



SBA - Corso di Formazione
*La valorizzazione del patrimonio culturale: la gestione dei
progetti di digitalizzazione e la digital library*

12 ottobre 2023

La conservazione del patrimonio culturale e dei prodotti della ricerca digitali

Chiara Storti



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Conservazione del digitale: definizione e concetti introduttivi



Digital Preservation and Nuclear Disaster: An Animation
@wepreserve

Una definizione

"Digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary".

Digital Preservation Handbook, 2nd Edition,
<https://www.dpconline.org/handbook>,
Digital Preservation Coalition © 2015.



Digital**Preservation**Coalition

La conservazione è un processo

Nel mondo analogico

insieme delle procedure che servono a rallentare il naturale decadimento fisico dei materiali che costituiscono un documento

Nell'ecosistema digitale

insieme delle procedure che servono a garantire

- **Autenticità**
- **Integrità**
- **Accessibilità**
- **Intellegibilità**

della risorsa nel periodo necessario o in quello stabilito

Glossario

AGID - “Linee Guida sulla formazione, gestione e conservazione dei documenti informatici” (Allegato 1)

Autenticità

Caratteristica in virtù della quale **un oggetto deve considerarsi come corrispondente a ciò che era nel momento originario della sua produzione**. Pertanto un oggetto è autentico se nel contempo è integro e completo, non avendo subito nel corso del tempo o dello spazio alcuna modifica non autorizzata. L'autenticità è valutata sulla base di precise evidenze.

Integrità

Caratteristica di un documento informatico o di un'aggregazione documentale in virtù della quale risulta che essi **non hanno subito nel tempo e nello spazio alcuna alterazione non autorizzata**. La caratteristica dell'integrità, insieme a quella della completezza, concorre a determinare la caratteristica dell'autenticità.

Leggibilità*

Caratteristica di un documento informatico che garantisce la **qualità di poter essere decodificato e interpretato** da un'applicazione informatica.

*in questa definizione sono ricompresi i concetti, in realtà distinti, di leggibilità e intellegibilità.

Accesso

Operazione che consente di prendere visione dei documenti informatici.

Un'altra definizione

*"the ability to sustain the accessibility, understandability, and usability of digital objects in the distant future regardless of changes in technologies and in the '**designated communities**' that use these digital".*

Digital Preservation in Libraries: Preparing for a Sustainable Future
by Jeremy Myntti, Jessalyn Zoom, ALA Editions, 2018.

Alcuni punti fermi

- Conservare il digitale non significa fare un backup.

Alcuni punti fermi

- Il processo di conservazione inizia al momento della formazione della risorsa.
- L'*archiviabilità* deve essere una condizione nativa della risorsa digitale.

Alcuni punti fermi

- La conservazione è un processo cooperativo.

La cooperazione è necessaria al fine di garantire sia l'efficienza e l'economicità nella gestione dei processi che la conservazione del maggior numero di risorse destinate alla conservazione.

Alcuni punti fermi

- La conservazione è un processo in divenire: muta in base al contesto tecnologico, culturale, amministrativo e normativo in cui si colloca.
- La conservazione è un processo che deve essere documentato.

Cosa conserviamo

Le risorse digitalizzate

- Le rappresentazioni informatiche di risorse nativamente analogiche



Le risorse native digitali

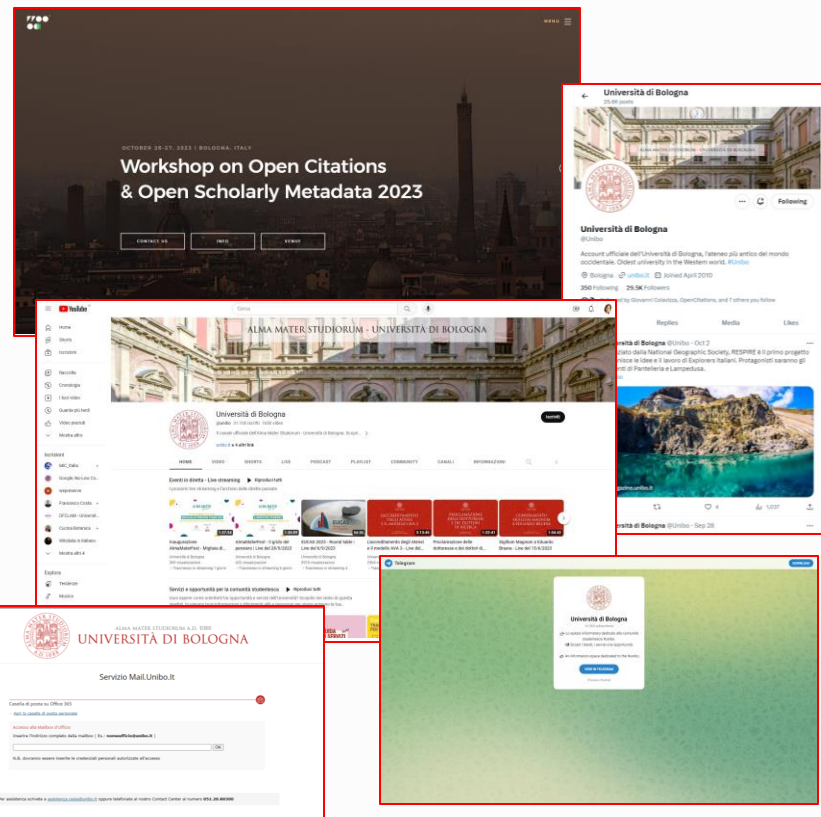
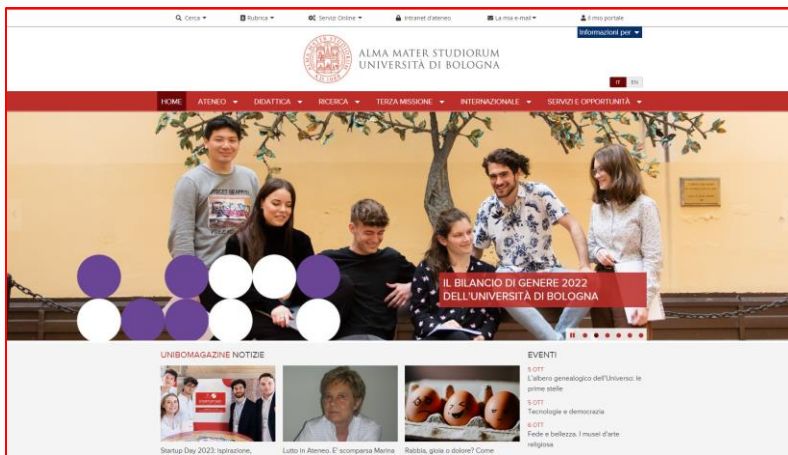
- Le risorse diffuse principalmente o esclusivamente in formato digitale

The screenshot shows the homepage of the AMSActa Institutional Research Repository. At the top left is the Alma Mater Studiorum University of Bologna logo. The main title is "AMSActa Institutional Research Repository" with the AlmaDL logo below it. A search bar is visible on the right. The central text describes AMSActa as an open institutional archive for researchers, supporting Open Access. Below this, there is a section for "ULTIMI DOCUMENTI" (Latest Documents) featuring a paper dandelion graphic. The first document listed is by Difore, Mariabella, Alamo, Yemshaw Ades (2022) titled "ECLPSE, WP1. Systems for agent simulation, Task 1.2. Synthesis and characterization of ECL active ultra-nanoparticles, VL University availability restricted".

The screenshot shows the homepage of the AMSDottorato Institutional Doctoral Theses Repository. It features the same Alma Mater Studiorum University of Bologna logo and title as AMSActa. A search bar is on the right. The central text describes the repository for doctoral theses. Below, the "ULTIMI TESI" (Latest Theses) section is visible, with the first entry by Chappuis, Giovanni (2022) titled "Interplay between national and EU law: time limits on cross-border joint proceedings and their impact on the free circulation of judgments".

The screenshot shows the homepage of AlmaDL Journals, Open Access Scientific Journals. It includes the AlmaDL logo and a search bar. The main text states that AlmaDL Journals offers assistance and support for the online publication of peer-reviewed open access journals. Below this, there are two featured journals: "Almatourism - Journal of Tourism, Culture and Territorial Development" and "Annali Sismondici". The Almatourism journal is described as an international platform for study and research on culture and territorial development. The Annali Sismondici journal is a peer-reviewed journal of the Association for Scientific Studies, publishing essays, unpublished documents, and reviews.

Le risorse informative del Web



Cosa dovremmo conservare

Oggetti digitali che hanno generalmente cinque attributi:

- contenuto (content)
- contesto (context)
- struttura (structure)
- aspetto (appearance)
- funzionalità (behavior /functionality)

Cosa conserviamo

*"Una parte del desiderio di bassa risoluzione è figlio di questa mediazione fra il nostro spirito archivistico e la contingenza del reale:
in altre parole non possediamo cantine sufficientemente grandi".*

M. Mantellini, *Bassa risoluzione*, Einaudi 2018

Cosa conserviamo

Anche nell'ecosistema digitale si presentano problemi di:

- spazio
- capacità di ordinamento delle risorse
- capacità di recupero delle informazioni

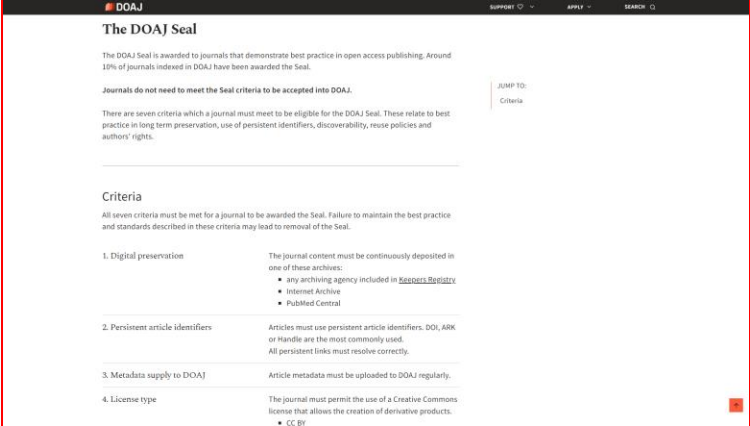
Cui si sommano problematiche emergenti come quella della sostenibilità ambientale.



Pozzo librario della BNCF

Cosa conserviamo?

- Quanto previsto (o consentito) da leggi, normative e regolamenti, sia nazionali che sovranazionali - in particolare in relazione al patrimonio culturale e alla ricerca scientifica > la **sfida legale**
- Quanto testimonia la *mission*, l'organizzazione e le attività di una istituzione > la **sfida organizzativa**
- Quanto è consentito sulla base delle risorse economiche, umane, strumentali a disposizione dell'istituzione o delle istituzioni deputate alla conservazione? > la **sfida economica**
- Quanto è consentito dalle tecnologie? > la **sfida tecnologica**



The screenshot shows the DOAJ Seal website. The main heading is "The DOAJ Seal". Below it, there is a paragraph explaining that the seal is awarded to journals demonstrating best practice in open access publishing. A note states that 10% of indexed journals have been awarded the seal. A section titled "Journals do not need to meet the Seal criteria to be accepted into DOAJ" follows. Below this, a "Criteria" section lists seven requirements for a journal to be eligible for the seal. The criteria are: 1. Digital preservation (content must be deposited in archives like Internet Archive or Pubmed Central), 2. Persistent article identifiers (DOI, ARK, or Handle), 3. Metadata supply to DOAJ, and 4. License type (must be Creative Commons CC BY).

DOAJ

The DOAJ Seal

The DOAJ Seal is awarded to journals that demonstrate best practice in open access publishing. Around 10% of journals indexed in DOAJ have been awarded the Seal.

Journals do not need to meet the Seal criteria to be accepted into DOAJ.

There are seven criteria which a journal must meet to be eligible for the DOAJ Seal. These relate to best practice in long term preservation, use of persistent identifiers, discoverability, reuse policies and authors' rights.

[JUMP TO: Criteria](#)

Criteria

All seven criteria must be met for a journal to be awarded the Seal. Failure to maintain the best practice and standards described in these criteria may lead to removal of the Seal.

- Digital preservation**

The journal content must be continuously deposited in one of these archives:

 - any archiving agency included in [Repecs Registry](#)
 - Internet Archive
 - Pubmed Central
- Persistent article identifiers**

Articles must use persistent article identifiers. DOI, ARK or Handle are the most commonly used. All persistent links must resolve correctly.
- Metadata supply to DOAJ**

Article metadata must be uploaded to DOAJ regularly.
- License type**

The journal must permit the use of a Creative Commons license that allows the creation of derivative products.

 - CC BY

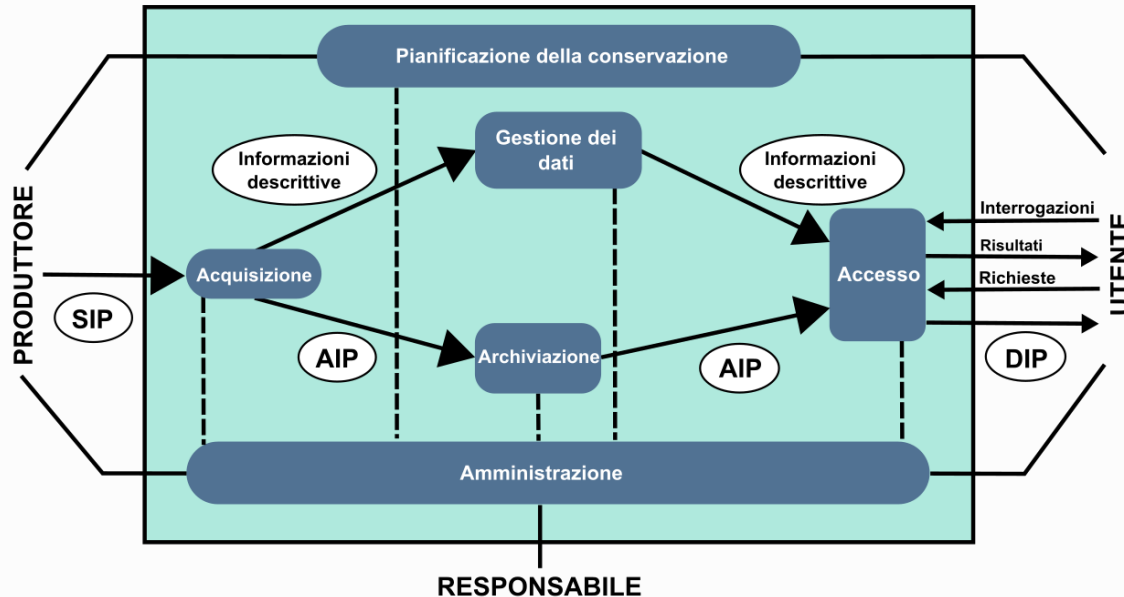
Cosa conserviamo

"La domanda da porsi non è tanto ciò che può essere conservato quanto ciò che non dovrebbe andare perso".

Digital Preservation Handbook, 2nd Edition, 2015

Modelli e standard di riferimento

Lo standard ISO 14721:2012 Space data and information transfer systems – Open archival information system (OAIS) – Reference model



Rappresentazione dell'
OAIS Environment

Di Mathieualexhache (original work); Mess (SVG conversion & Italian translation) - File:Schema fonctionnel modele oais.jpg, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=98896007>

Lo standard ISO 14721:2012

Space data and information transfer systems – Open archival information system (OAIS) – Reference model

*The term **open** refers to the fact that the reference model was developed and released in open public forums, in which any interested party was encouraged to participate. It does not refer to, or make any suppositions about, the level of accessibility associated with an archive.*

*An **archival information system** is ‘an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community’ (OAIS, 2012, 1-1)”*

*"The Open Archival Information System (OAIS) Reference Model: Introductory Guide" (2nd Edition)
By Brian Lavoie, DPC - Technology Watch Report, 2014 - ISSN 2048-7916 - DOI: <http://dx.doi.org/10.7207/twr14-02>*

Lo standard ISO 14721:2012

Space data and information transfer systems – Open archival information system (OAIS) – Reference model

Le responsabilità obbligatorie di un archivio OAIS

- Negoziare e **accettare informazioni adeguate dai produttori** dell'informazione.
- Riuscire ad **avere un controllo sufficiente dell'informazione** per raggiungere obiettivi di conservazione a lungo termine.
- **Definire l'ambito della comunità di utenti dell'archivio.**
- **Assicurarsi** che le informazioni conservate siano comprensibili in forma indipendente per la comunità di utenti, nel senso **che le informazioni possono essere comprese dagli utenti senza il supporto del produttore** di queste ultime.
- Seguire delle politiche e strategie documentate per **far sì che l'informazione venga conservata** a dispetto di ogni ragionevole eventualità, e che non si verifichino cancellazioni ad hoc.
- Rendere l'informazione conservata disponibile per la comunità di utenti, e **permettere la diffusione di copie autentiche dell'informazione** conservata nella sua forma originale, o in una forma che può essere riconducibile all'originale.

Lo standard ISO 14721:2012

Space data and information transfer systems – Open archival information system (OAIS) – Reference model

Individua **3 soggetti** che interagiscono tra loro e con l'archivio OAIS:

1. Produttore
2. Responsabile
3. Utente

Si tratta di ruoli "funzionali" non "organizzativi" quindi ciascun ruolo può essere ricoperto da una o più organizzazioni e ciascuna organizzazione può ricoprire ruoli diversi in diversi ambienti OAIS che, a loro volta, possono essere o meno collegati tra loro.

Lo standard ISO 14721:2012

Space data and information transfer systems – Open archival information system (OAIS) – Reference model

Individua **6 servizi di alto livello**:

1. Ingest
2. Archival Storage
3. Data Management
4. Preservation Planning
5. Access
6. Administration

7. [Common Services]

Lo standard ISO 14721:2012

Space data and information transfer systems – Open archival information system (OAIS) – Reference model

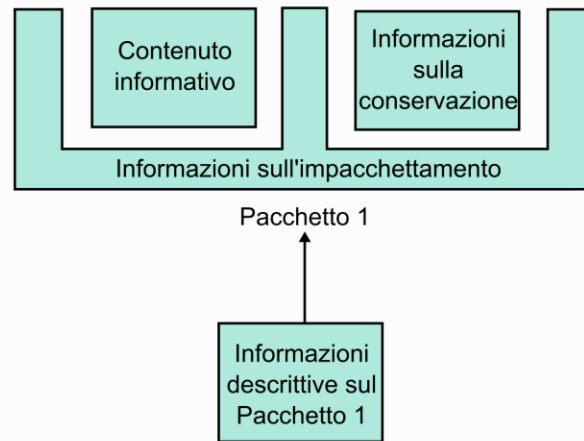
Il modello OAIS gestisce pacchetti informativi.

Un **pacchetto informativo*** è composto dall'oggetto da conservare e da tutti i metadati utili alla conservazione, all'accesso e alla intellegibilità dei dati nel lungo periodo.

Il pacchetto informativo ha 3 varianti:

- **SIP = Submission Information Package**
- **AIP = Archival Information Package** (che può essere specializzato in AIU e AIC)
- **DIP = Dissemination Information Package**

*si tratta di pacchetti "logici" non per forza "fisici"



Di Mess - Opera propria, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=98896009>

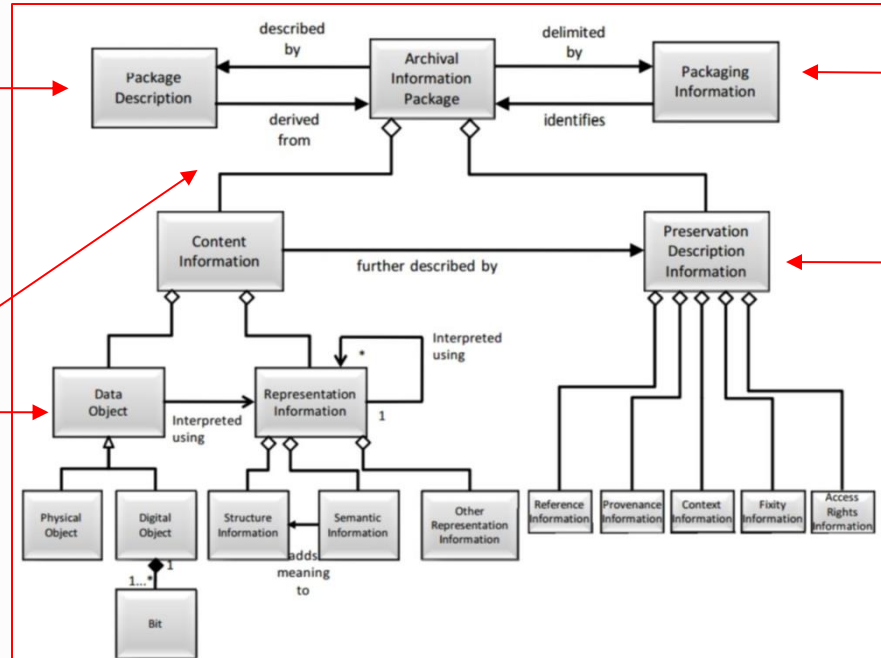
Lo standard ISO 14721:2012

Space data and information transfer systems – Open archival information system (OAIS) – Reference model

Per il recupero del pacchetto da parte dei *Consumers*

Tipicamente:

- [METS](#)
- [BagIT](#)
- [WARC](#)



Per tenere insieme "Content information" e "Preservation Description Information"

Tipicamente descritti in base allo [Standard PREMIS](#)

Figure 4-18: Archival Information Package (Detailed View)

https://wiki.dpconline.org/index.php?title=4.2.2.3_The_Archival_Information_Package

Lo standard ISO 14721:2012

Space data and information transfer systems – Open archival information system (OAIS) – Reference model

OAIS è un **MODELLO DI RIFERIMENTO** pertanto:

- ha un alto livello di astrazione: definisce concetti, non standard;
- non fornisce informazioni sulle tecnologie da adottare;
- è adattabile a qualsiasi contesto linguistico, culturale, legale ecc. e a qualsiasi dominio;
- potrebbe essere applicato alla gestione di qualsiasi oggetto informativo (teoricamente anche analogico!);
- la sua *compliance** da parte degli archivi potenzialmente facilita la cooperazione, consentendo di razionalizzare i costi della Conservazione.

Qualche esempio di **implementazione**:

- [The Inter-university Consortium for Political and Social Research \(ICPSR\)](#)
- [OCLC CONTENTdm®](#)
- [Rosetta](#)
- [LOCKSS](#)

Metadati per la conservazione:

lo standard PREMIS - PREservation Metadata Implementation Strategies

- Inizialmente definito da un gruppo di lavoro partecipato da OCLC e RLG, oggi lo standard è ospitato e mantenuto dalla Library of Congress.
- Definisce le entità coinvolte nel Processo di Conservazione e le relazioni tra di esse.
- Si compone di uno Schema e di un Vocabolario.
- È uno standard di alto livello (solitamente viene implementato in XML).

Metadati per la conservazione: lo standard PREMIS - PREservation Metadata Implementation Strategies

The PREMIS Data Model

The PREMIS Data Dictionary defines **semantic units**. Each semantic unit defined in the Data Dictionary is mapped to an entity that is organized within a simple data model. A semantic unit can therefore be understood as a property of an entity. The model defines four entities important in regard to digital preservation activities: Objects, Events, Agents and Rights.⁴ [Figure 1](#) provides a graphical illustration of the PREMIS Data Model.

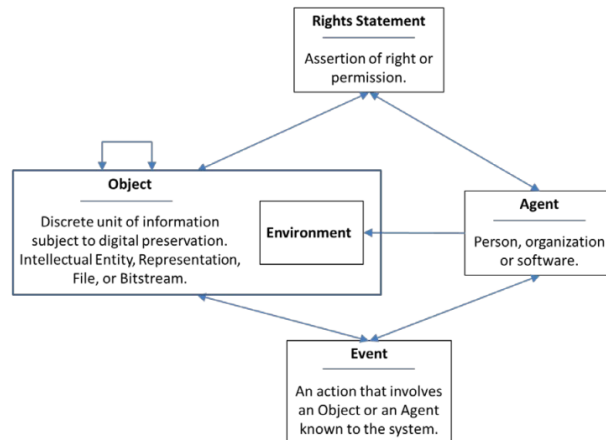


Figure 1: The PREMIS Data Model

PREMIS Data Dictionary for
Preservation Metadata,
Version 3.0

La codifica e lo scambio delle informazioni archiviate: lo standard METS - Metadata Encoding and Transmission Standard

Un documento METS è costituito da sette sezioni principali:

- `<metsHdr>` **Intestazione METS**: contiene i metadati che descrivono il documento METS stesso, includendo alcune informazioni quali autore, editore, ecc.;
- `<dmdSec>` **Metadati descrittivi**: contiene al suo interno i metadati descrittivi, o punta a metadati esterni specifici per ciascun dominio;
- `<amdSec>` **Metadati amministrativi**: contiene sia informazioni sui file creati, che conservano anche i diritti di proprietà intellettuale, sia metadati riguardanti l'oggetto di origine da cui deriva l'oggetto digitale, sia informazioni sulla provenienza dei file e sulle relazioni degli oggetti digitali;
- `<fileSec>` **Sezione file**: è una lista di tutti i file contenente le versioni elettroniche dell'oggetto digitale;
- `<structMap>` **Mappa strutturale**: descrive la struttura gerarchica, fisica o logica, dell'oggetto digitale e collega i file e i metadati loro associati;
- `<structLink>` **Link strutturali**: permette di gestire collegamenti ipertestuali tra nodi nella gerarchia definita nella Mappa strutturale;
- `<behaviour>` **Comportamento**: può essere usato per associare comportamenti (per esempio, la visualizzazione) con il contenuto dell'oggetto METS.

[MiC - PND](#)

[Linee Guida per la
digitalizzazione del patrimonio
culturale](#)

La gestione dei pacchetti informativi: il profilo METS-PREMIS

- Il METS è agnostico riguardo agli schemi di metadati descrittivi <dmdSec> o amministrativi <amdSec> che i suoi implementatori possono utilizzare.
- Il PREMIS è stato quindi registrato come schema di metadati riconosciuto da utilizzare per la definizione dei metadati di conservazione (da considerarsi come una specifica tipologia di metadati amministrativi) in METS.

Buone pratiche di conservazione del digitale

Un approccio basato sul ciclo di vita del digitale

"The first line of defense against loss of valuable digital information rests with the creators, providers and owners of digital information."

Waters, D and Garrett, J., 1996. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* commissioned by the Commission on Preservation and Access and the Research Libraries Group. Washington, DC: Commission on Preservation and Access.
Available: <https://www.oclc.org/content/dam/research/activities/digpresstudy/final-report.pdf>

Creare risorse archiviabili

- Scegliere **formati** di dati **standard** e **software** di creazione e gestione **consolidati e ben documentati**.
- Assicurarsi che i dati siano usabili ed accessibili: quanto facilmente li posso ricercare e vi posso accedere? **Dati accessibili ed usabili sono quasi sempre anche dati archiviabili!**
- Assicurarsi di **avere la titolarità legale per la gestione dei dati** e per effettuare tutte le operazioni necessarie alla loro conservazione (copie, conversioni ecc.).

Digital Asset Register

"Un Digital Asset Register sarà veramente utile nella valutazione dell'ampiezza e della significatività della collezione, nell'identificazione delle priorità e pianificazione di azioni di conservazione digitale."

- *Qual è l'argomento della collezione?*
- *Da dove proviene e qual è la sua funzione?*
- *Dov'è archiviata e quali tipi di supporti sono stati usati?*
- *Perché è in conservazione?*
- *Chi ne è il responsabile; chi sono gli utenti; a chi sono rivolti i dati?*
- *Come avviene l'accesso ai dati?*
- *Come potrebbero cambiare e crescere i dati in un futuro prossimo?*

Attività preliminari

1. Messa in sicurezza dei dati (dei flussi di bit)

- Scansione virus o malware
- Verifica completezza dei dati (rispetto anche ad Indici ecc.)
- Verifica integrità e qualità dei dati
- Sostituzione dei dati danneggiati e integrazione di quelli mancanti
- Normalizzazione dei formati (preferendo formati standard es. PDF/A, JPEG, TIFF ecc.)
- Creare una lista dei file presenti nella collezione contenente le informazioni essenziali quali
 - Denominazione
 - Posizione
 - Dimensione
 - Tipologia di formato
 - Checksum

Attività preliminari

2. Stabilizzazione dei dati

- Fare delle copie su supporti diversi da conservare in luoghi differenti e geograficamente distanti (sia online che offline):
 - almeno una copia deve essere facilmente accessibile per poterne verificare regolarmente l'integrità (verifica della *fixity*) e per essere eventualmente resa fruibile attraverso sistemi di replay;
 - le altre copie (*Dark Archive*) possono essere conservate in sistemi meno accessibili e ad alta affidabilità.

Attività preliminari

3. Verifica della fixity (integrità dei dati)

- Riesaminare i dati con regolarità:
 - ricalcolare i checksum e identificare i dati che risultano danneggiati;
 - recuperare le copie dei file danneggiati e ripararli, se possibile;
 - effettuare test di recupero dei backup.

Attività preliminari

4. Documentazione dei processi

- I contenuti della collezione
- Gli strumenti
- I flussi di lavoro

Anche per rendere indipendenti i processi dalle competenze / conoscenze di una singola persona!

Le principali strategie di conservazione

- Migrazione
- Emulazione
- Incapsulamento

La migrazione

*"Migration is a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects. It also retains the capability for clients to retrieve, display, and otherwise use the digital objects in the face of constantly changing technology. **Migration aims to overcome the threats of software obsolescence and form at obsolescence by a preservation system that periodically transforms digital information from one format to another.**"*

*"It further breaks down the migration into two more categories: **reversible migration and nonreversible migration**. This is a crucial distinction. The only way to guarantee that no information is lost when a migration is performed is the litmus test of a backward migration which re-creates the original object precisely"*

*Digital Preservation in Libraries: Preparing for a Sustainable Future
by Jeremy Myntti, Jessalyn Zoom, ALA Editions, 2018.*

L'emulazione

*"The computer science term emulation denotes a process in which one computer is used to reproduce the behavior of another computer with such fidelity that the emulation can be used in place of the original computer. Emulation is a proven technique that has been a part of mainstream computer science for decades. **The essential idea behind emulation is to access or run the original data/software on a new platform that “emulates” the original platform**".*

*"Thus, emulation has become very advanced and capable of running complete operating systems like Microsoft Windows, Apple's MacOS, or GNU/Linux, and supporting a wide array of device peripherals. However, **none of these solutions has been designed specifically for digital preservation. Emulation or virtualization software that works today provides no guarantee that it will operate under different conditions in the future.** Various techniques have been proposed to overcome this problem, such as chaining or migrating emulators".*

*Digital Preservation in Libraries: Preparing for a Sustainable Future
by Jeremy Myntti, Jessalyn Zoom, ALA Editions, 2018.*

L'incapsulamento

*"Encapsulation is a technique that “requires metadata to be bundled with, or embedded into, the digital object. The metadata allows the record to be intellectually understood and technologically accessed in the future. **This technique aims to counter obsolete file formats by encapsulating or grouping “details of how to interpret the digital bits in the object”** through the use of “physical or logical structures called ‘containers’ or ‘wrappers’ to provide a relationship between all information components, such as the digital object and other supporting information.”"*

"On its own, encapsulation cannot preserve digital records: “encapsulation is not a method that prescribes how digital documents will be reconstructed on the screen in future or how accessibility is preserved.” What it does is ensure that metadata about the object’s original relationships is packaged with it, to aid both preservation strategies such as migration or emulation, and future user interpretations (provenance, context, etc.)"

*Digital Preservation in Libraries: Preparing for a Sustainable Future
by Jeremy Myntti, Jessalyn Zoom, ALA Editions, 2018.*

Esternalizzare i servizi: cosa considerare

- Rapporto costi/efficienza.
- Vincoli legali (in particolare legati alla sicurezza e al trattamento dei dati, e alla tutela del diritto d'autore).
- Vincoli etici legati alla "affidabilità" (anche dal punto di vista comunicativo rispetto agli stakeholders) del fornitore.
- Possibilità di monitoraggio de processi.
- Scalabilità del sistema.
- Strategie e costi di uscita.

Audit e certificazione

- [Core Trust Seal](#)

Il Core Trust Seal rappresenta il primo passaggio, in un contesto di riferimento su scala globale, per la certificazione di repository affidabili. Questa valutazione del repository è costituita da 16 punti e può essere utilizzata come strumento di autovalutazione o revisione.

- [DIN 31644 Information and documentation - Criteria for trustworthy digital archives](#)

La versione estesa del processo di autovalutazione degli archivi digitali è una lista utile che è stata realizzata dai NESTOR21 sulla base di DIN 31644 Information and documentation - Criteria for trustworthy digital archives standard.

- [ISO16363: 2012 Audit and certification of trustworthy digital repositories](#)

ISO 16363 è un audit framework basato sulle evidenze per la conservazione digitale e consiste in più di 80 criteri che possono essere utilizzati per un'autoverifica o verifica esterna. I criteri usati nello standard sono riferiti all'intera organizzazione e non soltanto al sistema tecnico in cui il contenuto della collezione è archiviato.

Autovalutazione



Audit esterno

Il Digital Preservation Plan

Il Digital Preservation Plan dovrebbe includere:

1. la **strategia** (perché e per chi si conserva) > organizzazione
2. le **policy** di conservazione e accesso (come si conserva) > risorse
3. l'**operatività** (chi conserva) > tecnologie

Table 3.1 • Contents of each part of the three key digital preservation documents

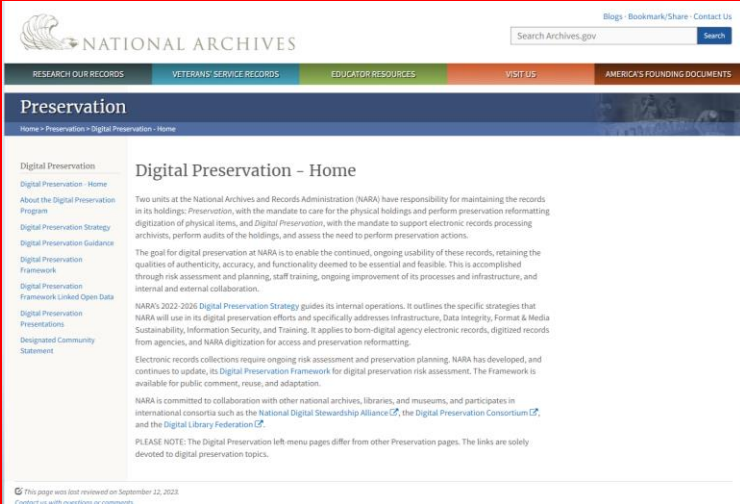
DOCUMENT	CONTENT	OWNER	FREQUENCY OF UPDATE
<i>Strategy</i>	<ul style="list-style-type: none">• Collaboration• Content Scope• Mandates• Policy/Strategy Review• Sustainability Planning	Senior Administrator	Aligns with organizational strategy update
<i>Policy</i>	<ul style="list-style-type: none">• Access and Use• Metadata or Documentation• Preservation Model/ Strategy• Preservation Planning• Rights and Restriction Management• Selection/Appraisal• Staff Training/Education	Manager	3-5 years
<i>Operations</i>	<ul style="list-style-type: none">• Accessioning and Ingest• Audit• Roles and Responsibilities• Security Management• Storage, Duplication, and Backup	Technical Lead	Annual

Digital Preservation in Libraries: Preparing for a Sustainable Future by Jeremy Myntti, Jessalyn Zoom, ALA Editions, 2018.

Il Digital Preservation Plan

Breve guida alla redazione del Digital Preservation Plan:

Liz Bishoff, [*Digital Preservation Plan. Ensuring Long Term Access and Authenticity of Digital Collections*](#), ISQ Information Standards Quarterly, Vol. 22, issue 2, 2010.



The screenshot shows the National Archives website page for Digital Preservation. The page features a navigation bar with links for Research Our Records, Veterans' Service Records, Educator Resources, Visit Us, and America's Founding Documents. The main content area is titled "Digital Preservation - Home" and includes a sidebar with links to Digital Preservation, About the Digital Preservation Program, Digital Preservation Strategy, Digital Preservation Guidance, Digital Preservation Framework, Digital Preservation Framework Linked Open Data, Digital Preservation Presentations, and Designated Community Statement. The main text discusses the National Archives and Records Administration's (NARA) responsibility for maintaining records, the goal of digital preservation, and the NARA 2022-2026 Digital Preservation Strategy. It also mentions the Digital Preservation Framework for digital preservation risk assessment and the Digital Library Federation.

<https://www.archives.gov/preservation/digital-preservation>

Alcune considerazioni sui costi della Conservazione digitale

- Se da una parte per una organizzazione è assai difficile quantificare nel dettaglio i costi della conservazione (seppure esistano degli [strumenti per la modellazione dei costi di gestione del ciclo di vita del digitale](#)) proprio perché si tratta di un processo soggetto a continui mutamenti, dall'altra **se la conservazione viene pianificata al momento della creazione di una risorsa, è possibile non solo avere un'idea più precisa dei costi, ma soprattutto evitare sprechi di risorse** dovuti ad una gestione non efficiente dei dati e delle risorse destinate alla conservazione stessa.
- Bisogna scendere a patti con l'evidenza che seppure la Conservazione sia un'attività di lungo periodo, **i suoi costi possono essere determinati con relativa precisione solo per un periodo di tempo determinato.**
- **Ciò che deve invece essere garantito "indefinitamente" è l'impegno a compiere ogni ragionevole sforzo per la conservazione.**

Strumenti e risorse utili

- [Community Owned digital Preservation Tool Registry \(COPTR\)](#)
- [DPC Technology Watch Publications](#)
- [Library of Congress - Recommended Formats Statement 2023-2024](#)
- [NESTOR E.V. – COMPETENCE NETWORK FOR DIGITAL PRESERVATION](#)
- [Sustainability of Digital Formats: Planning for Library of Congress Collections](#)

Il contesto normativo italiano e il rapporto con la legislazione sovranazionale

Gli ambiti normativi interessati

- Quali risorse possono o devono essere **conservate** e con quali modalità.
 - Deposito legale e diritto d'autore
- Quali risorse possono o devono essere **accessibili** e a quali condizioni.
 - Diritto d'autore, privacy e trattamento dei dati
- Quali risorse possono o devono essere **riutilizzabili** e a quali condizioni.
 - Diritto d'autore e licenze d'uso

Gli ambiti normativi interessati

- **Deposito legale**

In Italia: L. 106/2004 e D.P.R. 252/2006

- **Il diritto d'autore e le sue eccezioni** (nei paesi Civil law / Copyright e Fair use nei paesi di Common law)

In Italia: L. 633/1941, come modificata a seguito del recepimento della Direttiva UE 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale

- **La privacy e il trattamento dei dati**

In Italia: Regolamento UE 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati (GDPR) e D. Lgs. 196/2003 per quanto non modificato o abolito dal GDPR

Il diritto d'autore e le sue eccezioni

LEGGE 22 aprile 1941, n. 633

Protezione del diritto d'autore e di altri diritti connessi al suo esercizio.

CAPO V

Eccezioni e limitazioni

Art. 68.

2-bis. **Gli istituti di tutela del patrimonio culturale** di cui all'articolo 70-ter, comma 3, **per finalità di conservazione** e nella misura a tal fine necessaria, **hanno sempre il diritto di riprodurre e realizzare copie di opere o di altri materiali protetti, presenti in modo permanente nelle loro raccolte**, in qualsiasi formato e su qualsiasi supporto. È nulla qualsiasi pattuizione avente ad oggetto limitazioni o esclusioni di tale diritto.

Il diritto d'autore e le sue eccezioni

Art. 70-ter.

1. Sono consentite le riproduzioni compiute da organismi di ricerca e da istituti di tutela del patrimonio culturale, per scopi di ricerca scientifica, ai fini dell'estrazione di testo e di dati da opere o da altri materiali disponibili in reti o banche di dati cui essi hanno lecitamente accesso, nonché la comunicazione al pubblico degli esiti della ricerca ove espressi in nuove opere originali.

2. Ai fini della presente legge per estrazione di testo e di dati si intende qualsiasi tecnica automatizzata volta ad analizzare grandi quantità di testi, suoni, immagini, dati o metadati in formato digitale con lo scopo di generare informazioni, inclusi modelli, tendenze e correlazioni.

3. Ai fini della presente legge per istituti di tutela del patrimonio culturale si intendono le biblioteche, i musei, gli archivi, purché aperti al pubblico o accessibili al pubblico [...]

4. Ai fini della presente legge, per organismi di ricerca si intendono le università, comprese le relative biblioteche, gli istituti di ricerca o qualsiasi altra entità il cui obiettivo primario è quello di condurre attività di ricerca scientifica

[...]

6. Le copie di opere o di altri materiali realizzate in conformità al comma 1 sono memorizzate con un adeguato livello di sicurezza e possono essere conservate e utilizzate unicamente per scopi di ricerca scientifica, inclusa la verifica dei risultati della ricerca.

Il diritto d'autore e le sue eccezioni

Art. 70-quater.

1. Fermo restando quanto previsto dall'articolo 70-ter, **sono consentite le riproduzioni e le estrazioni da opere o da altri materiali contenuti in reti o in banche di dati cui si ha legittimamente accesso ai fini dell'estrazione di testo e di dati.**

L'estrazione di testo e di dati è consentita quando l'utilizzo delle opere e degli altri materiali non è stato espressamente riservato dai titolari del diritto d'autore e dei diritti connessi nonché dai titolari delle banche dati.

2. Le riproduzioni e le estrazioni eseguite ai sensi del comma 1 possono essere conservate solo per il tempo necessario ai fini dell'estrazione di testo e di dati.

Il trattamento dei dati personali

Regolamento UE 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati (GDPR)

Articolo 17

Diritto alla cancellazione («diritto all'oblio»)

1. L'interessato ha il diritto di ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano senza ingiustificato ritardo e il titolare del trattamento ha l'obbligo di cancellare senza ingiustificato ritardo i dati personali [...]

3. I paragrafi 1 e 2 non si applicano nella misura in cui il trattamento sia necessario:

a) per l'esercizio del diritto alla libertà di espressione e di informazione;

[...]

d) a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici conformemente all'articolo 89, paragrafo 1*, nella misura in cui il diritto di cui al paragrafo 1 rischi di rendere impossibile o di pregiudicare gravemente il conseguimento degli obiettivi di tale trattamento [...]

Il trattamento dei dati personali

*Articolo 89

Garanzie e deroghe relative al trattamento a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici

1. Il trattamento a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici è soggetto a garanzie adeguate per i diritti e le libertà dell'interessato, in conformità del presente regolamento. Tali garanzie assicurano che siano state predisposte misure tecniche e organizzative, in particolare al fine di garantire il L 119/84 IT Gazzetta ufficiale dell'Unione europea 4.5.2016 rispetto del **principio della minimizzazione dei dati**. Tali misure possono includere la pseudonimizzazione, purché le finalità in questione possano essere conseguite in tal modo. Qualora possano essere conseguite attraverso il trattamento ulteriore che non consenta o non consenta più di identificare l'interessato, tali finalità devono essere conseguite in tal modo.

Problemi aperti

Il complesso di eccezioni cui fanno riferimento sia la normativa sul diritto d'autore sia quella sul trattamento dei dati si riferisce, più o meno implicitamente, a risorse che gli istituti di conservazione e di ricerca hanno legittimamente acquisito.

La questione cruciale è quindi **come tali istituti possano legittimamente acquisire il patrimonio culturale e informativo prodotto da terzi, al fine di garantirne la conservazione e l'accesso nel lungo periodo .**

Il deposito legale delle risorse digitali: la situazione italiana

In Europa si inizia a discutere di Deposito legale del “digitale” tra il 2002 e il 2006...

In Italia, per la prima volta, la **L. 15 aprile 2004 n. 106** Norme relative al deposito legale dei documenti di interesse culturale destinati all'uso pubblico **prende in considerazione, tra le Categorie di documenti destinati al deposito legale (art. 4, c.1, lett. r), i documenti diffusi tramite rete informatica.**

Il successivo **D.P.R. 3 maggio 2006 n. 252** Regolamento recante norme in materia di deposito legale dei documenti di interesse culturale destinati all'uso pubblico definisce meglio i documenti diffusi tramite rete informatica come documenti trasmessi per via telematica, con qualunque rete mobile o fissa (art. 2, c. 1. lett. f n.2), e il loro deposito è trattato Capo VII - Deposito dei documenti diffusi tramite rete informatica.

Il deposito legale delle risorse digitali: la situazione italiana

Art. 37. Modalità di deposito e acquisizione dei documenti diffusi tramite rete informatica

1. **Le modalità di deposito dei documenti diffusi tramite rete informatica sono definite con successivo regolamento [...]**
2. **Il Ministero promuove forme volontarie di sperimentazione del deposito** [...] mediante la stipulazione di appositi accordi con i soggetti obbligati al deposito. Gli accordi definiscono le modalità tecniche del deposito prevedendo, ove possibile, anche forme automatiche di raccolta, secondo le migliori pratiche e conoscenze internazionali del settore.
3. Il Ministero, nella stipulazione degli accordi di cui al comma 2, assicura **prioritariamente la raccolta delle seguenti tipologie di documenti:**
 - a) documenti che assicurino la continuità delle collezioni già avviate, anche su supporti e mediante tecnologie tradizionali;
 - b) documenti concernenti la produzione scientifica delle università, dei centri di ricerca e delle istituzioni culturali;
 - c) documenti elaborati e messi in rete da soggetti pubblici;
 - d) documenti relativi a siti che si aggiornano con più frequenza, ovvero contenuti in siti che sono maggiormente citati da altri siti.

Il deposito legale delle risorse digitali: la situazione italiana

Il regolamento attuativo previsto dall'art. 37 c. 1 non è ancora stato emanato quindi...
in Italia il deposito legale dei documenti diffusi tramite rete informatica (compresi quindi siti e altre risorse Web) NON È OBBLIGATORIO.

Tuttavia, sulla base del mandato dell'art. 37 c. 2, dal 2010 è attivo il servizio **Magazzini Digitali...**

Gli ambiti normativi NON interessati

- L'unica normativa nazionale vigente in materia di “Conservazione del digitale”, derivante nella sostanza dalle previsioni degli art. 44 e 71 del D. Lgs. n. 82/2005 (CAD - Codice dell'Amministrazione digitale) prende in considerazione esclusivamente la conservazione dei documenti informatici, così come definiti nello stesso CAD: *documento elettronico che contiene la rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti.*
- Tuttavia, come ribadito anche al paragrafo 5.1 del documento, a cura di AGID, “Progetto Poli di conservazione definizione di un modello di riferimento per i Poli di Conservazione e della relativa rete nazionale”:

L'attuale normativa sulla conservazione digitale non prende in considerazione, se non marginalmente, la conservazione a fini storici di documenti e archivi digitali, così come non specifica le funzioni di fruizione del patrimonio documentale, insite invece nella definizione di “archivio” data dal Codice dei beni culturali e anche dallo standard OAIS (ISO 14721).

Gli ambiti normativi NON interessati

Ciò non impedisce, anzi rende maggiormente auspicabile, trovare forme di cooperazione a livello nazionale per la gestione di entrambi i processi, laddove la conservazione di queste diverse tipologie di risorse digitali, se da una parte differisce per finalità - non consentendo una gestione completamente integrata dei servizi -, dall'altra mostra punti di convergenza rispetto alle necessità tecniche e tecnologiche, incentivando l'opportunità di sopperirvi tramite la creazione di infrastrutture nazionali condivise.

Problemi aperti

La **Conservazione del Digitale** - e in particolare alcune sue componenti tra cui il Web and Social media archiving - **è una sfida globale**, che richiede risposte collaborative e coordinate a livello internazionale: **come è possibile conciliare questa esigenza con tradizioni legislative a volte molto diverse**, frutto di eredità storiche e culturali differenti, le cui norme necessariamente si possono applicare solo in ambiti ben definiti?

Problemi aperti

Conservare il digitale non significa solo conservare le singole risorse o gruppi di risorse **ma spesso implica anche**, per volontà o necessità, **la conservazione degli ambienti di produzione e fruizione delle risorse**, la cui conservazione potrebbe essere soggetta ad ulteriori limiti legali (es. licenze software, termini d'uso delle piattaforme ecc.).

Conservazione del digitale in Italia: progetti in corso

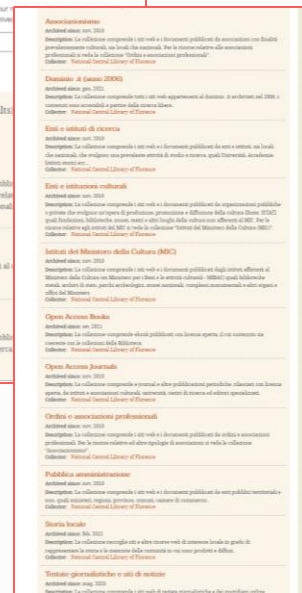
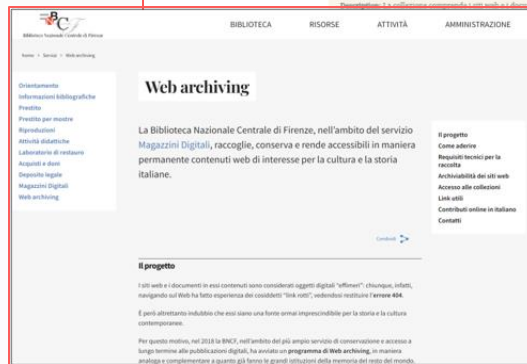
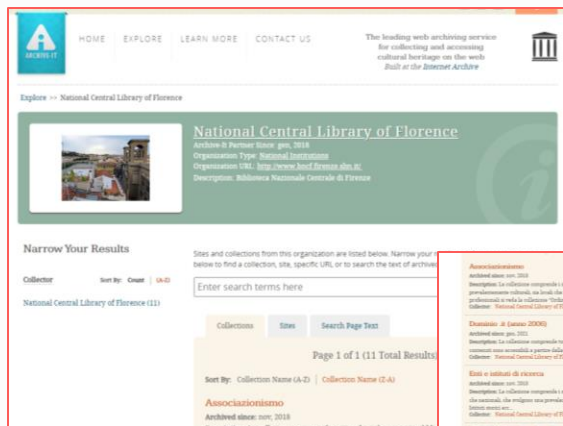
Il deposito legale delle risorse digitali: la situazione italiana

[Magazzini Digitali](#) è il servizio nazionale di conservazione e accesso a lungo termine alle pubblicazioni digitali:

- tesi di dottorato
- ebook (sperimentazione attualmente ferma)
- ejournal open access
- siti e altre risorse web ([Web e Social media archiving](#))

Il deposito legale delle risorse digitali: la situazione italiana

- Nel 2018 la BNCf ha attivato un servizio di [Web archiving](#) e archivia periodicamente quasi 500 siti web liberamente accessibili su [Archive-it](#).



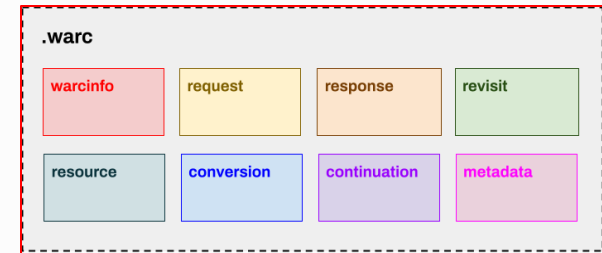
Il deposito legale delle risorse digitali: la situazione italiana

- I siti e i contenuti web sono soggetti alla normativa relativa al Diritto d'autore, che ne impedisce la copia non autorizzata.
- La normativa sul deposito legale digitale, tra le altre cose, dovrebbe regolare le eccezioni al Diritto d'autore per finalità storiche e culturali per quanto riguarda i siti web e le altre risorse digitali.
- In assenza di tale normativa, è possibile che le istituzioni depositarie trovino degli accordi con i produttori dei contenuti che regolino “depositi volontari” con opzioni *opt-in* o *opt-out*.
- **La BNCF porta avanti il servizio nazionale di Web archiving chiedendo un'adesione volontaria ai gestori dei siti (opt-in), che hanno anche facoltà di decidere se rendere le copie archiviate liberamente accessibili sul Web o soltanto dalla rete interna della Biblioteca.**

Il Web e social media archiving

Che cos'è il Web archiving?

- Il processo che consente di **catturare delle istantanee**, più o meno profonde, di porzioni di Web, ad intervalli di tempo definiti, di **archiviarle e renderle accessibili nel lungo periodo**.
- la tecnica di cattura più comune* si basa sull'utilizzo di **Web crawler**, il più diffuso è Heritrix (Open source)
- il formato di archiviazione più diffuso attualmente è il **WARC** (standard ISO)
- lo strumento di accesso più usato è la **Wayback Machine** (Open source)



File WARC Schema

<https://archive-it.org/post/the-stack-warc-file/>

La durata della vita delle risorse web

La durata media della vita di un sito web è 2 anni e 7 mesi

THE LIFESPAN OF A WEBSITE

How long does a website last?

In general, a typical web page can be expected to last **-90-100 days** before changing, moving, or disappearing completely.

- > In 2013, our colleagues at Old Dominion University determined that **over 10%** of event related content posted to social media platforms is lost after one year.
- > In 2014, a study by UCLA determined that **7-in-10** scholarly articles that include citations with hyperlinks suffer from *reference rot*.

Social Media Post Lifespan Snapshot

Every social networking platform is unique. We know that some platforms are primarily for search intent, while others share media content, and yet others are for photography enthusiasts. The length of time your content is likely to last also varies from platform to platform.

Here is a quick look at life spans on various social networks:

- TikTok – Instant Decay
- Twitter – 15-18 minutes
- Facebook – 6 hours
- LinkedIn – 24 hours
- Instagram – 48 hours
- YouTube – About 20 days

Brewster Kahle
@brewster_kahle

This is a book that has every footnote archived. So good. (and the book is free and wonderful to read).

Traduci il Tweet

Glyn Moody @glynmoody · 17h
the ebook version of "Walled Culture" uses 750 links to the indispensable @internetarchive to avoid link rot; so it seems only right that you can download "free" copies from the same great site, forever: archive.org/details/walled...

07:03 · 09 ott 22 · [Twitter Web App](#)

<https://blog.atisyre.com/lifespan-of-a-social-media-post-how-long-is-your-content-relevant/>

Comunicazione / informazione pubblica

Destinata al pubblico o di interesse pubblico

• Istituzionale

[\[L. 150/2000\]](#)

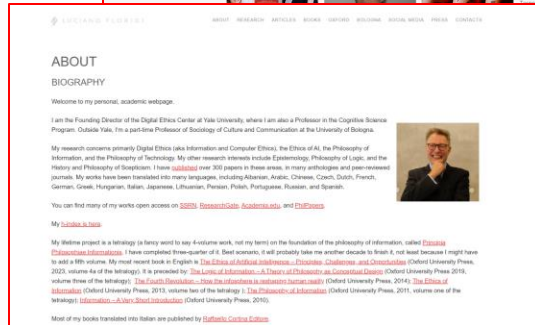
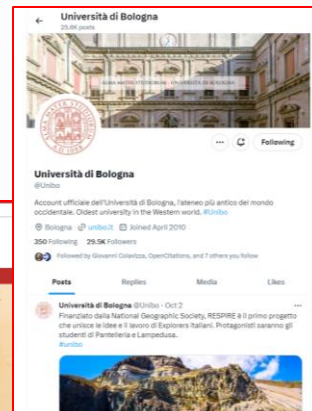
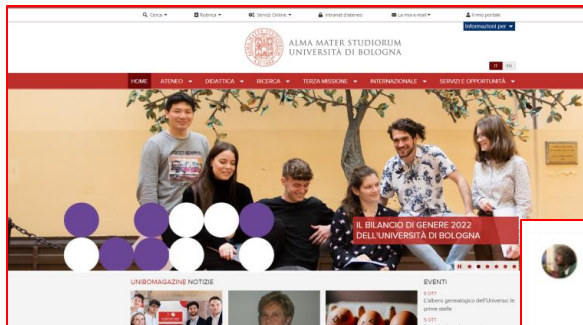
○ su spazi pubblici del Web

○ su spazi privati del Web

• Non istituzionale

○ su spazi pubblici del Web

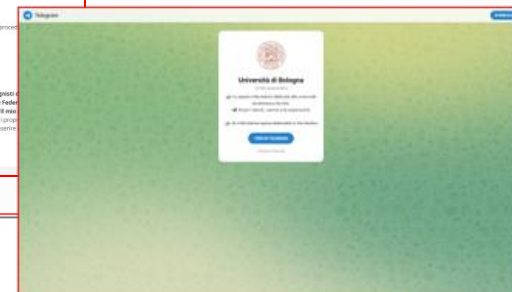
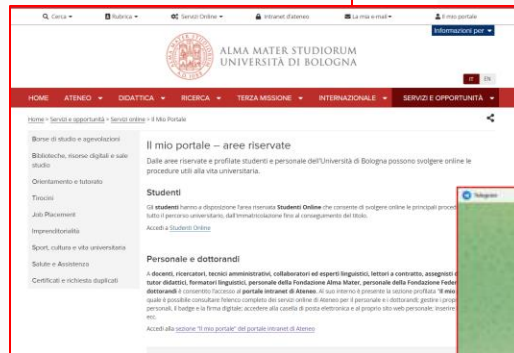
○ su spazi privati del Web



Comunicazione / informazione privata

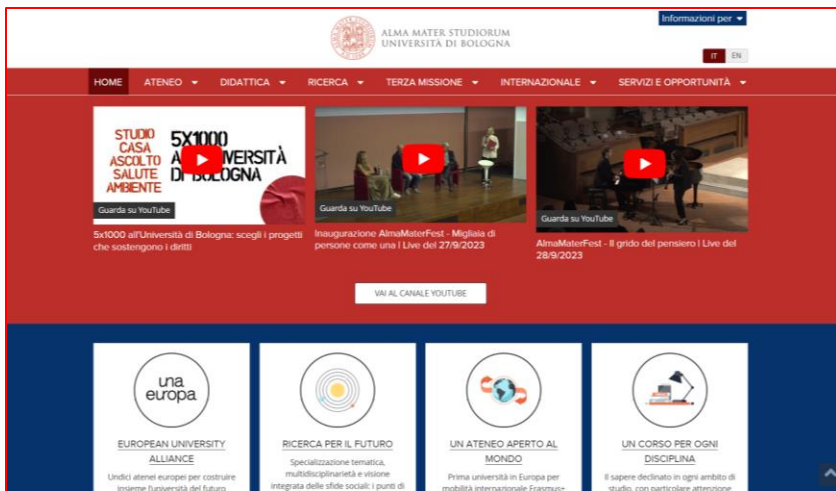
Non destinata al pubblico o non di interesse pubblico

- Istituzionale
- Non istituzionale



Comunicazione "ibrida"

Dal punto di vista dei contenuti e/o dei media utilizzati



Il Web archiving per l'Università e la ricerca scientifica

The image displays three overlapping screenshots of the Internet Archive website, illustrating the process of web archiving and search results. The top screenshot shows the main navigation and a search for 'Enti e istituti di ricerca'. The middle screenshot shows search results for 'Open Access Books', including a detailed view of a book titled 'Il culto medievale di san Vitaliano di Capua'. The bottom screenshot shows search results for 'Open Access Journals', including a detailed view of a journal issue titled 'CASPUR-CIBER Publishing - Sustainable publications'.

Top Screenshot: Internet Archive Home

- Navigation: HOME, EXPLORE, LEARN MORE, CONTACT US
- Text: The leading web archiving service for collecting and accessing cultural heritage on the web. Built at the Internet Archive.
- Search: Explore >> National Central Library of Florence >> Enti e istituti di ricerca

Middle Screenshot: Open Access Books

- Section: Open Access Books
- Collected by: National Central Library of Florence
- Archived since: set, 2021
- Description: La collezione comprende ebook pubblicati con licenza aperta, il cui contenuto è le collezioni della Biblioteca.
- Collector: National Central Library of Florence
- Section: Narrow Your Results
- Search: Enter search terms here
- Results: 3 Total Results
- Item: Title: Il culto medievale di san Vitaliano di Capua
- URL: <https://www.ereamosanvitaliano.it/ereimo22/libridigitali/publicazioni/C-Vultaggio-Il-culto-medievale-di-san-Vitaliano.pdf>
- Captured once on July 06, 2023
- Creator: Vultaggio, Claudia
- Publisher: Pubblicazioni dell'Eremo di San Vitaliano

Bottom Screenshot: Open Access Journals

- Section: Open Access Journals
- Collected by: National Central Library of Florence
- Archived since: nov, 2018
- Description: La collezione comprende e-journal e altre pubblicazioni periodiche, rilasciate con licenza aperta, da istituti e associazioni culturali, università, centri di ricerca ed editori specializzati.
- Collector: National Central Library of Florence
- Section: Narrow Your Results
- Search: Enter search terms here
- Results: 7,502 Total Results
- Item: Title: CASPUR-CIBER Publishing - Sustainable publications
- URL: <http://archivociberpublishing.cineca.it/issue/archive/>
- Captured 2 times between December 21, 2021 and December 27, 2021
- Item: Title: Bollettino del CILEA
- URL: <http://bollettino.cilea.it/issue/archive/>
- Captured 2 times between December 27, 2021 and December 27, 2021
- Publisher: CILEA
- Title: Preface

Il Web archiving per l'Università e la ricerca scientifica

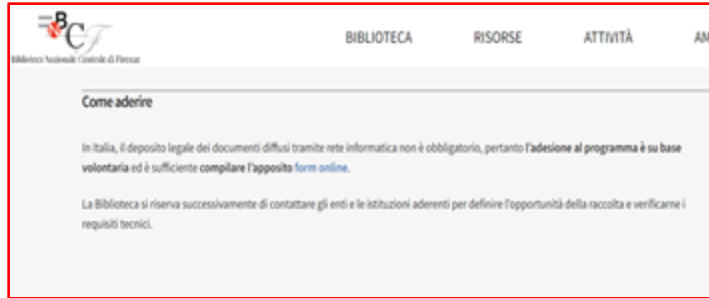
Table 5. Elsevier articles with link rot, lacking representative Mementos, infected by reference rot.

	1997	1998	1999	2000	2001	2002	2003	2004
Elsevier articles	39,480	41,110	39,027	38,494	39,314	37,232	35,971	37,439
with URIs	177	303	395	819	1,536	2,702	3,755	4,938
subject to								
Link rot	141	207	271	517	912	1,582	2,188	2,791
Not archived	169	277	363	624	1,163	2,105	2,581	3,280
Reference rot	169	278	364	626	1,170	2,117	2,603	3,326
	2005	2006	2007	2008	2009	2010	2011	2012
Elsevier articles	35,735	39,236	41,065	42,032	44,162	44,716	49,150	50,877
with URIs	5,828	7,285	8,600	9,792	10,874	11,273	12,845	13,523
subject to								
Link rot	3,052	3,707	4,126	4,524	4,774	4,682	5,205	5,109
Not archived	3,965	4,947	6,007	7,496	8,799	9,336	9,764	10,648
Reference rot	4,022	4,995	6,045	7,513	8,821	9,350	9,782	10,665

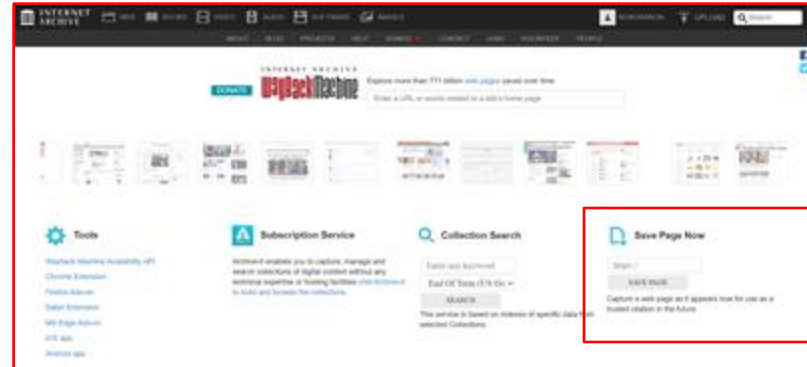
doi:10.1371/journal.pone.0115253.t005

Klein M, Van de Sompel H, Sanderson R, Shankar H, Balakireva L, et al. (2014) *Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot*. PLOS ONE 9(12): e115253. <https://doi.org/10.1371/journal.pone.0115253>

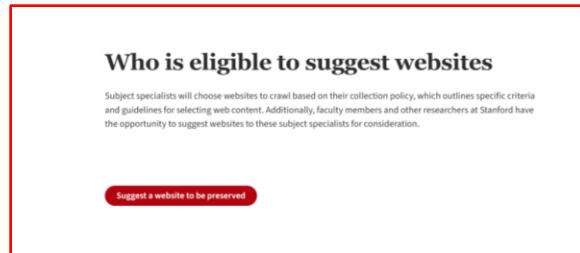
I servizi di archiviazione on-demand



<https://www.bncf.firenze.sbn.it/biblioteca/web-archiving/#cap-1>



<https://archive.org/web/>



<https://library.stanford.edu/web-archiving>



<https://archive.ph/>

L'archiviabilità delle risorse web



The screenshot shows the website of the Biblioteca Nazionale Centrale di Firenze. The header includes the BNCf logo and navigation links: Biblioteca, Risorse, Pubblicazioni, Attività, Amministrazione, Notizie, and CATALOGO. The breadcrumb trail is Home / Pagine / Extra. The main heading is 'Archiviabilità dei siti web'. The text discusses the importance of accessibility, performance, SEO, and W3C standards for web archiving. A sidebar on the right contains icons for home, search, and help. The footer lists 'Linee guida per la realizzazione di siti web archiviabili' and 'Validatore'.

BNCf Biblioteca Nazionale Centrale di Firenze

Biblioteca Risorse Pubblicazioni Attività Amministrazione Notizie CATALOGO

Home / Pagine / Extra

Archiviabilità dei siti web

Gli sviluppatori e i designer di siti web devono oggi tenere conto dei criteri di accessibilità, performance, SEO - Search Engine Optimization, compatibilità con gli standard del W3C e usabilità. Tuttavia, tra le buone pratiche emergenti c'è quella di prendere in considerazione anche i criteri di archiviabilità dei siti web.

Per **archiviabilità** si intende l'insieme delle caratteristiche che i contenuti, la struttura, le funzionalità e le interfacce di un sito web devono possedere perché il sito stesso possa essere conservato e reso accessibile nel lungo periodo con gli attuali strumenti di web archiving.

L Linee guida per la realizzazione di siti web archiviabili
L Validatore

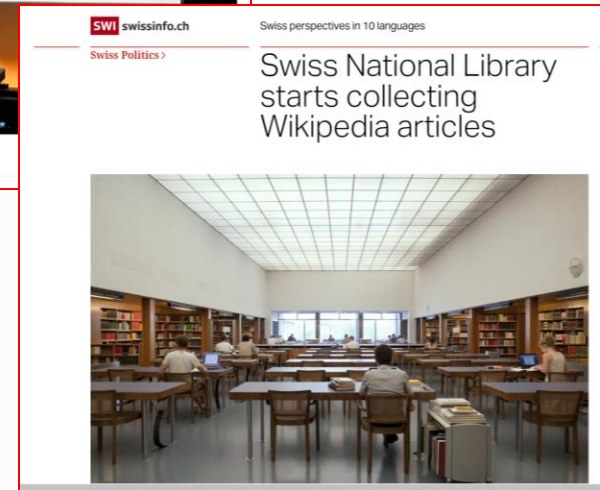
<https://www.bncf.firenze.sbn.it/biblioteca/archiviabilita-dei-siti-web/>

**La conservazione dei prodotti
della ricerca scientifica: progetti
consolidati**

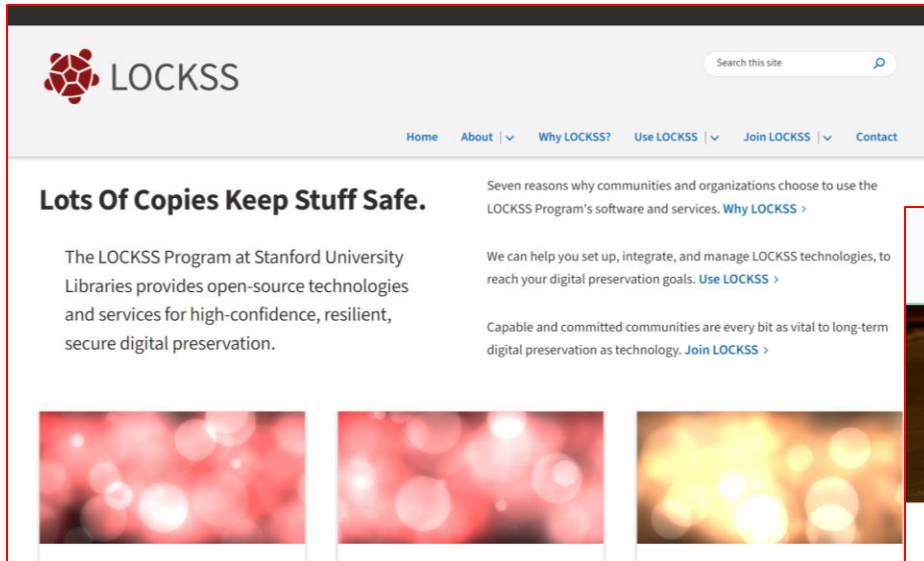
Una riflessione sull' *Open*

- L' *Openness* delle risorse è un fattore decisivo per la loro archiviabilità
- D'altra parte, costituisce anche un fattore di rischio: **chi è responsabile della conservazione di un contenuto open?**

Laakso, Mikael, Lisa Matthias, and Najko Jahn. 2021.
“*Open Is Not Forever: A Study of Vanished Open Access Journals.*”
Journal of the Association for Information Science and Technology
<http://arxiv.org/abs/2008.11933>



La rete LOCKSS e il servizio CLOCKSS



The screenshot shows the LOCKSS website homepage. At the top left is the LOCKSS logo, a red hexagonal shape composed of smaller hexagons. To its right is a search bar with the text "Search this site" and a magnifying glass icon. Below the logo and search bar is a navigation menu with links: Home, About (with a dropdown arrow), Why LOCKSS?, Use LOCKSS (with a dropdown arrow), Join LOCKSS (with a dropdown arrow), and Contact. The main content area features a large heading "Lots Of Copies Keep Stuff Safe." followed by a paragraph: "The LOCKSS Program at Stanford University Libraries provides open-source technologies and services for high-confidence, resilient, secure digital preservation." To the right of this paragraph are three columns of text. The first column says: "Seven reasons why communities and organizations choose to use the LOCKSS Program's software and services. [Why LOCKSS >](#)" The second column says: "We can help you set up, integrate, and manage LOCKSS technologies, to reach your digital preservation goals. [Use LOCKSS >](#)" The third column says: "Capable and committed communities are every bit as vital to long-term digital preservation as technology. [Join LOCKSS >](#)" At the bottom of the page are three blurred images: two with a red bokeh effect and one with a yellow bokeh effect.

<https://www.lockss.org/>



The screenshot shows the CLOCKSS website homepage. At the top left is the CLOCKSS logo, a clock face with the text "CLOCKSS" to its right. To its right is a search bar with a magnifying glass icon and two buttons: "Join CLOCKSS" and "Contact us". Below the logo and search bar is a navigation menu with links: About (with a dropdown arrow), What's in CLOCKSS? (with a dropdown arrow), Community (with a dropdown arrow), News, and CLOCKSSface. To the right of the navigation menu are social media icons for Twitter and LinkedIn. The main content area features a large heading "A Global Leader in Digital Preservation Services" over a background image of an open book. Below this heading is a paragraph: "We are CLOCKSS. A collaboration of the world's leading academic publishers and research libraries, CLOCKSS provides a sustainable online archive to ensure the long-term survival of digital scholarly content." Below this paragraph are two more paragraphs: "As global leaders in digital preservation services, we ensure that all books, journals, and knowledge entrusted to CLOCKSS will be safe. We preserve the scholarly record so it is protected and accessible for all, preserving the legacy of academic authors indefinitely." and "CLOCKSS is the only online archive that assigns Creative Commons Open Access licenses to all triggered publications to ensure they always remain open and available to everyone." At the bottom of the page is a dark blue footer with the text: "At CLOCKSS, through the services we provide to libraries and publishers, built over LOCKSS software and delivered as part of the broader LOCKSS community, we instill confidence in

IL PKP Preservation Network

SFU.CA

SFU PKP PUBLIC KNOWLEDGE PROJECT

ABOUT ▾ PUBLISHING SOFTWARE ▾ HOSTING & SERVICES ▾ RESEARCH COMMUNITY ▾

Home / PKP Preservation Network

PKP PRESERVATION NETWORK

PKP has developed the PKP Preservation Network (PKP PN) to digitally preserve OJS journals. The LOCKSS program offers decentralized and distributed preservation, seamless perpetual access, and preservation of the authentic original version of the content. The PKP PN ensures that journals that are not part of any other digital preservation service (such as CLOCKSS or Portico) can be preserved for long-term access.

See the [PKP Preservation Network Guide](#) for additional details about the PKP PN and how to use it. [This earlier discussion document](#) also provides a high level overview.

OVERVIEW

The PKP PN:

- provides preservation services for any OJS journal that has published at least one article and has an ISSN and is running OJS 3.1.2 or newer
- allows OJS journal managers to register their journals for inclusion in the PKP PN by simply enabling a plugin in their OJS instance and agreeing to the terms of the Preservation Agreement
- on registration of a journal, automatically determines if the registered journal meets the minimum inclusion requirements
- preserves the raw journal content as generated by OJS's existing export tools
- automatically harvests new content from registered journals and adds the content to the PKP PN, eliminating manual staff intervention at PN member sites
- provides access to preserved content after a "trigger event"

END-USER ACCESS TO JOURNALS IN THE PKP PN

<https://pkp.sfu.ca/pkp-pn/>

PROJECT JASPER

Project JASPER (JournAIS are Preserved foreVER) is an initiative to preserve open access journals. It was launched on [World Preservation Day 2020](#) and is in response to research that shows that online journals—both open and closed access journals—can just disappear from the internet. This happens because of a lack of awareness amongst smaller publishers around the need for long-term digital preservation and/or the resources to enroll a journal in a long-term digital preservation scheme.



Long-term archiving of research resources is of paramount importance for scholarship. Authors want to ensure their contributions to the scholarly record will be permanent. Scholars must be able to access all of the published research in their fields, both now and long into the future.

Project JASPER aims to close the gap in preservation coverage that currently exists among open access journals. It is a partnership led by [DOAJ](#) and including [CLOCKSS](#), [Keepers Registry](#), [Internet Archive](#) and PKP as participants. PKP is contributing to this project by developing better documentation and support efforts for existing OJS users to understand and implement the preservation mechanisms available to them (whether the PN, CLOCKSS, Internet Archive or another option) and by working with the other project partners to improve outreach and communications supporting these tools. We are also actively working with the group on making the path to having a viable preservation solution easier for OA journals.

Questions? Email us, or take a look at our [primary project page](#).

Portico

PORTICO

Why Portico Our work Coverage Join Portico News

For participants

Portico

Working with libraries and publishers to preserve scholarly content

Portico is a community-supported preservation archive that safeguards access to e-journals, e-books, and digital collections. Our unique, trusted process ensures that the content we preserve will remain accessible and usable for researchers, scholars, and students in the future.

We join with DPC and other member organizations in calls to preserve historical, cultural, and informational resources amidst the crisis in Ukraine.

[Read the statement](#)

NEWS

All News

June 21, 2023

ITHAKA announces new services for academic, research, and cultural institutions to share, preserve, and manage digital collections

May 23, 2023

Search for titles in the Portico archive

Search titles by keyword:

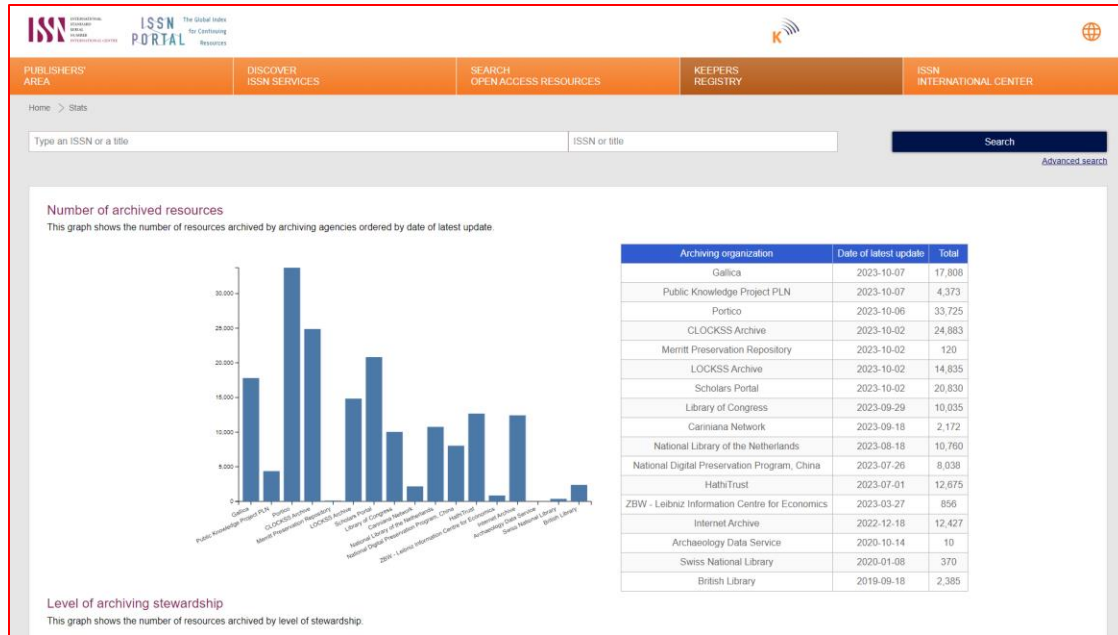
E-journals E-books Collections

Request a free holdings comparison

- Inform your decision about joining Portico
- Compare costs to other preservation solutions
- Manage the preservation status of

<https://www.portico.org/>

ISSN _ Keepers Registry



<https://keepers.issn.org/stats>

**Il PND: una occasione mancata
per la conservazione del
patrimonio digitale?**

Next Generation EU

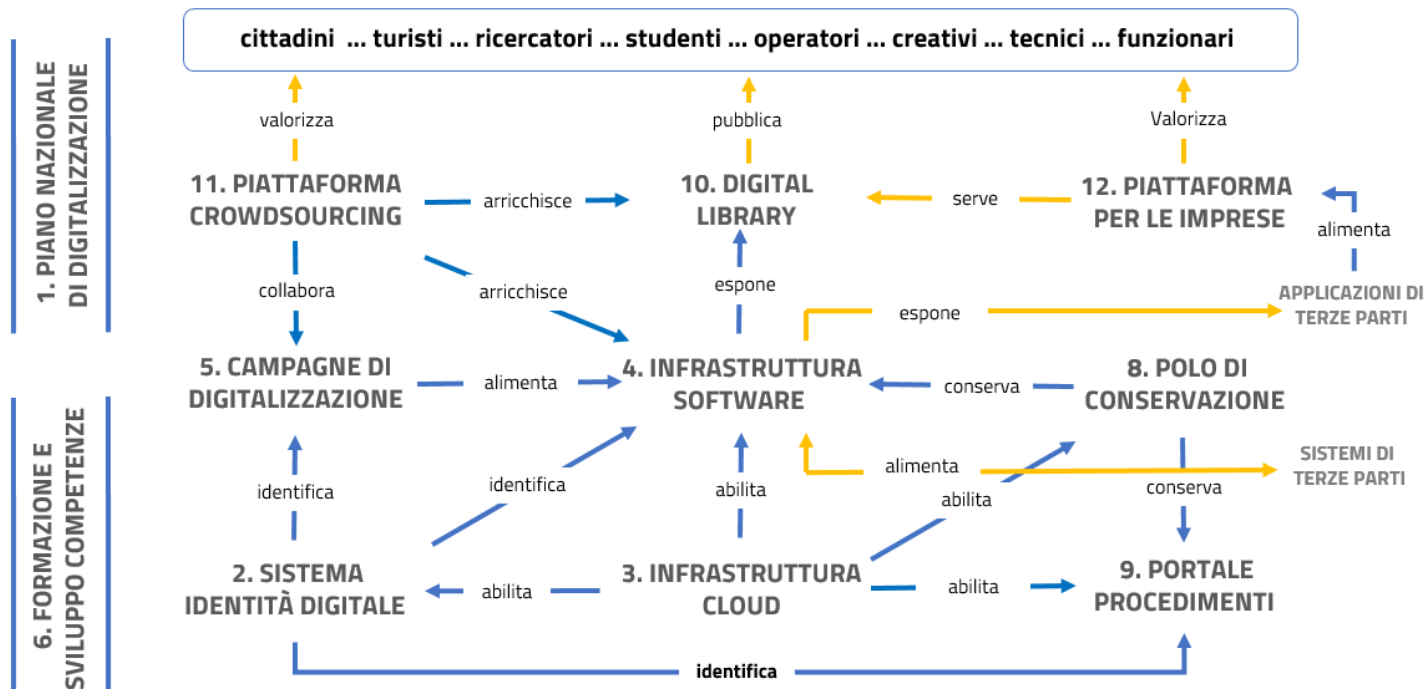
M1C3 – Turismo e Cultura 4.0

Strategie e piattaforme digitali per il patrimonio culturale

maggio 2023

RELAZIONI TRA I 12 SUB-INVESTIMENTI

Strategie e piattaforme digitali per il patrimonio culturale



Sub-investimento 1.1.4

Infrastruttura software del patrimonio culturale

Soggetto attuatore:	Istituto centrale per la digitalizzazione del patrimonio culturale – Digital Library, in collaborazione con Cineca
Importo stanziato:	73 milioni
Target:	rilascio servizi core entro dicembre 2023

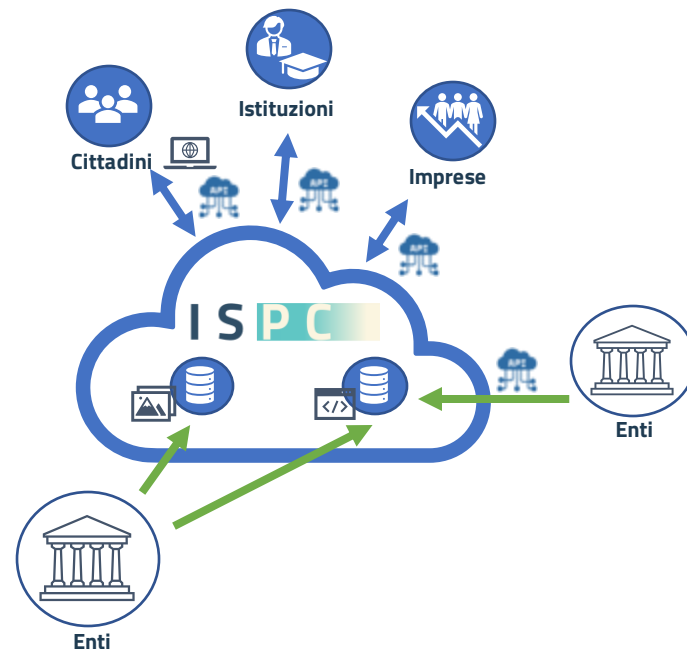
OBIETTIVI DEL PROGETTO

Infrastruttura software del patrimonio culturale

L'Infrastruttura Software del Patrimonio Culturale (**ISPC**) rappresenta il **primo spazio dati nazionale della cultura** in grado di ospitare in sicurezza tutto il patrimonio digitalizzato del Paese.

Nella ISPC il patrimonio è valorizzato ed arricchito per mezzo di **algoritmi innovativi** e tecniche di **Intelligenza Artificiale** e viene messo in correlazione con tutte le risorse culturali digitalizzate presenti, sia all'interno di ciascun dominio di appartenenza (artistico, archivistico, librario) sia tra domini diversi.

La ISPC nasce con **un'architettura scalabile e sicura basata su soluzioni Cloud** offrendo performance e livelli di disponibilità adeguati a sostenere credibilmente il ruolo di abilitatore nell'offerta di servizi digitali per il Panorama Culturale Italiano anche nella dimensione europea.



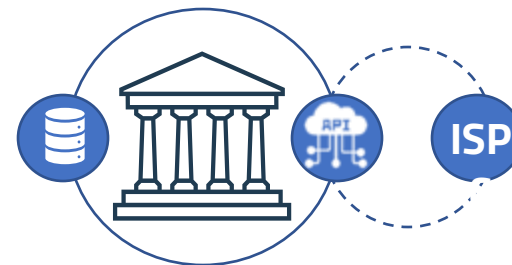
MODELLI DI COOPERAZIONE

Infrastruttura software del patrimonio culturale



Sistemi Integrati

Gli enti cooperanti che scelgono il modello dell'integrazione, **conferiscono i propri dati all'infrastruttura**, condividendone i servizi: **le risorse digitali sono** quindi «ospitate» nell'infrastruttura software e vengono memorizzate e conservate sui sistemi dell'infrastruttura, che ne gestisce il ciclo di vita. Tali sistemi continuano a gestire interamente i workflow di catalogazione dei beni culturali.

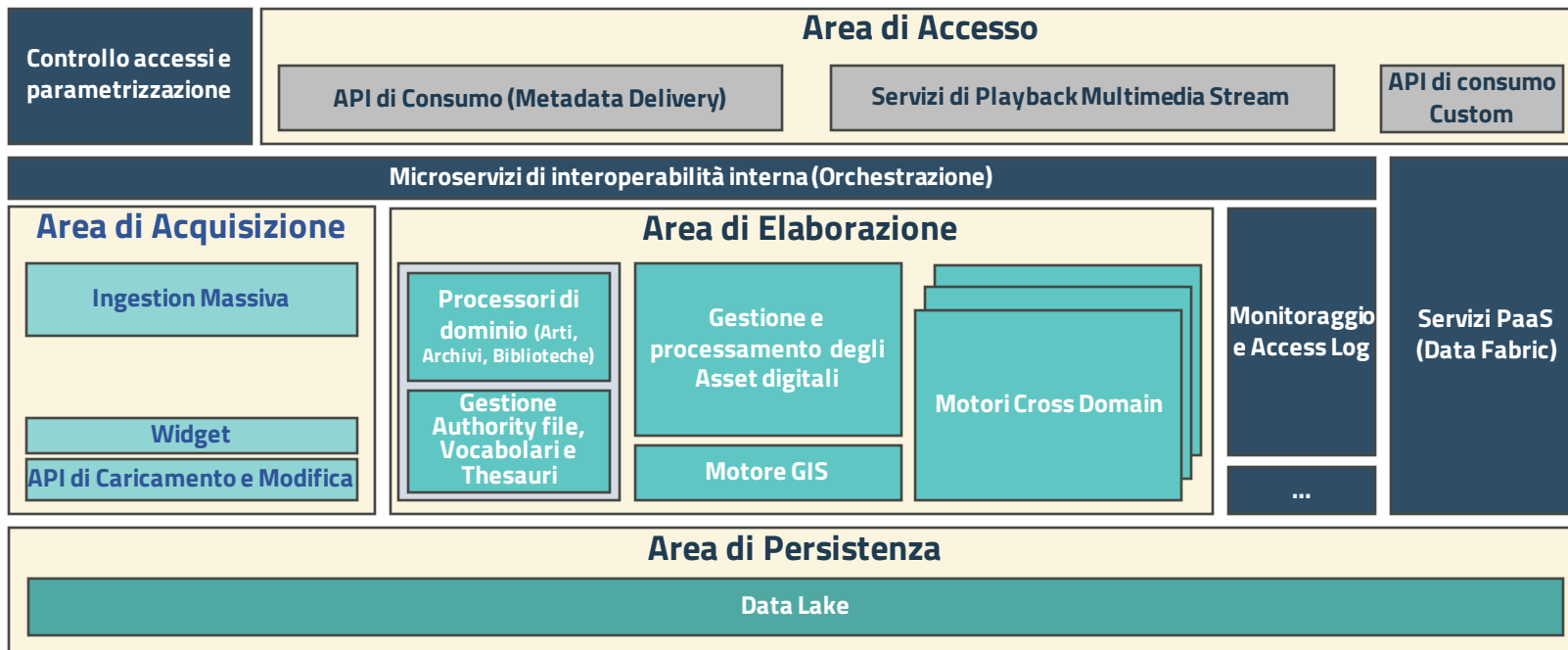


Sistemi Federati

Gli enti cooperanti che scelgono il modello della federazione, hanno sistemi informativi in grado di esporre in modo stabile ed efficiente le risorse digitali mediante **API standard**; **condividono con l'infrastruttura solo le descrizioni**, mantenendo le **risorse digitali nei sistemi di origine e «referenziandole»** nell'infrastruttura, mentre il ciclo di vita dei dati è gestito dall'ente nei propri sistemi.

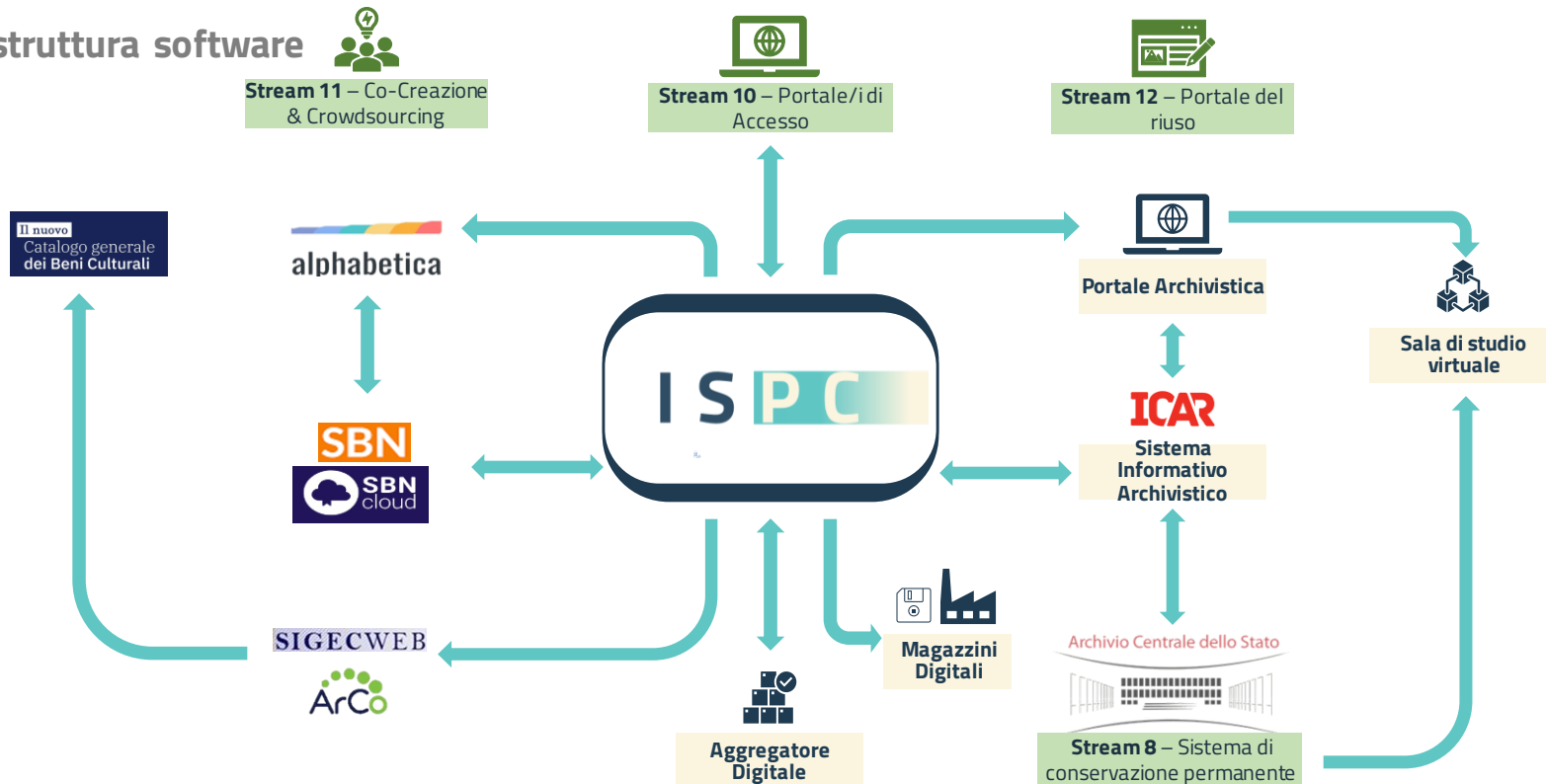
MACRO AREE LOGICO-FUNZIONALI

Infrastruttura software del patrimonio culturale



ECOSISTEMA DIGITALE DEL MINISTERO

Infrastruttura software



CRONOPROGRAMMA DEI RILASCI IN ESERCIZIO

Infrastruttura software del patrimonio culturale

➤ **Q2 2023**

Alpha Release

rilasciata agli **early adopters** del MIC

➤ **Q2 2024**

Core Release

➤ **Q4 2024**

Complementary Release

2022

2023

2024

**Conservare il Digitale:
le competenze necessarie**

Qualche considerazione sulle competenze

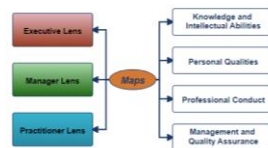
- Quali competenze deve possedere chi si occupa di conservazione del digitale?
Sarebbe necessario definire un profilo professionale "nuovo"?
 - Qual è il suo ruolo nel processo di conservazione? Quali sono le sue responsabilità?
- Potrebbe essere maggiormente utile definire quali competenze per una efficace attività di conservazione digitale debba possedere un'organizzazione nel suo complesso?

I profili professionali

- Il DigCurV Curriculum Framework è uno degli strumenti utili ad identificare i diversi gruppi di competenze necessarie a chi si occupa di conservazione, definiti in relazione al tipo di responsabilità che l'operatore ha all'interno dell'organizzazione.

Focusing on Specific Audiences

The framework defines separate skills lenses to match the specific needs of distinct audiences within digital curation in the shape of Executives, Managers, and Practitioners.



Executive Lens

The skills defined under the Executive Lens are the skills that will enable a digital curation professional to maintain a strategic view of digital curation, to understand the emerging challenges in digital curation for the cultural heritage sector, and to make informed funding decisions to meet these challenges.

Manager Lens

The skills defined under the Manager Lens are the skills that will enable a digital curation professional to plan and monitor execution of digital curation projects, to recruit and support project teams, and to liaise with a range of internal and external contacts within the cultural heritage sector.

Practitioner Lens

The skills defined under the Practitioner Lens are the skills that will enable a digital curation professional to plan and execute a variety of technical tasks, both individually and as part of a multi-disciplinary team. This includes understanding diverse issues relating to both digital curation in general, and to their specific area of cultural heritage.

<https://digcurv.gla.ac.uk/>

I profili professionali

- Un approccio "empirico" consiste nel verificare se le proprie competenze corrispondono a quanto viene richiesto dal mercato del lavoro!

The screenshot shows the Digital Preservation Coalition (DPC) website's job vacancies page. The header is green with the DPC logo and navigation links: ABOUT, NEWS, DIGITAL PRESERVATION, EVENTS, BLOG. A search icon is in the top right. The main content area is titled 'Job Vacancy' and includes a sub-header 'The DPC advertises relevant digital preservation vacancies for and on behalf of the digital preservation community.' Below this, three job listings are shown, each with a 'READ MORE' button:

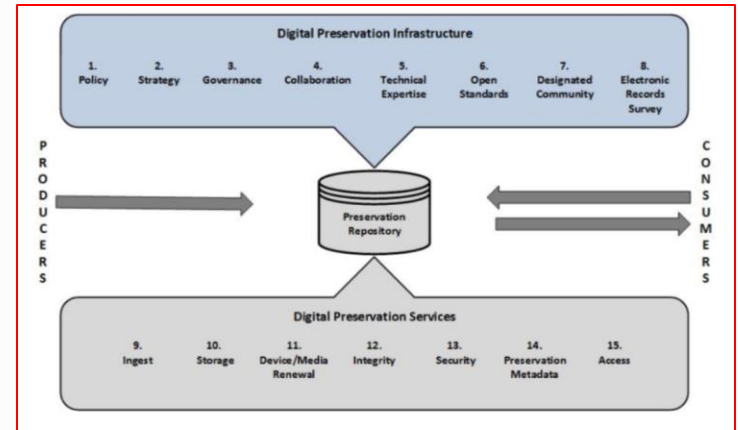
- Vacancy for Head of the Archives Unit at the International Atomic Energy Agency**
3 November 2023 | Austria-Vienna-Vienna-IAEA Headquarters | Net base salary starting at US \$64121 | Fixed Term
- Vacancy for Assistant Digital Preservation Analyst at Rockefeller Archive Center**
17 November 2023 | New York, US | \$ 66,422 - \$96,312 per year | Full-Time
- Vacancy for Director of Digital at the British Library**
12 October 2023 | Boston Spa/London | £110K | Fixed Term

The sidebar on the left contains a 'DPC JOB VACANCIES RSS FEED' button and a 'New Members of the DPC' section. The footer of the page includes the British Museum logo.

<https://www.dpconline.org/news/job-vacancies>

La "maturità digitale" delle organizzazioni

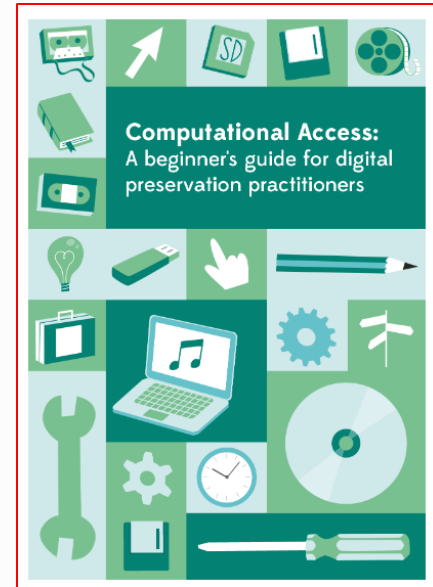
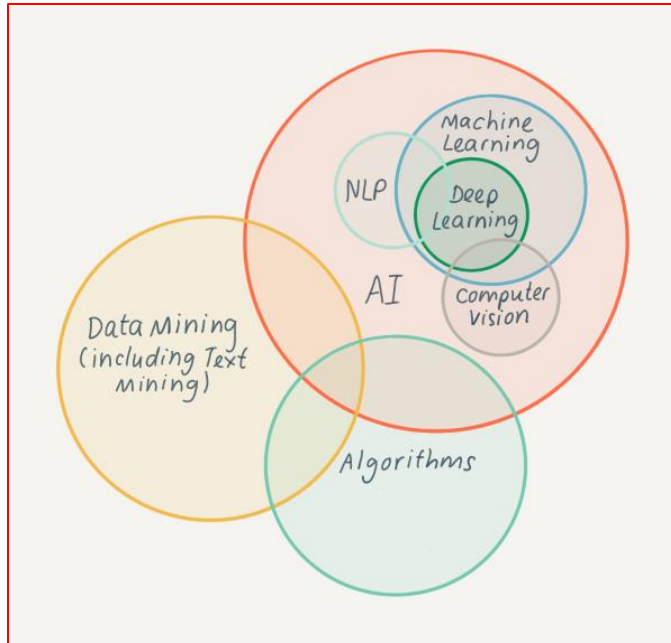
- Esistono diversi modelli per valutare la "maturità digitale" di una organizzazione ma pochi prendono sufficientemente in considerazione anche gli aspetti relativi alla Conservazione.



<https://www.securelyrooted.com/dpcmm>

Conservazione del Digitale e AI

Le Intelligenze Artificiali



[Digital Preservation Coalition](#)
[Computational Access: A beginner's guide for digital preservation practitioners](#)

L'accesso alle risorse e il Computational Access

L'accesso alle risorse digitali può essere migliorato dai sistemi di IA

- **Accesso da parte degli umani**

Necessità di mettere a punto sistemi di selezioni, indicizzazione e recupero dell'informazione efficienti e usabili, grazie anche a Machine learning, IA generativa ecc...

- **Accesso da parte delle macchine (o degli umani ma con l'ausilio di strumenti software "sostanziosi")**

Collections as Data > Computational Access

Es. [Archives Unleashed](#)

Gli archivi del web tra fake news e addestramento dei modelli di AI Generativa

"La profusione di informazioni false da parte di LLM [modelli linguistici di grandi dimensioni o Large Language Models] - a causa di disinformazione intenzionale, pregiudizi della società o allucinazioni - può potenzialmente mettere in dubbio l'intero ambiente informativo, minacciando la nostra capacità di distinguere i fatti dalla finzione..."

Così scrive OpenAI trattando dei rischi di GPT-4 [...]

E proprio a questo hanno pensato al Guardian quando qualcuno li ha contattati cercando un vecchio articolo sul sito. "Un giornalista è stato contattato da una persona che stava facendo delle ricerche per chiedere perché un particolare articolo di molti anni fa fosse stato tolto dal nostro sito", ha twittato Chris Moran, capo dell'innovazione editoriale nella famosa testata inglese. "Era passato un po' di tempo e il giornalista non se lo ricordava. Abbiamo scavato nei nostri sistemi e non ne abbiamo trovato traccia. La persona che lo chiedeva aveva usato ChatGPT per fare la ricerca...La cosa più probabile è che non sia mai esistito. Che sia stata un'allucinazione. Immaginate tutto ciò in un'area incline al complottismo. Queste allucinazioni sono comuni. Potremmo vedere molte teorie del complotto alimentate da articoli "cancellati" che non sono mai stati scritti".

Estratto dalla [Newsletter "Guerre di Rete"](#)

Carola Frediani – 25 marzo 2023

Gli archivi del web tra fake news e addestramento dei modelli di AI Generativa

I sistemi di **Intelligenza Artificiale Generativa** possono però anche, al contrario, essere utilizzati per migliorare l'interazione con la grande quantità di dati presenti negli archivi del Web e il recupero di informazioni da essi.

Al contempo, questi stessi sistemi possono essere migliorati utilizzando gli archivi del web come fonte di dati, per fornire risposte più accurate alle domande dei propri utenti.

*[Navigating Through Archived Websites:
From Text Matching to Generative AI-Enhanced Q&A](https://netpreserveblog.wordpress.com/)*
By Peter Chan, Web Archivist at Stanford Libraries
<https://netpreserveblog.wordpress.com/>
28 giugno 2023