

SUPSI

LAVORO DI DIPLOMA DI

FRANCESCA TORTI

DIPLOMA OF ADVANCED STUDIES
IN BIBLIOTECHE E SCIENZE DELL'INFORMAZIONE

ANNO ACCADEMICO 2019/2020

OPEN DATA E BIBLIOTECHE

RELATRICE

ORNELLA MONTI-JAUCH

Abstract

Con l'avvento dell'era digitale, il mondo delle biblioteche si è trovato di fronte ad un nuovo, enorme tema: l'Open Science e la sua gestione.

Da diversi anni ormai il movimento che sostiene la promozione di Open Science a livello globale ha interessato anche biblioteche e archivi, ponendo grandi interrogativi su come la figura del bibliotecario dovrà cambiare per rispondere ai bisogni della digitalizzazione e della promulgazione della Scienza Aperta.

L'Open Science si propone infatti di rendere accessibile a chiunque sia gli articoli scientifici che tradizionalmente si trovano su riviste scientifiche spesso molto costose (Open Access), sia tutto l'insieme dei dati che sono stati raccolti e studiati dai ricercatori negli atenei (Open Data).

In questo lavoro di diploma viene dapprima descritto il concetto di Open Science sia da un punto di vista formale sia da un punto di vista tecnico e, nella seconda parte, si illustra nel dettaglio quali saranno le sfide e i compiti dei bibliotecari nell'ottica dell'Open Data.

Sommario

Introduzione.....	p. 3
1. Open Science.....	p. 4
2. Open Data.....	p. 7
3. E-Government.....	p. 10
4. Le convenzioni europee: openAIRE.....	p. 12
5. Un illustre esempio di database “made in Switzerland”: ZENODO.....	p. 13
6. Un esempio concreto dell’utilizzo di Zenodo: il progetto enCompass.....	p. 15
7. Un esempio di dataset enorme: l’Open Data al CERN.....	p. 16
8. Altri aspetti tecnici per rendere i dati aperti.....	p. 17
9. Il ruolo delle biblioteche universitarie.....	p. 18
10. Incoraggiare la produzione di dati aperti: le piattaforme collaborative.....	p. 20
11. Formazione all’utilizzo dell’Open Data: alcuni spunti.....	p. 21
Conclusione.....	p. 24
Bibliografia e sitografia.....	p. 26

Introduzione

L'obiettivo di questo lavoro di diploma è quello di offrire una panoramica sul movimento, ancora in evoluzione, dell'Open Science ed in particolare dell'aspetto dell'Open Data.

Questo movimento sta prendendo sempre più piede nel panorama della ricerca universitaria, ma non solo: l'Open Science coinvolge la vita del cittadino non solo da un punto di vista prettamente scientifico, nel senso stretto del termine, ma anche della società – fanno parte dell'Open Science anche grandi progetti come l'E-Government, l'Open Source, l'Open Education e la Citizen Science.

Questo documento non ha chiaramente pretese di esaustività: l'argomento è infatti molto vasto e non è stato possibile discutere in questa sede di importanti questioni come le licenze, le basi legali, il diritto alla privacy o il copyright. Per lo stesso motivo, il panorama internazionale è menzionato solo in parte, per approfondire piuttosto il discorso europeo e svizzero. Non è stato perciò possibile entrare nel dettaglio di tutte le piattaforme Open Data disponibili in Svizzera e portare come esempio le prospettive di molte altre università oltre a quelle prese in considerazione. Un'ulteriore difficoltà riscontrata nella stesura di questo lavoro di diploma è stata quella di esporre in termini semplici e lineari aspetti tecnici e specifici normalmente descritti tramite un gergo specialistico, al fine di rendere il testo comprensibile anche a chi non ha mai avuto a che fare con il concetto di Open Data.

Il lavoro è dunque una panoramica iniziale che si propone di far luce sul significato e l'importanza dell'Open Science e dell'Open Data, con un accento alle buone pratiche dal punto di vista tecnico.

Infine, ma non meno importante, viene discusso il ruolo del bibliotecario e delle biblioteche universitarie all'interno del movimento – molto spesso infatti quando si parla di Open Data in ambito biblioteconomico la domanda principale dei bibliotecari è “noi cosa dobbiamo fare?”.

L'ultima parte del presente lavoro risponde, si spera, almeno in parte, a questo grande quesito.

1. Open Science

Il concetto di Open Science si basa su un semplice assunto, riportato sulla pagina SUPSI ad esso dedicata:

“Open Science indica tutte le azioni finalizzate a rendere i risultati e i processi del lavoro all'interno del mondo accademico più aperti. Open Science include Open Access, Open Data, Open Education, Open Innovation, Open Software, Open Hardware e Citizen Science.” (SUPSI Open Science, 2020)

In altre parole, Open Science si prefigge di rendere pubblici non solo i risultati (quindi gli articoli scientifici o i paper) ma anche ogni passo della ricerca come statistiche, dati raccolti, appunti di laboratorio, risultati computazionali e via dicendo, nel modo più fruibile possibile. In questo modo i dati sono sia riutilizzabili che verificabili in modo più esteso rispetto, ad esempio, alla classica peer review utilizzata dalle case editrici scientifiche.

Naturalmente, in Open Science rientrano tutte le discipline accademiche: sia quelle scientifiche sia quelle umanistiche.

Inoltre, non si limita a voler dialogare solo con i professionisti della ricerca, ma coinvolge diversi attori: uffici di supporto alla ricerca, enti pubblici, politici, istituzioni scientifiche, istituzioni universitarie e, non da ultime, le biblioteche.

Per capire di cosa parliamo quando usiamo il termine Open Data, è utile partire dal concetto e dallo sviluppo dell' Open Access riferito alle pubblicazioni scientifiche. Open Access e Open Data, infatti, sono le due grandi aree che costituiscono Open Science stessa.

Nella fattispecie, nel 2015 la SEFRI (la Segreteria di Stato per la ricerca, la formazione e l'innovazione) ha affidato a Swissuniversities il compito di elaborare una strategia d'azione a lungo termine riguardo all'accesso aperto (Open Access) alle pubblicazioni ed ai dati scientifici (Open Data). Questo perché l'Open Science non è meramente un obiettivo federale, ma si colloca nel contesto non solo europeo ma internazionale. Le università e i loro ricercatori devono, in poche parole, poter accedere liberamente ai risultati scientifici di altre università a livello globale:

“I risultati della ricerca basata su finanziamenti pubblici costituiscono un bene pubblico che può essere pienamente sfruttato soltanto se tutti hanno libero accesso ad esso senza restrizioni.” (“Strategia nazionale svizzera sull'accesso aperto”, 2020)

Nel documento “Strategia nazionale svizzera sull'accesso aperto” (2020), Swissuniversity ha infatti definito i seguenti punti:

- Un approccio potente e unificato

A fronte delle dimensioni ridotte della Svizzera, del suo sistema educativo e di ricerca estremamente decentrata, tutti gli stakeholder, i politici, le istituzioni di formazione superiore (con le rispettive biblioteche) e i finanziatori hanno unito le proprie forze per perseguire obiettivi comuni. Qualsiasi transizione richiederà tuttavia una forte volontà politica per promuovere l'OA a partire dai vertici delle istituzioni di ricerca, dagli attori politici di alto livello e dalle organizzazioni finanziatrici. Dovrebbero essere prese in considerazione anche delle alternative di collaborazione internazionale.

- Supporto e impegno delle comunità di ricerca

Una solida collaborazione di questo tipo richiede un grande senso di responsabilità da parte di tutto gli stakeholder, che devono fare la loro parte per assicurare il successo della strategia. L'impegno nel campo dell'OA e la consapevolezza delle comunità di ricerca sono essenziali ai fini di un cambiamento.

- Trasparenza e neutralità dei costi

Nel medio e nel lungo termine, l'OA non dovrebbe portare a un aumento delle spese di pubblicazione (neutralità dei costi) per la Svizzera nel suo insieme, anche se durante la fase di transizione vi saranno costi aggiuntivi. Un prerequisito per l'implementazione dell'OA è costituito da una panoramica completa e trasparente dei costi da parte delle organizzazioni di ricerca e finanziatrici. Fino a quando le comunità scientifiche continueranno a permettere alle grandi case editrici orientate ai profitti di dominare il settore delle pubblicazioni scientifiche, i vantaggi dell'Open Access saranno minacciati dagli alti prezzi delle pubblicazioni – non più per le licenze, ma per le spese di pubblicazione OA.

- Assicurare il controllo e la diversità del processo di produzione scientifica

L'attuale concentrazione del mercato e gli aumenti dei prezzi costituiscono ostacoli significativi a un'attività di pubblicazione equa, aperta, efficiente e finanziariamente sostenibile. Questa situazione richiede l'adozione di una strategia articolata. Occorre promuovere nuove e promettenti forme di pubblicazione, comprese quelle che vedono coinvolti gli stakeholder.

- Revisione del sistema di misurazione della qualità

Molti degli attuali sistemi di misurazione della qualità accademica, che dipendono in misura significativa da poche riviste, sono perlopiù considerati inaccurati e limitati (cfr. dichiarazione DORA) e andrebbero riconsiderati. In effetti, come già menzionato più sopra, i processi di valutazione e il guadagno in termini di reputazione determinano dove gli scienziati pubblicano i propri lavori. Inoltre, i criteri su cui si basano dovrebbero includere anche l'Open Access e l'Open Science. Tuttavia, anche nuovi sistemi di valutazione possono rivelarsi efficaci unicamente se sono supportati dalla comunità di ricerca. (“Strategia nazionale svizzera sull'accesso aperto”, 2020)

Il bisogno di costituire una rete ed una filosofia Open Science scaturisce da diversi fattori: come detto, il potenziale della ricerca può essere sfruttato pienamente solo se chiunque ha accesso ai risultati ed ai dati; d'altro canto, l'urgenza di implementare i sistemi Open Access nasce da un bisogno marcatamente economico.

Nello stesso documento, viene inoltre specificato il motivo per il quale si è sentito il bisogno di stabilire delle regole precise nel panorama Open Access:

“Le iniziative di OA sono state lanciate soprattutto come reazione delle comunità scientifiche agli sviluppi insostenibili e onerosi nel panorama delle pubblicazioni scientifiche. Nel corso degli ultimi decenni, le spese di abbonamento sono esplose e le case editrici orientate al profitto hanno svolto un ruolo dominante nella pubblicazione e nella divulgazione delle opere scientifiche . La loro posizione è basata sul fatto che i ricercatori ne apprezzano i contenuti, lavorano per loro come autori, revisori ed editori e spesso si sentono obbligati a pubblicare i loro lavori con loro – il tutto basandosi sugli attuali

meccanismi di valutazione e di reputazione – allo scopo di mantenere nel tempo opportunità di carriera ottimali. Di conseguenza, la maggior parte della ricerca finanziata pubblicamente rimane chiusa dietro a un paywall.” (“Strategia nazionale svizzera sull'accesso aperto”, 2020)

Notiamo quindi una discrepanza fondamentale tra i mandati di ricerca universitari sovvenzionati da fondi governativi ed il tentativo da parte delle case editrici scientifiche di lucrare sulla pubblicazione e la sua disseminazione.

2. Open Data

Il progetto FOSTER, una piattaforma di e-learning dedicata alla scienza aperta, definisce l'Open Data nel modo seguente:

“Open Data are online, free of cost, accessible data that can be used, reused and distributed provided that the data source is attributed and shared alike.” (FOSTER, 2020)

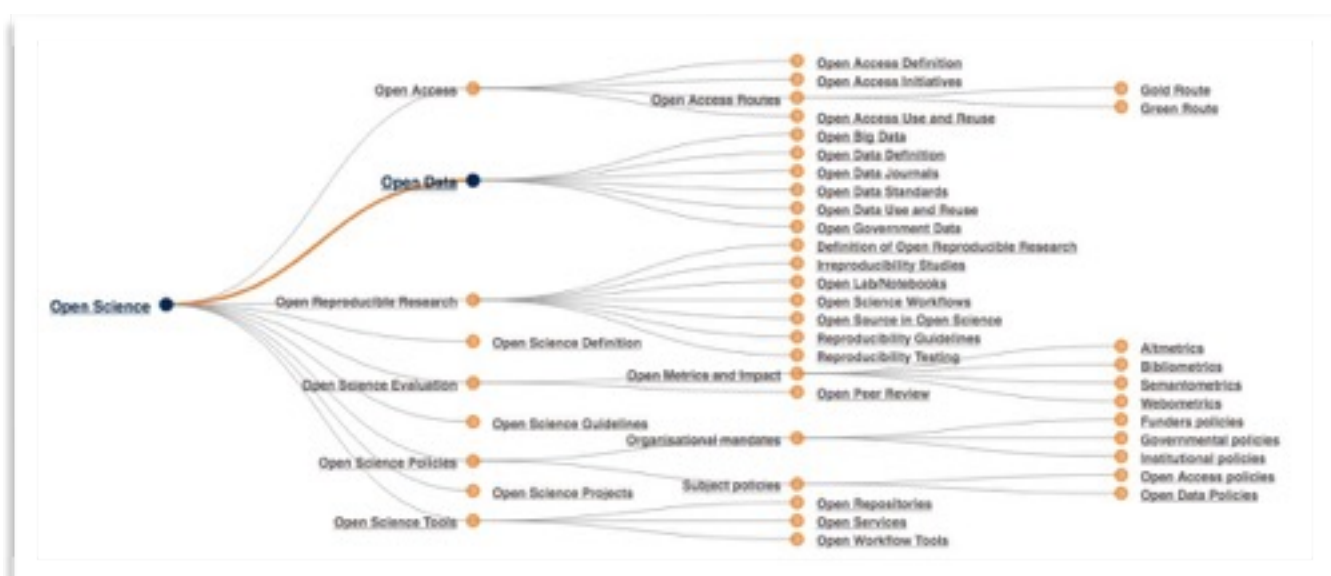


Immagine: FOSTER 2020

Foster è un importante punto di riferimento per comprendere le risorse disponibili attualmente nel panorama Open Science: si tratta infatti di un progetto finanziato dall'Unione Europea: a partire dal

2020 e per una durata di due anni, 11 partner provenienti da 6 paesi lavoreranno congiuntamente al fine di stabilire un set di pratiche e linee guida per rendere l'Open Science una pratica perfettamente definita tra i ricercatori europei.

Finora ci siamo occupati prettamente di articoli e pubblicazioni scientifiche. Veniamo dunque al concetto di Open Data.

L'Open Data riguarda una mole di dati, che specialmente nelle discipline STEM (Science, Technology, Engineering, and Mathematics) accompagnano articoli o pubblicazioni, custoditi su grandi database informatici o repository istituzionali. Proprio come succede per le riviste scientifiche, spesso i dati non sono liberamente accessibili ma vengono protetti in determinati modi: limitazione dei permessi di accesso ai database e repository, pagamento di quote per l'accesso e la consultazione, limitazioni alle licenze o ai diritti d'autore etc.

Open Data, quindi, si prefigge non solo di abbattere i costi legati alle pubblicazioni scientifiche ma anche e soprattutto di migliorare la ricerca, e quindi la scienza, tramite la condivisione dei dati.

Immaginiamo ad esempio una ricerca di laboratorio per creare un vaccino che curi il virus SARS-COV-2: se un laboratorio in una certa parte del mondo inizia a condurre dei test di un probabile vaccino, condividendo i dati di laboratorio altri ricercatori potranno non solo verificare che questi test siano attendibili ma ricrearli e dare un contributo concreto nell'avanzamento della ricerca e dello sviluppo di una cura. Si tratta di un esempio di collaborazione estremamente reale, i cui benefici sono evidenti.

Proprio come succede per le pubblicazioni scientifiche, il concetto di Open Data rigetta l'idea di restringere l'accessibilità ai dati:

“Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.” (FOSTER, 2020)

I dati devono quindi essere trattati come veri e propri beni comuni.

Entrando più in dettaglio, gli aspetti di maggiore importanza nel concetto di Open Data sono i seguenti:

- disponibilità e accesso: i dati devono essere disponibili nel loro complesso, per un prezzo non superiore ad un ragionevole costo di riproduzione, preferibilmente mediante scaricamento da Internet. I dati devono essere disponibili in un formato utile e modificabile;
- riutilizzo e redistribuzione: i dati devono essere forniti a condizioni tali da permetterne il riutilizzo e la redistribuzione. Ciò comprende la possibilità di combinarli con altre basi di dati;
- partecipazione universale: tutti devono essere in grado di usare, riutilizzare e redistribuire i dati. Non ci devono essere discriminazioni né di ambito di iniziativa né contro soggetti o gruppi. Ad esempio, la clausola ‘non commerciale’, che vieta l’uso a fini commerciali o restringe l’utilizzo solo per determinati scopi (es. quello educativo) non è ammessa. (Open Data Handbook, n.d.)

Un altro aspetto fondamentale è l'interoperabilità dei dati, così definita:

“L’interoperabilità è la capacità di diversi sistemi e organizzazioni di lavorare insieme (Inter-operare). In questo caso, è la capacità di combinare una base di dati con altre.

L’interoperabilità è importante perché permette a componenti diverse di lavorare insieme. L’abilità di rendere ciascun dato un componente e di combinare insieme vari componenti è essenziale per la costruzione di sistemi sofisticati.” (Open Data Handbook, n.d.)

È interessante soffermarsi anche sull'aspetto etico e sociologico del movimento Open Data.

Esso nasce dal bisogno di condivisione delle scoperte e del progresso, ed il suo fine ultimo e più alto è il raggiungimento di una vera e propria “knowledge society” (società della conoscenza), concetto già presente nella sociologia di inizio '900 . Negli ultimi vent'anni questo concetto ha preso una forma sempre più concreta, sebbene non sia ancora realizzato, ma già attorno al 1967 i sociologi Berger e Luckmann affrontavano la questione:

“The way that humans understand the world (...) involves learning, and this learning involves institutions and sets of established actions that are created through social action. Berger and Luckmann's (1967) argument is focused on the way that the social world is shaped by human action, yet also appears **objective and fact-like**. In forming this proposition, they show us that learning about the social world, and what is perceived as

nature, requires actors to 'go out' and learn, and that this learning is then embedded within actions, typifications, and institutions. **This process of knowing creates and results in data, information and knowledge** in an array of forms. What Berger and Luckmann (ibid.) demonstrate is that the world gains shape, becomes ordered and is made understandable when data, information and knowledge are coded in particular ways. There is, therefore, a relationship between what we know about the world, how we find out about the world and the ways in which we structure the data, information and knowledge.” (Wessels et al., 2017)

Ancora una volta è fondamentale sottolineare come gli Open Data riguardino il mondo della ricerca accademica *in toto*: è facile lasciarsi prendere dall'impressione che i dati di ricerca riguardino solo le discipline scientifiche in senso stretto (ovvero quelle discipline racchiuse nel gruppo STEM: Science, Technology, Engineering, Mathematics). Può darsi che l'equivoco nasca dal fatto che l'applicazione dei dati risultanti da ricerche STEM abbiano un'applicazione apparentemente più immediata rispetto, ad esempio, a quelli delle discipline definite “umanistiche” - pensiamo ai progressi della scienza medica o farmaceutica. In realtà, discipline come il diritto o le scienze politiche sono egualmente fondamentali nel costruire una società della conoscenza.

L'esempio che affronteremo di seguito è pertanto quello dell'E-Government.

3. E-Government

Partiamo da un semplice presupposto, presentato come introduzione alla Dichiarazione Universale dei Diritti Umani delle Nazioni Unite (10 dicembre 1948) sul sito dell'Amministrazione Federale:

“Gli archivi giocano un ruolo importante nella protezione dei diritti umani. Per poter perseguire le violazioni dei diritti umani, le testimonianze sono imprescindibili. Gli archivi proteggono e rendono accessibili atti e documenti. In tal modo contribuiscono alla ricerca della verità e costituiscono la base per l'amministrazione della giustizia e i risarcimenti. Oltre a ciò formano parte della memoria collettiva, che è di importanza centrale anche per tenere traccia delle trasgressioni dei diritti umani. Dagli anni '90 il diritto di accesso agli archivi ha assunto sempre maggiore importanza, in particolare nei paesi in fase di transizione verso la democrazia. I documenti degli archivi diventano così un mezzo

essenziale per la protezione dei diritti umani e la ricostruzione degli eventi politici e storici.” (Archivio Federale Svizzero [AFS], n.d.)

Si deduce, quindi, che l'accesso all'informazione ed ai relativi dati è di per sé un diritto umano. In effetti, l'articolo 27, capoverso 1, della Dichiarazione Universale dei Diritti Umani delle Nazioni Unite recita:

“Ogni individuo ha diritto di prendere parte liberamente alla vita culturale della comunità, di godere delle arti e di **partecipare al progresso scientifico ed ai suoi benefici.**” (Dichiarazione Universale dei Diritti Umani, 1948)

Su questa base, l'Amministrazione Federale Svizzera ha elaborato nel novembre 2018 la “Strategia sul libero accesso ai dati pubblici in Svizzera 2019-2023”, creando anche il portale opendata.swiss. Questo il mandato:

“Pubblicare i dati dell'amministrazione aperti e liberamente accessibili favorisce la trasparenza, la collaborazione e l'innovazione in tutti gli ambiti della società. Per garantire ciò in modo duraturo e sicuro, tutti i dati resi pubblici dai servizi federali sono progressivamente pubblicati sul portale opendata.swiss in formati aperti, gratuiti e leggibili elettronicamente («Open Government Data»). Inoltre, verranno messi a disposizione del pubblico anche altri dati già esistenti, ad esempio riferiti ai Cantoni, ai Comuni, alle città e alle imprese parastatali. Tuttavia, occorre tener conto del fatto che questo principio non sarà applicato nel caso in cui un interesse legittimo preponderante vi si opponesse, per esempio per ragioni di protezione di dati.” (opendata.swiss, n.d.)

4. Le convenzioni europee: OpenAIRE

Come abbiamo detto, i movimenti Open Science non sono limitati ai confini nazionali ma coinvolgono più nazioni.

Nel panorama europeo, il progetto openAIRE, che nel 2020 festeggia il suo decimo anno di attività, costituisce il punto di riferimento per questi movimenti, offrendo un network di repositories, archivi e journal Open Access (EOSC – European Open Science Cloud).

Il progetto svizzero per la promozione ed il sostegno dei dati open si colloca all'interno di questo quadro europeo che definisce così il bisogno di promuovere Open Science secondo le direttive del Consiglio Europeo (EC), organo collettivo che definisce le priorità politiche e legislative all'interno dell'Unione Europea:

“The European Open Science Cloud (EOSC) is grounded in the EC’s aim to promote the access and reuse of research data which comes out of publicly funded research. At present, there is fragmented access to research data, which exists, stored and is created in many different data centres, institutions and research centres across Europe. Open access to this data is not a given, and the content is not interoperable, restricting inter-disciplinary research. EOSC will solve this problem by providing easy access to this data, making publicly funded data open. It will provide one single point of free access, ensuring all databases are interoperable. The push therefore for a recognition of the benefits of Open Science - policies and infrastructures - is therefore key.” (OpenAIRE, 2020)

Il progetto openAIRE offre anche un repository dedicato, Zenodo, di cui parleremo più avanti.

Le basi legali sulle quali openAIRE si fonda riguardano la risoluzione del 16 febbraio 2017 del Parlamento Europeo, consultabile qui: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0052_EN.html?redirect (European Parliament resolution of 16 February 2017 on the European Cloud Initiative).

5. Un illustre esempio di database “made in Switzerland”: Zenodo

Zenodo è stato creato dai ricercatori del CERN nel 2013 per sostenere e promuovere l'Open Science, ed è un prodotto di openAIRE. È stato infatti commissionato dall'Unione Europea nell'ambito del progetto openAIRE. Zenodo offre dei database con capacità molto elevate dove è possibile caricare i propri dati di ricerca e offre avanzati tool per il Big Data Management. Nonostante sia stato sviluppato dal CERN, Zenodo è aperto ad ogni ambito disciplinare a livello mondiale. Di fatto, Zenodo dà la possibilità ai ricercatori di conservare in un unico luogo sicuro tutti i passi della ricerca di modo da poterli rendere pubblici una volta pubblicati i risultati:

“Publication may happen months or years after completion of the research, so collecting together all the research artefacts at that stage to publish openly is often challenging. Zenodo therefore offers the possibility to house closed and restricted content, so that artefacts can be captured and stored safely whilst the research is ongoing, such that nothing is missing when they are openly shared later in the research workflow.” (Zenodo, n.d.)

Quindi, Zenodo risolve il problema dei repository istituzionali che spesso offrono spazio limitato sui propri server. Il codice di programmazione di Zenodo, inoltre, è anch'esso Open Source: questo garantisce i principi fondamentali dell'Open Science per la collaborazione ed il miglioramento del programma stesso.

Non ci sono inoltre limiti sui formati o i tipi di dati caricati – Zenodo compie un passo avanti nel definire i dati “artefatti” e incoraggia i ricercatori a depositare nel database qualsiasi materiale digitale utile alla ricerca:

“To fully understand and reproduce research performed by others, it is necessary to have all the details. In the digital age, that means all the digital artefacts, which are all welcomed in Zenodo.

To be an effective catch -all, that eliminates barriers to adopting data sharing practices, Zenodo does not impose any requirements on format, size, access restrictions or licence.

Quite literally we wish there to be no reason for researchers not to share!

Data, software and other artefacts in support of publications may be the core, but equally welcome are the materials associated with the conferences, projects or the institutions themselves, all of which are necessary to understand the scholarly process.” (Zenodo, n.d.)

Da un punto di vista tecnico, è interessante notare che i dati di Zenodo sono ospitati su server ad alta performance progettati per venire incontro alla gestione dell'enorme quantità di dati prodotti da progetti di fisica come ad esempio gli esperimenti ATLAS, ALICE, CMS e LHCb raccolti nel CERN Data Centre che produce e conserva all'incirca 30 petabyte di dati all'anno (per farci un'idea, corrispondono a 1.2 milioni di dischi Blu-Ray).

Zenodo utilizza il software Invenio, un software open source che può essere modificato a seconda dei bisogni (Zenodo ha in effetti già aggiunto nuove parti di codice per migliorare la performance tecnica); Invenio è utilizzato al di fuori del CERN per la gestione di repository istituzionali da numerosi atenei, come ad esempio il Politecnico Federale di Losanna (EPFL). Nel 2019 è stato creato InvenioRDM, un pacchetto open source che semplifica l'uso del codice aggiuntivo Zenodo per la creazione di repository istituzionali.

Nel 2019 è stata anche annunciata una collaborazione con il repository Dryad. Questa collaborazione porterà ad un Data Management Flow più efficiente.

Si tratta quindi di uno dei migliori esempi di repository che va incontro alle esigenze dei ricercatori.

Le caratteristiche principali di Zenodo sono infatti:

Safe — your research is stored safely for the future in CERN’s Data Centre for as long as CERN exists.

Trusted — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.

Citeable — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.

No waiting time — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.

Open or closed — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.

Versioning — Easily update your dataset with our versioning feature.

GitHub integration — Easily preserve your GitHub repository in Zenodo.

Usage statistics — All uploads display standards compliant usage statistics

(Zenodo, n.d.)

6. Un esempio concreto dell'utilizzo di Zenodo: il progetto enCompass

L'Istituto Sostenibilità Applicata all'Ambiente Costruito (ISAAC) del Dipartimento Ambiente, Costruzioni e Design (DACD) della SUPSI, in collaborazione con alcuni ricercatori del Dipartimento Tecnologie Innovative (DTI), sta lavorando ad enCompass (<http://www.encompass-project.eu/>) un importante progetto finanziato da Horizon2020, il più grande programma di finanziamento per la ricerca istituito a livello europeo. Ecco la descrizione del progetto:

“Il progetto europeo enCOMPASS sta sviluppando un sistema basato su un’app per smartphone che legge i consumi energetici degli edifici e fornisce raccomandazioni e consigli personalizzati su come ridurli, senza compromettere il comfort abitativo. L’app enCOMPASS sfrutta la rete dei contatori intelligenti (smart meter) già installati da alcune aziende elettriche e si appoggia anche a una serie di sensori, installati appositamente negli edifici che partecipano al progetto, che rilevano la presenza delle persone nei locali e leggono automaticamente temperatura e umidità. I dati rilevati sono analizzati in tempo (quasi) reale, in modo da fornire agli utenti, sul proprio smartphone, informazioni semplici e comprensibili sulle proprie abitudini di consumo e suggerimenti su come ridurli.” (enCompass Project, n.d.)

Tre esperimenti pilota sono stati avviati nel 2018 in Germania, Grecia e Svizzera – in particolare, l'esperimento svizzero si tiene su territorio ticinese e coinvolge 100 famiglie selezionate secondo criteri statistici nel territorio di Contone.

Dei sensori sviluppati dai ricercatori ISAAC e DTI sono stati installati nelle case delle 100 famiglie partecipanti al progetto, che a loro volta hanno potuto testare una apposita app sui propri smartphone.

L'insieme dei dati generati da questo progetto (il dataset) è stato pubblicato il 20 aprile 2020 proprio su Zenodo. Sulla pagina del progetto, all'indirizzo <https://zenodo.org/record/3757899#.XyhW5xMzZE5> , è possibile visualizzare e scaricare l'intero dataset della ricerca, suddiviso per esperimento (Germania, Grecia e Svizzera) e per tipo di dato.

Essendo un progetto finanziato da Horizon2020, il dataset è automaticamente indicizzato anche in openAIRE.

A colpo d'occhio è inoltre possibile individuare il DOI assegnato al dataset ed altre informazioni importanti, come ad esempio da chi è stato finanziato il progetto.

Sono inoltre presenti i collegamenti (tramite DOI) agli articoli scientifici e agli atti dei congressi correlati al dataset.

7. Un esempio di dataset enorme: l'Open Data al CERN

Come abbiamo visto, Zenodo è un prodotto congiunto sviluppato dal CERN di Ginevra e adottato da openAIRE. Evidentemente, il CERN stesso utilizza questa piattaforma per la messa a disposizione dei propri dataset.

Se per il progetto enCompass parliamo di poche centinaia di megabyte, il discorso per il CERN e i suoi progetti è piuttosto diverso.

Al momento attuale, l'Open Data del CERN ospita e mette a disposizione in maniera totalmente open più di due petabyte di dati. Questa mole enorme di dati comprende dataset di dati grezzi (ad esempio i dati raccolti durante le collisioni dell'LHC), vari software open source, ambienti informatici specifici, documentazione di ricerca etc.

In questo caso specifico, è stato creato un portale dedicato ai dataset del CERN con una funzione di ricerca estremamente semplice (ricorda molto Google) e con collegamenti e spunti per trovare, esplorare ed utilizzare i dati. Naturalmente, pur trattandosi di un portale dedicato, l'intero Data Management è basato su Zenodo ed è ospitato sugli stessi server su cui sono ospitati tutti gli altri progetti. Ciò dimostra che Zenodo offre soluzioni sia per istituzioni e progetti di ricerca di dimensioni enormi, sia per progetti di ricerca molto piccoli – garantendo però la medesima qualità e la medesima tecnologia utile a trovare efficacemente i dati.

Il portale è consultabile al seguente link: <http://opendata.cern.ch/>

8. Altri aspetti tecnici per rendere i dati aperti

Rendere i dati aperti nel panorama Open Science è tutt'altro che semplice, sia dal punto di vista tecnico sia dal punto di vista legale.

In questa sede non è possibile soffermarsi dettagliatamente su ciascuno di questi aspetti, ma in generale possiamo riassumere alcuni requirements fondamentali per la corretta gestione dell'Open Data:

- disponibilità (availability): i dati devono essere disponibili in maniera gratuita;
- insieme di dati (in bulk data): i dati devono essere disponibili sottoforma di “insieme completo” o “pacchetto”. L'accesso ai dati deve comunque essere diretto;
- formato open e “machine readable”: il ri-utilizzo dei dati non deve essere soggetto a restrizioni di brevetto. Inoltre, i dati devono essere forniti in forma “machine-readable”, ovvero sottoforma di file facilmente scaricabili e visionabili da qualsiasi macchina;
- rilasciare dati grezzi: il rilascio di dati grezzi in tempo reale è preferibile al rilascio di dati perfezionati dopo mesi o anni;
- individuabilità: i dati devono essere facilmente individuabili. Per questo motivo sono da preferire piattaforme di alto livello e con un workflow ben funzionante (ad esempio Zenodo).

È importante inoltre menzionare che ogni Paese regola il proprio DMP (Data Management Plan) per determinare quale deve essere, ai sensi di legge, il “life cycle” dei dati di ricerca. Ciò significa che il modo in cui i dati devono essere generati, raccolti, documentati, condivisi e preservati non è casuale ma rientra in una prospettiva a lungo termine. In Svizzera, la SNSF (Swiss National Science Foundation) offre delle guide complete affinché i ricercatori possano soddisfare i criteri richiesti per pubblicare in Open Data. Tutti i ricercatori che svolgono ricerche finanziate da fondi pubblici sono tenuti ad attenersi al DMP di riferimento del proprio Paese.

9. Il ruolo delle biblioteche universitarie

Le biblioteche sono **facilitatori** nel concetto di Open Data.

Se il bibliotecario deve infatti già analizzare l'impact factor delle riviste scientifiche in abbonamento, egli dovrà assolutamente assumere anche il ruolo di supporto alla ricerca; il suo è quindi un ruolo che comprende da una parte la biblioteconomia classica e dall'altra le scienze dell'informazione:

“L'impact factor è tenuto in massimo conto dai bibliotecari scientifici nella selezione delle fonti da sottoscrivere. Tra molte cose destinate probabilmente a cambiare, la biblioteca si è già trasformata diventando un luogo virtuale. Le collezioni di libri e riviste sono tutt'al più mal sopportate dalle aziende sanitarie e dagli istituti di ricerca. Il grande cambiamento, però, è soprattutto nel ruolo del bibliotecario: quanto dobbiamo aspettare perché si diffonda e sia accettata la figura del documentalista a supporto delle attività di ricerca?” (Addis & De Fiore, 2019)

In quest'ottica, quindi, due sono gli aspetti fondamentali del lavoro bibliotecario in relazione agli Open Data, ovvero garantire che essi siano:

“ - **reperibili**: La prima cosa da fare per rendere i dati e i metadati riutilizzabili è renderli rintracciabili facilitandone la ricerca sia per gli esseri umani che per i computer. Il recupero automatico e affidabile di set di dati e servizi dipende dagli identificatori persistenti (PID) e dai metadati leggibili dalle macchine.

- **accessibili**: I (meta)dati devono poter essere recuperati attraverso il loro identificatore utilizzando un protocollo di comunicazione standardizzato e aperto, che includa eventualmente dei sistemi di autenticazione e autorizzazione. Inoltre, i metadati dovrebbero essere resi disponibili anche quando i dati non lo sono più.” (Il manuale per formatori della Scienza Aperta, 2019)

Andando ancora un po' nello specifico, il “Manuale per formatori della Scienza Aperta” (2019) offre tre regole fondamentali per rendere i dati aperti, reperibili ed accessibili:

- **scegliere la semplicità.** Cominciare con un progetto piccolo, semplice e veloce. Non è necessario aprire tutti i dati in una sola volta. Inizialmente va bene aprire anche un solo dataset, o anche una sua parte – naturalmente, più dati si aprono, meglio è. Da ricordare che è innovazione. Muoversi il più in fretta possibile è bene, perché significa prendere slancio e imparare dall'esperienza – innovare comporta successi ed errori, e non tutte le banche dati saranno utili.
- **coinvolgere gli utenti fin dall'inizio e coinvolgerli spesso.** Cercare presto e spesso il confronto con i potenziali utilizzatori dei dati fra cittadini, imprese o sviluppatori. Ciò aumenterà la rilevanza dell'iniziativa durante tutto il suo percorso. È essenziale tenere presente che gran parte dei dati non raggiungeranno gli utenti finali direttamente, ma tramite 'info-intermediari'. Queste sono le persone che prendono i dati e li trasformano o li remixano per la presentazione. Ad esempio, la maggior parte di noi non vuole o non ha bisogno di un grande database di coordinate GPS, preferiamo decisamente una mappa. Così coinvolgete da subito gli info-intermediari, in modo che essi possano riutilizzare e riadattare i vostri dati.
- **affrontare i timori e le incomprensioni diffuse.** Questo è importante soprattutto se lavori in o con grandi organizzazioni come le istituzioni governative. Nell'aprire i dati sorgeranno molte domande e timori. È importante (a) identificare le più rilevanti, e (b) darvi una risposta il più presto possibile.

Si evince che la collaborazione stretta e costante con i ricercatori è fondamentale. Dall'aiuto nella scelta del database fino alla scelta del tipo di dati da caricare, il bibliotecario deve poter fornire consulenze di alto livello. Non è quindi più possibile affidarsi solamente all'impact factor di una rivista per compiere delle scelte, ma occorre prendere in considerazione anche questi ultimi aspetti.

Per facilitare queste operazioni, nel 2014 sono stati elaborati i "Principi FAIR" (Findability, Accessibility, Interoperability, Reusability); due di essi sono già stati menzionati sopra, ma vale la pena descriverli tutti:

- **reperibili:** La prima cosa da fare per rendere i dati e i metadati riutilizzabili è renderli rintracciabili facilitandone la ricerca sia per gli esseri umani che per i computer. Il recupero automatico e affidabile di set di dati e servizi dipende dagli identificatori persistenti (PID) e dai metadati leggibili dalle macchine.
- **accessibili:** I (meta)dati devono poter essere recuperati attraverso il loro identificatore utilizzando un protocollo di comunicazione standardizzato e aperto, che includa eventualmente dei sistemi di autenticazione e autorizzazione. Inoltre, i metadati dovrebbero essere resi disponibili anche quando i dati non lo sono più.
- **interoperabili:** I dati devono poter essere combinati e utilizzati con altri dati o strumenti. Il formato dei dati deve pertanto essere aperto e interpretabile da vari strumenti, compresi altre basi di dati. Il concetto di interoperabilità si applica sia a livello di dati che di metadati. Ad esempio, i (meta)dati dovrebbero utilizzare un linguaggio che riprende i principi FAIR.
- **riutilizzabili:** In sostanza, i principi FAIR mirano ad ottimizzare il riutilizzo dei dati della ricerca. A tal fine, sia i metadati sia i dati devono essere descritti nel migliore dei modi perché possano essere replicati e/o combinati in contesti diversi. Il riutilizzo dei metadati e dei dati dovrebbe essere dichiarato con una/o più licenze chiare ed accessibili. (FORCE11, 2014)

Queste linee guida sono essenziali dal punto di vista del funzionamento specifico delle macchine piuttosto che da quello dei ricercatori. Utilizzandole, infatti, i motori di ricerca dei vari database o repository sono facilitati nel recupero di informazioni pertinenti.

Molte biblioteche universitarie, inclusa la biblioteca dell'Università di Leiden (Paesi Bassi), offrono corsi specifici rivolti ai ricercatori su tematiche legate all'Open Data, ad esempio: come scrivere un Data Management Plan, come pubblicare i propri dati, come scegliere i repository, etc.

10. Incoraggiare la produzione di dati aperti: le piattaforme collaborative

Le biblioteche sono, sempre di più, spazi di condivisione. Questo vale anche per i ricercatori, ed è importante che biblioteche e bibliotecari possano assistere i propri ricercatori ed offrire loro

soluzioni per lavorare al meglio in maniera collaborativa con altre istituzioni al fine di arrivare a produrre Open Data.

Un modo per sostenere i ricercatori in questo senso è la creazione di piattaforme collaborative:

“Le piattaforme collaborative online permettono ai ricercatori geograficamente lontani tra loro di entrare in contatto e di collaborare in maniera continuativa ad un lavoro di ricerca, condividendo oggetti di studio, idee ed esperienze. In genere, le piattaforme collaborative mettono a disposizione dei servizi online in un ambiente virtuale al quale più utenti possono collegarsi simultaneamente e lavorare alla stessa attività. Sono compresi sia ambienti di ricerca virtuali grandi (VRE) che comprendono una serie di strumenti atti a facilitare la condivisione e la collaborazione - tra cui forum virtuali e wiki, servizi di archiviazione collaborativa di documenti (file hosting) e strumenti disciplinari come l'analisi o la visualizzazione dei dati – così come anche singoli strumenti specifici che consentono ai ricercatori di lavorare insieme in tempo reale su particolari aspetti collegati alla ricerca (come la stesura o l'analisi).” (Il manuale per formatori della Scienza Aperta, 2019)

Ad esempio, la biblioteca dell'Università di Leiden (Paesi Bassi), offre un progetto di VRE (Virtual Research Environment) per i propri ricercatori: nello specifico i bibliotecari si occupano di creare un VRE *ad hoc* per ogni progetto di ricerca ma soprattutto offrono un training specifico per l'utilizzo del VRE e supporto tecnico.

11. Formazione all'utilizzo dell'Open Data: alcuni spunti

Siamo partiti dal presupposto secondo cui le biblioteche sono facilitatori nel processo di produzione e promozione dell'Open Science e, nello specifico, dell'Open Data.

Sia produzione che promozione, però, necessitano di un certo grado di formazione – una formazione che può essere gestita dai bibliotecari stessi e può coinvolgere, a seconda delle esigenze, sia i ricercatori che i cittadini, o entrambe le categorie.

I metodi di formazione sono svariati e richiedono un'analisi approfondita del target a cui sono rivolti. In generale, il bibliotecario dovrebbe poter offrire workshop, seminari e mini-corsi volti a rispondere alle esigenze del proprio pubblico di riferimento, oltre ad offrire un costante servizio di reference e supporto.

Un esempio interessante di attività svolta in biblioteca sul tema dell'Open Science sono gli “Open Science Meet-ups” organizzati dalla Biblioteca Universitaria di Göttingen:

“Il gruppo “Open Science” di Göttingen, costituito da ricercatori e bibliotecari sostenitori della Scienza Aperta nonché dello scambio reciproco di conoscenza, organizza regolarmente degli incontri nel corso dei quali si discute sui principali argomenti collegati alla Scienza Aperta. La rete raccoglie tutte le persone che all'interno dell'Università di Göttingen si interessa a vario titolo di Scienza Aperta ed è aperta a tutti gli interessati. Sono diventati degli appuntamenti molto popolari e riescono ad attrarre studiosi di svariate discipline, altamente motivati a condividere le proprie esperienze sull'open scholarship e ad imparare nuovi metodi, strumenti e pratiche. I relatori invitati di solito introducono l'argomento soltanto, mentre in maniera più approfondita viene poi trattato all'interno di gruppi più piccoli.” (Il manuale per formatori della Scienza Aperta, 2019)

Un altro esempio è quello del workshop “Open Science – what's in it for me?” tenutosi presso la biblioteca dell'Università di Torino nel 2018:

“Il workshop mira a fornire a ricercatori e funzionari amministrativi dei concreti modelli di strumenti e processi nell'ambito della Scienza Aperta, trasversali a tutte le diverse discipline perché inizino ad utilizzarli e a discuterne. Per questo, viene fornita una panoramica di pratiche e di strumenti della Scienza Aperta che vengono utilizzati nel corso del procedimento scientifico con esempi pratici, discussioni partecipate e coinvolgendo la platea. Il secondo giorno è dedicato alla messa in pratica e alla condivisione. A turno, i partecipanti esplorano e, laddove possibile, sperimentano o fanno uso di strumenti e di procedure. In gruppi di poche persone, o individualmente e o anche in sessioni plenarie molto vivaci. Durante la sessione finale, si discute sugli ostacoli e gli incentivi collegati al

passaggio alla Scienza Aperta prendendo come caso specifico il loro lavoro di ricerca.” (Il manuale per formatori della Scienza Aperta, 2019)

Conclusione

Open Science è il perfetto connubio tra biblioteconomia e scienze dell'informazione.

Stiamo assistendo ad un grosso cambiamento nella società, nel concetto di scienza ma anche nel mondo delle biblioteche: da una parte, la sempre maggiore digitalizzazione sta cambiando il modo in cui usufruiamo e gestiamo le risorse documentali; dall'altra, il concetto di terzo luogo si sta affermando sempre più in tutto il mondo.

Esempi concreti sono la Filanda di Mendrisio, un luogo in cui la biblioteca è di fatto un luogo terzo in cui la comunità in tutta la sua eterogeneità partecipa ai processi sociali, e il futuro polo bibliodocumentale del Dipartimento Ambiente, Costruzione e Design della SUPSI, di cui sono fieramente responsabile. In quest'ultimo caso non si tratta solo di un terzo luogo, ma anche di una vera e propria *agora* di conoscenze e saperi che includono biblioteca e Centro Documentazione Materiali nelle loro varie declinazioni. Va da sé, quindi, che all'interno della SUPSI, ed in particolare nel DACD, l'Open Data sarà un traguardo importantissimo non solo da raggiungere ma anche da implementare. Se vogliamo infatti diventare, citando un famoso titolo di biblioteconomia, una vera e propria piazza del sapere, l'Open Data dovrà avere un ruolo centrale nel futuro del dipartimento.

In questo Lavoro di Diploma mi sono concentrata sugli aspetti tecnici e formali dell'Open Data e dell'Open Science in particolare, trascurando la parte forse più poetica del movimento: riuscire a mettere in atto un cambiamento tale da creare una profonda collaborazione ed uno stretto scambio di sapere tra comunità scientifiche che portino, come ultima ragione, al miglioramento della società stessa *in toto*.

Come bibliotecari, abbiamo il dovere di facilitare e incoraggiare l'approccio alla cultura ed alla scienza, ad ogni livello biblioteconomico: dal servizio per le biblioteche per l'infanzia, passando per le biblioteche scolastiche, fino alle università.

Tutta l'antropologia si basa sull'assunto secondo cui l'umano è un animale sociale. Questo è vero in tutti gli ambiti della vita dell'essere umano, ergo anche la scienza ed il progresso tecnologico, il sapere e l'arte, devono essere sociali e condivisi proprio per la natura intrinseca dell'uomo.

In un interessante documentario sul CERN (che, guarda caso, ha sviluppato una delle migliori piattaforme Open Science attualmente esistenti) una frase in particolare mi ha colpita: l'intervistato sosteneva che il CERN riuscisse nel miracolo di far lavorare insieme scienziati provenienti da culture e religioni diverse, spesso in contrasto tra loro, al fine di perseguire uno scopo comune.

Scienziati palestinesi che lavorano al fianco di scienziati israeliani – questa è l'essenza dell'Open Science.

Lavorare insieme per un mondo migliore.

Bibliografia e sitografia

Addis, Antonio, & De Fiore, Luca. (aprile 2019). La disseminazione della ricerca. Forward. <https://forward.recentiprogressi.it/it/speciali/la-disseminazione-della-ricerca/>

AFS, Amministrazione federale svizzera. Dichiarazione Universale dei Diritti Umani delle Nazioni Unite, 10 dicembre 1948. <https://www.bar.admin.ch/bar/it/home/service---publikationen/publikationen/geschichte-aktuell/allgemeine-erklaerung-der-menschenrechte-durch-die-vereinten-nat.html>

Hagstrom, S. (3 settembre 2014). The FAIR Data Principles. FORCE11. <https://www.force11.org/group/fairgroup/fairprinciples>

opendata.swiss. (n.d.). Il nostro mandato. <https://opendata.swiss/it/il-nostro-mandato/>

Il Manuale per Formatori della Scienza Aperta. https://doi.org/10.15168/11572_246342

FOSTER. (n.d.). Open data definition. <https://www.fosteropenscience.eu/taxonomy/term/110>

enCOMPASS Project. (n.d.). <http://www.encompass-project.eu/pilots/swiss-pilot/>

Open Science Europe: Overview. OpenAIRE. Ultima consultazione 21 agosto 2020 <https://www.openaire.eu/open-science-europe-overview>

Zenodo Research. Shared. (n.d.). from <https://about.zenodo.org/>

SUPSI Open Science <https://www2.supsi.ch/cms/openscience/>



Questa pubblicazione, *Open Data e Biblioteche DAS BSI*, scritta da *Francesca Torti*, è rilasciata sotto Creative Commons Attribuzione – Non commerciale 3.0 Unported License.