

*"Le biblioteche hanno  
gestito i dati per  
secoli"*

**Marcy Strong**

(responsabile del servizio di metadati presso le biblioteche River Campus dell'Università di Rochester, New York)

# Il bibliotecario nell'era dei Big Data

Cosa sono i Big Data

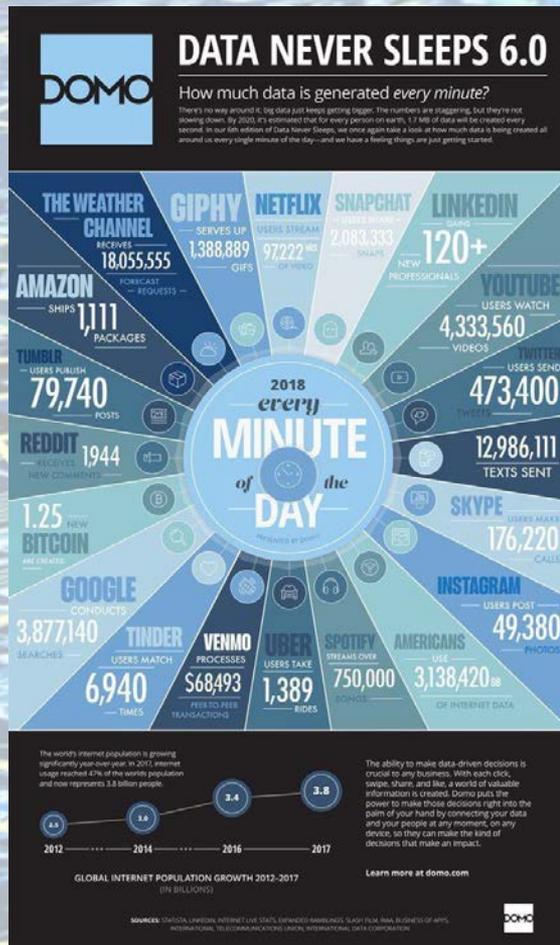
Big Data e Biblioteche

Il nuovo bibliotecario



# IL DILUVIO DEI DATI

# Il Diluvio, dati generati su internet ogni minuto (2018)



- **3.877.140 ricerche su Google**
- **4.330.560 video visualizzati su YouTube**
- **473.400 Twitter**
- **97.222 film in streaming su Netflix**

Il bibliotecario nell'era dei Big Data  
Bibliosan

# Big Data: definizione

## la regola 3V + 1V

- La definizione di big data è anche legata alle tre caratteristiche che i dati devono avere
  - **Volume**: nel dataset devono essere presenti grandi quantità di dati, nell'ordine degli zettabyte
  - **Varietà**: i dati provengono da fonti eterogenee, quindi hanno natura diversa e non strutturata. Oltre che valori numerici possono essere comprese immagini, parole, video, etc.;
  - **Velocità**: si riferisce alla velocità con cui i dati vengono generati, svolgendo l'analisi in tempo reale.
- Con l'aggiunta di una quarta V, il **Valore/Veridicità**
  - Dati devono essere affidabili e capaci di fornire analisi utili e interessanti

# Big Data, l'era dell'incertezza

- L'ossessione per l'esattezza è un costrutto dell'era analogica caratterizzata da una costante carenza di informazioni (small data)
- Rinuncia dell'esattezza a livello micro per comprensione generale (trend) dei fenomeni a livello macro: i Big Data trasformano le cifre in qualcosa di probabilistico piuttosto che di preciso
- Per massimizzare i benefici dei Big Data dobbiamo accettare un certo grado di confusione come ineludibile e non come un elemento di disturbo da eliminare

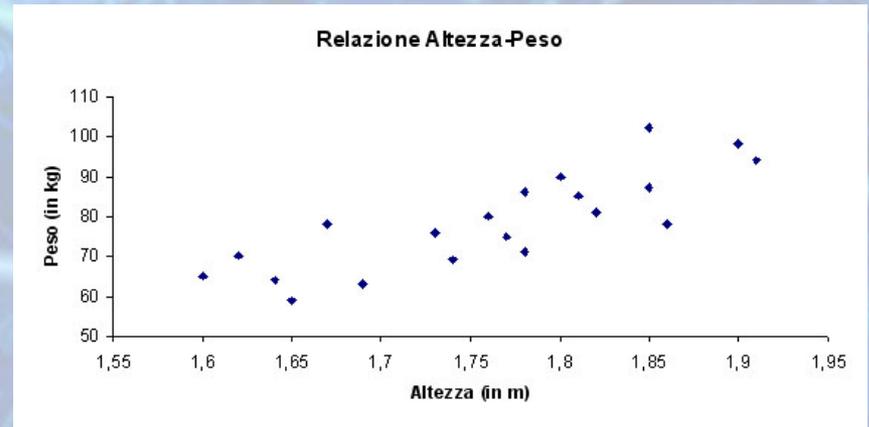
# L'era dell'incertezza... ma predittiva

***“lasciare cadere le proprie  
ossessioni sui fenomeni di  
causalità in cambio di  
correlazioni semplici in quanto  
in futuro interesserà sempre  
di più sapere il cosa accade e  
non il perché accade.....”***



# Correlazioni

- Per correlazione si intende una relazione tra 2 variabili statistiche tale che a ciascun valore della prima variabile corrisponda «con una certa regolarità» un valore della seconda
- Non è un rapporto di causa-effetto, ma semplicemente la tendenza di una variabile a variare in funzione di un'altra



# Il controverso caso Google Flu Trends

- Le correlazioni si sono basate sulle interrogazioni di Google fatte a seguito del clamore delle notizie sull'influenza.
- Di conseguenza il numero crescente delle persone interessate non corrispondeva al numero reale delle persone infette.
- Risultato: Google Flu Trends ha avuto risultati discordanti rispetto al conteggio reale effettuato dai Centri di controllo della malattia sul territorio.

# Un Big Data – Rischio: la privacy

- I Big Data possono rendere inefficaci i principali meccanismi tecnici e legali con cui oggi tentiamo di tutelare la privacy
- I Big Data scavalgano il «Consenso informato»
  - Finora le leggi che tutelano la privacy hanno lasciato il controllo del rischio agli individui consentendo loro di decidere se come e da parte di chi potevano essere processate le proprie informazioni personale
- Ora con i Big Data il grosso del valore dei dati si è spostato sull'utilizzo secondario, cioè su utilizzi che neanche si potevano immaginare al momento della raccolta
  - Per questi motivi deve essere pensato un sistema di tutela della privacy meno focalizzato sul consenso individuale e più sugli utilizzatori dei dati, cioè una “privacy tramite responsabilizzazione” di coloro che acquisiscono il diritto di sfruttare i dati.

# Big Data, ma con metodo scientifico

- Nel 2012, una ricerca di Danah Boyd e Kate Crawford “[Critical questions for big data](#)”, si incaricava di ridimensionare i “Big Data fundamentalismi” confutando che i Big Data “*possano offrire di per sé una più alta forma di intelligenza e conoscenza dando vita a intuizioni e/o rivelazioni prima impossibili*”
- L’accumulazione dei dati può rappresentare una preziosa fonte di informazione, ma senza una comprensione teorica, senza una metodologia è impossibile aggiungere tridimensionalità alle interpretazioni. In pratica, è impossibile studiare e capire i dati.

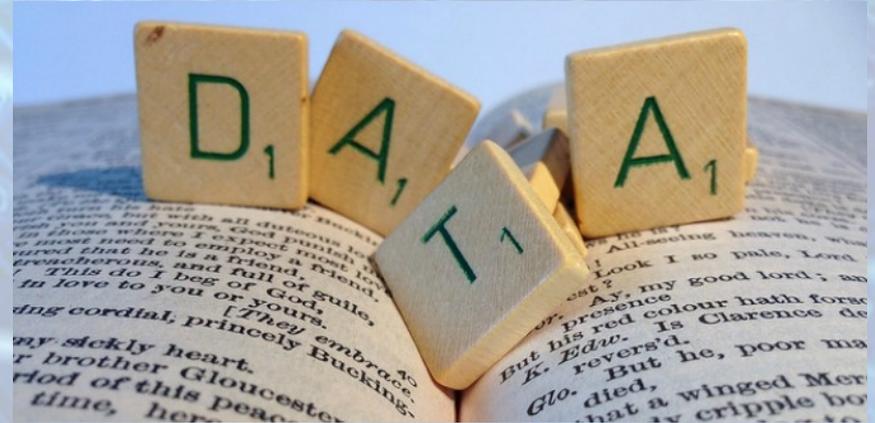


# Big Data e Biblioteche

# Big Data della ricerca scientifica

- Sono fatti, osservazioni o esperienze sulle quali è costruita e/o verificata una teoria. I dati possono essere numerici, descrittivi, sonori o visivi.
- Vengono raccolti, osservati e/o creati, per scopi di analisi e soprattutto per produrre risultati di ricerca originali

Gestirli e amministrarli in modo consapevole e sistematico è un imperativo dal quale ogni istituzione di ricerca pubblica o privata non può prescindere



# Settori dove è più richiesto il supporto delle biblioteche

- **Consulenza su tipologia e formato dati**
- **Stime per lo storage**
- **Supporto tecnico e legale (copyright) per i ricercatori**

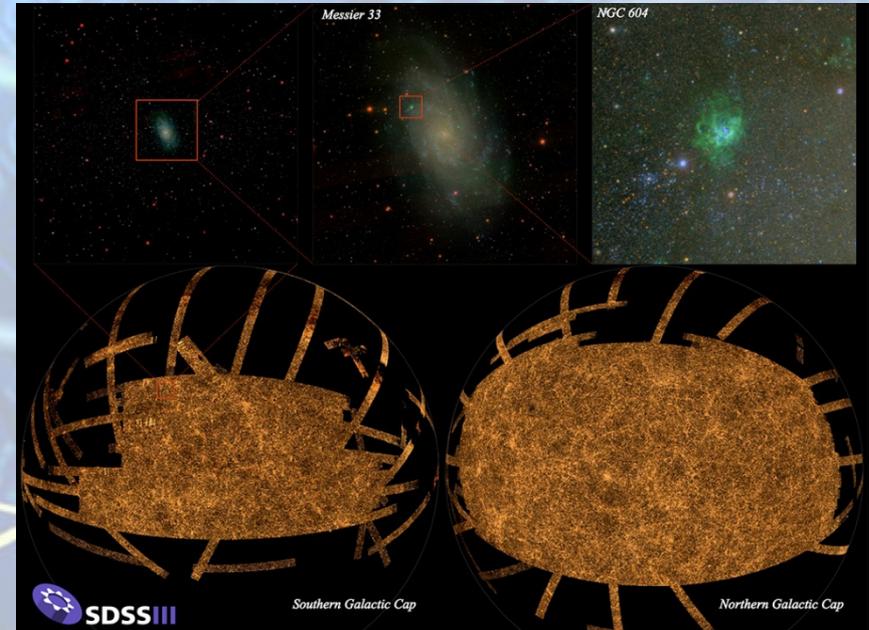
(Indagine università austriache e britanniche 2015: dati in linea con indagini anche in altri Paesi)

# Biblioteche e Big Data

- le biblioteche hanno cominciato a misurarsi con i Big Data offrendo supporto per la gestione dei dati della ricerca e progettando piattaforme per l'accesso pubblico agli Open Data.
- Negli USA, secondo uno studio dell'Università del Tennessee, su oltre 100 biblioteche universitarie, almeno il 40% è impegnato a sviluppare programmi per supportare gli scienziati nelle procedure di "big data curation"
- James R. Jacobs, bibliotecario di Stanford, sostiene che *"le biblioteche per le loro competenze negli standard dei metadati e nelle strategie di conservazione, possono candidarsi a partner fondamentali nei processi di costruzione di sistemi pubblici basati sugli Open (Big) Data"*
- il coordinatore del sistema bibliotecario scientifico dell'Università di New Mexico – William Michener – ha fatto osservare che nessuno meglio delle biblioteche può occuparsi della conservazione dei datasets scientifici generati con finanziamenti pubblici.

# Biblioteche e dati della ricerca

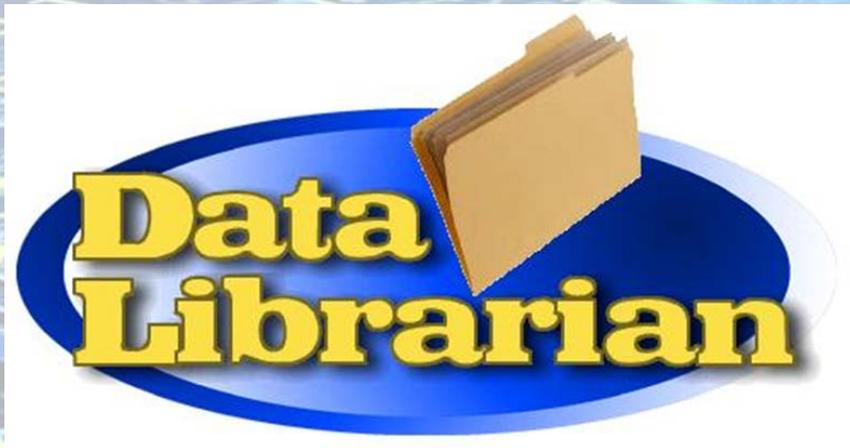
- Lo “Sloan Digital Sky Survey ” è un progetto astronomico molto importante che prevede la mappatura dello spazio profondo.
- Finora sono stati catalogati circa 100 milioni di stelle, 1 milione di galassie e 100 mila quasar.
- Un’impresa titanica con conseguente produzione di un’enorme massa di dati
- Per gestire la complessità di questi “big data” sono arrivate in aiuto le biblioteche della John Hopkins University di Baltimora



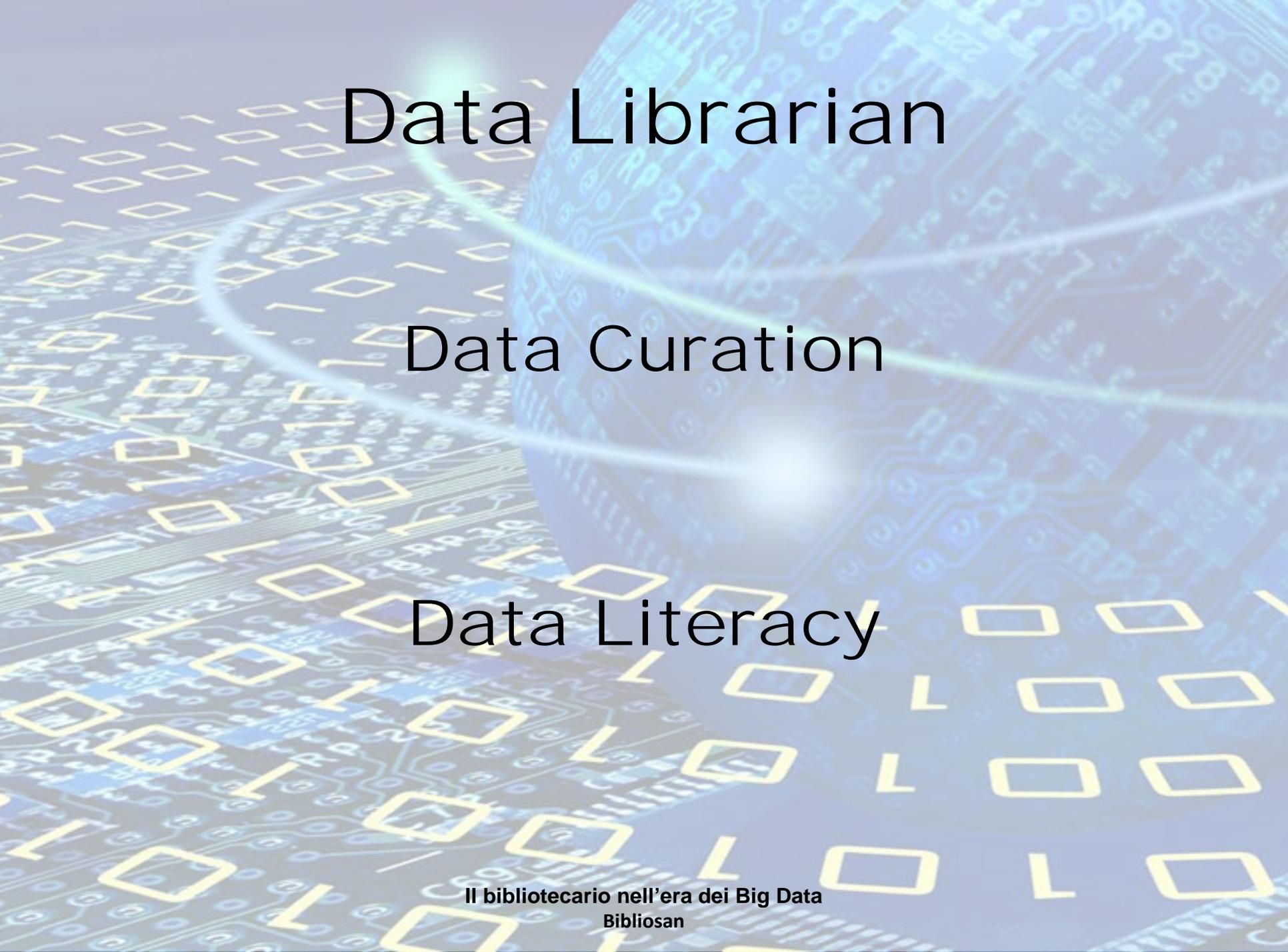
The background features a large, semi-transparent blue sphere on the right side, set against a light blue gradient. A glowing green line curves around the sphere. The foreground is filled with a grid of yellow squares, some of which are slightly offset, creating a sense of depth and movement. The overall aesthetic is clean, modern, and tech-oriented.

# Il nuovo bibliotecario

# La sfida, un “nuovo” bibliotecario?



- il ruolo del bibliotecario sembra destinato a cambiare drasticamente
- Già ora, nei paesi anglosassoni e del nord Europa, si sta affermando una nuova figura professionale: il “*liaison librarian*”
- ovvero un bibliotecario, che a stretto contatto con il mondo della ricerca universitaria, svolge attività qualificate di consulenza.
- Ma, questa nuova definizione, potrebbe presto essere a sua volta scavalcata, da un’altra ancora più nuova anch’essa figlia della rivoluzione dei Big Data
- quella di “*data librarian*”, vale a dire un bibliotecario anche e soprattutto capace di “affrontare” grandi masse di dati maneggiando gli strumenti della “*data curation*”.



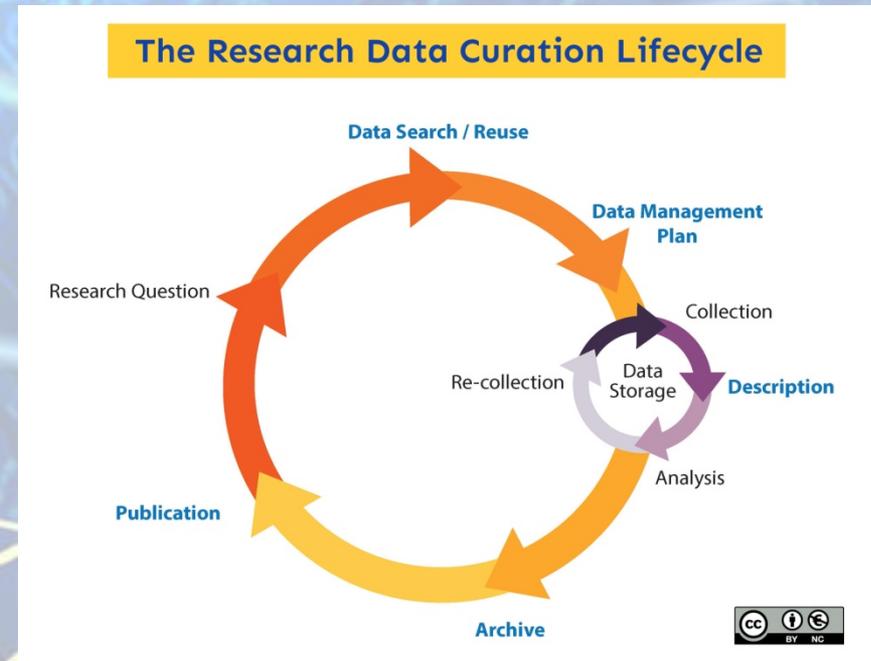
Data Librarian

Data Curation

Data Literacy

# Data Curation

Secondo l'Università dell'Illinois *Graduate School of Library and Information Science*, "data curation è la gestione dei dati al fine di scoprire, recuperare, mantenere, aggiungere valore, permettere il riutilizzo dei dati, garantendone contemporaneamente l'autenticazione, l'archiviazione, la gestione e la preservazione"



# Attività di DATA CURATION

- **Selezione**
- **Gestione**
- **Integrità**
- **Sicurezza**
- **Descrizione/Metadattazione**
- **Formati**
- **Archiviazione/storage**
- **Accesso ai dati**
- **Condivisione**
- **Riuso**
- **Ridistribuzione**
- **Conservazione a medio / lungo termine**

# Data Literacy

i dati sono inutili senza le competenze

- **la capacità di leggere, comprendere, creare e comunicare i dati come informazioni.**
- **è una vera e propria alfabetizzazione rispetto ai dati che si concentra sulle competenze coinvolte nel lavorare con essi.**
- **Poiché la raccolta dei dati e la condivisione dei dati sono routine e l'analisi dei big data è ormai sempre più frequente, diventa importante per ricercatori, studenti, cittadini essere alfabetizzati.**
- **Un servizio importante tra quelli svolti dalle biblioteche accademiche nella gestione dei dati sono i corsi di Data Literacy**
  - Finalizzati alla formazione dei ricercatori nella gestione dei dati frutto delle loro attività di ricerca
- **Le competenze associate ai corsi di Data Literacy sono:**
  - Integrazione, cura e interoperabilità dei dati creati durante il ciclo della ricerca



# Epilogo, Biblioteche e Big Data

una sfida tecnologicamente e  
concettualmente *complessa*

# Sfida per i sistemi bibliotecari

- **L'ingresso dei Big Data in Biblioteca potrà suscitare nuovi importanti significati con ricadute non soltanto tecnologiche ma anche organizzative**
- **Si profila una sfida che potrebbe determinare un vero e proprio rivolgimento degli attuali assetti bibliotecari**
- **Si può immaginare che l'impatto dei Big Data possa perfino spostare il tradizionale baricentro delle biblioteche dalle collezioni ai dati**
- **Con un ritorno – nell'organizzazione dei sistemi bibliotecari – a forme più centralizzate (più attrezzate nella gestione della complessità dei Big Data) con il conseguente declino dei modelli distribuiti basati su biblioteche di medio – piccole dimensioni.**

# Sfida per nuovi servizi

biblioteche: piattaforme per la gestione dei Big Data

- **Lo spostamento del «focus» dalle collezioni ai “Big Data”, non significa solo nuovi costi infrastrutturali**
- **Significa anche progettare e supportare nuovi servizi per la gestione e accesso ai dati:**
  - **Archiviazione**
  - **Ricerca**
  - **Condivisione**
  - **Trasferimento**
  - **Visualizzazione**
  - **Analisi**

# Sfida: dal documento ai dati, la destrutturazione del record bibliografico nei Big Data del web

- **Ci sono milioni e milioni di (Big data) bibliografici conservati negli OPAC delle biblioteche di tutto il mondo, ma non raggiungibili attraverso la Rete perché creati e registrati con formati non interoperabili nel Web**
- **le tecnologie del web però corrono verso il web semantico. Le nuove tecnologie si chiamano: Linked Data e RDF (Resource Description Framework). È la ricerca della massima granularità: ridurre tutta l'informazione del web in dati interoperabili e soprattutto processabili in maniera automatica dai computer connessi nella rete globale**
- **Con la tecnologia dei Linked Data il record bibliografico è destinato ad aprirsi al web e trasformarsi dinamicamente in una ri-composizione di dati provenienti non più solo dal dominio delle biblioteche ma dalle più disparate fonti del web**

# Grazie per l'attenzione

**Fabio Di Giammarco**

[fabio.digiammarco@beniculturali.it](mailto:fabio.digiammarco@beniculturali.it)

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