

Digeat N.1 - 19 Marzo 2024

La conservazione by default degli archivi digitali è in contrasto con il GDPR?

Di Antonella Deiana

Abstract

L'Intelligenza Artificiale dovrebbe essere guidata da strumenti archivistici? Parliamo infatti di una tecnologia che si spinge ben oltre l'information retrieval e che ormai può essere considerata a tutti gli effetti un elemento chiave nella progettazione della sedimentazione documentale, giocando un ruolo strategico per disegnare processi vitali, ma troppo spesso dimenticati. Come sappiamo, la corretta conservazione dei documenti digitali ha origine nella fase di produzione e la cura di questo ecosistema è indispensabile per ciascun Ente. Il clima di (ri)scoperta dell'importanza della conservazione documentale è cresciuto e la **maturazione del mercato di servizi digitali** non può proseguire ignorando il peso che l'intelligenza artificiale sta assumendo nel campo. Proviamo riflettere su come incastrare correttamente i tasselli di un complesso mosaico.

Indice

- L'ecosistema documentale: dalla produzione alla conservazione
- Scarto e IA: un binomio possibile per la valorizzazione del patrimonio informativo
- Lo scarto by design in luogo della conservazione by default
- La funzionalità dello scarto rispetto al trattamento dei dati
- Verso un'intelligenza artificiale "educata" alla cultura dell'archivio

Il lungo percorso di digitalizzazione nazionale ha visto fin dall'introduzione delle Linee Guida AgID una crescente attenzione nei riguardi degli archivi, non solo da parte del mondo della Pubblica Amministrazione, ma anche delle aziende private.

Ci si è resi conto che la corretta conservazione dei documenti digitali ha origine nella fase di produzione e la cura di questo ecosistema è indispensabile per ciascun Ente. Il clima di (ri)scoperta dell'importanza della conservazione documentale è cresciuto di pari passo con la **maturazione del mercato di servizi digitali**, con particolare riferimento ai conservatori, che hanno effettuato un notevole lavoro in termini di adeguamento dei metadati.

La sedimentazione dei documenti nelle PA e nelle aziende private sta evolvendo (in)consapevolmente verso l'individuazione di **ulteriori tipologie documentarie** che nascendo digitalmente ed essendo gestite attraverso workflow di firma digitale diventano nuovi oggetti di conservazione. Se fino a questo momento l'attenzione era posta sui documenti fiscali, adesso assistiamo al moltiplicarsi delle classi documentali, più o meno organizzate, da destinare alla conservazione.

L'ecosistema documentale: dalla produzione alla conservazione

Il contesto di produzione dei documenti rende più o meno semplice la successiva conservazione, in quanto **il processo non è sempre lineare**, ovvero diretto attraverso gli stessi applicativi. Se pensiamo a produzioni di documenti in ambito completamente destrutturato da un contesto di utilizzo degli applicativi, **la conservazione potrebbe (e spesso così è) essere considerata e percepita come un ulteriore carico di lavoro**. Infatti, la creazione di un documento con sistemi di video scrittura, la sua successiva trasformazione in un formato stabile e gli scambi che ne conseguono per la raccolta delle firme digitali, costituisce già un notevole lavoro, al quale deve essere necessariamente aggiunta la parte di metadattazione richiesta all'atto dell'upload del documento presso il conservatore (attività che potrebbe essere gestita automaticamente se il documento fosse generato con applicativi, ne è un esempio la tipologia dei contratti).

Un concetto sul quale si devono ancora effettuare passi avanti è la **fascicolazione**, già prevista nell'art. 65 del CAD, come **strumento strategico** (nonché indispensabile) per fornire informazioni sul **legame** che esiste tra i documenti relativi ad uno stesso procedimento/attività, per velocizzare il reperimento delle informazioni e per consentire lo scambio delle stesse con altri sistemi.

La struttura del fascicolo è **tutt'ora inutilizzata dai più**, in luogo, al contrario, delle cartelle di rete dove risiede una molteplicità di aggregazioni documentarie non strutturate.

Possiamo quindi dire che se il documento è **prodotto** secondo adeguate procedure, **gestito** all'interno di un contesto organizzato, anche mediante strumenti di protocollazione e **strutturato all'interno di fascicoli** (qualora non repertoriato), potremmo dirci in regola solo se a tutto questo sia associata la fase di conservazione.

Ancora, nelle organizzazioni che gestiscono correttamente una mole sempre crescente di documenti (dalla produzione fino alla conservazione), manca spesso un tassello importante: lo **scarto della documentazione** per la quale sono cessati i termini di ritenzione e che pertanto non rivestono più interesse dal punto di vista giuridico, amministrativo e storico. L'individuazione e la definizione dei **termini di conservazione** è spesso più complessa rispetto a quanto contrattualmente previsto con il conservatore digitale.

Scarto e IA: un binomio possibile per la valorizzazione del patrimonio informativo

Contrariamente a quanto si può immaginare lo "scarto" ha una funzione preservativa e nulla toglie all'archivio e alla sua integrità, ma **ne favorisce la corretta sedimentazione e la successiva fruizione**. Erroneamente siamo portati a preservare indistintamente tutto quello che è nativo digitale e che ha una firma, ma dall'altra parte ci scontriamo con l'esigenza di avere:

- archivi fruibili;
- costi nel tempo adeguati al loro mantenimento;
- accessibilità delle informazioni;
- risparmio di tempo nel reperire i documenti ricercati.

Nel Piano Triennale per l'informatica nella PA (ed. 2024-2026), al cap. 5 "Dati e Intelligenza Artificiale", l'Intelligenza Artificiale viene individuata prioritariamente come strumento volto "alla valorizzazione del patrimonio informativo". Attraverso l'analisi dei dati relativa ai documenti prodotti dalla pubblica amministrazione, potrebbe trasformarsi in uno strumento in grado di **guidare il processo di selezione di documenti che diventeranno oggetto di scarto** raggiunti i limiti di tempo sulla base dei Piani di Conservazione.

I documenti, indipendentemente dal supporto sul quale sono creati, sono **oggetto di selezione periodica**. Si ricorda che la procedura di scarto è prevista non solo dal GDPR, ma anche dal Codice

dei Beni culturali, previa autorizzazione delle Soprintendenze.

Nelle Linee Guida lo scarto dei documenti digitali (cap. 4.6) è addirittura previsto come “obbligo da parte delle Pubbliche Amministrazioni” che devono “individuare e pubblicare i tempi di versamento, le tipologie documentali trattate, i metadati, le modalità di trasmissione dei Pacchetti di Versamento e le tempistiche di selezione e scarto dei propri documenti informatici”.

In mancanza di selezione periodica **milioni di documenti sono destinati a rimanere nella memoria a tempo indeterminato**, mentre al contrario solo una ridotta percentuale necessita di essere sottoposta ad un processo di conservazione permanente.

Le logiche che in passato hanno guidato la sedimentazione dei documenti, difficilmente hanno contemplato all'atto dell'analisi di imponenti processi gestiti digitalmente, la possibilità di automatizzare le attività di selezione e di predisposizione di alert atti a favorire l'individuazione dei documenti scartabili. **In mancanza di questa pianificazione attualmente l'unica possibilità è procedere con delle selezioni manuali**, che impongono un notevole aggravio di tempo, per eseguire queste operazioni e spesso per questo si preferisce lasciare in conservazione.

In particolare per le PA la sedimentazione dei documenti nativi digitali e la loro successiva conservazione (nella maggior parte dei casi senza indicazioni specifiche inerenti ai termini di conservazione da inserire nei metadati) ha avuto inizio con l'entrata in vigore del DPR 445/2000 e con l'arrivo nel 2004 del **protocollo informatico**.

Lo scarto by design in luogo della conservazione by default

All'atto dell'introduzione del protocollo informatico, non sono state fatte le riflessioni che oggi rendono cogente la necessità di avere archivi funzionali e snelli. A distanza di 20 anni dalla sua introduzione in molti casi ci si rende conto che deve essere intrapresa una differente strada, atta anche a favorire lo **scarto by design** in luogo della conservazione **by default**.

Si rende necessaria l'associazione degli opportuni **metadati relativi a termini di conservazione**, non tanto delle serie omogenee di documenti, quanto piuttosto nella PA, nei singoli documenti relativi al “carteggio amministrativo” o nell'unità documentaria e questo nella prospettiva della **long preservation**, finalizzato alla valorizzazione del nucleo di documenti che devono essere conservati a tempo **indeterminato**.

Nelle aziende private la conservazione digitale prevalente è quella delle tipologie specifiche di documenti fortemente strutturati, che nella maggior parte dei casi vengono generati attraverso degli applicativi, es. Fatture attive e passive, LUL, CU, scritture contabili (registro IVA, libro giornale...).

Relativamente diffusa è invece la conservazione relativa a quello che potremmo definire “carteggio”.

Anche in questo caso, in processi più articolati, nei quali **non vengono selezionati in modo lineare intere annualità di documenti**, l'attività di selezione risulta essere gravosa, in funzione del fatto che **per un singolo processo spesso vengono utilizzati più applicativi**, all'interno dei quali restano archiviati gli stessi documenti che sono anche oggetto di conservazione. La selezione di quanto è in conservazione digitale deve necessariamente prevedere anche **l'eliminazione degli stessi documenti sui singoli applicativi rendendo in questo modo esponenziali le attività di selezione**, che sono peraltro poco lineari, non essendo automatizzate e pertanto possono essere suscettibili di un margine di errore.

La funzionalità dello scarto rispetto al trattamento dei dati

La finalità dello scarto è quella di avere un archivio funzionale rispettoso anche del trattamento dei dati. Ma purtroppo quello che è stato fatto fino a questo momento è irreversibile, pertanto, ci confrontiamo con una situazione di fatto che richiede non solo l'individuazione dei documenti che non sono più necessari (in quanto hanno perso il loro valore giuridico-probatorio, amministrativo e di gestione) e non rivestono alcun interesse dal punto di vista storico, ma anche una sempre maggior consapevolezza nei riguardi della gestione degli archivi correnti, per non incorrere nelle stesse problematiche.

La sedimentazione dei documenti, soprattutto nelle PA, ma non solo, nell'ambito di un archivio è guidata dal **titolario di classificazione** e dal **piano di conservazione**, entrambi gli strumenti hanno pari valore e non sono alternativi tra di loro. Se la sedimentazione organizzata in modo oggettivo è guidata dal titolario di classificazione, il piano di conservazione determina quello che sarà l'archivio, al termine delle operazioni di selezione periodica

Ripercorrendo la normativa le attività di scarto sono già previste dal DPR 445/2000, art. 68 attraverso la predisposizione del piano di conservazione, nelle Linee Guida si configura l'intero ciclo di vita del documento:

- produzione;
- gestione;
- conservazione;

Si ritiene talmente meritevole questo argomento da dedicargli un interno capitolo **“selezione e scarto dei documenti informatici”**.

Il responsabile del servizio per la gestione dei flussi documentali e degli archivi così come individuato nel dpr 445/2000 è “un dirigente ovvero un funzionario, comunque in possesso di idonei requisiti professionali o di professionalità tecnico archivistica acquisita a seguito di processi di formazione definiti secondo le procedure prescritte dalla disciplina vigente”, il quale opera d'intesa con il responsabile della conservazione digitale, che nelle PA (Linee Guida cap. 4.5), è individuato formalmente un “dirigente o un funzionario interno formalmente designato e in possesso di idonee competenze giuridiche, informatiche ed archivistiche”.

Per il settore del privato il responsabile della conservazione è un soggetto anche esterno all'organizzazione in possesso di idonee competenze giuridiche, informatiche ed archivistiche.

Rispetto alla gestione analogica dei documenti in entrambi i settori il responsabile della conservazione “definisce e attua politiche complessive del sistema di conservazione”.

Nelle PA spesso le due responsabilità sono coincidenti, anche se le competenze tecniche e informatiche necessarie non sono né da sottovalutare, né da improvvisare.

Non basta infatti conoscere i tempi di conservazione dei documenti per definire un manuale di conservazione.

La definizione dei termini di conservazione dei documenti, attraverso piani di conservazione o, per le aziende, di Manuali della conservazione, consente al responsabile della conservazione, di avere un quadro dettagliato (se definito bene a monte) delle attività legate allo scarto che devono poste in essere, **sempre valutando il contesto attraverso un confronto** con i differenti settori dell'azienda e della PA, con il DPO e nel caso di una PA le richieste di nulla osta alla Soprintendenza archivistica o alla Commissione di Vigilanza.

Infine “I documenti e le aggregazioni documentali informatiche sottoposti a scarto nel sistema di conservazione devono essere distrutti anche in tutti i sistemi gestiti dal Titolare dell'oggetto di conservazione”, solo con questa ultima attività di rimozione dei documenti non solo dall'ambiente di

conservazione, ma anche dagli ambienti di archiviazione potrà essere completata l'attività di selezione.

Verso un'intelligenza artificiale “educata” alla cultura dell'archivio

Da anni ormai penso all'intelligenza artificiale “educata” e al suo utilizzo per produrre degli elenchi di scarto e pertanto per selezionare i documenti che non necessitano di ulteriore conservazione. Le consuetudini del soggetto produttore, le modifiche agli strumenti archivistici, l'aggiornamento costante dei piani di conservazione si riflettono stabilmente in un archivio e questo servirà a guidare le informazioni che a seguito di un'analisi andranno a costituire il patrimonio di base relativo ad un apprendimento progressivo dell'IA.

Sarebbe bello pensare che possa esserci al pari della certificazione di processo per la digitalizzazione dei documenti, un processo di scarto certificato.

La difficoltà maggiore sarà la personalizzazione ovvero la standardizzazione, ogni PA e ogni azienda dovranno effettuare un'analisi preliminare per definire un set di metadati utili per la selezione partendo quindi dagli strumenti (titolario e piano di conservazione) finalizzata alla predisposizione dell'elenco di scarto. A tal fine particolare attenzione dovrà essere prestata: ai formati, alla classificazione, ai fascicoli e agli oggetti dei documenti .

Questa per me sarebbero i “desiderata” per tutta la documentazione pregressa, avendo però cura per gli archivi in formazione, di indicare “a monte” per ogni singolo documento o più esteso per tipologia documentale i termini di conservazione.

Se progettiamo i nuovi archivi dobbiamo **prevedere lo scarto by design**, sulla base delle differenti tipologie documentali e dei singoli documenti.

L'Intelligenza Artificiale dovrebbe essere guidata dagli strumenti archivistici, se questi sono stati adottati correttamente nell'ambito dell'organizzazione alla quale fanno riferimento possono giocare un ruolo strategico per disegnare processi vitali, ma troppo spesso dimenticati.