

# Il Progetto Arkive@unimi

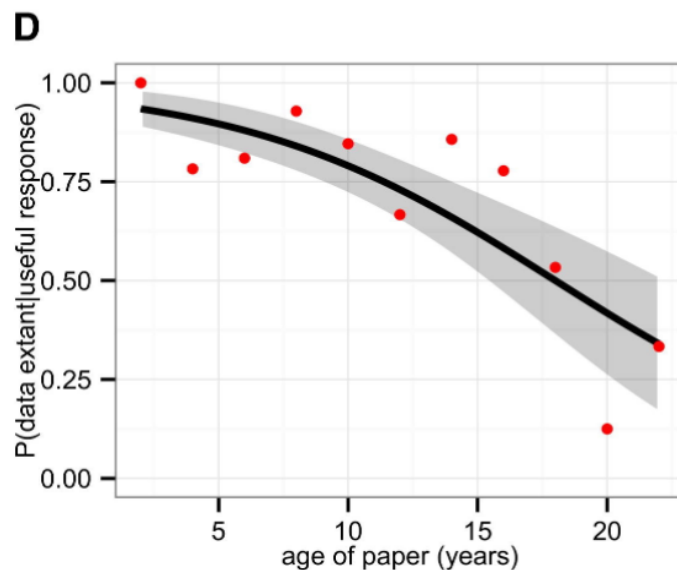
*Archiviazione e cura 'long term' dei contenuti digitali prodotti dall'Università degli Studi di Milano:  
un approccio olistico orientato alla sostenibilità e alla FAIRness*

Federica Zanardini  
Giorgio Bagnato  
Loredana Rollandi  
Matteo Zoppi

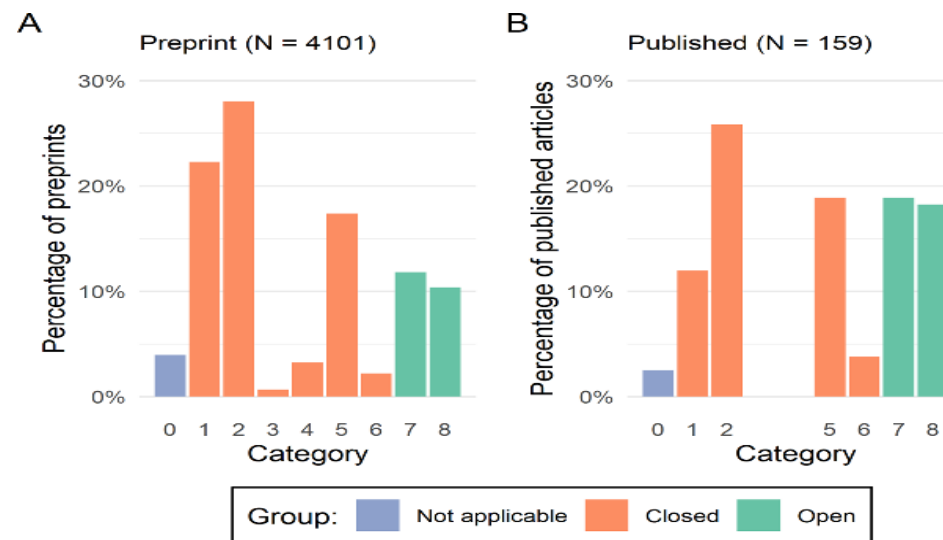
Università degli Studi di Milano

CONDIZIONI  
VISIONI  
Conferenza GARR 2022  
Palermo 18-20 maggio

# L'«estinzione» dei Dataset



Perdita annuale di dataset a partire dalla pubblicazione<sup>1</sup>



% dataset *closed* o non disponibili in articoli corredati da *availability statement* «open»<sup>2</sup>

<sup>1</sup>*The Availability of Research Data Declines Rapidly with Article Age.* Vines, Albert, et al. (2014) 1, Vol. Current Biology, 24 ( <https://doi.org/10.1016/j.cub.2013.11.014> )

<sup>2</sup>*A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts.* McGuinness LA, Sheppard AL (2021) PLoS ONE 16(5): e0250887. ( <https://doi.org/10.1371/journal.pone.0250887> )

# Accesso *long term* a collezioni digitalizzate beni culturali

- Non esistono studi sistematici analoghi nell'ambito Beni Culturali:

BC=collezioni digitalizzate dalle Università (biblioteche, archivi, musei, dipartimenti) nell'ambito di progetti di ricerca o terza missione, per lo più umanistici ma non solo (es. collezioni mineralogiche, erbari, strumenti scientifici antichi)

- Project Work <sup>1</sup> nell'ambito del master Polimi in Management dell'Università e della Ricerca:

→ 7 progetti in ambito BC a rischio perdita di accesso a causa di obsolescenza

- Obsolescenza formati, tecnologie e infrastrutture
- Grandi sforzi per recuperare piccoli progetti
- Mancanza di metadattazione adeguata

- *La biblioteca ideale di Giordano Bruno*
- *La Bibbia nel '500*
- *Imago Historiae*
- *Giorgio Vasari. Le vite - Edizione Giuntina e Torrentiniana*
- *La fortuna visiva di Pompei - Archivio di immagini e testi*
- *Monumenta Rariora*
- *BiViO - Biblioteca Virtuale Online*

→ Cause diverse ma problema simile: alto rischio perdita dati in formato digitale

<sup>1</sup>« *Gestione dei dati digitali della ricerca: linee guida per garantirne la produzione e conservazione sostenibile*» A.Isolani, L.Mezzina - Project Work Master SUM VIII edizione (2020)

# La situazione di UNIMI

Inizio 2021: questionario

- dipartimenti
- musei
- biblioteche
- Archivi



- Quanto storage? (pregresso, corrente e stime future)
- Quali formati?
- Quali soluzioni di archiviazione?
- Usati standard o linee guida?
- Usata metadattazione?

Risultati:

- Storage: 60PB per corrente e pregresso, 20PB crescita annuale
- Formati: enorme varietà
- Soluzioni di archiviazione: grande varietà e frammentazione
- Uso di standard e linee guida: quasi soltanto in ambito biblioteche e musei
- Uso di metadattazione: quasi soltanto in ambito biblioteche e musei

# Il progetto Arkive

- Grande quantità di storage
- Scalabilità
- Interoperabilità
- Automazione dei processi e delle operazioni di *data curation*
- Policy per la cura e la gestione dei dati
- Affidabilità, certificazione dell'archivio

# Le soluzioni

- Storage/scalabilità/sicurezza: 

- Cura dei dati e automazione dei processi:

iRODS

- Interoperabilità e *long term* preservation:

- Moduli base Eudat:
  - Autenticazione federata: B2ACCESS
  - Condivisione dei dati: B2SHARE

- Apertura dei dati:

- B2FIND

- Dati FAIR: integrazione con Dataverse ([dataverse.unimi.it](https://dataverse.unimi.it))



Dataverse 

# Prospettive di Arkive@unimi.it

- Federazione con rete Eudat/EOSC



- Integrazione con archivio digitale dei Beni Culturali del Ministero della Cultura



- Partecipazione a Europeaana



# Grazie

*[federica.zanardini@unimi.it](mailto:federica.zanardini@unimi.it)*

**CONDIZIONI**  
**VISIONI**  
Conferenza GARR 2022  
Palermo 18-20 maggio