



Accesso aperto ai dati della ricerca come vettore per la scienza aperta

Elena Giglia

I due giorni dedicati al tema *Open Access to research data as a driver for Open Science* (Atene, 15 e 16 gennaio 2015), evento conclusivo del progetto europeo RECODE - Policy RECommendations for Open Access to Research Data in Europe, sono stati un'occasione unica per riflettere sul tema dell'accesso aperto ai dati della ricerca come vettore per la scienza aperta.¹

Relatori di altissimo livello e una platea vivace hanno dialogato su valori, rischi, opportunità, questioni aperte, buone pratiche: scopo di questo lavoro è tentare di dare un'idea della complessità del panorama attuale e dell'importanza che il tema dei dati aperti rivestirà nell'immediato futuro.

1. Intelligent openness

La Open Science è considerata dai più come il nuovo paradigma della scienza: ci troviamo in piena fase di «creative destruction», ovvero, secondo l'economista Joseph Schumpeter, «the process of industrial mutation that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one» (Schumpeter, 1947, II, VII, 83). La transizione verso la Open

¹ Il video e le presentazioni saranno disponibili sul sito del progetto RECODE <http://recodeproject.eu> e su quello dell'EKT - Ente Nazionale di Documentazione Ellenica http://openaccess.gr/conferences/conference2015/index.dot?language_id=1. Nel corso del testo i riferimenti senza nota esplicita sono a citazioni nelle presentazioni dei relatori ad Atene.



Science è avviata, anzi, come ha sottolineato Mark Thorley dei Research Councils britannici, non c'è modo di tornare a quella che per antitesi possiamo definire la "closed science": siamo in momento cruciale, in cui si stanno definendo i paradigmi di questa nuova scienza; abbiamo quindi la pesante responsabilità di far sì che essi siano i più aperti e inclusivi possibile. Il volume *Science as an open enterprise*, curato dalla Royal Society nel 2012, è una lettura illuminante per la ricchezza dei contenuti e l'approccio innovativo della "intelligent openness": non è sufficiente mettere i dati online, occorre farlo in modo intelligente, come ha ricordato Geoffrey Boulton, neo presidente di CODATA. I dati devono essere ritrovabili, accessibili, valutabili e usabili, da cui l'importanza dei metadati ad essi associati:

Realising the benefits of open data requires effective communication through a more intelligent openness: data must be accessible and readily located; they must be intelligible to those who wish to scrutinise them; data must be assessable so that judgments can be made about their reliability and the competence of those who created them; and they must be usable by others. For data to meet these requirements it must be supported by explanatory metadata (data about data). As a first step towards this intelligent openness, data that underpin a journal article should be made concurrently available in an accessible database. We are now on the brink of an achievable aim: for all science literature to be online, for all of the data to be online and for the two to be interoperable (Royal Society 2013).

In questo scenario risulta centrale dal punto di vista tecnico il ruolo degli specialisti dell'informazione che da sempre hanno trattato i metadati - i bibliotecari -, anche se per una gestione effettiva dei dati è necessario creare all'interno di ogni ente una sinergia fra esperti di area informatica per database e software, di area biblioteche appunto per metadati e conservazione, di area amministrativa per aspetti legali

e finanziari legati ai *grants*: tutti costoro devono interagire con i ricercatori, portatori delle competenze relative alle pratiche delle singole discipline, ai requisiti, al processo di ricerca, all'analisi e visualizzazione dei dati. Dal punto di vista politico, invece, un ruolo determinante possono giocare sia gli enti di finanziamento della ricerca sia i comitati scientifici delle riviste nel richiedere, contestualmente alla pubblicazione degli articoli scientifici, anche la pubblicazione dei dati che costituiscono la sostanza dell'articolo in questione, ovvero, in ultima istanza, della ricerca finanziata.

L'editore PLoS, come spesso accade, è all'avanguardia anche in questo, avendochiesto fin da marzo 2014 il deposito dei dati per tutti gli articoli pubblicati. Iratxe Puebla, di PLoS ONE, ha presentato i primi riscontri relativi alla policy e a due servizi utili ai ricercatori: l'accordo con DRYAD per un workflow semplificato di deposito dei dati e lo sviluppo del progetto *Making Data Count: A Data Metrics Pilot Project*, verso una metrica appositamente studiata per i dati.² Sempre PLoS ha curato le *Recommendations for the Role of Publishers in Access to Data* (Lin e Strasser 2014), da cui vale la pena estrapolare gli otto punti chiave, che anche gli altri editori dovrebbero far propri:

- Establish and enforce a mandatory data availability policy.
- Contribute to establishing community standards for data management and sharing.
- Contribute to establishing community standards for data preservation in trusted repositories.
- Provide formal channels to share data.
- Work with repositories to streamline data submission.
- Require appropriate citation to all data associated with a publication —both produced and used.

² La policy delle riviste PLoS su Open Data si legge in <http://www.plosone.org/static/policies#sharing>; i termini per il deposito in DRYAD in <http://data.dryad.org/pages/journalIntegration>; il progetto Making data count in <http://escholarship.org/uc/item/9kf081vf#page-1>

- Develop and report indicators that will support data as a first-class scholarly output.
- Incentivize data sharing by promoting the value of data sharing.

2. Il valore dei dati aperti

Lo scambio e la circolazione della conoscenza è uno dei paradigmi principali della scienza; la rivoluzione digitale ne ha solo modificato gli strumenti, rendendo possibile una maggiore disseminazione, non il fine ultimo, come ha ricordato Jennie Larkin dei National Institutes of Health. L'idea dell'accesso aperto ai dati trova perfetta consonanza con lo statuto stesso della ricerca scientifica, che deve essere riproducibile: non esiste una "catena dei dati", è la catena stessa della logica scientifica che si fonda sui dati, per cui averli a disposizione contribuisce alla trasparenza e alla solidità della ricerca. In un contesto di scienza aperta, le parole d'ordine diventano allora «re-use, re-purpose, re-produce»: il riuso dei dati da differenti prospettive rende vero l'apologo di Rufus Pollock secondo cui «the coolest thing to do with your data will be thought of by someone else» (Pollock 2007). I dati, secondo l'ammonizione di Kevin Ashley del Digital Curation Centre britannico, sono «un record, non una verità»: chi fa ricerca non asserisce mai che essi siano "corretti" ma che siano appunto "ricerca" in sé, perché sempre confutabili, incrementabili, migliorabili con nuovi approcci. Chi tratta e cura i dati li mette a disposizione quindi di pubblico ed esperti in modo agnostico: il curatore può dire "questi sono i dati su cui la ricerca è stata condotta", oppure "sono i dati sulla cui base sono state prese le decisioni politiche", o ancora "questi sono i dati disponibili in questo momento", sta poi al ricercatore l'analisi e la disamina dei dati stessi, per corroborarli o confutarli, come è appunto nello statuto della scienza. Se i dati risultano scorretti, a maggior ragione è necessario che sia l'intero set sia la discussione che ne scaturisce siano pubblici, per trasparenza e controllo, oltre che - non dimentichiamolo - per il fine ultimo che resta l'avanzamento della conoscenza scientifica, la quale procede per tentativi.

Non sono solo gli scienziati a trarre beneficio dalla condivisione dei dati, ma anche i cittadini, direttamente o indirettamente grazie allo sviluppo di prodotti e servizi nati da dati aperti, e le imprese, perché i dati di base della ricerca sono linfa per la ricerca applicata e la creazione e commercializzazione di nuovi prodotti - il che genera, ovviamente, anche nuova occupazione. Il fondamentale rapporto curato da Research Data Alliance per la Commissione Europea a dicembre 2014, *The Data Harvest: how sharing research can yield knowledge, jobs and growth* (Research Data Alliance 2014), raccoglie fra l'altro cifre sul ritorno sugli investimenti, per cui valgono alcuni esempi: per ogni sterlina investita dal governo britannico nel progetto *Economic and Social Data Service*, il ritorno è stato di 5,40 sterline; il valore del riutilizzo creativo di dati del settore pubblico in Europa si aggirava intorno a 32 miliardi di euro già nel 2010; *The Human Genome Project* ha fruttato in nuove medicine, nuova occupazione, nuovi servizi un ritorno valutato in 1000 miliardi di dollari a fronte dei 3,8 investiti inizialmente. Il rapporto è anche una fonte preziosa di buone pratiche e di proposte operative, oltre che innovative, per cui vale la pena leggere le frasi conclusive, del tutto significative:

Numbers matter. A crowd behaves differently than a few individuals; an avalanche has different dynamics than a snowball; and today's rising volume of scientific data brings a change in quality, not just quantity. It changes the way science is performed, the problems it can tackle, and the speed and nature of the solutions it finds. It generates new products, services and jobs. It challenges old policies – and, in time, will change the way policy itself is made, in all spheres. This requires a bold response. We are starting to see it in major capitals around the world. We are also starting to see a collaborative approach, with the formation of the Research Data Alliance to help coordinate policies. But, equally, the challenges are great: It will take money to build the infrastructure, support for clever pioneers and entrepreneurs, and

creative solutions to the policy problems with data privacy and security. These are all issues which we urge the new European Parliament and Commission to take up, urgently. The strength of Europe's science base is at stake, surely. But more broadly, so is the Union's competitiveness globally – its ability to create jobs, attract investment, and maintain the social fabric Europe values. Exactly how this will all evolve, we don't yet know; but we do know it will be important. As the British Internet pioneer Tim Berners-Lee put it: "The Web as I envisaged it, we have not seen it yet. The future is still so much bigger than the past."

3. L'ecosistema dei dati aperti

Trovandoci nel momento di transizione verso la scienza aperta, ci manca la prospettiva storica per valutare con senso critico. Nell'ecosistema dei dati aperti, in fase di strutturazione, è importante però tener conto di alcuni moniti. Dal punto di vista culturale, bisogna evitare di creare unennesimo "knowledge divide": è necessario tenere in considerazione le situazioni di paesi in cui il livello di infrastrutture e di consapevolezza sul vantaggio di condividere i dati è minore, come ha ricordato Geoffrey Boulton. Il cambiamento in atto è epocale, ma si tratta in prima istanza di un cambiamento culturale, per cui John Wood di Research Data Alliance ha ribadito la necessità preliminare di creare un clima di fiducia reciproca per far crescere una cultura della condivisione. La resistenza al cambiamento è comune nelle fasi di passaggio: oggi, nel mondo accademico, il successo deriva dalla "pubblicazione" scientifica tradizionale; in un clima di *openness* i prodotti della ricerca possono essere invece molteplici e variegati (pubblicazioni, dati, software...): il timore di ciò che ancora non si conosce porta alla resistenza, come fu nel passaggio ricordato da Jennie Larkin fra le monete d'oro e la carta moneta, in cui la gente comune si chiedeva se davvero un pezzo di carta avrebbe avuto lo stesso valore... Risulta quindi evidente come il sistema di incentivi e

di riconoscimento per chi mette a disposizione i propri dati costituisca una spinta decisiva verso l'apertura.

Un'altra preoccupazione riguarda invece i costi: inutile nascondersi, secondo Mark Thorley, che i costi della Open Science possano essere maggiori rispetto a quelli della "closed" science attuale. Di più: spesso il rapporto costi/benefici non è immediatamente percepibile, perché il riuso dei dati aperti genera valore in un'area della società diversa da quella che ha sostenuto le spese iniziali, come ha ricordato Kevin Ashley. I costi sono elevati, per cui da più parti si è evocata una selezione dei dati da conservare, ma Geoffrey Boulton ha messo in guardia da una selezione troppo rigida dei dati da conservare nel lungo periodo: con una metafora mutuata dalla fisica, il "rumore" di oggi può essere il "segnale" di domani se interpretato correttamente o se posto di fronte a domande differenti. Anche nella suddivisione di ruoli e responsabilità occorre avere cautela: nell'ecosistema dei dati aperti in via di formazione, come in un sistema biologico, non si possono definire a priori tutte le funzioni, che si delineano via via che l'organismo cresce; è necessario però rispettare un approccio a più livelli: internazionale per l'armonizzazione delle politiche e la standardizzazione, un ruolo che CODATA e RDA svolgono egregiamente;³ nazionale, per esigenze specifiche legate alle singole realtà; disciplinare, perché ogni disciplina produce dati diversi e assegna loro un diverso valore, oltre ad avere un diverso grado di cultura della condivisione. Quello che è certo è che le potenzialità del web vanno sfruttate appieno, il web è infatti costituito da *links* e soprattutto da relazioni fra entità: un semplice pdf online – quale quello utilizzato oggi da gran parte delle piattaforme editoriali anche di prestigiosi editori commerciali – è del tutto restrittivo. La "pubblicazione scientifica" offre diverse ulteriori possibilità e soprattutto può essere valorizzata facendo leva sui legami possibili

³ CODATA, Committee on Data for Science and Technology, <http://www.codata.org/about-codata/our-mission>; RDA, Research Data Alliance, <https://rd-alliance.org>

nel web semantico, legando un concetto a un altro, e sfruttando tecniche quali *data mining* e *text mining*, come ha suggerito Phil Archer, del consorzio W3C. Anche rispetto alla qualità dei dati il discorso si fa più complesso e non può ridursi ad essere unidimensionale: è necessario allora concentrarsi su quegli aspetti della qualità che davvero ne accrescono il valore, secondo l'indicazione di Kevin Ashley.

In sostanza, va ripensato l'intero ecosistema della comunicazione scientifica, con un *focus* specifico sul sistema di incentivi ai ricercatori che condividono i propri dati e sui servizi utili a far rientrare i dati nel flusso corrente della comunicazione scientifica, quali l'adozione di standard di citazione, l'uso di identificativi univoci come il DOI fornito dal consorzio DataCite appositamente per i dataset, lo sviluppo di metriche per il conto e la tracciabilità delle citazioni ai dati.⁴

Un capitolo a parte riguarda il tema delle licenze: anche se i dati sono aperti e si trovano liberamente sul web, non sapere quale uso se ne possa fare di fatto costituisce una barriera. Unanimemente è stata sottolineata quindi l'importanza di associare le licenze ai dati – in particolare, le Licenze Creative Commons che sono digitali e armonizzate - per chiarire in modo inequivocabile gli usi permessi e per porre obblighi anche nei confronti degli utilizzatori: tutto questo va esplicitato, come sempre, a livello di metadati. Nel caso dei dataset, per permettere successivi servizi di *data mining*, vengono consigliate le Licenze Creative Commons 0 che consentono il più ampio riuso possibile.⁵ Il contesto europeo e internazionale della ricerca può generare poi casi di autori che ricadano sotto regimi legali differenti, o di contraddizioni fra le richieste degli enti finanziatori e clausole di riservatezza presenti nel contratto con terzi: ma, come ha ben detto

⁴ <https://www.datacite.org>.

⁵ Le Licenze Creative Commons 0 prevedono la rinuncia a qualsiasi diritto d'autore, <http://creativecommons.org/publicdomain/>

Rachel Finn di Trilateral consulting, la complessità degli aspetti etico legali non deve essere vista come una barriera ma come una opportunità per trovare soluzioni armoniche.

4. Le resistenze

A fronte di questioni aperte nella creazione di un ecosistema favorevole alla cultura della condivisione dei dati, sia a livello tecnico (standardizzazione, accessibilità, sicurezza e integrità dei dati, conservazione e cura, ricercabilità...), sia legale, sia culturale, c'è una resistenza di fondo da parte dei ricercatori, che si traduce in un repertorio di "scuse" che essi avanzano per non rendere aperti i propri dati: a questi, Kevin Ashley suggerisce soluzioni semplici ed efficaci, tratte dal blog di Carly Strasser, curatrice di dati alla California Digital Library (Strasser 2013):

- I miei dati possono essere male interpretati o fraintesi: a maggior ragione, vanno resi pubblici contestualizzandoli con tutte le corrette istruzioni e spiegazioni del caso, sarà utile per tutti
- Mi contatteranno per chiedermi cose e non ho tempo: ...ma la scienza è di per sé una grande conversazione e consiste esattamente nel discutere le proprie idee con gli altri...
- I miei dati non sono interessanti: il giudizio va lasciato agli altri, secondo il principio per cui ciò che per me è rumore per altri è un segnale
- Potrei usare i dati in una nuova pubblicazione: è una preoccupazione reale, ma si risolve facilmente associando ai dati un periodo di embargo per permettere il pieno sfruttamento
- Non sono certo di avere i diritti sui dati: questo si può chiarire risalendo nella catena delle responsabilità istituzionali, molto spesso è più facile di quanto sembri

- I miei dati sono troppo complicati da condividere: ... se sono complicati da condividere, allora è complicato anche replicare l'esperimento, siamo certi che si tratti di un vero avanzamento per la scienza?
- I miei dati non sono curati: i lettori li accetteranno per quello che sono, al ricercatore servirà da esperienza per migliorare la volta successiva
- Non è una priorità e sono molto impegnato: in realtà, il capitolo "Open Data" sta diventando una priorità per un numero crescente di enti finanziatori della ricerca e di comitati scientifici di riviste.

5. Top-down: il contesto europeo

La Commissione Europea, nel suo ruolo di finanziatore nel quadro di Horizon 2020, è uno degli enti che ha accelerato decisamente sulle politiche di apertura. L'Open Access è visto a Bruxelles come un tassello fondamentale della Open Science, che in modo unanime viene riconosciuta come il futuro della scienza in Europa per le sue implicazioni in termini di innovazione e competitività; l'accesso aperto alle pubblicazioni e ai dati, del resto, è la chiave di volta del cambiamento in atto nella comunicazione scientifica attuale, come risulta dalla Consultazione pubblica del 2013.⁶ Per la Commissione Europea l'accesso aperto costituisce lo strumento cardine per ottenere lo scopo ultimo, che resta la massimizzazione dell'impatto della ricerca finanziata con i fondi pubblici. Daniel Spichtinger, della Commissione, ha fornito alcune coordinate di riferimento di Horizon 2020 che, ricordiamo, prevede l'Open Access per tutte le pubblicazioni e un progetto pilota sui dati per alcune aree chiave, con poche definite clausole di *opt out* e la possibilità, sempre aperta, di adesione volontaria anche da parte di ricercatori di aree non coinvolte. Il

⁶ *EU Consultation on Open Research Data - Report*, https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/Report_2013-07-OpenResearchData-Consultation-FINAL1.pdf

riscontro incoraggiante è che la percentuale di chi ha optato per l'adesione volontaria supera quella di chi invece ha preferito non rendere aperti i dati pur appartenendo alle aree pilota.⁷ Il progetto pilota prevede che entro sei mesi dall'approvazione del progetto il ricercatore depositi il *Data Management Plan* per la gestione e conservazione dei dati. Una chiosa: è necessario, da parte di chi supporta la ricerca – bibliotecari *in primis* – acquisire fin da ora le competenze specifiche per la redazione coerente e la corretta gestione tecnica di un *Data Management Plan (DMP)*. Nel corso del dibattito è emerso come il DMP debba entrare a far parte del tessuto della ricerca e della comunicazione scientifica, entrando a far parte anche del processo di *peer review*, dato che solo gli esperti del settore sono in grado di giudicare l'appropriatezza di un piano di gestione e conservazione dei dati secondo le specifiche della disciplina. La Commissione stessa ha fornito le *Linee Guida sulla Gestione dei dati in Horizon 2020*,⁸ e il progetto europeo FOSTER ha dedicato larga parte del proprio sforzo di formazione proprio al tema degli Open Data:⁹ sarebbe utile, in Italia, attrezzarsi da subito creando le giuste sinergie fra chi ha competenze di infrastrutture, di metadati, di licenze, di servizi a valore aggiunto per gestire al meglio il deposito e l'accesso ai dati come risorsa preziosa e non come ennesimo carico burocratico sulle spalle del ricercatore.

⁷ Le aree interessate sono: Future and Emerging Technologies; Research infrastructures – part e-Infrastructures; Leadership in enabling and industrial technologies – Information and Communication Technologies; Societal Challenge: Secure, Clean and Efficient Energy – part Smart cities and communities; Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – with the exception of topics in the area of raw materials; Societal Challenge: Europe in a changing world – inclusive; innovative and reflective Societies; Science with and for Society: cfr. http://europa.eu/rapid/press-release_IP-13-1257_en.htm

⁸ *Guidelines on Data Management in Horizon 2020*, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

⁹ Il sito del progetto FOSTER contiene un elenco di risorse utili per corsi, presentazioni, promozione: <https://www.fosteropenscience.eu/foster-taxonomy/open-data>

Uno strumento utile a chi vuole aggiornarsi in tema di Open Science è il portale *Netvibes - Open Access to research*, curato dalla Biblioteca Centrale della Commissione Europea, che dà accesso a risorse su Open Access, Open Data, MOOCs, news tratte dal web, da mailing lists ma anche da Twitter, con #openaccess.¹⁰

Chi è interessato ai più recenti rapporti in tema di strategie di pubblicazione in Open Access, dello stato dell'arte sugli Open Data, sull'evoluzione delle politiche e del tasso di reperibilità dei lavori scientifici in Open Access può visitare la pagina di Science-metrix e consultare quelli commissionati dagli organi politici europei.¹¹

6. Bottom-up: buone pratiche da consorzi e data centres

Posto che, come ha sottolineato Max Wilkinson dell'University College di Londra, spesso il contesto in cui ci si muove per implementare una politica sull'accesso aperto somiglia in modo imbarazzante a una famiglia disfunzionale, in cui messaggi contrastanti provengono da soggetti diversi che a diverso titolo esercitano un qualche potere, ci sono enti di ricerca e di finanziamento che hanno una politica e *data centers* che hanno buone pratiche da condividere. Molti di loro erano presenti ad Atene, da ognuna di queste esperienze si possono trarre spunti o strumenti di grande utilità su diversi piani.

6.1 Piano politico

«Open by default»: questo è il principio della Legge greca sull'accesso aperto ai dati del settore pubblico, varata a ottobre 2014, che prevede come standard l'apertura, e la chiusura solo per quei dati per cui sussistano validi motivi. Il portale *Open data* lista già 75 dataset ed è aperto ai suggerimenti dei cittadini sui sets che riterrebbero utile

¹⁰ Netvibes-Open Access to research, <http://www.netvibes.com/open-access#Home>

¹¹ Science-metrix Reports, <http://www.science-metrix.com/en/publications/reports>

vedere pubblicati.¹² Sulla base di questi dati sono state create App come publicspending.net che traccia spese e pagamenti effettuati con denaro pubblico o servizi di ricerca full text come Cl@rity, improntato alla logica della massima trasparenza.¹³ Sul tema dei dati aperti nel settore pubblico i progetti di riferimento sono Share-PSI 2.0, curato dal consorzio W3C¹⁴ e Open Data Barometer, curato da Open Data Research Network.¹⁵

Che la scienza aperta sia un bene pubblico è una convinzione comune; la cifra innovativa contenuta nel messaggio congiunto di CODATA e ICSU - International Council for Science è che anche la valutazione della ricerca debba contribuire alla trasparenza e all'apertura. Perpetuare le metriche quali Impact Factor e *ranking* di reputazione delle riviste favorisce invece la chiusura e la competizione sfrenata più che l'apertura e il reale avanzamento della scienza:

Increasingly, those involved in the administration of research rely on metrics designed to assess the importance and impact of research as an aid to evaluation, with publication outputs in traditional scientific journals being the major focus. These metrics in turn affect the behaviour of researchers, such as their choice of journals, as they seek to maximize their performance as measured by the metrics used. They can contribute to the maintenance of high journal prices, promote intense competition rather than openness and sharing, and fail to recognise research contributions such as the production of datasets, software, code, blogs, wikis and forums (ICSU 2014).

A livello di Atenei, la *LERU Roadmap for Research Data*, presentata all'inizio del 2014 dalla LERU - League of European Research

¹² Open Data in Greece, <http://data.gov.gr/>

¹³ publicspending.net- where public money goes worldwide per la Grecia: <http://publicspending.net/paver.php?graph=greece>; Cl@rity, <https://diavgeia.gov.gr/>

¹⁴ Share-PSI 2.0, <http://www.w3.org/2013/share-psi/>

¹⁵ Open Data Barometer, <http://www.opendataresearch.org/project/2013/odb>

Universities, è un documento fondamentale. Basata sull'assunto per cui la politica, l'infrastruttura tecnologica e il supporto vanno offerti simultaneamente ai ricercatori per essere efficaci, la *Roadmap* si articola in 6 capitoli con relative minuziose raccomandazioni su temi quali Policy, Advocacy, Management, Costs, Infrastructure, Roles and Responsibilities (LERU 2014).

6.2 Piano economico

La condivisione dei dati massimizza il ritorno sugli investimenti, e la disseminazione è il fulcro della ricerca: questi sono due dei principi dei Research Councils britannici, che hanno stilato 7 *Common Principles* sulla base dei quali ogni Research Council definisce la propria politica¹⁶. Una decisa presa di posizione da parte degli enti di finanziamento della ricerca è la vera leva per la crescita della Open Science. Del resto, già nel 2007 l'OCDE riconosceva che

Access to research data increases the returns from public investment in this area; reinforces open scientific inquiry; encourages diversity of studies and opinion; promotes new areas of work and enables the exploration of topics not envisioned by the initial investigators (OECD 2007).

6.3 Piano operativo

Sul piano strettamente operativo è necessario creare un ambiente favorevole che faciliti i ricercatori nel rendere aperti i dati: questa la strada seguita dall'University College London per il proprio *Research Data Service*¹⁷, un servizio centralizzato per la cura e la conservazione

¹⁶ Segnalo l'elenco dei principi comuni disponibili in <http://www.rcuk.ac.uk/research/data/policy/> e anche il rapporto *Open Data Dialogue*, che pur essendo del 2012 contiene ancora spunti interessanti, <http://www.rcuk.ac.uk/Publications/policy/OpenData/>

¹⁷ UCL Research data service, <http://www.ucl.ac.uk/research-it-services/our-work/research-data-service>; il testo della policy si legge in J LIS.it. Vol. 6, n. 2 (May 2015). Art. #11130 p. 238

dei dati che si avvale delle sinergie fra diverse competenze. La politica, approvata nel 2012 e rivista ogni anno, prevede che la responsabilità del deposito sia in capo all'autore, mentre l'ente è responsabile della conservazione; il servizio è oneroso e vanno previsti finanziamenti adeguati. Va prevista una struttura informatica sia modulare, perché l'ambiente dei dati aperti è in rapida evoluzione e la struttura deve adeguarsi in modo rapido e flessibile.

Una vera miniera di strumenti utili dal punto di vista operativo è il sito del Digital Curation Centre britannico, attivo dal 2004 e vero punto di riferimento nel contesto dei dati aperti: le risorse segnalate vanno da articoli scientifici a set di metadati per singole discipline a guide pratiche a *checklist* a *tutorials* su come creare un adeguato Data Management Plan.¹⁸

Anche LIBER, l'Associazione delle Biblioteche europee di ricerca, che spinge le 400 biblioteche aderenti a supportare attivamente la Open Science (si veda lo *Statement* di fine settembre 2014),¹⁹ sul sito istituzionale offre molti strumenti utili, come gli 11 *case studies* sulla corretta gestione dei dati o le linee strategiche 2013 in cui largo spazio viene assegnato al tema della gestione dei dati:

We want “to enable the research library to become a central point of contact for researchers and faculty in the area of scholarly publishing and research data

http://www.ucl.ac.uk/isd/services/research_services/research-data/content-objects/researchdata/uclresearchdatapolicy

¹⁸ Data curation centre, <http://www.dcc.ac.uk>; Guide e checklist:

<http://www.dcc.ac.uk/resources/how-guides>; Risorse generali:

<http://www.dcc.ac.uk/resources>; Data Management Plans:

<http://www.dcc.ac.uk/resources/data-management-plans>

¹⁹ LIBER *Statement on Open Science*, Sept. 2014, http://libereurope.eu/wp-content/uploads/2014/09/LIBER_Statement-on-open-science-final.pdf

management at institutional level for all academic disciplines".²⁰

LIBER è anche molto attiva nel promuovere a livello europeo sia la revisione della normativa sul copyright in direzione più funzionale alla ricerca sia l'adozione di strumenti legali e licenze che favoriscano invece di impedire – come è attualmente – il *text mining* e il *data mining*, che consentirebbero all'Europa una più rapida innovazione tecnologica e scientifica.²¹

Dal punto di vista strettamente pratico, lo strumento di base adatto a quei ricercatori che non abbiano un archivio aperto in cui depositare i dati è l'archivio multidisciplinare e internazionale ZENODO, ospitato dal CERN e frutto del progetto europeo OpenAIREplus. Accetta ogni tipologia di record scientifico; contiene a oggi 392 dataset, tutti identificabili e citabili attraverso DOI. Al progetto ZENODO sono legati anche gli assai istruttivi *Principles of data management and sharing at European Research Infrastructures*.²²

6.4 Buone pratiche

L'Olanda, come spesso accade, è all'avanguardia, avendo già creato un servizio nazionale federato per la cura dei dati della ricerca, il Research Data Netherlands - RDNL.²³ Il modello si articola in servizi offerti selettivamente da "Front Office" e "Back Office". Per "Front Office" si intendono le singole strutture universitarie e di ricerca, e

²⁰ *Case studies on research data management*, 2014: <http://libereurope.eu/committee/scholarly-research/research-data-management-case-studies/>; LIBER Strategy 2013: <http://libereurope.eu/strategy/>

²¹ LIBER on copyright: <http://libereurope.eu/copyright-reform/>; LIBER on Text and Data mining: <http://libereurope.eu/text-data-mining/>

²² ZENODO, <https://zenodo.org/>. Da leggere anche i *Principles of data management and sharing at European Research Infrastructures*, <http://zenodo.org/record/8304#.VMilMSzN4>

²³ RDNL è un consorzio fra i principali enti olandesi che si occupano di ricerca, 3TU.Datacentrum, Data Archiving and Networked Services (DANS) and SURF sara: <http://www.researchdata.nl/en>.

loro compito è la promozione, la formazione e il supporto diretto ai ricercatori. Il “Back Office” è costituito dalle tre strutture nazionali consorziate, che si occupano della conservazione e della cura dei dati, oltre che della formazione del personale del “Front Office”.²⁴ Questo consente di valorizzare le diverse competenze e realizzare economie di scala. Il consorzio mette a disposizione anche una utilissima *checklist* per la selezione dei dati da conservare.²⁵

Sempre in Olanda è nato il *Data Seal of Approval*, a cura del DANS - Data Archiving and Networked Services. Si tratta di uno strumento utilissimo che certifica la qualità degli archivi di conservazione - o meglio - la loro aderenza ai principi esposti in 16 linee guida (per i produttori, i curatori, gli utilizzatori) applicate all’archivio che ne fa richiesta con uno strumento online; al termine, se l’archivio risulta coerente con le linee guida, viene assegnato il sigillo.²⁶ Lo scopo è quello di allargare la rete degli archivi di qualità e contribuire a creare quel clima di fiducia che è alla base della diffusione della cultura della condivisione dei dati. I principi cui ci sia attiene sono riassunti nell’acronimo FAIRR, ovvero:

- The research data can be Found on the Internet.
- The research data are Accessible, while taking into account relevant legislation with regard to personal information and intellectual property of the data.
- The research data are available in a usable format (Interoperable).

²⁴ Il modello operativo back office/front office è illustrato qui: http://www.researchdata.nl/fileadmin/content/RDNL_algemeen/Documenten/RDNL_FOBomodel-UK-web.pdf

²⁵ RDNL, Checklist for data preservation selection: <http://www.researchdata.nl/en/activities/data-management/selecting-research-data/#c6630>

²⁶ *Data Seal of Approval*, <http://www.datasealofapproval.org/en/>. Le linee guida prese in considerazione si trovano qui: http://www.datasealofapproval.org/media/filer_public/2013/09/27/guide_lines_2014-2015.pdf

- The research data are Reliable.
- The research data can be Referred to.²⁷

7. Le raccomandazioni sulle politiche di accesso aperto ai dati della ricerca - RECODE

Le *Policy recommendations for open access to research data*²⁸ costituiscono il prodotto finale del progetto RECODE e al tempo stesso un'ottima sintesi di tutti i punti fin qui trattati.

Il progetto ha preso in considerazione cinque *case studies* di dati aperti in altrettante discipline (fisica, sanità, bioingegneria, ambiente e archeologia) e li ha incrociati con quattro grandi sfide aperte, ovvero i valori e gli ecosistemi relativi agli attori coinvolti nella catena della ricerca (finanziamento, creazione, disseminazione, cura, uso), le preoccupazioni legali ed etiche, le infrastrutture tecnologiche e il contesto istituzionale. Sono emersi due risultati preliminari: la mancanza di un ecosistema coerente per i dati e la mancanza di attenzione verso la specificità delle pratiche di ricerca, dei processi, della raccolta dei dati. Posto il comune riconoscimento dei dati come bene comune e del loro valore in termini di un più rapido avanzamento scientifico e dei vantaggi economico-sociali ad esso connessi, è risultato chiaro però come i diversi attori coinvolti abbiano una percezione diversa del percorso da intraprendere verso l'apertura dei dati, e questo genera catene di valori in conflitto fra di loro e mette in atto processi contrastanti.

Questi risultati hanno portato a dieci raccomandazioni generali e a quattro liste di raccomandazioni specifiche per singoli *stakeholders*

²⁷ Data Seal of Approval Booklet
http://www.datasealofapproval.org/media/filer_public/2013/09/27/dsa-booklet_1_june2010.pdf

²⁸ Il documento è disponibile sul sito del progetto RECODE, e verrà presto tradotto in altre lingue: http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf

(finanziatori, enti di ricerca, gestori dei dati, editori). Vediamo nel dettaglio le dieci generali:

1. Sviluppare politiche allineate e comprensive in relazione all'accesso aperto ai dati della ricerca: le politiche devono essere coerenti con le priorità dei singoli Stati membri ma allineate ai documenti europei di riferimento, ovvero le Raccomandazioni del 2012 e le linee di Horizon 2020.²⁹
2. Assicurare un finanziamento adeguato all'accesso aperto ai dati della ricerca, in particolare per quanto riguarda la conservazione sul lungo periodo
3. Sviluppare politiche e iniziative che offrano un riconoscimento ai ricercatori che rendano disponibili in modo aperto dati di qualità
4. Identificare gli attori chiave e i networks di ricerca che possano lavorare in modo collaborativo alla creazione di un ecosistema sostenibile per i dati aperti, evitando la duplicazione di sforzi
5. Programmare la conservazione e la cura dei dati aperti nel lungo periodo e in modo sostenibile
6. Sviluppare soluzioni tecniche e infrastrutture comprensive e collaborative che consentano l'accesso e la conservazione dei dati aperti
7. Sviluppare standard scientifici e tecnici per dati di qualità
8. Richiedere l'uso di licenze aperte e armonizzate

²⁹ Raccomandazione della Commissione del 17 luglio 2012 sull'accesso all'informazione scientifica e sulla sua conservazione, <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32012H0417&rid=1>; Horizon 2020, documenti ufficiali: <http://ec.europa.eu/programmes/horizon2020/en/official-documents>

9. Affrontare in modo sistematico le questioni etiche e legali che nascono dall'accesso aperto ai dati della ricerca
10. Supportare la transizione verso i dati aperti della ricerca con adeguata formazione curricolare sia per i ricercatori sia per i tecnici che curano i dati.

Il progetto RECODE dedica particolare attenzione agli enti di ricerca in quanto sede primaria in cui vengono prodotti i dati. Anche se le politiche Open Access sono cresciute in via esponenziale, esse sono a tutt'oggi centrate quasi sempre sulle pubblicazioni, mentre sono ancora relativamente poche le politiche che trattano espressamente dei dati. Ecco dunque le raccomandazioni specifiche:

1. Sviluppare una apposita strategia istituzionale per i dati della ricerca, in cui l'accesso aperto sia la norma, collaborando strettamente con i ricercatori per intercettare specifiche esigenze di servizi e strutture e adeguarsi alle pratiche delle singole discipline;
2. Perseguire attivamente la collaborazione intra-istituzionale e inter-istituzionale per creare sia un ecosistema sostenibile sia le infrastrutture necessarie per l'accesso aperto e la conservazione a lungo termine dei dati: mettere a fattor comune le diverse competenze permette economie di scala
3. Includere l'accesso aperto a dati di alta qualità come un formale criterio premiante ai fini della progressione di carriera; a questo va collegata l'adozione di un sistema di metriche adeguato alla nuova realtà dei dataset
4. Curare la formazione e l'educazione dei ricercatori e dei tecnici per accrescere le competenze sulla gestione dei dati e su una ricerca fortemente basata sui dati
5. Creare consapevolezza sui vantaggi dei dati aperti per la ricerca e prevedere adeguati riconoscimenti

6. Supportare la comunità dei ricercatori con servizi di consulenza etica e legale.

Credo non si possa non concludere che, anche dal punto di vista professionale per la comunità legata alle biblioteche, il capitolo dei dati aperti rappresenti il futuro, e quindi una grande opportunità di sviluppo. Ma occorre attrezzarsi da subito, per evitare il “knowledge divide” con paesi più sensibili e quindi già più avanti nel percorso.

References

ICSU - International Council for Science. 2014. *Open access to scientific data and literature and the assessment of research by metrics*

<http://www.icsu.org/general-assembly/news/ICSU%20Report%20on%20Open%20Access.pdf>

LERU - League of European Research Libraries. 2014. *Roadmap for Research Data*.

http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf

Lin, Jennifer and Carly Strasser. 2014. “Recommendations for the Role of Publishers in Access to Data”. *PLoS Biology* 12(10):e1001975. doi:10.1371/journal.pbio.1001975. Accessed January 28, 2015, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4211645/>

OECD. 2007. *Principles and Guidelines for Access to Research Data from Public Funding*. <http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>

Research Data Alliance. 2014. *The Data Harvest: How sharing research data can yield knowledge, jobs and growth*. <https://europe.rd->

alliance.org/documents/publications-reports/data-harvest-how-sharing-research-data-can-yield-knowledge-jobs-and

Royal Society. 2012. *Science as an open enterprise*.
<https://royalsociety.org/policy/projects/science-public-enterprise/Report/> .

Schumpeter, Joseph. 1947 [3rd ed]. *Capitalism, Socialism and Democracy*. Part II, Chapter VII, p.83. New York : Harper.

Strasser, Carly. 2013. "Closed data...excuses, excuses". *Data Pub Blog*, April 24. <http://carlystrasser.net/closed-data-excuses-excuses/>

Pollock, Rufus. 2007. Blog post. Accessed January 28, 2015, <http://rufuspollock.org/misc/>

ELENA GIGLIA, Università degli Studi di Torino.
elena.giglia@unito.it.

Giglia, E. "Accesso aperto ai dati della ricerca come vettore per la scienza aperta". *JLIS.it*. Vol. 6, n. 2 (May 2015): Art: #11130. DOI: 10.4403/jlis.it-11130

ABSTRACT: The two-days conference "Open Access to research data as a driver for Open Science" (Athens, 15-16 January 2015), as the final event of the RECODE project (Policy RECommendations for Open Access to Research Data in Europe) has been a unique occasion to think about Open Access to research data as a way to move towards Open science.

High-level speakers and an active audience have discussed on values, risks, opportunities, open issues, best practices; the aim of this paper is to deal with the complexity of these issues and to give a taste of the importance of open data in the near future.

KEYWORDS: *open data, open access, access to research data, policies, data management plan*

Submitted: 2015-02-03

Accepted: 2015-02-12

Published: 2015-03-29

