

Dati aperti con LODE

Antonella De Robbio e Silvia Giacomazzi

Abstract

Unlike the case of Open Access and Open Source movement, in the Open Data context there is no general agreement regarding the definition of the term. The Open Data Movement, like other Open philosophies, aims the goal of making immediately "accessible to anyone the primary research data, without limitation of copyright, patent or other control mechanisms". In this work we try to analyze some interesting statements made by different stakeholders operating in the context of the Open Data, in order to clarify a system of behaviors, often subject to ambiguity and sometimes manipulation. The experts are divided in considering Open Data a derivation of Open Government or a derivation of Open Science, but in both cases the context is the digital environment. Between these two Open territories, we could place a third place, that connects them and that is called Open Bibliographic Data or OpenBiblio: this middle place includes OPACs, databases and open archives. Speaking of Open Data means to place the defining elements within the framework of the Open Knowledge, in a dimension of knowledge that must analyse also the issue of Open Data licenses, which are different from the best-known Creative Commons. The Open Data shares its objectives and developments, in particular with reference to the Open Linked Data project, which deals with interoperability between different data sources and formats. Hence the title chosen for this work: Open Data con LODE, LinkedOpenData, where the final "e", for us, means something immersed, "embedded" within the system itself. The real meaning belongs to the context of the Open Library, and in particular belongs to the Recommendations LODE BD, drawn up by the AIMS team of FAO, and it applies the concept of LOD-enabled bibliographic data, where the "e" has the meaning of something "enabled". It's not sufficient to apply a license in order to make a resource "open", but the resource must also be "open" in terms of real and effective interoperability, avoiding the use of licenses only as something fashionable. In this work our aim is not to be completely exhaustive in the definition process, but our purpose is simply to analyse the issue within a framework that seems sometimes rich as well as chaotic.

Why are libraries around the world devoting time and resources to releasing their bibliographic data under an open licence? What's in it for them and what are the costs and practical issues involved? [...]
One of the possibilities that open bibliographic data offers is the chance for libraries and indeed anyone to reuse the data to build innovative services for researchers, teachers, students and librarians.

JISC Open Bibliographic Data Guide

<http://obd.jisc.ac.uk/>

1. Introduzione all'Open Data

A differenza di quanto accade per i movimenti Open Access o Open Source, per i quali sono state formulate definizioni incluse in dichiarazioni condivise a livello internazionale e che – per certi

versi - sono entrate ormai nel linguaggio comune, per l'Open Data non vi è ancora un accordo generale nella definizione del termine. Il movimento Open Data, al pari di altre filosofie "open", si propone l'obiettivo di rendere immediatamente "accessibili a chiunque i dati primari della ricerca, senza limitazione di copyright, brevetti o altri meccanismi di controllo". [1] Una buona definizione di partenza di cosa siano l'Open Data, l'Open Content, e l'Open Information è stata formulata dall'Open Knowledge Foundation [2].

Con questo lavoro tentiamo di analizzare alcune formulazioni interessanti provenienti da attori diversi che operano nei contesti dell'Open Data, anche per chiarire un sistema di comportamenti, soggetto ad ambiguità e talvolta a strumentalizzazione. Il duplice approccio definitorio rispecchia i due rami dai quali si origina l'attuale assetto dell'Open Data nel suo complesso, che comunque è in continua evoluzione. Da una parte lo vediamo incanalato entro un assetto di trasparenza amministrativa, strettamente connesso con forme di cittadinanza attiva o democrazia diretta; dall'altra lo possiamo collocare entro una sempre più ampia apertura scientifica, oltre i territori dell'Open Access. Gli esperti in materia, infatti, si dividono nel considerare l'Open Data come derivazione dell'Open Government o dell'Open Science, ma in entrambi i casi la sfera è quella del contesto digitale. Parliamo di cittadinanza di rete e conseguente fruizione di dati Open, per i dati sull'impatto ambientale, o di e-Science laddove la comunicazione scientifica usa i dati grezzi per condividere dati (in particolare di laboratorio) al fine di raggiungere risultati mirati ad uno specifico contesto geografico e in breve tempo.

Nel mezzo di questi due territori Open, potremmo collocare un terzo luogo, che li collega entrambi e che comunque riguarda strettamente la meta informazione e che viene definita Open Bibliographic Data o anche OpenBiblio: questa terra di mezzo comprende i cataloghi delle biblioteche (OPAC), i database e le basi di conoscenza e gli archivi aperti.

Parlare di Open Data significa collocare gli elementi definitivi entro il quadro del movimento Open Knowledge, in una dimensione di conoscenza aperta che non può prescindere da un'analisi consapevole delle licenze Open Data, diverse dalle più note Creative Commons. L'Open Data infatti ne condivide gli obiettivi, gli sviluppi, in particolare in riferimento al progetto Open Linked Data, che si occupa dell'interoperabilità tra dati di diversa origine e formato. Di qui il titolo scelto per questo lavoro: *Open Data con LODe*, LinkedOpenData, appunto, con la "e" finale che per noi sta a significare qualcosa di immerso o incorporato entro il sistema stesso, e quindi nel senso di *embedded*. Il significato reale lo si trova invece entro il contesto Open Biblio, e in particolare nelle *Raccomandazioni LODe BD* redatte dal team AIMS della FAO [3], e ricomprende il concetto di *LOD-enabled bibliographic data*, dove la "e" assume il significato di qualcosa di "abilitato". Non basta applicare una licenza per rendere "open" una risorsa, ma occorre anche "aprirla" in termini di interoperabilità reale ed effettiva per evitare che l'uso di licenze sia solo di facciata, quasi una moda.

In altri termini, come vedremo in seguito, il concetto di LOD-enabled bibliographic data si riferisce a dati e servizi intesi come *data e service provider* i quali presentano numerose questioni specifiche correlate alle strategie di codifica a vario livello, ad esempio:

- quali standard di metadati devono essere seguiti al fine di pubblicare i dati collegati?
- qual è l'insieme minimo di proprietà che un insieme di dati dovrebbe includere per essere significativo entro un contesto di condivisione?
- c'è un modello di metadati o profilo di applicazione che può essere direttamente adottato per la produzione di dati (in particolare dal nostro database locale)?
- se il vocabolario controllato che abbiamo usato è disponibile come Linked Data, quale modello di valore dovremmo scambiare tramite il nostro repository, in particolare, le forme letterali rappresentano un concetto oppure l'URI identifica il concetto?

- come dovrebbe essere codificato il nostro dato di modo che possa migrare da un database locale a un insieme di dati di tipo LinkedData?

Non ci proponiamo con questo lavoro di essere esaustivi nel trattare ambiti e definizioni, ma il nostro scopo è semplicemente quello di sviscerare la questione entro una realtà che appare a tratti ricca quanto caotica.

2. Introduzione al concetto di dato verso i territori dell'Open Data

Prima di immergerci nei paesaggi e territori dell'Open Data è necessaria una piccola digressione sul concetto stesso di "dato", spesso sottovalutato nella sua vastità e confuso con il concetto di informazione. I dati - diversi dalle informazioni - devono essere analizzati, organizzati, strutturati ed elaborati per divenire informazione [4]. L'informazione può essere dunque definita come lo scarto di qualità dal dato grezzo, il quale deve anche essere strutturato secondo una sintassi che lo renda interpretabile. Gregory Bateson [5] dice che "l'informazione è una differenza che fa differenza" [6]. Una goccia di pioggia che cade sul terreno di per sé non contiene, per noi, informazione, è un semplice dato, ma se questa goccia ci cade sul naso l'informazione diviene misura d'effetto, dice Bateson. Riprendendo la definizione assodata:

Un dato (dal latino datum che significa letteralmente fatto) è una descrizione elementare, spesso codificata, di una cosa, di una transazione, di un avvenimento o di altro. L'elaborazione dei dati può portare alla conoscenza di un'informazione. Ogni tipo di dato dipende dal codice e dal formato impiegati. I dati possono essere conservati su diversi mezzi (o supporti) fisici: cartaceo, magnetico (floppy disk), ottico (CD) ed altri supporti. I dati possono presentarsi sotto diverse forme: numero, lettera dell'alfabeto, immagine, suono, ed altri. In informatica, il termine indica un valore che può essere elaborato e/o trasformato da un automa o meglio da un elaboratore elettronico. Il dato rappresenta l'oggetto specifico su cui interviene l'esecutore dell'algoritmo. Le memorie di massa consentono di salvare i dati in modo permanente. Il processo di registrazione dei dati in una memoria si chiama memorizzazione o archiviazione. I dati possono essere conservati in file o nei database. [7]

La distinzione tra dato e informazione apparentemente sembra chiara: il dato si ferma a costituire la materia prima che, una volta elaborata, può generare un'informazione ben strutturata. I dati si diversificano tra loro per tipologia di codificazione, per la loro sintassi, formato e supporto di archiviazione-conservazione. Sempre secondo Bateson, tra i dati, possono essere individuati due livelli di valore: esistono dati semplici/elementari, che fanno parte della quotidianità (segno, gesto, parola, numero) e dati che appartengono ad un grado di strutturazione più elevato, sono composti di dati semplici. Potremmo individuare alcune tipologie di dato a seconda dell'ambito di ricerca dal quale esso deriva, dalla sua formulazione e/o codifica, dalla funzione e dalla tipologia del contenitore che lo ospita, ma per lo scopo del presente lavoro possiamo raggrupparli, semplificando, in quattro categorie:

- dato grezzo;
- dato di ricerca (risultato);
- dato fattuale (formule matematiche, chimiche, dati ambientali...);
- metadato (i dati di un database, dato bibliografico, etc.).

Il dato grezzo deriva dalla misurazione strumentale di un fatto, che viene registrato e conservato per poi essere interpretato in maniera tale da creare informazione e significati. Il dato grezzo è ciò che costituisce la conoscenza prima di ogni elaborazione. Un dato grezzo è la mappatura di un territorio, la registrazione di un'immagine, ecc. Il dato di ricerca è il risultato di un'osservazione sperimentale, raccolto e archiviato per essere successivamente analizzato e interpretato. Il dato fattuale si presenta sotto forma di formule matematiche, chimiche, biologiche ed è frutto della realtà di fatto codificata secondo criteri particolari. Il metadato, secondo l'etimologia stessa della parola, va oltre il dato e si

costituisce come un dato strutturato che descrive altri dati, al fine di reindirizzare l'interrogatore verso l'identificazione dell'informazione che va ricercando. Sono dati relativi ad altri dati (dati-oggetto), sono meta informazioni, tipici del contesto OpenBiblio.

3. Definizioni e origine dell'Open Data

Alcune scuole di pensiero ritengono che l'Open Data derivi dalla dottrina dell'Open Government [8], la quale prevede dati aperti nella Pubblica amministrazione, ai fini di una più ampia trasparenza dei dati pubblici, ma soprattutto nell'ottica di una democrazia partecipata dei cittadini raggiungibile attraverso l'uso delle nuove tecnologie della comunicazione [9]. Nel presente lavoro non ci occuperemo di questo fronte - se non per un breve paragrafo relativo alla situazione in Italia che tratta anche di Open Government- in quanto esiste già dell'ottima letteratura a riguardo.

Altre scuole di pensiero ritengono invece che l'Open Data scaturisca dalla ricerca scientifica, in particolare dalle scienze cosiddette dure e quindi di chiara derivazione dal movimento dell'open acces. A seguito dei movimenti entro le comunità scientifiche che si sono auto-organizzate nel web 2.0, numerosi sono stati gli scienziati, in particolare in ambito biomedico, convinti dei benefici di un approccio "aperto" e partecipativo ai risultati della ricerca (sia positivi, sia negativi). Uno dei pionieri dell'Open Data è Peter Murray-Rust, biochimico all'Università di Cambridge che traccia una interessante distinzione tra il concetto di Open Data nelle "small science" e "large science":

Si riconosce che la pratica nella pubblicazione e riutilizzo dei dati varia notevolmente tra le diverse discipline. Alcune di queste, come le bioscienze, hanno una lunga tradizione nel richiedere dati da pubblicare e aggregare in banche dati a finanziamento pubblico. Altre discipline, o "scienze grandi", hanno ben sviluppato politiche del riutilizzo dei dati e richiedono questi ultimi da telescopi, satelliti, acceleratori di particelle, sorgenti di neutroni, ecc., affinché divengano universalmente disponibili per il riuso. In queste aree le norme sono spesso sufficienti per la pratica dell'accesso aperto ai dati. Nelle "scienze piccole" invece l'unità di ricerca è il laboratorio o l'individuo (con il termine "scienze piccole" non si intende riflettere l'importanza della disciplina che può essere numericamente vasta). Queste discipline tipicamente sfociano in molte pubblicazioni indipendenti che riportano esperimenti individuali (*hypopublication* è un neologismo che esprime la natura disarticolata di tali informazioni). [10]

Murray-Rust insomma, nel definire l'Open Data, chiede di considerare il fatto che il riutilizzo dei dati può variare molto a seconda del campo di ricerca dal quale tale movimento prende vita. Il riutilizzo dei dati è infatti più diffuso nelle cosiddette "scienze ampie" rispetto a quelle "piccole". Queste ultime fanno riferimento al singolo laboratorio e quindi molto spesso danno luogo a disgregazione e frammentazione più che aggregazione e condivisione.

Anna Gold, a proposito di dati, afferma che "i dati sono la moneta della scienza, anche se le pubblicazioni sono ancora la valuta di possesso. Per essere in grado di scambiare dati, comunicarli, estrarli, riutilizzarli, e revisionarli è essenziale per la produttività scientifica, collaborazione e scoperta". [11] Seguendo il filone che ritiene il movimento Open Data di ispirazione Open Science, Anna Gold lo descrive come essenziale al progresso scientifico stesso. La motivazione è presto data nella definizione di Wikipedia, dove si legge: "con Dati aperti [...] si fa riferimento ad una filosofia che è al tempo stesso una pratica. Essa implica che alcune tipologie di dati siano liberamente accessibili a tutti, senza restrizioni di copyright, brevetti o altre forme di controllo che ne limitino la riproduzione". [12]

4. La definizione dell'Open Data Foundation

L'Open Data Foundation (OdaF) [13] si spinge oltre nella definizione, rintracciando le caratteristiche essenziali dell'Open Data :

L'Open Data deve rispondere alle esigenze sia del produttore sia della comunità di utenti. Esso ha le seguenti caratteristiche: 1. Adatto per l'uso e scopo: uniforme, coerente, esaustivo, tempestivo, pertinente, fornendo risposte a domande reali, 2. Accessibile: può essere scoperto, accessibile e compreso, 3. Sicuro e protetto: l'accesso fisico ai dati effettivi è controllato e gli intervistati sono protetti in conformità ai principi e alle leggi statistiche, 4. Ben documentato: i dati sono accompagnati da elevata qualità dei metadati che soddisfa gli utenti occasionali, ricercatori e produttori. Cattura informazioni sufficienti per replicare il lavoro (standard replica) e si basa su standard accettati di qualità e completezza. [14]

Per l'ODaF dunque l'Open Data ricopre un ruolo che va in una duplice direzione, quella di chi pubblica il dato e quella di chi usufruisce dell'informazione grezza, con caratteristiche di precisione, coerenza, sicurezza. Senza queste caratteristiche che determinano la condivisione, l'universalità, l'apertura del dato informativo non si può parlare di dati aperti.

Partendo da questo presupposto è necessario soffermarsi sul significato stesso del termine "conoscenza aperta", che sembra sottostare alla definizione di Open Data, e che acquisisce, a seconda degli ambiti, diverse gradazioni e tonalità, in archi più o meno ampi d'azione. La definizione di "open knowledge" che viene tracciata dall'Open Knowledge Foundation (OKF o OKFN) [15] genera la gamma delle varie licenze "aperte" – delle quali parleremo più avanti - disponibili e previste entro l'Open Database Licence, gestito dal Open Data Commons [16].

OKFN partecipa inoltre a numerosi progetti tra i quali [17]:

- The Comprehensive Knowledge Archive Network
- Open Knowledge Definition
- Open Shakespeare
- Open Economics
- Open Text Book (Libri di testo aperti)
- Public Domain Works (Opere di Dominio Pubblico)
- Public Domain Calculators (Calcolatori di Dominio Pubblico)
- Open Knowledge Forums
- Information Accessibility Initiative
- Open Geodata
- Guide to Open Data licensing (Guida alle licenze di dati aperti)
- The annual Open Knowledge Conference.

Tali progetti mirano tutti all'apertura della conoscenza come viene intesa dalla OKF.

La versione 1.0 della definizione di conoscenza aperta [18] esplicita il termine "conoscenza" in queste tre tipologie di materiale informativo: i contenuti semplici (musica, film, libri), i dati (scientifici, storici, geografici, statistici, ecc.) e l'informazione tratta dal settore pubblico. Come viene specificato, questo esclude la questione dei software che rimanda ad altri movimenti Open. Il termine "opera" ricopre, nella definizione offerta, il ruolo di elemento di conoscenza che viene trasferito da chi pubblica a chi visualizza e riutilizza. Più opere aggregate possono costituire un "pacchetto informativo" che rappresenta in sé un'opera unica. Per l'OKF "un'opera è aperta se la relativa modalità di distribuzione soddisfa le seguenti condizioni:

1. Accesso
2. Ridistribuzione
3. Riutilizzo

4. Assenza di restrizioni tecnologiche
5. Attribuzione
6. Integrità
7. Nessuna discriminazione di persone o gruppi
8. Nessuna discriminazione nei settori d'attività
9. Distribuzione della licenza
10. La licenza non deve essere specifica per un pacchetto
11. La licenza non deve limitare la distribuzione di altre opere".

5. Open bibliographic data (OpenBiblio)

La corrente Open investe anche la questione dei dati bibliografici e il mondo bibliotecario. Se l'Open Data è generato da due filoni quasi paralleli, Open Government da una parte e e-Science dall'altra, le cui radici si situano nel mondo "open", l'Open Biblio Data è un anello di congiunzione che recupera filosofie, strategie e dimensioni "open" e connette mondi diversi mettendo al centro l'informazione come bene comune. Nasce così l'*Open Bibliographic Data Working Group* in seno all'*Open Knowledge Foundation*, nucleo del movimento [Open bibliography and Open Bibliographic Data](#), in breve Open Biblio. [19] Open Biblio mira a promuovere una cultura solidamente aperta al fine di ottenere, da chi produce meta-informazione in ambito biblioteconomico o delle scienze dell'informazione, dati bibliografici aperti e disponibili liberamente per l'uso e il riuso da parte di chiunque e per qualsiasi scopo.

I dati bibliografici in rete, che per definizione descrivono una risorsa fisica o digitale, richiede un URI, utile all'identificazione della risorsa nel Web, e un URL, utile alla sua localizzazione. Questa parte del dato bibliografico ne costituisce la base primaria, alla quale vanno a sommarsi altri dati amministrativi o legali (per esempio quelli sui diritti di proprietà) esterni al web. Per ottenere dati bibliografici aperti e facilmente condivisibili e riutilizzabili, sono stati formulati quattro principi fondamentali [20]:

1. quando si pubblicano dati bibliografici si deve usare una licenza esplicita e robusta;
2. è necessario usare una licenza o una deroga appropriata al tipo di dati;
3. se davvero si vuole che i dati prodotti siano realmente usati e arricchiti da altri, devono essere aperti nelle modalità definite dalla Open Definition; in particolare, non devono essere usate la clausola non-commerciale e altre clausole restrittive;
4. ove possibile, è preferibile porre esplicitamente i dati bibliografici nel Pubblico Dominio tramite licenza PDDL (Public Domain Data Licence) o CC0 (Creative Commons Zero).

Chi sottoscrive tali principi [21], come produttore o titolare di dati, deve aderire ad una dichiarazione esplicita di intenti in merito al riuso degli stessi, nonché la scelta di licenze appropriate ai dati che produce o per i quali detiene diritti.

In generale la questione Open Data non viene compresa in maniera adeguata. Le motivazioni possono essere principalmente due: o vi è un rifiuto di tutte le filosofie Open, tendenza proveniente dal fronte "conservatore" presente purtroppo anche nell'ambiente bibliotecario, o la filosofia Open

viene sfruttata come etichetta di facciata o talvolta anche strumentalizzata. In più c'è da sottolineare la completa confusione che si genera attorno alle correnti Open e in maniera ancor più accentuata attorno alle correnti Open Data. Si confonde quello che è "accessibile" nel senso che sta in rete con ciò che invece è riutilizzabile in modo più meno pieno, e l'esempio degli OpenData è emblematico. Un catalogo in rete, un OPAC è ovviamente accessibile a chiunque, ma ciò non significa che i suoi dati possano essere scaricati, in blocco o solo in parte, e riutilizzati per costruire nuovi cataloghi o rielaborati per esempio per scopi commerciali. Scaricare una bibliografia da un OPAC – funzione ormai comune a qualsiasi interfaccia di buon livello - a seguito di una ricerca catalografica, non significa che quel catalogo aderisca alla corrente OpenData.

Di aperture di cataloghi a tutto tondo se ne sta parlando ormai da un paio d'anni: nel blog di Bonaria Biancu [22] recuperiamo infatti un interessante post del gennaio 2009 focalizzato su un articolo del Guardian che comparava le nuove politiche di OCLC, delle quali parleremo in seguito, con la scelta operata dai Library Services dell'Università di Huddersfield, che rilasciano i dati di utilizzo del proprio patrimonio e li mettono disponibili gratuitamente online con Licenza Open Data Commons [23].

Se per un catalogo collettivo nazionale come SBN l'applicazione di un'eventuale licenza di tipo Pubblico Dominio come la PDDL o la CC0 [24] potrebbe anche essere fattibile, considerando che la corrente Open Government potrebbe aiutare a contestualizzarne il significato, e considerando che SBN è un catalogo davvero pubblico e condiviso, per altri cataloghi basati su ILS proprietari la questione potrebbe essere più complicata. I sistemi ILS commerciali si basano su ulteriori software spesso proprietari, come i SGBD (sistemi di gestione dei dati bibliografici), i quali sono ceduti dai produttori in licenza d'uso sulla base del numero di accessi contemporanei.

Ora, il fatto che non tutti abbiano chiara le differenze tra licenze diverse (per esempio tra una licenza PDDL o CC0 e quella ODC-BY [25]) è indicativo: molta strada si deve ancora fare non tanto in termini di "educazione" all'uso delle licenze corrette, ma piuttosto verso un cambiamento o meglio una trasformazione delle architetture mentali rispetto alla condivisione di dati, fatti, informazioni. Non si tratta di aspetti banali, perché le sfumature che presentano le licenze del mondo OpenDataCommons richiedono grande attenzione e una seria analisi previa su quali politiche si vogliano adottare per quel determinato strumento o servizio, sia esso un OPAC, un database o un archivio aperto (repository). Ma ancor prima è necessario agire sulla consapevolezza della reale portata che la filosofia OpenData comporta.

Alcune licenze vietano l'utilizzo dei dati per scopi commerciali. Se si dota un archivio aperto di una licenza che vieta gli usi commerciali, si impedirà l'indicizzazione da parte dei servizi cosiddetti database A&I (Abstracting and Indexing) che hanno lo scopo precipuo di indicizzare la letteratura scientifica di settore. I database A&I sono servizi prettamente commerciali, precludere il riuso dei dati di un archivio aperto significa limitare l'impatto scientifico delle produzioni intellettuali di quel determinato archivio, per ragioni talvolta puramente "ideologiche". Se per un catalogo il problema della indicizzazione di tipo A&I non si pone, per un archivio aperto (disciplinare o istituzionale) essere o non essere citati è un fattore di notevole rilevanza e di conseguenza non essere indicizzati significa non ottenere impatto citazionale. Per questo entro il contesto OpenBiblio è necessaria un'accurata analisi di tutte le categorie chiamate in causa: i cataloghi (OPAC), gli archivi aperti o repository (istituzionali o disciplinari), le banche dati e le basi di dati di conoscenza.

I repository hanno due layers: uno riguarda i metadati, l'altro i contenuti sul quale non possiamo a priori applicare licenze in quanto la titolarità dei diritti spetta agli autori che depositano nell'archivio. Il JISC ha sviluppato per conto della *Resource Discovery Taskforce* un utilissimo documento presente sul sito "Guide to Open Bibliographic Data", [26] focalizzato su casi o esempi

di applicazione possibili di Open Bibliographic Data come oggetto di studio. La guida cerca di chiarire in termini generali, ma anche entro specifici contesti, 17 studi di casi concreti di utilizzo di dati [27], considerando alcuni fattori comuni:

- come licenziare i dati;
- questioni legali da considerare;
- costi potenziali e eventuali risparmi;
- implicazioni pratiche in termini di processi, sforzi e abilità;
- formati dei dati e altre opzioni tecniche.

I casi coprono aspetti concreti che ciascuna biblioteca potrebbe già aver incontrato o che potrebbe pianificare per uno sviluppo dei servizi bibliotecari verso l'utenza. La guida nell'insieme fornisce un potenziale e razionale effetto di stimolo verso azioni concrete basate sull'OpenData, con esempi chiari e documentati [28]. Un altro sito che elenca esempi concreti di cataloghi e/o servizi bibliotecari che hanno aderito pienamente alla filosofia OpenData è quello fornito dal registro CKAN (Comprehensive Knowledge Archive Network) [29], dove spicca anche tutto il catalogo della biblioteca del CERN i cui dati sono stati licenziati per il pubblico dominio con licenza PDDL e CC0.

Se per i dati catalografici o per quelli dei repository la questione OpenData, seppur nelle complessità esposte sopra, è in qualche modo matura, per le banche dati l'OpenData è una partita tutta da giocare, o meglio una scommessa se non un azzardo dei prossimi anni. Difatti nelle banche dati, siano esse bibliografiche o fattuali, il problema di un'applicazione di licenze più o meno aperte si scontra con la questione squisitamente europea del *diritto sui generis*: laddove il produttore è prettamente commerciale, allora agisce la legge sul copyright delle banche dati, il noto *diritto sui generis* [30].

E' indubbio che la titolarità dei diritti - di tipo economico - di una banca dati sta in capo al suo costituente, che ha investito in termini di risorse, hardware e software, denaro, tempo, persone; ma se pensiamo a banche dati di conoscenza dove diversi sono gli attori e i soggetti che concorrono alla loro costituzione, allora la faccenda si complica. Di chi sono i diritti sui dati creati da dipendenti di enti di ricerca, università, istituzioni su piattaforme proprietarie che vengono incrementate sulla base di un lavoro collettivo seppur in un tempo differito? Mancano licenze chiare a proposito, e quindi è quasi azzardato proporre in questa fase licenze OpenData se non è chiaro il processo correlato alla titolarità.

6. Open data per il catalogo WorldCat: le politiche di OCLC

Emblematico è l'esempio del dibattito che si sta conducendo entro OCLC [31]: l'accesa discussione tra i membri del Consorzio è proprio in merito a quale licenza adottare sui dati bibliografici prodotti a seguito delle attività di catalogazione partecipata, ed è solo una conseguenza di decisioni politiche che devono essere attentamente vagliate a monte. [32]

Come segnala Andrea Marchitelli nel suo blog [33], dopo oltre vent'anni OCLC modifica le politiche di condivisione dei record di WorldCat e di conseguenza cambiano le condizioni d'uso dei record catalografici. Rispetto alle precedenti linee guida [34] il nuovo schema di policy [35] non è poi così rivoluzionario. Non vi è una reale apertura verso il pubblico dominio, ma sembra quasi un'operazione che mira "a formalizzare in una qualche forma di licenza le politiche di condivisione e riuso dei record". Il blog segnala anche un raffronto sinottico [36] tra le due versioni della policy, la vecchia e la nuova, preparato da Tim Spalding di LibraryThing, che aiuta a capire come alla fine il cambiamento sia stato solo "di facciata".

Ogni servizio o strumento bibliografico che deriva da attività catalografiche condivise - per esempio entro un catalogo collettivo - può avere peculiarità che possono comportare adozione di licenze differenziate, proprio perché differenti possono essere i gradi di *Openness* per quanto riguarda ciò che l'utente finale può fare o non fare. WorldCat [37], il più grande catalogo bibliografico del mondo, è uno dei tanti prodotti generati dall'attività catalografica di OCLC che comunque prevede anche altri servizi commerciali, tra cui prestito interbibliotecario, catalogazione derivata, reference, ecc... [38].

Sono poche ad oggi le biblioteche, come quelle afferenti all'Università del Michigan, che hanno iniziato a fornire record catalografici originali in MARC21 e/o in MARCXML sotto licenza Creative Commons Zero (CC0) [39]. Del resto recentemente OCLC sta prestando grande attenzione a questi sviluppi, di particolare interesse anche in seguito all'adozione delle nuove politiche sui dati di OCLC [40] – una sorta di codice di buone pratiche - che coprono non tanto e non solo lo stesso catalogo WorldCat, ma anche i metadati immagazzinati nei cataloghi bibliotecari, la cui fonte era il database bibliografico WorldCat.

Le politiche sui dati di OCLC intendono incoraggiare un ampio uso e riuso dei dati bibliografici di WorldCat, sostenendo una piena attuazione e un utilizzo anche a lungo termine dei servizi del catalogo come la presentazione, la catalogazione e la condivisione delle risorse. A tal fine, le politiche sui dati definiscono un quadro comportamentale di autogoverno orientato alla sostenibilità del modello WorldCat, garantendo l'erogazione dei suoi servizi, i risultati che tali servizi producono, e una sostenibilità dello stesso Consorzio come organizzazione nel corso del tempo.

Essenzialmente, la politica OCLC conferisce ampi diritti sia ai membri del Consorzio sia a soggetti terzi di utilizzare, riutilizzare, trasferire i metadati di WorldCat, mentre si richiede ai membri di OCLC di esercitare tali diritti nel contesto di precise responsabilità. In particolare, si chiede ai membri di rispettare le politiche, assicurando la consapevolezza delle clausole contenute, e facendo tutti gli sforzi ragionevoli in merito all'attribuzione dei dati al Consorzio, laddove opportuno, e di compiere ogni ragionevole sforzo per assicurare che il riutilizzo sia coerente con l'intento delle politiche, dei regolamenti delle comunità che gravitano in OCLC, ma soprattutto rispettando gli scopi pubblici che la rete bibliotecaria si prefigge dal suo sorgere [41].

A seguito delle numerose richieste a proposito delle licenze open data, negli ultimi mesi OCLC ha cominciato ad analizzare in modo approfondito la questione, che presentava non poche complessità. Alla fine OCLC ha optato per la licenza Open Data Commons Attribution (ODC-BY) [42], il cui approccio sembra, in generale, essere più calzante alla tipologia di dati di un catalogo bibliotecario rispetto ad altre licenze open. ODC-BY ha la caratteristica – determinata dall'etichetta BY che significa attribuzione – di richiedere espressamente che l'attribuzione della titolarità dei dati sia sempre mantenuta e rispettata, e quindi risulta perfettamente compatibile con gli obblighi delle clausole contenute nella sezione 3B delle nuove politiche sui dati. Inoltre, il framework della licenza ODC permette ad un insieme di regole collettive di essere collegate alla licenza stessa.

OCLC pensa che sia possibile elaborare un'implementazione della nota della licenza ODC-BY e suggerire ai membri del consorzio di voler rilasciare i dati dei loro cataloghi sotto una licenza aperta, strutturata coerentemente con le politiche di WorldCat. D'altro canto, OCLC è preoccupata per il rilascio di consistenti quantità di dati bibliografici derivanti da WorldCat sotto licenze di tipo più aperto, come la CC0 e la ODC-PDDL [43], licenze che sono incompatibili con i diritti dei dati in WorldCat, con le clausole di responsabilità e con gli intenti delle politiche, e i regolamenti delle comunità che afferiscono al consorzio. In particolare, le politiche richiedono ai membri di OCLC di esercitare i loro diritti nei confronti dei dati bibliografici e dei prodotti derivati da WorldCat, nel

contesto di certe responsabilità precise, come garantire sempre un'appropriata attribuzione in merito alla titolarità dei dati e garantendo che il successivo riutilizzo sia

- a. coerente con gli scopi pubblici di OCLC e con le sue politiche;
- b. a supporto della sostenibilità a lungo termine di WorldCat.

Le licenze CC0 e PDDL non richiedono attribuzione e non pongono limiti a successivi riutilizzi di qualsiasi tipo, in quanto con l'adozione di tali licenze esplicitamente si dichiara di rinunciare a tutti gli altri diritti o ai crediti relativi alla titolarità o menzione d'autore di tutti i dati bibliografici del catalogo. Per le ragioni su esposte OCLC preferisce orientarsi su una licenza ODC-BY, una via di mezzo tra la licenza [Open Database License \(ODC-ODbL\) - Attribution Share-Alike for data/databases \[44\]](#), più restrittiva anche in termini di interoperabilità, e le licenze di pubblico dominio come la PDDL o la CC0 sulle quali OCLC nutre parecchie riserve.

Questo è ad oggi l'orientamento di OCLC, il quale intende formulare una proposta concreta per una linea d'azione basata su reali richieste provenienti dal mondo bibliotecario e dell'informazione, cercando di ottenere input aggiuntivi dai membri di OCLC e dalle comunità portatrici di interessi attorno all'argomento licenze open data e al modello Linked Open Data [\[45\]](#). Poiché l'argomento licenze Open Data è relativamente nuovo e non vi è ancora una tradizione di buone pratiche sulle quali contare, OCLC ha impegnato un avvocato con una notevole esperienza in questo settore [\[46\]](#), per una consulenza sulle questioni generali su casi d'uso particolari.

7. Open Data in Italia

In Italia, sulla scia di altre iniziative governative straniere, nel giugno del 2010 si annunciava l'apertura di un portale web italiano per l'Open Data ispirato ad esperienze anglosassoni, [\[47\]](#) Ma già nel maggio dello stesso anno la Regione Piemonte [\[48\]](#) ne realizzava uno, diventando così pioniera nel settore della trasparenza dei dati. All'iniziativa piemontese seguiva quella sarda, indirizzata all'apertura dei dati cartografici. Un passo importante nella direzione dell'Open Data è stato compiuto dal Foromez nell'ottobre del 2010, dando vita alla Italian Open Data Licence (IODL v. 1.0) [\[49\]](#), la prima licenza italiana rivolta ai dati pubblici aperti, allineata con il modello Creative Commons 2.5 e Open Data Commons. Ad oggi, le più importanti attività in direzione Open Data partono da associazioni, centri di competenza, community [\[50\]](#), anche se a livello governativo resta ancora molta strada da fare.

Numerosi sono gli articoli in letteratura riguardo l'Open Government, e anche i siti nati con l'intento di fornire informazioni aperte ai cittadini, anche se manca a tutt'oggi – almeno in Italia – una reale capacità dei cittadini stessi di auto-organizzarsi in forme di democrazia partecipata e cittadinanza diretta. Il primo esperimento davvero operativo di democrazia diretta che si è concretizzato nella stesura collettiva della legge fondamentale di uno Stato è partito ad aprile in Islanda, quando il Consiglio costituzionale di Reykjavík ha cominciato a trasmettere in streaming sul suo sito web le riunioni settimanali del Consiglio, pubblicando le bozze degli articoli di legge sottoponendole al giudizio popolare:

In un Paese dove due terzi dei 320 mila abitanti hanno un profilo su Facebook il salto nell'agorà virtuale è stato naturale. Pareri e suggerimenti hanno inondato pagine e canali aperti sui diversi social network [...]. La nuova Carta, che trasforma anche le procedure di elezione dei parlamentari e di nomina dei giudici, dovrebbe essere pronta entro luglio, ferma restando la possibilità di prorogare la chiusura dei lavori di un mese. A quel punto il testo potrà essere sottoposto a referendum confermativo anche senza correzioni del Parlamento. [\[51\]](#)

In Italia comunque il movimento Open Data [\[52\]](#) è relativamente giovane, ma rappresenta una realtà vitale e attiva. A testimonianza di ciò è importante parlare del portale *Spaghetti OpenData*

[53] costruito da un gruppo di cittadini italiani con l'intento di promuovere il rilascio di dati pubblici in formato aperto, in modo da renderne facile l'accesso e il riuso (Open Data). Questo portale riunisce i link ai dataset aperti italiani e agli strumenti necessari all'organizzazione dei dati messi a disposizione. Il sito è progettato per essere integrato con CKAN (Comprehensive Knowledge Archive Network) [54], luogo per lo stoccaggio e la distribuzione dei dati, che combina le caratteristiche di un catalogo, di un indice di pacchetti (con registrazione automatica, ricerca e installazione di materiale) e di un wiki (soluzione che permette a chiunque di aggiungere materiali). Progetto di ulteriore rilevanza è quello del portale LOD [55]:

Il Web ci permette di collegare documenti correlati. Allo stesso modo ci permette di collegare i dati correlati. Il termine Linked Data si riferisce ad un insieme di best practice per la pubblicazione e la connessione di dati strutturati sul web. Le principali tecnologie che supportano Linked Data sono URI (uno strumento generico per identificare le entità o concetti nel mondo), HTTP (un meccanismo semplice ma universale per il recupero delle risorse, o le descrizioni delle risorse) e RDF (un generico modello grafico basato su dati con cui strutturare collegare i dati che descrive le cose nel mondo) [56].

Ciò che ha condotto a questo movimento ormai ben avviato è il desiderio di ottenere dal settore amministrativo, sia pubblico che privato, la massima trasparenza. I protocolli per il trattamento di questa enorme mole di dati però tardano ad arrivare. È necessario, come già detto in precedenza, pensare non solo alla pura pubblicazione sul web, ma al processo di integrazione dei dati nel passaggio dall'Open Data al Linked Open Data.

LinkedOpenData.it è un progetto che, nel contesto italiano, nasce in appoggio alle iniziative Open Data con lo scopo di pubblicare i dati grezzi in Linked. Il portale vuole costituire un punto di aggregazione e correlazione tra dataset, rendendo disponibili gli strumenti stessi per l'interrogazione e la rielaborazione dei dati. Il passo in più, che distingue l'Open Data e l'Open Linked Data è la possibilità di creare relazioni semantiche non solo nel contesto interno ma anche con entità esterne. L'attività più complessa nel Linked Data è la scelta dell'ontologia, o vocabolario, più adatta alle esigenze del set di dati. La rappresentazione in formati specifici richiede utilizzo di più vocabolari che vanno scelti a seconda della loro diffusione di utilizzo da parte dell'utenza. Il Linked Open Data Italia costituisce un punto di riferimento per quelle attività di stoccaggio dati per la condivisione libera, promuovendo queste azioni in tutti i loro aspetti.

Attività degna di nota è quella promossa da un piccolo gruppo [57] di persone interessate al tema OpenBiblio in Italia, che opera entro la Open Knowledge Foundation Italia [58], e del quale le scriventi fanno parte, in connessione con Karen Coyle [59]. Il gruppo si occupa della promozione dei *Principles for Open Bibliographic Data* [60] in Italia, della manutenzione dei pacchetti su CKAN.net, della gestione del servizio *Bibliographica*, un catalogo aperto di opere culturali, del movimento Open Access (OAI) nell'ambito delle biblioteche con il quale è opportuno creare un collegamento. Negli incontri, che si svolgono tutti in remoto via conferenze skype e con strumenti come PiratePad, si è parlato dei possibili fronti d'azione per il movimento OpenBiblio in Italia: mondo della ricerca scientifica, open government data e trasparenza amministrativa, dati bibliografici aperti: cataloghi e banche dati bibliografiche composte essenzialmente di metadati [61].

8. La filosofia che sottende l'Open Data

Gli Open Data fanno riferimento a dati contenuti in banche dati disciplinari diverse, come ad esempio quelle appartenenti all'ambito matematico, biomedico, chimico, genetico, cartografico, anagrafico o governativo. La difficoltà nel diffondere i dati provenienti da queste sfere amministrative e di ricerca deriva perlopiù da motivazioni commerciali: dati, metadati, formule, sono posseduti e gestiti da organizzazioni di stampo privato o pubblico che vedono nella diffusione

del materiale informativo a loro disposizione una minaccia per la propria attività. Spesso questi enti mantengono il controllo su tale materiale attraverso forme restrittive di accesso, che limitano le possibilità di riutilizzo dei dati. I sostenitori dell'Open Data affermano invece la necessità di eliminare queste restrizioni, che costituiscono un vero e proprio limite al bene della comunità. Oltre all'eliminazione delle restrizioni di accesso, che permetterebbe la pubblicazione dei dati, il gruppo Open Data chiede che vi sia la possibilità, senza altre particolari forme di autorizzazione, di riutilizzare questo materiale informativo.

Una rappresentazione tipica della necessità dell'apertura dei dati viene proposta in una dichiarazione di John Wilbanks, direttore esecutivo dello Science Commons: "Numerosi scienziati hanno sottolineato con ironia che proprio nel momento storico in cui disponiamo delle tecnologie per consentire la disponibilità dei dati scientifici a livello globale e dei sistemi di distribuzione che ci consentirebbero di ampliare la collaborazione e accelerare il ritmo e la profondità della scoperte [...], siamo occupati a bloccare i dati e a prevenire l'uso di tecnologie avanzate che avrebbero un forte impatto sulla diffusione della conoscenza. [62]

Secondo i sostenitori del movimento Open Data esistono diverse motivazioni per le quali è necessario rendere libero l'accesso ai dati. In primis, è doveroso ammettere l'appartenenza dei dati allo stesso genere umano: il materiale raccolto in ambiti di interesse genetico, medico, ambientale, etc. è da considerarsi patrimonio dell'umanità e come tale va reso disponibile non solo all'accesso ma anche al riutilizzo. L'universalità dell'informazione viene poi chiamata in causa anche per quanto riguarda i dati che derivano dalla pubblica amministrazione: tale materiale si è costituito e ha preso forma grazie al sostegno della cittadinanza e al denaro pubblico, che ne ha permesso la raccolta e la gestione. Questa massa di dati finanziata dai contribuenti deve essere disponibile agli stessi. Per quanto riguarda l'evoluzione nel campo della ricerca in ogni ambito di studi, le motivazioni che spingono alla richiesta di dati aperti sono altrettanto fondamentali: è indubbio che l'accesso al materiale informativo sia propulsore fondamentale allo sviluppo delle attività umane, pratiche ed intellettive.

9. Quanto aperti devono essere i dati?

Dopo aver chiarito il concetto di dato e il raggio d'azione della filosofia Open Data, è possibile chiarire anche gli aspetti fondanti che descrivono un set di dati come aperto:

I dati aperti devono essere indicizzati dai motori di ricerca; devono essere disponibili in un formato aperto, standardizzato e leggibile da una applicazione informatica per facilitare la loro consultazione ed incentivare il loro riutilizzo anche in modo creativo; devono essere rilasciati attraverso licenze libere che non impediscano la diffusione e il riutilizzo da parte di tutti i soggetti interessati [63].

E' indubbio che alla base dell'Open Data vi sia sempre un'etica simile a tutta una serie di altri movimenti e comunità di sviluppo "open", ad esempio:

- Open Source [64], che riguarda le licenze di distribuzione di programmi e applicazioni.
- Open Content [65], facilmente confondibile con l'Open Data, ma mentre l'Open Data concentra l'attenzione alla ricerca scientifica e ai dati che le competono, l'Open Content si rivolge alla creatività delle opere.
- Open Knowledge, che si ispira all'Open Content e mira alla "diffusione della conoscenza con le seguenti caratteristiche: libera (di proprietà di ogni essere umano), accessibile (raggiungibile da ogni essere umano), comprensibile (presentata nel modo più semplice possibile, e attraverso la definizione esaustiva di ogni significato coinvolto), universale (accessibile in ogni lingua)" [66].

- Open Access [67], che si concretizza nella disponibilità dei contenuti digitali online. L'Open access si rivolge in maggior parte agli articoli di riviste e alle ricerche svolte in campo universitario, consentendone la pubblicazione e l'accesso/utilizzo gratuito.
- Open Library, consistente in progetti che mirano a creare pagine web per ogni libro pubblicato, permettendo l'accesso libero all'opera, la condivisione, la diffusione della stessa. Nello specifico l'*Open Library Project* nasce da Internet Archive, in collaborazione con la California State Library e la Kahle Austin Foundation.

L'Open Data Commons [68] è un raccoglitore di strumenti legali per aiutare la comunità nella pubblicazione di dati aperti. Esso è costituito, infatti, da una serie di licenze che derivano da un progetto di raccolta dell'Open Knowledge Foundation per una più ampia diffusione dell'apertura dei dati. La licenza più conosciuta, quella che rappresenta in maniera più significativa gli intenti della OKF, è la Open Database Licence che, a partire dai dati cartografici, è diventata la licenza più diffusa anche in ambito di dati pubblici. Essa offre la possibilità di copiare, distribuire e utilizzare i set di dati contenuti in un database e la possibilità di rimaneggiare tali dati creando nuovo materiale informativo. Ciò che viene chiesto in cambio è la possibilità di poter utilizzare le opere derivate dal database madre, la garanzia del rispetto della licenza per la diffusione dei derivati (quello che viene definito come share-alike) e la garanzia di non limitazione all'uso delle versioni adattate. L'accesso riguarda la possibilità di disporre dell'opera nella sua interezza in modo facile e diretto, preferibilmente in un formato accessibile ad ogni esigenza.

Questo riconduce alla questione della redistribuzione, intesa come la libertà da ogni tipo di licenza che limiti la vendita o l'offerta gratuita dell'opera. Le licenze - norme secondo le quali l'opera viene resa disponibile - devono essere prive di qualsivoglia limitazione agli interventi di modifica, consentendo il libero intervento e successiva distribuzione dell'opera rimaneggiata. Per questi stessi motivi, l'opera deve essere disponibile in un formato accessibile, evitando ogni restrizione di tipo tecnologico che renda il lavoro limitato a pochi. Le licenze possono comportare comunque la necessità di citazione del creatore dell'opera a patto che non risulti in alcun modo onerosa. Un'altra richiesta lecita nelle licenze è la diversificazione del nome o del numero di versione rispetto all'opera madre alla quale ci si ispira. Le norme che indicano le modalità di riutilizzo non devono discriminare in questa attività nessun individuo, gruppo o settore di attività e non devono comportare alcun obbligo aggiuntivo di sottoscrizione.

Le restrizioni all'accesso non devono riguardare nemmeno la possibilità o meno di poter usufruire di materiale estratto da pacchetti informativi più grandi, e non implica di per sé la restrizione di tutte le opere collegate ad un particolare materiale. Questa, in sintesi, è una lista delle caratteristiche che un'opera deve possedere per essere considerata "aperta". Oltre a fornire gli elementi definitivi per validare l'apertura del materiale, l'OKF offre una descrizione dell'Open Data in questi termini:

La definizione può essere riassunta nella dichiarazione che "Un pezzo di contenuti o dati è aperta, se qualcuno è libero di utilizzare, riutilizzare e ridistribuirlo – soltanto soggetto, al massimo, all'obbligo di attribuzione e share-alike". [69]

Si fa riferimento alle prime tre caratteristiche in precedenza elencate: accesso, redistribuzione e riutilizzo, aggiungendo l'accento alla possibilità massima dell'attribuzione, anch'essa componente fondamentale per dichiarare un'opera aperta.

10. Il ruolo del web semantico nell'Open Data e il concetto di Linked Data

Parlando di Dati Aperti si deve menzionare la struttura che ne ha favorito lo sviluppo, ovvero il Semantic Web [70], un'evoluzione del Web attuale. Esso ha lo scopo di creare strumenti applicativi

in grado di rintracciare, utilizzare, integrare con maggiore facilità dati di provenienza diversa. E' proprio in questo contesto che si sviluppano il movimento Open Data e l'iniziativa Linked Data [71]. Quest'ultima nasce dal diretto interesse di Tim Berners-Lee, inventore del World Wide Web, con l'obiettivo di incoraggiare la pubblicazione sul Web dei "raw data" o "dati grezzi": "Il paradigma Linked Data definisce delle regole per la pubblicazione dei dati sul Web in modo che essi possano essere facilmente individuati, incrociati e manipolati dalle macchine. [72]

Questo conduce verso il Semantic Web, ovvero un Web come un unico grande database globale e distribuito, interrogabile dalle macchine indipendentemente dalla provenienza dei dati. Il termine web semantico, coniato da Tim Berners-Lee, porta all'arricchimento del classico World Wide Web e quindi alla nascita di un nuovo ambiente dove i documenti si integrino a dati ed informazioni che ne delineino il contesto semantico, in un formato che consente la facile elaborazione automatica. [73]

Se il movimento Open Data intende eliminare gli ostacoli principali alla libera condivisione dei dati, le barriere di tipo economico, ma anche sociale e culturale, Linked Data mira piuttosto ad abbattere la barriera tecnologica che impedisce la libera condivisione dei dati. Lo scopo è ottenere, per mezzo dell'Open Data e del Linked Data, un Web ancor più libero e aperto, definibile come Open Linked Data (LOD).

Principalmente, il web semantico è una tecnologia Web 3.0, un modo di collegare i dati tra i sistemi o le entità che permette interrelazioni tra dati disponibili in tutto il mondo sul web. In sostanza, si segna un cambiamento nel modo di pubblicare i dati, da documenti leggibili dall'uomo (HTML) a documenti leggibili dalla macchina. Ciò significa che le macchine possono fare qualcosa in più. Oggi, gran parte dei dati ci viene consegnata nella forma di pagine web, ossia documenti HTML che sono legati tra loro attraverso l'uso di collegamenti ipertestuali. Gli esseri umani o le macchine sono in grado di leggere questi documenti, ma di solito nella ricerca di parole chiave in una pagina, le macchine hanno difficoltà a estrarre un significato dai documenti stessi. [74]

Per Linked Data si intende l'uso del web per collegare i dati correlati che non risultavano connessi tra loro in precedenza, e per abbassare le barriere al collegamento di dati già connessi con metodi diversi [75]. I Linked Data, o dati collegati, non sono altro che un aspetto del web semantico, un metodo di esporre, condividere e connettere dati attraverso URI deferenziabili [76]. Infatti, l'aspetto che contraddistingue i Linked Data [77] è l'elaborazione di meccanismi condivisi per l'accesso e l'interrogazione dei dati in maniera tale da permettere il collegamento e l'interrogazione di più dataset correlati per mezzo di linguaggi e protocolli standardizzati e noti.

L'ambizione più alta del movimento è la creazione di uno spazio unico di condivisione dati dove l'utente e il software possano trovare libertà di movimento alla ricerca di informazioni per la condivisione e il riutilizzo privo di barriere. Un forte vantaggio che spinge verso il progetto Linked Data è costituito dall'integrazione dei materiali-dati: chi pubblica set di dati può infatti ricondurre il suo insieme a elementi già esistenti, evitando fastidiose duplicazioni. Un altro vantaggio importante riguarda i creatori di siti e applicazioni web *mash-up*, i quali potranno linkare il dato informatico senza sovraccarichi eccessivi, ottenendo inoltre informazioni sempre aggiornate. L'utilizzo di vocabolari comuni permetterà infine di raggiungere la standardizzazione dei set di dati creando maggior omogeneità e chiarezza.

11. Verso gli Open Linked Data (LOD)

Il passaggio da Open Data a Open Linked Data (LOD) non può prescindere da un'analisi del problema del formato [78]. Come già detto più volte in precedenza, questo passo verso l'Open

Il Linking Open Data è un progetto che nasce grazie al W3C SWEO [80]. Tale progetto ha lo scopo di creare una rete di dati aperti e disponibili a tutti, con la particolarità dell'interconnessione e dell'interoperabilità degli stessi dataset pubblicati [81]. Il formato adottato è l'RDF [82], che ha la caratteristica fondamentale di consentire l'interoperabilità tra diverse applicazioni che si scambiano informazioni sul Web, e che permette la fluente accessibilità a insiemi di dati relazionati tra loro, così da poter passare da un dato informativo all'altro senza dover preoccuparsi della provenienza diversificata degli stessi. Tutto questo facilita il recupero dell'informazione non solo per l'utente-uomo ma anche per l'utente macchina.

I ricercatori del campo scientifico, economico, statistico e sociale necessitano un gran numero di dati, spesso presentati in modo caotico, ma ben strutturati, secondo standard che rispecchiano l'interesse di studio che è loro proprio (XML, JSON [83]), in maniera tale da permettere la facile combinazione con altri set di dati. Al contrario chi lavora con le macchine, chi programma software di settori specifici, non cerca il dato di massa ma il dato aggiornato e flessibile all'esigenza particolare nel suo formato (APIs [84]).

L'utente "di base" ha un'esigenza altra rispetto a quelle presentate in precedenza: non vuole il dato di massa, non vuole il dato strutturato tecnicamente, ma compie la ricerca per interesse particolare, vuole poter ordinare l'elenco risultante dalla query in maniera tale da rintracciare all'istante l'informazione. Ciò che più interessa a chi compie questa ricerca-base è l'accessibilità diretta, la visualizzazione semplice.

Per chi pubblica il dato è altrettanto fondamentale poter agire in maniera organizzata, al fine di conoscere l'attività di condivisione che segue tale pubblicazione. In una realtà ideale sarebbe possibile conoscere le modalità di riutilizzo dei dati, la quantità di citazione, di commento, di riutilizzo nel linkaggio, di integrazione in software e applicazioni. Questi sono gli scopi fondamentali del dato aperto e connesso, le direzioni nelle quali muoversi: la reperibilità, l'accessibilità, l'interattività, l'integrabilità.

Secondo alcuni ricercatori del campo scientifico, il Linked Data non è sufficiente a rispondere alle esigenze di studio del settore. Nell'interessante scritto *Why Linked Data is Not Enough for Scientists* [85], alcuni autori australiani spiegano quali siano le difficoltà che i dati connessi incontrano nel settore della ricerca scientifica. I dati scientifici rappresentano una porzione significativa del Linked Open Data cloud e la scienza, di per sé, trae beneficio dalla capacità di fusione e condivisione dei dati che ciò permette. Tuttavia, la semplice pubblicazione dei dati legati nella nuvola non soddisfa necessariamente le esigenze di riutilizzo. La pubblicazione ha requisiti di provenienza, qualità, credito, attribuzione, metodologia al fine di fornire la riproducibilità che consente la validazione dei risultati.

Attraverso l'uso di http URI e infrastruttura Web, Linked Data offre un meccanismo standardizzato per la pubblicazione dei dati strutturati, con la possibilità di esplorare e raccogliere nella navigazione risorse esterne. È l'utilizzo di RDF ad offrire la possibilità di deduzione per il recupero e l'interrogazione delle informazioni. Che i dati siano collegati non è esplicitamente previsto, tuttavia è un modello comune per descrivere la struttura di questi oggetti di ricerca. Questo però non dice molto su come sia possibile organizzare, gestire e consumare l'informazione.

"Linked Data fornisce una piattaforma per la condivisione e pubblicazione dei dati, ma il fatto di pubblicare i nostri dati come dati collegati non è sufficiente per sostenere e facilitare il loro riutilizzo". [86] Le perplessità riguardano anche la scelta appropriata di un set di dati dalla nuvola Linked Data, la mancanza di espressività nei set di dati utilizzati (limitando la scelta del sistema di calcolo utilizzabile) e la mancanza di schemi di mapping tra insiemi di dati. La preoccupazione

prevalente riguarda però la necessità di un comune modello di aggregazione. "Si noti che questo non è inteso come una critica all'approccio ai dati collegati, ma semplicemente una constatazione del fatto che una struttura aggiuntiva e formati appropriati di metadati sono necessari nella parte superiore del substrato Linked Data per supportare l'interpretazione e il riutilizzo di tali dati. Inoltre vi è la necessità da parte dei metadati di collegare la struttura delle risorse di ricerca alla funzione del processo di ricerca stesso". [87]

Quello che manca, secondo questa visione, è quindi un meccanismo che descrive l'aggregazione delle risorse e attraverso questa descrizione renda il contributo di queste risorse per le indagini e le relazioni che intercorrono tra queste, per cogliere il valore aggiunto della collezione e permettere il riutilizzo attraverso lo scambio di un singolo oggetto.

12. Dati aperti con LODE

Mentre Linked Open Data si riferisce ad un insieme di pratiche per la pubblicazione, la condivisione e l'interconnessione di dati strutturati all'interno del contesto del web semantico, l'acronimo "LODe" sta per Linked Open Data enabled, ovvero una abilitazione all'attribuzione di "dati aperti e legati". Come sopra descritto, le tecnologie che accompagnano queste procedure consistono nell'URI, che identifica le entità/concetti, nell'RDF, che struttura e collega le descrizioni degli oggetti, e nell'HTTP per recuperare le risorse e le descrizioni che ne derivano.

Per quanto riguarda queste operazioni, non vi è però approccio univoco in una realtà che sta sviluppando sempre più modelli standard per i metadati diversi tra loro, per scopi e realtà differenti. Per questo motivo c'è bisogno di individuare standard univoci che riescano ad esprimere l'impatto di prontezza nel campo LOD di ogni set di dati che si intende condividere e aprire al riutilizzo. La mancanza di univocità porta infatti i fornitori di dati e di servizi per i dati a domandarsi quale sia lo standard più adatto agli scopi, qual è l'unità minima che detta il valore per la condivisione di un set di dati, se vi è un modello di metadato universalmente applicabile e valido in ogni circostanza, qual è l'unità di scambio (forma letterale di rappresentazione o URI di identificazione), ecc.

Uno strumento utile, che tenta di dare delle risposte ai professionisti dell'informazione nel decidere la modalità più consone alla codifica dei dati in possesso, fornendo al contempo anche strumenti valutativi nella scelta delle strategie di codifica per la produzione di dati LOD-e, è il documento noto come *Raccomandazioni LODE BD* [88], redatto dal team AIMS FAO. Riconoscendo il fatto che i produttori di dati bibliografici di solito usano strutture di dati bibliografici differenti che, come previsto da LinkedData, possono essere situate in diverse fasi, LODE offre più insiemi di punti d'azione e suggerimenti di codifica corrispondenti a ciascuna delle proprietà.

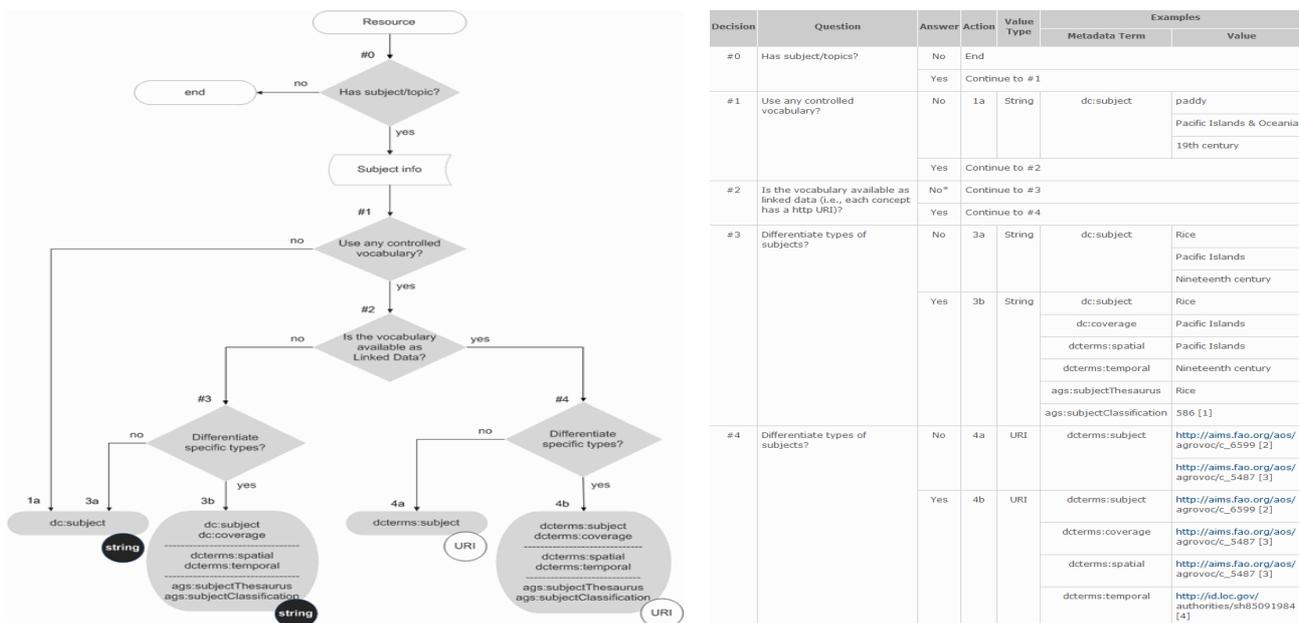
Ci sono cinque principi chiave nelle Raccomandazioni LODE [89]:

1. facilitare il processo decisionale relativo alla codifica dei dati ai fini dello scambio e del riutilizzo dei dati;
2. incoraggiare l'uso dei dati dei file di autorità, dei vocabolari controllati e degli standard di codifica della sintassi laddove possibile, al fine di migliorare la qualità dell'interoperabilità e l'efficacia dello scambio dell'informazione;
3. promuovere l'uso di standard di metadati consolidati, nonché l'emergente set di vocabolari LODE proposti entro la Comunità Linked Open Data;
4. incoraggiare l'uso delle risorse URI come valori di dati entro i metadati stessi quando disponibili;
5. fornire uno strumento che sia aperto a suggerimenti di nuove proprietà e termini di metadati, in base alle esigenze della comunità LOD così come alle esigenze dei produttori dei dati.

Il team AIMS della FAO ha come obiettivo principale quello di fornire agli utenti servizi di organizzazione della conoscenza una serie di strumenti per la gestione dei metadati e dei rispettivi standard di codifica, e linee guida metodologiche per queste operazioni a esperti nella gestione dell'informazione e agli esperti di information technology, al fine di costruire sistemi per la facile sincronizzazione e condivisione dei dati.

Il sito <aims.fao.org> permette la comunicazione e la collaborazione tra le comunità globali agricole, compresi i fornitori di informazioni, gli istituti di ricerca, le istituzioni educative e il settore privato, attraverso forum e blog. I principi base su cui poggiano queste raccomandazioni consistono essenzialmente nella promozione di standard consolidati per i metadati, nell'incoraggiare all'uso di vocabolari controllati che permettano la massima interoperabilità, oltre che all'uso del sistema identificativo URI, nel facilitare il processo decisionale per l'individuazione del modello di codifica, nel fornire supporto alle esigenze emergenti nella comunità Linked Data.

Per quanto riguarda i dati bibliografici vi è la necessità di individuare uno strumento di riferimento nella scelta di strategie di codifica appropriate. Si entra così nel contesto dei LODe-BD (Linked Open Data enabled – Bibliographic Data). Infatti, per tentare di dare una risposta all'esigenza di univocità nella pubblicazione di dati bibliografici è necessario analizzare con accuratezza le esigenze che si esprimono nel campo dei metadati bibliografici. LODe-BD utilizza diagrammi di flusso per presentare alberi decisionali per le proprietà dei metadati.



Esempio di alberi decisionali per Soggetti [90]

Esso risponde alla richiesta di un modello di codifica specifico per i dati bibliografici, al fine di produrre correlazione e scambio di dati. Il produttore di contenuti bibliografici che intenda pubblicare un database dovrà tener conto di entità e relazioni specifiche (tema, risorse, agente), necessariamente coinvolte nella descrizione della risorsa bibliografica. Le proprietà che secondo questo modello sono da considerarsi significative e utili per la pubblicazione di dati LODready sono:

1. Titolo
2. Entità responsabile
3. Caratteristiche fisiche
4. Collocazione
5. Soggetto
6. Descrizione del contenuto
7. Proprietà intellettuale
8. Uso
9. Relazione tra documenti/agenti

La scelta, nella preparazione di metadati LODready, deve ricadere su standard ampiamente utilizzati nella comunità e su vocabolari LODe sempre più diffusi all'interno di quest'ultima.

<u>dc:</u>	<u>Dublin Core Metadata Element Set</u>
<u>dcterms:</u>	<u>DCMI Metadata Terms</u>
<u>bibo:</u>	<u>Bibliographic Ontology</u>
<u>agls:</u>	<u>AGLS Metadata Standard (Australian Government Locator Service)</u>
<u>ags:</u>	<u>AgMES (Agricultural Metadata Element set)</u>
<u>eprint:</u>	<u>Eprints Terms, UKOLN</u>
<u>marcrel:</u>	<u>MARC List for Relators</u>

Standard diffusi [91]

In merito alla scelta dello standard per i metadati, specifico è il ruolo degli "alberi decisionali" che forniscono assistenza nel processo di selezione e si presentano come diagrammi di flusso che individuano le proprietà rilevanti in ciascuno dei nove gruppi. Ogni diagramma di flusso, infatti, a partire dalla specifica risorsa, presenta i principali punti decisionali e offre supporto per la soluzione dei principali problemi di codifica dei metadati.

Ogni albero decisionale è progettato per facilitarne la selezione delle strategie appropriate nelle più svariate situazioni ed esigenze. Al termine di ogni diagramma di flusso ci sono insiemi di termini alternativi per la selezione dei metadati. Ogni grafico è seguito da spiegazioni testuali, con note ed esempi e tabelle esplicative. Così si presentano le strutture promosse per la diffusione sempre più organizzata di sistemi volti alla totale apertura e fruibilità dei dati LODe.

Non è più solo importante fornire i dati in un formato aperto, ma risulta fondamentale munire i dati degli step utili a comprendere gli sviluppi che sono intervenuti su tali dati, rendendo anche graficamente i processi intervenuti sui dati stessi tramite visualizzazioni in mappe concettuali e diagrammi di flusso. Se tutto il processo risulta trasparente, il dato può essere riutilizzato nelle sue singole fasi, tramite link adeguati. Il web 2.0 e l'uso di feed possono essere di un certo interesse non solo in termini di trasparenza di dati pubblici o provenienti dal mondo della ricerca, ma nella manipolazione stessa del dato pubblico, inteso come bene comune.

Chiudiamo con una citazione tratta da un post di Gavin Starks pubblicato sul blog di Tim O'Reilly [92]:

Data is not binary: Why open data requires credibility and transparency. Major difference between open and public data is [that with open data] you have the ability to re-use it. Data in document format is effectively useless. By making [data] open [...] people can analyze, compare, and benchmark it, and find patterns that you did not realize.

Antonella De Robbio, CAB Centro di Ateneo per le Biblioteche - Universita' degli Studi di Padova e-mail: antonella.derobbio@unipd.it

Silvia Giacomazzi, Padova, e-mail silvia.giacomazzi@unipd.it.

Note

[1] <http://wiki.openarchives.it/index.php/Open_Access,_Open_Data,_E-research>.

[2] L'Open Knowledge Foundation è un'organizzazione non-profit fondata nel 2004, con lo scopo di promuovere la conoscenza aperta, nel tentativo di render sempre più ampia ed organizzata la condivisione di dati di diversa natura (dati statistici, geografici, bio-medici, ecc.), per renderli utilizzabili, riutilizzabili, ridistribuiti per scopi sociali benefici. L'Open Knowledge Foundation sostiene l'apertura di una serie di iniziative, in particolare, quelle legate all'Open Data. Esso copre: a) l'ambito scientifico, storico, geografico; b) l'ambito dei contenuti, come musica, film, libri; c) l'ambito governativo e delle altre informazioni amministrative. I dati sono contenuti all'interno del campo di applicazione della Open Knowledge Definition, che allude al protocollo Science Commons per l'implementazione dell'accesso aperto ai dati. Cfr. <http://it.wikipedia.org/wiki/Open_knowledge>.

[3] Da Wikipedia: "gli Standard per la Gestione dell'Informazione sull'Agricoltura (AIMS), sono un'iniziativa promossa dalla FAO allo scopo di migliorare la coerenza tra i sistemi informativi agricoli. Primo obiettivo del progetto è la creazione di un 'clearing house' per gli standard di gestione dell'informazione. Ciò servirà a rendere interoperabili i sistemi informativi agricoli esistenti e quelli nuovi e a condividere e promuovere l'adozione di metodologie, standard e applicazioni comuni".

[4] Se poi tale elaborazione che consente al dato di divenire informazione sia frutto della mente umana o anche di un'elaborazione automatica, è ulteriore questione da analizzare anche nell'ottica delle nuove frontiere che riguardano l'analisi automatica dei dati, come l'analisi testuale.

[5] Studioso di cibernetica, psicologo, linguista britannico.

[6] Gregory Bateson - Mary Catherine Bateson, *Dove gli angeli esitano. Verso un'epistemologia del sacro*, Milano, Adelphi, Milano, 1989.

[7] Da Wikipedia, <<http://it.wikipedia.org/wiki/Dato>>.

[8] "Con l'espressione 'Open Government' - letteralmente 'governo aperto' - si intende un nuovo concetto di Governance a livello centrale e locale, basato su modelli, strumenti e tecnologie che consentono alle amministrazioni di essere "aperte" e "trasparenti" nei confronti dei cittadini. In particolare l'Open government prevede che tutte le attività dei governi e delle amministrazioni dello stato debbano essere aperte e disponibili, al fine di favorire azioni efficaci e garantire un controllo pubblico sull'operato" (<http://it.wikipedia.org/wiki/Open_government>).

- [9] A questo proposito vogliamo citare l'iniziativa di *innovatoripa* (<<http://www.innovatoripa.it/>>), rete sociale composta da persone che condividono l'interesse per il tema dell'innovazione e del cambiamento nella pubblica amministrazione italiana.
- [10] Peter Murray-Rust, *Open Data in Science*. Available from Nature Precedings (2008), traduzione a cura delle autrici, <<http://hdl.handle.net/10101/npre.2008.1526.1>>.
- [11] Anna Gold, *Cyberinfrastructure, Data, and Libraries*, Part 1: *A Cyberinfrastructure Primer for Librarians*, "D-Lib Magazine", 13 (2007), n. 9/10, <<http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>>, traduzione a cura delle autrici.
- [12] Da Wikipedia, <http://it.wikipedia.org/wiki/Dati_aperti>.
- [13] L'Open Data Foundation (ODaF) è una fondazione statunitense senza scopo di lucro che ha lo scopo di promuovere l'adozione globale di metadati standard e lo sviluppo di soluzioni open-source per la gestione e l'uso di dati statistici. L'OdaF si concentra sul miglioramento dei dati, l'accessibilità dei metadati e la qualità generale a sostegno della ricerca, il processo decisionale, la trasparenza in materia sociale e nelle scienze economiche (SBE).
- [14] <http://www.opendatafoundation.org/ODaF_brochure.pdf>, traduzione a cura delle autrici.
- [15] Fondazione no profit con lo scopo di promuovere l'apertura di contenuti e di dati, creata il 24 maggio 2004 da Peter Murray-Rust, <<http://okfn.org/>>.
- [16] <<http://www.opendatacommons.org/>>.
- [17] Informazioni desunte dalla pagina Facebook <<http://www.facebook.com/pages/Open-Knowledge-Foundation/138241222868769>>.
- [18] La definizione di conoscenza aperta è stata tradotta in ben 23 lingue diverse. La versione italiana è stata curata da Primavera De Filippi, Andrea Glorioso e Juan Carlos De Martin dello NEXA, Center for Internet & Society, del Politecnico di Torino, <<http://www.opendefinition.org/okd/italiano/>>.
- [19] <<http://openbiblio.net/>>.
- [20] <<http://openbiblio.net/principles/it/>>.
- [21] <<http://openbiblio.net/principles/endorse/>>.
- [22] <<http://bonariabiancu.wordpress.com/tag/oclc/>>.
- [23] Si veda l'articolo dall'eloquente titolo sul blog di Dave Pattern (*Self plagiarim is style*), citato sempre da Bonaria Biancu, <<http://www.daveyp.com/blog/archives/528>>. I dati di utilizzo si possono recuperare in modalità completamente aperta e libera, <<http://library.hud.ac.uk/data/usagedata/>>.
- [24] <<http://gs-service-bookdata.web.cern.ch/gs-service-bookdata/announcement.html>>.
- [25] <<http://www.opendatacommons.org/licenses/>>.
- [26] <<http://obd.jisc.ac.uk/>>.
- [27] <<http://obd.jisc.ac.uk/navigate>>.
- [28] <<http://obd.jisc.ac.uk/examples>>.
- [29] <<http://ckan.net/tag/library>>.

[30] Per approfondimenti si rimanda a Antonella De Robbio, *La tutela giuridica delle banche nel diritto d'autore e nei diritti connessi*, 1999, <<http://eprints.rclis.org/bitstream/10760/4012/1/dbthesis.pdf>>.

[31] OCLC (Online Computer Library Center) è un'organizzazione non-profit che agisce a livello mondiale e che fornisce alle biblioteche aderenti servizi per il recupero, la catalogazione e la conservazione dei libri.

[32] L'interessante dibattito può essere seguito a questo link: <<http://blog.okfn.org/2011/06/06/oclc-worldcat-rights-and-responsibilities-and-open-data-licensing/>>, mentre le relative policy si possono visionare qui

<<http://www.oclc.org/worldcat/recorduse/policy/default.htm>>.

[33] <<http://blog.andreamarchitelli.it/2008/11/nuova-policy-per-luso/>>.

[34] Le vecchie linee guida furono stilate nel 1987:

<<http://www.oclc.org/support/documentation/worldcat/records/guidelines/default.htm>>.

[35] <<http://www.oclc.org/worldcat/catalog/policy/default.htm>>.

[36] <http://www.librarything.com/wiki/index.php?title=OCLC_Policy_Changes&diff=11748&oldid=11747>.

[37] Rete globale di servizi alle biblioteche che permette alle stesse di essere maggiormente interoperabili e aperte.

[38] Tra i quali FirstSearch, servizio di reference basato sull'aggregazione di banche dati di OCLC e commerciali, o ECO, Electronic Collections Online. Vedi le informazioni sul sito di IFNET, <<http://www.ifnet.it/oclc/>>.

[39] <<https://creativecommons.org/about/cc0>>.

[40] *WorldCat Rights and Responsibilities for the OCLC Cooperative*,

<<http://www.oclc.org/worldcat/recorduse/policy/default.htm>>.

[41] <<http://www.oclc.org/about/purpose/default.htm>>.

[42] <<http://www.opendatacommons.org/licenses/by/summary/>>.

[43] <<http://www.opendatacommons.org/licenses/pddl/summary/>>.

[44] <<http://www.opendatacommons.org/licenses/by/>>.

[45] A proposito di progetti Linked Open Data in seno a OCLC, si rimanda alle due esperienze di piattaforme prototipali VIAF: The Virtual International Authority File (<http://viaf.org/>) e Dewey Decimal Classification/Linked Data <http://dewey.info/>>, spazi sperimentali che usano sottoinsiemi di dati da WorldCat.

[46] Si tratta di Jordan Hatcher, definito su twitter come "geek, gamer and IP/technology lawyer with a penchant for open licensing. Founder, OpenDataCommons.org; Boardmember, okfn.org" (<<http://twitter.com/#!/jordanhatcher>>).

[47] A riguardo si veda l'articolo di [Fabio Deotto](#) pubblicato su "Wired" del 13 settembre 2010, *Open Government: in Italia si comincia*, <<http://daily.wired.it/news/politica/open-government-in-italia-si-comincia.html>>.

[48] <<http://www.dati.piemonte.it/>>.

[49] <<http://www.formez.it/iodl/>>.

[50] Un esempio si trova a <<http://www.datagov.it/>> iniziativa che nasce da parte di un gruppo di esperti di diritto e di nuove tecnologie, funzionari pubblici e privati, docenti universitari ed altri componenti della società civile che si sono riuniti nell'Associazione Italiana per l'Open Government con l'obiettivo di sensibilizzare cittadini, imprese ed

Amministrazioni e promuovere l'attuazione di strategie di Open Government nel nostro Paese. Per il manifesto dell'Open Government si vada <<http://www.datagov.it/il-manifesto/>>.

[51] <<http://www.quotidianamente.net/cronaca/fatti-dallestero/la-nuova-costituzione-islandese-si-scrive-via-facebook-e-twitter-1593.html>>.

[52] Christian Morbidoni - Michele Barbera - Federico Ruberti, *LinkedOpenData.it: una piattaforma italiana per i dati "aperti" e "collegati"*, <<http://www.linkedopendata.it>>.

[53] <<http://www.spaghettiopendata.org/>>.

[54] <<http://it.ckan.net/about>>.

[55] <LinkedOpenData.it>.

[56] <<http://linkeddata.org/faq>>.

[57] [Open Knowledge Foundation Italia](http://OpenKnowledgeFoundationItalia.org)>, coordinato da Stefano Costa, <<http://it.okfn.org/>>. Gli incontri sono ospitati sul sito

<<http://okfnpad.org/openbiblio-italia>>; chiunque volesse partecipare può iscriversi liberamente.

[58] Al link <<http://it.okfn.org/>> si possono trovare news, eventi, report e materiali vari, che testimoniano l'attività del gruppo di lavoro Open Knowledge Foundation Italia.

[59] Oltre a Karen Koyle, bibliotecaria all'Università di Berkeley, membro del gruppo di lavoro internazionale OKFN sui dati bibliografici, fanno parte del gruppo italiano anche Andrea Zanni (Wikimedia Italia e Alma DL, Università di Bologna), Francesca Di Donato (Università di Pisa e Associazione Linked Open Data), Karen Coyle (Università di Berkeley emembro del gruppo di lavoro internazionale OKFN sui dati bibliografici), Stefano Costa (Università di Siena e coordinatore italiano di Open Knowledge Foundation).

[60] Documento tradotto in italiano.

[61] Relazioni al blog: <<http://it.okfn.org/2011/05/30/openbiblio-italia-prima-puntata/>>.

[62] Da Wikipedia, <http://it.wikipedia.org/wiki/Dati_aperti>.

[63] <http://it.wikipedia.org/wiki/Dati_aperti>.

[64] "In informatica, open source (termine inglese che significa sorgente aperto) indica un software i cui autori (più precisamente i detentori dei diritti) ne permettono, anzi ne favoriscono il libero studio e l'apporto di modifiche da parte di altri programmatori indipendenti. Questo è realizzato mediante l'applicazione di apposite licenze d'uso. La collaborazione di più parti (in genere libera e spontanea) permette al prodotto finale di raggiungere una complessità maggiore di quanto potrebbe ottenere un singolo gruppo di lavoro. L'open source ha tratto grande beneficio da Internet, perché esso permette a programmatori geograficamente distanti di coordinarsi e lavorare allo stesso progetto" (<http://it.wikipedia.org/wiki/Open_source>).

[65] "Il concetto di opera libera (dal francese 'Œuvre libre') o di opera a contenuto libero (dall'inglese 'Open Content') trae la sua ispirazione da quello di Open Source (sorgente libera): la differenza sta nel fatto che in un'opera di contenuti su Internet ad essere liberamente disponibile ed utilizzabile non è il codice sorgente del programma software che li genera, ma i contenuti editoriali generati dal programma, quali testi, immagini, musica e video. Non tutti i contenuti appartenenti a questa categoria sono liberamente disponibili e riproducibili allo stesso modo. Spesso i termini e le condizioni di utilizzo, riproduzione e di modifica dei contenuti, vengono stabiliti da una licenza di pubblicazione che viene scelta dall'autore (<http://it.wikipedia.org/wiki/Contenuto_libero>).

[66] <http://it.wikipedia.org/wiki/Open_knowledge>.

[67] <http://it.wikipedia.org/wiki/Accesso_aperto>.

[68] <<http://www.opendatacommons.org/>>.

[69] <<http://www.opendefinition.org/>>.

[70] <<http://it.okfn.org/2011/02/19/linkedopendata-it-una-piattaforma-italiana-per-i-dati-%E2%80%9Caperti%E2%80%9D-e-%E2%80%9Ccollegati%E2%80%9D/>>.

[71] Per quanto riguarda la differenza tra Semantic web e Linked data si hanno pareri che si discostano un po', comunque una visione diffusa è che il Semantic Web è costituito da Linked Data, ossia il Semantic Web è il tutto, mentre Linked Data è la parte. Tim Berners-Lee, inventore del web e persona autore dei termini Semantic Web e Linked Data, ha spesso descritto Linked Data come "Semantic Web done right" (<<http://linkeddata.org/faq>>).

[72] <<http://it.okfn.org/2011/02/19/linkedopendata-it-una-piattaforma-italiana-per-i-dati-%E2%80%9Caperti%E2%80%9D-e-%E2%80%9Ccollegati%E2%80%9D/>>.

[73] <http://semanticweb.org/wiki/Main_Page>.

[74] <<http://www.linkeddatatools.com/semantic-web-basics>>.

[75] <<http://linkeddata.org/home>>.

[76] Un "dereferenceable Uniform Resource Identifier" o "dereferenceable URI" è un meccanismo di recupero delle risorse che utilizza uno dei protocolli di Internet per ottenere una copia o rappresentazione della risorsa che identifica. In altri contesti non-dereferenceable, come XML Schema, l'identificatore di namespace è ancora un URI, ma questo è semplicemente un identificativo. Nel modello Linked Data, la rappresentazione assume la forma di un documento (di solito HTML o XML) che descrive la risorsa che l'URI identifica (<http://en.wikipedia.org/wiki/Dereferenceable_Uniform_Resource_Identifier>).

[77] <<http://it.okfn.org/2011/02/19/linkedopendata-it-una-piattaforma-italiana-per-i-dati-%E2%80%9Caperti%E2%80%9D-e-%E2%80%9Ccollegati%E2%80%9D/>>.

[78] <<http://www.titticimmino.com/2010/07/27/linked-open-data-cui-prodest-pensieri-sparsi-su-dati-piemonte-it/>>.

[79] <<http://lod-cloud.net/>>.

[80] Semantic Web Education and Outreach (SWEO) Interest Group,
<<http://semanticweb30.wordpress.com/2009/01/25/linking-open-data-e-dbpedia/>>.

[81] Per citare un'importante progetto compreso nel LOD, Dbpedia lavora sugli articoli raccolti in Wikipedia, offrendone al web la versione RDF (fondamentale per l'interrogazione e il riutilizzo) e strutturando connessioni tra domini diversi. Il progetto crea una rete informativa sostanziosa, arricchita ancor più da risorse appartenenti a sorgenti diverse (Geonames).

[82] RDF, o Resource Description Framework, è uno strumento per la codifica, lo scambio e il riutilizzo di metadati strutturati. Esso si basa su alcuni presupposti fondamentali: sottende il fatto che qualsiasi elemento può essere identificato da un particolare URI (Universal Resource Identifier); richiede un linguaggio il meno espressivo possibile per la descrizione e la definizione degli elementi; fa capo all'enunciato "Qualunque cosa può dire qualunque cosa su qualunque cosa".

[83] JavaScript Object Notation semplice formato per lo scambio di dati che utilizza convenzioni proprie dei linguaggi di programmazione della famiglia del C, come C, C++, C#, Java, JavaScript, Perl, Python, ecc...

[84] Si dice che un software è dotato di una API quando fornisce un meccanismo attraverso il quale i programmi esterni sono in grado di comunicare con esso permettendo lo scambio di dati. API permette lo scambio di dati tra siti web, applicazioni e organizzazioni, mentre l'origine dei dati rimane nello stesso luogo. Le applicazioni create con le API sono comunemente chiamati *mashup*, combinando i dati e le funzionalità di diversi programmi. Per esempio, le API consentono di afferrare le statistiche da un sito web e la pubblicazione su un altro sito web. In questo modo non è necessario accedere a siti web singolarmente, ma si possono immediatamente combinare dati tratti da siti diversi.

[85] Sean Bechhofer et al., *Why Linked Data is Not Enough for Scientists*, Sixth IEEE e-Science conference (e-Science 2010), Brisbane, 2010, <<http://eprints.ecs.soton.ac.uk/21587/>>, traduzione a cura delle autrici.

[86] Ibid.

[87] Ibid.

[88] <<http://aims.fao.org/lode/bd>>.

[89] Per una più puntuale trattazione della tematica, si rimanda alla presentazione tenuta all'*euroCRIS Meeting di Bologna* il 26-27 aprile 2011 da Imma Subirats (*FAO of the United Nations*) e Marcia Zeng (*Kent State University*) dal titolo "*Metadata: Concentrating on the data, not on the scheme*", <http://www.eurocris.org/Uploads/Web%20pages/members_meetings/201105%20-%20Bologna,%20Italy/Italian%20session%20-%20Imma%20Subirats.ppt>.

[90] Ibid.

[91] Ibid.

[92] <<http://radar.oreilly.com/2010/06/data-is-not-binary.html>>.