

---

---

# **La Biblioteca Nazionale Centrale di Firenze quando la cultura è aperta e accessibile**

— Giovanni Bergamin, BNCF —  
(20 giugno 2017)

---

---

# Parlerò di:

- qualche nota critica sui progetti di digitalizzazione in BNCF
- qualche domanda aperta:
  - digitalizzazione di massa o progetti “artigianali”?
  - obiettivi: testo ricercabile o “immagine fedele”?
- qualche direzione:
  - valorizzare di più quello che abbiamo già

# Progetti di digitalizzazione in BNCF - 1

- Dal 1992 (Sperimentazione SDIEF) a ARSBNI (fine anni 90)
- L'obiettivo è arricchire il record bibliografico attraverso la digitalizzazione dei frontespizi, indici ecc (bianco e nero e OCR) -> aumentare le possibilità di accesso ai documenti
- Vengono - in generale - presi in considerazione i libri correnti (quelli descritti della BNI)

# Progetti di digitalizzazione in BNCF - 2

Nel corso degli anni molti progetti hanno riguardato principalmente:

- manoscritti
- libri antichi
- carte geografiche
- fotografie

con l'obiettivo dominante dell'immagine "fedele"

## Progetti di digitalizzazione in BNCF - 3

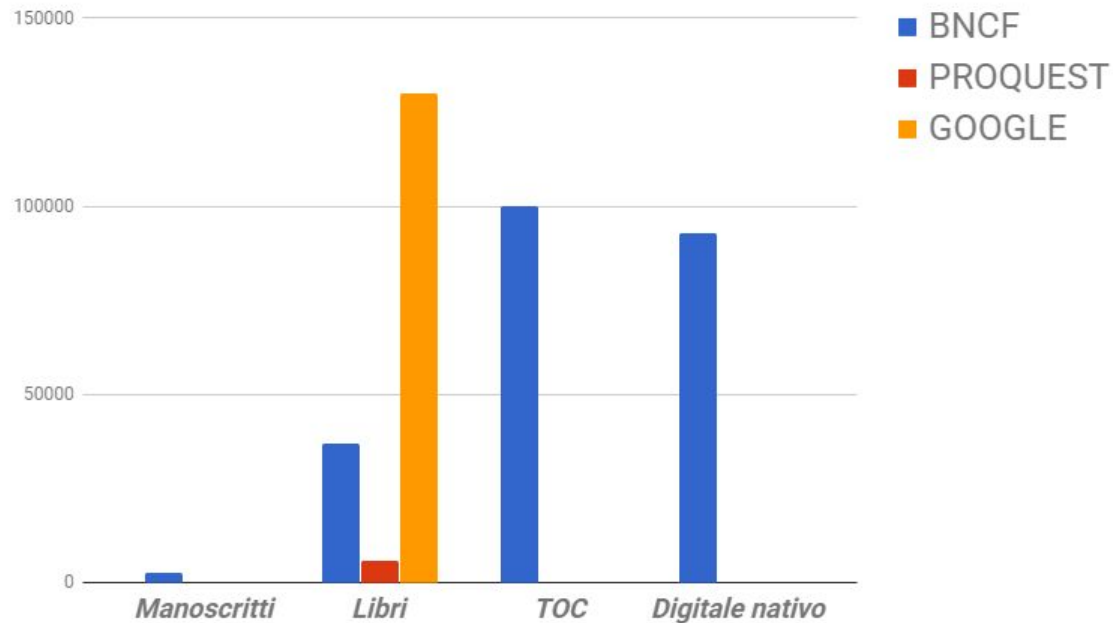
Due progetti di digitalizzazione “di massa”

- Google Books (dal 2012 - in corso)
- Proquest EEB - Early European Books (dal 2010 - in corso)

Una sperimentazione di digitalizzazione “wiki” o collaborativa per quanto riguarda l’OCR (accordo BNCF, Wikimedia Italia)

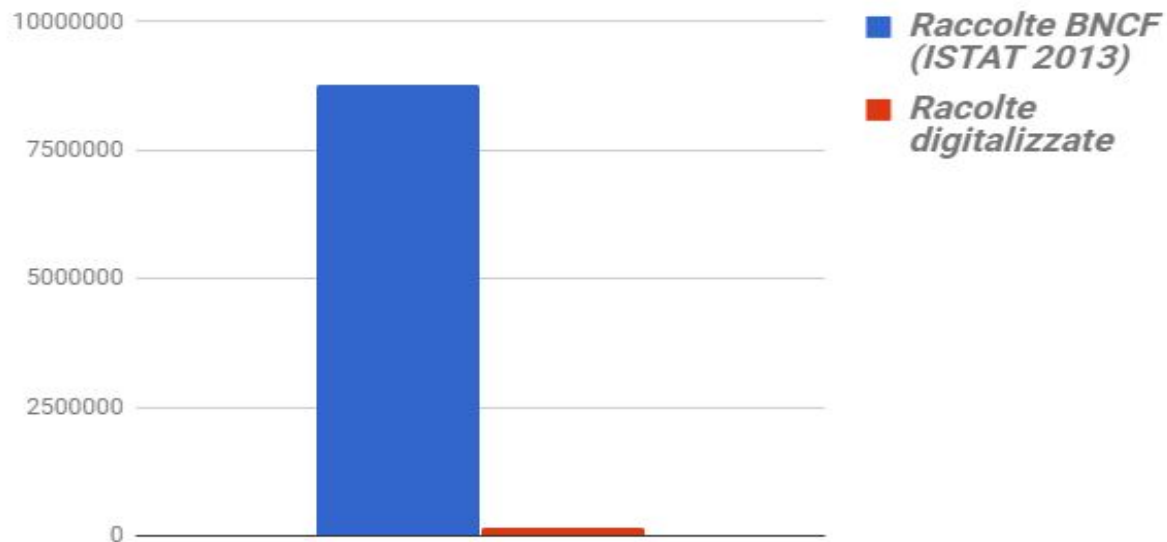
# Ordini di grandezza - 1

Digitale in BNCF e Raccolte - statistiche a fine 2016 (misurato in unità bibliografiche)



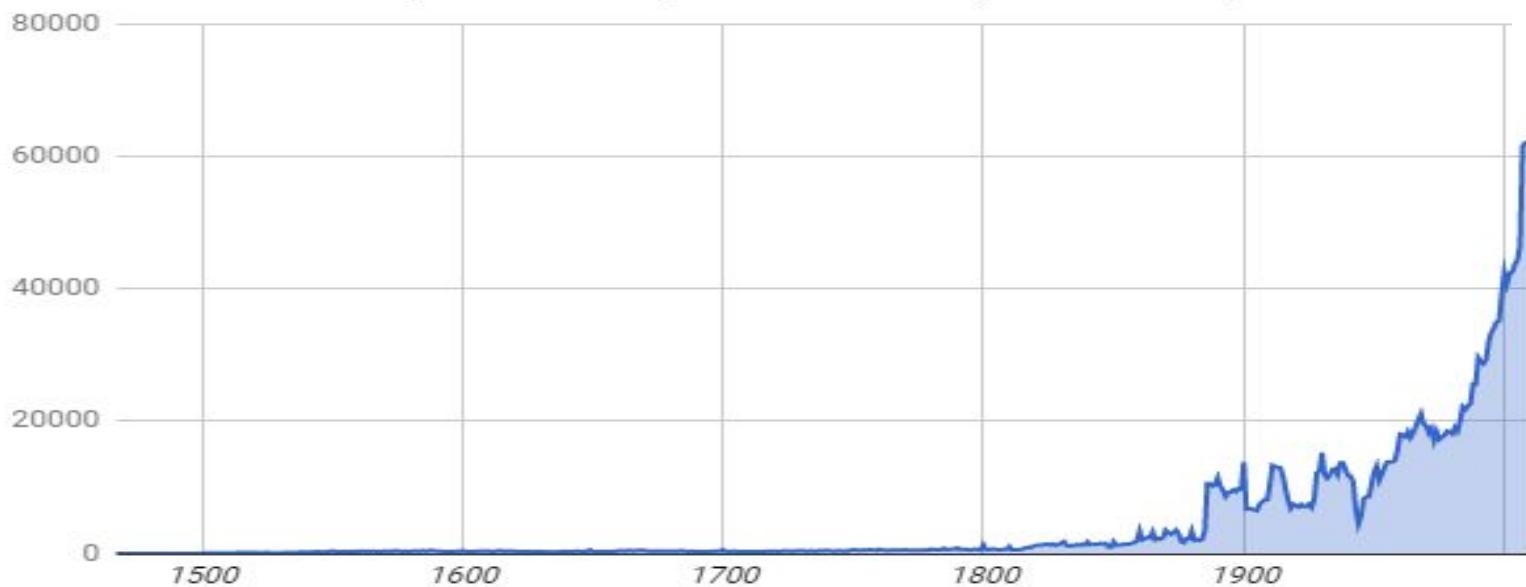
# Ordini di grandezza - 2

## Raccolte BNCF (ISTAT 2013) e Digitalizzate



# Ordini di grandezza - 3

Libri: distribuzione per anno di pubblicazione (fino al 2010)





# Che cos'è la digitalizzazione di massa?

- digitalizzazione su scala industriale (non solo su larga scala);
- digitalizzazione di intere biblioteche/raccolte (senza selezionare i singoli documenti)

*fonte: Karen Coyle*

# Google Books: pagine del web e pagine dei libri

- Inizio 2004
- Fine prevista 2020
- The Google Books aim is the Google aim:
  - “organize the world's information and make it universally accessible and usable” --- so the **content of all published books** has to be searchable together with the **content of all web** pages
- Totale stimato (da Google) di libri al mondo: 129.864.880
- Totale libri digitalizzati a oggi: ~30 M (non sono disponibili statistiche ufficiali)

# Confronto Google Books - Proquest - 1

	<i>Google Books</i>	<i>Proquest EEB</i>
Copertura	fino al 1877	fino al 1700
Obiettivo principale	ebook con testo "liquido" e ricercabile	immagine "fedele"

## Confronto Google Books - Proquest - 2

	<i>Google Books</i>	<i>Proquest EEB</i>
Tipo progetto	nazionale	BNCF
Costi a carico di	Google eccetto selezione e mov. in BNCF	tutto a carico di Proquest
Luogo di trattamento dei libri	non in BNCF (Centro Google in Italia)	in BNCF

## Confronto Google Books - Proquest - 3

	<i>Google Books</i>	<i>Proquest EEB</i>
Risultati	copia per BNCF disponibilità su Google Books in pubblico dominio	copia per BNCF e disponibilità senza costi per tutti gli IP italiani (fino al recupero dei "costi", poi pubblico dominio)

# Google Books opportunità >> e criticità << (punti discussi nella letteratura professionale)

- >> la copertura
- << l'alta percentuale dei libri in inglese (vicina al 50%)
- >><< è un necessario patto con il diavolo (cit.)
- >> le tecnologie usate non hanno concorrenti (es. OCR: solo il wiki collaborativo può migliorare i risultati)
- << controllo di qualità eseguito dalle macchine
- << esclusione per determinate dimensioni

# Google Books - domande aperte

- Potevano (possono) esserci alternative?
- Le biblioteche potevano (possono) chiedere di più?

# Uno scambio di tweet recente



massimo mantellini ✓

@mante

Following



Sento le imprecazioni di mia moglie in  
lontananza. **Archivio di Stato di Firenze**  
digitalizza dentro interfacce in flash inusabili

In risposta a @mante

HTML5! Date loro la notizia: esiste l'HTML5!

---



# La fruizione delle immagini richiede un adeguato aggiornamento tecnologico

- Oltre alla standardizzazione dei formati e dei metadati (MAG, METS) diffusamente applicata
- Occorre tener presente che le applicazioni *passano* e i dati (immagini e metadati) *restano* (cit.)

# I servizi attuali che pubblicano le digitalizzazioni hanno molti limiti ...

- Spesso gestiscono più copie della stessa immagine
  - alta risoluzione (digital preservation)
  - risoluzione per il web
  - miniatura (thumbnail)
  - versioni particolari per servizi particolari (es. zoom)
  - ...
- Le immagini sono accessibili solo attraverso un software e ogni servizio (biblioteca digitale) ha il suo; le applicazioni *imprigionano* le immagini

## L'utente invece vorrebbe ...

- vedere ad alta risoluzione anche i dettagli di una immagine (*zoom*)
- vedere le immagini da personal computer, dal tablet, dallo smartphone (*adattabilità*)



## L'utente invece vorrebbe ... - 2

- *mettere a confronto* immagini che provengono da servizi differenti
- *citare* con un indirizzo URL una immagine o un dettaglio di una immagine
- *manipolare, annotare ecc ...*



# Le biblioteche digitali vorrebbero ...

- Mantenere il controllo delle proprie immagini
- Semplificare la gestione degli oggetti digitali





## Le biblioteche digitali vorrebbero ... - 2

- Non legare il proprio servizio solo a un determinato software: i dati sono più importanti delle applicazioni



# iiif - International Image Interoperability Framework

Una comunità (Biblioteche digitali e sviluppatori di software)  
che:

- *definisce e sviluppa API*  
(Application Programming Interface = le regole con le quali differenti applicazioni possono scambiarsi dati)



## iiif - 2

- sviluppa *software basati su queste API*
- *mette a disposizione contenuti interoperabili* (usabili e riusabili in differenti contesti e non legati a un determinato software)

Entro il 2017 la TECA BNCF sarà accessibile anche con funzionalità IIIF





**Grazie**