

Vuoi diventare un bibliotecario dei dati della ricerca?

Introduzione

Il seminario *Vuoi diventare un bibliotecario dei dati di ricerca?* si è tenuto il 15 marzo 2019 all'interno del Convegno delle Stelline. L'incontro è stato organizzato dall'International Master DILL (Università di Parma)¹ con OpenAIRE, RDA Italia, IOSSG (Italian Open Science Support Group).² Il seminario ha fatto seguito al precedente organizzato da IOSSG e l'International Master DILL alle Stelline lo scorso anno,³ che aveva messo insieme bibliotecari digitali e bibliotecari dei dati con un focus sulle diverse competenze. Lo scopo del Seminario è stato quello di chiarire le attività di cui attualmente si occupa il bibliotecario dei dati e capire di quale formazione e aggiornamento professionale abbia bisogno per acquisire le competenze richieste.

La Commissione europea (2016) ha stimato che nei prossimi dieci anni servirà almeno mezzo milione di professionisti per i servizi di supporto alla ricerca, tuttavia ci sono ancora ostacoli al loro riconoscimento e insufficienze nella formazione professionale. Il seminario ha cercato di rispondere alle seguenti domande: Quale esperienza può essere definita come buona pratica? Quali ostacoli e problemi devono essere ancora superati dal bibliotecario dei dati? Quale formazione e aggiornamento sono indispensabili per poter migliorare le proprie competenze?

Scienza aperta è buona scienza

Il contesto di riferimento del bibliotecario dei dati è la Scienza aperta (Open Science) che rappresenta un approccio alla ricerca collaborativo, trasparente e aperto. Rientrano nell'ambito della Scienza aperta un insieme di attività nel ciclo della ricerca e comportamenti collaborativi dei ricercatori come le pubblicazioni open access, open data, open peer review e open research. Comprende anche la citizen science (scienza

dei cittadini) o, più in generale, le attività partecipative in cui i non specialisti si impegnano direttamente nella ricerca. Scienza aperta non definisce quindi solo la proprietà pubblica della scienza, in particolare quella che è finanziata con denaro pubblico, ma comprende l'integrità, la riproducibilità e la qualità della ricerca. In altre parole, Scienza aperta è buona Scienza.

Fino a poco tempo fa, la ricerca scientifica aperta, necessaria alla condivisione e al riutilizzo dei dati, era limitata. Le pratiche di condivisione sono diventate più diffuse con l'avanzare della tecnologia digitale, sono state facilitate dalle politiche dei finanziatori di ricerca pubblica e sono aiutate dai servizi di supporto alla Scienza aperta delle istituzioni di ricerca. Ognuno di questi fattori ha contribuito a ciò che viene comunemente definito "il diluvio di dati". La maggior parte delle discussioni sulla crescita dei comportamenti di ricerca basati sulla condivisione e riutilizzo di dati aperti si sono concentrate sulla dimensione dei Big data. Vi sono altre questioni relative all'accesso aperto ai dati che non sono state ampiamente discusse, come l'aumento di nuovi profili di intermediari e nuove tipologie di prodotti di ricerca, oltre le pubblicazioni tradizionali. Il bibliotecario dei dati è uno dei nuovi profili come parte integrante – se non centrale – dell'infrastruttura che serve ai ricercatori per realizzare la Scienza aperta.

Elena Giglia (Università di Torino), Ilaria Fava (Göttingen State and University Library) e Emma Lazzeri (CNT ISTI) hanno descritto il contesto europeo della Scienza aperta ed i punti di riferimento che i bibliotecari dei dati possono oggi trovare in OpenAIRE e nei nodi di servizio RDA che formano i pilastri dell'infrastruttura cloud su cui si basa EOSC (European Open Science Cloud).

Elena Giglia nella sua presentazione dal titolo *Open Science, Open Access: cosa succede in Europa* ha illustrato le strategie della Commissione europea per realizzare l'infrastruttura EOSC che renderà possibile ottenere pubblicazioni e dati aperti FAIR (Findable, Accessible, Interoperable, Reusable) entro il 2022. Il processo sembra tuttavia lento e sono stati evidenziati sia ostacoli culturali che organizzativi, come ad esempio il riconoscimento⁴ dei risultati della ricerca inclusivo dei dati aperti. In Europa la conversazione è ora su "PlanS" che vuole sveltire la transizione all'Accesso aperto e i "Trasformative agreements" che cercano

di trasformare il modello di business delle pubblicazioni scientifiche dall'abbonamento ora prevalente all'accesso aperto necessario per realizzare la condivisione dei risultati di ricerca.

In Italia, dopo la L. 112/2013 che ha avviato una prima iniziativa politica per l'accesso aperto alle pubblicazioni scientifiche, è ora in discussione in Parlamento una proposta di legge (chiamata "Gallo" dal nome del relatore) che include i dati di ricerca, come prima base giuridica di un'infrastruttura nazionale, a partire dalla rete esistente dei depositi istituzionali.

Il mandato della Commissione europea per la Scienza aperta dichiara: "As open as Possible, as Closed as necessary". In pratica, la Commissione europea chiede ai docenti e ricercatori coinvolti in progetti per cui viene chiesto e ottenuto un finanziamento pubblico di depositare i risultati della ricerca, incluse le pubblicazioni scientifiche, in depositi ad accesso aperto e di produrre il Data Management Plan dei dati.

OpenAIRE è un'infrastruttura europea che vuole accompagnare le istituzioni di ricerca verso la Scienza aperta e rende disponibili linee guida e standard, come anche servizi e formazione. Ad esempio sono disponibili due toolkit per gli aspetti politici e giuridici della Scienza aperta: Policy toolkit, Legal toolkit. RDA (Research Data Alliance) riunisce in modo volontario quei ricercatori e docenti innovatori che già condividono i dati apertamente. L'alleanza serve per costruire ponti sia sociali, costruendo comunità, che tecnici, diffondendo buone pratiche. RDA è strutturato in *working group* e *interest group* ed organizza *plenary meetings* ogni sei mesi e *meeting nazionali* dove sono presentati l'aggiornamento su nuovi risultati e casi di studio. I soci italiani di RDA sono quasi quattrocento. Emma Lazzeri e Ilaria Fava hanno presentato il Progetto "RDA Europe 4" che intende essere il pilastro della realizzazione di Open science in Europa attraverso la creazione di una serie di nodi nazionali, di cui ISTI CNR di Pisa è il nodo italiano.

Buone pratiche

Data curation è un termine usato da molti anni dalla comunità professionale, ma rimane una pratica nuova ed emergente per molti bibliotecari che offrono un servizio di supporto per la gestione dei dati presso università e centri di ricerca in tutto il mondo. L'o-

biiettivo della cura dei dati è rendere i dati accessibili e utilizzabili a chiunque, in qualsiasi momento, ovunque e per qualsiasi scopo. Il servizio di supporto alla gestione dei dati di ricerca non è però un semplice “restyling” della conservazione per nuove tipologie di documenti dinamici, ma implica un ruolo del bibliotecario dei dati che si estende a tutte le fasi del ciclo della ricerca, a cominciare dal disegno e implementazione del progetto di ricerca. La caratteristica più importante della *data curation* è che rappresenta un sistema socio-tecnologico: sono necessarie nuove partnership dei bibliotecari con docenti e ricercatori insieme all’infrastruttura tecnologica formata da strumenti e sistemi informativi.

La visione del servizio delle biblioteche accademiche è stata chiaramente disegnata da ARL:

In 2033, the research library will have shifted from its role as a knowledge service provider within the university to become a collaborative partner within a rich and diverse learning and research ecosystem. ARL (2015).

Anche in Italia sono nate delle esperienze pionieristiche, collegate al gruppo IOSSG. Le buone pratiche di alcune delle maggiori università italiane sono state illustrate durante il Seminario da Paola Galimberti (Università di Milano), Elena Bianchi (Università di Padova), Marialaura Vignocchi (Università di Bologna). Alberto Rovelli (IULM) ha presentato un’esperienza di una università specializzata.

La caratteristica che contraddistingue il servizio di queste università italiane è stata definita “single point of entry”. Questo modello dei servizi di supporto ai dati aperti, elaborata dal Gruppo IOSSG, evidenzia il ruolo sociale del bibliotecario dei dati, che deve collaborare con gli altri uffici dell’università coinvolti nel ciclo della ricerca.

Vignocchi ha criticato la forte spinta alla frammentazione delle competenze e delle infrastrutture che c’è negli atenei. Nell’Università di Bologna in particolare il primo obiettivo è stato quello di concordare un modello gestionale per erogare servizi per l’utenza accademica attraverso un “protocollo” organizzativo tra aree coinvolte nel supporto all’Accesso aperto alle pubblicazioni e dati della ricerca (senza risorse aggiuntive!). Partendo dal ciclo della ricerca, si è cercato di costruire un flusso di lavoro (protocollo inter-area) in cui le diverse competenze sono state condivise.

Capire i bisogni dei docenti è una necessità per avviare un nuovo servizio di supporto. Tutte le esperienze presentate hanno cominciato partendo da indagini sui comportamenti dei ricercatori, cercando di capire dove attualmente pubblicano i loro risultati di ricerca. Paola Galimberti ha presentato i risultati delle interviste fatte dall’Università di Milano nel 2017 per capire quali dati venivano prodotti, chi ne era responsabile e quale fosse il grado di condivisione e apertura. I risultati hanno dimostrato la necessità di offrire nuovi servizi di supporto ma anche la necessità di un indirizzo istituzionale chiaro.

Un aspetto importante sono quindi le policy delle università, che rappresentano il mandato politico locale per offrire servizi di supporto alla gestione dei dati. Le Università di Milano, Padova e Bologna hanno utilizzato il modello predisposto da IOSSG e ispirato a LEARN Policy for Research Data Management.⁵ Sono risultate evidenti nel seminario anche le criticità che ancora ostacolano i bibliotecari dei dati. La percezione in Italia per la Scienza aperta è ancora soprattutto legata all’Accesso aperto agli articoli nei periodici, incluso il pagamento per pubblicare in Accesso aperto con editori predatori.

Dal punto di vista tecnologico, i servizi del bibliotecario dei dati si basano sull’infrastruttura dei depositi istituzionali.

Alberto Rovelli (IULM) ha parlato del deposito della IULM chiamato Apeiron (dal significato di “infinito” ma anche “incompiuto”) avviato nel 2000. In una prima versione i metadati delle risorse erano disponibili in OPAC in una sezione separata, dal 2011 i metadati sono inseriti in DSpace che è integrato in IRIS. La classificazione usata è quella basata sulle aree disciplinari MIUR e su una classificazione interna della ricerca. Il deposito delle pubblicazioni è obbligatorio solo per le tesi di dottorato.

Come strumento per la gestione dei dati di ricerca a Milano è stato utilizzato Dataverse,⁶ a Padova Eprints3.⁷ A Bologna sono stati usati i depositi e sistemi informativi esistenti a supporto della gestione dei progetti (SSRD-ARIC⁸ e IRIS⁹).

Le infrastrutture nuove o esistenti sono state arricchite con altri strumenti software. Il deposito di Padova consente l’autenticazione integrata con lo Shibboleth dell’Ateneo, come anche l’attribuzione del DOI, tramite la convenzione CRUI con il Consorzio internazionale Datacite; sono anche possibili connessioni tra i set di dati e gli articoli, sia quelli depositati

in IRIS sia quelli nel sito degli editori. Per l'indicizzazione viene usata la classificazione ERC (European Research Council).

In tutte le esperienze descritte, sono i ricercatori che caricano i set di dati e inseriscono i metadati, che vengono poi validati e pubblicati dai bibliotecari dei dati. I dati sono resi visibili attraverso i sistemi di discovery esistenti come R3Data (Registry of Research Data Repositories)¹⁰ e OpenDoar (Directory of Open Access Repositories).¹¹

Competenze emergenti

Le esperienze presentate hanno dimostrato che la gestione dei dati di ricerca spinge al rinnovamento e alla riqualificazione dei servizi bibliotecari. Tutte le esperienze presentate hanno aperto e mantengono anche una presenza virtuale nel ciclo della ricerca con una pagina all'interno del sito della ricerca o nel sito dello SBA (Sistema bibliotecario di ateneo).

La presentazione della Vignocchi ha così definito le competenze in campo:

...alle competenze tecniche va affiancata la cultura della collaborazione, la profonda conoscenza dell'istituzione, l'esperienza maturata nel trattare con le risorse e i progetti di ricerca, l'atteggiamento orientato all'utente, la capacità relazionale, l'autonomia lavorativa e la capacità decisionale, la formazione permanente.

Queste competenze evidenziano il ruolo sociale dei bibliotecari dei dati che offrono ai ricercatori servizi di supporto fin dalle fasi di progettazione e implementazione della ricerca. In particolare i bibliotecari dei dati svolgono alcune preziose attività che si possono sintetizzare come in Tabella 1.

Il ruolo dei bibliotecari dei dati è definito come *data*

| |
|---|
| Identificare e localizzare data sets |
| Identificare software e codice per il riutilizzo dei dati |
| Fornire supporto per l'accesso e le condizioni di riuso dei dati |
| Fornire consigli e supporto su strumenti informatici, di analisi e di discovery |
| Fornire supporto per la citazione dei dati e il referencing |

Tabella 1 - Attività del bibliotecario dei dati nella fase di progettazione e implementazione

curation con il compito di curare le collezioni di dati nei depositi nelle diverse fasi del ciclo della ricerca. In sintesi lo si potrebbe descrivere come in Tabella 2.

| |
|--|
| Garantire la corretta configurazione e gestione dei repository, la loro registrazione e la loro certificazione in collaborazione con ICT |
| Garantire la conservazione a lungo termine dei dati |
| Garantire adeguato back up e storage dei dati: data security |
| Assicurare la disseminazione |
| Sviluppare servizi per la fruizione dei dati |

Tabella 2 - Attività del bibliotecario dei dati nella fase di gestione e accesso ai dati di ricerca

Il bibliotecario dei dati affronta prima i problemi sistemici piuttosto che i singoli strumenti tecnologici all'interno del sistema informativo.

Presso l'Università di Bologna e Padova i sistemi bibliotecari hanno fatto formazione dello staff coinvolto nei servizi di supporto. La nuova generazione di bibliotecari dei dati può trovare offerte formative specifiche, come il Master DILL. Per il personale in servizio, il seminario ha stimolato chi voglia avviare dei servizi all'autoformazione: sono tanti i materiali accessibili in rete, a cominciare da quelli raccolti da RDA Italia e OpenAire. Il seminario ha anche evidenziato l'importanza della capacità di fare rete, come dimostra IOSSG e RDA.

Conclusioni

La Scienza aperta stimola un'estensione del ruolo del bibliotecario, integrato nel ciclo della ricerca e non più limitato al flusso sequenziale di acquisizione, organizzazione e accesso delle pubblicazioni.

Le buone pratiche descritte durante il seminario dimostrano che anche in Italia bisogna facilitare la Scienza aperta e la condivisione dei risultati di ricerca. L'approccio necessario è quello sistemico definito da Borgman (2015) come sistema socio tecnologico. Non bisogna partire cioè dalla granularità della collezione, ma dai bisogni degli utenti (autori e riutilizzatori) e dal contesto dell'istituzione (Corrall 2012). Di conseguenza, la comprensione del ruolo definito dalla *data curation*, come un insieme di pratiche socio-tecnologiche, e i titoli formali con i quali i bi-

bliotecari dei dati operano tipicamente (come data scientist, curatore dati di ricerca ecc.), possono variare considerevolmente nel tempo, tra diversi contesti istituzionali e attraverso i confini internazionali. Non esistono ancora sistemi di certificazione e accreditamento per questi nuovi profili. A livello internazionale, ci sono diverse qualifiche, nuovi profili e diversi livelli di servizio che sono stati indicati in diverse fasi del ciclo della ricerca (illustrati da DataOne in Figura 1).

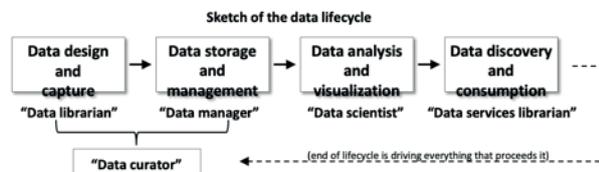


Figura 1 - Profili professionali nel ciclo di vita dei dati
Fonte: DataONE <https://www.dataone.org/data-life-cycle>

Il flusso (o meglio ciclo perché iterativo) della ricerca indica il bibliotecario dei dati insieme al gestore dei dati (anche chiamati Data curator) che offrono servizi nelle prime fasi della ricerca di creazione e memorizzazione dei dati di ricerca, il Data scientist e il Data services librarian sono attivi nelle fasi di scoperta dei dati e di ri-uso, in particolare l'analisi e la visualizzazione dei dati. Il ruolo del bibliotecario dei dati, come profilo emergente, può attrarre una nuova generazione di bibliotecari e anche rivitalizzare l'immagine dei bibliotecari in servizio. Tuttavia, nella discussione, una domanda ha smorzato gli entusiasmi dei pionieri e organizzatori del seminario: "Non è il ruolo dei bibliotecari gestire i dati di ricerca! Perché le biblioteche universitarie devono assumersi il compito di servizi di supporto al ciclo della ricerca?" La risposta necessaria a questa domanda è che sono cambiate le esigenze degli utenti, le biblioteche accademiche devono soddisfare nuovi bisogni e per ottenere questo risultato devono integrarsi nei flussi di lavoro dei ricercatori e contribuire al successo dell'istituzione universitaria in un momento di grande trasformazione. Questo continua evoluzione per essere sempre adeguati (anzi anticipare) i bisogni degli utenti è il significato vero di "la Biblioteca è un organismo che cresce" di Ranganathan.

ANNA MARIA TAMMARO

International Master DILL Università di Parma
annamaria.tammaro@unipr.it

RIFERIMENTI BIBLIOGRAFICI

- ARL (2015) *Strategic Thinking + Design Initiative report*.
 BORGMAN, C. (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World*, Cambridge: MIT Press.
 CORRALL, S.(2012). *Roles and responsibilities: Libraries, librarians and data*. In G. Pryor (Ed.) *Managing Research Data* (pp.105-133). London: Facet.
 DATAONE: <https://www.dataone.org/data-life-cycle>
 TAMMARO A.M. (2018), *Una proposta non sovversiva*, Biblioteche oggi (36) pp. 65-71.

NOTE

- ¹ <https://www.unipr.it/node/17891>.
- ² <https://sites.google.com/view/iOSSG/home>.
- ³ Nel seminario del 2018 il bibliotecario dei dati era stato definito così: "I bibliotecari dei dati offrono servizi di supporto a ricercatori e ri-utilizzatori dei dati di ricerca, nel processo di creazione, conservazione, accesso e uso dei dati". Un breve resoconto del seminario pubblicato in: Tammaro A.M. (2018) *Una proposta non sovversiva*, Biblioteche oggi (36) pp. 65-71.
- ⁴ A questo scopo la Commissione europea ha pubblicato un rapporto sulla valutazione della ricerca nel 2017: *Evaluation of research career fully acknowledging Open science practice*: http://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport.pdf.
- ⁵ Informazioni sul Progetto LEARN accessibili al link: <http://learn-rdm.eu/en/about>.
- ⁶ <https://dataverse.org>.
- ⁷ <https://www.eprints.org/uk/index.php/eprints-software/>
- ⁸ <https://www.unibo.it/it/ateneo/organizzazione/amministrazione-generale/10010/3288>.
- ⁹ <https://www.cineca.it/it/content/IRIS>.
- ¹⁰ <https://www.re3data.org>.
- ¹¹ <http://v2.sherpa.ac.uk/opensoar>.

DOI: 10.3302/0392-8586-201904-058-1