e degli eventi privati, l'Istituto nazionale di statistica definisce una classificazione volta all'attribuzione di un codice ATECO»;

17. all'articolo 14, comma 2, il primo periodo è sostituito dal seguente: «I produttori che per la prima volta immettono sul mercato pile e accumulatori



nel territorio italiano sono obbligati ad iscriversi in via telematica soltanto una volta al Registro nazionale presso la Camera di commercio di competenza»;

18. l'articolo 11 è abrogato;

19. è abrogato l'articolo 13, commi 2 e 3;

20. all'articolo 425 del codice penale, il numero 5) è abrogato;

21. all'articolo 4, numero 2), le parole: «di persona minore degli anni 21 o» sono soppresse;

22. all'articolo 58, comma 1, numero 9), sono aggiunte, in fine, le parole «limitatamente alle norme riguardanti l'espropriazione»;

23. al terzo comma dell'art. 83 è aggiunto, in fine, il seguente periodo: «La procura si considera apposta in calce anche se rilasciata su foglio separato che sia però congiunto materialmente all'atto cui si riferisce»;

24. è abrogato il capoverso seguente all'alea introduttivo del primo comma dell'articolo 3 del regio decreto 7 giugno 1923 n. 1252;

25. all'articolo 425, alea, del codice penale, le parole: “dai due articoli precedenti” sono sostituite dalle seguenti: “dagli articoli 423 e 424”;

26. nell'articolo 46, comma 1, del decreto legislativo 30 luglio 1999, n. 300, sono soppresse le lettere a) e b);

27. al comma 1, lettera e), secondo periodo, tra la parola: “durata” e la parola: “quadriennale” è inserita la seguente: “massima”;

28. all'articolo 6: 1) al comma 3 le parole: «nonché compiti di tutorato e di didattica integrativa» sono soppresse;

29. all'articolo 2, lettera b) e all'articolo 6, comma 3, le parole «nonché compiti di tutorato e di didattica integrativa» sono soppresse;

30. all'articolo 13, secondo comma, sono soppresse le parole da: “telegrafici e radioelettrici” fino a: “servizi telefonici”; nella rubrica sono soppresse le parole: “e delle telecomunicazioni”.

### 3.2. Modalità di interrogazione di GPT-3

Il compito richiesto a GPT-3 è stato quello di eseguire il comando impartito in linguaggio naturale e cioè quello di applicare disposizioni di modifica testuale esplicita ad un testo normativo, senza l'uso di meta informazioni per la struttura formale (suddivisione in partizioni) e funzionale (indicazioni semantiche delle disposizioni) del testo.

La sperimentazione è stata effettuata nel periodo compreso tra novembre 2022 e febbraio 2023.

L'interrogazione di GPT-3 è avvenuta in modalità *Edit*, che ha permesso di verificare le capacità del modello in modo più puntuale. Per tutte le prove la procedura di interrogazione, infatti, è stata quella di inserire nell'area di *Input* la porzione di testo normativo da modificare (novellando) e in *Instructions* la disposizione di modifica (novella).

Si è proceduto, inoltre, sempre in modalità *Zero-shot*, ovvero senza fornire alcun esempio a supporto.

I parametri di generazione utilizzati sono stati i seguenti: modello: text-davinci-001; temperature: 0.7; Top P: 1.

È stato usato il modello *Davinci-001* in quanto indicato da OpenAI come modello specializzato proprio per modificare un testo, a fronte di una precisa istruzione fornita.

La *temperature* di 0.7 è il valore standard di casualità (*randomness*) nella generazione dell'output già suggerito nell'interfaccia del *Playground*, così come il valore 1 di *Top probability* (*Top P*).

### 3.3. Analisi dei risultati

A fronte di 30 prove effettuate, ovvero 30 disposizioni di modifica sottoposte a GPT-3 come prima e unica istruzione (*Zero-shot*) per essere applicate alle relative porzioni di testo da modificare, 26 casi sono stati eseguiti perfettamente.

Di seguito, in Figura 2, un esempio di sostituzione di parole in periodo e in Figura 3 un esempio di modifica multipla comprendente una sostituzione e una integrazione.

The screenshot shows the OpenAI Playground interface. On the left, there is a 'Get started' sidebar with instructions. The main area is titled 'Playground' and contains an 'Input' field with the following text:

Art. 8.  
Obiettivi di raccolta  
1. Ai fini del presente decreto, la percentuale di raccolta delle pile e degli accumulatori viene calcolata per la prima volta in relazione alla raccolta effettuata nel corso dell'anno 2011. Fatta salva l'applicazione del decreto legislativo 25 luglio 2005, n. 151, i dati annuali relativi alla raccolta e alle vendite comprendono pile e accumulatori incorporati in apparecchi.  
2. Al fine di realizzare un sistema organico di gestione delle pile ed accumulatori portatili che riduca al minimo il loro smaltimento insieme al rifiuto urbano misto, entro la data del 26 settembre 2012 dovrà essere conseguito, anche su base regionale, un tasso di raccolta minimo di pile ed accumulatori portatili pari al 25 per cento del quantitativo immesso sul mercato; tale tasso di raccolta dovrà raggiungere, entro il 26 settembre 2016, il 45 per cento del quantitativo immesso sul mercato.  
3. Le percentuali di raccolta di pile e accumulatori portatili sono calcolati annualmente dall'Istituto superiore per la protezione e la ricerca ambientale, di seguito: «ISPRA», secondo il piano di cui all'allegato I,

The 'Output' field shows the same text, but with the words 'della raccolta' and 'della raccolta' replaced by 'della raccolta' and 'della raccolta' respectively. The right-hand control panel shows the 'Engine' set to 'text-davinci-edit-001', 'Temperature' set to 0.7, and 'Top P' set to 1. A 'Submit' button is visible at the bottom of the input field.

Figura 2: Sostituzione di parole in periodo.

The screenshot shows the OpenAI Playground interface. On the left, there is a 'Get started' sidebar with instructions. The main area is titled 'Playground' and contains an 'Input' section with a text area containing a legal article snippet. Below the input is an 'Instructions' section with a list of modifications to be applied to the text. On the right, there is a 'Mode' sidebar with settings for 'Engine' (text-davinci-edit-001), 'Temperature' (0.7), and 'Top P' (1). A 'Submit' button is located at the bottom of the input area.

Figura 3: Modifica multipla comprendente una sostituzione e una integrazione.

Diverse considerazioni, invece, possono essere fatte analizzando l'esecuzione delle istruzioni da parte di GPT-3 nelle restanti 4 prove effettuate.

La prova numero 5 (vedi elenco di cui al par. 3.1), ad esempio, prevedeva una modifica multipla comprendente una soppressione e una sostituzione di parole in un articolo.

La novella richiesta era la seguente: «All'articolo 3, comma 2, lettera b) dopo le parole: “una di quelle” la parola: “attività” è soppressa e le parole: “attività economiche ATECO” sono sostituite dalle seguenti: “attività economiche ATECO 2007”».

In questo caso GPT-3, come rappresentato in Figura 4, ha eseguito la sostituzione ma non ha effettuato la soppressione della parola.

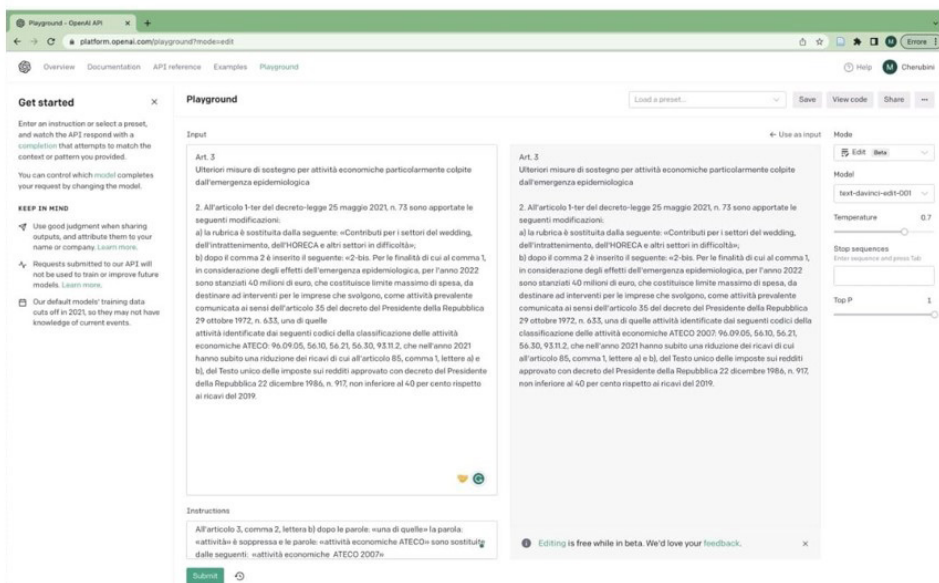


Figura 4: Sostituzione di parole in periodo.

Così nel caso della prova numero 14, la novella prevedeva una modifica multipla, contenente due integrazioni di parole: «All'articolo 41 della Costituzione sono apportate le seguenti modificazioni: a) al secondo comma, dopo la parola: “danno” sono inserite le seguenti: “alla salute, all’ambiente,”; b) al terzo comma sono aggiunte, in fine, le seguenti parole: “e ambientali”».

Il modello GPT-3 non ha effettuato l’inserimento delle parole “alla salute, all’ambiente,” nella loro giusta posizione, ovvero subito dopo la parola “danno”, ma, anche dopo ulteriori prove, le ha inserite sempre nella parte finale del comma (Fig. 5).

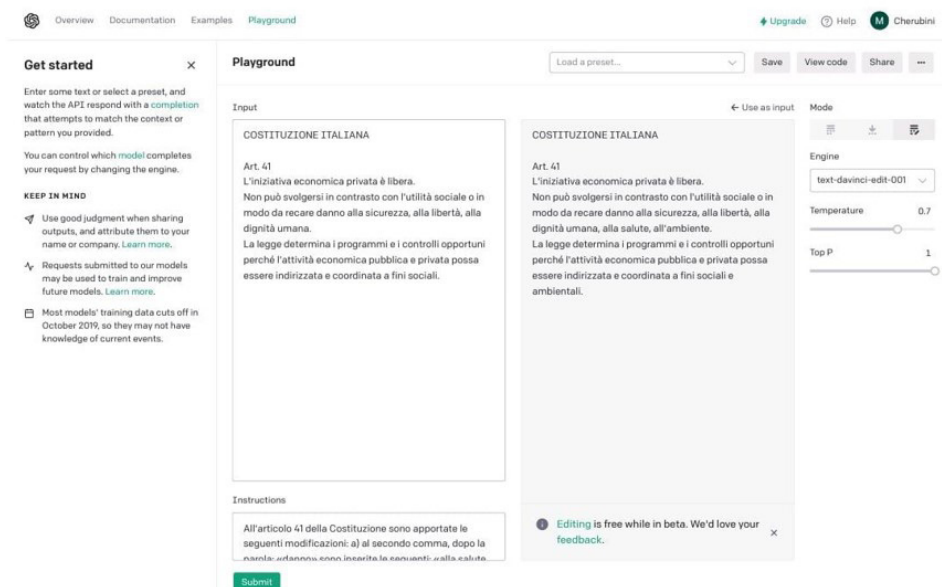


Figura 5: Modifica multipla (integrazione di parole).

In un ulteriore caso, prova numero 25, si prevedeva la soppressione di un periodo dopo un'alinea.

Il testo della norma originario, ovvero l'articolo 3 del Regio Decreto 7 giugno 1923, n. 1252, era il seguente:

Art. 3.

La Gazzetta sarà divisa in due parti: nella prima saranno pubblicate tutte le leggi e decreti che dovranno essere inseriti nella Raccolta ufficiale delle leggi e decreti del Regno d'Italia, e, su richiesta del Ministro proponente, d'accordo col Ministro guardasigilli, gli altri decreti Reali o Ministeriali di carattere speciale, nonché tutte le normali e le circolari esplicative dei provvedimenti legislativi, eccetto quelle di carattere riservato. Vi saranno pure pubblicate le disposizioni emanate dalle Presidenze delle due Camere, dalla Presidenza dal Consiglio dei ministri e dal Ministero della Real Casa; nella seconda saranno pubblicati gli annunci e gli altri avvisi prescritti dalle leggi o dai regolamenti del Regno. Le due parti del giornale saranno stampate in fogli o parti di foglio separati e distinti, la prima con l'intestazione «Gazzetta ufficiale – Parte prima», la seconda Gazzetta ufficiale – Seconda parte – Foglio delle inserzioni». Esse potranno essere poste in vendita anche separatamente.

La modifica, introdotta dalla Legge 11 dicembre 1984, n. 839 prevede che: «È abrogato il capoverso seguente all'alinea introduttivo del primo comma dell'articolo 3 del regio decreto 7 giugno 1923 n. 1252».

Il combinato disposto delle due norme produce, quindi, il seguente risultato: «Art. 3. La Gazzetta sarà divisa in due parti: nella seconda saranno pub-

blicati gli annunci e gli altri avvisi prescritti dalle leggi o dai regolamenti del Regno. Le due parti del giornale saranno stampate in fogli o parti di foglio separati e distinti, la prima con l'intestazione «Gazzetta ufficiale – Parte prima», la seconda Gazzetta ufficiale – Seconda parte – Foglio delle inserzioni». Esse potranno essere poste in vendita anche separatamente».

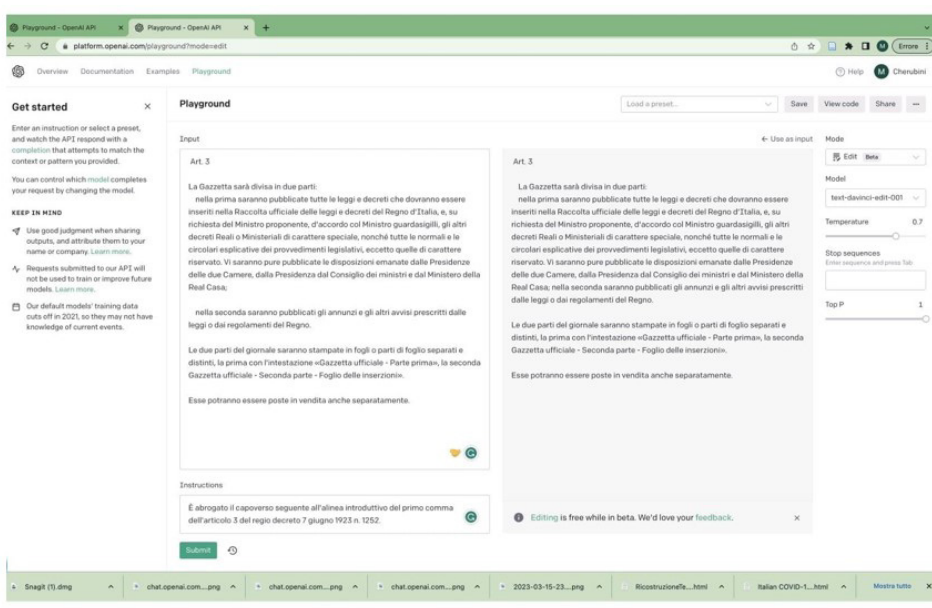


Figura 6: Abrogazione di capoverso seguente all'alinea (istruzione non eseguita correttamente).

Lo strumento ha prodotto, invece, in prima battuta, il risultato mostrato in Figura 6:

Si può supporre che l'istruzione non sia stata eseguita correttamente in quanto il concetto di alinea è molto tecnico e non univoco per gli stessi giuristi. Si riporta di seguito la definizione di alinea ripresa dal vocabolario Treccani:

Latinismo (di solito abbreviato in al.) usato promiscuamente dai giuristi sia nel senso di comma sia in quello di capoverso, sia per indicare ciascun accapo nelle enumerazioni interne di un singolo comma, con riferimento a testi di legge (con le prime due accezioni, l'uso risulta ambiguo, in quanto, per es., art. 121 al. 4° può significare comma 4° o capoverso 4°, che tra loro non si corrispondono; non è invece ambiguo comma 3° al. 4°). Nell'uso di altri scrittori, corrisponde in genere a capoverso (Treccani, n.d.).

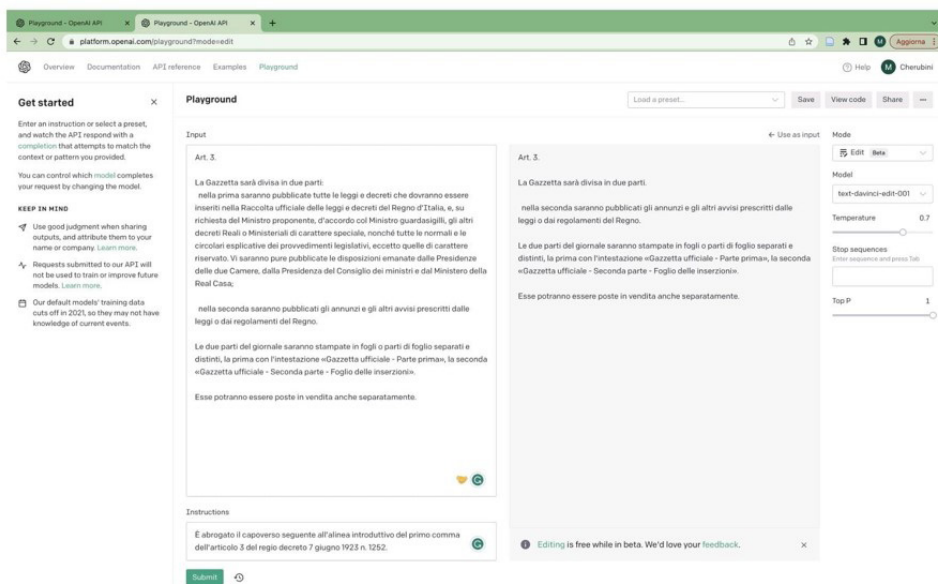


Figura 7: Abrogazione di capoverso seguente all'alea (comando eseguito correttamente).

Dopo varie prove, tuttavia, il risultato prodotto da GPT-3 è stato quello sperato (Fig. 7).

Un altro caso in cui il modello non ha correttamente effettuato le modifiche richieste è stato quello relativo alla modifica multipla che comportava una soppressione di parole in un comma e una soppressione di parole in una rubrica (prova numero 30).

L'articolo 13 del Decreto del Presidente della Repubblica 29 marzo 1973, n. 156, *Approvazione del testo unico delle disposizioni legislative in materia postale, di bancoposta e di telecomunicazioni*, era così formulato:

#### Art. 13

##### Contravvenzioni in materia postale e delle telecomunicazioni

Per le contravvenzioni punibili con la sola pena dell'ammenda è ammessa l'oblazione in sede amministrativa prima che sia iniziato il dibattimento, per somme non inferiori al minimo dell'ammenda.

La competenza a decidere sulla domanda di oblazione spetta, rispettivamente, ai direttori provinciali delle poste e delle telecomunicazioni per le contravvenzioni in materia di servizi postali, di bancoposta, telegrafici e radioelettrici e agli ispettori di zona dell'Azienda di Stato per i servizi telefonici in materia di servizi telefonici.

La novella contenuta nell'articolo 218 del Decreto legislativo del 1 agosto 2003, n. 259, *Codice delle comunicazioni elettroniche*, recita: «all'articolo 13, secondo comma, sono soppresse le parole da: “telegrafici e radioelettrici” fino



a: “servizi telefonici”; nella rubrica sono soppresse le parole: “e delle telecomunicazioni”».

A seguito della suddetta modifica, dunque, il testo dell’articolo 13 del DPR 156/1973 risulta oggi così formulato:

Art. 13

Contravvenzioni in materia postale

Per le contravvenzioni punibili con la sola pena dell’ammenda è ammessa l’oblazione in sede amministrativa prima che sia iniziato il dibattimento, per somme non inferiori al minimo dell’ammenda.

La competenza a decidere sulla domanda di oblazione spetta, rispettivamente, ai direttori provinciali delle poste e delle telecomunicazioni per le contravvenzioni in materia di servizi postali, di bancoposta.

Contrariamente a ciò, il modello GPT-3 non ha effettuato la soppressione di parole in un comma, mentre non ha avuto problemi a farlo nella rubrica dell’articolo (Fig. 8).

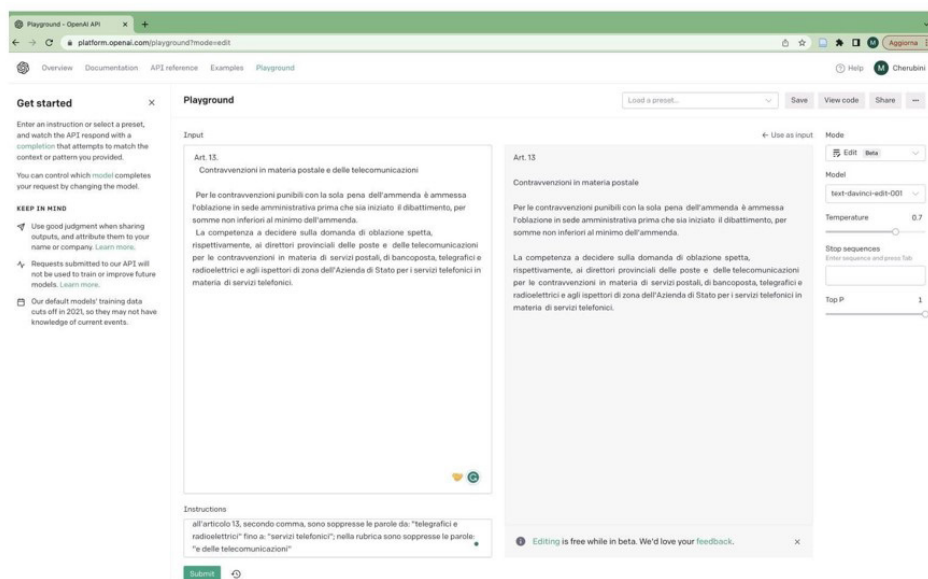


Figura 8: Modifica multipla (soppressioni di parole in comma e in rubrica).

#### 4. Conclusioni e sviluppi futuri

La sperimentazione dello strumento ha evidenziato risultati sostanzialmente positivi nell’eseguire i comandi impartiti in linguaggio naturale, relativi all’applicazione di varie tipologie di modifica testuale esplicita su un corpus di testi di legge.



Come osservato, infatti, a fronte di 30 prove effettuate, relative a 30 tipologie di disposizioni di modifica testuale esplicita utilizzate dal legislatore, in 26 casi il comando è stato eseguito in modo corretto, generando il testo coordinato della disposizione.

I risultati ottenuti dal modello risultano positivi anche considerando che la sperimentazione è stata condotta in modalità *Zero-shot* e, dunque, come primo e unico tentativo di applicazione delle novelle al testo da modificare.

È importante, poi, ricordare che il modello GPT-3 utilizzato è stato addestrato principalmente sulla lingua inglese e solo parzialmente sull'italiano e che il linguaggio del legislatore è un linguaggio settoriale altamente specialistico.

Tuttavia la sperimentazione ha sofferto di un primo limite originario, dato dal fatto che non è stato possibile sottoporre allo strumento interi testi di legge e relative norme di modifica, a causa del limitato numero di caratteri sottoponibili in una singola sessione di lavoro.

Riguardo agli errori commessi nell'applicazione delle novelle questi possono categorizzarsi nei seguenti tipi:

- difficoltà di interpretazione di termini tecnico-giuridici quali “alinea”;
- difficoltà nell'individuare la giusta posizione in cui applicare, o comunque applicare correttamente, soppressioni e integrazioni di parole, quando la posizione è identificata dalla novella usando formule del tipo “dopo la/e parola/e ...” e “le parole da ... fino a ...”.

La sperimentazione lascia intravedere possibili campi di applicazione per l'uso dello strumento nel campo della comunicazione di informazioni giuridiche a supporto della comprensibilità dei testi di legge, sia da parte dei cittadini, che del personale delle istituzioni che hanno a che fare con l'elaborazione di testi normativi a vari livelli.

Premesso che gli uffici legislativi delle istituzioni centrali dispongono di strumenti e competenze adeguate per gestire il lavoro di ricostruzione dei testi coordinati delle norme, bisogna comunque tenere presente il lavoro non banale che tale gestione comporta.

I sistemi basati su LLM potrebbero essere utili anche in materia di gestione di emendamenti durante l'iter di formazione delle leggi e potrebbero essere di supporto anche in istituzioni, come quelle degli enti locali, ove si producono e modificano regolamenti complessi e rilevanti (basti pensare ai regolamenti edilizi), senza spesso avere a disposizione risorse umane e dispositivi specifici per questo tipo di trattamento linguistico-informatico.

Pur restando il limite di parole sottoponibili al modello in ogni sessione di lavoro, ulteriori sperimentazioni dovrebbero riguardare:

- altre tipologie di modifica testuale esplicita;
- l'interrogazione in modalità *few-shot learning* nei casi non si sia mostrata utile la modalità *Zero-shot*.

Il lavoro di sperimentazione del modello linguistico GPT-3 per la *summarization* di testi giuridici è già stato descritto (Cherubini et al. 2023) e sono in corso nuove sperimentazioni per verificare l'utilità dello strumento per agevolare la comprensione di testi giuridici da parte dei cittadini.

Le sperimentazioni potranno, inoltre, essere effettuate non soltanto tramite GPT-3, ma anche utilizzando e comparando le prestazioni di altri modelli linguistici pre-addestrati, che saranno posti a disposizione proprio per verificarne i risultati nei più diversi task.

## Riferimenti bibliografici

- Aydin, Murat, and Hakan Yaman. 2020. "Domain knowledge representation languages and methods for building regulations." In *Advances in Building Information Modeling: First Eurasian BIM Forum, EBF 2019, Istanbul, Turkey, May 31, 2019, Revised Selected Papers*, edited by Salih Ofluoglu, Ozan Onder Ozener, and Umit Isikdag, 101-21. Berlin: Springer.
- Bach, Ngo Xuan, Nguyen Thi Thanh Thuy, Dang Bao Chien, Trieu Khuong Duy, To Minh Hien, and Tu Minh Phuong. 2019. "Reference extraction from Vietnamese legal documents." In *Proceedings of the 10th International Symposium on Information and Communication Technology*, 486-93. New York: Association for Computing Machinery.
- Biagioli, Carlo. 2009. *Modelli funzionali delle leggi. Verso testi legislativi autoapplicativi*. Firenze: European Press Academic Publishing.
- Bolioli, Andrea, Luca Dini, Pietro Mercatali, and Francesco Romano. 2002. "For the automated mark-up of Italian legislative texts in XML." In *Legal Knowledge and Information Systems. Jurix 2002*, edited by Trevor J.M. Bench-Capon, Aspasia Daskalopulu, and Radboud Winkels. 21-30. Amsterdam: IOS Press.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. "Language models are few-shot learners." *Advances in neural information processing systems* 33: 1877-1901.
- Cherubini, Manola, Francesco Romano, Andrea Bolioli, Nazareno De Francesco, e Irene Benedetto. 2023. "La summarization di testi giuridici: una sperimentazione con GPT-3." *Rivista italiana di informatica e diritto* 9 (1): 1-14.
- Cherubini, Manola, e Francesco Romano. 2022. "Legiferare con l'Intelligenza Artificiale." *Journal of Ethics and Legal Technologies* 4 (1): 41-64. <https://doi.org/10.14658/pupj-JELT-2022-1-4>.

- Corasaniti, Giuseppe. 2022. *Data science e diritto. Certezze digitali e benefici del dubbio*. Torino: Giappichelli.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.
- 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* 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, e Enrico Francesconi. Firenze: European Press Academic Publishing.
- Garcia-Constantino, Matias, Katie Atkinson, Danushka Bollegala, Karl Chapman, Frans Coenen, Claire Roberts, and Katy Robson. 2017. "CLIEL: context-based information extraction from commercial law documents." In *Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law*, 79-87. New York: Association for Computing Machinery.
- Garofalakis, John, Konstantinos Plessas, and Athanasios Plessas. 2016. "Automatic identification, extraction and application of textual amendments in Greek legislative texts." In *Legal Knowledge and Information Systems, JURIX 2016*, edited by Floris Bex, and Serena Villata, 187-90. Amsterdam: IOS Press.
- GitHub. n.d. "Languages by word." Consultato il 10 gennaio 2023. [https://github.com/openai/gpt-3/blob/master/dataset\\_statistics/languages\\_by\\_word\\_count.csv](https://github.com/openai/gpt-3/blob/master/dataset_statistics/languages_by_word_count.csv).
- Lesmo, Leonardo, Alessandro Mazzei, e Daniele Radicioni. 2008. "Estrazione automatica di informazioni relative alle modificazioni normative." In *I fondamenti cognitivi del diritto*, a cura di Raffaele Caterina, 65-87. Bologna: GEDIT Edizioni.
- Maat, Emile, and Tom van Engers. 2003. "Mission impossible?: Automated norm analysis of legal texts." In *Legal Knowledge and Information Systems, JURIX 2003*, edited by Daniele Bourcier, 143-44. Amsterdam: IOS Press.
- Macey-Dare, Rupert. 2023. "ChatGPT & Generative AI Systems as Quasi-Expert Legal Advice Lawyers-Case Study Considering Potential Appeal Against Conviction of Tom Hayes." <https://ssrn.com/abstract=4342686> or <http://dx.doi.org/10.2139/ssrn.4342686>.
- OpenAI. n.d.a. "Examples." Consultato il 10 gennaio 2023. <https://beta.openai.com/examples/>.

- OpenAI. n.d.b. "API Reference." Consultato il 10 gennaio 2023. <https://platform.openai.com/docs/guides/completion/editing-text>.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. "Language models are unsupervised multitask learners." *OpenAI blog* 1 (8).
- Sartor, Giovanni, Monica Palmirani, Enrico Francesconi, e Maria Angela Biasiotti, a cura di. 2011. *Legislative XML for the Semantic Web. Principles, Models, Standards for Document Management*. Berlin: Springer.
- Treccani. n.d. "Alinea." Consultato il 10 gennaio 2023. <https://www.treccani.it/vocabolario/alinea>.
- Vincenzi, Federico. 2019. "Intelligenza artificiale, machine learning, deep learning." In *Tecnologia e diritto. Informatica giuridica, II*, a cura di Giovanni Ziccardi, e Pierluigi Perri. Milano: Giuffrè.

# COVID-19 e lavoro terminologico: riflessioni a posteriori

Nataschia Ralli, Isabella Stanizzi, Marlies Alber\*

**Abstract:** In Italy, the state of emergency related to the spread of SARS-CoV-2 ended on 31 March 2022, after more than two years. It had lasted for 26 months. Throughout this period, the Italian government issued over nine hundred legal provisions and a large number of public announcements (video messages, live streams, press releases, etc.). The result was a *terminological pandemic* (Ralli 2020). In this paper, we analyze the impact of the pandemic on terminology and terminology work. In particular, we highlight the most relevant characteristics of the COVID-19 terminology (i.e., dynamism, variation, and ambiguity), focusing on the ensuing challenges from a methodological point of view. We also deal with terminology from a contrastive point of view by comparing the terminology used in the German-speaking countries with the terminology chosen for South Tyrol. Such choices were often aimed at limiting the regionalization of German as a minority language in this province.

**Keywords:** COVID-19 Terminology, Legal Terminology, Terminological work, Communication, Bistrot.

## 1. Introduzione

Tre anni fa, nel 2020, un nuovo tipo di coronavirus ha fermato il mondo. Molti paesi hanno disposto il lockdown: scuole e attività non essenziali sono state chiuse, eventi pubblici cancellati, assembramenti vietati, mezzi di trasporto ridotti. Distanziamento sociale e mascherina sono diventate le due regole base per combattere la pandemia. Ma se il mondo si è fermato, la lingua, invece, ha continuato a evolversi, adattandosi e modellandosi al nuovo contesto: parole e termini legati al coronavirus hanno cominciato a diffondersi, quasi alla stessa velocità del virus stesso, attraverso i social media, i quotidiani (cartacei e digitali), TV, radio, broadcast, ecc.

Ogni ambito della sfera pubblica e privata ne è stato travolto: la popolazione doveva essere informata sulla nuova infezione, su come prevenirla, su come

---

\* Eurac Research – Istituto di linguistica applicata, Bolzano, Italia. [nataschia.ralli@eurac.edu](mailto:nataschia.ralli@eurac.edu), [isabella.stanizzi@eurac.edu](mailto:isabella.stanizzi@eurac.edu), [marlies.alber@eurac.edu](mailto:marlies.alber@eurac.edu).

rallentarne la trasmissione, su come proteggersi e proteggere gli altri dal contrarre il virus (es. lavarsi le mani regolarmente, mantenere una distanza di sicurezza interpersonale di almeno 1 metro), ecc. (Ministero della Salute 2020a). Il risultato è stato un “sovraffollamento comunicativo” (CENSIS 2021, 3) e una “pandemia terminologica” (Ralli 2020).

Nell'emergenza sanitaria da COVID-19 è apparsa evidente la rilevanza di una comunicazione chiara, coerente e corretta (Sala e Scaglioni 2020, 9): le istituzioni dovevano fornire informazioni accurate della situazione pandemica e motivare l'adozione di comportamenti adeguati a «favorire la possibilità di contrastare meglio l'emergenza, suscitando un clima generale di fiducia» (CNB 2022, 11). Con l'intento di fornire un contributo ad un uso chiaro, corretto ed uniforme della terminologia in ambito giuridico, l'Istituto di linguistica applicata di Eurac Research si è da subito (marzo 2020) attivato per elaborare la terminologia COVID-19, dapprima in modo puntuale per risolvere questioni dubbie relative a singoli termini, poi in modo tematico, estendendo la ricerca all'intero ambito che la COVID-19 è andata a mano a mano costruendosi. I risultati di questo lavoro sono pubblicati nel Sistema informativo per la terminologia giuridica *bistro* (Eurac Research 2016).

Il presente articolo si apre con una breve disamina sul lavoro terminologico e il metodo applicato da parte dell'Istituto sopramenzionato (par. 1.1.). Nel paragrafo 2 si ripercorre la storia della pandemia attraverso i provvedimenti normativi emanati durante lo stato di emergenza. Il paragrafo 3 presenta alcuni aspetti della terminologia che si è sviluppata intorno al virus; descrive inoltre l'impatto dell'emergenza sanitaria sul lavoro terminologico, soprattutto dal punto di vista metodologico. Nel paragrafo 4, sulla scorta di esempi concreti, si descrive la terminologia dell'emergenza in chiave contrastiva, mostrando quanto la comparazione giuridica sia fondamentale per individuare analogie e discrepanze fra istituti giuridici appartenenti a ordinamenti diversi, al fine di renderli in modo corretto nell'altra lingua.

### 1.1. Aspetti metodologici: il lavoro terminologico

L'attività terminologica dell'Istituto di linguistica applicata segue i tradizionali principi terminologici (Arntz et al. 2014) e il metodo comparativo applicato alle scienze giuridiche (Sacco 1992; Sandrini 1996; Mayer 2000; Sacco 2018).

Fondamentalmente si inizia con delle considerazioni preliminari sulla metodologia di lavoro che può variare a seconda del dominio, dell'obiettivo e dei destinatari della raccolta terminologica. Si prosegue poi con la delimitazione e strutturazione del settore in esame. Per avere una visione sistematica dell'ambito trattato, si creano dei glossari, nei quali si raggruppano termini concettualmente correlati e facenti parte di uno stesso (sotto)dominio. Successivamente,

si raccoglie la documentazione necessaria fornita da fonti autorevoli e attendibili. Nell'ambito della terminologia giuridica si tratta solitamente di manuali, testi normativi, provvedimenti giudiziari e fonti Internet istituzionali. Dalla documentazione si ricavano i termini e le informazioni su di essi (es. definizioni). I termini (sinonimi e varianti ortografiche compresi) vengono individuati e selezionati sulla base dei seguenti parametri (Ralli e Stanizzi 2018, 181):

- grado di specializzazione, ossia il livello di tecnicità e la pertinenza dei termini al dominio considerato;
- frequenza, tenendo presente che anche termini con un basso numero di occorrenze vengono presi in considerazione, se accertata la specificità al dominio;
- pertinenza dei termini alla struttura, all'orientamento e allo scopo di *bistrotro*.

Si prosegue poi con l'analisi contrastiva fra l'ordinamento italiano, quale ordinamento di riferimento, e gli ordinamenti di lingua tedesca, ossia Austria, Germania e Svizzera. Solitamente si parte dalla definizione del concetto italiano, a cui possono corrispondere una o più denominazioni, e si opera un confronto con gli ordinamenti giuridici sopra menzionati al fine di individuare un potenziale equivalente. Da questo raffronto possono emergere analogie o discrepanze fra i concetti indagati. In presenza di equivalenze inesistenti e, dunque, di lacune terminologiche (Drewer e Schmitz 2017, 22) si procede con una proposta di traduzione che possa essere facilmente adottata nel sistema linguistico esistente, in modo da venire accettata ed effettivamente utilizzata anche nelle diverse forme e situazioni comunicative. Infine, si tiene anche conto del diritto comunitario e di quello internazionale. In questo caso, però, non viene fatta una comparazione, ma si resta su un piano sovranazionale. I risultati delle analisi confluiscono nelle schede terminologiche che, oltre a contenere definizioni, presentano anche contesti e, quando necessario, note a livello della denominazione o del concetto.

La terminologia elaborata viene pubblicata in *bistrotro* con l'intento di agevolare la comunicazione transfrontaliera e, al contempo, promuoverne un uso corretto e uniforme, rendendola accessibile a tutti coloro che, a vario titolo, sono chiamati a redigere o tradurre in italiano e tedesco<sup>1</sup>.

Nel prosieguo del presente contributo vedremo come la pandemia ha influito su alcuni di questi aspetti, soprattutto per ciò che riguarda la selezione dei termini e la definizione dei concetti.

---

<sup>1</sup> Gran parte della terminologia, fra cui quella relativa a COVID-19, viene elaborata nell'ambito del progetto di collaborazione fra l'Istituto di linguistica applicata di Eurac Research e l'Ufficio Questioni linguistiche della Provincia autonoma di Bolzano: ConsTerm 2.0 (Eurac Research 2023).



## 2. I temi della pandemia: il coronavirus nei provvedimenti normativi

Il 31 marzo del 2022, dopo oltre 2 anni, si è concluso in Italia lo stato di emergenza legato alla diffusione del SARS-CoV-2. Lo stato di emergenza è durato 26 mesi. Nel corso di questo periodo si sono susseguiti più di novecento provvedimenti normativi (Openpolis 2023) e innumerevoli comunicati istituzionali (sotto forma di videomessaggi, dirette Facebook, comunicati stampa, ecc.).

L'emergenza sanitaria e socioeconomica è stata affrontata ricorrendo a una vasta gamma di provvedimenti normativi: ordinanze dei Ministeri e degli altri organi ed enti coinvolti (es. Ministero della Salute, Ministero dell'Istruzione, Protezione Civile, Regioni, Comuni), circolari, decreti ministeriali (DM), decreti legge (DL) e decreti del Presidente del Consiglio dei Ministri (DPCM).

Questo considerevole flusso di provvedimenti e comunicazioni istituzionali si è tradotto in una notevole produzione terminologica caratterizzata, nel complesso, da un estremo dinamismo e da un'accentuata varietà. Analizzando i termini che ricorrono nei provvedimenti normativi principali, è possibile rintracciare le caratteristiche che hanno contrassegnato la terminologia del periodo, ripercorrendo allo stesso tempo la storia della pandemia.

Con la delibera del Consiglio dei Ministri del 31 gennaio 2020 è stato dichiarato in Italia lo «stato di emergenza in conseguenza del rischio sanitario connesso all'insorgenza di patologie derivanti da agenti virali trasmissibili». Nella delibera si parla del nuovo coronavirus (2019-nCov). Il virus non ha ancora un nome ufficiale, tantomeno la sua malattia, li acquisteranno in seguito, intorno alla metà di febbraio: SARS-CoV-2 il nome assegnato al nuovo coronavirus, COVID-19 quello della malattia (Ministero della Salute 2023). Prima di questo momento la COVID-19 ha rappresentato una minaccia confinata alla Cina, ben lontana dal cambiare le vite dei cittadini di tutto il mondo ed i principali provvedimenti del mese di gennaio (F.Q. 2021) riguardano, di conseguenza, i viaggi da e verso la Cina, le procedure e i controlli sanitari presso gli aeroporti (Ministero della Salute, ordinanze 25 e 30 gennaio 2020). Nei provvedimenti si parla, a titolo esemplificativo, di *nuovo coronavirus, 2019-nCov, sorveglianza sanitaria e profilassi internazionale* (Ralli 2021; Pacini, n.d.).

A partire dalla dichiarazione dello stato di emergenza inizia una intensa attività normativa tesa a strutturare l'organizzazione per far fronte all'emergenza pandemica. Febbraio, marzo e aprile 2020 sono i mesi in cui si registra la produzione di atti più numerosa<sup>2</sup>. Il Governo inizia a disporre le *misure di contenimento e gestione dell'emergenza* per contrastare il diffondersi dell'epidemia, misure che diventeranno via via più stringenti con l'aggravarsi della situazione sanitaria. Il *divieto di allontanamento*, le *misure di cautela*, la *quarantena*, la *quarantena con sorveglianza attiva*, la *permanenza domiciliare* (DL 23 febbraio

<sup>2</sup> A febbraio 2020 furono pubblicati 67 atti Covid, a marzo 103, ad aprile 65 (Openpolis 2023).



2006, n. 6) sono dapprima indirizzate ai comuni di Lombardia e Veneto, sede dei primi focolai (DPCM 23 febbraio 2020), e, successivamente, estese all'intero territorio nazionale (DL 25 marzo 2020, n. 19) (Pacini, n.d.; Decreto del Presidente del Consiglio dei Ministri 4 marzo 2020). La popolazione inizia a familiarizzare con le *misure di prevenzione igienico sanitarie*, il *lavoro agile*, la *didattica a distanza*, le *mascherine chirurgiche* e i *tamponi molecolari* (o PCR), mentre la limitazione alla libertà di circolazione fa sorgere i primi dubbi relativi all'interpretazione di alcuni concetti come, ad esempio, l'individuazione delle *situazioni di necessità o urgenza* che legittimano gli spostamenti. Il DPCM del 26 aprile 2020 consente gli spostamenti per incontrare i *congiunti* purché venga rispettato il *divieto di assembramento* e il *distanziamento interpersonale* di almeno un metro. Inoltre, è sempre obbligatorio usare protezioni delle vie respiratorie «nei luoghi chiusi accessibili al pubblico, inclusi i mezzi di trasporto e comunque in tutte le occasioni in cui non sia possibile garantire continuamente il mantenimento della distanza di sicurezza» (art. 3, c. 2), ricorrendo anche alle cosiddette *mascherine di comunità*. Come spiega il DPCM (art. 3, c. 3) si tratta di «mascherine monouso o mascherine lavabili, anche auto-prodotte, in materiali multistrato idonei a fornire una adeguata barriera e, al contempo, che garantiscano comfort e respirabilità, forma e aderenza adeguate che permettano di coprire dal mento al di sopra del naso». Inizia così un graduale allentamento delle *misure di contenimento* e ripartono le attività commerciali, di ristorazione e i servizi turistici seppure nel rispetto delle norme di *distanziamento sociale* (DL 16 maggio 2020, n. 33; DPCM 17 maggio 2020 e DPCM 11 giugno 2020). Nell'estate del 2020 l'attenzione si sposta sul livello europeo e sulla crisi economica generata dall'emergenza sanitaria (Ralli 2021). Si lavora alla creazione di un *Piano per la ripresa dell'Europa*, anche attraverso l'istituzione di *Next generation Eu* e del *Dispositivo per la ripresa e la resilienza* con l'obiettivo di rilanciare l'economia dell'UE (Magnani 2021). I nuovi termini generano una certa confusione terminologica e non è sempre facile districarsi tra essi per individuarne esattamente il significato (par. 3.3.).

In vista dell'avvio dell'anno scolastico 2020/2021 vengono predisposte una serie di misure finalizzate alla prevenzione del contagio nelle scuole. Nei documenti di riferimento<sup>3</sup> si parla, ad esempio, di *cruscotto informativo e didattica digitale integrata*, dei *Patti educativi di comunità* e del *referente scolastico per COVID-19*. Accanto ai termini della medicina, biologia, economia, confluiscono così anche i termini relativi all'ambito scolastico nel patrimonio terminologico della pandemia, arricchendolo e rendendolo particolarmente variegato (par. 3.1.).

---

<sup>3</sup> Nello specifico: “Piano scuola 2020-2021– Indicazioni operative per la gestione di casi e focolai di SARS-CoV-2 nelle scuole e nei servizi educativi dell'infanzia” dell'Istituto Superiore di Sanità (n. 58/2020) e “Linee guida sulla didattica digitale integrata” (DM 89/ 2020).

Nell'autunno del 2020 l'andamento dell'epidemia mostra un aggravamento: accanto alle nuove chiusure si affianca il meccanismo della suddivisione del paese in aree diverse a seconda del livello di rischio, determinato in base all'andamento dei contagi e alla resilienza delle strutture sanitarie (DPCM 3 novembre 2020) (Servizi e uffici del Senato della Repubblica e della Camera dei deputati 2021). Con il DL del 23 febbraio 2021, n. 15 le aree geografiche di cui sopra diventano le note *zone colorate* (bianca, gialla, arancione e rossa) riprendendo i termini comunemente usati nel dibattito pubblico. Seppure con successive modifiche e aggiustamenti dei parametri di individuazione, le zone colorate rimarranno fino alla fine dello stato di emergenza.

Nei primi mesi del 2021 parte la campagna vaccinale in Italia e viene adottato il "Piano strategico nazionale dei vaccini per la prevenzione delle infezioni da SARS-CoV-2" (decreto del Ministero della Salute 2 gennaio 2021). Il vocabolario COVID-19 si arricchisce di termini ulteriori come *catena di approvvigionamento*, *procedura di rolling review*, *ciclo vaccinale*, *dose addizionale*, *dose booster*, *catena del freddo*, *vaccino a mRNA*.

Nella primavera del 2021 viene introdotto l'obbligo per tutti i lavoratori del settore pubblico e privato di possedere e di esibire, su richiesta, la *certificazione verde COVID-19* conseguita in seguito a vaccinazione, guarigione o tampone (cosiddetto *Green pass base*) (DL 22 aprile 2021, n. 52) (Camera dei deputati 2022). L'aggravamento dell'epidemia nell'autunno 2021 e nei primi mesi del 2022 porta all'applicazione di regolamentazioni sempre più restrittive: viene ampliato l'obbligo di vaccinazione a diverse categorie professionali. Compaiono il *Super Green Pass* o *green pass rafforzato* e il *Green Pass Booster* come requisiti per accedere a determinati servizi o luoghi o partecipare a determinati eventi. Nel corso dei mesi successivi la situazione migliora e si arriva così al termine dello stato di emergenza e al "progressivo rientro nell'ordinario" pur nella consapevolezza che «nonostante la cessazione dello stato d'emergenza, persistano comunque esigenze di contrasto del diffondersi della pandemia da COVID-19» (DL 24 marzo 2022, n. 24).

### 3. Aspetti della terminologia COVID-19 e impatto sul lavoro terminologico

La pandemia è faccenda complessa, non riguardando solo il virus, le sue varianti, il rischio che comporta, ma anche i comportamenti dei cittadini, le ricadute sul sistema economico e sociale, nonché il rapporto tra l'origine del virus e la spoliatura delle risorse del pianeta. Ciò implica che la comunicazione istituzionale, se vuole essere trasparente, non può e non deve semplificare, ma, viceversa, illustrare la complessità in quanto tale. Oltre alla complessità, la comunicazione istituzionale deve tener conto dell'incertezza relativa ai dati scientifici e alla loro interpretazione e comunicarla adeguatamente (CNB 2022, 15-16).

A partire da gennaio 2020 la nostra lingua si è dovuta modellare, riadattare e plasmare ad un nuovo contesto che era incerto, mutevole e imprevedibile, dando origine ad un quadro terminologico alquanto complesso. Nel prosieguo si illustrano alcuni aspetti che hanno influito sul lavoro terminologico a livello metodologico.

### 3.1. Trasversalità ed eterogeneità del lessico vs selezione dei termini

La convivenza con l'infezione da SARS-CoV-2 ha generato una terminologia che toccava ogni aspetto della nostra vita, individuale e collettiva, descrivendola e modificandola in itinere (Ralli 2020). Se però l'apparato concettuale iniziale era limitato alle scienze mediche, includendovi anche aspetti dell'epidemiologia e della biologia, con l'avanzare rapido del virus tale apparato è andato rapidamente crescendo, accogliendo al suo interno termini provenienti da altri ambiti, come dimostrano i numerosi decreti emessi durante il periodo della crisi sanitaria incentrati di volta in volta su ambiti differenti: dall'economia al lavoro, dalla scuola alle politiche sociali e altri ancora (par. 2).

Accanto a termini provenienti dal mondo sanitario (es. *terapia intensiva*), epidemiologico (es. *focolaio*) e medico (es. *cloroquina*) troviamo termini che provengono da altri contesti specialistici come quello della prossemica (es. *distanza sociale*), della legislazione sociale (es. *contributi previdenziali e assistenziali*), universitaria e scolastica (es. *DAD*), del diritto pubblico (es. *DPCM*), comunitario (es. *autorizzazione all'immissione condizionata*) internazionale (es. *emergenza di sanità pubblica di rilevanza internazionale*). Non mancano termini dal diritto del lavoro (es. *ammortizzatori sociali*) e dalla sicurezza sul lavoro (es. *FF2*). Inoltre, i testi normativi emessi contenevano obblighi e divieti volti a contenere la diffusione del virus, sollevando anche questioni di diritto penale. Pertanto, una condotta imprudente o un'inosservanza dei divieti e degli obblighi previsti durante l'emergenza sanitaria potevano integrare una serie di reati di varia entità (Bordandini 2020; Domenici 2020): dall'illecito amministrativo dato dal mancato rispetto delle misure urgenti (es. violazione dell'obbligo di sospensione dell'attività produttiva) al reato di *manovre speculative su merci*, nel caso in cui un commerciante avesse messo in vendita prodotti di prima necessità come mascherine e disinfettanti a prezzi elevati, al reato di *lesioni personali dolose*, se il virus fosse stato trasmesso volontariamente causando così ad un'altra persona la malattia, o di *delitto colposo contro la salute pubblica* commesso da «chi diffonde germi patogeni provocando un'epidemia» (Domenici 2020).

Accanto a parole di nuova formazione (es. *covidiot*, *apericorona*), il linguaggio comune si è quindi arricchito anche di espressioni provenienti dal mondo specialistico. In alcuni casi si è assistito ad una risemantizzazione, ossia ad un'attribuzione di un nuovo significato a termini già esistenti. Ne è un esem-

pio il verbo *tamponare* e *tamponarsi* che nell'emergenza sanitaria ha assunto anche il significato di "sottoporre / sottoporsi a tampone" (De Vecchis 2022). In altri casi, invece, il termine è stato recepito nel suo significato originale, riportandolo però alla crisi sanitaria e legandolo così ad essa. Si pensi ad esempio ad *assemblamento* (par. 3.5.) Altri termini specialistici, invece, sono stati adattati al nuovo contesto creando espressioni più funzionali e immediate nel loro significato rispetto al vocabolario istituzionale, talvolta con «conseguenze negative sull'univocità della categorizzazione e della comunicazione» (Prandi 2010, 72) (par. 3.4). Ne è un esempio *congedo parentale* che è stato trasformato in *congedo per emergenza COVID-19*, di cui si è poi diffusa la forma contratta *congedo COVID-19* o, semplicemente, *congedo COVID* (INPS 2020).

Eterogeneità, trasversalità e duttilità del lessico hanno rappresentato da subito una sfida enorme per il lavoro terminologico: necessità comunicative si sono scontrate con i principi terminologici che prevedono, come primo passo, la delimitazione e la sistematizzazione del dominio in cui individuare, successivamente, i concetti cardini e le denominazioni che li designano. Questo principio lo si riesce però ad applicare quando il dominio in esame è già definito e circoscritto, cosa che invece non era fattibile per COVID-19: il suo apparato concettuale si è sviluppato di pari passo con l'evolversi della pandemia, le misure di contenimento, l'acquisizione di conoscenza sul virus, ecc. Un apparato che è diventato sempre più grande e sempre più eterogeneo, difficile da delimitare, strutturare e sistematizzare data la sua portata e l'impatto sulla quotidianità e la società in generale (Ralli 2021). Si doveva comunicare in maniera rapida e chiara. Di conseguenza, la terminologia doveva essere elaborata in tempo reale e in modo descrittivo, registrando i termini che venivano utilizzati dalle istituzioni e dai media, varianti e sinonimi compresi.

Si è quindi reso necessario trovare un compromesso: coniugare una terminologia di tipo sistematico con una terminologia di tipo testuale e puntuale (Drewer e Schmitz 2017, 26). Nei primi due mesi il lavoro terminologico è stato di tipo puntuale e proattivo: si è cercato di far fronte alla mole terminologica, elaborando i termini "del momento", senza dunque seguire un criterio sistematico. Ciò al fine di «prevenire la dispersione terminologica e agevolare una comunicazione chiara, corretta e affidabile» (Ralli e Lenz 2022, 27). Osservando a posteriori, tali termini, che toccavano più ambiti specialistici, fuoriuscendo spesso da quello giuridico, rappresentano la base della terminologia COVID-19 (es. *assemblamento*, *autocertificazione*, *didattica a distanza*, *distanza di sicurezza interpersonale*, *immunità*, *isolamento fiduciario*, *tampone PCR*). In alcuni casi, si è trattato di termini che sono stati al centro del dibattito pubblico (es. *Stati generali*). In altri casi, invece, i termini selezionati sono stati il risultato di un'estrazione manuale da circolari del Ministero della Salute, da ordinanze e decreti di emergenza. Diventando la COVID-19 oggetto di normativa, i termini medici, biologici, ecc. sono diventati di fatto

parte integrante del lessico giuridico e, dunque, della terminologia giuridica. Si pensi ad esempio alle tipologie di *caso* (*confermato*, *probabile*) definiti dalla (Decisione di esecuzione UE 2018/945) e ripresi dalla Circolare n. 0007922 del 09/03/2020 del Ministero della Salute (2020b). Attraverso il ricorso alla definizione, il legislatore stabilisce i confini applicativi dei singoli concetti e ne specifica i contenuti. Ne consegue che «nel momento in cui il legislatore decide di accogliere un termine all'interno della normativa, definendolo e disciplinandolo, esso assume rilevanza giuridica» (Ralli e Stanizzi 2008, 64).

### 3.2. Urgenza, immediatezza e transitorietà

Immediatezza e urgenza sono stati i tratti distintivi della terminologia COVID-19, soprattutto nei primi mesi dell'emergenza sanitaria (Ralli 2021). La pandemia ci ha colpito inaspettatamente e, in brevissimo tempo, ha influenzato gran parte della nostra vita quotidiana. Di fronte a una situazione nuova, anche il linguaggio deve inevitabilmente adattarsi. Avere a disposizione parole e termini adeguati è un presupposto fondamentale e ineludibile per poter comunicare in merito a circostanze e situazioni. Se tuttavia, da un lato, la lingua si sviluppa in modo velocissimo, dall'altra ci vuole molto più tempo per standardizzare e consolidare le norme linguistiche.

Nel caso di COVID-19 l'uso linguistico è andato a mano a mano adattandosi automaticamente alla situazione che la pandemia stessa aveva generato, circondandoci di nuove parole e costringendoci a confrontarci con nuovi concetti. Nei primi mesi della crisi sanitaria sono pertanto nati nuovi concetti e nuovi termini. Di conseguenza, nell'arco di brevissimo tempo, sono entrati nell'uso linguistico generale anche nuove denominazioni, in parte sinonimiche, in parte imprecise o fuorvianti. Ci si è quindi trovati in una situazione in cui la terminologia era ancora in divenire e in pieno processo di sviluppo e standardizzazione, ma, al contempo, già ampiamente presente nell'uso linguistico.

Questo aspetto ha avuto un ruolo rilevante nel lavoro terminologico: le schede terminologiche di *bistro* si compongono di termini utilizzati nei testi normativi nazionali, regionali e provinciali così come, più in generale, nei testi istituzionali. Tuttavia, nel caso di COVID-19, la discussione mediatica e istituzionale era spesso preliminare all'approvazione dei testi normativi stessi (leggi, decreti, ordinanze). A ciò si aggiunge un ulteriore aspetto: in una situazione di emergenza, i testi normativi vengono emanati in tempi molto brevi e, in tempi altrettanto brevi, possono intervenire modifiche e aggiornamenti. Così può accadere che anche le definizioni vengano adattate al nuovo contesto oppure che i concetti vengano denominati in altro modo. Il risultato di questa evoluzione può essere anche la compresenza di più definizioni di uno stesso concetto, cosa che, a sua volta, può essere fonte di confusione e incertezza.

Un esempio al riguardo è dato da *lockdown* (Ralli 2021). Si tratta di un anglicismo utilizzato a gennaio 2020 da parte dell'OMS (Organizzazione Mondiale della Sanità) e dei media per indicare le misure messe in atto in Cina nella città di Wuhan (Corbolante 2020) per contenere la diffusione del coronavirus. A marzo 2020 *lockdown* comincia a diffondersi nel lessico italiano per indicare i «provvedimenti di emergenza ora [20 marzo 2020] in atto nel paese, anche se finora non è usato nelle comunicazioni del governo che invece privilegia misure di contenimento» (Corbolante 2020).

Ad aprile 2020 il termine sembra aver mutato significato e indicare «confinamento o segregazione, anche se l'uso prevalente, nella comunicazione italiana di questi giorni [2 aprile 2020] sembra alludere piuttosto alla chiusura forzosa degli esercizi commerciali e delle fabbriche, e finisce quindi per equivalere a serrata o chiusura obbligatoria o obbligata» (Marazzini 2020). Nell'arco di due settimane, il termine sembra avere assunto un significato più specifico: la chiusura obbligata di alcuni esercizi è di fatto un tipo di misura di contenimento (Ralli 2020). Questa incertezza concettuale si traduce naturalmente con altrettanta incertezza sul piano della comprensione e interpretazione. A distanza di due anni il Dizionario italiano Treccani (2022, 612) definisce *lockdown* come «periodo di isolamento in casa più o meno ristretto, imposto alle persone per limitare il rischio di contrarre o diffondere un virus, con riferimento in particolare a quello noto come COVID-19».

In altri casi può succedere che nascano termini che sono utilizzati in un certo periodo, per poi scomparire. Si pensi ad esempio a *Fase 2* o *coprifuoco*.

Tutto questo ha comportato una verifica assidua, costante e a brevi intervalli di tempo della terminologia, sia a livello della definizione che della denominazione<sup>4</sup>, tanto da rendere «il lavoro terminologico un costante *work in progress*» (Ralli 2021). Si pensi al termine relativo al certificato dell'Unione Europea attestante che una persona 1) è stata vaccinata contro la COVID-19, 2) ha ricevuto un risultato negativo a un test o 3) risulta guarita dalla COVID-19 (*bistro*). Inizialmente, la denominazione in uso era *certificato verde digitale*, sostituita poi da *certificato COVID digitale dell'UE* nel corso del 2021. Si tratta di una denominazione con un riferimento più concreto e chiaro al concetto e al contesto, ma proprio per questo anche più restrittivo. In teoria, un *certificato verde digitale* potrebbe riferirsi a una qualsiasi malattia. *Certificato COVID digitale dell'UE* è, invece, univocamente legato all'emergenza sanitaria da COVID-19 e, conseguentemente, non può essere utilizzato al di fuori di questo contesto.

---

<sup>4</sup> È opportuna una precisazione: la revisione e l'aggiornamento periodici sono attività fondamentali per garantire la qualità della scheda terminologica e dei suoi contenuti (Früh e Tamás 2021, 17).



### 3.3. Terminologia comunitaria

La straordinarietà dell'emergenza sanitaria da COVID-19 ha reso necessari prese di decisione e interventi rapidi da parte delle istituzioni, non solo a livello locale e nazionale, ma anche europeo. Per fronteggiare le considerevoli conseguenze sanitarie, economiche e sociali e fornire il necessario sostegno finanziario, l'Unione Europea ha cominciato a elaborare proposte di intervento sin dai primi mesi della pandemia (MEF 2021, 1) e che hanno preso il nome di *Piano per la ripresa europea*. A maggio 2020 hanno avuto inizio i lavori per l'accordo sul futuro quadro e sui programmi settoriali che lo accompagnavano. Questi lavori si sarebbero dovuti concludere all'inizio dell'autunno affinché il nuovo bilancio a lungo termine potesse essere pienamente operativo e guidare la ripresa dell'Europa dal 1° gennaio 2021. Alcuni concetti erano definiti pertanto in modo ipotetico e in maniera generica (Tab. 1) non essendo questi ancora stati implementati, ma solo previsti. Inoltre, era ancora assente un regolamento che li disciplinasse e definisse nel concreto, ma solo una comunicazione (442/2020 della Commissione europea). Trattandosi di concetti per i quali, in alcuni casi, era necessario trovare degli escamotage linguistici per evitare formulazioni generiche e ipotetiche, si è riflettuto se fosse il caso di elaborare questa terminologia. Data la grande discussione politica e mediatica intorno al *MES* e al *Recovery Plan*, la decisione è risultata in qualche modo obbligata e del tutto in linea con gli obiettivi della nostra raccolta terminologica (par. 1.1.), ossia agevolare la comunicazione, garantendo un uso chiaro, coerente e corretto della terminologia.

Definizione dal documento "Domande e risposte sul bilancio dell'UE per la ripresa: il dispositivo per la ripresa e la resilienza", 28 maggio 2020 (Commissione Europea 2020)	Il nuovo dispositivo per la ripresa e la resilienza fornirà un sostegno finanziario su larga scala per riforme e investimenti intrapresi dagli Stati membri, allo scopo di attenuare l'impatto a livello sociale ed economico della pandemia da coronavirus e di rendere le economie dell'UE più sostenibili, resilienti e meglio preparate per le sfide poste dalle transizioni verde e digitale.
Definizione rielaborata e pubblicata in <i>bistro</i>	Strumento volto a fornire un sostegno finanziario su larga scala per riforme e investimenti intrapresi dagli Stati membri, allo scopo di attenuare l'impatto a livello sociale ed economico della pandemia da COVID-19 e di rendere le economie dell'UE più sostenibili, resilienti e meglio preparate per le sfide poste dalle transizioni verde e digitale.

Tabella 1: Definizione di *dispositivo per la ripresa e la resilienza*.

### 3.4. Denominazioni concorrenti

I media, ma anche Internet, sono soprattutto i mezzi attraverso cui la terminologia si diffonde giungendo anche a persone non esperte. Questo aspet-

to è emerso chiaramente nella comunicazione durante la pandemia da COVID-19: la comunicazione era onnipresente, in qualsiasi situazione e registro.

Solitamente, i termini tecnici vengono usati da esperti di un determinato settore, laddove un termine dovrebbe essere assegnato a un solo concetto, così come un concetto dovrebbe essere assegnato a un solo termine, secondo il principio terminologico della univocità (Wüster 1985, 1; Soglia 2002, 10). Tuttavia, un lessico di specialità non rimane confinato alla cerchia specialistica, ma si differenzia sul piano diastratico in funzione dei suoi utenti (Prandi 2010, 80) e «a seconda del punto di vista che caratterizza una determinata situazione comunicativa» (Giampreti e Sintuzzi 2010, 228). In questo senso il punto di vista può essere influenzato anche da altri settori in cui tale lessico compare (Giampreti e Sintuzzi 2010, 227). Ciò è particolarmente evidente in quegli ambiti che hanno un forte impatto sulla quotidianità e che vedono coinvolto un vasto bacino d'utenza. In questi casi si possono spesso osservare denominazioni concorrenti che vengono utilizzate in situazioni comunicative diverse a seconda dei parlanti, del punto di vista e dello scopo. Questo aspetto emerge chiaramente nelle situazioni di emergenza nelle quali i «media locali e nazionali si fanno portavoce della comunicazione istituzionale» (Sfardini 2020, 64). Affinché il messaggio possa essere recepito in modo chiaro e corretto da tutta la popolazione, accade che venga utilizzata una terminologia più semplice. I concetti più complicati possono così venire affiancati da «espressioni non tecniche, appartenenti al linguaggio comune, di più facile e immediata comprensione» (Giampreti e Santuzzi 2010, 228). Si pensi ad esempio ai termini *isolamento* e *autoisolamento* utilizzati dai media, mentre nei decreti si leggeva di *isolamento fiduciario*.

In questo contesto particolare rilevanza è data dalla “motivazione” della denominazione (Arntz et al. 2014, 58). Se nella comunicazione tra esperti, la denominazione dovrà rispecchiare il più possibile il concetto e i termini ontologicamente correlati rispettando il principio dell'univocità, nella comunicazione generale gli effetti concreti sulla popolazione stanno al centro dell'attenzione. Un esempio al riguardo è dato dal termine *test rapido* in relazione al *test antigenico rapido* o al *test sierologico*. Ad uno primo sguardo emerge subito un aspetto che contravviene ad una delle regole terminologiche: l'ambiguità. La denominazione *test rapido* è una forma breve che si è diffusa nel linguaggio comune per indicare sia il *test antigenico rapido* sia il *test sierologico*. In questo senso la forma ellittica è molto generica e, conseguentemente, meno precisa. Senza un preciso riferimento al settore di appartenenza, *test rapido* potrebbe riferirsi ad ogni tipo di test, anche al di fuori dell'ambito medico. La denominazione non fornisce, pertanto, alcuna indicazione sulla modalità, la procedura, la valutazione o lo scopo del test. In questo caso, la motivazione principale si basa sull'impatto diretto sulle persone interessate. Quest'ultime, infatti, non devono né occuparsi di fare il prelievo né di analizzare il campione



prelevato. La caratteristica principale risiede nel risultato che è disponibile in tempi brevi, costituendo la motivazione della denominazione. È evidente che una denominazione di questo tipo è imprecisa e, di conseguenza, può essere fonte di confusione: sebbene il *test antigenico rapido* e il *test sierologico* siano concetti correlati, presentano comunque caratteristiche diverse, a partire dalla loro procedura. In questi casi è, quindi, fondamentale analizzare i contesti d'uso e, se necessario, sottolineare nella raccolta terminologica che si tratta di espressioni della lingua comune.

### 3.5. Vaghezza e contestualità dei concetti vs difficoltà di definizione e di traduzione

Chi si occupa di terminologia giuridica sa che ha a che fare con un linguaggio complesso: termini vaghi perlopiù astratti, concetti ambigui difficili da definire. Anche termini, che all'apparenza possono sembrare semplici, racchiudono in sé delle complicazioni, quando devono essere interpretati alla luce di norme giuridiche specifiche. È il caso ad esempio di *abitazione*.

Il concetto di *abitazione* ha creato non pochi grattacapi alla popolazione durante la pandemia. Sono numerosi, infatti, i provvedimenti normativi del periodo in cui ricorre il termine *abitazione* e, in tutti i casi, esso è legato alle limitazioni agli spostamenti. O meglio, i provvedimenti che introducono restrizioni agli spostamenti, considerano l'abitazione come il luogo in cui è comunque sempre consentito rientrare o, viceversa, il luogo dal quale non ci si può spostare. È così che ciascuno di noi si è trovato, suo malgrado, ad affrontare una delle prime problematiche del lavoro terminologico: la ricerca del significato di un concetto. Al tempo del coronavirus questo si è tradotto in una serie di domande a cui i cittadini chiedevano di dare risposta. La seconda casa può essere considerata *abitazione*? Lo studente universitario che vive abitualmente nella città sede dell'università può fare rientro nell'abitazione dei genitori? E altre ancora.

In assenza di una precisa definizione giuridica di *abitazione*, il Governo, consapevole della difficoltà di inquadrare esattamente il concetto, ha cercato di chiarirne il significato attraverso le FAQ pubblicate sulle pagine del proprio portale Internet. E così l'abitazione è stata identificata con il luogo in cui si abita di fatto con una certa continuità e stabilità (quindi per periodi continuati durante l'anno, anche se limitati) o con abituale periodicità e frequenza (per esempio, solo alcuni giorni della settimana: per lavoro, studio o altre esigenze), escludendo tuttavia dalla nozione di *abitazione* le seconde case utilizzate per le vacanze.

Per quanto concerne il lavoro terminologico in ambito giuridico, quella data dal Governo è una definizione settoriale e, quindi, eccessivamente circo-

stanziata. L'esclusione delle seconde case dal novero delle abitazioni e il riferimento ai criteri della stabilità e continuità dell'abitare sono elementi legati allo specifico contesto delle misure di emergenza nel periodo della pandemia e, pertanto, non possono essere accolti nella definizione di *abitazione* in una raccolta terminologica che si propone di definire i concetti, partendo dalla ricerca del loro significato più generale all'interno di uno specifico settore del diritto.

Accanto al problema derivante dall'assenza di una definizione giuridica univoca di *abitazione* si è posto un altro problema che possiamo identificare come uno scostamento di significato tra il termine giuridico, come esisteva prima dell'avvento del coronavirus, e il modo in cui è stato interpretato successivamente ad esso. È il caso ad esempio di *assembramento*. *Assembramento* è un termine che ricorre nell'ambito della sicurezza pubblica e indica le adunate di più persone avvenute senza una preventiva decisione, tali da potersi definire accidentali. Il termine rileva principalmente in relazione ai poteri delle autorità di imporre limitazioni al suo esercizio quando ricorrono determinate circostanze (Ciannamea 2017, 2–3). Durante il periodo del coronavirus il termine è diventato famoso in relazione ai *divieti di assembramento* imposti dai provvedimenti del Governo. In questo caso il Governo non ha specificato ulteriormente il termine, e questo ha portato a innumerevoli dubbi su come dovesse essere interpretato. Il significato di *assembramento* alla luce dei provvedimenti normativi emanati durante la pandemia è, a grandi linee, quello di raggruppamento di persone sia in luoghi pubblici sia privati vietato dalla legge per motivi di salute pubblica. Anche in questo caso, tuttavia, come abbiamo visto prima nel caso di *abitazione*, questo preciso significato non è stato accolto all'interno della nostra raccolta terminologica proprio perché legato a un periodo e a circostanze specifiche e, dunque, soggetto a cambiamenti di significato. Ad oggi, infatti, ad esempio gli assembramenti non sono vietati.

C'è un ulteriore aspetto legato alla definizione su cui è necessario soffermarsi: nell'ambito del lavoro terminologico multilingue, se un termine è definito in modo vago nella lingua di partenza, sarà probabilmente più difficile renderlo correttamente nella lingua di arrivo. In altre parole, più chiaro sarà il concetto, meno perplessità ci saranno nella sua resa verso un'altra lingua. Se infatti un concetto e il termine che lo designa sono poco chiari già nella lingua di partenza, ci saranno perplessità nella sua resa verso altre lingue con il rischio di generare confusione ed interpretazioni errate da parte dei destinatari. Prendiamo come esempio *caso di necessità*. Nel marzo 2020 questo termine è stato al centro del dibattito pubblico vista la genericità e la soggettività del suo significato, tanto che sono dovute intervenire testate nazionali e siti giuridici, come “La legge per tutti”, per specificare i tipi di spostamenti consentiti dalle nuove misure (Ralli 2020; Ubbriaco 2020). Tale termine è stato reso con *Notstand* (letteralmente: emergenza) nel tedesco altoatesino. Tuttavia, se, ad esempio, portare fuori il cane poteva rientrare nel *caso di necessità*, lo stesso non poteva

essere applicato per *Notstand*. Una volta compreso meglio il concetto italiano, la resa verso il tedesco si è orientata sul termine *Notwendigkeit*, ossia necessità, che rispecchia sicuramente meglio il termine italiano.

#### 4. Terminologia COVID-19 e comparazione

Date le finalità di *bistro* (par. 1.1.), elemento centrale del lavoro terminologico dell'Istituto di linguistica applicata è la comparazione giuridica fra l'ordinamento italiano e gli ordinamenti di lingua tedesca, ossia Germania, Austria e Svizzera. Da questo punto di vista, il coronavirus ci ha posto di fronte a delle sfide non indifferenti: l'Italia è stato il primo paese europeo ad essere colpito dalla pandemia e, conseguentemente, uno dei primi paesi a sviluppare la terminologia intorno ad essa. Nelle prime settimane dell'emergenza sanitaria, è risultato complesso individuare equivalenti o traducendo da adottare per la lingua tedesca in Alto Adige (Ralli 2021): nei paesi di lingua tedesca si parlava certamente già di coronavirus, ma ancora in riferimento alla situazione italiana. Nei paragrafi successivi si approfondiscono alcuni aspetti legati al raffronto terminologico tra i paesi di lingua tedesca e alle scelte terminologiche adottate in Alto Adige per evitare una eccessiva regionalizzazione del linguaggio giuridico.

##### 4.1. Armonizzazione della terminologia intralinguistica

Per evitare di creare nuovi termini tedeschi per l'Alto Adige, nella prima fase della pandemia che ha visto l'Italia come protagonista, si è fatto riferimento alla normativa dell'Austria, della Germania e della Svizzera già esistente in materia. A questo riguardo si è cercato nelle fonti normative vigenti in materia di infezioni ed epidemie, che sono:

- la legge austriaca sulle epidemie (*Epidemiegesetz 1950 – EpiG*);
- la legge federale tedesca per la prevenzione e il contrasto delle malattie infettive (*Gesetz zur Verhütung und Bekämpfung von Infektionskrankheiten beim Menschen – Infektionsschutzgesetz, IfSG*);
- la “Legge federale svizzera sulla lotta contro le malattie trasmissibili dell'essere umano” (*Bundesgesetz über die Bekämpfung übertragbarer Krankheiten des Menschen – Epidemiegesetz, EpG*).

Tuttavia, l'evoluzione della pandemia e la sua espansione geografica hanno portato anche in Austria, Germania e Svizzera alla nascita e al progressivo sviluppo di un vocabolario proprio della pandemia. Questo si è tradotto nella necessità di verificare periodicamente gli equivalenti individuati per il tedesco in Alto Adige, aggiornandoli all'occorrenza e adeguandoli alla terminologia degli altri paesi di lingua tedesca per evitare che le denominazioni tedesche

altoatesine acquisissero una connotazione troppo locale. Si è cercato in questo modo di favorire l'armonizzazione con la terminologia COVID-19 utilizzata negli altri paesi tedesofoni (Ralli 2021).

Un esempio al riguardo è dato dal termine *isolamento* che esisteva già prima dello scoppio della pandemia da COVID-19. In un primo momento all'interno della raccolta terminologica è stato inserito il termine *Absonderung* che appariva nelle leggi sopra citate e sembrava essere l'equivalente adatto per rendere il concetto italiano di *isolamento*. Con il diffondersi a macchia d'olio della pandemia in tutta Europa e la contestuale nascita e sviluppo di un vocabolario specifico per descriverla, si è potuto constatare che in Germania, Austria e Svizzera non si parlava però quasi mai di *Absonderung*, sebbene fosse il termine utilizzato dalla loro normativa in materia, ma semmai di *Isolation* e *Isolierung*, come già di fatto accadeva anche nelle comunicazioni istituzionali altoatesine. Si è reso quindi necessario fare una ricerca più approfondita ricorrendo anche a fonti non giuridiche. Dato il tema, un punto di riferimento importante da consultare è il dizionario specialistico *Infektionsschutz und Infektionsepidemiologie* dell'Istituto tedesco Robert Koch. A differenza delle fonti giuridiche, vi si trova il termine *Isolierung* che viene definito come una forma particolare di *Absonderung* (RKI 2015, 78). In questa pubblicazione del 2015, *Absonderung* è inteso come un iperonimo che comprende l'isolamento (*Isolierung*), da un lato, e la quarantena (*Quarantäne*), dall'altro. Ciò dimostra che in ambito medico-epidemiologico si distingue tra la separazione fisica di persone effettivamente infettate (*Isolierung*) e la separazione fisica di persone potenzialmente infettate (*Quarantäne*). Questa distinzione, che identifica due situazioni che richiedono comportamenti differenti, si è diffusa soprattutto nel linguaggio comune e divulgativo della pandemia. A livello normativo, tale differenziazione la ritroviamo nella suddetta legge federale svizzera: *Absonderung* rappresenta effettivamente l'equivalente del termine italiano *isolamento*, mentre per *quarantena* si utilizza il termine *Quarantäne*. Nelle leggi austriaca e tedesca sopramenzionate si è, invece, mantenuto il termine *Absonderung* per indicare entrambe le forme di separazione. A fronte di queste ricerche, *Isolierung* è stato aggiunto in *bistro*; il termine *Absonderung* è stato mantenuto per la Svizzera, come sinonimo di *Isolierung*, ma eliminato per Austria e Germania.

Altre volte il processo di adattamento del lessico ha seguito una direzione opposta: il lessico si è sviluppato a partire dai paesi di lingua tedesca ed è stato recepito in un primo momento in Alto Adige per poi espandersi successivamente al resto d'Italia<sup>5</sup>. Questo ha portato alla necessità di individuare

---

<sup>5</sup> L'Alto Adige gode di un'ampia autonomia e gli sono riconosciute diverse competenze legislative e amministrative. Inoltre, la vicinanza geografica e culturale con i paesi tedesofoni, in particolare con l'Austria, fa sì che a volte quest'ultima funga da modello di riferimento per la scelta di alcune misure legislative o regolamentari. Ciò è accaduto anche al tempo della pandemia, in cui l'Alto Adige ha preso in prestito alcune misure adottate in Austria per applicarle

una terminologia italiana univoca per esprimere i concetti tedeschi. Si pensi ad esempio al *test antigenico nasale rapido* introdotto nelle scuole altoatesine nella primavera del 2021. Sull'esempio dell'Austria, la Provincia di Bolzano ha introdotto tali test per poter consentire la ripresa e il funzionamento della scuola primaria in presenza. Nelle comunicazioni dell'Azienda sanitaria locale in lingua italiana si leggeva di *test rapidi antigenici per utilizzo in autosomministrazione* o di *autotest antigenici nasali*. Quando questa pratica si è diffusa anche nelle altre province e regioni italiane, si è svolto un controllo per verificare quale fosse la denominazione che veniva utilizzata correntemente nelle altre parti d'Italia in modo da poterla accogliere nella raccolta terminologica.

#### 4.2. Discrepanze terminologiche a livello inter- e intralinguistico

Sotto l'aspetto traduttivo, ci sono ulteriori fattori di cui tenere conto. Ogni paese è dotato di un proprio apparato concettuale e di proprie strutture cognitive (Sandrini 1996, 138; Šarčević 1997, 232). Operando un'analisi contrastiva, anche fra ordinamenti che parlano la stessa lingua, ci si trova spesso di fronte a equivalenze parziali o nulle che talvolta si tramutano in lacune terminologiche. A tal riguardo è opportuno chiedersi se sia possibile adottare la stessa denominazione per ordinamenti giuridici diversi o se, invece, non sia preferibile individuare una soluzione alternativa (es. proposta di traduzione) per evitare confusione o interpretazioni errate. Riprendendo l'esempio di *lockdown* (par. 3.2.), ogni paese ha adottato misure per contrastare la diffusione del virus che potrebbero rientrare approssimativamente nel concetto di *lockdown*. Tuttavia, l'implementazione di tali misure (es. tipologia, durata, condizioni, eccezioni) è stata realizzata in modi diversi dai vari paesi.

Da un punto di vista terminologico si hanno quindi due livelli di analisi (Palermo e Pfössl 1997, 17; Ralli 2007, 55; Ralli, Stanizzi, e Wissik 2007, 285-286; Chiochetti e Ralli 2016, 107):

- interlinguistico, vale a dire tra ordinamenti giuridici che parlano lingue diverse. Si verifica se due o più concetti appartenenti a ordinamenti giuridici diversi presentano le medesime caratteristiche fondamentali, valutandone le possibili similitudini e discrepanze;
- intralinguistico, ossia tra ordinamenti giuridici che usano la stessa lingua. In questa fase si verifica se due termini linguisticamente identici coincidono anche nel contenuto, se sono cioè concettualmente equivalenti.

Nel primo caso, da un raffronto, ad esempio, con l'Austria emerge una linea soft e una dura nell'attuazione delle misure di contenimento che hanno portato alla genesi di due concetti distinti: *weicher Lockdown* (= lockdown soft) e *harter Lockdown* (= lockdown duro). In Italia questa distinzione non

---

direttamente nel proprio territorio.

è stata fatta. Da un punto di vista concettuale sussiste quindi un rapporto di inclusione (Arntz et al. 2014, 147; Drewer e Schmitz 2017, 21) fra il concetto italiano e quello austriaco, essendo il primo più ampio rispetto al secondo.

Nel secondo caso, in un raffronto fra l'ordinamento italiano espresso in lingua tedesca per l'Alto Adige e l'ordinamento austriaco emerge che la denominazione omografa *Lockdown* può essere fuorviante: il termine tedesco altoatesino, riferendosi all'ordinamento giuridico italiano, documenta esclusivamente il concetto giuridico italiano; ne presenta pertanto le stesse caratteristiche. Ne consegue che anche fra il termine tedesco altoatesino e quello austriaco sussiste un rapporto di inclusione.

Benché il concetto di *lockdown* sia frutto della stessa circostanza, questo si è evoluto in maniera non uniforme a seconda dei paesi e delle misure di contenimento adottate, costituendosi di un proprio apparato concettuale, strettamente legato all'ordinamento che lo ha prodotto. Coincidenza denominativa, infatti, non significa necessariamente coincidenza concettuale.

## 5. Conclusioni

Univocità, chiarezza e accuratezza sono requisiti fondamentali della terminologia e, in particolare, della terminologia giuridica su cui si concentra il presente articolo. Solo una comunicazione pubblico-istituzionale, che riduce il più possibile il ricorso a termini ambigui, dal significato incerto, si rileva una comunicazione efficace. Nel periodo caratterizzato dall'emergenza sanitaria da COVID-19 è parso da subito evidente l'importanza di una comunicazione chiara e coerente verso il cittadino. Tuttavia, tale consapevolezza si è scontrata con una realtà incerta e un quadro terminologico complesso (lessico eterogeneo, mutevole, impreciso). Anche il lavoro di elaborazione terminologica ha dovuto fare i conti con tale complessa realtà e si è dovuto in parte rimodellare sulle nuove esigenze. Si sono moltiplicati gli sforzi per risalire al significato di concetti spesso troppo vaghi, per ricercarne le corrispondenze in altre lingue e altri paesi, per aggiornare e modificare quasi in tempo reale la terminologia, in modo da renderla sempre attuale. In alcuni casi si è ampliata la ricerca terminologica a fonti documentali non prettamente giuridiche, accogliendo termini provenienti da media locali e nazionali. Lo scopo è sempre stato quello di contribuire a un uso chiaro e uniforme della terminologia, nella consapevolezza che il problema di un corretto impiego della stessa nei tempi dell'emergenza da COVID-19 non investe solo gli organi pubblici, ma si ripercuote nei cittadini chiamati a conformarsi a regole di comportamento, talvolta oscure, e in generale nell'opinione pubblica che osserva e discute l'operato degli organi di gestione.

In questo contesto confidiamo che il procedimento sviluppato per definire, contestualizzare, sistematizzare e tradurre la terminologia relativa all'emer-

genza sanitaria si sia tradotto in un prodotto utile e adeguato rispetto ai suoi scopi principali: mettere a disposizione del pubblico la terminologia al fine di promuoverne un uso corretto e contribuire a rendere la comunicazione più efficace.

## Ringraziamenti

Un sentito ringraziamento va a Chiara Messina della Sezione di terminologia dei Servizi linguistici centrali della Cancelleria federale svizzera con cui è stato avviato uno scambio metodologico e terminologico durante la pandemia.

## Riferimenti bibliografici

- Arntz, Reiner, Heribert Picht, und Klaus-Dirk Schmitz. 2014. *Einführung in die Terminologearbeit*. 7 [1. Aufl. 1989 von Arntz/Picht.]. Hildesheim: Olms.
- Bordandini, Denaura. 2020. “Le misure restrittive COVID-19 quali illeciti amministrativi.” *Altalex*, 30 aprile 2020. <https://www.altalex.com/documents/news/2020/04/30/misure-restrittive-covid19-quali-illeciti-amministrativi>.
- Camera dei deputati. 2022. “Obbligo di possesso della certificazione verde Covid-19 per i lavoratori”. 28 marzo 2022. <https://temi.camera.it/leg19/temi/obbligo-di-possesso-della-certificazione-verde-covid-19-per-i-lavoratori.html>.
- CENSIS – Centro Studi Investimenti Sociali. 2021. “Disinformazione e fake news durante la pandemia: il ruolo delle agenzie di comunicazione. Rapporto Ital Communications – CENSIS.” Roma: Fondazione CENSIS. [https://www.censis.it/sites/default/files/downloads/Rapporto%20Ital%20Communications-Censis\\_def.pdf](https://www.censis.it/sites/default/files/downloads/Rapporto%20Ital%20Communications-Censis_def.pdf).
- Chiocchetti, Elena, und Natascia Ralli. 2016. “Ein Begriff, zwei Sprachen, unterschiedliche (Rechts)Kulturen.” In *Terminologie und Kultur. Akten des Symposiums, Mannheim, 3-5. März*, herausgegeben von Petra Drewer, Felix Mayer, und Klaus-Dirk Schmitz, 103–12. München: Deutscher Terminologie-Tag e.V. <https://hdl.handle.net/10863/8128>.
- Ciannamea, Fabrizio. 2017. “Libertà di riunione e possibili limitazioni. Con uno sguardo al Decreto Minniti e alla direttiva del Ministero dell’Interno sulle manifestazioni urbane.” *Giurisprudenza penale Web* 10. [https://www.giurisprudenzapenale.com/wp-content/uploads/2017/10/ciannamea\\_gp\\_2017\\_10.pdf](https://www.giurisprudenzapenale.com/wp-content/uploads/2017/10/ciannamea_gp_2017_10.pdf).



- CNB – Comitato Nazionale per la Bioetica. 2022. *La comunicazione istituzionale nell'ambito della pandemia: aspetti bioetici*. 17 marzo 2022. Presidenza del Consiglio dei Ministri. [https://bioetica.governo.it/media/4414/p147\\_2022\\_comunicazione-istituzionale-in-pandemia\\_it.pdf](https://bioetica.governo.it/media/4414/p147_2022_comunicazione-istituzionale-in-pandemia_it.pdf).
- Commissione Europea. 2020. “Domande e risposte sul bilancio dell’UE per la ripresa: il dispositivo per la ripresa e la resilienza.” European Commission, 28 maggio 2020. [https://ec.europa.eu/commission/presscorner/api/files/document/print/it/qanda\\_20\\_949/QANDA\\_20\\_949\\_IT.pdf](https://ec.europa.eu/commission/presscorner/api/files/document/print/it/qanda_20_949/QANDA_20_949_IT.pdf).
- Corbolante, Licia. 2020. “Lockdown vs distanziamento sociale.” *Terminologia etc.: Terminologia, localizzazione, traduzione e altre considerazioni linguistiche*. Blog post. 20 marzo 2020. <https://www.terminologiaetc.it/2020/03/20/origine-significato-lockdown-pandemie/>.
- “Decisione di esecuzione (UE) 2018/945 della Commissione del 22 giugno 2018 relativa alle malattie trasmissibili e ai problemi sanitari speciali connessi da incorporare nella sorveglianza epidemiologica, nonché alle pertinenti definizioni di caso.” *Gazzetta Ufficiale dell’Unione Europea*, L 170/1. <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32018D0945&from=EN>.
- Decreto del Presidente del Consiglio dei Ministri 4 marzo 2020. “Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell’emergenza epidemiologica da COVID-19, applicabili sull’intero territorio nazionale.” *Gazzetta ufficiale*, 4 marzo 2020, no. 55, <https://www.osservatoriosullefonti.it/emergenza-covid-19/fonti-governative/decreti-del-presidente-del-consiglio-dei-ministri/2797-emcov-dpcm-4marzo2020>.
- Decreto-legge 23 febbraio 2021, no. 15. “Ulteriori disposizioni urgenti in materia di contenimento e prevenzione dell’emergenza epidemiologica da COVID-19.” *Gazzetta Ufficiale*, 23 febbraio 2021, no. 45. <https://www.osservatoriosullefonti.it/emergenza-covid-19/fonti-governative/decreta-zione-d-urgenza/3906-emcov-dl-15-2021>.
- De Vecchis, Kevin. 2022. “Alla fine abbiamo imparato a tamponarci anche da soli. Note su tamponare e tamponarsi.” Accademia della Crusca, 9 maggio 2022. <https://accademiadellacrusca.it/it/consulenza/alla-fine-abbiamo-imparato-a-tamponarci-anche-da-soli-note-su-tamponare-e-tamponarsi/15133>.
- Domenici, Mauro. 2020. “Coronavirus: FAQ in ambito penale.” *Altalex*, 20 marzo 2020. <https://www.altalex.com/documents/news/2020/03/20/coronavirus-faq-in-ambito-penale>.

- Drewer, Petra, und Klaus-Dirk Schmitz. 2017. *Terminologiemanagement: Grundlagen – Methoden – Werkzeuge. Kommunikation und Medienmanagement*. Berlin: Springer Vieweg.
- Eurac Research. 2016. “Bistro – Sistema Informativo per La Terminologia Giuridica.” Consultato il 30 maggio 2023. <https://bistro.eurac.edu>.
- Eurac Research. 2023. “ConsTerm 2.0. Supporto scientifico in ambito terminologico.” Consultato il 30 maggio 2023. <https://www.eurac.edu/it/institutes-centers/istituto-di-linguistica-applicata/projects/consterm-20>.
- F. Q. 2021. “Un anno in stato di emergenza, il 31 gennaio 2020 il governo firma la delibera che prepara l’Italia all’epidemia di coronavirus.” *Il Fatto Quotidiano*, 31 gennaio 2021. <https://www.ilfattoquotidiano.it/2021/01/31/un-anno-in-stato-di-emergenza-il-31-gennaio-2020-il-governo-firma-la-delibera-che-prepara-l-italia-al-lepidemia-di-coronavirus/6085375>.
- Früh, Beate, and Dóra Mária Tamás. 2021. “Quality Evaluation of Termbases.” *Edition: Fachzeitschrift Für Terminologie* 2/21: 11–23.
- Giampreti, Claudia, e Samantha Sintuzzi. 2010. “Sinonimia e variazione nel lessico del trasporto aereo.” In *Terminologia a colori*, a cura di Franco Bertaccini, Sara Castagnoli, e Francesca La Forgia, 227–38. Studi Interdisciplinari su Traduzione, Lingue e Culture 10. Bologna: Bononia University Press.
- INPS. 2020. “Circolare n. 45 del 25 marzo 2020 – Congedo per Emergenza COVID-19 in favore dei lavoratori dipendenti del settore privato, dei lavoratori iscritti alla gestione separata di cui all’articolo 2, comma 26, della legge 8 agosto 1995, n. 335, e dei lavoratori autonomi. estensione permessi retribuiti di cui all’articolo 33, commi 3 e 6, della legge n. 104/1992, per i lavoratori dipendenti del settore privato. Istruzioni operative.” <https://www.assolombarda.it/servizi/lavoro-e-previdenza/documenti/circolare-inps-n-45-del-25-marzo-2020>.
- Magnani, Alberto. 2021. “Next Generation EU, cos’è e come funziona.” *Il Sole 24 ORE*, 4 marzo 2021. <https://www.ilsole24ore.com/art/next-generation-eu-cos-e-e-perche-l-europa-deve-correre-fondi-la-ripresa-covid-A-DIKpzMB>.
- Marazzini, Claudio. 2020. “In margine a un’epidemia: Risvolti linguistici di un virus – II puntata.” Accademia della Crusca. 02 aprile 2020. <https://accademiadellacrusca.it/it/contenuti/in-margine-a-unepidemia-risvolti-linguistici-di-un-virus-ii-puntata/7914>.

- Mayer, Felix. 2000. "Terminographie im Recht: Probleme und Grenzen der Bozner Methode." In *Linguistica Giuridica Italiana e Tedesca / Rechtslinguistik des Deutschen und Italienischen*, a cura di Daniela Veronesi, 295–306. Padova: UNIPRESS.
- MEF – Ministero dell'Economia e delle Finanze – Dipartimento della Ragioneria generale dello Stato – Servizio Studi Dipartimentale. 2021. *La risposta delle Istituzioni europee alla pandemia di Covid-19. Il programma NGEU e i Piani nazionali*. Roma: Studi e Pubblicazioni MEF – Ragioneria generale dello Stato. [https://www.rgs.mef.gov.it/\\_Documenti/VERSIONE-I/Pubblicazioni/Strumenti\\_e\\_Metodi/note\\_tematiche\\_SeSD/Rapporto-SeSD-La-risposta-delle-Istituzioni-europee-alla-pandemia-di-Covid19.pdf](https://www.rgs.mef.gov.it/_Documenti/VERSIONE-I/Pubblicazioni/Strumenti_e_Metodi/note_tematiche_SeSD/Rapporto-SeSD-La-risposta-delle-Istituzioni-europee-alla-pandemia-di-Covid19.pdf).
- Ministero della Salute. 2020a. "Covid-19. Raccomandazioni per la Prevenzione." [https://www.salute.gov.it/imgs/C\\_17\\_opuscoliPoster\\_443\\_allegato.pdf](https://www.salute.gov.it/imgs/C_17_opuscoliPoster_443_allegato.pdf).
- Ministero della Salute. 2020b. Direzione Generale della Prevenzione Sanitaria – Ufficio 5 Prevenzione delle Malattie Trasmissibili e Profilassi Internazionale. Circolare 0007922 del 09/03/2020. Consultato il 30 maggio 2023. <https://www.certifico.com/component/attachments/download/21483>.
- Ministero della Salute. 2023. "Covid-19: Cosa Sono SARS-CoV-2 e Covid-19." <https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioFaqNuovoCoronavirus.jsp?id=257&lingua=italiano>.
- Openpolis. 2023. "Coronavirus, l'elenco completo degli atti." <https://www.openpolis.it/coronavirus-lelenco-completo-degli-atti/>.
- Pacini, Fabio. a cura di. 2020. "Elenco d.P.C.M." *Osservatorio sulle fonti*. Consultato il 30 maggio 2023. <https://www.osservatoriosullefonti.it/emergenza-covid-19/fonti-governative/decreti-del-presidente-del-consiglio-dei-ministri>.
- Palermo, Francesco, e Eva Maria Pfösl. 1997. *Normazione linguistica e tutela minoritaria / Minderheitenschutz durch Sprachnormierung*. Accademia Europea: Bolzano.
- Prandi, Michele. 2010. "Lessico naturale e lessici di specialità: Tra descrizione e normalizzazione." In *Terminologia a colori*, a cura di Franco Bertaccini, Sara Castagnoli, e Francesca La Forgia, 53–84. Studi Interdisciplinari su Traduzione, Lingue e Culture 10. Bologna: Bonomia University Press.
- Ralli, Natascia. 2007. "Die Normierung aus terminographischer Sicht." In *Italienisch-Deutsche Studien Zur Fachlichen Kommunikation*, herausgegeben von Dorothee Heller, und Piergiulio Taino, 2: 53–64. Deutsche Sprachwissenschaft in Italien. Frankfurt: Peter Lang.

- Ralli, Natascia. 2020. "Il o la COVID-19? Dell'articolo non v'è certezza... o forse sì?" *Eurac Research – Science Blogs "Covid-19 And Beyond"*. Blog post. 12 maggio 2020. <https://blogs.eurac.edu/covid-19/dellarticolo-non-ve-certezza>.
- Ralli, Natascia. 2021. "Terminologia COVID-19 e bistro: caratteristiche e problematiche." Contributo presentato al *Convegno Ass.I. Term 2021 Ieri e oggi: la terminologia e le sfide delle Digital Humanities*, Verona 10 dicembre 2021.
- Ralli, Natascia, e Annette Lenz. 2022. "Quando la collaborazione istituzionale funziona: Il Progetto ConsTerm." In *Risorse e Strumenti per l'elaborazione e la diffusione della terminologia in Italia*, a cura di Elena Chiochetti, e Natascia Ralli. Bolzano: Eurac Research. <https://doi.org/10.57749/wtfr-y339>.
- Ralli, Natascia, e Isabella Stanizzi. 2008. "Il dietro le quinte della normazione." In *Normazione, armonizzazione e pianificazione linguistica/ Normierung, Harmonisierung und Sprachplanung/ Normalisation, harmonisation et planification linguistique*, 61–74. Bolzano: Eurac.
- Ralli, Natascia, e Isabella Stanizzi. 2018. "Il linguaggio giuridico tedesco in Alto Adige. Evoluzione delle politiche terminologiche". *AIDAInformazioni* 36 (numero speciale): 169–89.
- Ralli, Natascia, Isabella Stanizzi, e Tanja Wissik. 2007. "Der Bologna-Prozess und die Folgen für die italienische und österreichische Universitätsterminologie." In *Proceedings of the Conference 'Översättningsteori, Fackspråk Och Flerspråkighet'*, edited by Nina Niemelä and Esa Lehtinen, 34: 281–92. Vaasa: VAKKI.
- RKI – Robert Koch-Institut. 2015. *Infektionsschutz und Infektionsepidemiologie. Fachwörter – Definitionen – Interpretationen*. Berlin. [https://www.rki.de/DE/Content/Service/Publikationen/Fachwoerterbuch\\_Infektionsschutz.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/Service/Publikationen/Fachwoerterbuch_Infektionsschutz.pdf?__blob=publicationFile).
- Sacco, Rodolfo. 1992. *Introduzione al diritto comparato*. Torino: UTET.
- Sacco, Rodolfo. 2018. *Sistemi giuridici comparati*. Torino: UTET.
- Sala, Marianna, e Massimo Scaglioni, a cura di. 2020. *L'altro Virus. Comunicazione e disinformazione al tempo del COVID-19*. Milano: Vita e Pensiero. <http://hdl.handle.net/10807/161283>.
- Sandrini, Peter. 1996. *Terminologiarbeit im Recht. Deskriptiver Begrifforientierter Ansatz vom Standpunkt des Übersetzers. IITF-Series 8*. Wien: TermNet.
- Šarčević, Susan. 1997. *New Approach to Legal Translation*. The Hague: Kluwer Law International B.V.

- Servizi e uffici del Senato della Repubblica e della Camera dei deputati, a cura di. 2021. “Misure urgenti per fronteggiare l'emergenza epidemiologica da COVID-19 e per l'esercizio in sicurezza di attività sociali ed economiche D.L. 105/2021 – A.S. 3223.” *Dossier XVIII Legislatura* 415. <https://documenti.camera.it/Leg18/Dossier/Pdf/D21105.Pdf>.
- Sfardini, Anna. 2020. “Come comunicare la pandemia? Credibilità e fiducia delle fonti istituzionali nell'informazione italiana sul Covid-19.” In *L'altro Virus. Comunicazione e disinformazione al tempo del COVID-19*, a cura di Massimo Scaglioni, e Marianna Sala, 63–74. Milano: Vita e Pensiero. <http://hdl.handle.net/10807/161283>.
- Soglia, Susanna. 2002. “Origine, Sviluppo e Tendenze Della Terminologia Moderna.” In *Manuale di Terminologia*, a cura di Marella Magris, Maria Teresa Musacchio, Lorenza Rega, e Federica Scarpa, 10–25. Milano: Hoepli.
- Treccani. 2022. *Dizionario dell'italiano Treccani. Parole da leggere*. Roma: Istituto dell'enciclopedia italiana.
- Ubbriaco, Denise. 2020. “Coronavirus: quando posso uscire di casa?” *La Legge per Tutti*, 11 marzo 2020. [https://www.laleggepertutti.it/375334\\_coronavirus-quando-posso-uscire-di-casa](https://www.laleggepertutti.it/375334_coronavirus-quando-posso-uscire-di-casa).
- Wüster, Eugen. 1985. *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. Infoterm.

# Handwritten Text Recognition as a digital perspective of Archival Science

Salvatore Spina\*

**Abstract:** The digital divide in the Humanities scientific field represents a heated debate between traditionalists (analogue scholars) and digital humanists. While some progress has been made towards a dialogue between the two sides, friction persists. Technology, including the development of artificial intelligence tools and algorithms, is not a threat to humanities research but a solution to problems. However, society has changed dramatically, and new generations require new communicative products and systems. The debate becomes increasingly wearying, but “analogues” do not consider the mutation of reality’s interpretive patterns and prefer to rely on the “death of traditional Humanities statutes”. The closed-mindedness towards Information technology and communication (ICT) tools, such as Handwritten Text Recognition (HTR), makes no sense. This study aims to demonstrate the potential of automatic transcription as a helpful tool for archival fields and research.

*Keywords:* Transkribus, Biscari Archive, Italian Administrative model, Filemaker, Digitization.

## 1. Introduction

If we were to talk about the *digital divide*, indeed, the world of the Humanities is a valid testing ground for the exemplification of the debate that pits, on the one hand, traditionalists (analogue scholars) against digital humanists. The heated and profound debate often shows a position of fear and anxiety about the future of research and does not look for mediation. Of course, some progress has been made in this direction, and a likely dialogue – always calibrated on the absence of trust – could mediate the ideas of both parties. However, there remains constant friction between the parties involved. The underlying assumptions behind the positions can probably be poorly understood, especially by those who look at technology as an inadequate solution to a problem that, actually, has been solved since the invention of the wheel, thanks to a mechanical-technological invention. Just think of Pascaline. If technology has always been the solution to all those problems dictated by

---

\* Università degli Studi di Catania, salvatore.spina@unict.it

human resource inadequacy, then the development of artificial intelligence tools, neural systems and algorithms should not be considered a menacing weapon for eliminating the human operator/worker. However, in the face of innovation, the “day of judgment” and the “embodied evil” have been the only weapons available to “traditionalists”.

«[Is p]assing, O men, Satan the great»<sup>1</sup> (Carducci 1964) – with these words, for example, trains were greeted as they clattered through Europe, which had recently invented the railway infrastructure. Society has changed dramatically, not in terms of government and political systems – which remain expressions of business interests devoid of concern for the community – but in the demands of new generations for communicative products that have, beyond deliberate consequences, put ancient disciplines to the corner, unable to keep up with humans evolution towards the structuring of a new cognitive entity that, in an epigenetic key, will be perennially connected (it is *Homo-Loggatus*) (Spina 2022a). The Turing Machine and the Internet network will log every intellectual and rational activity. Evolution does not go in a direction different from this just stated. For this reason, the debate among humanists becomes increasingly wearing, especially for “analogues,” whose position does not take into account a mutation of reality’s interpretive patterns. Thus, rather than promoting the evolution of discipline statutes, trying to make the new language of Sciences – Information Technology – its own, it is preferred to appeal to the “death of humanities research” (Valacchi 2021) or to the impossibility, for instance, of archivists, to move in the direction of building the Big Data complexes of History.

In this one-sided debate perspective, the strong rejection of ICT tools such as HTR makes no sense. For this reason, the study conducted on the part of the “Correspondence” section of the Paternò-Castello family archive aims to demonstrate the potential of the Transkribus technology and, at the same time, be a clear invitation to create an automatic transcription model that could become an essential tool for the correct digitisation of the Italian archives’ heritage.

Digitalisation represents a frontier, but that, in many ways, is not viewed from the correct perspective. Despite clear ideas between those who fear (because they do not know) computer technologies and those who see it as a logical, technologically valuable tool for scientific research, the meaning that humanists attribute to the term “digitisation” is still incorrect. Therefore, clarification is mandatory: “To digitise” means translating analogue pieces of information into machine-readable language. To digitalise is to “encode” any statement from one language to another. Remembering, therefore, the events that marked the development of the first computers (in the 1930s), the cor-

<sup>1</sup> «Come di turbine / L’alito spande: / Ei passa, o popoli, / Satana il grande. / Passa benefico / Di loco in loco / Su l’infrenabile / Carro del foco».



rect digitalisation required the creation of punch cards on which information was formalised in binary code or in a “proposition” that the mainframe could analyse. Fortunately, nowadays, “to digitalise” does not mean “writing in binary code”, but this does not mean that data must be encoded in digital text to apply various computational tools.

“Digitisation” and “digitalisation” are often used interchangeably but have different meanings<sup>2</sup>. “Digitisation” is often seen as merely “taking a picture” of an archival document and uploading it to a database.

Many projects that involve the digital acquisition of archival documentation beyond metadata do not operate in this direction; that is, complete digital texts are not adequately realised. This deficit, of course, becomes an obstacle to a more profound use of the semantic web.

This viewpoint is limited as it fails to take into account that a picture, in the digital environment, cannot be read by an artificial intelligence platform in the same way historians read the text contained within it. For a computer, an image is simply a data set, such as colour codes, dimensions, and dpi. Consequently, a proper digitisation process of an archival document requires more than photographic acquisition if we aim to create its digital version. The latter necessitates the accurate transcription of every word present in the image to compile the document’s digital edition.

Therefore, starting from the late 1950s with the rise of Tom Dimond, the idea of developing technologies for automatic recognition of handwritten text became increasingly (Adamek, O’Connor, and Smeaton 2007; Albertin et al. 2016; Dunley 2018; Spina 2022b) intense:

In the last five years, much thought and effort have gone into the development of printed character-recognition devices. Varying degrees of success have been achieved. In some cases, ingeniously distorted typefaces have been required. One might wonder why all this interest exists. The answer is simple. Character-recognition devices help reduce the cost of getting information into forms that computers can understand (Dimond 1957).

As we know – of course – when we delve into the field of archival or historical research, the most severe difficulties in using computer technologies are related to the existence of certain limitations: archival documentation mainly consists of manuscripts, a significant challenge for using computer tools to study events in the Modern Age – a situation where challenges of Archival Science and archivists occur.

The debate on the usefulness of computer tools seems not to affect the principles of this discipline. Despite the scientific and technological dynamism

---

<sup>2</sup> (1) “Digitization,” which refers to the conversion of analog information into a digital format; (2) “Digitalization,” which involves the native creation of information in a digital format; (3) “Digital transformation,” which is the process aimed at the exclusive adoption of technology in production processes.

sm, archivists take refuge behind statutory principles that seem to make this scientific field resistant to the needs of the research world.

What emerges from monographs and many scientific publications is a desire to defer the confrontation with computer technology, which could improve the activities of archives and archivists, accusing historians and scholars of fragmenting documentary resources when published on some websites (Vallacchi 2021) – this is because the multiplication of users through web access requires a reassessment of the paradigms of archival science and a stance that makes the difficulties of archivists in realising digitalisation projects adequate to the demands of the scientific community and, on the other hand, from computer technology, which considers the simple photographic acquisition of documents useless.

While the Italian National Archival System is advertised as one of the best products for digitising the national archival heritage, we face millions of photographs that remain meaningless for the semantic web neural system and the everyday user, who cannot perform an analytical search – this is because no transcription phase follows the photograph, neither human (reasonably impossible) nor automated.

HTR computer tools seem to be little known to the most prepared archivists, whose function, according to the dictates of their discipline, remains to preserve and not to make the archival heritage accessible through the Internet.

The limits are undoubtedly evident. It will not be possible to digitise and transcribe every Italian archival document quickly. However, it is also true that Artificial intelligence like Transkribus and the creation of increasingly performing transcription models could boost the creation of a digital heritage that can guarantee concrete access to the sources of our Past.

So, thanks to the Transkribus artificial intelligence platform (Erwin 2020; Kahle et al. 2017; Massot, Sforzini, and Ventresque 2018; Milioni 2020; Muehlberger et al. 2019), this limit begins to waver, as demonstrated by the test I carried out on the archival documentation of the Paternò Castello family, Princes of Biscari.

## 2. Biscari Archive and HTR Transkribus

Consider the assumption: a text depicted in a picture – in a technological process that, nowadays, is digitising objects, places, and men – cannot be read by artificial intelligence tools in the same way historians read the exact text. To a computer, a picture is a set of other kinds of data, such as colour code – e.g., #0000 is ‘black’, #654321 is ‘brown’, #FF0000 is ‘red’, and so on –, definition, number of pixels, and location information if the camera has Global Positioning System (GPS). Thus, a proper digitisation process would involve more than just taking pictures of historical documents.

Despite the limitations mentioned, there are still important questions to be addressed when it comes to digitising archival heritage: Is taking a photo the right way to digitise it? Can a computer read handwritten texts? How can computer tools and artificial intelligence enhance the profession of historians and archivists?

Alan Turing already posed the question: “Can machines think?” (Turing 1950) ... a question that has sparked debates and, above all, perspectives that, to this day, place scholars on opposing theoretical fronts based on unscientific and unconvincing foundations. In our case, the question is much simpler: “Can machines read?”

If we considered reading solely from the perspective of mechanising the act of recognising a character, the answer would undoubtedly be affirmative. However, everything changes when computers, hardware, and software do not operate on machine-readable texts.

Nevertheless, starting from the concept of “learning” – which characterises the life of humans who, after proper and adequate schooling, can recognise written signs and reproduce them – even a software or artificial intelligence tool can do the same – it is the “machine learning” (Alpaydin 2020; Bishop 2006; Collobert and Weston 2008; Gori 2018) –, while still maintaining a distinct semantic level, in which the machine remains deficient.

A study of the Biscari Archive in Catania can help address these issues and demonstrate the feasibility of using HTR technology in historical research.

Most archivists believe that the digital methodology consists of scanning or photographing some archival document and uploading it to the official archive website. However, a photo – as previously stated – cannot be read by artificial intelligence tools the same way scholars read the text of the image. If we want to conduct an archival or historical study by analysing some depicted text, no computer can figure out all the information it carries because it is inaccessible even to the Turing Machine. If this were the case, the computer could only reproduce words without relation to reality because language is a code structured on non-explicit elements, inconsistency between terms and actions and other aspects that computers cannot process. Furthermore, even if a computer can analyse a digital text, it cannot read and recognise a handwritten word, not even a single consonant or vowel.

The Biscari Archive has 2,000 folders with hundreds of thousands of sheets documenting legal disputes, political decisions, and business and personal letters. The “Correspondence” section consists of some 42,493 papers, in which, in folder 1.642, we can find 591 sheets, constituting 366 letters and a manuscript by Emile Rousseau. So, to test the capabilities of Transkribus AI, it was decided to use these 367 documents.

The acquisition process was performed using a Nikon D610 camera with an AF-S Nikkor 24-120mm f/4G ED VR lens. The photos were collected in

a database created using Claris FileMaker 19 software, as no website was planned for this project stage. After adding metadata, the photos were converted to PDF format to reconstruct the 366 letters, and 53 were uploaded to the Transkribus server for automatic transcription.

Transkribus, powered by Java technology, allows for creating workflows based on deep neural networks that can be trained to recognise specific handwriting, and to achieve the best results, texts must be written by the same hand. For this reason, 53 letters were selected, consisting of 28 letters from Michele Maria Paternò to Princess Anna Maria Morso Bonanno and 25 letters from Marquis Giovanni Fogliani Sforza D’Aragona to various recipients.

With the help of Gephi software (Bastian, Heymann, and Jacomy 2009), 53 letters were pinpointed and selected – consisting of 28 letters from Michele Maria Paternò to Princess Anna Maria Morso Bonanno and 25 letters from Marquis Giovanni Fogliani Sforza D’Aragona to various recipients (Fig. 1) – and sorted into two separate PDF files, which were then uploaded to the Transkribus server.

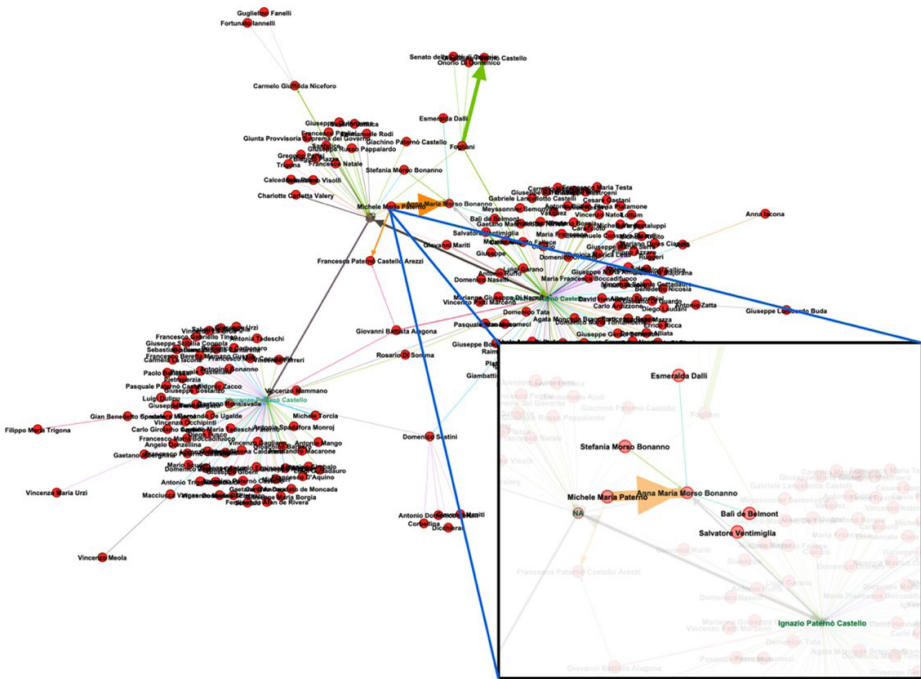


Figure 1: Correspondence between Michele Maria Paternò and Anna Maria Morso Bonanno.

Thanks to the small corpus size, it allowed for quick and accurate automatic transcription. Transkribus aim to support historical research by transcribing large volumes of handwritten texts. However, the Italian scientific community

is still reluctant to use state-of-the-art IT tools for historical research, leading to a delay in digitising historical heritage. In contrast, Finland has embraced AI for historical research by involving students and scholars in training models to minimise errors in transcriptions. One of the core features of Transkribus is the ability to involve the broader community in digital transcription projects, making the goal of Public History achievable.

### 3. Outcome

The automatic transcription of the letters was done without any prior training to evaluate the accuracy of the public Transkribus AI model. The “Italian Administrative Hands 1550-1700” model (Midura 2020)<sup>3</sup>, created by Jake Dyble, Antonio Iodice, Sara Mansutti and Rachel Midura, with a Character Error Rate (CER) of 9.15%, was used for the transcription. This result is considered excellent for a public model in the automatic transcription of manuscripts in Italian archives, which hold the most diverse historical heritage in Europe due to various writing styles, languages, and institutions – e.g., Latin in the Vatican, Spanish in the Kingdom of Sicily and the Duchy of Milan, Italian in the Grand Duchy of Tuscany.

The model was applied when the two PDF files were uploaded to the Transkribus server. The error rate for Marquis Fogliani’s letters was 7% (e.g., 6 out of 82 words were not transcribed correctly in the first letter) and remained constant throughout. However, the error rate was much higher, over 10%, for the 28 letters written by Michele Maria Paternò, due to incorrect Italian words and words unknown to the model. The graphemes “V.S.” and “S.M.” were also not transcribed, even though they were present in the text image segmentation, resulting in a higher error rate and an inability for the user to correct it.

---

<sup>3</sup> The “Italian Administrative Hands” model features a variety of Italian-language documents only from State Archives in Milan (from the *Carteggio delle cancellerie dello Stato*, *Atti di Governo*, and *Registri delle cancellerie dello Stato* collections), Venice (from the *Compagnia dei corrieri*, *Senato*, and *Inquisitori di Stato* collections), Florence (fond *Mediceo del Principato*, series *Relazioni con stati italiani ed esteri*), Pisa (from the *atti civili* sub-series of the archive of the *Consoli del Mare di Pisa*), and Genoa (from the *Notai Giudiziari* and the *Conservatori del Mare* collections).

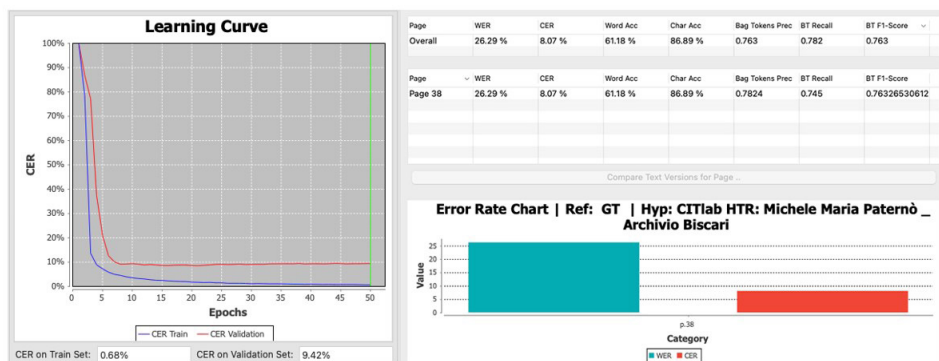


Figure 2: Correspondence between Michele Maria Paternò and Anna Maria Morso Bonanno.

Therefore, a new model was trained based on the “Italian Administrative” one, using the 24-page “ground truth” as the starting point (pages 1-5, 7-9, 11-18, 20-24, 26, 27, 35). The CER of the Training Set dropped to 0.68% and that of the Validation Set to 9.42% after the training and formation of the “Michele Maria Paternò\_Archivio Biscari\_Model” (Fig. 2). Comparing the “ground truth” and automatic transcription using the latter model showed a CER of 9.88% (Fig. 3). The validation set had a relatively high percentage of 35.57%, indicating the need for training on a more extensive “ground truth” of at least 50 pages. Nevertheless, the “Michele Maria” model resulted in a transcript with a CER of 8.07%/26.29 and minor errors that do not impede scholarly inference but make computational analysis unreliable.



1-1 # Messina 13 Guglio 1782  
 1-2 # Riv.ma S.ra Principessa  
 1-3 # 373  
 1-4 # ~~Mon o~~ Non ò termini che ~~dastino~~ bastino a ringraziarvi ~~dell~~ dell'  
 1-5 # a geteccosa premuva, ch avete affettuosa premura, ch avete perme, giacchè prendete ~~agitan~~ agitazne  
 1-6 # Da dremuoti da Tremuoti per la mia ~~bersona~~ persona. Io non posso; ma ~~dedio~~ Iddio vi  
 1-7 # ~~pieo~~ rimunievare, può rimunierare, e vi ~~rimenerira~~ rimunerirà. Quindi vi dico di levarvi  
 1-8 # da ~~soleitudine~~, sollecitudine, che non ~~se s'è~~ inteso ~~pui più~~ niente; ~~l~~ lo vi assicuro  
 1-9 # che sto di ~~beona~~ buona salute alla quale giovan molto le piacevoli  
 1-10 # ~~noti gie~~ notizie della vostra, e del Principe vro ~~conforte~~, consorte, e di tutta  
 1-11 # la vra Famiglia, che caramente riverisco.  
 1-12 # Me pesa molto che il ~~prodevo~~ prodotto de ~~frementi hia~~ frumenti sia  
 1-13 # ~~boato~~ stato così scarso, che ~~nia fatto~~ n'è Fatto mettervi il ~~presso~~ prezzo di ~~gd: Vapto 3:15:8a~~  
 1-14 # ~~puor de fuor de'~~ Caricatori, e che ~~giorsi~~ gli orzi costino ad ~~ste: 1:25:~~ col timo  
 1-15 # re di qualche ~~acemento~~ aumento  
 1-16 # Mi pesa altresì che non ~~o ò~~ notizia da darvi, perchè  
 1-17 # ne da Napoli, ne da Palermo me ne son venute in ~~quasta~~ questa  
 1-18 # ~~Pattimana~~ Paprete Settimana. Saprete che da Palermo ~~fravani~~ stavan per partire ~~gi gli~~  
 1-19 # ~~Geiadecchi~~ Sciadecchi colle Parsone destinate allo ~~Spurno~~ Spurgo dell'Isoletta  
 1-20 # di ~~lampedusa~~: Paprete ~~pire~~ Lampedusa: Saprete pure che si vuole per sicuro il ritor  
 1-21 # no dl Vicere ~~baracciolo~~ Caracciolo in ~~Du~~ Ott.re, e che vuol passare da ~~Gue Que~~  
 1-22 # ~~va~~, sta, e forse poi da codesta e veder ~~l'Esna~~ l'Etna. Or ~~piee più~~ di tanto  
 1-23 # non ò che dirvi in rapporto a ~~notinie~~ notizie avviate.

Compare Advanced Compare

Reference: (Correct Text) 05.10.22 15:01:15 - salvatore.spina@unict.it - In Progress Use current

Hypothesis: (HTR Text) 05.10.22 15:02:54 - salvatore.spina@unict.it - In Progress Use current

Previous Advanced Compare Results

Created	Status	Queries	Duration	Scope	Type	Results
05.10.22 15:02:06	Completed	Page(s) : 38   Option : Quick Compare   R 0.51 sec.		Document 1020 HTR		CER/WER: 8,07%/26,29%

Figure 3: Comparison set.

After testing the new model, it was HTR the remaining part of the corpus, testing both the Michele Maria Paternò model and the Italian administrative, obtaining, with the latter, a transcription with minimal errors. Once exported, ChatGPT – a Large Language Model (LLM) technology by OpenAi (a digital project by Elon Musk and Sam Altman in 2015) (Fig. 4) – corrected the transcription managing to obtain a digital corpus on which it was possible to apply additional text analysis tools, such as Keyphrase Digger (Moretti, Sprugnoli, and Tonelli 2015), which allowed to dig new historical information.



Messina 8 Maggio 1781 Il Bali Paternò Rivma Sig.a Principessa Infinito è il contento che mi ha fatto sperimentare la vra pregma del 4. andante, pella notizia, che mi dona di vra perfetta salute, e di quella ancora del PnpeVroe vostri Figli. Io sono nel godimento dell'uqual beneficio, e sono tutto intento per eseguire qualunque vro comando Vi ringrazio di quanto mi favorisce col tanto degno Pre Puglieri, dal quale ho ricevuto due lettere col passato ordinario la prima è con quello di qta Settimana la seconda, e siccome non potei riscontrarlo oggi otto, costi lo fò questa sera



Ecco la correzione del testo:

Messina, 8 maggio 1781

Il Bali Paternò, Rivma Sig.a Principessa, infinitamente contento sperimento la vera premura del 4. andante alla notizia che mi dona di vostra perfetta salute e di quella dei vostri figli. Sono nel godimento dello stesso desiderio e sono tutto intento ad eseguire qualsiasi vostro comando. Vi ringrazio per quanto mi favorite con il tanto degno Pre Puglieri, dal quale ho ricevuto due lettere col passato ordinario; la prima è di questa

Figure 4: Correction in ChatGPT.

#### 4. Transkribus and Filemaker

Beyond conflicting positions, the digital divide brings out the need for an innovative approach to humanities studies. Of course, information technology can boost digital asset organisation workflows and those to be digitised. Technology allows the development of heterogeneous approaches and produces digital documentation that can be analysed, organised, and structured differently. Digitalisation is total control of the archival and historical sources.

The study conducted at the Biscari Archive demonstrates how these technologies can interact and assist historians in reconstructing past events with more comprehensive and accurate information. Transkribus allows for the export of transcriptions in PDF files, which can be analysed, processed, and organised into databases, enabling searches through a user interface and relational system that highlights new information.

Thus, Transkribus is a critical tool for building the Big Data of History (Kaplan and di Lenardo 2017) and aligns with Gardin's belief that the creation of databases and ontologies must be accompanied by the proper encoding and formalisation phase, which converts historical texts into a computable format.

The 367 transcriptions from the Biscari Archive were entered into a database created with FileMaker 19, a software that facilitates the creation of a relational structure between the data. FileMaker 19.5 has also added a new JSON function, allowing for the differentiation of numbers and text and using system libraries, Optical Character Recognition (OCR), and HTR images in

a “container field” (Fig. 5). This addresses the need to digitise photographic sources and prepares the databases for creating machine-readable sources.

Transkribus highlights the importance of upgrading historians’ skills in a digital environment, focusing on the most crucial aspects of source analysis: organisation and transcription.

The collections created by Transkribus users represent a valuable archival heritage beyond automated transcription capabilities. Despite being often wrongly considered disconnected fragments from the primary documentation, these collections demonstrate the perspective that Archival Science should follow to meet the needs of the highly technological society. Additionally, by opening up to society (crowdsourcing), there will be opportunities for more profound training, possibly creating a model that can transcribe the Italian archival corpus. Furthermore, the transcriptions can be exported in formats that text analysis tools can use, and, as demonstrated in our case, PDFs can be obtained that can be inserted into databases, platforms, and websites for dynamic indexing and searching. The 367 documents will be added to the website “Biscari Epistolography – Archivio Biscari Archivio di Stato di Catania” (Spina 2023), providing a means of disseminating knowledge to all connected users. The dynamism of the PDFs will also provide a text search and export function (Fig. 6), taking advantage of the full potential of computer tools for data analysis.

The image displays two screenshots of a digital archive interface. The top screenshot shows the document's metadata page for 'ASCI\_Biscar\_L1642'. It includes fields for 'Mittente (Source)' (Michele Maria Paternò) and 'Destinatario (Target)' (Anna Maria Morso Bonanno). Below this is a table with columns: 'Soggetti', 'Luoghi', 'Eventi', 'Keyword', and 'Note'. The bottom screenshot shows the document's transcription page, featuring a scanned image of a handwritten letter on the left and a transcription area on the right. The transcription area has a 'Testo' tab and a 'Descrizione breve del documento' field. At the bottom of the transcription page are buttons for 'Salva PDF', 'Importa testo da PDF', and 'Esporta testo'.

Figure 5: The database, Transkribus and Filemaker.

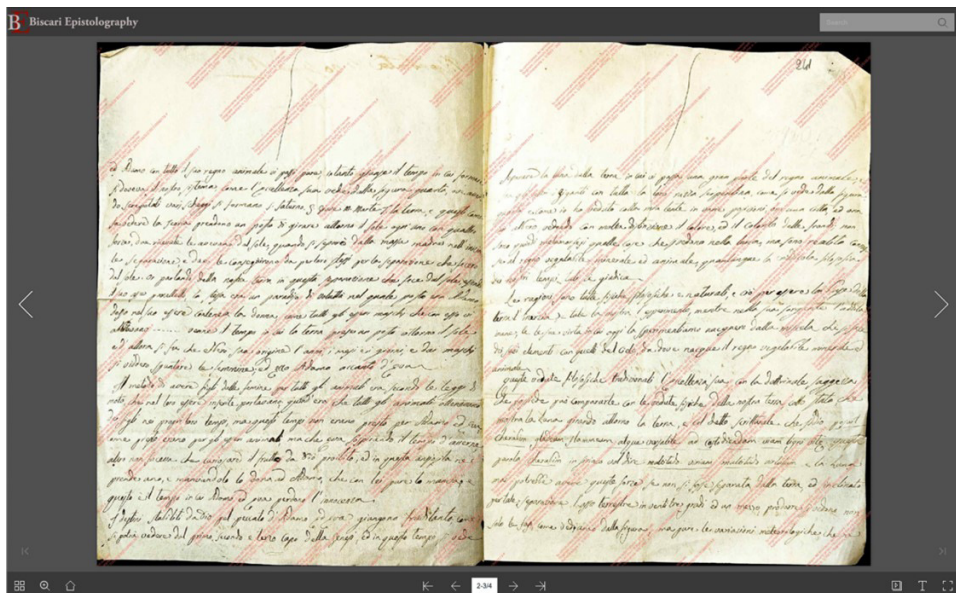


Figure 6: Pietro Musumeci flipbook on “www.biscariepistolography.it” website.

## 5. Achievement

The concept of “digitisation” has become widespread in our society. However, if digitisation can be considered a space-time acceleration towards a dimension of participation, communication, versatility, and usability, surely Archival Science can redefine itself within this assumption. Ordering is the most significant activity of the archivists’ craft, and this requires them to give a controlled and scientific sense to a community that, as a whole, is participating in the construction of its heritage to be entrusted to the care of the archives. Indeed, anxiety arises when referring to contemporaneity. Today, every single document finds space in sedimentation that is based on digital archives, which, in addition to their traditional function, acquires the function of an open-source structure and rapid access. The discourse, although similar, takes on an even more pressing tone when it comes to historical archives, that is, documentary complexes on which, through the web, everyone wants to organise research. The paper, indeed, is there and remains there, at complete disposal. However, if the archivist is also the one who holds the keys to reconstructing the time of events, he must reasonably accept that everything that is not on the web does not exist. The post-modern, digital, digitised cultural structure is based on a clear epigenetic principle: the world changes our being in relationships and the vision of spaces, and this became an appropriate request from a community that constantly lives on the web and interpersonal communi-

cation platforms – this requires a stance that is not destructive but wants to enhance the discipline of archival science and the profession of the archivist to meet the will of web users to access their documentary heritage (current and historical). Therefore, the possibility of using tools such as Transkribus must become a “necessity”. Not only to allow the scientific community to deepen its research but to decode complex documentary texts into a machine-readable digital structure.

Moreover, the task – which is currently in the hands of researchers who put some small documentary series on the network, extracted from certainly larger corpora – of this digitisation must be foundational to the vocation of archivists, who, like everyone else nowadays, can no longer exempt himself from computer training that enhances his traditional function. To this, there is a need to create automatic transcription models or boost the “Italian Administrative” with new training that features Italian-language documents from archives in the South of Italy, which can be an open-source tool for the Italian archival system.

Furthermore, despite the reluctance of the Italian archivist community, Transkribus is one of the best tools to open up to society. Whilst the international archivists’ community aspires to build enough digitised historical heritage, the Italians – apart from the case of the (Fondazione Banco di Napoli, n.d.) and the (Archivi Storici e Biblioteche Istituto Suor Orsola Benincasa, n.d.) – are failing to break out of the constraints of the legal protection of archival heritage, leading to a delay in digitising historical heritage. In addition, Italy lacks a scientific mindset to ensure the dissemination of its national archival heritage, unlike, for example, Finland, where, thanks to the “National Archives project” (The National Archives of Finland 2023), the Central Government wanted to restore to its community the archival heritage related to the Second World War, the judicial registers and property inventories of the Finnish nobility (Fig. 7 and Fig. 8).

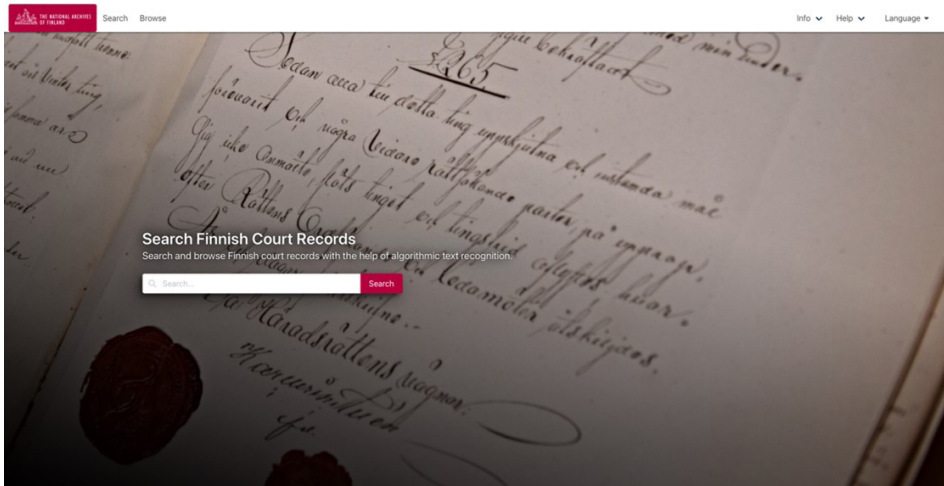


Figure 7: The National Archives of Finland website.

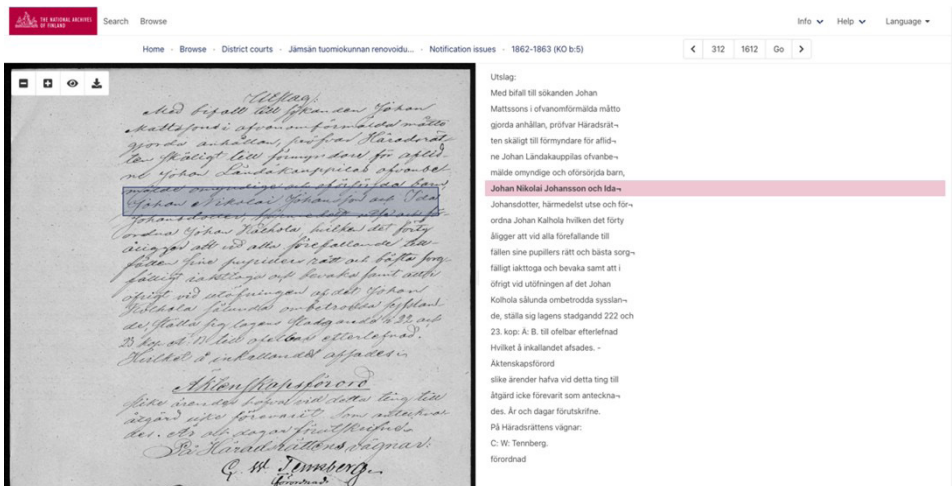


Figure 8: An example of a digital edition of the National Archives of Finland's document.

To do this, it involved students and scholars in training a model to obtain transcriptions with minimal margins of error. Detailed and in-depth training will enhance artificial intelligence tools, enabling Transkribus to automatically transcribe the entire human heritage in all languages and eras of the Past.

A further core feature of Transkribus is that digital transcription projects of archival heritages can be open to the entire community, who thus participate in the training phases, thus realising the grand goal of Public History.



## Bibliographic References

- Adamek, Tomasz, Noel E. O'Connor, and Alan F. Smeaton. 2007. "Word Matching Using Single Closed Contours for Indexing Handwritten Historical Documents." *International Journal of Document Analysis and Recognition* 9 (2): 153–65. <https://doi.org/10.1007/s10032-006-0024-y>.
- Albertin, Fauzia, Alessandra Patera, Iwan Jerjen, S. Hartmann, Eva Peccenini, Frédéric Kaplan, Marco F.M. Stampanoni, Rolf Kaufmann, and G. Margaritondo. 2016. "Virtual Reading of a Large Ancient Handwritten Science Book." *Microchemical Journal* 125 (March): 185–89. <https://doi.org/10.1016/j.microc.2015.11.024>.
- Alpaydin, Ethem. 2020. *Introduction to Machine Learning*. Massachusetts Institute of Technology.
- Archivi Storici e Biblioteche Istituto Suor Orsola Benincasa. n.d. Consultato il 10 febbraio 2023. <https://archivistoriciisob.transkribus.eu/>.
- Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. "Gephi - The Open Graph Viz Platform." <https://gephi.org/>.
- Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. New York, NY: Springer International Edition.
- Carducci, Giosuè. 1964. *Tutte Le Poesie, Juvenilia, Levia Gravia, A Satana, Giambi Ed Epodi, Intermezzo, Rime Nuove, Odi Barbare, Rime E Ritmi, Canzone Di Legnano*. Segrate: Rizzoli. <https://www.ibs.it/tutte-poesie-juvenilia-levia-gravia-libri-vintage-giosue-carducci/e/2560035385391>.
- Collobert, Ronan, and Jason Weston. 2008. "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning." In *Proceedings of the 25th International Conference on Machine Learning*, 160–67. ICML '08. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/1390156.1390177>.
- Dimond, Tom L. 1957. "Devices for Reading Handwritten Characters." In *Papers and Discussions Presented at the December 9-13, 1957, Eastern Joint Computer Conference: Computers with Deadlines to Meet*, 232–37. IRE-ACM-AIEE '57 (Eastern). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/1457720.1457765>.
- Dunley, Richard. 2018. "Machines Reading the Archive: Handwritten Text Recognition Software." The National Archives Blog. The National Archives. March 19, 2018. <https://blog.nationalarchives.gov.uk/machines-reading-the-archive-handwritten-text-recognition-software/>.



- Erwin, Brittany. 2020. "Digital Tools for Studying Empire: Transcription and Text Analysis with Transkribus." *Not Even Past*. November 6, 2020. <https://notevenpast.org/digital-tools-for-studying-empire-transcription-and-text-analysis-with-transkribus/>.
- Fondazione Banco di Napoli. n.d. Consultato il 10 febbraio 2023. <http://www.fondazionebanconapoli.it/en/>.
- Gori, Marco. 2018. *Machine Learning. A Constraint-Based Approach*. Burlington, Massachusetts: Morgan Kaufmann Publishers. <https://doi.org/10.1016/C2015-0-00237-4>.
- Kahle, Philip, Sebastian Colutto, Günter Hackl, and Günter Mühlberger. 2017. "Transkribus. A Service Platform for Transcription, Recognition and Retrieval of Historical Documents." In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 19–24. <https://doi.org/10.1109/ICDAR.2017.307>.
- Kaplan, Frédéric, and Isabella di Lenardo. 2017. "Big Data of the Past." *Frontiers in Digital Humanities* 4. <https://doi.org/10.3389/fdigh.2017.00012>.
- Massot, Marie-Laure, Arianna Sforzini, and Vincent Ventresque. 2018. "Transcribing Foucault's Handwriting with Transkribus." <https://hal.archives-ouvertes.fr/hal-01913435>.
- Midura, Rachel. 2020. "Italian Administrative Hands." Early Modern Digital Itineraries. July 21, 2020. <https://emdigit.org/tool/2020/07/21/italian-administrative-hands.html>.
- Milioni, Nikolina. 2020. "Automatic Transcription of Historical Documents. Transkribus as a Tool for Libraries, Archives and Scholars." PhD Diss., Uppsala Universitet. <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-412565>.
- Moretti, Giovanni, Rachele Sprugnoli, and Sara Tonelli. 2015. "Digging in the Dirt: Extracting Keyphrases from Texts with KD." In *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-It 2015*. <https://doi.org/10.4000/books.aaccademia.1518>.
- Muehlberger, Guenter, Louise Seaward, Melissa Terras, Sofia Ares Oliveira, Vicente Bosch, Maximilian Bryan, Sebastian Colutto, et al. 2019. "Transforming Scholarship in the Archives through Handwritten Text Recognition. Transkribus as a Case Study." *Journal of Documentation* 75 (5): 954–76. <https://doi.org/10.1108/JD-07-2018-0114>.
- Spina, Salvatore. 2022a. *Digital History. Metodologie Informatiche per La Ricerca Storica*. Napoli: Edizioni Scientifiche Italiane.

- Spina, Salvatore. 2022b. “Historical Network Analysis & Htr Tool. Per un approccio storico metodologico digitale all’archivio Biscari di Catania.” *Umanistica Digitale* 14: 163–81. <https://doi.org/10.6092/issn.2532-8816/15159>.
- Spina, Salvatore. 2023. *Biscari Epistolography*. Ministero della Cultura, MIC|MIC\_AS-CT|03/02/2023|0000173-P. <https://www.biscariepistolography.it/>.
- Turing, Alan Mathison. 1950. “Computing Machinery and Intelligence.” *Mind* LIX (236): 433–60. <https://doi.org/10.1093/mind/LIX.236.433>.
- Valacchi, Federico. 2021. *Gli archivi tra storia uso e futuro*. Milano: Editrice Bibliografica.

# Quelles sont les Fake News qui ont circulé lors de la pandémie de Covid-19 ?

Une étude menée sur Twitter sur la thématique des vaccins du  
01/02/2021 au 28/02/2021

Marc Tanti\*

**Abstract:** The Covid-19 pandemic has generated a wave of false information on social networks that has given rise to the concept of infodemic. Our article presents an analysis of publications concerning Covid-19 vaccines and immunization collected on Twitter during the period from February 01, 2021 to February 28, 2021. It highlights the resurgence of fake news.

*Keywords:* Covid-19, Vaccine, Twitter, Fake News, Mixed methodology.

## 1. Introduction

Même si ce concept est considéré comme récent, les *fake news*, informations sciemment conçues comme fausses, erronées ou fallacieuses afin d'emporter l'adhésion du récepteur ou du public, ont toujours existé (Giry 2017). Dans le domaine des régimes « miracles », les exemples sont nombreux. Dans le domaine de la vaccination, on peut par exemple citer l'idée largement répandue selon laquelle la vaccination Rougeole/Oreillons/Rubéole serait responsable de maladies comme l'autisme.

Le virus Sars-Cov2 a émergé en Chine dans le marché de Wuhan le 16 novembre 2019 et a entraîné l'épidémie de Covid-19 qui a été mondiale et sans précédent.

Avec l'avènement des Réseaux Sociaux Numériques (RSN), les données sur le virus et la maladie ont littéralement explosées. Dans cette pléthore d'informations, quelle part prennent les *fake news* ? Aussi, notre article a pour objectif principal d'analyser sur le média social particulier, Twitter, quelle part ce type d'information prend dans un corpus collecté durant la période allant du 01/02/2021 au 28/02/2021, en se focalisant sur une thématique, la vaccination et les vaccins. Notre objectif secondaire est de mener une discussion

---

\* Centre d'Epidémiologie et de Santé Publique des Armées CESPAM-UMR 1252-SESSTIM-INSERM-IRD-AMU, Marseille, France ; Institut Méditerranéen des Sciences de l'Information et de la Communication IMSIC-AMU, Marseille, France. mtanti@gmx.fr.

autour de ce concept dans le cadre de notre travail au regard des travaux antérieurs.

## 2. État de l'art

On retrouve dans la littérature, notamment internationale, de nombreux travaux qui ont étudié les *fake news* diffusées lors de l'épidémie de Covid-19.

Par exemple, on peut citer une lettre publiée par Shimizu dans "The Lancet" le 11 février 2020 (Shimizu 2020) où l'auteur indique que l'épidémie a entraîné au Japon l'émergence de *fake news* à l'origine d'une importante xénophobie contre les patients et visiteurs chinois. Il évoque notamment la propagation sur Twitter du mot-dièse #ChineseDontComeToJapan et des messages présentant les visiteurs chinois comme des personnes sales, insensibles et même bioterroristes (Shimizu 2020).

On peut aussi citer une revue systématique publiée récemment qui a analysé l'impact sur la santé de la diffusion sur les réseaux sociaux de ces fausses nouvelles (Rocha et al. 2021). Ce travail met en avant que les *fake news* diffusées sur Facebook, WhatsApp ou encore Twitter ont provoqué dans la population des troubles psychologiques, de la peur, des dépressions et de la fatigue. Conscients que ce phénomène constitue un véritable problème de santé publique, les gouvernements ont mis en œuvre des stratégies pour contenir la désinformation et les *fake news* pendant la pandémie (González-Padilla et Tortolero-Blanco 2020).

D'autres travaux ont tenté de modéliser les mécanismes de propagation de ces fausses nouvelles sur les RSN. On peut par exemple citer le travail d'Apuke et Omar qui a mis en évidence que l'altruisme, la socialisation et le partage d'informations étaient les facteurs les plus importants qui prédisaient le partage de *fake news* sur la Covid-19 dans les médias sociaux (Apuke et Omar 2021). Un autre travail met aussi en avant le manque de connaissances, le divertissement, le passe-temps et la faible confiance dans le gouvernement et les médias (Balakrishnan et al. 2022). Ce travail met en exergue que le phénomène de propagation était plus important chez les personnes peu instruites, les hommes et les jeunes. Les principales *fake news* relevées concernent les vaccins, le virus, les traitements et la prévention (Balakrishnan et al. 2022).

D'autres travaux ont tenté de dresser une typologie de ces *fake news* sur les RSN. On peut mettre en avant le travail de Monnier qui met en évidence une forte politisation des échanges informationnels autour de la Covid-19 (accusations entre États, etc.) et les clivages, politiques et idéologiques que l'épidémie a attisée (Monnier 2020). Ce travail met en avant la consolidation de la pratique du *fact-checking*, malgré les limites de celle-ci (Bigot 2019) et l'engagement de certains RSN qui ont décidé de s'attaquer frontalement au problème de la désinformation par la suppression de messages problématiques

(Nunes 2012). En plus d'agir contre cette désinformation nuisible, l'article de Schuetz rajoute que ces initiatives de *fact-checking* aident les usagers des RSN à identifier des informations précises et vérifiées dans ce flux torrentiel, notamment sur la façon de se protéger contre la maladie (Schuetz, Sykes, et Venkatesh 2021).

Dans le même cadre, on peut aussi citer l'article de Carrion-Alvarez et Tijerina-Salina qui met en avant les lobbies dans cette forte politisation des échanges, notamment anti-vaccinaux et qui utilisent les médias sociaux tels que YouTube pour diffuser des vidéos trompeuses ou falsifiées (Carrion-Alvarez et Tijerina-Salina 2020). Les auteurs soulignent dans la politisation des débats le rôle important des personnalités publiques tels que les célébrités, les présentateurs et les politiciens (Carrion-Alvarez et Tijerina-Salina 2020). Dans le même sens, plusieurs présidents comme Donald Trump aux États-Unis, Andres Manuel López Obrador au Mexique et Jair Bolsonaro au Brésil ont diffusé sur les RSN de nombreuses *fake news* (Carrion-Alvarez et Tijerina-Salina 2020). On peut citer l'injection sécurisée de désinfectants pour traiter la maladie voir même certains présidents ont nié la pandémie elle-même (Carrion-Alvarez et Tijerina-Salina 2020).

### 3. Approche méthodologique

Le présent article suit une méthodologie d'analyse mixte triangulée qualitative et quantitative, associant des études descriptives et lexicométriques (Fortin 2010 ; Creswell et Plano Clark 2011 ; Gallagher et Marceau 2020) et conduite sur le réseau social Twitter durant la période allant du 01/02/2021 au 28/02/2021.

#### 3.1. Choix méthodologique d'analyse sur Twitter

Pour rappel, les RSN sont des plateformes en ligne au sein desquels les utilisateurs (par exemple : une personne ou un groupe de personnes, une entreprise ou encore une structure publique) peuvent créer un profil et développer un réseau de contacts personnels qui leur permettent d'être connectés à d'autres utilisateurs (Rissoan 2011). Selon Cardon, les RSN ne sont pas de simples espaces publics où les utilisateurs racontent leur vie à tout le monde (Cardon 2019). Ce sont des espaces de visibilité paramétrables qui encouragent les internautes à créer des représentations de leur identité adaptées aux différents publics que les plateformes leur permettent de rassembler autour de leurs comptes (Cardon 2019). Les formats de visibilité variés offerts par les plateformes invitent les participants à mobiliser des facettes identitaires multiples, à

s'engager dans des interactions qui ne sont pas de même nature, à se lier à des réseaux sociaux composés selon des principes très différents (Cardon 2019).

Cardon dresse une typologie des RSN en quatre familles, selon le degré de visibilité que les réseaux donnent aux profils de leurs utilisateurs (Cardon 2019). Il distingue ainsi les familles de RSN paravent, clair-obscur, phare et mondes virtuels (Cardon 2019).

La première famille paravent n'est pas la plus importante. Mais elle joue un rôle particulier. S'y rangent les sites de rencontres où la visibilité des profils est organisée derrière un paravent. L'identité affichée est très réaliste et les critères retenus, objectifs et durables, sont ceux de l'identité civile (photo, âge, localisation, mensurations, revenus, etc.) mais on ne s'y découvre que progressivement (Cardon 2019).

La deuxième famille clair-obscur est la plus importante. Elle regroupe un ensemble de réseaux dont la propriété commune est de créer une visibilité en clair-obscur, une zone grise ou bien une visibilité privée-publique (Cardon 2019). Dans cette famille de services, les internautes s'exposent tout en se cachant. Ils affichent une identité narrative en racontant leur journée, en livrant leurs sentiments, leur vie avec les amis, leurs aventures de vacances, leurs démêlés avec les parents ou les professeurs. Avant que Facebook ne vienne s'imposer, chaque pays a développé son propre réseau en clair-obscur : Cyworld en Corée du Sud, Friendster aux États-Unis, Bebo au Royaume-Uni, Orkut au Brésil. Facebook est aujourd'hui le principal réseau en clair-obscur dans la plupart des pays ; il n'y a guère que dans les régimes autoritaires que la fonctionnalité du clair-obscur (se montrer tout en se cachant) est prise en charge par une autre plateforme, tels Vkontakte en Russie ou Wechat en Chine. En France, la visibilité en clair-obscur a d'abord été portée par Skyblog. Aujourd'hui, outre Facebook, elle s'observe notamment sur Snapchat et Whatsapp (Cardon 2019).

Une troisième famille de plateformes sociales, celle dite phare par Cardon, donne une visibilité beaucoup plus large au profil des participants (Cardon 2019). C'est Myspace qui a inventé ce format qualifié de phare, par opposition au clair-obscur : tout y est visible par tous. Cette famille de réseaux sociaux est apparue de façon inattendue à une époque, le début des années 2000, où il paraissait évident que les réseaux d'amis en ligne devaient rester privés. En 2003 Friendster était une plateforme en clair-obscur utilisée par les jeunes Américains, où l'on ne pouvait voir que les profils de ses amis. En général sur ce RSN, on n'y publie pas pour susciter la reconnaissance, l'amour, le rire de ses proches à propos d'événements ordinaires vécus dans la vie quotidienne. Au contraire, on y affiche une identité tendue vers un centre d'intérêt afin de pouvoir le partager avec des inconnus : la musique sur Myspace, les chaînes personnelles sur Youtube, les photos sur Flickr ou Pinterest, les activités professionnelles sur LinkedIn, les informations sur Twitter.

La quatrième famille de réseaux sociaux en ligne est celle des mondes virtuels : jeux vidéo à univers persistants tels World of Warcraft, mondes virtuels de type Second Life. Ici, les identités des participants sont publiques. En revanche elles ont été tellement façonnées, sculptées et fabriquées (pseudos, avatars, jeux avec l'identité) qu'elles dissimulent l'identité réelle des personnes (Cardon 2019).

Aussi, nous avons choisi de mener notre analyse sur le RSN Twitter, pour plusieurs raisons. Ce réseau informationnel créé en 2006 repose sur le principe du microblogage qui consiste à publier de courts messages, en l'espèce des unités informationnelles limitées à 280 caractères appelés Tweets (Perez 2021). Cela permet de commenter en temps réel les événements, personnels ou d'actualité (Kaplan et Haenlein 2011). La grande majorité des comptes Twitter sont publics permettant un accès facilité aux messages sans condition de suivi réciproque (Marwick et Boyd 2011) contrairement à Facebook où le contenu est généralement visible pour une liste d'amis approuvés (Waterloo et al. 2018). En effet, dans la troisième famille de RSN dits *phares* et dont fait partie Twitter, il existe une relation étroite entre ce que l'on montre de soi et la visibilité que la plateforme accorde aux publications des utilisateurs (Cardon 2019). Au vu de tous, les participants ne se connectent pas entre eux parce qu'ils se connaissent, mais parce qu'ils ont des goûts, des opinions en commun. Une tout autre logique se met en place lorsque la visibilité est ouverte : dans cette famille de réseaux sociaux c'est le partage de contenus qui domine (Cardon 2019), d'où nos choix portés sur ce RSN.

### 3.2. Corpus collecté

Le corpus collecté sur Twitter durant la période d'étude a été obtenu via le logiciel Radarly® commercialisé par Linkfluence (Lembert 2012). Ce logiciel permet de capter l'intégralité des messages originaux (Tweets et réponses à des Tweets), sans Retweets pour éviter le bruit, publiés sur le réseau social sélectionné correspondant à une requête soumise. Il donne également l'information sur le nombre d'« actions d'engagement », c'est-à-dire le nombre d'interactions avec le Tweet : Retweets, commentaires, likes, clics sur les liens, partages par compte, concernant le sujet. Nous avons, dans notre requête, tenté de nous rapprocher du langage naturel utilisé sur Twitter et avons opté pour une collecte en langue française. La requête utilisée a été la suivante :

((covid\* OR corona\* OR ncov OR sars\*) AND (vaccin\* OR COVID vaccination OR coronavaccin)) OR ((vaccin\* OR corona\* OR ncov OR sars\*) AND ((astra AND zeneca) OR astrazeneca OR AZ OR covishield OR moderna OR sinopharm OR cansino OR convidencia OR coronavac OR sinovac OR pfizer OR biontech OR cominarty OR cominaty OR sputnik OR sputnik OR covivac OR kovivac OR epivaccorona OR covaxin OR janssen OR johnson OR « J&J »))



Nous avons ainsi récolté 109 681 Tweets bruts. Après suppression des doublons, le corpus a été constitué de 101 615 Tweets. Les variables d'intérêt pour ce corpus sont :

- la date de publication du Tweet ;
- le nom du compte ayant publié ;
- le pays d'origine du compte ayant publié ;
- les actions d'engagement précitées ;
- le texte du Tweet.

### 3.3. Analyses menées

#### 3.3.1. Analyses descriptives

Elles ont permis d'identifier quelques caractéristiques sur les sources utilisées afin de contextualiser les données collectées. Ainsi, des tableaux croisés dynamiques réalisés à partir du corpus ont permis d'obtenir trois types d'informations :

- le nombre de publications quotidiennes durant le mois de février 2021 qui permet notamment l'identification des pics informationnels ;
- la répartition géographique de l'information, par pays, cartographiée via QGIS, un logiciel libre de système d'information géographique. Nous nous sommes notamment basés sur l'information de localisation des comptes fournie par Radarly ;
- les 10 acteurs communiquant sur la thématique les plus actifs. Il s'agissait notamment des comptes Twitter ayant le plus publié sur la vaccination contre la Covid-19 durant la période d'étude. Une analyse complémentaire a été réalisée afin d'identifier les 10 comptes ayant suscité le plus d'actions d'engagement.

#### 3.3.2. Analyses lexicométriques

Une analyse textuelle a été conduite afin d'identifier les thématiques traitées dans le corpus. Cette analyse a été menée à partir du logiciel IRaMuTeQ version 0.7 alpha 2 (Ratinaud 2014) qui est une implémentation libre de la méthode Alceste de Max Reinert (Reinert 1990). IRaMuTeQ permet l'analyse statistique de données textuelles en soumettant des scripts au logiciel R et en affichant les résultats dans son interface graphique. Bien que le logiciel propose certaines options de modifications des textes, telles que leur passage en minuscule ou la lemmatisation, il a fallu cependant compléter par un prétraitement d'harmonisation des données. Pour extraire les thématiques traitées et avoir des informations sur leur poids dans le corpus, nous avons appliqué la

classification hiérarchique descendante (CHD), dite Méthode Reinert. Cette méthode, initialement utilisée pour des corpus de taille réduite, est également applicable à de gros corpus, comme dans notre cas (Ratinaud et Marchand 2012).

Ayant la particularité d'avoir des textes courts (Tweets), nous avons opté pour une CHD simple sur textes (Tweets) entiers. Afin d'identifier les thématiques auxquelles se rapporte chaque classe, nous avons effectué des allers-retours entre les formes significatives listées dans le profil de chaque classe et le texte, via le concordancier dans les segments de texte/textes de la classe qui affiche le contexte d'utilisation du vocabulaire spécifique. Nous avons donc décrit le profil de chaque classe lexicale tout en proposant un segment de texte (Tweet pour notre corpus) significatif. La CHD nous a donné des informations concernant le Chi2 d'association entre les classes et les modalités des variables telles que l'auteur de l'information.

Nous avons également effectué une analyse manuelle des contenus dans les thématiques retrouvées (Negura 2006). Pour ancrer de manière théorique cette approche, le terme « contenu » désigne l'information transmise à travers le discours (Maingueneau 2009). L'école de Palo Alto l'oppose au terme « relation ». On distingue ainsi deux types de contenu : le contenu manifeste et le contenu latent, empruntés au domaine de la psychologie freudienne. Le contenu manifeste est ce qui est explicitement exprimé : opinions, croyances... En ce sens, étudier le contenu d'un discours consiste à faire ressortir son contenu manifeste : les thèmes les plus souvent abordés, les mots clés, les prises de position et les arguments invoqués pour les justifier (Clauzard 2017). Dans ce cadre, selon Albig, l'analyse de contenu est « une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste des communications, ayant pour but de les interpréter » (Albig 1952). Selon Mucchielli, analyser le contenu d'un document ou d'une communication, « c'est rechercher les informations qui s'y trouvent, dégager le sens ou les sens de ce qui y est présenté, formuler, classer tout ce que contient ce document ou cette communication » (Mucchielli 2006). Pour Mucchielli, dans « l'analyse de contenu, l'essentiel est le sens et non la forme ». Le tableau 1 synthétise les étapes de la méthodologie appliquée (Tab. 1).

Langue	Nombre de Tweets	Analyses descriptives	Analyses thématiques
Français	101 615	<ul style="list-style-type: none"> <li>- Nombre de publications quotidiennes</li> <li>- Provenance géographique</li> <li>- Acteurs les plus actifs</li> </ul>	<ul style="list-style-type: none"> <li>-Identification de classes (classification hiérarchique descendante)</li> <li>-Analyse manuelle de contenus</li> </ul>

Tableau 1 : Récapitulatif du corpus collecté et méthodologie d'analyse appliquée.

## 4. Résultats

### 4.1. Dynamique spatio-temporelle et principaux acteurs de l'information

Nos résultats permettent de retrouver 5 pics majeurs de publications (2, 8, 12, 18 et 25 février) et un pic intermédiaire (5 février), l'ensemble représentant 27,78% du corpus. A contrario, on retrouve 5 périodes plus creuses (1, 7, 14, 21, 28) (Fig. 1).



Figure 1 : Répartition par jour de février 2021 des 101 615 Tweets du corpus.

Les données sur la provenance géographique des Tweets sont disponibles pour 73,4% des Tweets du corpus (Fig. 2) et sont dépendantes, en partie, des informations fournies par les personnes twittant. Le trio des pays en tête sont la France, le Canada et la Belgique et correspondent à l'origine de 82,7% des publications du corpus dont 73,3% pour la France. Les Tweets proviennent à 89,3% de pays dont le français est la langue officielle et à 91,8% de pays membres de l'OIF. Ils émanent de cinq continents : l'Europe (81,8%), l'Amérique (8,6%), l'Afrique (7,2%), l'Asie (2,2%) et l'Océanie (0,2%).

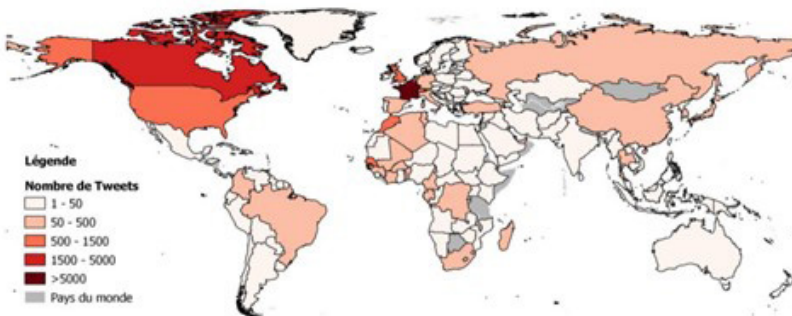


Figure 2 : Répartition par pays des Tweets du corpus.

Parmi les dix comptes Twitter du corpus ayant le plus publié sur la thématique de la vaccination contre la Covid-19 en février 2021 (Fig. 3), seuls deux sont certifiés par Twitter : la chaîne d'informations française BFMTV

et le groupe de presse belge francophone Sudpresse. Trois autres comptes se présentent comme étant des sites d'informations et d'actualités (Mxit Infos, FR24 News France, W38777Y équivalent à Wnews) ou un site agréant les principales sources de la presse en France, en Suisse, au Canada et au Luxembourg (Titrespresse.com). Les quatre comptes restants sont des comptes personnels dont celui d'un médecin généraliste (Dr Jean-Jacques Fraslin). Ces 10 comptes, parmi plus de 36 000, représentent 5% de l'ensemble des Tweets publiés.

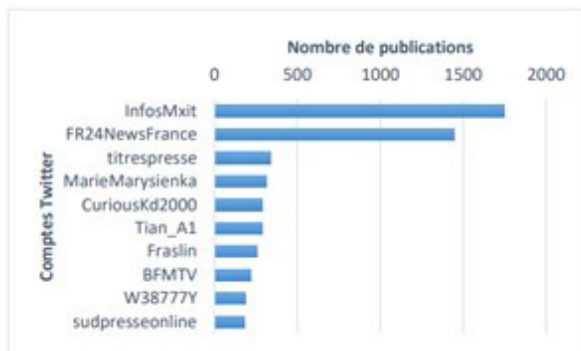


Figure 3 : Représentation graphique des 10 comptes Twitter ayant suscité le plus d'actions d'engagement sur le sujet en février 2021.

On peut noter dans nos résultats que ce ne sont pas ces comptes qui ont suscité le plus d'actions d'engagement sur le mois de février 2021 (Fig. 4). Nous retrouvons deux comptes certifiés : ceux de l'AFP (Agence France Presse) et de la députée Martine Wonner. Parmi les autres comptes, deux proposent des brèves sur l'actualité : Conflits France et Mediavenir. Trois ont un lien avec le monde de la santé : Alexandra Henrion-Caude, généticienne et directrice de l'Institut de Recherche SimplissimA; le collectif de médecins Laissons les prescrire et le compte papanours60 qui se présente comme professeur agrégé d'université en immunopathologie.

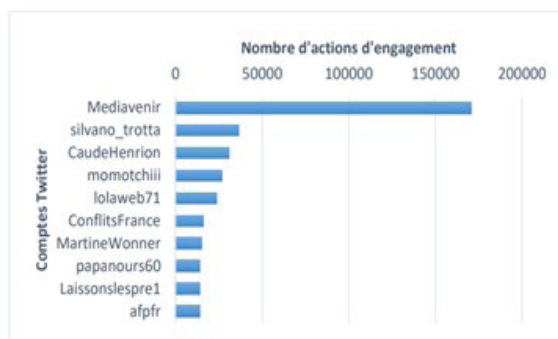


Figure 4 : Représentation graphique des 10 comptes Twitter ayant suscité le plus d'actions d'engagement sur le sujet en février 2021.

## 4.2. Thématiques traitées et contenus

La CHD de l'analyse lexicométrique aboutit sur 11 mondes lexicaux dans lesquels 81,1% des Tweets sont classés. Le poids de chaque classe s'échelonne de 4 à 14,8% des Tweets classés (Fig. 5). On retrouve deux branches du corpus.

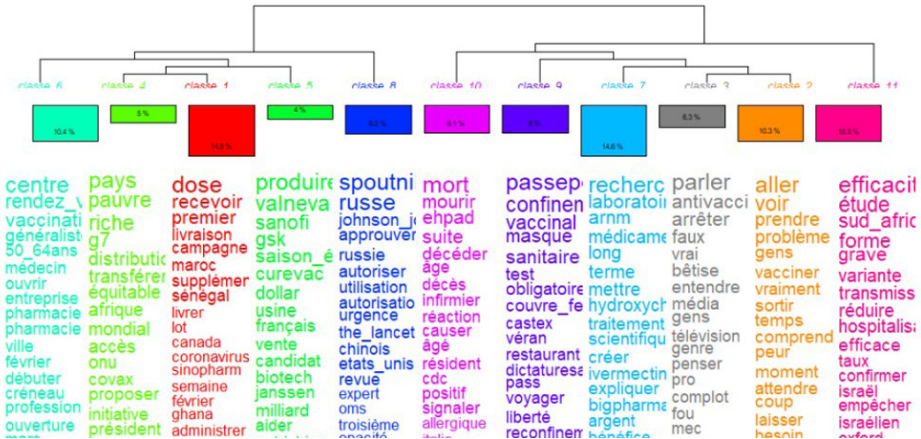


Figure 5 : Dendrogramme du corpus Twitter avec le vocabulaire spécifique associé à chaque classe (IRaMuTeQ).

La première branche du corpus (42,4% des Tweets classés), celle de droite, traite principalement de la disponibilité et de l'accessibilité des vaccins.

– Le corpus renvoie ainsi aux autorisations sanitaires de mise sur le marché : classe 8 (8,2% des Tweets classés) associée notamment à Wnews.

– Il fait également référence à la dimension industrielle des vaccins et aux capacités de production : classe 5 (4%) liée, par exemple, à Mxit Infos.

Cette branche comprend également la dimension politique des vaccins (19,8%) avec :

– l'accès équitable aux vaccins: classe 4 (5%) principalement associée à l'AFP ;

– la mise en œuvre de cet accès au niveau mondial avec le suivi des livraisons et des campagnes de vaccination: classe 1 (14,8%) liée notamment à FR24 News France ;

– enfin, cette première branche contient des éléments stratégiques et opérationnels sur les campagnes de vaccination au niveau local: classe 6 (10,4%) associée, par exemple, à Mxit Infos.

Dans ce corpus, notre analyse de contenus retrouve principalement des données factuelles, des décisions politiques et stratégiques relayées dans et par les médias.

On retrouve une deuxième branche de corpus qui comprend des sujets de scepticisme par rapport aux vaccins et à la vaccination (57,6% des Tweets

classés). C'est dans ce corpus que notre analyse de contenus met en évidence des *fake news* sur les vaccins et la vaccination.

Dans ce corpus, on retrouve :

- la classe 3 (3,6%) qui met en avant l'existence d'une opposition d'opinions autour du vaccin lui-même ;
- la classe 2 (10,3%) qui renvoie de manière générale à l'hésitation vaccinale avec les questionnements et les raisons liés à la décision de se faire vacciner ou non, notamment en lien avec le risque vaccinal.

Ces sujets de scepticisme sont notamment liés à :

- l'efficacité des vaccins, notamment en présence de variants du virus: classe 11 (10,3%) liée, par exemple, au compte du Dr Frasin ;
- la pertinence des vaccins qui sont opposés, en termes d'utilité et de dangerosité, aux traitements anti-Covid19: classe 7 (14,6%) associée notamment à Wnews ;
- les « mesures » de lutte contre la Covid-19 dans lesquelles est intégrée la vaccination et qui sont considérées comme une atteinte aux libertés fondamentales: classe 9 (8%) ;
- la sécurité des vaccins avec des effets délétères sur la santé qui sont mis en avant dans la classe 10 (8,1%), notamment liée aux comptes momotchiii et lolaweb71.

Le Tableau 2 synthétise les principaux résultats retrouvés (Tab. 2).

5 pics d'informations	3 principaux pays d'origine	20 comptes influenceurs	11 thématiques principales
2, 5, 8, 12, 18 et 25 février	<ul style="list-style-type: none"> <li>• France</li> <li>• Canada</li> <li>• Belgique</li> </ul>	<ul style="list-style-type: none"> <li>• Comptes officiels (BFMTV, Sudpresse, AFP, MartineWonner)</li> <li>• Comptes non officiels d'informations et d'actualités (Mxit Infos, FR24 News France, Wnews, Titrespresse.com, Conflits France et Mediavenir)</li> <li>• Collectif de médecins (Laissons les prescrire)</li> <li>• Comptes personnels (dont 1/3 a un lien avec le monde de la santé)</li> </ul>	<ul style="list-style-type: none"> <li>• Suivi mondial des campagnes vaccinales</li> <li>• Pertinence des vaccins versus alternatives</li> <li>• Éléments stratégiques et opérationnels des campagnes de vaccination</li> <li>• Questionnements généraux liés à la décision vaccinale et raisons du choix effectué</li> <li>• Efficacité vaccinale et impacts des variantes</li> <li>• Autorisations sanitaires de mise sur le marché Sécurité des vaccins et effets sanitaires délétères</li> <li>• Dispositif de lutte contre la Covid-19 et atteintes aux libertés fondamentales</li> <li>• Accès équitable aux vaccins dans le monde</li> <li>• Industrialisation des vaccins</li> <li>• Existence d'une opposition d'opinions</li> </ul>

Tableau 2 : Principaux résultats retrouvés à partir de l'analyse lexicométrique.



On peut par exemple citer dans ce corpus, la propagation d'une *fake news* affirmant le décès de Mauricette âgée de 78 ans après avoir été la première française vaccinée contre la Covid-19 et qui a été hyper-médiatisée à la télé. Dans les Tweets relayés par les comptes personnels, mais aussi par les comptes officiels et non officiels d'informations, la vieille dame serait décédée peu après la vaccination d'une crise cardiaque (Fig. 6).



Figure 6 : Fake News autour du décès de Mauricette après avoir été la première française vaccinée.

Cette *fake news* a, par la suite, été démentie par des initiatives individuelles et collectives de Fact-Checking (Fig. 7).

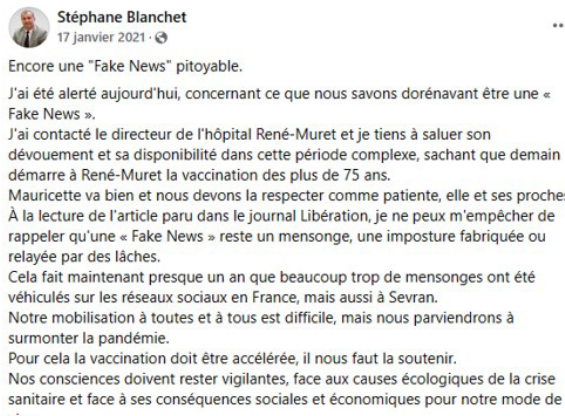


Figure 7 : Démenti de la *fake news* autour du décès de Mauricette.

Mais cette *fake news* a suscité à son tour une controverse autour du vaccin et de la vaccination (Fig. 8).





Figure 8 : Controverse autour de la vaccination générée par la *fake news*.

Elle a également entraîné la circulation de théories du complot (Fig. 9).

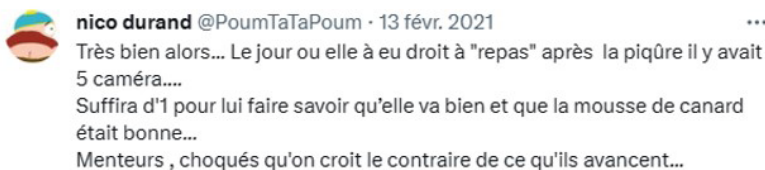


Figure 9 : Théories du complot autour de la *fake news*.

## 5. Discussions

Nos travaux retrouvent durant la période d'étude allant du 01 au 28 février 2021, après une analyse lexicométrique sur 81,1% des Tweets classés parmi 101 615 Tweets collectés, 57,6% de Tweets comprenant des sujets de scepticisme par rapport aux vaccins et à la vaccination et dans lequel une analyse de contenus met en évidence des *fake news* sur les vaccins et la vaccination. Les sujets relevant de ces types d'informations traitent notamment de la pertinence des vaccins versus d'autres alternatives, de questionnements liés à la décision vaccinale et l'efficacité des vaccins, la mise sur le marché, la sécurité de ces produits, des effets sanitaires délétères ou encore de l'industrialisation. On relève notamment une opposition d'opinions sur ces thématiques.

Comme premier point de discussion, ces résultats corroborent les résultats retrouvés dans la littérature internationale, notamment les travaux de Balakrishnan et al. cités dans notre état de l'art et qui mettent en avant que les principales *fake news* retrouvées durant la pandémie concernent les vaccins, le virus, les traitements et la prévention (Balakrishnan et al. 2022).

Comme autre point de discussion, il nous paraît légitime de retrouver dans notre travail et dans le cadre de cette pandémie, la diffusion de *fake news* sur les vaccins et la vaccination dans le corpus collecté durant la période d'étude. En effet, à la date d'analyse, sont pratiquées les premières vaccinations avec peu de recul sur les vaccins à ARN messenger administrés, l'existence de beaucoup de doutes et d'incertitudes.

De plus, dans d'autres épidémies récentes, on retrouve la diffusion de *fake news* dans les médias sociaux dont Twitter sur les vaccins et traitements. On

peut par exemple citer lors de l'épidémie d'Ebola 2014, la diffusion de *fake news* autour des traitements administrés (Jin et al. 2014). Il faut s'interroger sur les motivations de cette diffusion. Nos travaux mettent en évidence leur relai via des comptes personnels qui n'ont pas forcément de liens avec la santé, mais aussi via des comptes officiels de sources d'informations ou d'actualités (BFMTV, Sudpresse, AFP, MartineWonner) ou encore non officiels (Mxit Infos, FR24 News France, Wnews, Titrespresse.com, Conflits France et Mediavenir). Quelle est la motivation de ces relais ? Quelle est le circuit de diffusion ?

Comme autre point de discussion, les *fake news* ont été popularisées par Donald Trump durant la campagne électorale de 2016 pour qualifier ses opposants et également les journaux qui lui sont critiques, la formule de « fake news » – fausses nouvelles – s'est diffusée au travers des réseaux sociaux et des médias à partir de l'année 2017, pour atteindre un pic d'attention en 2018 (Allard-Huver n.d.). Même si l'expression « fake news » est celle qui est plus retrouvée dans le vocabulaire courant, il est intéressant de se pencher sur sa traduction officielle décidée par la Commission d'enrichissement de la langue française : *infox* (Commission d'enrichissement de la langue française 2018). Ainsi, le Journal Officiel, précise la définition de ce nouveau phénomène comme : « information mensongère ou délibérément biaisée, répandue par exemple pour favoriser un parti politique au détriment d'un autre, pour entacher la réputation d'une personnalité ou d'une entreprise, ou encore pour contredire une vérité scientifique établie » (Commission d'enrichissement de la langue française 2018). Même si les concepts sont liés, il est important aussi de souligner la différence entre rumeur et *fake news*. La rumeur s'avère un mode de régulation du social dont l'énoncé peut être véridique. Les *fake news* sont, quant à elles, des artefacts, des fausses nouvelles, dont la nature même est d'être fausse, en tout ou partie, quand bien même elles prétendent être exactes et peuvent prendre les apparences symboliques de la « vérité » (Peyron-Bonjan 2018) ou des codes d'énonciation des médias dominants (Jeanret et Souchier 2005). L'expression anglo-saxonne « fake news » précise bien cette notion de « fausseté de l'information » de manière délibérée. L'objectif étant la création et le partage délibéré d'informations fausses (Wardle 2017). Par ailleurs, la durée de vie des rumeurs est généralement relativement courte (Reumaux 1990). En revanche, la durée de vie des *fake news* est mal quantifiée et difficilement évaluable. Toutefois, il semble que par leur caractère instrumental, elles puissent résister davantage encore que les rumeurs aux démentis (Kapferer 1990). Aussi, nos analyses mettent en évidence dans les Tweets relevés, ce caractère délibérément faux de fausses nouvelles, les différenciant de la rumeur dont l'énoncé peut être véridique.

Nos travaux corroborent les données de la littérature internationale, notamment ils mettent en avant cette forte politisation des échanges et ces cli-

vages idéologiques que l'épidémie a attisé (Monnier 2020). Mais ils mettent aussi en évidence que les *fake news* sont loin de se limiter au seul domaine politique (Allard-Huver n.d. ; 2019). Elles contaminent l'ensemble de l'espace public, s'attaquant tout autant à la réputation des individus ou des marques, qu'aux faits scientifiques, dans notre cadre d'étude (Allard-Huver 2019). Mais pour revenir à nos questionnements, pour quels objectifs ? Au profit de quels intérêts, dans le cadre de notre étude ? Des réponses peuvent venir des lobbies en lien avec la défiance et l'hésitation vaccinale. Ce qui corroborerait les résultats retrouvés dans la littérature internationale sur la Covid-19 qui mettent en avant le rôle des lobbies dans cette forte politisation des échanges sur les RSN (Carrion-Alvarez 2020 et Tijerina-Salina). Ce phénomène de lobbies anti-vaccination dans politisation de débats en santé n'est pas nouveau. Dans la littérature, on retrouve, le concept de lobbies anti-vaccin dans d'autres débats de santé, par exemple dans le cadre de la vaccination anti-HPV (Human Papilloma Virus) (Dubé et al. 2021). On peut également citer en 2009–2010 dans le cadre de la campagne de vaccination contre la grippe A (H1N1) qui s'est soldée par un échec cuisant avec seulement 8% de la population vaccinée pour un objectif de couverture de 70% de la population (Ward, Guille-Escuret, et Alapetite 2019).

Nos travaux mettent aussi en évidence, durant notre période d'étude, le développement d'initiatives de Fact checking qui consiste à vérifier la véracité de propos tenus par des responsables politiques ou d'autres personnalités, au sein de tribunes offertes à ces personnalités (interviews radio ou TV, meetings, médias sociaux etc.) (Bigot 2017). Cet exercice journalistique semble nouveau. Mais en réalité, il revisite largement une autre version, plus ancienne, du *fact-checking*, née aux États-Unis. C'est en effet généralement aux années 1920 que l'on fait remonter l'histoire du *fact-checking*, lorsque le magazine "Time" recrutait, à ses débuts, les premiers *fact-checkers* de la presse magazine américaine, bientôt suivi par "The New Yorker", puis par l'ensemble des autres titres (Cornu 1994). L'appellation *fact-checking* de ces dernières années a ainsi opéré un glissement pour désigner, non plus la vérification exhaustive et systématique des contenus journalistiques a priori, mais le contrôle ponctuel de citations publiques a posteriori. Notre travail sur ce point va dans le sens des données de littérature. En effet, selon l'AFP, les publications associées au coronavirus sur les médias sociaux n'ont cessé de prendre de l'ampleur. L'activité de Fact-checking initiée par cette agence, qui se concentre sur les rumeurs les plus virales diffusées sur Facebook ou Twitter, fait ainsi état d'une augmentation considérable du nombre de données analysées avec 60 rumeurs pour le seul mois de mars 2020 et un pic relevé le 18 mars. Entre le 27 janvier 2020 et le 16 avril 2020, 115 rumeurs ont été analysés par l'AFP, l'écrasante majorité de ces contenus s'avèrent faux ou trompeurs (AFP 2020). Plus globalement, ce résultat corrobore les données de la littérature qui mettent en avant la conso-

lvalidation des pratiques de *fact-checking* durant cette épidémie (Monnier 2020 ; Schuetz, Sykes, et Venkatesh 2021).

En conclusion, on peut faire le même constat pour les épidémies récentes. Des *fake news* ont circulé sur les RSN. Par exemple, lors de la dernière grande épidémie d’Ebola 2014–2015 en Afrique de l’Ouest, de nombreuses *fake news* ont circulé sur Twitter et ont fait l’objet d’études. Celle de Jin et al. a notamment mis en évidence la circulation de *fake news* liées au traitement de la maladie (Jin et al. 2014).

On peut noter en conclusion générale que cette vague de *fake news* sur les réseaux sociaux a donné naissance le 02 février 2020 au concept d’infodémie par Tedros Adhanom Ghebreyesus, directeur général de l’OMS (cité par Mesquita 2020).

## 6. Conclusions

Nos travaux retrouvent durant la période d’étude allant du 01 au 28 février 2021, après une analyse lexicométrique sur 81,1% des Tweets classés parmi 101 615 Tweets collectés, 57,6% de tweets comprenant des sujets de scepticisme par rapport aux vaccins et à la vaccination et dans lequel une analyse de contenus met en évidence des *fake news* sur les vaccins et la vaccination. Les sujets relevant de ce type d’information traitent notamment de la pertinence des vaccins versus d’autres alternatives, de questionnements liés à la décision vaccinale et l’efficacité des vaccins, la mise sur le marché, la sécurité, les effets sanitaires délétères ou encore l’industrialisation. On relève notamment une opposition d’opinions sur ces thématiques. Aussi, nos travaux corroborent en grande partie les données retrouvées dans la littérature dans le cadre des grandes épidémies. On peut par exemple citer celle d’Ebola 2014–2015 qui a vu la diffusion de nombreuses *fake news* sur les traitements de la maladie sur le réseau Twitter (Jin et al. 2014). Nos travaux souffrent de nombreuses limites, notamment d’une période d’étude courte d’un mois. Ils peuvent se compléter par des analyses quantitatives pour déterminer les quantités de Tweets relevant de *fake news*. Ils peuvent aussi être complétés par l’étude des mécanismes de leur diffusion sur le réseau Twitter, notamment à partir des grands relais identifiés dans notre travail.

## Références bibliographiques

AFP – Agence France Presse. 2020. “Covid-19 : retrouvez ici tous nos articles de vérification.” 17 février 2020. <https://factuel.afp.com/le-coronavirus-les-verifications-faites-par-lafp>.

- Albig, William. 1952. "BERELSON, BERNARD. Content Analysis in Communication Research. Pp. 220. Glencoe, Ill.: The Free Press, 1952. \$3.50". *The ANNALS of the American Academy of Political and Social Science* 283 (1): 197–98. <https://doi.org/10.1177/000271625228300135>.
- Allard-Huver, François. n.d. "Fake news, post-vérité & infox." [http://allardhuver.fr/publications/Fake News/](http://allardhuver.fr/publications/Fake%20News/).
- Allard-Huver, François. 2019. "Fake news." In *Comprendre la culture numérique*, sous la direction de Pauline Escande-Gauquié et Bertrand Naivin, 58–66. Dunod.
- Apuke, Oberiri Destiny, and Bahiyah Omar. 2021. "Fake news and COVID-19: modelling the predictors of fake news sharing among social media users." *Telematics and Informatics* 56: 101475. <https://doi.org/10.1016/j.tele.2020.101475>.
- Balakrishnan, Vimala, Wei Zhen Ng, Mun Chong Soo, Gan Joo Han, and Choon Jiat Lee. 2022. "Infodemic and fake news – A comprehensive overview of its global magnitude during the COVID-19 pandemic in 2021: A scoping review." *International Journal of Disaster Risk Reduction* 78: 103144. <https://doi.org/10.1016/j.ijdr.2022.103144>.
- Bigot, Laurent. 2017. "Le fact-checking ou la réinvention d'une pratique de vérification." *Communication & langages* 2 (192) : 131–56.
- Bigot, Laurent. 2019. *Fact-checking vs fake news. Vérifier pour mieux informer*. Paris : INA.
- Cardon, Dominique. 2019. "Typologie des réseaux sociaux en ligne." In *Culture numérique*, sous la direction de Dominique Cardon, 152–63. Paris : Presses de Sciences Po.
- Carrion-Alvarez, Diego, and Perla X. Tijerina-Salina. 2020. "Fake news in COVID-19: A perspective." *Health Promot Perspect* 10 (4) : 290–91.
- Clauzard, Philippe. 2017. *Séminaire d'initiation méthodologique à la Recherche*. Université de la Réunion. ESPE. <https://fr.slideshare.net/philip61/seminaire-methodorechercheclauzard>.
- Commission d'enrichissement de la langue française. 2018. "Recommandation sur les Équivalents Français à Donner à l'expression Fake News." Texte 113. *Journal Officiel de la République française*, no. 0029 du 04 octobre 2018.
- Cornu, Daniel. 1994. *Journalisme et vérité*. Genève : Labor et Fides.
- Creswell, John W., and Vicki L. Plano Clark. 2011. *Designing and Conducting Mixed Methods Research (2nd Ed.)*. Thousand Oaks, CA: Sage. SAGE Publications Inc.

- Dubé, Ève, Jeremy K. Ward, Pierre Verger, and Noni E. MacDonald. 2021. "Vaccine hesitancy, acceptance, and anti-vaccination: trends and future prospects for public health." *Annual Review of Public Health*, 42: 175–91.
- Fortin, Marie-Fabienne. 2010. *Fondements et Étapes du Processus de Recherche : Méthodes Quantitatives et Qualitatives*. Montréal, Canada : Chenelière Éducation.
- Gallagher, Frances, et Mélanie Marceau. 2020. "La recherche descriptive interprétative - Exploration du Concept de la Validité en tant qu'impératif social dans le contexte de l'évaluation des Apprentissages en Pédagogie des Sciences de la Santé." In *Méthodes Qualitatives, Quantitatives et Mixtes, 2e Édition : Dans la Recherche en Sciences Humaines, Sociales et de la Santé*, sous la direction de Marc Corbière et Nadine Larivière, 5–32. Québec : Presses de l'Université du Québec.
- Giry, Julien. 2017. "Étudier les théories du complot en sciences sociales : enjeux et usages." *Quaderni* 94 (3) : 5–11.
- González-Padilla, Daniel A., and Leonardo Tortolero-Blanco. 2020. "Social media influence in the COVID-19 Pandemic." *International Braz J Urol* 46: 120–24. <https://doi.org/10.1590/S1677-5538.IBJU.2020.S121>.
- Jeanneret, Yves, et Emmanuel Souchier. 2005. "L'énonciation éditoriale dans les écrits d'écran." *Communication et langages* 145 : 3–15.
- Jin, Fang, Wei Wang, Liang Zhao, Edward R. Dougherty, Yang Cao, Chang-Tien Lu, and Naren Ramakrishnan. "Misinformation Propagation in the Age of Twitter." *IEEE Computer* 47(12): 90–94.
- Kapferer, Jean-Noel. 1990. *La Rumeur : Le plus vieux media du monde*. Paris : Éditions Le Seuil.
- Kaplan, Andreas, and Michael Haenlein. 2011. "The early bird catches the news: Nine things you should know about micro-blogging." *Business Horizon* 54: 105–13. <https://doi.org/10.1016/j.bushor.2010.09.004>.
- Lembert, Sylvain. 2012. "Radarly : un outil pour gérer l'e-reputation et le Community Management d'une marque." *Webmarketing.com*. 28 septembre 2012. <https://www.webmarketing-com.com/2012/09/28/15785-radarly-un-outil-pour-gerer-le-reputation-et-le-community-management-dune-marque>.
- Mangueneau, Dominique. 2009. *Les termes clés de l'analyse du discours*. Paris : Éditions du Seuil.
- Marwick, Alice E., and Danah Boyd. 2011. "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience." *New Media & Society* 13: 114–33.



- Mesquita, Claudio T., Anderson Oliveira, Flávio Luiz Seixas, and Aline Paes. 2020. "Brazil - Infodemia, Fake News and Medicine: Science and The Quest for Truth." *International Journal of Cardiovascular Sciences* 33 (3): 203–05. <https://doi.org/10.36660/ijcs.20200073>.
- Monnier, Angeliki. 2020. "Covid-19 : de la pandémie à l'infodémie et la chasse aux fake news." *Recherches & éducations, Quelle éducation avec la Covid-19*. <https://doi.org/10.4000/rechercheseducations.9898>.
- Mucchielli, Roger. 2006. *L'analyse de contenu. Des documents et des communications*. Montrouge : Éditions E.S.F.- Entreprise moderne d'édition.
- Negura, Lilian. 2006. "L'analyse de contenu dans l'étude des représentations sociales." *SociologieS* 1 : 1–21. <https://doi.org/10.4000/sociologies.993>.
- Nunes, Joao. 2012. "Health, Politics and Security." *e-cadernos CES* 15:142–64.
- Perez Sarah. 2021. "Twitter officially expands its character count to 280 starting today." *TechCrunch*. 7 novembre 2017. <https://techcrunch.com/2017/11/07/twitter-officially-expands-its-character-count-to-280-starting-today/>.
- Peyron-Bonjan, Christiane. 2018. "De l'inexorabilité des fake news dans le monde contemporain." In *L'Information d'actualité au prisme des fake news*, sous la direction de Alexandre Joux et Maud Pélissier, 23-45. Paris : Éditions L'Harmattan.
- Ratinaud, Pierre. 2014. "Iramuteq." V. 07 alpha 2. <http://www.iramuteq.org/>.
- Ratinaud, Pierre, et Pascal Marchand. 2012. "Application de la méthode ALCESTE aux « gros » corpus et stabilité des « mondes lexicaux » : analyse du « CableGate » avec IRAMUTEQ." In *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles (JADT) 2012*, 835–44.
- Reinert, Max. 1990. "Alceste une méthodologie d'analyse des données textuelles et une application : Aurelia De Gerard De Nerval." *Bulletin de méthodologie Sociologique* 26 : 24–54.
- Reumaux, Françoise. 1990. "Traits invariants de la rumeur." *Communication* 52 : 144–159.
- Rissoan, Romain. 2011. *Les réseaux sociaux : Facebook, Twitter, LinkedIn, Viadeo, Google+ : Comprendre et maîtriser ces nouveaux outils de communication*. Saint-Herblain : Eni Editions.
- Rocha, Yasmim M., Gabriel Acácio de Moura, Gabriel Alves Desidério, Carlos H. de Oliveira, Francisco D. Lourenço, and Nicolete Larissa de Figueiredo. 2021. "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review." *Journal of Public Health* 9: 1–10.



- Schuetz, Sebastian W., Tracy Ann Sykes, and Viswanath Venkatesh. 2021. "Combating COVID-19 fake news on social media through fact checking: antecedents and consequences." *European Journal of Information Systems* 30 (4): 376–88.
- Shimizu, Kazuki. 2020. "2019-nCoV, fake news, and racism." *The Lancet* 395 (10225) : 685–86.
- Ward, Jérémy, Paul Guille-Escuret, et Clément Alapetite. 2019. "Les « antivaccins », figure de l'anti-Science." *Déviance et Société* 2 (43): 221–51.
- Wardle, Claire. 2017. "Fake News – La complexité de la désinformation." *First Draft*. 17 mai 2017. <https://firstdraftnews.org/articles/Fake-News-la-complexite-de-la-desinformation-2/>.
- Waterloo, Sophie F., Susanne E. Baumgartner, Jochen Peter, and Patti M. Valkenburg. 2018. "Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp." *New Media & Society* 20 (5): 1813–831.

# Se l'archivio è artificiale. Verso uno *ius archivi* partecipativo?

Federico Valacchi\*

«Mi spiace Dave.  
Purtroppo non posso farlo».  
(HAL 9000)

**Abstract:** The article, starting from consolidated methodological observations, aims to evaluate the concept of digital archive in a critical and comparative key, analyzing sedimentation processes that give rise to peculiar and “artificial” aggregations, such as invented archives and participatory archives. An attempt is also made to reflect on the consequences of the multiplication of potential producers and on the idea of democratization of history, with particular reference to the role of the archival discipline in these scenarios.

**Keywords:** Invented archives, Participatory archives, Archival context, Archives, Digital archives.

## 1. Dal vincolo al wiki - mondo

Cosa sia un archivio lo sappiamo bene. O almeno crediamo di saperlo quando recitiamo a memoria le formule più accreditate per descriverne la natura e la sostanza. Le definizioni oscillano nel tempo e nello spazio, ma alla fine tendono a collimare e a sublimarsi, da Casanova agli standard di descrizione archivistica. Il motore immobile che fa girare questo universo definitorio è il concetto di *archivio in senso proprio*, risultato di una lunga analisi giuridica, storica e filologica delle ragioni per cui un archivio viene prodotto, conservato e usato. È un'idea che muove dal prerequisito dell'organicità e si fonda sul rapporto serrato tra produttore/produzione, sedimentazione, uso e conservazione. La figura del soggetto produttore, nella sua limpida univocità, giganteggia. Conoscere il *creator* significa giustificare l'esistenza dell'archivio “organico” e, al tempo stesso, garantirne identificazione, affidabilità e fruizione. Una lunga tradizione e un'adeguata giurisprudenza ci consegnano dunque una solida percezione della sedimentazione documentaria, caratterizzata da un robusto pragmatismo.

---

\* Dipartimento di Scienze della Formazione, Beni Culturali e Turismo, Università di Macerata (MC). federico.valacchi@unimc.it.

Altrettanto consolidata, e si potrebbe dire inevitabile, è l'immanenza della polifunzionalità, secondo la quale nel tempo gli archivi cambiano faccia e finalità. I ritmi del ciclo vitale sembrano incontestabili e, per quanto con molte possibili deviazioni, collocano i fondi archivistici dentro a flussi di collaudata armonia.

Cosa succede, però, se per qualche accadimento una costruzione documentaria non risponde più a questi canoni e smarrisce la sua granitica organicità a vantaggio di fisionomie e strutture molto più evanescenti? Cosa succede, cioè, quando le ragioni e i mezzi della produzione e della trasmissione non sono più gli stessi da cui scaturiscono definizioni consolidate e faticosamente condivise? Sono domande che è opportuno farsi quando constatiamo come la definizione di archivio, in determinati contesti, possa risultare eccentrica rispetto a quella conquistata dalla nostra pur rispettabile tradizione.

L'archivio in senso proprio cui noi ci riferiamo non è un τόπος universalmente e necessariamente condiviso. In contesti giuridici, culturali e tecnologici diversi dal nostro, in particolare nel mondo anglosassone, le cose possono infatti andare diversamente. Ormai da tempo, anzi, sono andate diversamente e non mancano definizioni e concettualizzazioni alternative a quelle a noi più vicine (Gilliland e McKemmish 2014). In particolare, l'uso politico, attivo e democratico dell'archivio ha dato luogo a complessi documentari di natura eterogenea e decisamente "impura", nei quali le ragioni genericamente pubbliche prevalgono su quelle giuridiche e culturali in senso stretto.

In questi peculiari percorsi si ridefiniscono alla base le ragioni e le modalità della produzione e dell'organizzazione dei sistemi di fonti. Parlare di *invented archive* o di *participatory archive*, come vedremo, crea i presupposti per una progressiva crisi dell'univocità della provenienza e fa tremare il castello classificatorio dell'archivio *proprio*, a tutto vantaggio di una dilagante artificiosità della sedimentazione.

A ben guardare, poi, ci sarebbe anche di più, nel moltiplicarsi di aggregazioni decisamente spurie ma capaci di assumere rilevanza documentaria, proprio in ragione di una matura pervasività e della patente di affidabilità ricavata dai diritti di una condivisione fortemente plurale (Catalani e Feliciati 2018)<sup>1</sup>. È il *wiki-mondo*, dove «tutto si regge sui volontari che aggiungono, modificano, controllano, pubblicano le informazioni di ogni singola voce» (Fasano 2011, 23) e dove, in maniera per certi versi ideologica, la conoscenza può/deve essere il prodotto di estese comunità di *prosumer* (Rosenzweig 2006).

Questa appassionata partecipazione è cosa buona e giusta. Va sicuramente assecondata e perfezionata, ma ne vanno esplorate anche le zone d'ombra. Dietro a tutto questo, infatti, non c'è solo la lieta novella della democrazia partecipativa del sapere, ma un progressivo decentramento della produzione

<sup>1</sup> Al tema è dedicato l'intero fascicolo di (*JLIS.it* 2018).

che si spinge fino a livello individuale e, alle estreme conseguenze, enfatizza il ruolo del singolo dentro alle logiche logorroiche dei social. Gli strumenti con cui si costruisce e si trasmette l'informazione, nel momento in cui si ramificano, creano una serie di problemi che vanno anche oltre la necessaria verifica dell'autorevolezza della fonte e dei dati (Balbi e Ortoleva 2023). Quando si moltiplicano le sorgenti aumenta il rumore di fondo e prendono corpo catene documentali sempre più personalizzate, più o meno strutturate, più o meno qualificate, più o meno degne di sopravvivere.

Si potrebbe certo limitarsi ad obiettare che tutto ciò non ha nulla a che vedere con l'archivio dell'archivistica e continuare lungo la strada consueta, visto che su quel fronte ancora moltissimo c'è da fare. Se però l'archivistica non è tassonomia fine a se stessa, ma ha l'ambizione di contribuire alla costruzione e alla trasmissione di una conoscenza per quanto possibile affidabile, tirare dritto non è possibile, sarebbe un grave tradimento deontologico.

Il confronto con un'ebollizione documentaria ormai diffusa è inevitabile, e sembra inevitabile anche l'esigenza di valutare se e come il metodo e le prassi possano reagire a queste provocazioni, magari dentro a costruttive alleanze con altre discipline e con i loro saperi (Tomasi 2022).

## 2. L'emozione di un momento

Per valutare al meglio certe espressioni documentarie *atipiche*, è utile soffermarsi prima di tutto sul lato che potremmo definire "emotivo" della produzione archivistica. Gli archivi, comunque li si voglia valutare, portano con se forti dosi di soggettività. Sono sempre interpretazioni della realtà, non la realtà in purezza. La pragmaticità dei documenti è meramente congiunturale ed è condizionata dai sentimenti di chi li produce e dai personalissimi obiettivi di chi li usa. La teoria postula l'avalutatività ma si tratta di un vezzo deontologico, o poco più, perché ogni istante del ciclo vitale è permeato di volontà. Le carte tendono a subire le conseguenze della volubilità di un fato avverso o, più spesso, di un preciso disegno di costruzione della memoria rappresentativa. In qualsiasi archivio, insomma, bisogna fare i conti con l'imprevedibilità e con l'emotività della vita vera da cui esso scaturisce.

Le emozioni, quindi, sono parte integrante degli archivi e hanno un ruolo non secondario nelle modalità di percezione, interpretazione e uso dei documenti, ma bisogna intendersi prima di tutto sul significato della parola emozione, almeno limitatamente ai suoi "usi archivistici".

Secondo il dizionario un'emozione è «una viva sensazione suscitata da qualcosa» o, meglio, «una viva impressione, che può essere un'eccitazione positiva o un senso sgradevole di turbamento» (Treccani, n.d.).

Le diverse sfumature della definizione sembrano compatibili con tutta una serie di manifestazioni archivistiche che possono generare sensazioni di questo

tipo. L'emozione archivistica di cui parliamo qui non è esattamente quella che scaturisce da certi luoghi comuni, ovvero da un generico fascino degli archivi da cui poi tutto discenderebbe, in un crescendo di entusiasmo euristico o di incontenibile noia, a seconda delle sensibilità e degli approcci.

Gli archivi possono suscitare queste emozioni *dirette*, quando rispondono a un determinato bisogno di ricerca, ma anche dar luogo a emozioni *indirette*, nel momento in cui li si organizza con l'obiettivo deliberato di creare "eccitazione positiva" o "turbamento". La polifunzionalità e l'eterogenesi dei fini possono giustificare usi e percezioni che vanno oltre il significato immediato della parola scritta. L'emozione dell'archivio sta allora innanzitutto qui, nella consapevolezza che l'archivio è molto più grande di se stesso, e può alimentare processi incrementali che vanno oltre le aride ragioni della produzione, generando conoscenze ampie e imprevedibili.

Esistono poi archivi, o sedicenti tali, nei quali l'emozione e l'emotività hanno un valore programmatico che ne fa le sorgenti stesse della produzione. L'emozione non scaturisce dall'archivio ma, al contrario, l'archivio è il risultato dell'emotività – individuale o, più spesso, collettiva – che si traduce nel tentativo di orientare e giustificare le conseguenze di ciò che "documenta". È un archivio che si palesa risorgendo senza soluzione di continuità dalle sue ceneri. Un archivio nel quale la scintilla fatale della produzione non scocca dal bisogno di un flusso documentale a sostegno di una prassi in divenire, ma dalla volontà di memorizzare e interpretare cose già avvenute.

La spontaneità involontaria della *universitas rerum* si trasforma in architetture costruite ex post. Nascono archivi asincroni, congegnati per rispondere a precisi bisogni delle diverse comunità designate, nei quali i documenti, adeguatamente organizzati e trattati, possono generare flussi emotivi di portata anche considerevole e influenzare ad arte il sentimento pubblico. In questo senso l'emozione non è la meraviglia di fronte al residuo fisso del tempo andato. È un'emozione per certi versi assai poco emotiva, il risultato di modelli di gestione e uso dei complessi documentari che risentono di una molteplicità di fattori e che possono dare una molteplicità di risultati.

Quando la memoria è un'impressione, più che un processo di sedimentazione, l'emozione ha già prevalso sull'analisi e i giochi sono fatti.

### 3. L'archivio è l'uso che se ne fa e il mezzo che lo produce

L'archivio su cui si è costruito questo castello ermeneutico nasce "spontaneamente" dalla realtà, è un dato di fatto indiscutibile, frutto semplicemente della catena fattuale che lo alimenta. È anche vero, d'altra parte, che la presunta neutralità dei fondi archivistici ha da sempre assecondato molti bisogni eccentrici alle ragioni teoriche della produzione, a tutto vantaggio di quelle dell'uso (Mineo 2020). Sono le istanze degli utenti a qualificare il potenziale

altrimenti inesperto dell'archivio (Valacchi 2020). L'eterogenesi dei fini, in questa lettura, muove dalla naturalezza innata della produzione verso la forte soggettività della fruizione.

Negli archivi in senso proprio le ragioni del contesto sono abbastanza forti da compensare l'eccentricità dell'utilizzazione. Un documento che appartenga a un fondo descritto e ordinato, comunque lo si usi, porta con sé il suo imprinting ed esiste sempre la possibilità di ricondurlo al suo status originario, ostacolando macchinazioni informative più o meno perverse. Se l'archivio è figlio legittimo del suo produttore, in altre parole, la provenienza lo qualifica e lo certifica, anche se non ne indirizza necessariamente l'utilizzazione.

Quando invece l'archivio si trasfigura, e le ragioni della produzione cedono il passo a quelle di una artificiosità a trazione digitale, il ragionamento si complica. Se l'archivio storico diviene archivio *per una storia*, al servizio di punti di vista particolari, cambiano le regole del gioco e non è semplice dare evidenza contestuale a fonti disperse nei meandri di arbitrari montaggi ad personam.

Archivio è una parola per tutte le stagioni, oggi più che mai. Gli archivi contemporanei sono soggetti a trasformazioni strutturali che annacquano sensibilmente il rigore e il significato del concetto tecnico/scientifico da cui siamo partiti, dilatandoli fino alle conseguenze, estreme, nebulose e per nulla convincenti, del "tutto è archivio" (Giannachi 2021).

Il polimorfismo documentario è ormai molto marcato, oscilla in un range che va dagli archivi *convenzionali* alle imprevedibili performances dell'archivio digitale, nelle sue molte possibili accezioni. Sono davvero archivi questi *archivi* genericamente digitali? La questione non è banalmente etimologica e non si tratta neppure di difendere a oltranza una "denominazione di origine protetta". Di sicuro la parola archivio non basta più a presidiare un dominio sempre più vasto e un uso del termine sempre più fluttuante. Vengono infatti chiamati archivi diverse tipologie di insiemi di dati, riconducibili non tanto alla *proprietà* dell'archivio quanto all'idea di raccolta o collezione. In un crescendo sostenuto dalla duttilità digitale, si tende allora ad assimilare le invenzioni archivistiche digitali al collezionismo o all'organizzazione per materie. Si tenta cioè di prevenire obiezioni e preoccupazioni metodologiche ribadendo che, come al solito, il sole non splende su cose nuove: «but though the tools of communication belong to the 21st century, participatory practice is invariably informed and shaped by traditions of antiquarian collecting and "serious leisure" research activity [...]» (Benoit e Eveleigh 2019, 1).

Questi fenomeni, invece, non sono né consueti né innocenti e vanno valutati nella loro originalissima essenza. Le stesse ragioni di natura filosofica, antropologica, sociale e culturale che stanno alla base di molti artifici digitali dimostrano che non c'è automatica continuità con il passato. Se è vero, infatti, che il collezionismo antiquario è da sempre una malattia infantile dell'archivistica, in nessun momento storico esso ha potuto fare affidamento su mezzi

di legittimazione e trasmissione tanto potenti. Il mezzo fa la differenza. La selettiva ostensione digitale ha una forza destinata a prevalere sulla paludata interezza analogica, innescando veri e propri processi di rimozione. Non a caso, è stata coniata l'espressione "hidden collections", soprattutto per individuare «materials that cannot be found in the online catalog» (Yakel 2005, 95).

L'archivio non nasce più *dall'azione* ma *per l'obiettivo*, è evidenza meditata dei fatti e non più loro conseguenza. Nel momento in cui svanisce l'innocenza originaria (per quanto presunta) cambiano le carte in tavola. Il filmato documentario è il risultato di un montaggio funzionale a sostenere fatti ed azioni nuove e concrete. Si altera il ritmo di secolari flussi documentali, mentre gli oggetti si intrecciano tra loro per produrre risultati *atipici* ma capaci di penetrare in profondità nell'immaginario collettivo. Così facendo si creano le condizioni per una ridondante memoria on demand, gravida di conseguenze non sempre auspicabili. La potenza di calcolo, che si traduce in una smisurata capacità di immagazzinamento e trasmissione, finisce con l'alimentare un paradossale malessere postmnemonico, nel quale la quantità rischia di uccidere la qualità e l'effettiva usabilità.

Viene da chiedersi se davvero questa immane mole di dati ci è indispensabile e se davvero ha un senso la rincorsa ad accumularne volumi sempre più ingenti per nutrire il moloch dei big data. Forse no e forse la velocità un po' allucinata delle macchine non è l'ottica giusta con cui valutare la realtà.

Come si suol dire, però, questo è un altro discorso...

#### 4. L'invenzione dell'archivio

Le allettanti opportunità binarie, ogni giorno più performanti, hanno ben presto popolato i sogni dei ricercatori di obbedienti archivi digitali, lasciando intravedere nuovi modi di scrivere la storia (Vitali 2004). Ci si è chiesti innanzitutto in che modo, loro tramite, si potesse superare la fatica dell'archivio, intesa come lunga e incerta approssimazione ai contenuti. Di conseguenza, si è molto ragionato sulla possibilità di assecondare i desideri della ricerca, magari creando contenitori di oggetti digitali ad hoc, a partire dalle domande che la ricerca stessa si poneva (Rosenzweig 2005).

Il concetto di *archivio digitale*, nella sua potenza polisemica, ha molte possibili declinazioni e altrettante accezioni. Non tutte corrispondono al modello dell'archivio in senso proprio, sia pure sedimentatosi in assenza di carta. L'espressione non ha quindi nulla di neutro e, al contrario, ha suscitato e suscita più di una cautela tra gli archivisti per le conseguenze che può avere (Vitali 2006).

L'archivio digitale di cui parliamo qui non esiste in natura, ma, per così dire, è un organismo geneticamente modificato, inventato in laboratorio. È *artificiale*, appunto.



Per parte sua, un *invented archive*, se vogliamo tentare di definirlo, è un'aggregazione di oggetti digitali, generalmente svincolati dai loro soggetti produttori e sottoposti poi ad un montaggio fortemente soggettivo. È caratterizzato da sistemi di relazioni che scaturiscono dalle finalità rappresentative e non dalla necessità della classificazione e del vincolo. Questi archivi, come sempre ma più di sempre, possono contribuire alla costruzione di specifiche interpretazioni della realtà. Il punto di vista sostituisce il vincolo. Se il vincolo è la conseguenza logica, necessaria e involontaria dell'azione che si manifesta in sequenze di documenti, il punto di vista mette in relazione tra loro i documenti a sostegno del pensiero che si forma intorno all'azione e alle sue conseguenze.

L'invenzione ha qualcosa in più della pertinenza e dell'ordinamento per materie perché, più che ad organizzare per trovare un dato, punta ad accumulare per dimostrare un pensiero, e magari giustificarne le conseguenze. Inventare l'archivio significa orientare le risposte documentarie, con una leggerezza che gli archivi in senso proprio, nella loro atavica austerità, non si possono permettere.

La chiave di lettura di questi archivi è il loro obiettivo tematico, con buona pace del principio di provenienza. Conservano e propongono documenti avulsi dai loro produttori e forzosamente ricontestualizzati. Il produttore dell'archivio è sostituito dal *soggetto costruttore*. Il processo non è indolore. Si passa dal *creator* al *builder*, con quello che ne può conseguire anche in termini di riposizionamento della descrizione archivistica. L'opinione finisce col prevalere sui fatti e il punto di vista sul contesto.

Detto tutto questo, e adottando tutte le precauzioni del caso, non è possibile formulare un giudizio di astratto rifiuto di questi sistemi di fonti solo perché essi non corrispondono a un canone consolidato. Nel mondo complicato dei documenti l'evidenza è infatti il primo metro di giudizio. Se una sedimentazione documentaria si manifesta con vigore, qualunque ne siano le fattezze, non conta che essa rispecchi o meno canoni o generici principi filologici: quella sedimentazione esiste e la si deve valutare e considerare in quanto tale. L'archivistica non può, né potrà mai, condizionare la sedimentazione degli archivi, tanto meno di quelli digitali. Il problema non è se queste risorse debbano essere valutate, ma è come valutarle dentro a simili contesti iperattivi.

In tutti questi casi contestualizzare significa dare conto soprattutto del profilo del soggetto aggregatore, nel tentativo di evidenziare e spiegare le ragioni dell'invenzione. La descrizione archivistica si sposta sopra e fuori dalla sedimentazione e diventa più che altro interpretazione critica e istruzione per l'uso, dal momento che sembra difficile ipotizzare approcci descrittivi che presuppongano un'azione di sintesi ex post. L'archivio inventato, del resto, non diventa mai propriamente un archivio storico, dal momento che fluttua nel web e resta costantemente aperto a ulteriori contributi. È pressoché impossibile cristallizzarlo a un momento dato.

Per amministrare a dovere queste complessità si sconfinano di necessità nel campo della gestione allargata della conoscenza, attingendo a modelli di descrizione multidimensionali e integrati, capaci di dare conto delle peculiarità dei diversi sistemi. Entrano in gioco competenze integrate e modelli descrittivi a loro volta integrati ed elastici, di area Library and Information Science (LIS), potremmo spingerci a dire (Valacchi 2022).

Se vogliamo scendere sul terreno concreto, un caso ormai classico e sufficientemente consolidato di archivio inventato è il (The September 11 Digital Archive, n.d.). Questo ambizioso progetto, dal malcelato retrogusto patriottico, è stato inizialmente finanziato dalla Alfred P. Sloan Foundation e concepito dall'*American Social History Project* e dal *Roy Rosenzweig Center for History and New Media* (Silveira 2018).

Il grande serbatoio digitale ha preso forma a partire dall'onda emotiva sollevata dagli attacchi agli USA del 2001 e si configura esplicitamente come un contenitore destinato a un pubblico da ascoltare, informare e orientare. Più precisamente, *The September 11 Digital Archive* si propone di conservare e rendere fruibili i materiali utili per salvaguardare e rappresentare la storia degli attentati "and its aftermath".

Evocando le conseguenze fin dalla presentazione del progetto si fa un immediato ed esplicito riferimento al senso politico del progetto stesso. L'archivio crea i presupposti documentari ad azioni conseguenti ai fatti che documenta. C'è una naturale continuità tra la raccolta e gli eventuali avvenimenti successivi che gli stessi documenti contribuiscono a generare, spiegare e giustificare.

Si ha motivo di credere che il senso dell'intera operazione non sia semplicemente evocativo ma, in una sorta di outing collettivo, punti piuttosto a curare una ferita e a costruire e giustificare un'adeguata risposta. Non a caso, nella sezione delle FAQ, si fa riferimento esplicito alla *National Commission on Terrorist Attacks Upon the United States (la 9-11 Commission)* istituita «to prepare a full and complete account of the circumstances surrounding the September 11, 2001 terrorist attacks, including preparedness for and the immediate response to the attacks» (Frequently Asked Questions about 9/11, September 11 Digital Archive, n.d.).

L'archivio, chiamiamolo così, si segnala per il forte polimorfismo tipologico delle risorse documentarie conservate, non disgiunto da una grande varietà di provenienze e accumula: «more than 150,000 digital items, a tally that includes more than 40,000 emails and other electronic communications, more than 40,000 first-hand stories, and more than 15,000 digital images» (The September 11 Digital Archive, n.d.).

Significativa anche la scelta di organizzazione dell'informazione, che non rinuncia al concetto nominale di albero (si parla di *collection-tree*) ma sostituisce gli abituali modelli gerarchici con strutture tipologiche (*art, audio, video* per esempio), funzionali (*organizations*) o conservative (istituzioni della me-

moria) (Collections, September 11 Digital Archive, n.d.). Da segnalare senz'altro la sezione *Personal Accounts*, popolata da una molteplicità di testimonianze individuali che esaltano la vocazione partecipativa. Qui la soggettiva del montaggio risulta in tutta la sua evidenza (Personal Accounts, September 11 Digital Archive, n.d.).

Nel suo insieme questa "invenzione" determina importanti ricadute metodologiche e suggerisce più di una riflessione sull'incerto futuro dell'archivistica e della ricerca storica: «The *Archive* is also using these events as a way of assessing how history is being recorded and preserved in the twenty-first century and as an opportunity to develop free software tools to help historians to do a better job of collecting, preserving, and writing history in the new century» (About, September 11 Digital Archive, n.d.).

L'idea secondo la quale tocchi agli storici a fare il lavoro di raccolta e conservazione delle loro stesse fonti allontana definitivamente gli archivi inventati e partecipativi dal concetto di archivio in senso proprio. Sembra che ci si dimentichi infatti di quella separazione dei poteri che da sempre ha garantito sia il lavoro degli storici che quello degli archivisti. L'archivio, come lo abbiamo sempre conosciuto, è infatti una risorsa neutra e polivalente, sia pure con i suoi limiti, mentre in questo caso i sistemi di fonti sono costruiti ad uso e consumo di chi poi li dovrà consultare, con un pericoloso tasso di soggettività e in un conflitto di interessi molto più che latente.

## 5. Tutti insieme appassionatamente: gli archivi partecipativi

L'archivio inventato evolve per sua natura verso quello partecipativo (Benoit e Eveleigh 2019). «Participatory archives acknowledge that multiple parties have rights, responsibilities, needs and perspectives with regard to the record. They are recreated by, for and with multiple communities, according to and respectful of community values, practices, beliefs and needs [...]» (Gilliland e McKemmish 2014, 78). La sedimentazione partecipativa prevede quindi la collaborazione di una pluralità di attori nella costruzione dell'archivio, con il fine di dar luogo a sistemi documentari fortemente connotati e, anzi, *unici* nel loro genere: «participatory archiving is a form of collaborative practice in which archivists, historians, and community members *work together* to document local and community histories and build unique archival collections» (Stull, n.d.). La distinzione tra produzione e uso in questa visione tende evidentemente a scomparire: chi produce usa e viceversa, mentre si costruiscono *archival collections* peculiari, *unique* appunto, sulla cui neutralità contestuale è lecito nutrire qualche dubbio. L'archivio partecipativo è un party a cui tutti i vicini sono invitati, ma, come ogni buon party, può avere esiti imprevedibili.

L'idea di una gestione collettiva della memoria pubblica basata sull'originalità creativa di una raccolta di documenti è molto forte, oltretutto molto post-

moderna (Cook 2001). Mette in gioco categorie archivistiche consolidate e le rimodula secondo gli interessi delle diverse comunità di riferimento. L'obiettivo dichiarato e, almeno in apparenza, "democratico", è quello di «incorporate multiple voices, reach diverse audiences, and encourage popular participation in presenting and preserving the past» (Roy Rosenzweig Center for History and New Media, n.d.a).

L'archivio partecipativo a questo livello intercetta e probabilmente lenisce alcune sofferenze della public history. È l'esito di un percorso lungo il quale la logica di una certa storia pubblica può creare le premesse per la nascita di sistemi di fonti che obbediscano ai suoi requisiti di fondo. La public history, che fatica a trovare risposte davvero esaurienti in archivi che non sono pubblici nel senso che i *public historian* si attendono, può esaltarsi nei sistemi partecipativi e può forse essere uno strumento utile per valutare e fronteggiare questi fenomeni. Su questo terreno, del resto, l'archivistica ha bisogno di alleati: può sollevare il problema ma non ha tutte le armi per costruire adeguate contromisure. Ammesso naturalmente che si concordi sul bisogno di contromisure.

La memoria pubblica e collettiva è sicuramente un valore, ma in certe sue declinazioni radicali ha qualcosa di paradossalmente orwelliano. La comunità (ovvero *il pubblico*) è il vero e inarrestabile Grande Fratello, in un'entasi crescente dove i valori e i pregi della partecipazione rischiano di sconfinare in una sorta di demagogia documentaria fuori controllo. Alle sue estreme conseguenze l'utente è il centro insindacabile dell'universo: «radical user orientation means that the archive is oriented and reoriented to its users all the time» (Huvila 2008, 17).

A questo livello che si manifesta il problema non secondario dell'autorevolezza e dell'affidabilità dei modelli conservativi: «thanks to the potential of Web 2.0 technologies, diverse modes of interaction and participation have emerged, minimizing institutional boundaries between cultural agents and users» (Alaoui 2019, 470). In questa lettura l'abbattimento della distanza vera o presunta tra gli agenti culturali "istituzionali" e gli utenti sembra essere un valore non negoziabile. È legittimo nutrire qualche perplessità al riguardo, proprio in ragione dell'esigenza di fare affidamento su un'autorevolezza riconosciuta e condivisa della conservazione. Superare il confine in maniera disordinata può infatti semplicemente alimentare l'incertezza e spingere verso una potenziale anarchia della conoscenza, agli antipodi di equilibrati processi di democratizzazione del sapere.

Non si tratta naturalmente di difendere a oltranza un modello o tanto meno una corporazione. La dimensione partecipativa è senza alcun dubbio un forte valore aggiunto che sprigiona grandi energie di archivistica attiva. Il punto è piuttosto intendersi sul senso del partecipare. Questa partecipazione è condivisione di un sistema etico o è costruzione strumentale di risorse condi-

visive limitate ad obiettivi specifici? Alimenta processi incrementali di rimodulazione dell'identità o si limita a uno story telling tutto sommato circoscritto?

L'archivio è pubblico per vocazione, su questo non c'è dubbio. Ma in questi casi ci si deve chiedere cosa si intenda con "pubblico", cioè se la pubblicità sia un attributo dell'uso o della produzione. La moltiplicazione generalizzata delle provenienze da un lato enfatizza l'aspetto pubblico, ma dall'altro può creare i presupposti per una memoria generalista dentro alla quale va in sofferenza la qualità dei dati, finora garantita da attori identificabili e da parametri e procedure condivise.

La libertà è senza dubbio partecipazione ma la partecipazione non è immediata garanzia dei diritti e dei doveri che ci rendono compatibilmente liberi dentro a società complesse. La libertà in una democrazia è un esercizio quotidiano che passa anche da affidabili processi di legittimazione documentaria. In tempi di partecipazione e di intelligenze artificiali più o meno invasive è indispensabile valutare criticamente l'affidabilità dei dati e farlo non è semplice retroguardia o resistenza al cambiamento, ma un preciso dovere etico.

È appena il caso di notare, infine, come sullo sfondo si profili anche la questione della *long time preservation* di risorse che, in quanto digitali, devono fare i conti con la loro obsolescenza e con l'esigenza di affidabili politiche conservative. La consueta fatica della conservazione digitale può accentuarsi in una logica che sembra propensa a ridimensionare sensibilmente il ruolo degli "agenti culturali" e anche questo è un problema da tener presente.

Fatte tutte queste considerazioni critiche, non si può negare l'esistenza e l'importanza di realtà che in molti casi aggiornano e orientano il potenziale intrinseco alla parola archivio. Il richiamo al metodo in quanto tale non serve a negare, quindi, ma costituisce un possibile termine di paragone. Una lettura comparativa può offrire l'opportunità di valutare e collocare gli archivi partecipativi nel quadro di una storia e di una cultura di lungo periodo, considerandoli una tappa del processo evolutivo e non una frattura profonda con il passato.

Nella sua concretezza, un uso interessante dell'idea di archivio partecipativo è emerso durante la pandemia da Covid-19. La costruzione di risorse documentarie su questo tema ha avuto innanzitutto un ruolo importante sul piano della condivisione di dati e informazioni utili nella battaglia contro il virus. Le ricadute più efficaci – in linea con la fisionomia partecipativa – si colgono nel ruolo che potremmo definire psicologico e umanitario di questi archivi. Essi hanno infatti contribuito ad alimentare una sensazione di comunità e di empatica condivisione della sofferenza, in un periodo di drammatico isolamento degli individui. In questi archivi della resilienza individuale e collettiva si è manifestato «the power of collaborative documenting, storytelling, and memory-making to foster connectivity, increase empathy, and build resilience» (Bushey 2023, 2379). Al tempo stesso, negli archivi della pandemia è

venuta sedimentandosi la coscienza storica di un evento planetario, forse uno dei pochi esempi di memoria universalmente condivisa. La pandemia, tra le sue molte conseguenze, ha influenzato anche l'azione dei soggetti conservatori, riposizionandone almeno parzialmente gli assetti e i protagonisti: «the outbreak of the pandemic created unprecedented challenges for galleries, libraries, archives and museums (...) The COVID-19 pandemic served as a wake-up call to rethink how cultural production and consumption are organized and articulated with different sets of actors» (GLAMMONS, n.d.).

Esistono anche altri esempi dell'uso partecipativo dell'archivio, più vicine a una lunga continuità d'uso. Sono manifestazioni che prendono corpo soprattutto in una dimensione tesa alla "democratizzazione della storia". Esempi di grande interesse al riguardo sono quelli disponibili nel portale del *Roy Rosenzweig Center for History and New Media*, un soggetto molto attivo e quotato in questo ambito (Roy Rosenzweig Center for History and New Media, n.d.b).

Alcuni di questi archivi mettono al centro della loro attenzione precisi periodi storici, offrendo panoramiche interessanti e di sicura valenza didattica. È ad esempio il caso di *Liberty, Equality, Fraternity: Exploring the French Revolution* (Roy Rosenzweig Center for History and New Media, n.d.c) o di *Making history of 1989* (Roy Rosenzweig Center for History and New Media, n.d.d), dedicato alla caduta del comunismo nell'est Europa. Questi due "archivi" raggiungono sicuramente i loro obiettivi "democratici" e svolgono un ruolo importante nella diffusione della conoscenza basica, ma non vanno esenti dai limiti di cui si diceva sopra. Allo stesso modo vanno segnalati tra le risorse disponibili supporti alla storia digitale locale (Roy Rosenzweig Center for History and New Media, n.d.e), alla *data driven history* (Roy Rosenzweig Center for History and New Media, n.d.f), alla public history (Roy Rosenzweig Center for History and New Media, n.d.g) e alla didattica della storia (Roy Rosenzweig Center for History and New Media, n.d.h).

Ognuno di questi progetti documentari è il risultato di costruzioni partecipative e tematiche. Molto spesso, e sicuramente nel caso del *Roy Rosenzweig Center*, l'istituzione che li coordina gode di indubbia autorevolezza. Un'autorevolezza che paradossalmente è per certi aspetti un pericolo ulteriore, perché contribuisce a rendere affidabili a prescindere questi archivi artificiali, cui sembra al contrario mancare il respiro realmente ugualitario e democratico dell'archivio in senso proprio.

L'obiettivo che tali progetti perseguono è decisamente ambizioso e almeno apparentemente collocato alle porte di Utopia: «we use digital media and computer technology to democratize history» (Rosenzweig Center for History and New Media, n.d.a). Si ritiene cioè di poter orientare e ridefinire la fisiologia dei processi di sedimentazione, conservazione e uso per liberare gli archivi da presunti condizionamenti verticali e autoritari.



Per certi versi questo approccio introduce nell'uso degli archivi le distinzioni che passano tra le democrazie rappresentative e quelle dirette. In linea generale non si fa fatica a condividere l'idea dell'archivio come strumento di democrazia, ma lo strumento funziona quando è solida e qualificata garanzia di efficienza, trasparenza e certificazione dei diritti e dei doveri. L'affidabilità è il requisito indispensabile dell'archivio al servizio della democrazia. Nel caso degli archivi partecipativi, invece, la democrazia è il risultato di un'accumulazione ad ampio spettro, esercitata negli spazi di una *ἀγορά* nella quale è difficile stabilire effettive responsabilità.

Non si può negare che gli obiettivi siano sulla carta assolutamente virtuosi. Resta però qualche perplessità su una conservazione costruita "dal pubblico e per il pubblico", che produce archivi deliberatamente orientati a mettere in discussione modelli conservativi e istituzioni evidentemente percepiti come ostacoli alla democrazia della conoscenza.

Quando si perde di vista il senso giuridico ed etico dell'archivio, l'allargamento verso il basso della produzione documentaria e l'ampliamento della platea dei produttori può non bastare a garantire affidabilità e democrazia in senso stretto. Siamo di fronte a un'idea molto parziale del ruolo dell'archivio e il rischio di generalizzazione e decontestualizzazione è alto, per quanto l'offerta sia molto ricca e sicuramente intrigante.

Questi processi, in ultima analisi, non sono ingenuamente partecipativi. Ne scaturisce infatti un flusso emotivo ben definito, un punto di vista di cui chi organizza la risorsa è consapevole, quando si apre alla partecipazione del pubblico in quanto parte integrante del percorso di produzione. Il rapporto classico di produzione/uso/conservazione si rompe e viene messa in discussione una consolidata filiera nella quale la dialettica tra produttori, conservatori, mediatori e utenti scandiva tempi e responsabilità capaci di garantire in pieno la contestualizzazione e la fiducia. Il ruolo degli utenti con i loro bisogni non si manifesta più nel processo di mediazione ma fin da quello di sedimentazione, con conseguenze che sembrano non indifferenti sulla neutralità dell'archivio o, almeno, sulla sua equidistanza dai diversi punti di vista. Il potere secolare degli archivi si trasfigura e la collettività acquisisce uno *ius archivi* universale che gli permette di «create, change, destroy, share and keep records, in all forms, however and wherever» (Mackay 2019).

## 6. Conclusioni

Gli archivi sono il risultato di una sedimentazione, più o meno profonda. Ogni archivio ha il proprio codice genetico e una storia alle spalle che ne condiziona la portata reale. Sappiamo bene che non sono contenitori di essenze fattuali surgelate in attesa di uno dei molti possibili usi, ma una rappresentazione delle società, delle culture e delle tecnologie che li producono.



In tempi di crescita esponenziale di potenti automatismi informativi, e di sovrapposizione sempre più marcate tra intelligenza umana e artificiale, il problema della costruzione dell'archivio si pone con forza particolare. Non basta evocare scenari apocalittici per difendere l'umanità dalla ribellione di macchine perverse. Anche quel poco di buono di HAL 9000 alla fine è stato neutralizzato, e quella stessa umanità può difendersi da sola, se accetta un confronto critico e sereno con le sue tecnologie.

Discutere oggi sul senso della parola archivio, in fondo, significa proprio tentare di portare acqua alla causa di una coscienza che sia digitale ma critica. Il futuro non si può cancellare ma lo si può prevedere e perfino governare. Alla radice del potere delle macchine, vero o presunto che sia, rimangono pulsioni molto umane e forse è solo il nostro cattivo esempio a incoraggiare la ribellione delle Artificial Intelligence (AI).

I casi su cui ci siamo soffermati, con diverse caratteristiche, diversi obiettivi e diverse modalità, rappresentano esempi efficaci di come, con l'aiuto delle Information and Communication Technologies (ICT), soggettività, emotività e ricerca del consenso possano condizionare sedimentazioni documentarie che popolano il web di precisi e pregiudiziali punti di vista.

Nel mondo partecipativo la parola archivio lievita e si manifesta una dicotomia tra archivi "propri" e archivi "artificiali", mentre prendono corpo processi di mediazione indotta, sconosciuti in precedenza (Papy 2021). Molti complessi documentari etichettati come "archivi" perdono la loro assertività e tutti gli attributi che li hanno storicamente definiti e distinti. Divengono spazi negoziati e di negoziazione che liberano energie documentarie altrimenti represses. Il sapore politico e antropologico di natura partecipativa ridefinisce in senso attivo il concetto di archivio e il termine con cui lo si evoca. Queste forme di conoscenza, insieme al senso di una parola, mettono in discussione l'idea stessa di una società gerarchica e rigidamente strutturata.

L'archivio partecipativo è espressione di una palingenesi digitale che sembra mettere gli individui al posto delle istituzioni, in un ribaltamento di prospettive davvero rivoluzionario, nei (molti) limiti della sua reale sostenibilità. La potenza del medium lo rende ancora più visibile, efficace e apparentemente affidabile, soprattutto per utenti non particolarmente accorti. Non lo si può negare. Esiste e va tenuto in considerazione. La rimozione non è l'arma più efficace per contenerne le conseguenze.

Nel momento in cui l'archivio esce da se stesso e si genera e rigenera con mirabile e pervasiva rapidità tutto si complica. L'idea di *creator*, come abbiamo detto, si confonde con quella di costruttore o, meglio, di costruttori. Alle estreme conseguenze nell'archivio partecipativo, tutti e nessuno sono il soggetto produttore. Nel mondo che ormai condividiamo con le intelligenze artificiali probabilmente inventare gli archivi, senza aspettare pazientemente che si sedimentino, è inevitabile, ma questo non significa accettare l'archivio digitale

come un dato di fatto inoppugnabile. È ancora possibile, anzi, doveroso, esercitare un controllo, o almeno formulare delle valutazioni. Per farlo sembra proprio che dobbiamo assecondare il cambiamento senza opporre inutili resistenze.

L'archivistica – e non solo l'archivistica, ovviamente – deve fare i conti con le invenzioni partecipative. Deve cioè continuare a fare al meglio il suo mestiere, anche dentro a scenari documentari che ne incrinano le certezze disciplinari. Se ci fermassimo al metodo e alle nostre sicurezze si potrebbe passare oltre: questi non sono certo archivi. Si tratterebbe però di un'omissione di soccorso incompatibile con la nostra deontologia. Come abbiamo detto, invece, se queste risorse esistono bisogna cercare di conviverci, apprezzandone la funzione, valorizzandole e utilizzandole criticamente. Cercare di interpretarle non significa negarle o combatterle, ma credere ancora e ancora di più nel metodo archivistico, inteso come sistema comparativo di prassi e di valori. Sullo sfondo risuonano ancora le parole di Filippo Valenti: «compito dell'archivista non è di fornire dei “dati” allo studioso ma dei “documenti” da “interpretare”» (Vitali 2022, 206).

La descrizione rimane al riguardo la nostra arma più efficace. Giunti a questo punto, però, si avvia a divenire un complicato processo di sintesi che deve necessariamente appoggiarsi alle macchine e ai metadati che saremo capaci di far loro generare. Le diverse forme di produzione “sociale” devono essere governate e coordinate, per evitare un'anarchia della conoscenza. Occorre agire sulla consapevolezza del lettore e sulla sua capacità di discernere ciò che è di valore da ciò che non lo è. La struttura stessa del Web, in un'ottica di democratizzazione del sapere, delega agli intermediari questo compito. Ciò significa che per gli archivisti si manifesta la necessità di ripensare almeno in parte il loro profilo, modulando i processi di mediazione in una logica di connessione filologica fra la produzione anarchica ed emotiva e una fruizione consapevole e cosciente. La mediazione non è più soltanto una decodifica dei contesti, ma un processo educativo che si confonde nella *information literacy*<sup>2</sup>.

Ancora una volta l'archivistica è chiamata ad attraversare i suoi confini più lontani per sopravvivere: «finché non raggiungerai la verità, non potrai correggerla. Ma se non la correggi non la raggiungerai. Nel frattempo, non rassegnarti» (Saramago 2017, 8).

## Riferimenti bibliografici

About, September 11 Digital Archive. n.d. Consultato il 30 maggio 2023.  
<https://911digitalarchive.org/about>.

---

<sup>2</sup> Devo la forma definitiva delle conclusioni al confronto consueto con Francesca Tomasi che ringrazio.

- Alaoui, Siham. 2019. Recensione di *Participatory Archives: Theory and Practice*, a cura di Edward Benoit e Alexandra Eveleigh. *The American Archivist* 83, no. 2 (fall/winter 2000): 470-74.
- Balbi, Gabriele, e Peppino Ortoleva. 2023. *La comunicazione imperfetta. Ostacoli, equivoci, adattamenti*. Milano: Einaudi.
- Benoit, Edward, and Alexandra Eveleigh. 2019. *Participatory Archives: Theory and Practice*, edited by Edward Benoit and Alexandra Eveleigh. London: Facet Publishing.
- Bushey, Jessica. 2023. "A Participatory Archives Approach to Fostering Connectivity, Increasing Empathy, and Building Resilience during the COVID-19 Pandemic." *Heritage* 2023 6(3): 2379-93. <https://doi.org/10.3390/heritage6030125>.
- Catalani, Luigi, e Pierluigi Feliciati. 2018. "Wikipedia, le biblioteche e gli archivi / Wikipedia, Libraries and Archives." *JLIS.it* 9 (3): I-III. <https://doi.org/10.4403/jlis.it-12510>.
- Collections, September 11 Digital Archive. n.d. Consultato il 30 maggio 2023. <https://911digitalarchive.org/collection-tree>.
- Cook, Terry. 2001. "Archival science and postmodernism: new formulations for old concepts." *Archival Science* 1: 3-24. <https://doi.org/10.1007/BF02435636>.
- Da Silveira, Pedro T. 2018. "From Instant History to the Infinite Archive: Digital Archiving, Memory and the Practical Past at the Roy Rosenzweig Center for History and New Media." *Historiein*, 17(2). <https://doi.org/10.12681/historiein.10964>.
- Fasano, Giusi. 2011. "Cronache." *Corriere della sera*, 6 agosto 2011, 23, come citato in Treccani. n.d. "wiki-mondo." Consultato il 30 maggio 2023. [https://www.treccani.it/enciclopedia/wiki-mondo\\_%28altro%29/](https://www.treccani.it/enciclopedia/wiki-mondo_%28altro%29/).
- Frequently Asked Questions about 9/11, September 11 Digital Archive. n.d. Consultato il 30 maggio 2023. <https://911digitalarchive.org/faqs#response>.
- Giannachi, Gabriella. 2021. *Archiviare tutto. Una mappa del quotidiano*. Torino: Treccani.
- Gilliland, Anne J., and Sue McKemmish. 2014. "The role of participatory archives in furthering human rights, reconciliation and recovery." *Atlanti: Review for Modern Archival Theory and Practice*, 24: 78-88.
- GLAMMONS. n.d. "Resilient, sustainable and participatory practices: Towards the GLAMs of the commons." Consultato il 30 maggio 2023. <https://glammons.eu/>.

- Huvila, Isto. 2008. "Participatory archive: towards decentralised curation, radical user orientation and broader contextualisation of records management." *Archival Science* 8 (1): 15-36. [https://www.researchgate.net/publication/226936883\\_Participatory\\_archive\\_Towards\\_decentralised\\_curation\\_radical\\_user\\_orientation\\_and\\_broader\\_contextualisation\\_of\\_records\\_management](https://www.researchgate.net/publication/226936883_Participatory_archive_Towards_decentralised_curation_radical_user_orientation_and_broader_contextualisation_of_records_management).
- JLIS.it* 9 (3) 2018, <https://www.jlis.it/index.php/jlis/issue/view/9>.
- Mackay Hannah. 2019. "The Participatory Archive: Designing a spectrum for participation and a new definition of the participatory archive." <https://medium.com/@mackayhjc/the-participatory-archive-designing-a-spectrum-for-participation-and-a-new-definition-of-the-964bc1b0f987>.
- Mineo, Leonardo. 2020. "Le ordinaroni in serie a proprio ed altrui vantaggio. Collections of autographs and archival science." *JLIS.it* 11(1):130-50. <https://doi.org/10.4403/jlis.it-12587>.
- Papy, Fabrice. 2021. "Innovations numériques anthropocentrées pour le web des données et des documents : des perspectives d'émergence pour des communautés à orientation épistémique ?" *AIDAinformazioni*, 3-4 (luglio-dicembre): 67-82. <https://doi.org/10.57574/596509004>.
- Personal Accounts, September 11 Digital Archive. n.d. Consultato il 30 maggio 2023. <https://911digitalarchive.org/collections/show/278>.
- Rosenzweig, Roy. 2005. "Digital Archives Are a Gift of Wisdom to Be Used Wisely." *The Chronicle of Higher Education* 42. <https://rrchnm.org/essay/digital-archives-are-a-gift-of-wisdom-to-be-used-wisely/>.
- Rosenzweig, Roy. 2006. "Can History be Open Source? Wikipedia and the Future of the Past." *The Journal of American History*, 1: 117-46. <https://www.chronicle.com/article/digital-archives-are-a-gift-of-wisdom-to-be-used-wisely/>.
- Roy Rosenzweig Center for History and New Media, n.d.a. "Our History." Consultato il 30 maggio 2023. <https://rrchnm.org/our-history/>.
- Roy Rosenzweig Center for History and New Media, n.d.b. "Our work." Consultato il 30 maggio 2023. <https://rrchnm.org/our-work/>.
- Roy Rosenzweig Center for History and New Media, n.d.c. "Liberty, Equality, Fraternity: Exploring the French Revolution." Consultato il 30 maggio 2023. <https://rrchnm.org/portfolio-item/liberty-equality-fraternity-exploring-the-french-revolution>.
- Roy Rosenzweig Center for History and New Media, n.d.d. "Making the History of 1989." Consultato il 30 maggio 2023. <https://chnm.gmu.edu/1989/>.

- Roy Rosenzweig Center for History and New Media, n.d.e. "Creating Local Linkages." Consultato il 30 maggio 2023. <https://rrchnm.org/portfolio-item/creating-local-linkages/>.
- Roy Rosenzweig Center for History and New Media, n.d.f. "Data driven History." Consultato il 30 maggio 2023. <https://rrchnm.org/portfolio-category/data-driven-history/>.
- Roy Rosenzweig Center for History and New Media, n.d.g. "Public History." Consultato il 30 maggio 2023. <https://rrchnm.org/portfolio-category/public-history/>.
- Roy Rosenzweig Center for History and New Media, n.d.h. "Education." Consultato il 30 maggio 2023. <https://rrchnm.org/portfolio-category/education/>.
- Saramago, Josè. 2017. *Storia dell'assedio di Lisbona*, Milano: Feltrinelli.
- Stull, Rob. n.d. "What is Participatory Archiving?" Consultato il 30 maggio 2023. <https://ropa.umb.edu/what-is-participatory-archiving>.
- The September 11 Digital Archive. n.d. Consultato il 30 maggio 2023. <https://911digitalarchive.org/>.
- Tomasi, Francesca. 2022. *Organizzare la conoscenza: digital humanities e web semantico. Un percorso tra archivi, biblioteche e musei*. Milano: Editrice Bibliografica.
- Treccani. n.d. "Emozione." Consultato il 30 maggio 2023. <https://www.treccani.it/enciclopedia/emozione>.
- Valacchi, Federico. 2020. "L'archivistica del nulla." *AIDAinformazioni*, 1-2 (gennaio-giugno): 191-214.
- Valacchi, Federico. 2022. "The parts and the whole. Integrate knowledge." *JLIS.it* 13(3): 1-11. <https://doi.org/10.36253/jlis.it-477>.
- Vitali, Stefano. 2004. *Passato digitale. Le fonti dello storico nell'era del computer*. Milano: Mondadori.
- Vitali, Stefano. 2006. "Come si 'diventa digitali negli archivi'." *Bibliotime IX* (1), <https://www.aib.it/aib/sezioni/emr/bibtime/num-ix-1/vitali.htm>.
- Vitali, Stefano. 2022. "Archivistica e filosofia: le riflessioni di Valenti su informatica ed ermeneutica." In *Filippo Valenti, un intellettuale in archivio: le parole le carte, i libri*, a cura di Antonella Mulè, Angelo Spaggiari e Gilberto Zacchè. *Pubblicazione degli archivi di Stato Saggi 122*. Roma: Ministero della cultura.
- Yakel, Elizabeth. 2005. "Hidden collections in archives and libraries." *OCLC Systems & Services: International digital library perspectives* 21(2): 95-99. <https://doi.org/10.1108/10650750510598675>.

## Note e rubriche





## Recensione “L’innovazione in biblioteca” di Ferruccio Diozzi

Claudia Lanza\*

Il concetto dell’innovazione inteso come evoluzione che rende necessari cambiamenti radicali negli assetti sociali e culturali della società è la base di partenza da cui si muovono le pagine del recente volume di Ferruccio Diozzi *L’innovazione in biblioteca* edito da Editrice Bibliografica e pubblicato nel 2022. Autore di un elevato numero di pubblicazioni nel campo della Biblioteconomia e Scienza dell’Informazione, Diozzi riconferma l’importanza di analizzare la funzione delle biblioteche come luoghi fisici e digitali, *phygital*, da cui estrarre la conoscenza e come leve di gestione dei cambiamenti avvenuti nella società, in questo caso all’interno di una cornice storica contraddistinta da un progresso tecnologico che accelera e cambia l’accesso alle informazioni. Infatti, sulla base del legame indissolubile tra società e innovazione, l’autore sottolinea esplicitamente l’esigenza di interpretare i processi evolutivi sociali e culturali come spinte da cui possano emergere nuovi approcci in risposta alla dinamicità tecnologica. Diozzi offre in apertura una panoramica sulle diverse assunzioni di significato ricoperte dal termine “innovazione” partendo in primis dal contesto tecnologico. In questo primo excursus l’autore riflette in modo dettagliato sull’influsso dell’innovazione sull’introduzione di nuovi beni o servizi così come sui modelli di creazione di nuove tecniche di fruizione di nuovi prodotti per far fronte alle esigenze di utenze sempre più eterogenee e mutate nelle modalità di ricezione e consumo dei beni immateriali.<sup>1</sup>

L’analisi viene poi ampliata da un’accurata descrizione del quadro sociale su cui si basa la spinta all’evoluzione prendendo in considerazione il concetto di economia circolare, una chiave di lettura ripresa a più battute nel corso del volume, come strumento di soluzioni nuove e più sostenibili per rispondere ai bisogni delle collettività. Completano lo scenario altre due tipologie di innovazione: la cosiddetta *open innovation* che supporta, attraverso l’uso di

---

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

competenze esterne, il miglioramento dei processi e dei prodotti, e la *disruptive innovation* che rende possibile il superamento di metodi non più funzionali nel presente tramite l'introduzione di nuovi modelli di creazione e ricezione di prodotti e servizi. Su quest'ultimo passaggio l'autore si sofferma maggiormente per definire l'impatto del concetto dell'innovazione sulle biblioteche e, in particolare, sul contributo che essa ha avuto nel potenziamento della messa a disposizione delle risorse informative.

Nei capitoli che seguono la prima parte, dedicata ad una panoramica delle diverse tipologie di innovazione e all'introduzione sulla connessione tra tecnologie e biblioteche, l'autore enfatizza più concretamente gli elementi costituenti di nuove politiche per l'innovazione nel contesto biblioteconomico in seguito ai mutamenti avvenuti nel corso degli ultimi anni. L'attuale momento storico contrassegnato dalla pandemia da Covid-19 è uno degli scenari che viene puntualmente evidenziato nel corso del volume come un fenomeno che ha comportato non solo la trasformazione dei contesti lavorativi da luoghi fissi fisici a soluzioni interconnesse da remoto, ma anche, e parallelamente, nuove forme di creazione, gestione e fruizione dei dati, di conseguenza riducendo l'attavico divario digitale che diversificava le competenze degli utenti. Prendendo spunto dall'osservazione di una nuova necessaria conversione da luoghi fisici a ambienti digitali nella configurazione delle posizioni lavorative, una parte del volume è dedicata alla problematica relativa alla gestione del capitale umano, gli "infomedieri", figure che secondo l'autore dovranno aumentare la propria autodeterminazione, o *empowerment*, per migliorare la funzione di erogazione dei servizi delle biblioteche. A supporto dell'analisi che l'autore porta avanti sulle tecniche di miglioramento della consapevolezza della specificità professionale delle figure operanti nell'ambito biblioteconomico, un ulteriore elemento su cui si sofferma è il livello di competenze trasversali che un gruppo organico all'interno di una biblioteca dovrebbe possedere. Grazie a una serie di *soft skills* si rende possibile, secondo l'autore, la definizione di un contesto lavorativo caratterizzato da condizioni organizzative più efficienti che possano consentire l'affermazione di un modello di *business continuity*, riprendendo un concetto centrale del suo volume *Come garantire la continuità operativa in biblioteca* (Diozzi 2021), che supporti la continuità di processi gestionali e operativi.

Alcuni esempi applicativi vengono poi integrati dall'autore per riuscire a contestualizzare al meglio le nuove forme di interazione e accesso alle informazioni nelle biblioteche che diventano sempre di più luoghi digitali. Si pone l'attenzione altresì all'importanza delle biblioteche accademiche nei processi di valorizzazione dei beni pubblici presenti nel campo accademico, nei processi di disseminazione delle informazioni e condivisione dei risultati di ricerche universitarie, nella formazione continua e nelle forme di collaborazione con enti privati.

Risulta, infine, utile l’inserimento a fine volume di un Glossario che raggruppa i termini più significativi utilizzati dall’autore con le corrispettive definizioni che agevolano i processi interpretativi in riferimento al tempo in cui viene pubblicato il testo. Sarebbe stato interessante includere un approfondimento sulle nuove modalità di estrazione della conoscenza all’interno del sistema bibliotecario con alcuni esempi di tecnologie che ad oggi vengono utilizzate per l’accesso alle informazioni.

In sintesi, nel volume si evidenzia in modo chiaro come le nuove tecnologie abbiano comportato lo sviluppo di nuovi paradigmi nel rispondere alle esigenze della collettività. In questo scenario innovatore le biblioteche assumono un ruolo importante nella gestione dei processi di cambiamento pur con determinate sfide che l’autore interpreta come slanci da cui partire per rinnovarsi e rafforzare le proprie potenzialità relazionali sia nella medesima coordinazione del capitale umano che con gli utenti finali.

### **Riferimenti bibliografici**

Diozzi, Ferruccio. 2021. *Come garantire la continuità operativa in biblioteca*. Milano: Editrice Bibliografica.

Diozzi, Ferruccio. 2022. *L’innovazione in Biblioteca*. Milano: Editrice Bibliografica.



## TeKaMoLo ovvero le lingue come KOS, e viceversa

Claudio Gnoli\*

Tra i sistemi per l'organizzazione della conoscenza (KOS), soggetti, tesauri e classificazioni sono stati chiamati nella seconda metà del Novecento anche linguaggi di indicizzazione, e ancora oggi in francese vengono detti *langages documentaires*. Queste espressioni ci ricordano che, dopotutto, i KOS sono un tipo particolare di linguaggi, parzialmente artificiali perché destinati a particolari scopi, ma pur sempre dotati di molte caratteristiche comuni alle lingue parlate.

D'altra parte, già le lingue sono il primo dei modi con cui gli umani organizzano la conoscenza, suddividendo l'esperienza in un numero finito di concetti espressi da parole. Parlare è infatti prima di tutto assegnare a ciascun fenomeno un nome (*febbre, poiana, assemblea condominiale...*), il quale appartiene ad una rete implicita di gerarchie e associazioni. Studiata dalla semantica lessicale (Jezek 2005), essa assomiglia parecchio ai reticoli più formalizzati di concetti e relazioni che sottendono una classificazione o un'ontologia. I sostantivi della lingua vengono poi collegati con verbi, casi, preposizioni e congiunzioni che esprimono relazioni, analoghe alle combinazioni di concetti nei KOS.

Queste osservazioni suggeriscono tra l'altro che l'organizzazione della conoscenza si possa applicare anche alle reti semantiche studiate dalla linguistica, e magari ad una didattica basata sul riconoscimento delle strutture grammaticali come quella di M. Michela Sebastiani ed Emanuela Valenzano (2016): direzioni che probabilmente meritano di essere esplorate più a fondo.

Quello che nelle lingue naturali non c'è, ed è il motivo per cui risulta necessario sviluppare dei KOS, sono soprattutto tre cose:

(1) il controllo terminologico, che elimina le ambiguità dovute alle sinonimie e alle omonimie nella maggior parte dei KOS (con l'eccezione delle

---

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

parole-chiave sciolte o tag) associando ciascun concetto ad un termine univoco e offrendo opportuni rinvii (*bozzaghi*: vedi *poiane*);

(2) relazioni paradigmatiche esplicite, che permettono di percorrere a piacere le reti di concetti verso l'alto (termini più generici, classi superiori), verso il basso (termini più specifici, sottoclassi) o lateralmente (termini associati, faccette);

(3) un ordine sistematico, implicito o controllato da una notazione, che permette di visualizzare e scorrere i concetti in modi coerenti e utili, anziché secondo le bizzarrie dell'alfabeto.

Queste tre componenti permettono di passare dalle lingue naturali a quei linguaggi per scopi particolari che sono appunto i KOS. Ma la natura linguistica di questi ultimi rimane, e sebbene spesso trascurata può riemergere in diversi aspetti: ben lo sapevano autori come Maurice Coyaud (1966), W.J. Hutchins (1975) o Jacques Maniez (1987) sul quale è apparsa recentemente una voce dell'enciclopedia ISKO.

Anche Gérard Cordonnier (1951), teorico delle classificazioni sintetiche, si muoveva a cavallo fra lingue artificiali e naturali immaginando un sistema capace di funzionare come una lingua parlabile, caratteristica vantaggiosa anche per memorizzarne più facilmente la notazione. A questo scopo occorrerebbe che i simboli, che solitamente possono consistere di qualunque stringa di caratteri, risultassero pronunciabili. Questa funzione è stata realizzata da Douglas e Joy Foskett nella London Education Classification, prevedendo simboli che siano sempre sequenze consonante-vocale/i-consonante: così *RidMenLim* significa "metodo diretto, francese, scuola secondaria". Un'alternativa è progettare liberamente la notazione ottimizzandola per le sue consuete funzioni di ordinamento, e solo successivamente associare a ciascun simbolo un fonema o una sillaba che risulti pronunciabile in qualsiasi combinazione, perché ad esempio termina sempre per vocale: così nella Integrative Levels Classification (ILC) la faccetta *sb2e* "cibi, ristorazione" si può pronunciare /sabantœ/. Non a caso questo KOS è stato ispirato anche dalla precedente esperienza di sviluppo di una lingua logica chiamata liva. Simili problemi sono stati affrontati nel passato dagli sviluppatori di lingue ausiliarie internazionali, come l'esperanto, o di lingue filosofiche quale quella del vescovo seicentesco John Wilkins, che ha aspetti molto vicini alle classificazioni bibliografiche come già notò il grande Brian Vickery (1953).

Come abbiamo accennato, le relazioni che nelle lingue naturali sono espresse dai diversi tipi di complementi corrispondono sostanzialmente alle faccette e ai ruoli dei KOS sintetici: si può rendersene conto leggendo la guida Nuovo Soggettario (Biblioteca Nazionale Centrale di Firenze 2021). In questi KOS infatti ricorrono certe categorie fondamentali (parti, proprietà, processi, agenti, spazio, tempo...) che hanno degli equivalenti nei sistemi di casi di mol-

te lingue flessive, e meglio ancora nei casi profondi identificati dal linguista Charles Fillmore.

Talvolta nelle lingue naturali si ritrova perfino il principio classificatorio dell'ordine di citazione delle categorie: come le faccette vanno elencate secondo varianti della famosa formula PMEST di Ranganathan, così scrivendo in tedesco occorre seguire la formula equivalente TeKaMoLo (tempo, causa, modo, luogo) nel disporre i complementi indiretti e gli avverbi. Sono ammesse eccezioni solo quando si vuole enfatizzare uno dei complementi: una possibilità che ha senso anche nei KOS, dove promuovere una faccetta più vicino all'inizio della stringa ha l'effetto di renderla più rilevante ai fini dell'ordinamento sequenziale delle voci: così l'ordine standard tipo-parte-processo di "cibi, con uova, fritti" si può modificare in "cibi, fritti, con uova" nel caso che in un determinato documento la faccetta processuale risulti più rilevante.

Già Paul Otlet immaginava che la sua Classificazione decimale universale potesse essere impiegata linguisticamente, per esempio nei telegrammi, costruendo espressioni con classi qualificate da suffissi per soggetti, verbi e complementi come *35.633<sup>001</sup>*; *629.12313<sup>063.043</sup>*; *35.521.6<sup>002</sup>*; *615.935<sup>064</sup>*. che vorrebbe significare "il servizio sanitario trasporterà con imbarcazioni a vela i lavoratori volontari colpiti da dissenteria" (Dousa 2010; Institut international de bibliographie 1906)<sup>1</sup>.

Una differenza importante rispetto alle lingue naturali è che solitamente i KOS non esprimono affermazioni, del tipo "il cibo fritto aumenta il colesterolo", bensì classi senza tempo di soggetti, che vengono infatti espresse al plurale: "cibi fritti", dove tuttalpiù potremmo esplicitare il significato sottinteso "questo documento tratta di cibi fritti". Nell'esempio linguistico invece il tema ossia ciò di cui si parla ("il cibo fritto") è completato da un rema ("aumenta il colesterolo"), cioè da un'affermazione relativa ad un fatto osservato nella realtà: oltre alla classe si esprime dunque una sua relazione denotabile concretamente (deissi) perché si verifica qui o altrove o è relativa a un individuo particolare ("questo cibo") magari dotato di un nome proprio ("la torta di compleanno di Diana"). Nell'indicizzazione solitamente si intende esprimere soltanto il tema, ritenendo che sia esso ad interessare gli utenti che consultano gli indici. Tale assunto è stato però messo in discussione da un articolo di Bella Hass Weinberg (1988), secondo cui anche le specifiche affermazioni che un testo fa sul proprio tema possono essere rilevanti per gli studiosi.

Fra le relazioni espresse nelle classificazioni, Rick Szostak ritiene che in una prospettiva interdisciplinare le più importanti per gli utenti siano quelle di tipo causale. In fin dei conti, la conoscenza cerca innanzitutto di identificare i nessi causali fra i fenomeni: da che cosa un fenomeno è influenzato, da che cosa è derivato, ecc. Ad esempio, nella ILC è possibile costruire espressioni

---

<sup>1</sup> Sono grato a Tom per le segnalazioni.



come *fps85m(30sb(5oy))* che letteralmente significa “steroli, aumento, causato da cibi, friggere” e può essere recuperata sia cercando la notazione per “steroli” che cercando quella per “friggere”.

Per la verità, questo esempio non è ancora propriamente equivalente ad una frase del linguaggio naturale: per renderlo tale occorre aggiungere un deittico ossia un riferimento variabile alla realtà esterna, come *-V* “questi” o *-Y* “effettivamente”: si può allora avere *fps85m(30Ysb(5oy))* “steroli, aumento, è causato da cibi, friggere”. Di questo passo potremmo realizzare gli auspici di Wilkins e di Otlet e metterci a parlare in una lingua classificata.

## Riferimenti bibliografici

- Biblioteca Nazionale Centrale di Firenze. 2021. *Nuovo soggettario: guida al sistema italiano di indicizzazione per soggetto*, 2nd ed. Roma: Associazione Italiana Biblioteche; Firenze: Biblioteca nazionale centrale di Firenze. <https://doi.org/10.53263/9788878123465>.
- Cordonnier, Gérard. 1951. “Classification, classement, rangement, et sélection.” *CNOF: revue mensuel de l'organisation* 25 (5-6): 19-31.
- Coyaud, Maurice. 1966. *Introduction à l'étude des langages documentaires*. Paris: Klincksieck.
- Dousa, Thomas. 2010. “IORG Paul Otlet the UDC and the Origins of the Concept of A.” YouTube video. [youtu.be/hQaF5xJTgRc](https://youtu.be/hQaF5xJTgRc).
- Hass Weinberg, Bella. 1988. “Why Indexing Fails the Researcher.” *The Indexer: The International Journal of Indexing* 16: 3-6.
- Hutchins, William John. 1975. *Languages of Indexing and Classification: a linguistic study of structures and functions*. Stevenage, Eng: P. Peregrinus.
- Institut international de bibliographie. 1906. “Bulletin de l'Institut international de bibliographie.” no. 11: 92-99.
- Jezek, Elisabetta. 2005. *Lessico*. Bologna: Il Mulino.
- Maniez, Jacques. 1987. *Les langages documentaires et classificatoires*. Paris: Les éditions d'organisation.
- Sebastiani, M. Michela, e Emanuela Valenzano. 2016. *Educare alla lettura con i marcatori visivi*. Roma: Aracne.
- Vickery, Brian C. 1953. “The significance of John Wilkins in the history of bibliographical classification.” *Libri* 2: 326-43.



# AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 41

N. 1-2 – gennaio-giugno 2023

## Contributi

STEFANO ALLEGREZZA

*“Effimeri ma non troppo”*

CARLOS ALBERTO ÁVILA ARAÚJO

*A veritistic turn in information science?*

DIMITRI BRUNETTI

*Le raccomandazioni di Alba per la descrizione dei corrispondenti e la redazione dell'oggetto nel protocollo informatico*

MANOLA CHERUBINI,

FRANCESCO ROMANO, ANDREA BOLIOLI

*Intelligenza artificiale e informazione giuridica*

NATASCIA RALLI, ISABELLA STANIZZI,

MARLIES ALBER

*COVID-19 e lavoro terminologico: riflessioni a posteriori*

SALVATORE SPINA

*Handwritten Text Recognition as a digital perspective of Archival Science*

MARC TANTI

*Quelles sont les Fake News qui ont circulé lors de la pandémie de Covid-19 ?*

FEDERICO VALACCHI

*Se l'archivio è artificiale. Verso uno ius archivi partecipativo?*

## Note e rubriche

CLAUDIA LANZA

*Recensione “L'innovazione in biblioteca” di Ferruccio Diozzi*

CLAUDIO GNOLI

*TeKaMoLo ovvero le lingue come KOS, e viceversa*



mundaneum

In copertina

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

ISBN 979-12-5965-292-8



9 791259 652928

ISSN 1121-0095



9 770112 100950