

Archiviazione dei dati della ricerca nelle scienze sociali: quali strategie per incentivare le pratiche?

Filippo Accordino, Daniela Luzi, Fabrizio Pecoraro

IRPPS-CNR

Data Archive for Social Sciences in Italy (DASSI)

f.accordino@irpps.cnr.it

KEYWORDS: data archives, data sharing, data curation, open research data

1. Introduzione /contesto

Il Dipartimento di Scienze Umane e Sociali, Patrimonio Culturale (DSU) e l'Istituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS) del Consiglio Nazionale delle Ricerche, sono coinvolti insieme all'Università degli Studi di Milano-Bicocca nella realizzazione del Data Archive for Social Sciences in Italy (DASSI).

DASSI, service provider nazionale dell'infrastruttura europea CESSDA-ERIC, è pronto a offrire nuove opportunità alla ricerca nell'ambito delle scienze sociali, facilitando anche la possibilità di condurre studi comparativi tra Paesi.

La condivisione dei dati della ricerca, pratica collocata nell'alveo dell'open science, non costituisce una semplice pubblicazione di oggetti addizionali a corredo di paper scientifici, ma rappresenta un vero cambio di paradigma capace di ridefinire metodi e approcci nella ricerca. Questo richiede un profondo mutamento nelle pratiche della ricerca, nelle competenze necessarie e nella consapevolezza circa le opportunità da cogliere. Gli archivi di dati rappresentano la migliore soluzione in grado di mediare tra produttori dei dati e chi li riutilizza (Borgman et al., 2019), sostenendo l'innovazione nella ricerca e l'impatto sulle politiche pubbliche.

La pubblicazione dei dati della ricerca raccolti o elaborati è una condizione essenziale per assicurare trasparenza e credibilità nel lavoro dei ricercatori. Reperire in un archivio certificato i dati e le informazioni che ne garantiscono la comprensibilità, permette la riproducibilità della ricerca e di incrementarne la fiducia da parte della comunità scientifica (Borgman, 2012). La pubblicazione dei dati ne favorisce il riuso: la possibilità che ricercatori diversi da chi li ha raccolti possano utilizzarli per rispondere a nuove domande di ricerca, aumentandone il valore e l'impatto (Bishop & Kuula-Luumi, 2017; Kenfield et al., 2022; Pasquetto et al., 2017).

Il mutamento delle pratiche di ricerca verso un diffuso ricorso al deposito dei dati richiede l'attuazione di una serie di azioni da parte dei numerosi attori coinvolti nel processo di creazione e condivisione: autori dei dati, infrastrutture di ricerca e archivi, riviste scientifiche e finanziatori.

Gli archivi di dati, oltre a fornire servizi per il deposito e il reperimento dei dati, garantendone anche la conservazione a lungo termine, sono impegnati anche in attività di comunicazione, sensibilizzazione e formazione. La consapevolezza del potenziale derivante dalla condivisione dei dati deve essere accompagnata dalla diffusione di specifiche competenze nella gestione e descrizione dei dati della ricerca, indispensabili a facilitare il percorso di deposito.

Il contributo intende far riflettere sulle sfide necessarie a un nuovo approccio alla gestione dei dati della ricerca, raccogliendo spunti circa le strategie possibili per sensibilizzare sempre di più tutta la comunità scientifica e per incentivare il deposito dei dati nell'archivio DASSI.

2. Problema di ricerca con indicazione di domande di ricerca /spunti di riflessione / domande stimolo;

Le competenze necessarie a una gestione dei dati della ricerca compatibile con le pratiche di condivisione, deposito negli archivi e riuso sono specifiche e differenti rispetto a quelle dell'analisi o raccolta dei dati (Fischer et al., 2022; White et al., 2013). Inoltre, si tratta di un'attività time-consuming (Perry & Netscher, 2022) che il ricercatore deve essere disposto a sostenere.

Le conoscenze da acquisire comportano, da un lato, la necessità di aggiornare i programmi formativi sin dai corsi universitari, in modo da fornire le conoscenze minime su concetti e strumenti (data management plan, metadati, standard, dizionari controllati, cataloghi). Dall'altro, nuove professionalità specifiche (Tammaro et al., 2019), dovrebbero essere sempre più presenti all'interno delle università e degli enti di ricerca.

La letteratura scientifica internazionale ha affrontato varie volte il tema degli incentivi, motivazioni e percezioni dei ricercatori nell'ambito della condivisione dei dati (Darch & Knox, 2017; Tenopir et al., 2015; Zenk-Möltgen et al., 2018). La sfida per ogni archivio di dati è quella di incrementare il proprio patrimonio al fine di offrire alla comunità designata una vasta e attrattiva disponibilità di dati riutilizzabili. Gli archivi di dati sono costantemente impegnati nella diffusione di buone pratiche che incentivino la possibilità di gestione dei dati della ricerca basati sul paradigma FAIR (Wilkinson et al., 2016).

Il contributo intende far riflettere sulle tematiche proposte, focalizzando l'attenzione sulle pratiche. Quali le strategie migliori per far comprendere in modo rapido e diffuso alla comunità scientifica il potenziale della condivisione e del riuso? In che modo incentivare le pratiche di deposito dei dati negli archivi? Le specifiche competenze per la documentazione del dato e dell'intero lifecycle come possono trovare ampia diffusione, a partire dai corsi universitari e di dottorato, per estendersi a tutta la comunità scientifica?

3. Approccio proposto

Le esperienze di ricerca condotte dal gruppo DASSI-CESSDA, la realizzazione del deliverable sulle strategie di outreach da parte del Working Group "Policy and Documentation", nonché la programmazione delle attività presenti e future dell'archivio italiano, sono la base per la formulazione di alcune raccomandazioni che coinvolgono i diversi destinatari coinvolti:

- Aumentare la consapevolezza sull'importanza della condivisione dei dati, e sulle opportunità offerte dal riuso e dagli archivi, all'interno della comunità scientifica tutta;
- Incoraggiare le buone pratiche di gestione dei dati della ricerca in tutte le fasi del data lifecycle (Tenopir et al., 2011), in modo da predisporre i dati alla pubblicazione attraverso un archivio;
- Ribadire il riconoscimento del dato quale prodotto della ricerca a tutti gli effetti (Data Citation Synthesis Group, 2014), incentivando anche le migliori pratiche di citazione (Ball & Duke, 2015), da eseguire in maniera accurata, al fine di attribuire credito agli autori (Tenopir et al., 2015);
- Richiedere la pubblicazione dei dati delle ricerche finanziate e nelle pubblicazioni scientifiche.

In particolare, DASSI si impegna a:

- Stabilire relazioni con grandi produttori di dati di particolare interesse e con alto potenziale di riuso (organizzazioni ed enti di ricerca o università);
- Monitorare e restare in contatto con i progetti di ricerca in cui i dati sono prodotti;
- Collaborare con infrastrutture di ricerca e repositories sviluppati da università e istituzioni, promuovendo il collegamento tra pubblicazioni e dati, e l'interoperabilità tra cataloghi;
- Promuovere le proprie attività e le nuove acquisizioni, attraverso attività specifiche di comunicazione.

4. Discussione e conclusioni

Il lavoro portato avanti dagli archivi di dati va oltre la semplice fornitura di un servizio di deposito e consultazione, accogliendo istanze scientifiche di diffusione delle buone pratiche. Pertanto, riveste una notevole importanza stabilire le strategie più adatte anche rispetto al contesto italiano, affinché la condivisione dei dati della ricerca possa divenire una pratica generalmente diffusa e in grado di produrre un importante impatto.

Rerences

Ball, A., & Duke, M. (2015). *'How to Cite Datasets and Link to Publications'*. DCC How-to

Guides. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides>

Bishop, L., & Kuula-Luumi, A. (2017). Revisiting Qualitative Data Reuse: A Decade On. *SAGE*

Open, 7(1), 215824401668513. <https://doi.org/10.1177/2158244016685136>

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078.

<https://doi.org/10.1002/asi.22634>

- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70(8), 888–904. <https://doi.org/10.1002/asi.24172>
- Darch, P. T., & Knox, E. J. M. (2017). Ethical perspectives on data and software sharing in the sciences: A research agenda. *Library & Information Science Research*, 39(4), 295–302. <https://doi.org/10.1016/j.lisr.2017.11.008>
- Data Citation Synthesis Group. (2014). *Joint Declaration of Data Citation Principles*. Force11. <https://doi.org/10.25490/A97F-EGYK>
- Fischer, C., Hirsbrunner, S. D., & Teckentrup, V. (2022). Producing Open Data. *Research Ideas and Outcomes*, 8, e86384. <https://doi.org/10.3897/rio.8.e86384>
- Kenfield, A. S., Woolcott, L., Thompson, S., Kelly, E. J., Shiri, A., Muglia, C., Masood, K., Chapman, J., Jefferson, D., & Morales, M. E. (2022). Toward a definition of digital object reuse. *Digital Library Perspectives*, 38(3), 378–394. <https://doi.org/10.1108/DLP-06-2021-0044>
- Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. *Data Science Journal*, 16, 8. <https://doi.org/10.5334/dsj-2017-008>
- Perry, A., & Netscher, S. (2022). Measuring the time spent on data curation. *Journal of Documentation*, 78(7), 282–304. <https://doi.org/10.1108/JD-08-2021-0167>
- Tammaro, A. M., Matusiak, K. K., Sposito, F. A., & Casarosa, V. (2019). Data Curator's Roles and Responsibilities: An International Perspective. *Libri*, 69(2), 89–104. <https://doi.org/10.1515/libri-2018-0090>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, 6(6), e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among

Scientists Worldwide. *PLOS ONE*, 10(8), e0134826.

<https://doi.org/10.1371/journal.pone.0134826>

White, E., Baldrige, E., Brym, Z., Locey, K., McGlinn, D., & Supp, S. (2013). Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*, 6(2).

<https://doi.org/10.4033/iee.2013.6b.6.f>

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship.

Scientific Data, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science.

Journal of Documentation, 74(5), 1053–1073. <https://doi.org/10.1108/JD-09-2017-0126>