

Conservazione a lungo termine di contenuti digitali

alias: Preservation

Vittore Casarosa, ISTI-CNR e Università di Pisa

- IDC (International Data Corporation) has released a report on the ever-growing “datasphere” (what it calls the collective world’s data), and just like the recent Cisco study, the numbers are staggering
- IDC predicts that the collective sum of the world’s data will grow from 33 zettabytes this year (2018) to a 175 ZB by 2025, for a compounded annual growth rate of 61 percent
- If one were able to store 175 ZB onto BluRay discs, then you would have a stack of discs that could get you to the moon 23 times (David Reinsel, senior vice president at IDC)

<https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html>

- Oggi oltre 5 miliardi di consumatori interagiscono con i dati quotidianamente
- Entro il 2025 i consumatori saranno 6 miliardi (il 75% della popolazione mondiale)
- Nel 2025, ciascuna persona connessa avrà almeno una interazione coi dati ogni 18 secondi
- Molte di queste interazioni avvengono attraverso i miliardi di dispositivi IoT (Internet of Things) connessi in tutto il mondo, che ci si aspetta che creino oltre 90 ZB di dati nel 2025

<https://www.seagate.com/it/it/our-story/data-age-2025/>

Scala dei contenuti digitali

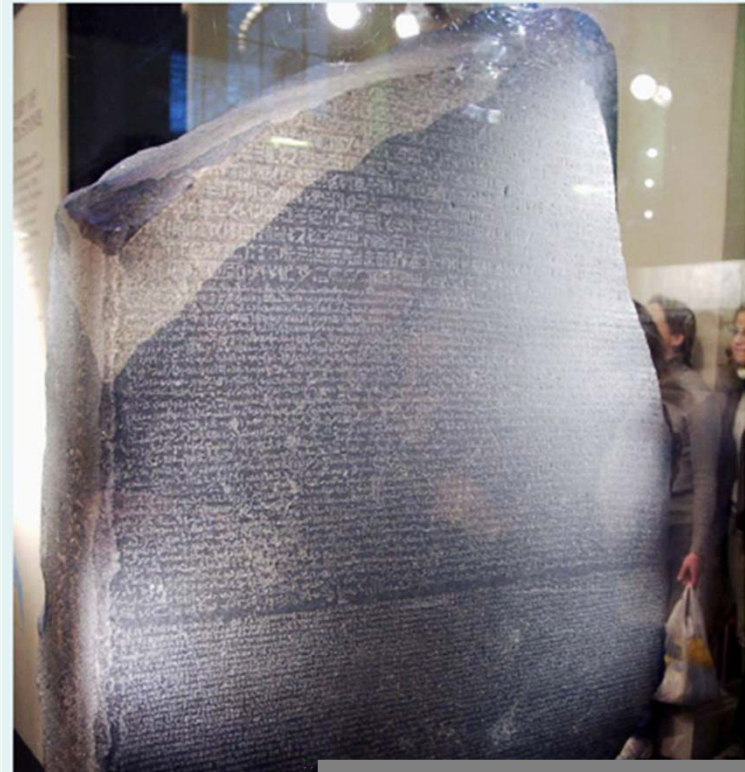
1000	k	kilo
1000 ²	M	mega
1000 ³	G	giga
1000 ⁴	T	tera
1000 ⁵	P	peta
1000 ⁶	E	exa

1 GB = 1000 MB

1 TB = 1000 GB

1 EB = un milione TB

1 ZB = un miliardo TB



Supporto

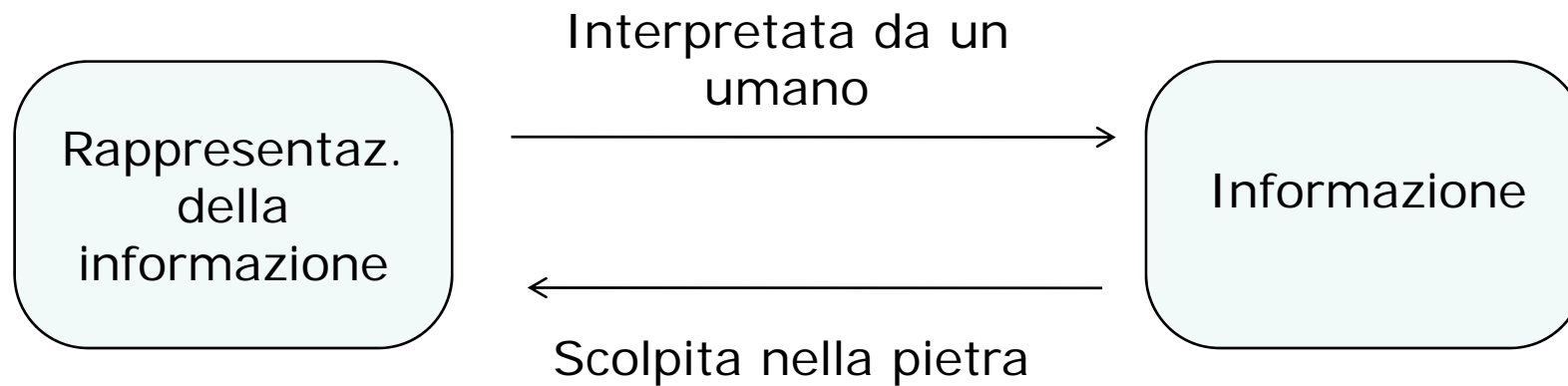
Materiale solido
(granodiorite)
114 x 72 x 28 cm
760 kg

Codifica

Caratteri leggibili da un
umano, in tre lingue
(geroglifici, demotico, greco
antico)

Come si ottiene l'informazione ?

Un umano, capace di
leggere (almeno) uno dei
tre testi scritti





Hardware

- Memoria (RAM, chiavetta, hard disk, ...)
- Visualizzazione (schermo, stampante, ...)

Software

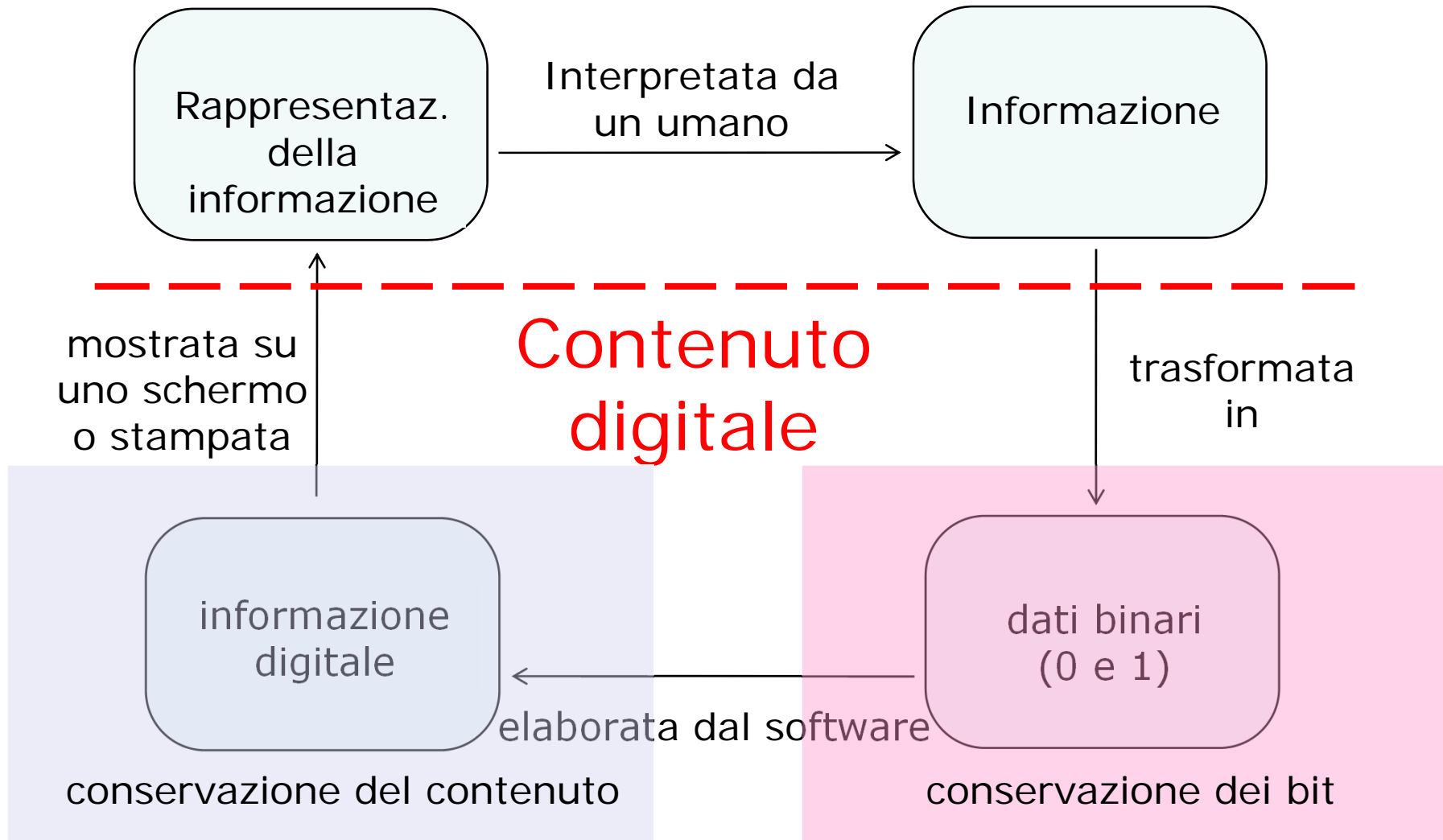
- Sistema Operativo
- Applicazioni (web browser, text editor, ...)

Codifica

- Machine-readable: dati binari
- Human-readable: caratteri, immagini

Come si ottiene l'informazione ?

Un umano, capace di interpretare quanto mostrato sullo schermo



- Cosa è un oggetto digitale ?
 - Alcuni sono la “rappresentazione digitale” di oggetti fisici, come libri, giornali, posta, video, ecc. I
 - Altri non hanno una vera controparte fisica, come ad esempio pagine Web, mappe annotate (GIS), modelli 3D di strutture chimiche, ecc.
- Tre punti di vista differenti per un oggetto digitale
 - Fisico: i bit memorizzati su un dispositivo di memoria
 - Logico: definisce come i bit vengono usati dal software in modo da rendere l’oggetto digitale “usabile” da un umano (o un altro software)
 - Concettuale: quale e’ il significato (il contenuto) di un oggetto digitale nel “mondo reale”

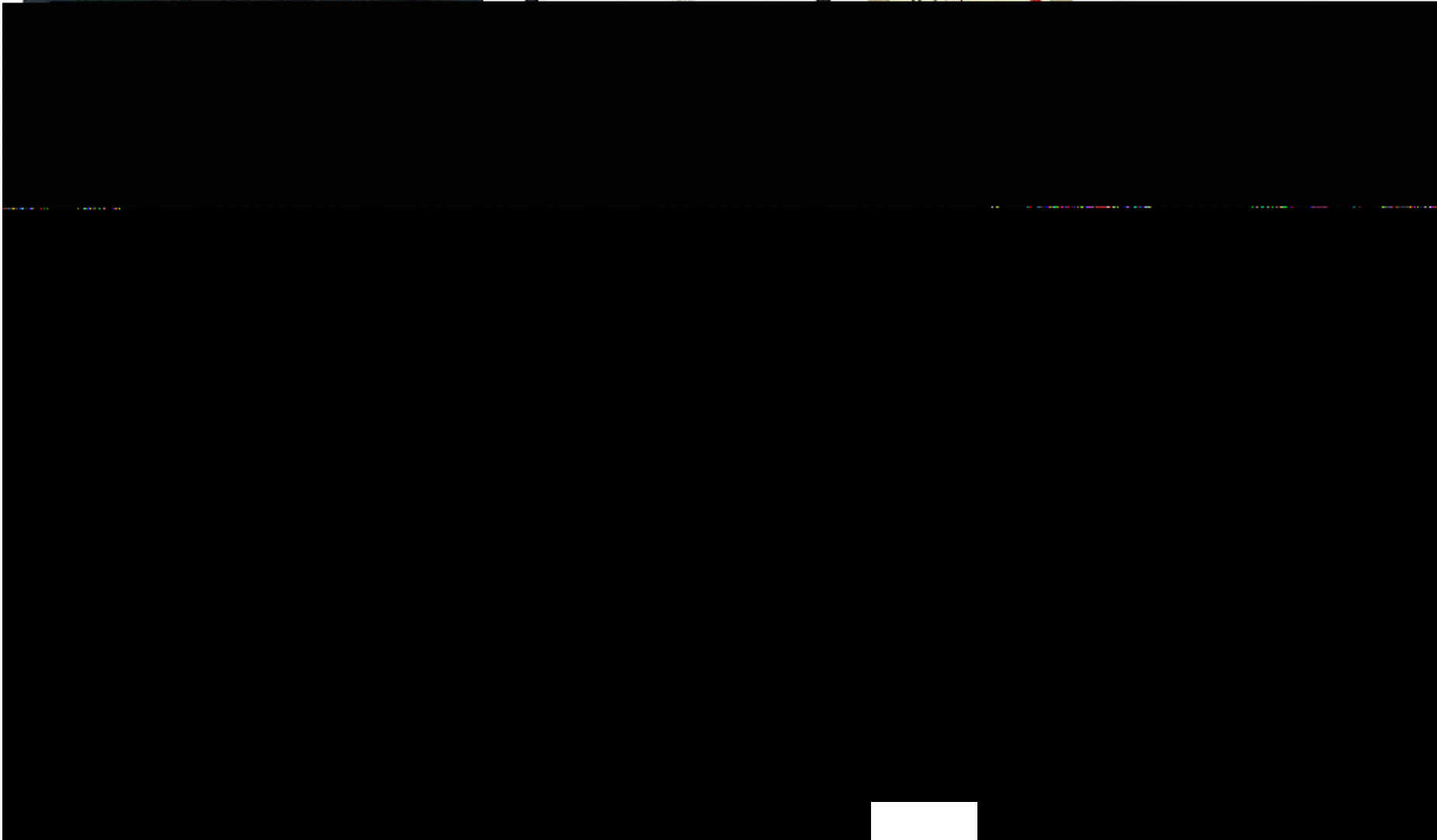
- Obsolescenza del supporto fisico
- Obsolescenza dei dispositivi di lettura

Conservazione dei bit

- Obsolescenza del formato
- Obsolescenza del software
- Perdita del contesto

Conservazione dell'informazione

Chi li ha visti ?



Durata del supporto fisico

Media	Estimated Lifespan
Magnetic data (tapes)	Up to 10 years
Nintendo cartridge	10-20 years
Floppy disk	10-20 years
CDs and DVDs	5-10 unrecorded, 2-5 recorded
Blu-Ray	Not certain, probably over 2-5 recorded
M-Disc	1,000 years (theoretically)
Hard disk	3-5 years
Flash storage	5-10 years or more (depends on write cycles)

<https://blog.storagecraft.com/data-storage-lifespan/>

- Migrazione (per obsolescenza del supporto fisico)
 - Copiare regolarmente i dati su un nuovo supporto fisico
 - Necessità di automatizzare il processo
- Musei dell'hardware (per obsolescenza dei dispositivi di lettura)
 - Costoso
 - Spesso dipende da altro hardware obsoleto

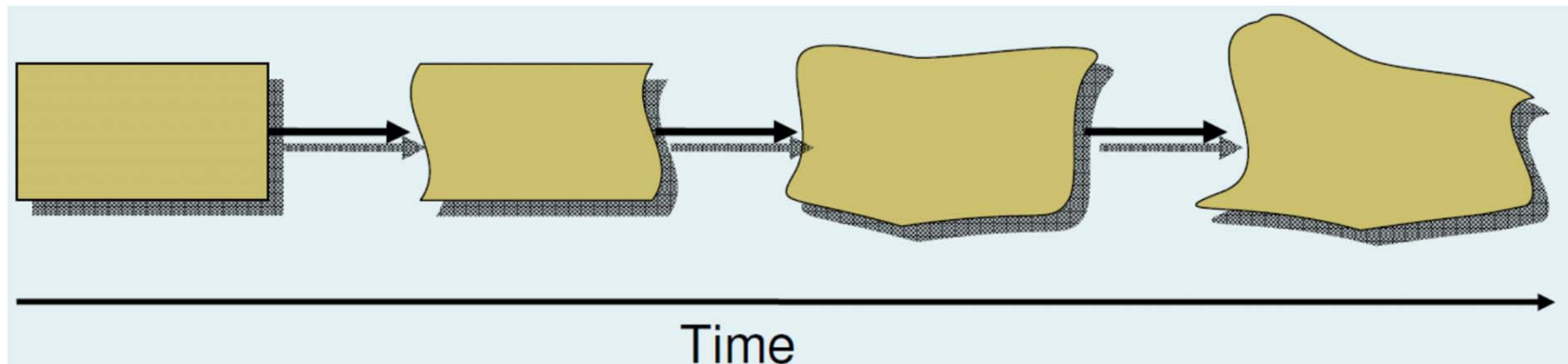
- **Problemi**

- I diversi formati dei dati (tipi di file) non sono supportati per sempre
- Non si sa più come interpretare la stringa di bit

- **Soluzioni**

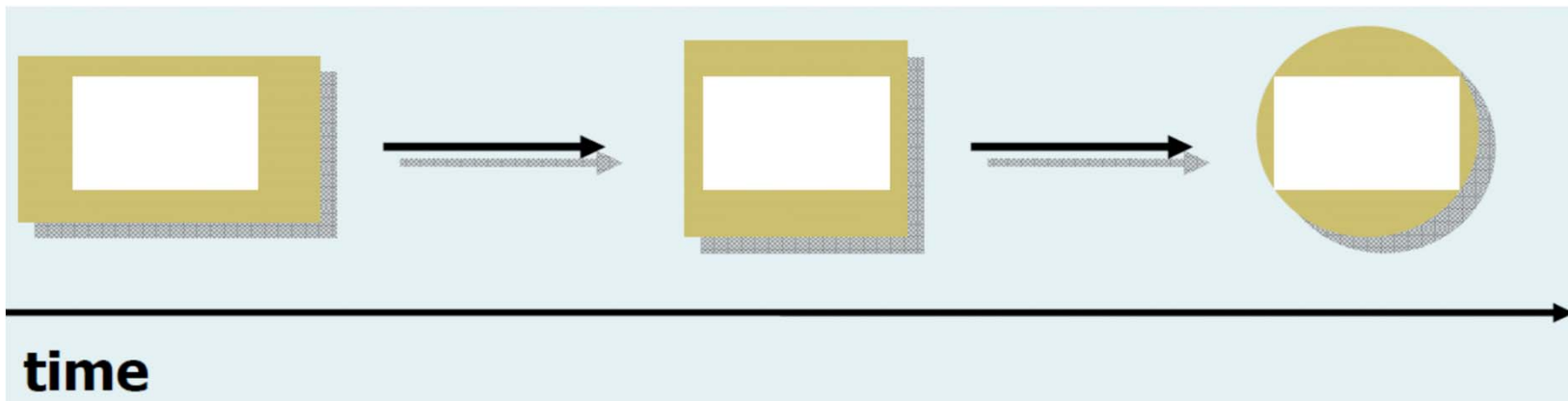
- Migrazione
 - Convertire le stringhe di bit in un formato più recente, possibilmente “open and standardized”
- Emulazione
 - Creare un ambiente “virtuale” che riproduca esattamente l’ambiente originale

- La migrazione cambia l'oggetto digitale
 - L'oggetto digitale è disponibile nell'ambiente attuale
 - Il riutilizzo dell'informazione è semplice



- Problemi
 - Si possono introdurre delle inconsistenze
 - Si può perdere qualche funzionalità
 - È difficile garantire la qualità dell'oggetto migrato

- Emulazione significa cambiare l'ambiente
 - Ricreare con software l'ambiente "originale", incluse le sue funzionalità
 - L'oggetto digitale non viene cambiato



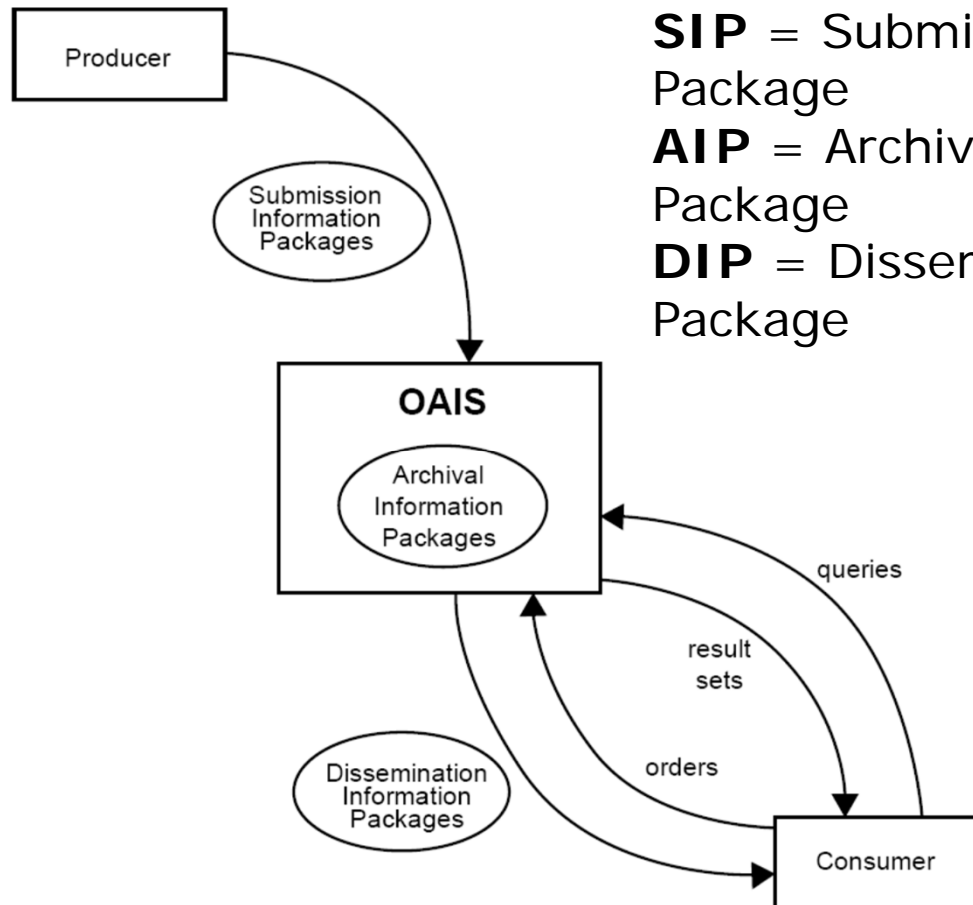
- Problemi
 - Tecnicamente più impegnativo
 - Necessaria conoscenza dettagliata dell'ambiente originale
 - Necessità di migrare l'emulatore

Open Archival Information System

- “Open” significa che il modello di riferimento deve essere disponibile per tutti
- “Archive” si riferisce a una "organization that intends to preserve information for access and use by a **designated community**”
- “Information” si riferisce a quanto conservato nell’archivio in quanto necessita di una “conservazione a lungo termine”
 - “Lungo termine” significa lungo abbastanza perche’ ci siano preoccupazioni sull’impatto del cambiamento delle tecnologia

- ISO 14721:2012 (anche CCSDS 650.0-M-2)
 - OAIS – a reference model for what is required for an archive to provide long-term preservation of digital information
 - <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- ISO 16363:2013 (anche CCSDS 652.0-M-1)
 - Audit and certification of **trustworthy digital repositories** – sets out comprehensive metrics for what an archive must do, based on OAIS
 - <https://public.ccsds.org/Pubs/652x0m1.pdf>
- ISO 16919:2014 (anche CCSDS 652.1-M-2)
 - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories – specifies the competencies and requirements on auditing bodies
 - <https://public.ccsds.org/Pubs/652x1m2.pdf>

Interazioni con un archivio OAIS

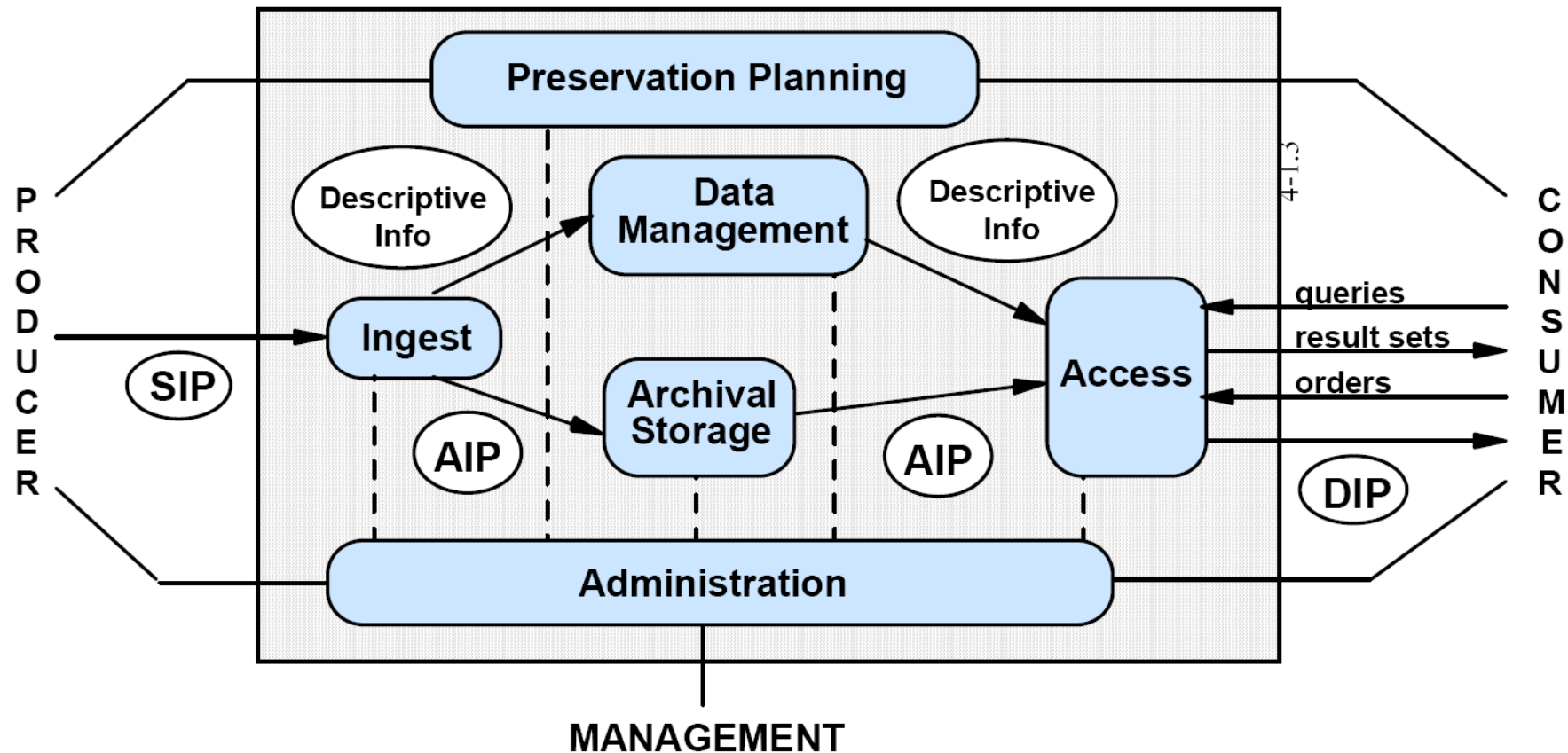


SIP = Submission Information Package

AIP = Archival Information Package

DIP = Dissemination Information Package

Modello funzionale di archivio OAIS



- Il contenuto di un sito Web in generale è transitorio e cambia spesso, ma può essere di interesse la sua “conservazione”
- Impegnativa dal punto di vista tecnico, in quanto può contenere molti tipi di oggetti digitali (pagine HTML, immagini, video, suoni, blogs, informazioni amministrative, ecc)
- Impegnativa dal punto di vista legale, in quanto può portare a duplicare contenuti illegali (diffamazione, pornografia, ecc.)
- Ciò nonostante, di estremo interesse in quanto la maggior parte dei contenuti d un sito Web in genere non ha controparte fisica
- Archiviare un sito Web significa archiviare una “foto” di tutte le sue pagine fata ad un certo punto del tempo
- Internet Archives sta archiviando siti Web dal 1996, e li rende accessibili tramite la “WayBack Machine”

WayBack Machine nel 2007

The screenshot shows the Internet Archive Wayback Machine interface in a Mozilla browser window. The browser's address bar displays the URL `http://www.archive.org/web/web.php`. The page features a navigation menu with links for [Web](#), [Moving Images](#), [Texts](#), [Audio](#), [Software](#), [Education](#), [Patron Info](#), and [About IA](#). Below the menu is a search bar with the text "Wayback Machine" and a "GO!" button. The main content area is divided into several sections:

- About the Wayback Machine:** A text block explaining the service, with the phrase "Browse through 85 billion web pages archived from 1996 to a few months ago." circled in red. It also includes a link to <http://archive.bibalex.org>.
- The Wayback Machine:** A section with a text input field containing "http://", a "Take Me Back" button, and a link to [Advanced Search](#).
- Around the World in 2 Billion Pages:** A text block announcing a 2 billion page web crawl in 2007, with a link to [collection](#).
- Web Archiving Services:** A section with the ARCHIVE-IT logo and the slogan "ARCHIVING THE INTERNET FOR FUTURE GENERATIONS COLLECT IT, MANAGE IT, SEARCH IT...ARCHIVE-IT".

The browser's status bar at the bottom shows "Done" and various system icons.

Internet Archive Wayback Machine - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop Search Print

INTERNET ARCHIVE
WayBackMachine

Enter Web Address: All [Adv. Search](#) [Compare Archive Pages](#)

Searched for <http://www.uwe.ac.uk/> 636 Results

Note some duplicates are not shown. [See all.](#)
 * denotes when site was updated.
 Material typically becomes available here 6 months after collection. [See FAQ.](#)

Search Results for Jan 01, 1996 - Aug 30, 2007

1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
0 pages	6 pages	3 pages	6 pages	18 pages	0 pages	1 pages	18 pages	117 pages	204 pages	69 page
	Jan 03, 1997 * Jan 31, 1997 Jan 31, 1997 * Mar 15, 1997 * Apr 27, 1997 Dec 10, 1997 *	Jan 29, 1998 * Feb 24, 1998 May 29, 1998 *	Jan 17, 1999 * Jan 25, 1999 Jan 29, 1999 * Feb 08, 1999 * Apr 21, 1999 * Apr 27, 1999 *	Feb 29, 2000 * Mar 01, 2000 * Mar 02, 2000 * Mar 03, 2000 * Apr 07, 2000 * Apr 14, 2000 * Apr 18, 2000 Apr 20, 2000 May 02, 2000 May 10, 2000 * May 11, 2000 * May 19, 2000 * Jun 19, 2000 * Jun 21, 2000 * Jun 22, 2000 * Aug 16, 2000 * Aug 24, 2000 *		Dec 21, 2002 *	Jan 25, 2003 * Jan 27, 2003 Feb 02, 2003 * Mar 21, 2003 * Apr 02, 2003 * Apr 22, 2003 * Apr 24, 2003 * May 27, 2003 * Jun 12, 2003 * Jun 19, 2003 * Jul 19, 2003 * Jul 24, 2003 * Aug 01, 2003 * Sep 25, 2003 * Oct 08, 2003 * Oct 28, 2003 * Dec 03, 2003 *	Jan 22, 2004 Feb 05, 2004 * Mar 22, 2004 * Apr 27, 2004 * Jun 02, 2004 * Jun 03, 2004 * Jun 04, 2004 * Jun 08, 2004 Jun 10, 2004 * Jun 12, 2004 * Jun 14, 2004 * Jun 15, 2004 Jun 16, 2004 Jun 18, 2004 Jun 19, 2004 * Jun 23, 2004 * Jun 24, 2004	Jan 03, 2005 Jan 11, 2005 * Jan 13, 2005 Jan 14, 2005 Jan 15, 2005 Jan 20, 2005 Jan 21, 2005 Jan 23, 2005 Jan 25, 2005 Jan 26, 2005 * Jan 30, 2005 Feb 03, 2005 Feb 04, 2005 Feb 05, 2005 Feb 06, 2005 Feb 07, 2005 Feb 11, 2005	Jan 01, 2006 Jan 02, 2006 Jan 03, 2006 Jan 04, 2006 Jan 06, 2006 Jan 06, 2006 Jan 07, 2006 Jan 08, 2006 Jan 10, 2006 Jan 11, 2006 Jan 12, 2006 Jan 15, 2006 Jan 17, 2006 Jan 18, 2006 Jan 27, 2006 Jan 28, 2006 Jan 31, 2006

WayBack Machine nel 2019

INTERNET ARCHIVE
WayBack Machine

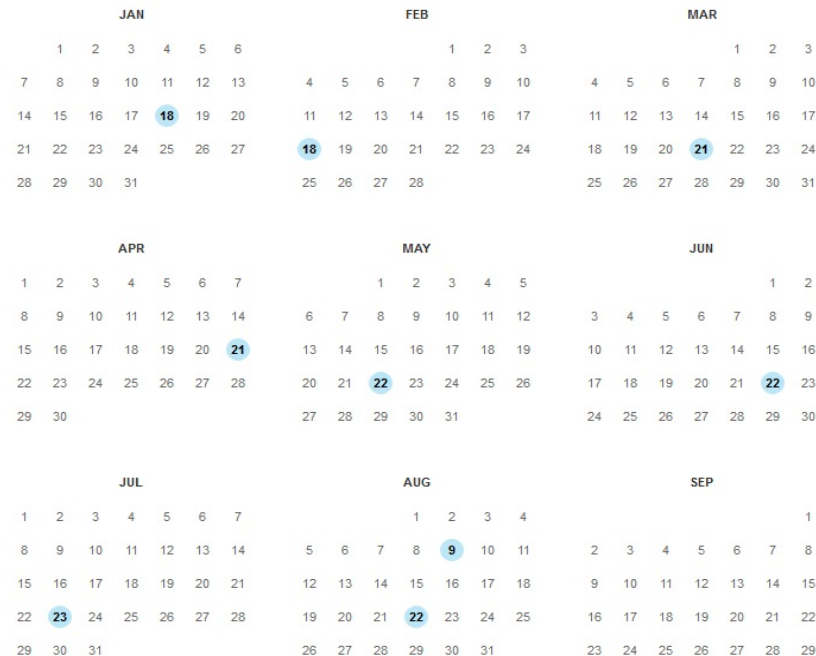
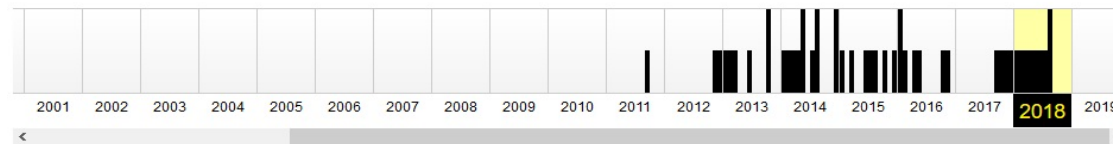
Explore more than 349 billion web pages saved over time

http://www.mab-italia.org

Find the Wayback Machine useful? [DONATE](#)

Saved 47 times between September 16, 2011 and August 22, 2018.

[Summary of mab-italia.org](#) · [Site Map of mab-italia.org](#)



INTERNET ARCHIVE WayBackMachine <http://www.mab-italia.org/index.php> Go FEB MAR JUN 2012 15 2013 2014



Q cerca...

HOME CHI SIAMO COMITATI TERRITORIALI STATI GENERALI 2012

Comunicati

EMERGENZA TERREMOTO IN EMILIA ROMAGNA



Comunicato
congiunto
AIB - ANAI - ICOM

LETTERA DI INTENTI DI MAB



Per il rilancio del sistema culturale italiano
28 marzo 2012

Eventi e iniziative

Stati generali dei professionisti del patrimonio culturale

Archivi, biblioteche e musei: agenda per un futuro sostenibile
22 - 23 novembre 2012
Milano, Palazzo Lombardia

MAB Italia

[Leggi tutto...](#)

Appello contro la soppressione delle competenze degli Enti locali sulla cultura

Un appello di MAB Italia contro la sottrazione a Province e Comuni delle competenze fondamentali sulla cultura
Roma-Milano, 5 dicembre 2012

[Leggi tutto...](#)

ANTEPRIMA Grand Tour Cultura 2012-2013: luoghi "comuni" della creatività

Dal 21 dicembre 2012 al 20 gennaio 2013.
Il MAB Marche e l'Assessorato alla Cultura della Regione Marche, propongono GRAND TOUR CULTURA sul tema Musei - Archivi - Biblioteche: luoghi "comuni" della creatività, per coinvolgere le istituzioni culturali marchigiane in attività che possano sostenere, sollecitare e sviluppare la creatività dei territori. È possibile scaricare di seguito l'elenco, in formato pdf., delle iniziative organizzate dalle Istituzioni che hanno aderito all'ANTEPRIMA, che si svolge dal 21 dicembre 2012 al 20 gennaio 2013.

[Leggi tutto...](#)

Musei, archivi e biblioteche della Campania in dialogo

Martedì 11 dicembre 2012, ore 10.00

MARCHE

AREA RISERVATA

Nome utente

Password

Ricordati di me

Login

[Password dimenticata?](#)

[Nome utente dimenticato?](#)

CALENDARIO EVENTI

Marzo 2013

L	M	M	G	V	S	D
25	26	27	28	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

I PROSSIMI APPUNTAMENTI

Nessun evento

Grazie per l'attenzione

Vittore Casarosa
casarosa@isti.cnr.it