

OPEN DATA: ORIGINI E PROSPETTIVE

Andrea Rossato

1. Free Culture, Open Data

Le origini di ciò che va sotto il nome di *Open Data* sono da rintracciarsi in quel movimento, di intellettuali e di attivisti, che andò germogliando attorno all'idea che il software dovesse essere *libero*, o quanto meno *aperto*.

Si sono brevemente descritti, nelle pagine che precedono, i primi sussulti di questo movimento e se ne sono mostrate le connessioni con i mutamenti istituzionali che portarono all'estensione della tutela offerta dal *copyright* ai programmi per elaboratore elettronico sul finire degli anni settanta.¹ Si è insistito sulla figura di Stallman, senz'ombra di dubbio una delle personalità di maggior spicco nel definire i fondamenti etici e giuridici del movimento,² ma non si devono dimenticare altre esperienze, anch'esse legate principalmente all'ambito accademico, che hanno concorso ad approntare il quadro giuridico che rende possibile la creazione di quell'appartenenza collettiva, o *digital commons* che il software libero rappresenta.

Un altro centro nevralgico nella fondazione del movimento è infatti rappresentato dalla *University of California at Berkeley* ove prese vita un gruppo di ricerca che finirà per produrre programmi, collettivamente noti come *Berkeley Software Distribution* (BSD), che verranno distribuiti con licenze molto permissive e finiranno con il costituire un'intera famiglia di sistemi operativi interamente liberi.³

¹ Mi riferisco al mio capitolo *Sulla natura dei beni comuni digitali* nel presente volume.

² Si veda R. STALLMAN, *Free Software, Free Society: Selected Essays*, Boston, 2002, <<http://www.gnu.org/philosophy/fsfs/rms-essays.pdf>> (consultato: ottobre 2013).

³ Sul punto si veda M.K. MCKUSICK, *Twenty Years of Berkeley Unix. From AT&T-Owned to Freely Redistributable*, in C. DiBONA, S. OCKMAN, M. STONE (a cura di),

La principale differenza tra l'esperienza di Stallman e della sua creatura, la *Free Software Foundation*, e quella californiana, oltre che sui fondamenti etico-politici che caratterizzerebbero il software libero, attiene al quadro giuridico all'interno del quale i programmi vengono distribuiti: la versione, che potremmo definire *forte*, del software libero di Stallman è saldamente incentrata sul concetto di *copyleft*, in virtù del quale ogni opera derivata deve essere distribuita mediante la medesima licenza dell'opera originale. Il software appartenente all'ambiente BSD, di contro, è distribuito con una licenza molto liberare che consente di produrre, a partire da quello, opere derivate cosiddette proprietarie, distribuite, cioè, con licenze tipiche del mondo commerciale che negano agli utenti, ad esempio, la possibilità di redistribuire il programma o di trarne, a loro volta, ulteriori opere derivate.

È questo modello pluralistico di modalità distributive anche molto differenti tra loro che l'iniziativa *Creative Commons* cerca di sussumere in un apparato di licenze modulare e ricomponibile: attribuzione o meno della paternità (pubblico dominio); possibilità o meno di consentire riutilizzi commerciali; possibilità o meno di permettere la creazione di opere derivate; *copyleft* o non *copyleft* (*ShareAlike* nella terminologia *Creative Commons*). Ogni singola opzione può essere combinata in modo indipendente con le altre, dando quindi la possibilità di creare un elevato numero di licenze che possano adattarsi ad ogni tipo di esigenza distributiva.

Il quadro giuridico creato nell'ambito del movimento del software libero e sistematizzato all'interno dell'iniziativa *Creative Commons* aveva quale termine di riferimento il diritto d'autore e il copyright, ma il suo successo non poteva che ripercuotersi anche in settori contigui quali, appunto, quello dei dati e delle basi e banche di dati.

L'esigenza di una disciplina giuridica di riferimento che potesse adattarsi alle caratteristiche tipiche delle banche dati, e al loro specifico inquadramento ad opera del diritto, incominciò ad avvertirsi, ad esempio, all'interno di quelle comunità che, sull'esempio di *Wikipedia*, andavano costruendo in modo collaborativo basi di dati – si pensi al-

Open Sources: Voices from the Open Source Revolution, Cambridge, Mass., 1999, <<http://www.oreilly.com/catalog/opensources/book/toc.html>> (consultato: ottobre 2013).

l'esperienza di *OpenStreetMap*. In mancanza di licenze apposite si riutilizzavano licenze create per la scrittura collaborativa di programmi per elaboratore o di veri e propri testi, quali appunto quelli predisposti per la celebre enciclopedia.

Il problema di un tale approccio, però, attiene alla natura specifica dei dati e delle loro banche: se, in taluni specifici casi, la banca dati e i singoli dati che la compongono, sono soggetti alla disciplina del diritto d'autore, le cose non sempre stanno in questi termini: i dati geografici, e le mappe che li esprimono, sono, ad esempio, difficilmente inquadrabili nelle categorie di opere intellettuali protette, mancando essi di quel requisito di originalità necessario per accedere alla tutela e rappresentando anzi la quintessenza di ciò che s'intende con la parola *fatto*⁴. Si aggiunga la disciplina cui essere soggiaciano nell'ambito dell'Unione europea, ove la Direttiva 96/9/CE del Parlamento europeo e del Consiglio, dell'11 marzo 1996, relativa alla tutela giuridica delle banche dati istituisce il diritto cosiddetto *sui generis* che, sino alla versione 4.0 delle licenze *Creative Commons*, non veniva disciplinato da alcun tipo di licenza libera che non fosse stata appositamente predisposta.⁵

La notevole forza attrattiva del modello libero ed aperto, come potremmo definirlo, che dal software libero si espande prima alle altre opere dell'intelletto per il tramite dell'iniziativa *Creative Commons* e poi contagia quasi ogni ambito connesso alla c.d. proprietà intellettuale, dalla produzione e pubblicazione scientifica alla difesa del sapere tradizionale (*traditional knowledge*), se da un lato condiziona la scelta del quadro giuridico all'interno del quale avvengono alcune esperienze spontanee di creazione collaborativa di basi di dati, dall'altro induce anche ad una riflessione su quale debba essere lo *status* dell'enormità di dati prodotti dalle pubbliche amministrazioni.

⁴ Nell'ordinamento statunitense la protezione di una banca dati è soggetta, per via di una giurisprudenza consolidata, al requisito dell'originalità dei dati o della loro organizzazione. Il *leading case* in materia è rappresentato da *Feist Publications, Inc., v. Rural Telephone Service Co.*, 499 U.S. 340 (1991).

⁵ Sul punto rimandiamo al capitolo della presente opera di Thomas Margoni, *Open Government Data. Dati aperti nel settore pubblico: quadro normativo, licenze, ed esempi*.

Ciò che si suol definire *Government Data*, il prodotto della raccolta, dell'organizzazione e della conservazione di dati nel settore pubblico, avviene ad opera della pubblica amministrazione nell'esercizio delle proprie funzioni e queste, unitamente ad ogni loro traccia, dovrebbero, in virtù di un generale interesse alla trasparenza, poter essere soggette ad ogni forma di ispezione e controllo da parte della collettività. Inoltre, date la loro origine e la loro titolarità, si può agevolmente argomentare come essi dovrebbero essere liberamente riutilizzabili proprio da parte di coloro che per la loro generazione hanno pagato. In altri termini non si può negare come essi possano essere visti come una forma di appartenenza collettiva, una sorta di bene comune per via del fatto di essere stato creato mediante il ricorso a risorse generate dalla fiscalità generale.

2. *Le motivazioni dell'apertura dei dati pubblici*

L'analisi economica tradizionalmente spiega la necessità di un istituto quale il diritto d'autore mostrando come la creazione intellettuale debba essere incentivata mediante la predisposizione di una sorta di monopolio legale che permetta di combattere i problemi tipici della creazione e produzione dei beni pubblici.⁶ Con «bene pubblico» non si allude tanto a una categoria giuridica quanto, invece, a una nozione puramente economica: essi sono quei beni che presentano un consumo non rivale e un elevato costo di esclusione, caratteristiche che si ritiene siano possedute dalle opere dell'intelletto.

Più di recente taluno ha posto in dubbio una siffatta ricostruzione: Lemley, ad esempio, ritiene la proprietà intellettuale una forma di internalizzazione, per il mezzo del diritto, di esternalità puramente positive, e rigetta pertanto l'idea secondo la quale una tale categoria di beni dovrebbe essere protetta dai fenomeni di *free riding*.⁷

⁶ W.M. LANDES, R.A. POSNER, *An Economic Analysis of Copyright Law*, 18 *J. Legal Stud.* 325 (1989); ID., *The Economic Structure of Intellectual Property Law*, Cambridge, Mass., 2003.

⁷ M.A. LEMLEY, *Property, Intellectual Property, and Free Riding*, 83 *Tex. L. Rev.* 1031 (2004).

Anche rifiutando l'impostazione di Lemley, il quale nota come nell'ambito di una teorica proprietaria l'internalizzazione delle esternalità positive non rientri mai tra compiti che si ritengano propri dell'ordine giuridico, nel caso dei dati prodotti dalle pubbliche amministrazioni la necessità di incentivarne la produzione mediante forme di esclusività non trova fondamento alcuno. Si pensi, ad esempio, al fatto che tutte le opere dell'intelletto prodotte dal governo statunitense e da tutte le sue agenzie ricadono, *ex lege*, nel pubblico dominio.⁸

Non solo: il progredire delle tecnologie informatiche ha fatto maturare la consapevolezza che molti dei dati prodotti dalle pubbliche amministrazioni potrebbero essere riutilizzati, ricombinati ed elaborati al fine di fornire servizi con un elevato valore aggiunto. Si pensi a come i dati meteorologici, combinati con dati sulla viabilità e dati geografici possano fornire preziose informazioni ai viaggiatori. Gli esempi potrebbero essere innumerevoli.⁹

Con *Open Data*, e *Open Government Data* nel settore pubblico, ci si intende quindi riferire alla possibilità di dare vita a una sorta di ambiente informativo in cui i dati siano liberamente riutilizzabili al fine di creare, ad opera di soggetti privati, servizi che le pubbliche amministrazioni in genere non forniscono. Si tratterebbe, pertanto, di una forma di valorizzazione di questa specifica categoria di *asset* a titolarità pubblica.

Un tale obiettivo non rappresenta invero l'unica motivazione adducibile per l'apertura dei dati detenuti dalle pubbliche amministrazioni: in suo favore si possono anche citare esigenze di trasparenza, di verifica e controllo delle attività pubbliche, quel che, ancora attingendo a quel complesso di fenomeni culturali che trovano la loro origine nei movimenti del software libero, va sotto il nome di *Open Government*.¹⁰

⁸ 17 U.S.C. § 105.

⁹ Sul punto si veda il capitolo finale *Linked Open Trentino* nel presente volume.

¹⁰ Si veda D. ROBINSON, H. YU, W.P. ZELLER, ET AL., *Government Data and the Invisible Hand*, 11 *Yale J. L. & Tech.* 160 (2009); si vedano inoltre H. YU, D.G. ROBINSON, *The New Ambiguity of "Open Government"*, 59 *UCLA L. Rev. Disc.* 178 (2012); J. WEINSTEIN, J. GOLDSTEIN, *The Benefits of a Big Tent: Opening Up Government in Developing Countries: A Response to Yu & Robinson's The New Ambiguity of "Open Government"*, 60 *UCLA L. Rev. Disc.* 38 (2012); T. PEIXOTO, *The Uncertain Relation-*

L'esigenza di rendere disponibili e riutilizzabili i dati è avvertita anche in ambito scientifico, ove la condivisione dei risultati prodotti dalle attività di ricerca è il presupposto di ogni ulteriore progresso scientifico – e, quindi, tecnologico. La protezione dei dati, in tal caso, può condurre ad una vera e propria *tragedy of the anticommons*¹¹.

Il tema degli *Open Data* in ambito scientifico, e con specifico riferimento alle ricerche finanziate dal settore pubblico, è stato esplicitamente affrontato anche dalla *Organisation for Economic Co-operation and Development* la quale, nel 2007, ha adottato gli *OECD Principles and Guidelines for Access to Research Data from Public Funding*¹² Tali principi e linee guida, che facevano seguito ad una *Declaration on Access to Research Data from Public Funding* formulata nell'ambito dell'Organizzazione il 31 gennaio 2004,¹³ muovono dalla constatazione di come l'accesso «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».

Vi è ovviamente una relazione stretta tra i dati scientifici aperti e la più generale questione dell'*Open Access*, dell'accesso aperto, cioè, ai prodotti della ricerca scientifica.¹⁴ Un esempio particolarmente evidente

ship Between Open Data and Accountability: A Response to Yu and Robinson's The New Ambiguity of "Open Government", 60 UCLA L. Rev. Disc. 200 (2013).

¹¹ Il problema degli *open data* in ambito scientifico è stato al centro della discussione, nel 2006, della *Commons in Science Conferenze*, intitolata *Creating a Vision for Making Scientific Data Accessible Across Disciplines* <http://www.spatial.maine.edu/icfs/>. Il problema della relazione tra protezione dei dati e *anticommons* è stato affrontato nella relazione di Paul David (<http://www.spatial.maine.edu/icfs/slides/David.pdf>). Si veda anche il paper presentato da Tom Dedeurwaerdere, *The Institutional Economics of Sharing Biological Information*, 2006 (<http://www.spatial.maine.edu/icfs/Dedeurwaerdere-Bio%20research.doc>). Sulla tragedia degli anti-comuni in generale si veda, ovviamente, M.H. HELLER, *The Tragedy of the Anticommons: Property in the Transition From Marx to Markets*, 111 *Harv. L. Rev.* 621 (1998).

¹² Disponibili al seguente indirizzo: <http://www.oecd.org/science/sci-tech/38500813.pdf>.

¹³ C(2004)31/REV1.

¹⁴ Per una trattazione dettagliata rinviamo ovviamente al capitolo di R. CASO, *Scientific knowledge unchained: verso una policy dell'università italiana sull'Open Access*, nel presente volume. Si vedano inoltre R. Caso (a cura di), *Pubblicazioni scientifi-*

è rappresentato da quell'iniziativa promossa dal governo statunitense e che va sotto il nome di *PubMed Central* (PMC).¹⁵ PMC è un archivio digitale dello *U.S. National Institutes of Health* (NIH) che consente l'accesso al testo completo degli articoli scientifici nel campo biomedico e delle scienze della vita. La pubblicazione di questi articoli in PMC avviene in genere su base volontaria, ma è obbligatoria nel caso di ricerche che siano state finanziate dal NIH. Simili iniziative sono state adottate anche in altri paesi e sono state raggruppate in *PubMed Central International*, un network di *repository* che condividono le tecnologia sviluppata dal governo americano e che includono collegamenti a basi di dati aperti quali UniProt, una banca dati di sequenze proteiche liberamente accessibile; la Protein Data Bank, una banca dati aperta di strutture tridimensionali di molecole biologiche; e molte altre.

3. Open Data tra tecnologia e diritto

L'obiettivo dell'apertura dei dati, particolarmente quelli delle pubbliche amministrazioni, può essere perseguito unicamente nel caso in cui due condizioni siano soddisfatte: una di ordine tecnologico e l'altra di ordine più propriamente giuridico.

Per quel che concerne la condizione tecnologica essa attiene alla concreta riutilizzabilità dei dati, la quale dipende dai formati usati per la loro distribuzione e dalla presenza di metadati, di descrizioni, vale a dire, del contenuto dei dati stessi, che ne consentano una rielaborazione anche automatizzata.

Come si vedrà più dettagliatamente nelle pagine a seguire, *Linked Data* è una metodologia, fondata su tecnologie aperte e sviluppate nell'ambito di quell'iniziativa conosciuta come *web semantico*, imma-

che, diritti d'autore e Open Access: atti del convegno tenuto presso la Facoltà di Giurisprudenza di Trento il 20 giugno 2008, Trento, 2009, <<http://eprints.biblio.unitn.it/00001589>> (consultato: ottobre 2013); R. Caso, F. Pupppo (a cura di), *Accesso aperto alla conoscenza scientifica e sistema trentino della ricerca: atti del convegno tenuto presso la Facoltà di Giurisprudenza di Trento il 5 maggio 2009*, Trento, 2010, <<http://eprints.biblio.unitn.it/00001821/>> (consultato: ottobre 2013).

¹⁵ Accessibile a questo indirizzo: <http://www.ncbi.nlm.nih.gov/pmc/>.

ginata per svolgere la funzione di rendere disponibili, *on-line*, dati in formati e modalità che ne rendano possibile l'estrazione, l'interrogazione, il collegamento e il riuso, anche automatizzati¹⁶.

Per quel che concerne invece la condizione giuridica del riuso dei dati, a differenza di quanto avviene per le opere dell'intelletto protette dal diritto d'autore, in questo caso le cose sono più complesse.

Il *copyright*, quanto meno nel suo contenuto più squisitamente economico, è un istituto del diritto privato che, prima di ogni altro, è stato sottoposto ad un processo di globalizzazione che ne ha, per molti versi, reso omogenee le specifiche epifanie nazionali¹⁷. Se taluni dei suoi aspetti, quanto meno a livello declamatorio, possono apparire anche molto distanti – si pensi alle diversità fra *fair use* e *libere utilizzazioni* – a livello operativo le differenze tendono ad affievolirsi fin quasi a scomparire. Ciò spiega il successo di modelli distributivi fondati su licenze – quasi sempre scritte in lingua inglese e pensate ad opera di giuristi appartenenti alla tradizione di *common law* – utilizzate in quasi tutti gli ordinamenti del mondo.

Nel caso delle banche dati, come già visto, abbiamo invece delle differenze più marcate, specificamente tra ordinamenti appartenenti all'Unione europea e gli Stati Uniti. Se infatti negli USA le banche dati sono protette solo nel caso in cui i dati, o la loro organizzazione, siano qualificabili come *an original work of authorship*, in Europa il diritto *sui generis* protegge la banca dati, a prescindere dai requisiti che ne rendano possibile la tutela da parte del diritto d'autore, in quanto frutto di «un investimento rilevante sotto il profilo qualitativo o quantitativo»¹⁸.

Un tale diritto *sui generis*, specifico degli ordinamenti europei, non è stato in genere tenuto in considerazione nella stesura delle licenze

¹⁶ Per una introduzione si veda T. BERNERS-LEE, *Design Issues: Linked Data*, 2006, <<http://www.w3.org/DesignIssues/LinkedData.html>>; per un'introduzione alle tecnologie del web semantico correlate con *Linked Data* si veda A. ROSSATO, *Tendenze evolutive nello spazio digitale*, in G. PASCUIZZI (a cura di), *Diritto e tecnologie evolute del commercio elettronico*, Padova, 2004, 19-57.

¹⁷ A tal proposito si veda J. BRAITHWAITE, P. DRAHOS, *Global Business Regulation*, Cambridge, 2000.

¹⁸ Si vedano i *considerando* 7, 8 e 9 e l'art. 7 della Direttiva 96/9/CE.

libere utilizzate per la distribuzione di software o altre opere dell'intelletto, quali quelle nell'ambito dell'iniziativa *Creative Commons*. Solo nella versione 3.0 di quest'ultime, e unicamente nella traduzione nelle lingue dei paesi appartenenti alla UE, il diritto *sui generis* era menzionato, ed unicamente al fine di rinunciarvi completamente e senza sottoporre la rinuncia ad alcuna condizione.¹⁹ Le cose, come vedremo più dettagliatamente nelle pagine a seguire, sono cambiate con la successiva versione 4.0.

La mancanza di una specifica disciplina del diritto *sui generis*, lì ove fosse disponibile, nelle principali licenze libere nate nell'ambito di ciò che molti chiamano *Free Culture* ha indotto taluni a creare delle licenze specifiche. Un esempio è rappresentato dalla *Open Data Commons Open Database License* (ODbL), utilizzata nel progetto collaborativo *OpenStreetMap*.

Un simile approccio presenta però alcune difficoltà che attengono alla proliferazione delle licenze e al correlato problema della loro reciproca compatibilità, fenomeno già conosciuto e lungamente dibattuto nell'ambito del software libero ed *open source*. Molto spesso i problemi di compatibilità non sono reali, anche se in taluni casi, per quanto rari, essi si possono presentare ed essere molto difficili da risolvere. In ogni caso la presenza di licenze differenti e il problema della loro compatibilità nel caso dei dati, che dovrebbero essere non solo riusabili ma anche componibili per dare origine a servizi fondati sull'aggregazione di banche dati diverse, aumenta i costi di ricerca sulle condizioni giuridiche del loro uso. L'incertezza che ne deriva può avere un effetto disincentivante.

¹⁹ Ad esempio, nella versione italiana della licenza CC BY-NC-ND 3.0, si legge: «[q]ualora l'Opera concessa in licenza includa una o più banche dati sulle quali il Licenziante è titolare di un diritto sui generis ai sensi delle norme nazionali di attuazione della Direttiva 96/9/CE sulle banche dati, il Licenziante rinuncia a far valere il diritto corrispondente» (<http://creativecommons.org/licenses/by-nc-nd/3.0/it/legalcode>).

