

# Fare Open Science nelle scienze umane

Bianca Gualandi

[bianca.gualandi4@unibo.it](mailto:bianca.gualandi4@unibo.it) - <https://orcid.org/0000-0001-8202-8493>

Area Servizi alla Ricerca  
Università di Bologna

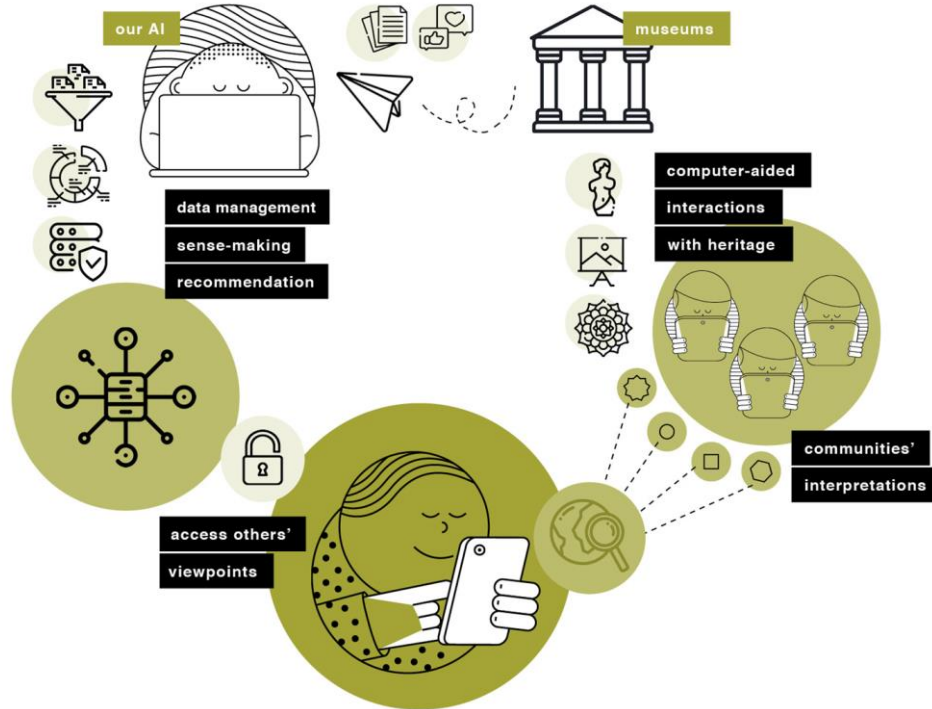
Silvio Peroni

[silvio.peroni@unibo.it](mailto:silvio.peroni@unibo.it) - <https://orcid.org/0000-0003-0530-4305>

Digital Humanities Advanced Research Centre (/DH.arc)  
Dipartimento di Filologia Classica e Italianistica  
Università di Bologna



# Un esempio concreto: il progetto SPICE



<https://spice-h2020.eu/>

- SPICE è un progetto che si concentra sullo sviluppo di strumenti e metodi per supportare la Citizen Curation
- Citizen Curation: permettere alle persone di **utilizzare tecnologie specifiche** (per esempio, il loro smartphone o il loro computer) per selezionare dipinti, sculture, e altri oggetti museali, **al fine di condividere le loro interpretazioni e riflessioni su questi oggetti** con gli altri
- Cinque casi di studio (musei in Finlandia, Irlanda, Israele, Italia, e Spagna)

# Quali “dati”?

- Data Management Plan: <https://doi.org/10.5281/zenodo.7186477>
- Tipi di dato gestiti:
  - Dataset: risposte di questionari, trascrizione di interviste, dati strutturati (ad esempio tabellari)
  - Software
- Obiettivo dell’infrastruttura tecnica (Linked Data Hub): permettere agli utenti di interagire con gli oggetti in un museo, che di fatto diventano intermediari e abilitano l’interazione con altre persone interessate allo stesso oggetto culturale o ad oggetti culturali simili
- Tecnologie, strumenti, e linee guida utilizzate per lo sviluppo:
  - FAIR
  - TRUST
  - Linked Data
  - SOLID



# Riuso dei dati generati dall'infrastruttura

- Promoting the widest possible access to digitised public domain material as well as the widest possible reuse of the material for non-commercial and commercial purposes  
*Commission Recommendation of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation, no. 2011/711/EU, Official Journal of the European Union. <http://data.europa.eu/eli/reco/2011/711/oj>*
- Alcuni problemi relativi al materiale digitalizzato e alle annotazioni degli utenti:
  - Il materiale digitalizzato potrebbe essere **soggetto a copyright** che potrebbe non permettere al museo che lo conserva di condividerlo liberamente usando il Linked Data Hub
  - Alcuni musei **percepiscono dei ricavi** dalla vendita di questo materiale digitalizzato per pubblicazioni commerciali, e potrebbero percepire questa condivisione libera sull'infrastruttura come una **minaccia alla loro sostenibilità economica**
  - Anche quando il materiale digitalizzato è nel pubblico dominio, l'oggetto *ibrido* creato sovrapponendo questo materiale con il nuovo contenuto condiviso dagli utenti attraverso le piattaforme digitali sviluppate all'interno del progetto **preconfigura situazioni non scontate di Data Rights Management (DRM)**, in particolare considerando che i dati relativi agli oggetti culturali e i dati (ad esempio, le annotazioni) degli utenti vengono di fatto prodotti e condivisi con dei processi diversi che, spesso, sono governati dall'istituto culturale coinvolto



# Scienza Aperta nelle Scienze Umane

- Grazie all'intervento deciso dell'Europa (per esempio, in Horizon Europe) nel considerare l'Open Science come un pilastro portante della ricerca e non soltanto come una sua appendice, l'Università di Bologna ha iniziato un piano di "rinnovamento" che tocca sia la didattica (un esempio è la laurea magistrale in Digital Humanities and Digital Knowledge, fortemente orientata alla gestione del "dato" nelle Scienze Umane) e il modo in cui si fa ricerca nei vari dipartimenti, sia STEM sia SSH
- Contrariamente a contesti STEM, dove è stato fatto sin dalla messa a punto dei principi FAIR un grande lavoro di identificazione della specificità dei vari tipi di dato, nel contesto delle Scienze Umane si è sempre percepita una diffidenza sul tema e, pragmaticamente, una difficoltà nel declinare una definizione di "dato" che fosse rappresentativa per l'area
- Idea di ricerca: coinvolgere un campione rappresentativo di alcune aree Umanistiche (in questo caso, per opportunità, il Dipartimento di Filologia Classica e Italianistica) per investigare come il concetto di "dato" viene declinato in progetti esistenti relativi alle Scienze Umane



# Il concetto di dato nelle scienze umane

Domande di ricerca:

1. **Per implementare i principi FAIR nelle scienze umane, come definiamo il dato della ricerca?**
2. **Cosa pensano i nostri colleghi del termine dato, anche rispetto ad altri termini come “materiali” della ricerca?**
3. **Come gestiscono attualmente i dati della ricerca?**

Metodologia:

- Lunghe interviste di persona o in video (30-90 minuti)
- “Parlami del tuo progetto”
- “Quali **materiali** hai raccolto o creato ex-novo nel corso della tua ricerca?”
- Solo alla fine → “Considereresti questi materiali come **dati della ricerca**?”



Metodologia riutilizzata da Thoegersen, J. L. (2018), “‘Yeah, I guess that’s data’: data practices and conceptions among humanities faculty”, *Libraries and the Academy*, Vol. 18 No. 3, pp. 491-504.  
<https://doi.org/10.1353/pla.2018.0030>



# Discipline considerate

Intervistati **19 ricercatori**, uno per ogni area scientifico-disciplinare rappresentata all'interno del Dipartimento di Filologia Classica e Italianistica @UniBO.

Nell'analisi, abbiamo poi semplificato e raggruppato i ricercatori in 5 gruppi:

<i>Filologia e critica letteraria</i> – studiano innanzitutto testi o documenti	12
<i>Linguistica e lingue</i> – si occupano di linguaggio, fenomeni linguistici e insegnamento delle lingue	4
<i>Storia dell'arte</i>	1
<i>Informatica</i>	1
<i>Archivistica</i>	1

# Analisi

Abbiamo condotto un'**analisi qualitativa** dei dati:

1. Registrazione audio  
delle interviste



Previo consenso informato

2. Trascrizione e  
anonimizzazione (manuale)



Gualandi, B., Pareschi, L., and Peroni, S. (2022), What do we mean by “data” in the arts and humanities? Interview transcripts (University of Bologna, FICLIT) and qualitative data coding, Zenodo, <https://doi.org/10.5281/zenodo.6123290>

3. Utilizzo di un approccio  
Grounded Theory  
**usando software OS**



Curtain, C. (2022), QualCoder version 2.8, Software Heritage, available at:  
<https://archive.softwareheritage.org/swh:1:dir:e24c62cf6c1fd583406f9f38a9de2d102249531d>





# RQ1: I dati della ricerca nelle scienze umane

*Domanda:* “Quali materiali hai raccolto o creato ex-novo nel corso della tua ricerca?” → abbiamo categorizzato le risposte in **13 tipologie di dato**:

<b>Tipi di dato</b>	<b>Creati da (tot. 19)</b>	<b>Utilizzati da (tot. 19)</b>
<i>Pubblicazioni</i>	18	15
<i>Altre fonti primarie</i> (e.g., manoscritti, opere d'arte)	0	18
<i>Rappresentazioni digitali di fonti primarie</i> (e.g., facsimile, fotografie)	4	8
<i>Cataloghi, basi di dati e altri strumenti di ricerca</i> (perlopiù bibliografica)	2	9
<i>Eventi</i> (e.g., conferenze, esibizioni)	6	0
<i>Siti web statici</i> (e.g., sito di progetto)	4	0
<i>Software</i>	2	2
<i>Documentazione</i>	3	0
<i>Infrastrutture digitali</i> (e.g., piattaforme online, app)	3	0
<i>Dati personali</i>	2	0
<i>Corpora linguistici</i>	2	0
<i>Standard</i> (e.g., per interoperabilità)	0	2
<i>“Born-digital artifacts”</i> (e.g., tag, associazioni, testi prodotti dagli utilizzatori di infrastrutture digitali – vd sopra)	1	1



# RQ2: Siamo tutti d'accordo che si tratti di "dati"?

*Domanda:* "Considereresti i materiali che hai elencato come dati?"



Risposta	Intervistati (19)
<i>Sì</i>	14
<i>No</i>	2
<i>Non so</i>	3



**Riserve sul termine**  
espresse da 3 ricercatori  
(che però hanno risposto sì)

*Domanda:* "Come definiresti un dato della ricerca?"



Molta **indecisione**, abbiamo pesato le varie affermazioni e attribuito un voto da 1 a 5 a ciascun intervista



Voto	Definizione di dato	Intervistati (19)
5	"Tutto è un dato"	2
4	Definizione molto ampia, che include fonti primarie, interpretazioni, metodologie e risultati della ricerca	4
3	Definizione ampia, che include almeno una tra le seguenti: fonti primarie, interpretazioni, metodologie e risultati	5
2	Definizione restrittiva, che però può includere fonti primarie, testi critici e/o varianti testuali	2
1	Definizione molto restrittiva: "informazione grezza", "elemento quantitativo"	3
n/a	Nessuna risposta, oppure impossibile assegnare un voto	3



# RQ3: La gestione dei dati di ricerca oggi

La stragrande maggioranza degli intervistati **conserva tutti i dati**, per utilizzarli nell'insegnamento o in progetti futuri

La **pubblicazione aperta dei dati** (incluse le pubblicazioni) **non avviene quasi mai**

Facsimili e fotografie delle fonti primarie, specie letterarie, sono molto utilizzate durante l'analisi ma quasi mai pubblicate

Documentazione e standard condivisi sono usati pochissimo

! La **collaborazione e il lavoro di squadra** sono considerati fondamentali – sebbene una piccola minoranza parli della necessità di una certa segretezza, visti i lunghi tempi di pubblicazione e i sistemi attuali di valutazione

! Emerge comunque la **estrema importanza delle metodologie** → ma raramente formalizzate, se non all'interno di pubblicazioni



# Conclusioni

Tema di un articolo appena accettato dal Journal of Documentation. Pre-print disponibile qui: Gualandi B., Pareschi L., Peroni S. (2022), "What do we mean by 'data'? A proposed classification of data types in the arts and humanities". arXiv.org. <https://doi.org/10.48550/arXiv.2205.06764>

Le **pubblicazioni** sono un elemento imprescindibile della ricerca nelle scienze umane → non solo derivate dai dati, ma **dati esse stesse** (teniamo a mente che 12/19 intervistati sono filologi!)

Gli intervistati **non hanno trovato il termine "dato" particolarmente problematico** → solamente 3 su 19 hanno espresso dubbi o resistenza

Nonostante la carenza di "workflow" formalizzati e documentati, dalle 19 interviste è emersa la **centralità delle metodologie** → sembra esserci ampio spazio per una discussione in questo senso

Lo studio **va replicato su più larga scala**: al momento dà una visione molto di parte delle scienze umane (cfr. sezione metodologia)

Si è dimostrato comunque uno strumento utile per **coinvolgere i ricercatori** sul tema dei dati e della loro gestione → quasi tutti gli intervistati hanno accolto la conversazione con grande interesse

