

“Data steward” per i dati FAIR

Valentina Pasquale¹, Emma Lazzeri², Elena Giglia³

¹Istituto Italiano di Tecnologia, ²Consortium GARR, ³Università degli Studi di Torino

Abstract. I dati sono il fondamento di ogni ricerca scientifica e pertanto devono essere conservati e condivisi, quando possibile, per una scienza solida e riproducibile. Enti di ricerca in tutto il mondo si stanno dotando di centri di competenza in open science e data management per supportare i ricercatori nel rendere FAIR i dati della ricerca, attraverso l'introduzione dei data steward. Lo scopo di questo lavoro è inquadrare lo stato dell'arte della definizione della figura professionale del data steward e delle sue competenze, con particolare riferimento ai recenti e futuri sviluppi in Italia

Keywords. open science, FAIR data, research data management, data steward, EOSC

Introduzione: perché il data steward

I dati sono il fondamento di ogni ricerca scientifica. Pertanto devono essere conservati e, quando possibile, condivisi per una scienza solida e riproducibile. Inoltre, “l'utilizzo intelligente dei dati, ivi compreso il loro trattamento attraverso applicazioni di intelligenza artificiale, può trasformare tutti i settori dell'economia” (Direttiva EU 1024/2019). L'intelligenza artificiale ha però bisogno di dati “intelligenti”, che possono essere (ri)utilizzati solo se accompagnati da metadati e rappresentati per renderli comprensibili e sfruttabili dalle macchine.

A questa esigenza rispondono i principi FAIR (Findable, Accessible, Interoperable, Reusable o anche Fully Artificial Intelligence Ready) (Wilkinson et al., 2016). Enti di ricerca accademica e privata in tutto il mondo si stanno dotando di centri di competenza in data stewardship per i dati FAIR.

Nella visione della Commissione Europea e di Horizon Europe, i dati sono al centro della ricerca e dell'innovazione: per affrontare le grandi sfide del futuro (cambiamento climatico, economia sostenibile, etc.) occorre che siano disponibili in modo FAIR e aperto a ricercatori ed esperti. Per questo la Commissione ha investito in EOSC, European Open Science Cloud, un “web di dati e servizi interconnessi” i cui blocchi costitutivi saranno i dati FAIR. EOSC sarà uno “spazio di condivisione” in cui chi produce e chi ri-usa dati potrà collaborare, creando nuove idee e soluzioni.

È stato stimato che nei prossimi anni per realizzare e operare EOSC saranno necessari 500.000 data steward. In questo contesto, è necessario destinare il 5% dei fondi di ricerca a una corretta gestione dei dati, inclusa l'assunzione dei data steward (Mons, 2020). In questa prospettiva, istituti di ricerca e istruzione universitaria devono impegnarsi a formare questa nuova figura professionale in modo strutturato e ad integrarla nel supporto alla ricerca. Scopo di questo lavoro è inquadrare lo stato dell'arte della definizione della figura professionale del

data steward e delle sue competenze, con particolare riferimento ai recenti sviluppi in Italia.

1. Le competenze del data steward

Il data steward è una nuova figura professionale competente nella pianificazione e gestione responsabile dei dati della ricerca durante tutto il loro ciclo di vita. Diversamente dal data scientist o dal research software engineer, il ruolo del data steward è ancora poco conosciuto e caratterizzato. Stewardship può essere tradotto con “amministrazione, gestione”, ma perde la sua connotazione di “pianificazione e gestione responsabile ed etica delle risorse” (Wikipedia).

La professionalizzazione dei data steward e la creazione di centri di competenze sono individuati nella EOSC Strategic Research and Innovation Agenda quali fattori critici di successo per EOSC. Alcuni progetti quali FAIRsFAIR e EOSC Pillar, nonché EOSC Executive Board Skills and Training Working Group chiuso nel 2020, e la Research Data Alliance (RDA) stanno lavorando alla definizione di framework di competenze del data steward. RDA Interest Group for Professionalizing Data Stewardship ha diversi task group su integrazione della data stewardship nelle istituzioni, formazione e profili di carriera, definizione e certificazione delle competenze e sostenibilità finanziaria.

L’Olanda è la nazione che per prima ha inserito i data steward nelle istituzioni, a cominciare da TU Delft nel 2017 (Teperek and Plomp, 2019). Sempre in Olanda, nel National Programme Open Science è stato finanziato un progetto per la definizione delle competenze e della formazione dei data steward (Jetten et al., 2021) che ha prodotto una classificazione dei vari profili, ovvero Research, Policy e Infrastructure (Scholtens et al., 2019). Per ogni profilo sono stati definiti funzioni, competenze e risorse formative.

In Danimarca nel 2019 il National Coordination of Data Stewards Education (Wildgaard et al., 2020), con università ed industria, ha analizzato e sintetizzato lo stato della formazione sulla gestione dei dati di ricerca. Interessanti sono la definizione di data steward come figura-ponte fra le istituzioni, le infrastrutture e l’utente finale, e il riconoscimento della molteplicità di funzioni che il data steward può svolgere, a seconda del contesto in cui sia impiegato. Sono stati individuati 4 ruoli: agente di cambiamento, creatore, analista, amministratore. Questi 4 ruoli hanno un tratto comune: la competenza di base sui dati di dominio, alla quale vanno aggiunte le competenze più tecniche e trasversali. Il data steward è quindi una figura di professionalità elevata, che in primo luogo sa come trattare i dati di una specifica disciplina. Nella formazione dei data steward la parola chiave è flessibilità e un bilanciamento fra hard e soft skills.

Nell’ambito del progetto FAIRsFAIR è stata condotta un’analisi delle competenze richieste al data steward dal mondo del lavoro (Demchenko & Stoy, 2021). Nell’analisi viene evidenziata la molteplicità di queste, sintetizzata nella definizione di 7 estensioni al modello di riferimento per la Data Science, EDISON Data Science Framework, e in un elenco dettagliato di elementi da inserire nei curricula universitari.

2. La situazione in Italia

In Italia, come in molti Paesi, non esiste ancora un percorso formativo specifico e strut-

turato per il data steward. Tuttavia, molte sono le attività formative su Open Science e gestione dei dati che vengono proposte a varie figure professionali, tra cui i ricercatori. Negli ultimi anni progetti europei e iniziative, quali OpenAIRE, RDA, EOSC-Pillar, GO FAIR, OPERAS, hanno organizzato corsi ed eventi di formazione. Molti sono gli esempi di corsi trasversali offerti agli studenti di dottorato in varie università (Torino, Pisa, Scuola Normale Superiore, Genova, etc.).

In questo scenario uno dei principali attori in Italia è ICDI (Italian Computing and Data Infrastructure), il tavolo di lavoro creato dai rappresentanti di alcune tra le principali Infrastrutture di Ricerca e Digitali italiane per promuovere sinergie al fine di ottimizzare la partecipazione italiana alle attuali sfide europee, tra cui EOSC, la European Data Infrastructure e high-performance computing. Nel 2020 ICDI ha avviato la creazione del primo "Centro di Competenze" per EOSC, l'Open Science e la gestione dei dati FAIR (Lazzeri et al., 2021).

Il Centro di Competenze si pone come unico punto di riferimento nazionale per supporto, consulenza e formazione sui temi Open Science, FAIR data e EOSC, professionalizzazione dei data steward, ed infine servizi e strumenti per il FAIR by design tenendo conto delle specificità di settore. Un valore aggiunto è la creazione di percorsi formativi dedicati grazie alla rete di esperti di settore.

3. Conclusioni e prospettive future

L'Italia ha accumulato ritardo verso la realizzazione dell'Open Science e la formazione strutturata dei data steward. Una spinta positiva può venire dalla pubblicazione del Piano Nazionale per la Scienza Aperta, per ora solo anticipato nella sezione 6.2 del Programma Nazionale per la Ricerca 2021-2027. Il Piano prevede 4 assi: testi, dati della ricerca, valutazione della ricerca e coinvolgimento delle comunità. Nella sezione sui dati della ricerca si auspica la creazione di un data center nazionale che funga da centro di competenze.

Occorre un coordinamento nazionale sotto l'egida del Ministero dell'Università e della Ricerca, che promuova le pratiche di Open Science e crei sinergie sulla gestione dei dati con il Ministero della Salute per i dati sanitari e il Ministero dello Sviluppo Economico per la strategia sull'intelligenza artificiale.

Riferimenti bibliografici

- Demchenko, Y., & Stoy, L. (2021), Research Data Management and Data Stewardship Competences in University Curriculum, 2021 IEEE Global Engineering Education Conference (EDUCON), 1717–1726. <https://doi.org/10.1109/EDUCON46332.2021.9453956>
- Jetten, M., Grootveld, M., Mordant, A., Jansen, M., Bloemers, M., Miedema, M., & Gelder, C. W. G. van. (2021), Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship, Zenodo. <https://doi.org/10.5281/zenodo.4623713>
- Lazzeri, E., Tanlongo, F., Pavone, G., Alpi, F., Ansuini, A., Bertazzon, E., Bonaccorsi, D., Cappelluti, F., Casati, S., Castelli, D., Cippitani, R., Colcelli, V., Costantini, A., Cozzini, S., Degl'Innocenti, E., Di Donato, F., Di Giorgio, S., Fava, I., Fiore, S., ... Zane, D. (2021), Competence Centre ICDI per Open Science, FAIR, ed EOSC - Mission, Strategia e piano

d'azione, Zenodo <https://doi.org/10.5281/zenodo.5071055>

Mons, B. (2020), Invest 5% of research funds in ensuring data are reusable. *Nature*, 578(7796), 491–491. <https://doi.org/10.1038/d41586-020-00505-7>

Scholtens, S., Jetten, M., Böhmer, J., Staiger, C., Slouwerhof, I., Geest, M. van der, & Gelder, C. W. G. van. (2019), Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise, Zenodo <https://doi.org/10.5281/zenodo.3474789>

Teperek, M., Plomp, E. (2019), The role and value of data stewards in Universities: a TU Delft case study on data stewardship, Zenodo <http://doi.org/10.5281/zenodo.2684278>

Wildgaard, L., Vlachos, E., Nondal, L., Larsen, A. V., & Svendsen, M. (2020), National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum), Zenodo <https://doi.org/10.5281/zenodo.3609516>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. 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>

Autrici

Valentina Pasquale valentina.pasquale@iit.it

Valentina lavora come "Research Data Management Specialist" all'Istituto Italiano di Tecnologia, dove coordina i servizi di supporto per la gestione dei dati della ricerca e la scienza aperta. Coordina la GO FAIR Data Stewardship Competence Centers Implementation Network. Prima di occuparsi di dati della ricerca, ha ottenuto un dottorato in "Humanoid Technologies" all'Università di Genova nel 2010 e lavorato per diversi anni nel campo delle Neuroscienze Computazionali.

Emma Lazzeri emma.lazzeri@garr.it

Emma lavora per Consorzio GARR alla definizione di strategie e alla formazione e informazione in tema di scienza aperta. Coordina la Task Force di ICDI per il Centro di Competenza Nazionale per l'Open Science, FAIR ed EOSC. È coordinatrice dei nodi italiani di OpenAIRE e della Research Data Alliance. Ha conseguito il dottorato di ricerca in Innovative Technologies-Telecommunications presso Scuola Superiore Sant'Anna e un Master in Ingegneria delle telecomunicazioni presso l'Università di Pisa.

Elena Giglia elena.giglia@unito.it

Elena è responsabile dell'Unità di Progetto Open Access all'Università di Torino. Svolge un ruolo attivo in gruppi di lavoro nazionali e internazionali sull'Open Science. Ha fatto parte presso il MIUR della Commissione che ha redatto il nuovo Piano Nazionale per la Scienza Aperta. È stata membro del gruppo di lavoro "Open Access e nuovi indicatori" presso ANVUR (2013-2014). Partecipa attivamente a convegni su Open Science e svolge una intensa attività di formazione, informazione e promozione.