

Open data steward: bibliotecari e alfabetizzazione ai dati aperti

di Damiano Orrù

Introduzione

Nella realtà di ogni giorno siamo circondati da strumenti e servizi che possono favorire la produttività e semplificare problematiche più frequenti nella vita di molti cittadini. Alcuni trovano utile il robot aspirapolvere, altri il navigatore sul cellulare o in automobile, alcuni prediligono i pagamenti *contactless*, altri le *app* elimina code, alcuni si affidano all'*app* Immuni, altri usufruiscono dei servizi per la casa domotica (la gestione remota e digitale di illuminazione, telecamere, riscaldamento, dispensa, cancelli automatici ecc.).

Tutti questi nuovi protagonisti delle nostre esistenze hanno lo stesso cuore, definito in gergo informatico 'algoritmo'. Si parla di 'intelligenza artificiale', forse in modo sproporzionato rispetto a ciò che effettivamente viene gestito e offerto. Questo cuore tecnologico può pulsare e sopravvivere solo se alimentato dal plasma digitale: i dati, disponibili in modo adeguato per formati, quantità, tempestività e interoperabilità.

Il bibliotecario non evolve in tecnico informatico o in ingegnere programmatore, ma deve saper intuire e distinguere:

- i confini dell'automatismo dell'elaborazione informatica;
- i potenziali conflitti o lacune di informazioni;
- gli strumenti attendibili per il supporto del cittadino.

Questi ambiti evoluti dell'information literacy trovano applicazione nella gestione dei dati aperti in modo etico e sostenibile, dove il bibliotecario diviene *open data steward*. Le potenzialità e le premesse per questo cambiamento sono di stretta attualità anche in ottica di sviluppo sostenibile generato dal *citizen engagement* in biblio-

DAMIANO ORRÙ, Università degli studi di Roma "Tor Vergata", Facoltà di Economia, Biblioteca "Vilfredo Pareto", e-mail orru@economia.uniroma2.it.

I contenuti di questo lavoro sono approfonditi nella tesi di master Damiano Orrù; Gloria Fiorani, *Data stewardship in biblioteca per la scienza sostenibile e aperta* [tesi di master]. Roma: Università degli studi di Roma "Tor Vergata", 2020, <<https://zenodo.org/record/3779063>>. Da giugno 2020 è disponibile il blog *Open data steward* dedicato alla tesi e ai contenuti correlati, con possibilità di interazione con cittadini, bibliotecari e esperti tramite i canali Facebook, Twitter e Instagram, <<https://opendatasteward.altervista.org>>.

Ultima consultazione siti web: 24 settembre 2020.



teca, favorendo la condivisione e l'accesso sia ai dati attendibili sia alla produzione scientifica. In particolare, l'alfabetizzazione ai dati contribuisce al Traguardo 4.6 dell'Agenda 2030 delle Nazioni unite per lo sviluppo sostenibile (da ora in avanti: Agenda 2030): «assicurarsi che tutti i giovani e una parte sostanziale di adulti, uomini e donne, raggiungano l'alfabetizzazione e l'abilità di calcolo»¹ nell'Obiettivo 4 "Educazione di qualità".

La *data stewardship* può essere supportata dalla certificazione etica SA 8000, non solo in ambito accademico, ma anche in tutte le realtà lavorative e produttive. Nel mondo delle biblioteche la *data stewardship* della ricerca accademica si basa su archivi, pratiche e strumenti potenzialmente favorevoli al Traguardo 16.10 dell'Agenda 2030: «garantire l'accesso del pubblico alle informazioni e proteggere le libertà fondamentali, in conformità con la legislazione nazionale e con gli accordi internazionali»² nell'Obiettivo 16 "Pace, giustizia e istituzioni solide".

Data steward per la ricerca

La connessione tra sostenibilità e ruolo formativo della biblioteca è ribadita tra le funzioni del *Nuovo manifesto per le biblioteche digitali*:

Le biblioteche digitali gestiscono i dati e i prodotti della ricerca scientifica, ne favoriscono la libera circolazione e ne supportano l'accesso aperto ai fini della diffusione universale della conoscenza. Le biblioteche digitali considerano proprio compito fondamentale la gestione efficace dei dati e dei prodotti della ricerca scientifica, compresi i dati grezzi e i metadati, il loro mantenimento nel tempo e l'accrescimento del loro valore (*Data Curation*) tramite l'interazione fra i dati stessi e con gli utenti; condividono i principi della *Open Science* e in particolare le prassi sviluppate a partire dalla Dichiarazione di Berlino, ai fini della massima diffusione dei risultati originali della ricerca, specie se finanziata pubblicamente; supportano a questi fini la realizzazione degli obiettivi dell'accesso libero e senza restrizioni, della distribuzione illimitata, dell'interoperabilità e dell'archiviazione a lungo termine e favoriscono il mantenimento degli standard qualitativi della validazione dei contenuti e della buona pratica scientifica³.

Il punto di partenza e filo conduttore è la *data stewardship*⁴, analizzata valutando:

- i ruoli, le funzioni e le responsabilità di un *data steward*;
- la gestione sistematica, sostenibile e responsabile dei dati a beneficio del pubblico.

Il bibliotecario come *data steward*⁵ ha il compito di promuovere il cambiamento culturale verso la ricerca condivisa, agendo su chi produce e su chi riusa i dati. Si

1 Cfr. United Nation. Development Programme, *The 2030 Agenda for sustainable development*, target 4.6, <<https://open.undp.org/sdg/targets/4/6>>.

2 *Ivi*, target 16.10, <<https://open.undp.org/sdg/targets/16/10>>.

3 Cfr. Associazione italiana biblioteche. Gruppo di lavoro sulle biblioteche digitali, *Nuovo manifesto per le biblioteche digitali*. 5 maggio 2020, punto 26, <<https://www.aib.it/struttura/commissioni-e-gruppi/gruppo-di-lavoro-biblioteche-digitali/2020/82764-nuovo-manifesto-per-le-biblioteche-digitali>>.

4 Cfr. *Steward*, in *Merriam-Webster dictionary*, <<https://www.merriam-webster.com/dictionary/steward>>.

5 Cfr. Christine Staiger [et al.], *Data stewards function landscape and its stakeholders*. 12 giugno 2019, <<https://zenodo.org/record/346052#.XkBO52hKhhk>>, DOI: 10.5281/zenodo.346052.

richiede uno sforzo anche per agevolare il trasferimento tecnologico necessario per la scienza aperta, tramite la sinergia a tutti i livelli operativi, finanziari e decisionali dell'istituzione. In tempi ragionevoli è necessario raggiungere una massa critica, supportata da un ruolo proattivo dei ricercatori e dei settori ICT (*information and communication technology*).

Inoltre il *data steward* dovrà sostenere piattaforme di scambio per informazioni pratiche, per domande frequenti, per soluzioni replicabili, anche tramite gruppi di discussione di mutuo sostegno.

In questa ottica possono essere particolarmente utili le iniziative proposte a livello internazionale tra cui Research Data Alliance⁶, European open science cloud⁷, Italian open science support group⁸ ecc.

Il *data steward* risponde con le sue competenze alle esigenze tipiche dei DMP (*data management plan*), supportando cittadini e ricercatori che condividono:

- come vengono generati e/o reperiti i dati;
- come vengono utilizzati, elaborati e organizzati i dati;
- come sono protetti i dati e le persone interessate;
- come vengono descritti e documentati i dati, il codice e gli elementi ausiliari;
- come vengono memorizzati e protetti i dati e per quanto tempo saranno conservati;
- come vengono assegnati la paternità e il credito dei dati;
- come vengono conservati i dati;
- come, se e in quali termini i risultati della ricerca possono essere condivisi.

Se esistono variazioni sulla struttura di archiviazione, sui formati, sui software è necessario documentare in modo accurato come si modifica il processo e come si può consentire nel futuro la replicabilità dell'elaborazione/esperimento. In questa visione oltre ai dati è fondamentale dettagliare chiaramente la metodologia, la conservazione e le condizioni di riuso.

Standard internazionali e certificazione *stewardship*

Nella valutazione della *stewardship* per l'architettura hardware/software dei *repository* esiste il certificato Coretrustseal⁹, una certificazione internazionale che attesta la rispondenza tecnica ai principi FAIR (*findable, accessible, interoperable, reusable*) nella verifica della sostenibilità per la conservazione digitale nel lungo periodo. In Italia la piattaforma Phaidra¹⁰ del Sistema bibliotecario di ateneo dell'Università degli studi di Padova ha raggiunto questo obiettivo tramite la verifica della gestione dell'archivio nel garantirne la sostenibilità per la conservazione digitale nel lungo periodo.

In Italia non esiste una certificazione o un sistema di accreditamento che possa rilasciare la certificazione *data stewardship*. Il bibliotecario evolve da *data curator* a *data steward* al fine di migliorare la gestione dei dati per la scienza aperta. Questo comporta la necessità di adattare le competenze e le professionalità alle esigenze dell'istituzione, della disciplina, del paese e della regione in cui si svolge la ricerca.

6 Research data alliance, <<https://www.rd-alliance.org>>.

7 European open science cloud (EOSC), <<https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>>.

8 Italian open science support group (IOSSG), <<https://sites.google.com/view/iossg/home>>.

9 Cfr. <<https://www.coretrustseal.org/why-certification/certified-repositories>>.

10 Cfr. <<https://bibliotecadigitale.cab.unipd.it/news/phaidra.-ultime-novita>>.

Il ricercatore è il principale *data stakeholder* all'interno della comunità scientifica, legata a *partnership* tra atenei (soprattutto nei progetti europei) e tra organizzazioni di ricerca concorrente come i ministeri, le aziende, i consorzi, le fondazioni, le onlus (organizzazione non lucrativa di utilità sociale). Diviene essenziale stabilire una cooperazione pragmatica, che si affianchi alla condivisione di competenze, requisiti, servizi e protocolli, formati, modelli e standard che travalicano i confini geografici e interdisciplinari.

Si parla di una nuova concezione della *data stewardship* definita *FAIR research data stewardship* volta a condividere:

- buone pratiche FAIR, ottimizzare le tecnologie (*app* e codice open source) e i *repository*;
- modelli di metadati FAIR (generici o specifici di una disciplina);
- strategie per gestire la privacy dei dati, le licenze, la sicurezza e il copyright;
- test e *feedback* degli strumenti finanziati;
- competenze necessarie e i requisiti per la gestione dei dati;
- strumenti di *e-learning*, corsi e programmi di formazione;
- incentivi approvati per supportare la ricerca FAIR;
- strategie per rendere sostenibile a lungo termine questo processo¹¹.

Lo standard internazionale SA 8000 sancisce i requisiti per un comportamento eticamente corretto delle imprese e della filiera di produzione verso i lavoratori. Nel 2020 oltre 4.483 organizzazioni (2.092.945 lavoratori, in 60 paesi e 57 settori produttivi)¹² hanno scelto SA 8000 per costruire un ambiente di lavoro sostenibile analizzando:

- lavoro infantile;
- lavoro obbligato;
- salute e sicurezza;
- libertà di associazione, diritto alla contrattazione collettiva;
- discriminazione;
- pratiche disciplinari;
- orario di lavoro;
- retribuzione;
- sistema di gestione.

Questo standard basa la valutazione dell'organizzazione sul coinvolgimento delle parti interessate (*stakeholder* interni ed esterni), con l'apporto del personale in servizio a ogni livello/sede. Le aziende sono parte dei *data stakeholder*, quanti necessitano di dati disponibili in tempi rapidi e con piattaforme interoperabili e basate su *policy* e standard di ingresso e uscita nel flusso di gestione informativo: la *data governance*.

Alla base vi è l'orientamento a migliorare la competitività del mercato su base volontaria degli *stakeholder*, tramite la rilevazione di eticità della propria filiera produttiva e del proprio ciclo produttivo. Nell'ottica della sostenibilità divengono fondamentali temi come: il rispetto dei diritti umani, il rispetto dei diritti dei lavoratori, la tutela contro lo sfruttamento dei minori, le garanzie di sicurezza e salubrità sul

¹¹ Data Stewardship Competence Centres (DSCC), <<https://www.go-fair.org/implementation-networks/overview/dscc>>.

¹² Dati al 20 febbraio 2020, cfr. Social Accountability Accreditation Services, *SA8000 certified organisations: certification statistics*. <<http://www.saasaccreditation.org/certifacilitieslist>>.

posto di lavoro¹³. In questa 'analisi di bilancio sociale' gli attori coinvolti sono fornitori, clienti, consumatori e loro associazioni, gruppi ambientalisti e umanitari, organizzazioni sindacali, *mass media*, il bacino di utenza locale/nazionale.

I riferimenti tecnici sono sintetizzati dagli allegati: indicatori di performance (documento che definisce le aspettative minime di performance di un'organizzazione) e documento guida SA 8000 (interpretazione dello standard per chiarire come applicarlo in sede di audit). Questo approccio si basa su interviste per ottimizzare la gestione informativa al fine di realizzare i vantaggi competitivi legati a:

- un incremento della credibilità, trasparenza e dell'immagine aziendale sul mercato di riferimento, un beneficio reputazionale;
- una maggiore fiducia da parte dei consumatori, miglioramento del rapporto con le istituzioni e le organizzazioni sociali;
- un controllo dell'eticità e della correttezza sociale nella catena dei fornitori;
- un miglioramento del clima aziendale: i lavoratori si sentono tutelati dall'impresa in cui operano e maggiormente coinvolti nel raggiungimento degli obiettivi;
- un miglioramento della comunicazione interna ed esterna, mediante rapporti resi pubblicamente disponibili¹⁴.

La *data stewardship* si occupa della gestione etica e responsabile del dato o del set di dati, focalizzando l'architettura della base dati e i relativi processi di automazione, al fine di favorire:

- l'uso coerente dei dati di gestione delle risorse;
 - una più facile mappatura dei dati tra sistemi informatici e lo scambio di documenti;
 - una riduzione dei costi associati alla migrazione di architetture di servizio.
- Quindi gli strumenti di *data stewardship* sono utilizzati nei progetti di *data quality* e *master data management* con lo scopo di eliminare entità duplicate e migliorare la riusabilità e la qualità dei dati aziendali¹⁵.

Data literacy per il cittadino

Negli ultimi anni si è molto discusso di dati e fonti affidabili per un razionale processo decisionale, volto ad ampliare le capacità statistiche e la fruizione dei dati da parte di persone comuni, come ad esempio per i dati in ambito alimentare, sanitario, civico, sociale e ambientale. Come ribadito tra i principi del *Nuovo manifesto per le biblioteche digitali*:

Le biblioteche digitali interagiscono con le proprie comunità. Le biblioteche digitali basano il loro rapporto con gli utenti sul metodo dell'interazione, favoriscono la partecipazione attiva degli utenti alla propria comunità, anche attraverso la creazione di nuove risorse, e cercano di parlare il loro linguaggio. Supportano inoltre l'attivazione di percorsi dedicati allo sviluppo della competenza informativa degli utenti stessi, intesa come capacità di ricerca indipendente,

13 Cfr. *La responsabilità sociale e etica*, <<http://www.sa8000.info>>.

14 CSQA certificazioni, *SA 8000*, <<https://www.csqa.it/CSQA/Norme/Responsabilita-Sociale/SA-8000>>.

15 Annalisa Casino, *Stewardship: politiche e pratiche per una gestione etica delle risorse*, prefazione di Russ Martin. Roma: Aracne, 2014, p. 85.

di analisi, valutazione critica, riuso efficace delle diverse fonti informative, e di comprensione di come l'informazione è prodotta e valutata, anche al fine di produrre nuova conoscenza¹⁶.

In applicazione di questo principio, a contrasto della disinformazione online attraverso l'alfabetizzazione alla ricerca e alla verifica delle informazioni in rete, si segnala il blog *Biblioverifica*¹⁷. L'iniziativa nasce nel 2017 nell'interazione di 'bibliovolontari' (bibliotecari e archivisti), tramite i *social* a supporto dei cittadini, promuovendo le pratiche di «fact-checking e l'utilizzo di dati aperti, fonti certificate, strumenti di ricerca gratuiti, strategie di ricerca per agevolare l'accesso all'informazione sostenibile in termini di accuratezza, tracciabilità, indipendenza, legalità, imparzialità»¹⁸.

Anche negli Stati Uniti la disinformazione viene contrastata con strumenti aperti e gratuiti come portali, blog e MOOC (*massive open online course*, corsi online aperti su larga scala). Dal blog *Metaliteracy* scaturisce il corso gestito da Thomas P. Mackey e Trudi E. Jacobson, "Empowering yourself in a post-truth world", fruibile dalla piattaforma gratuita Coursera, tramite brevi contenuti testuali e video, quiz di autovalutazione integrati da un forum di discussione con i tutor¹⁹. Per *metaliteracy* si intende un insieme di pratiche e contenuti tramite i quali i bibliotecari possono supportare studenti e docenti, anche in ambito accademico, nella comprensione delle informazioni e dei dati²⁰. In questi anni l'analisi e la verifica delle fonti e dell'affidabilità delle informazioni in rete sono divenute indispensabili per combattere la disinformazione, educando i cittadini al pensiero critico²¹.

La ricerca accademica è al servizio dei cittadini, primi finanziatori e primi *stakeholder* dei dati aperti con impatto sociale, spingendo al riuso e alla collaborazione in modo rapido e agile con imprese e società civile. Per diffondere questo tipo di pratiche e strumenti esistono piattaforme di contenuti e complessità diverse, rivolte al pubblico del web, particolarmente utili a dottorandi e ricercatori:

- tutorial "Mantra"²², progetto del Research Data Service in Information Services dell'University of Edinburgh²³;
- MOOC "Research data management and sharing" dell'University of North Carolina at Chapel Hill e della University of Edinburgh²⁴;

16 AIB. GBDIG, *Nuovo manifesto per le biblioteche digitali* cit., punto 7.

17 Blog ideato e gestito con il supporto di Paola Coppola, direttrice della Biblioteca di area economica "Vilfredo Pareto" dell'Università degli studi di Roma "Tor Vergata", <<http://www.biblioverifica.cloud>>.

18 Damiano Orru; Paola Coppola, *#Biblioverifica*, il blog a supporto del cittadino fact-checker, «Biblioteche oggi trends», 5 (2019), n. 2, p. 62-65, <<http://www.bibliotecheoggi.it/trends/article/view/1005>>, DOI: 10.3302/2421-3810-201902-062-1.

19 Tom Mackey è docente di Arts and media presso il Suny Empire State College (New York), Trudi E. Jacobson, bibliotecaria responsabile dell'Information Literacy Department della State University of New York, cfr. <<https://www.coursera.org/learn/empowering-yourself-post-truth-world>>.

20 Cfr. *Metaliterate learning for the post-truth world*, edited by Thomas P. Mackey, Trudi E. Jacobson, foreword by Troy A. Swanson. Chicago: American Library Association, 2019, p. 146.

21 Cfr. Greta Lugli, *Una panoramica sulla biblioteca pubblica d'oggi*. Brescia: Temperino rosso, 2019, p. 75.

22 Cfr. <<https://mantra.edina.ac.uk>>.

23 *Edinburgh research data blog*, <<http://datablog.is.ed.ac.uk>>.

24 Cfr. <<https://www.coursera.org/learn/data-management>>.

- MOOC “Essentials 4 data support”²⁵;
- webinar “Eurodoc open science ambassador training” dell’European Council of Doctoral Candidates and Junior Researchers²⁶;
- MOOC “Open science: sharing your research with the world”²⁷ della Technische Universiteit Delft Library²⁸ all’interno del progetto 4TU.researchdata²⁹;
- MOOC “Open science MOOC”³⁰ articolato in dieci moduli, composti da video, articoli di ricerca, set di ‘dati fittizi’ e codice sulla piattaforma ideata e fondata da Jonathan Tennant³¹, paleontologo e ricercatore indipendente presso l’Institute for Globally Distributed Open Research and Education, scomparso prematuramente il 9 aprile 2020;
- Tutorial “Data carpentry”³² integrato nelle iniziative Library carpentry³³ e Software carpentry³⁴, spazi formativi per la gestione, la pulizia e l’elaborazione di dati.

In una visione di ricerca sostenibile ed etica nascono partnership in cui i *data steward* possono essere protagonisti:

- nella scelta dei partner nel coinvolgere la comunità;
- nella definizione di ruoli e compiti interni all’organizzazione;
- nel definire interviste per raccogliere dati;
- nella divulgazione e comunicazione dei risultati;
- nelle collaborazioni per la sostenibilità dei progetti nel lungo periodo.

I cittadini possono accedere a un vasto repertorio di dati aperti, che possono essere compresi e riutilizzati se si dispone delle pubblicazioni che utilizzano o riusano i dati della ricerca. Per questo traguardo è fondamentale l’apertura delle pubblicazioni scientifiche.

Per la consultazione delle riviste ad accesso aperto è disponibile il portale DOAJ (Directory of open access journals), creato nel 2003 dalla Lund University, repertorio online di riviste di alta qualità ad accesso aperto e sottoposte a *peer review*.

La missione di DOAJ è quella di aumentare la visibilità, l’accessibilità, la reputazione, l’uso e l’impatto delle riviste di ricerca accademica di qualità, con revisione paritaria, ad accesso aperto a livello globale, indipendentemente dalla disciplina, dalla geografia o dalla lingua. DOAJ lavorerà con editori, editori e proprietari di riviste per aiutarli a comprendere il valore dell’editoria e degli standard delle migliori pratiche e ad applicarli alle proprie operazioni. DOAJ

25 Cfr. <<https://datasupport.researchdata.nl/en>>.

26 Cfr. <<http://eurodoc.net/ambassadors>>.

27 Cfr. <<https://www.edx.org/course/open-science-sharing-your-research-with-the-world>>.

28 TU Delft Library, <<https://www.tudelft.nl/library>>.

29 4TU.researchdata, <<https://researchdata.4tu.nl/en/about-4turesearchdata/organisation>>.

30 Cfr. <<https://opensciencemooc.eu>>.

31 Alla memoria di Jonathan Tennant è dedicata la tesi di master da cui è tratto questo articolo.

32 Cfr. <<https://datacarpentry.org>>.

33 Cfr. <<https://librarycarpentry.org/blog>>.

34 Cfr. <<https://software-carpentry.org>>.

si impegna a essere indipendente al 100% e a mantenere tutti i suoi servizi e metadati liberi da usare o riutilizzare per tutti³⁵.

Per l'accesso aperto a e-book scientifici con *peer review* è disponibile DOAB (Directory of open access books), con la possibilità di scaricare liberamente capitoli o monografie a testo integrale, tramite editori accademici in accesso aperto. Tra le discipline maggiormente presenti troviamo le scienze umane, giuridiche e sociali. Il portale è stato creato da Oapen (Open access publishing in European networks), attualmente gestito con OpenEdition, CNRS (Centre national de la recherche scientifique) e Aix-Marseille Université.

L'obiettivo principale di DOAB è aumentare la rilevabilità dei libri Open Access. Gli editori accademici sono invitati a fornire a DOAB i metadati dei loro libri Open Access. I metadati saranno raccolti per massimizzare la diffusione, la visibilità e l'impatto. Gli aggregatori possono integrare i record nei loro servizi commerciali e le biblioteche possono integrare la directory nei loro cataloghi online, aiutando studiosi e studenti a scoprire i libri. La directory è aperta a tutti gli editori che pubblicano libri accademici sottoposti a peer review in Open Access e devono contenere il maggior numero possibile di libri, a condizione che tali pubblicazioni siano in Open Access e soddisfino gli standard accademici³⁶.

Citizen engagement tramite i dati aperti

Alla base della reperibilità dei dati vi è l'interoperabilità e l'indicizzazione tramite *tag* specifici nel vocabolario strutturato Schema.org³⁷, compatibile con le applicazioni IoT (*internet of things*), spingendo verso la standardizzazione dei metadati, che nel futuro potrà supportare algoritmi AI (intelligenza artificiale). La rete offre strumenti di ricerca gratuiti come Google dataset search³⁸, che consente lo scarico di serie storiche aperte e condivise.

Il cittadino avrà diversi gradi di fruibilità del dato online, come misurabile nella classificazione ideata da Tim Berners-Lee³⁹, inventore del protocollo HTTP del world wide web, promotore dei linked data. Il grado di apertura dei dati è valutabile in cinque livelli in scala crescente (come per le strutture ricettive o le piattaforme di acquisto online) con diversi benefici strutturati dall'Open Data Institute⁴⁰.

Se il cittadino può visionare, stampare o copiare uno o più dati, disponibili con licenza aperta⁴¹, questi ottengono una stella di gradimento.

Per raggiungere la seconda stella i dati devono essere strutturati tramite *app* o software proprietari chiusi, come Microsoft Excel.

35 Cfr. <<https://doaj.org>>.

36 Cfr. <<https://www.doabooks.org>>.

37 Cfr. *IoT and Schema.org: getting started*, release 3.2, edited by Dan Brickley, <<https://schema.org/docs/iot-gettingstarted.html>>.

38 Cfr. Davide Castelvocchi, *Google unveils search engine for open data: the tool, called Google dataset search, should help researchers to find the data they need more easily*, «Nature», 561 (2018), n. 7722, p. 161-162, <<https://www.nature.com/articles/d41586-018-06201-x>>, DOI: 10.1038/d41586-018-06201-x.

39 Cfr. 5 star open data. 31 agosto 2015, <<https://5stardata.info/en/>>.

40 Cfr. <<http://theodi.org/>>.

41 Cfr. <<https://creativecommons.org/publicdomain/zero/1.0/deed.it>>.

La terza stella è disponibile solo se i dati sono scaricabili in formato CSV (*comma-separated value*), aperto non proprietario.

Diversamente, il formato URI (*uniform resource identifier*) consente la tracciabilità universale e invoca dei dati, valutati con quattro stelle per la fruibilità condivisa in ogni ambiente connesso al web. Un esempio di questa codifica è RDF (*resource description framework*), che è alla base dell'interoperabilità tra applicazioni, consentendo il passaggio al livello massimo.

La valutazione massima di cinque stelle è attribuita ai linked open data connessi con altre serie di dati esterni, consentendo l'interoperabilità simultanea di dati correlati congiunti e dialogando con lo stesso schema strutturato.

Il progetto dati.gov nasce con l'obiettivo di divulgare la cultura dei dati aperti, accrescere la trasparenza della Pubblica amministrazione. Il portale è gestito e sviluppato dall'AgID (Agenzia per l'Italia digitale) all'interno delle strategie di *open government*, articolate in venti azioni⁴², in parte finalizzate allo sviluppo inclusivo e sostenibile, all'interoperabilità e allo scambio dei dati pubblici con gli *stakeholder*.

Nel 2012 viene divulgato il portale Open data UE⁴³ per condividere i dati dei paesi e delle istituzioni europee nel rispetto delle normative nazionali ed europee, per consentire il riutilizzo libero e gratuito da parte della comunità internet. In Italia Dati open⁴⁴ indicizza fonti di dati esterne per convertirli in formato RDF, con cui è possibile interagire a livello di linked open data per consentire un accesso uniforme a dati in tabelle, grafici o mappe. UNdata⁴⁵ è stato lanciato nel 2005 nell'ambito di un progetto denominato Statistics as a public good (Statistica come bene pubblico) con l'obiettivo di fornire un accesso gratuito alle statistiche globali, di alfabetizzare alla statistica gli utenti del web. Il Fondo monetario internazionale (FMI) offre accesso alle statistiche prodotte tramite il portale IMF Data Portal⁴⁶, unitamente a strumenti statistici grafici, per favorire l'analisi dei dati personalizzate. Uno strumento di monitoraggio finanziario è Sistema informativo sulle operazioni degli enti pubblici (Siope)⁴⁷, dove il cittadino può tracciare incassi e pagamenti effettuati da tutte le amministrazioni pubbliche (enti, ministeri, comuni, regioni, municipi, atenei, ASL ecc.), tramite la sinergia tra la Ragioneria generale dello Stato, la Banca d'Italia e l'Istat. Dal 2015 Istat⁴⁸ offre la piattaforma linked open data per accedere e navigare i dati interoperabili direttamente tramite il vocabolario RDF-data cube.

Negli ultimi anni sono nate iniziative che coinvolgono i cittadini come scienziati per coinvolgerli in sperimentazioni di dati con il fine di agevolare la diffusione della ricerca scientifica, la partecipazione e la comprensione della scienza e della tecnologia tra il pubblico in generale. Un esempio è Scistarter⁴⁹, comunità online dedi-

42 Cfr. Vincenzo Patruno, *Strategia digitale 2025, IA e open data: come vincere l'inerzia dell'Italia*, «Agendadigitale.eu», 14 febbraio 2020, <<https://www.agendadigitale.eu/cittadinanza-digitale/strategia-digitale-2025-ia-e-open-data-come-vincere-linerzia-dellitalia>>.

43 Cfr. <<https://data.europa.eu>>.

44 Cfr. <<http://www.datiopen.it>>.

45 Cfr. <<https://data.un.org>>.

46 Cfr. <<http://data.imf.org>>.

47 Cfr. <<https://www.siope.it>>.

48 Cfr. <<https://www.istat.it>>.

49 Cfr. <<https://scistarter.org>>.

cata al miglioramento dell'esperienza scientifica dei cittadini con ricercatori e partner scientifici. Scistarter nasce presso l'University of Pennsylvania, nel 2014, in *partnership* con la School for the Future of Innovation in Society dell'Arizona State University, creando un sistema di gestione e identificazione degli utenti coinvolti attraverso piattaforme e discipline diverse.

Public engagement dei ricercatori

Alcuni atenei e consorzi hanno sviluppato piattaforme *ad hoc* per condividere anche *dataset*, software e metodi di riproducibilità degli studi e delle ricerche nel rispetto dei principi FAIR e delle limitazioni della normativa europea GDPR (*General data protection regulation*).

Per i ricercatori i principali vantaggi dell'archiviazione dei dati della propria ricerca possono essere:

- la conservazione a lungo termine di dati nei formati e con i metadati necessari;
- la condivisione con tutta la rete internet oltre ogni limite geografico;
- la tracciabilità dei dati tramite DOI (*digital object identifier*);
- la visibilità dei contenuti per collaborazioni del settore disciplinare e società civile;
- la possibilità di ricevere *feedback* da specialisti della materia.

I ricercatori, come anche i cittadini, possono reperire dati da diversi *repository* interoperabili indicizzati nel Registry of research data repositories⁵⁰ di DataCite, attivo dal 2012, finanziato dalla German Research Foundation⁵¹, in *partnership* con biblioteche accademiche ed editori.

Nei Paesi Bassi la Technische Universiteit Delft ha dato un ruolo centrale alla figura del *data steward* in ogni facoltà, per supportare tecnicamente i ricercatori, tra cui si distinguono i *data champion*⁵². Si tratta di una comunità che consente un approccio disciplinare aperto ai cambiamenti sistemici in un'ottica di collaborazione con gli *stakeholder* accademici e della società civile.

Esistono diverse piattaforme che agevolano la diffusione e l'interoperabilità dei dati della ricerca; la più nota è Zenodo, scelta nel programma Horizon Europe dalla Commissione europea come una delle piattaforme per il deposito di dati e pubblicazioni.

La piattaforma Zenodo nasce nel 2013, gestita dal Conseil européen pour la recherche nucléaire (CERN)⁵³ per OpenAIRE (Open access infrastructure for research in Europe)⁵⁴ per agevolare l'autoarchiviazione ai ricercatori che dispongano dei diritti di pubblicazione in *open data*. Una seconda piattaforma molto utilizzata per archiviare dati è Figshare, con la possibilità di caricare set di dati, tracciando gli autori tramite ORCID ID e le pubblicazioni tramite DOI, con statistiche di visualizzazione e di condivisione dai *dataset*. In ambito biomedico è molto utilizzata Dryad, per l'accesso gratuito dei dati pubblicati. Il caricamento dai *dataset* è a pagamento, per un costo per autore che parte da 80,00 USD, che include l'assegnazione del DOI, il monitoraggio delle citazioni dei propri lavori, limitati periodi di embargo per i loro dati

50 Cfr. <<http://www.re3data.org>>.

51 Cfr. <<http://www.dfg.de>>.

52 Cfr. Connie Clare, *Writing articles about data champions*. 4 settembre 2019, <<https://zenodo.org/record/3386060>>, DOI: 10.5281/zenodo.3386060.

53 Cfr. <<http://home.web.cern.ch>>.

54 Cfr. <<http://www.openaire.eu/it>>.

nella fase di *peer review*. Come estensione del *social network* omonimo, Mendeley data consente la condivisione di dati aperti della ricerca. Il *repository* aperto e gratuito è nato nel 2015, con la finalità di incrementare l'esposizione delle pubblicazioni e monitorare l'utilizzo dei loro dati.

Anche la Harvard University ha creato un suo *repository* di dati di ricerca: DataVerse⁵⁵ portale *open source* dell'Institute for Quantitative Social Science, per consentire il caricamento e la consultazione di dati di qualsiasi disciplina.

Ogni cittadino può iniziare a interessarsi dei dati della ricerca tramite Research Data Alliance⁵⁶, organizzazione internazionale, i cui membri collaborano allo sviluppo di infrastrutture e attività, volti a ridurre gli ostacoli esistenti alla condivisione e allo scambio di dati, per accelerare l'innovazione derivante dai dati a livello mondiale.

Conclusioni

Questa ricerca è frutto della stretta collaborazione con Annalisa Casino, presidente della cooperativa di servizi Eticae - Stewardship in action, specialista in gestione etica delle risorse per la certificazione internazionale SA 8000, cuore del progetto e cardine per la sostenibilità e responsabilità imprenditoriale. Per il successo nelle organizzazioni basate sull'apertura, la condivisione dei dati è frutto del potere di avviare, facilitare e coordinare in modo proattivo le collaborazioni in modo sistematico, sostenibile e responsabile.

Il mondo delle biblioteche, utilizzando la scienza aperta come leva per lo sviluppo sostenibile, nel prossimo futuro può coinvolgere realtà come l'AIB, la Rete delle università per lo sviluppo sostenibile (RUS), l'IFLA e l'ASVIS (Alleanza italiana per lo sviluppo sostenibile), per proporre al Social Accountability International (SAI) nuove basi per una nuova certificazione di *stewardship* basata su requisiti tecnici, enfatizzando come i dati aperti e condivisi possano favorire il perseguimento all'interno dell'Agenda 2030 per i seguenti traguardi:

- Traguardo 4.6 «assicurarsi che tutti i giovani e una parte sostanziale di adulti, uomini e donne, raggiungano l'alfabetizzazione e l'abilità di calcolo»⁵⁷, nell'Obiettivo 4 "Istruzione di qualità";
- Traguardo 10.2 «entro il 2030, potenziare e promuovere l'inclusione sociale, economica e politica di tutti, a prescindere da età, sesso, disabilità, razza, etnia, origine, religione, status economico o altro»⁵⁸ nell'Obiettivo 10 "Ridurre le disuguaglianze";
- Traguardo 16.6 «sviluppare istituzioni efficaci, responsabili e trasparenti a tutti i livelli»⁵⁹ e Traguardo 16.10 «garantire l'accesso del pubblico alle informazioni e proteggere le libertà fondamentali, in conformità con la legislazione nazionale e con gli accordi internazionali»⁶⁰, entrambi nell'Obiettivo 16 "Pace, giustizia e istituzioni solide";
- Traguardo 17.6 «migliorare la cooperazione Nord-Sud, Sud-Sud e quella triangolare in ambito regionale ed internazionale e l'accesso alla scienza, alla tecnologia e all'innovazione e migliorare la condivisione delle conoscenze sulle condizioni reci-

55 Cfr. <<https://dataverse.harvard.edu>>.

56 Cfr. <<https://www.rd-alliance.org/professionalizing-data-stewardship-rda-community>>.

57 Cfr. UNDP, *The 2030 Agenda* cit., target 4.6, <<https://open.undp.org/sdg/targets/4/6>>.

58 *Ivi*, target 10.2, <<https://open.undp.org/sdg/targets/10/2>>.

59 *Ivi*, target 16.6, <<https://open.undp.org/sdg/targets/16/6>>.

60 *Ivi*, target 16.10, <<https://open.undp.org/sdg/targets/16/10>>.

procamente concordate, anche attraverso un maggiore coordinamento tra i meccanismi esistenti, in particolare a livello delle Nazioni Unite, e attraverso un meccanismo di facilitazione globale per la tecnologia»⁶¹, Traguardo 17.7 «promuovere lo sviluppo, il trasferimento, la disseminazione e la diffusione di tecnologie eco-compatibili ai paesi in via di sviluppo a condizioni favorevoli, anche a condizioni agevolate e preferenziali, come reciprocamente concordato»⁶² e Traguardo 17.9 «rafforzare il sostegno internazionale per l'attuazione di un sistema di costruzione delle capacità efficace e mirato nei paesi in via di sviluppo per sostenere i piani nazionali di attuazione di tutti gli obiettivi di sviluppo sostenibile, anche attraverso la cooperazione nord-sud, sud-sud e triangolare»⁶³, nell'Obiettivo 17 "Partnership per gli obiettivi".

Questo processo di certificazione nel tempo potrebbe realizzare un sistema di monitoraggio che consideri non solo la tecnologia utilizzata, ma anche i processi e i contenuti interoperabili per l'effettiva *stewardship* dei dati, per l'accesso aperto da parte di cittadini e imprese.

Il bibliotecario, con adeguate competenze, potrebbe contribuire a facilitare e supportare nuove collaborazioni, tra cittadini e mondo produttivo, basate sulla disponibilità di informazioni in tempo reale, con elevati livelli di interoperabilità, reperibili liberamente in rete.

Questo approccio collaborativo⁶⁴ può generare sinergie e interazioni nuove, creando *partnership* tra settore privato e dati della ricerca, condividendo metodologie, piattaforme e strumenti che permettano una condivisione produttiva e responsabile dei dati.

La *data stewardship* viene definita come: «l'insieme delle attività per preservare e implementare il contenuto informativo, l'accessibilità e l'usabilità dei dati. Questa concezione della ricerca enfatizza le pratiche e le funzioni volte a creare una supervisione adatta a garantire la qualità e l'uso coerente del prodotto e/o per fornire informazioni a valore aggiunto»⁶⁵.

Il ruolo del *data steward* è trino: principalmente collaborativo con gli *stakeholder*, in secondo luogo è protettivo dei dati sensibili, in ultima istanza proattivo verso potenziali interazioni e sviluppi legati ai traguardi raggiungibili in base al campo di azione.

61 *Ivi*, target 17.6, <<https://open.undp.org/sdg/targets/17/6>>.

62 *Ivi*, target 17.7, <<https://open.undp.org/sdg/targets/17/7>>.

63 *Ivi*, target 17.9, <<https://open.undp.org/sdg/targets/17/9>>.

64 Cfr. (Re-)Defining the roles and responsibilities of data stewards for an age of data collaboration: research, methods and good practice, <<http://www.data4sdgs.org/resources/re-defining-roles-and-responsibilities-data-stewards-age-data-collaboration>>.

65 Ge Peng [et al.], *Scientific stewardship in the open data and big data era: roles and responsibilities of stewards and other major product stakeholders*, «D-lib magazine», 22 (2016), n. 5-6, <<http://www.dlib.org/dlib/may16/peng/05peng.html>>, DOI: 10.1045/may2016-peng.

Articolo proposto il 13 giugno 2020 e accettato il 29 settembre 2020.

ABSTRACT AIB studi, vol. 60 n. 2 (maggio/agosto 2020), p. 311-323. DOI 10.2426/aibstudi-12123
ISSN: 2280-9112, E-ISSN:2239-6152

DAMIANO ORRÙ, Università degli studi di Roma "Tor Vergata", Facoltà di Economia, Biblioteca "Vilfredo Pareto", e-mail orru@economia.uniroma2.it.

Open data steward: bibliotecari e alfabetizzazione ai dati aperti

I *data steward* bibliotecari possono proporre un aggiornamento dello standard SA 8000 integrando come requisito trasversale la '*data stewardship* per la scienza aperta' basata sui principi dei *FAIR data* e i dettami del GDPR per gli open data, perseguendo i traguardi dell'Agenda 2030 delle Nazioni unite: Traguardo 4.6 "Assicurarsi che tutti i giovani e una parte sostanziale di adulti, uomini e donne, raggiungano l'alfabetizzazione e l'abilità di calcolo"; Traguardo 16.6 "Sviluppare istituzioni efficaci, responsabili e trasparenti a tutti i livelli"; Traguardo 16.10 "Garantire l'accesso del pubblico alle informazioni e proteggere le libertà fondamentali, in conformità con la legislazione nazionale e con gli accordi internazionali".

La *data stewardship* in ambito scientifico oltrepassa la semplice accessibilità e usabilità del dato, puntando a nuovi livelli di interoperabilità con nuovi sistemi e piattaforme.

Nell'evoluzione di *app* e strumenti legati all'IoT (*internet of things* o internet delle cose) sono determinanti algoritmi alimentati dai dati. Si parla quindi di IoD (*internet of data*) come fonte per lo sviluppo sperimentale di applicazioni integrate e innovative nel campo ambientale, sanitario, energetico, con implementazioni anche nelle *smart city*. Per una corretta e trasparente gestione del flusso informativo microsistemico e macrosistemico diviene indispensabile condividere e aggregare fonti diverse e interoperabili.

Il bibliotecario come *data steward* ha il compito di promuovere il cambiamento culturale verso la ricerca condivisa, agendo su chi produce e su chi riusa i dati. Per questo obiettivo serve agevolare il trasferimento tecnologico necessario per la scienza aperta, tramite la sinergia tra *stakeholder* e produttori di dati.

Open data steward: librarians and open data literacy

Librarian data stewards can propose an update of the SA 8000 standard by integrating as a cross-cutting requirement the data stewardship for open science based on the European FAIR data guidelines and the GDPR directives for open data, pursuing the Agenda 2030 targets: Target 4.6 "Ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy"; Target 16.6 "Develop effective, accountable and transparent institutions at all levels"; Target 16.10 "Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements".

In the evolution of IoT (internet of things) related apps and tools, data driven algorithms are crucial. We therefore speak of IoD (internet of data) as a source for the experimental development of integrated and innovative applications in the environmental and energy services field, the healthcare sector, and also in the context of Smart Cities. For a correct and transparent management of the micro and macrosystemic information flow it becomes essential to share and aggregate different and interoperable sources.

The librarian as data steward has the task of promoting cultural change towards shared research, acting on those who produce and reuse data. To this end, it is necessary to facilitate the technology transfer necessary for open science, through the synergy between stakeholders and data producers.