

# Open Data

Paola Zamperlin

Settembre 2015

Licenza Creative Commons Italia 2.5  
Attribuzione - Condividi allo stesso modo

In questo corso, tra le altre cose, si è parlato di Open Source, di [Software Libero](#), di [Open Content](#), è il momento di parlare allora di Open Data.

Il Prof. Andreas ed io abbiamo pensato che fosse utile concludere con una panoramica su questo tema, anche per concludere un percorso che avete iniziato proprio scoprendo quale mondo si nascondeva dietro l'etichetta "Open".

Anche in questo caso, lo spirito che ci guida non è quello di presentare un argomento in maniera esaustiva e apodittica, ma quello di fornirvi dei nuclei concettuali e dei riferimenti bibliografici per costruire insieme una riflessione, fortemente inclinata verso le vostre specifiche esigenze "di campo", come avete imparato a conoscere in questi mesi, sperando che, proprio in virtù della vostra partecipazione diretta, ciò possa apportare un arricchimento professionale tangibile.

Entriamo ora nel vivo dell'argomento.

Come altri più sapientemente di noi compresero un tempo, *Natura non facit saltus*, la natura non fa salti, ma si comporta come una grandezza continua. Potremmo anche noi ora, *mutatis mutandis*, dire che oggi parliamo di Dati aperti o Open data, proprio perché qualche decennio fa qualcuno ha intrapreso un cammino introducendo il concetto di openness nel mondo del software, ma che lentamente è uscito dai confini disciplinari ed è diventato il caposaldo di un'ideologia culturale molto più ampia e pervasiva.

Sarà a molti di voi capitato di incappare nel proprio girovagare in rete in questa relativamente nuova etichetta, Open Data appunto, vuoi come "oggetto" disponibile e scaricabile, vuoi come tema al centro di un dibattito.

Ma per andare con ordine, sperando di non annoiare chi già è avvezzo alla questione e cercando di soddisfare la curiosità di chi invece non ha avuto l'occasione di imbattearsi, vorrei dare alcune basilari informazioni su che cosa andiamo ad esplorare.

Partiamo allora da una definizione: che cosa intendiamo e che cosa intende il legislatore quando ricorre all'espressione Open Data (o dati aperti)?

Secondo la cosiddetta Open Definition, *I dati aperti sono dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque, soggetti eventualmente alla necessità di citarne la fonte e di condividerli con lo stesso tipo di licenza con cui sono stati originariamente rilasciati.* (Cfr. [Definizione di Conoscenza Aperta, Versione 2.0](#))

In Italia, con la [Legge 17 dicembre 2012, n. 221](#) (che converte il DL 179 2012 noto come decreto Crescita 2.0 - PDF 4.1 MB) è stata finalmente data una definizione formale, poi inserita nel [Codice dell'Amministrazione Digitale](#) (di seguito CAD), art. 68, secondo la quale si definiscono (comma 3):

- a. *formato dei dati di tipo aperto, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;*
- b. *dati di tipo aperto, i dati che presentano le seguenti caratteristiche:*
  1. *sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato;*
  2. *sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono*

*adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi metadati;*

- 3. sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione. L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. In ogni caso, l'Agenzia, nel trattamento dei casi eccezionali individuati, si attiene alle indicazioni fornite dalla direttiva 2003/98/CE del Parlamento europeo e del Consiglio, del 17 novembre 2003, sul riutilizzo dell'informazione del settore pubblico, recepita con il decreto legislativo 24 gennaio 2006, n. 36.*

In generale, detto per inciso, al di là del tema OpenData, vi consiglio di dare un'occhiata, se non al testo integrale del CAD, almeno ad alcune sintesi utili come:

[Sintesi introduttiva al CAD](#) (aggiornato al 2013)

[Il nuovo CAD in pillole, ForumPA](#) (aggiornato 2011)

Cercando di sintetizzare quindi, possiamo dire che un dato, o meglio un insieme di dati, è aperto se risponde ad alcune caratteristiche essenziali quali l'essere rilasciato secondo licenze che ne permettano l'utilizzo da chiunque lo voglia anche per finalità di business e sia disponibile in un formato aperto, cioè non proprietario, e pertanto leggibile da un'applicazione informatica e riutilizzabile secondo le necessità dell'utente, ed infine sia corredato da opportuni metadati.

Perché ci è sembrato che l'argomento potesse essere interessante? Provo a spiegare per punti:

1. Come insegnanti, oltre che come cittadini, siete in qualche modo obbligati a una presa di coscienza del cambiamento che è in atto negli ultimi anni, tecnico, giuridico, culturale, non sempre pacifico e lineare, sull'accessibilità, anche gratuita, ai dati prodotti dalle pubbliche amministrazioni, che la legge ha sancito essere un dovere da parte di queste ultime (cfr. CAD, art. 52, con il quale si stabilisce il principio dell'"open data by default" per il quale tutti i dati e documenti prodotti dalle pubbliche amministrazioni, nel caso in cui non rilasciati con specifici termini d'uso, vengono considerati alla stregua di dati licenziati in modalità aperta).
2. Dal momento che si assiste a questa sempre maggiore produzione e distribuzione di dati e vi si può accedere liberamente, è doveroso chiedersi se e come questi possano essere impiegati nella didattica di una scuola che è chiamata a interrogarsi costantemente sui cambiamenti che la società fronteggia. In altre parole, possono essere utili queste moli di dati nella didattica? aiutano i ragazzi a costruire nuova conoscenza, acquisire competenze e abilità particolari, acuire senso critico nell'esame della fonte, ecc. o al contrario costituiscono rumore?

Per stimolare in voi la riflessione su questi due punti, ho pensato di farvi "sporcare le mani" cercando di rispondere assieme ad alcune delle seguenti domande:

- a. quali sono i dati prodotti e accessibili? dove si trovano? in quali formati sono distribuiti? sotto quali licenze?
- b. che cosa significa metadato? riflettiamo sull'importanza di saper descrivere bene una risorsa in rete
- c. con quali applicazioni posso gestire e utilizzare i dati a cui ho accesso?

d. come valuto i dati a cui accedo?

Intanto per prima cosa vi suggerisco un paio di riferimenti bibliografici utili, se non indispensabili: il primo è [Il manuale degli Open Data](#), traduzione in italiano di The Open Data Handbook, redatto dalla Open Knowledge Foundation e il testo di Simone Alibrandi, [Il fenomeno open data. Indicazioni e norme per un mondo di dati aperti](#), Versione 1, febbraio 2014. Utile lettura anche le Linee guida per i siti web della PA. Vademecum Open Data Come rendere aperti i dati delle pubbliche amministrazioni <http://www.funzionepubblica.gov.it/media/982175/vademecumopendata.pdf> (PDF 269 KB)

Cominciamo dal punto a.

In Italia nell'ottobre del 2011 è stato inaugurato il portale italiano dei dati aperti: [dati.gov.it](http://dati.gov.it), che segna una tappa importante in questo percorso di accesso ai dati, creato sul modello degli omologhi siti di matrice anglosassone, si pensi tra i primi al U.S. Government's open data ([www.data.gov](http://www.data.gov)) o al britannico [data.gov.uk](http://data.gov.uk), entrambi rilasciati nel 2009. Il portale italiano risponde agli stimoli dati dalla strategia europea per gli Open Data all'interno dell'[Agenda digitale europea](#)

Ma già precedentemente a questa data, in Italia si cominciava a parlare di accesso aperto al dato in connessione con il progetto OpenStreetMap, nato per la creazione e condivisione gratuita di dati cartografici, proprio per ovviare alle restrizioni legali o tecniche, che nella maggior parte dei casi ne impedivano l'uso o il riuso, poiché, contrariamente a quanto comunemente si pensa, gran parte dei dati geografici vengono rilasciati liberamente solo per usi parziali.

Sono state di fatto alcune Regioni o amministrazioni comunali che per prime hanno avuto la sensibilità di aggiornare le proprie procedure interne riguardanti la produzione e gestione dei dati e conseguentemente a predisporre delle piattaforme per il pubblico accesso.

Citiamo ad esempio il sito della Regione Piemonte (<http://www.dati.piemonte.it/>) rilasciato nel 2010, il sito della Regione Emilia Romagna (<http://dati.emilia-romagna.it/>), 2011, di Veneto, Lombardia e Lazio nel 2012 (<http://dati.veneto.it/>; <https://www.dati.lombardia.it/>; <https://dati.lazio.it/it>), Trentino, 2013 (<http://dati.trentino.it/>), Comune di Firenze (<http://opendata.comune.fi.it/>).

Attualmente il sito di riferimento da cui partire nella nostra esplorazione è il portale dell'Agenzia per l'Italia Digitale (<http://www.agid.gov.it/>) che fa da collettore di altri progetti istituzionali legati al tema dell'innovazione e crescita digitale, tra cui segnalo il già richiamato [dati.gov.it](http://dati.gov.it), il Repertorio Nazionale Dati Territoriali, che si prefigge di essere "lo strumento per ricercare, attraverso i metadati, i dati territoriali - e i relativi servizi - disponibili presso le Pubbliche Amministrazioni, per valutarne l'idoneità allo scopo e ottenere le opportune indicazioni sulle loro condizioni di accesso e utilizzo (<http://www.rndt.gov.it/RNDT/home/index.php>) e il portale Basi di dati, il Catalogo che contiene le informazioni relative alle basi di dati "non territoriali", comunicate dalle amministrazioni competenti (per le basi di dati territoriali il catalogo di riferimento è il RNDT) e risponde all'obiettivo di "facilitare la diffusione e la conoscenza dei dati della P.A. e favorisce, la condivisione dei dati tra pubbliche amministrazioni e il riutilizzo degli stessi secondo i principi dell'open data. Questo catalogo è anche una delle componenti della Piattaforma nazionale delle Comunità intelligenti" (<http://basidati.agid.gov.it/catalogo/>).

Altre iniziative interessanti, che vi voglio ricordare sono:

DatiOpen.it, il portale italiano dell'Open Data, un'iniziativa indipendente che si prefigge di rendere gli open data usabili dal grande pubblico, attraverso la raccolta e documentazione della maggiore quantità possibile di dati open italiani e permettendo la visualizzazione direttamente dal sito dei dati, in tabelle, grafici e mappe. (<http://www.datiopen.it/>).

Spaghetti OpenData, che anima una community di "cittadini italiani interessati al rilascio di dati pubblici in formato aperto, in modo da renderne facile l'accesso e il riuso (open data)". (<http://www.spaghetiopendata.org/>).

Open Data Hub, un catalogo con i dati aperti di organizzazioni pubbliche e private, che offre la possibilità di ricercare attraverso un motore di ricerca trasversale e specializzato oltre 13000 dataset pubblicati da organizzazioni pubbliche e private (<http://www.sciamlab.com/opendatahub/>).

Linked Open Data, associazione senza fini di lucro i cui membri sono grandi "fan delle tecnologie Web e Semantic Web e convinti che rendere di dominio pubblico e facilmente fruibili (riusabili) i dati grezzi, istituzionali e non, sia un importante passo in una società moderna e apra nuove ed interessanti prospettive sia sul piano della trasparenza, ma soprattutto su quello dell'offerta di servizi innovativi ai cittadini ed alle imprese"(<http://www.linkedopendata.it/>).

OpenGeoData, associazione nata con l'obiettivo di "diffondere la cultura dei dati geografici, di stimolare il loro riuso e di convincere le Pubbliche Amministrazioni a renderli aperti". (<http://www.opengeodata.it/>).

E questo per dare una risposta ai primi due quesiti (quali sono i dati prodotti e accessibili? dove si trovano?)

Cerco ora, senza tediarvi troppo, di rispondere agli altri del punto a (in quali formati sono distribuiti? sotto quali licenze?).

Riguardo ai formati vi rimando al [Manuale sugli OD, § Rendere i dati disponibili \(Aspetti tecnici\)](#) in cui si chiarisce che dovendo essere i dati aperti anche dal punto di vista tecnico, essi devono soddisfare le seguenti caratteristiche: essere "Available", "In bulk" (disponibili come insieme completo) e "In an open, machine-readable format" cioè processabili da una macchina poiché "(...) fornire i dati in formato "machine-readable" consente un loro maggior riutilizzo. Per chiarire ciò, si consideri il caso di statistiche pubblicate come documenti PDF (Portable Document Format), spesso utilizzati per la stampa di alta qualità. Anche se queste statistiche possono essere lette da esseri umani, è molto difficile renderle utilizzabili dai computer e questo limita pesantemente la capacità da parte di altri di riutilizzare quei dati".

Inoltre i dati devono essere distribuiti in formati non proprietari, ma al contrario "devono essere codificati in formati aperti e pubblici, sui quali non vi siano entità (aziende o organizzazioni) che ne abbiano il controllo esclusivo. Sono preferibili i formati con le codifiche più semplici e maggiormente supportati" (Vademecum OpenData, p. 21). Un importante concetto da tenere presente è il passaggio dal cosiddetto "raw data" al "linked open data". Cito sempre dal Vademecum, p. 22, "Qualora l'Amministrazione si trovi nella possibilità di scegliere il formato nel quale pubblicare i dati da aprire, è bene che identifichi formati che siano rispondenti al maggior numero di requisiti possibile. Se tuttavia la scelta fosse tra il pubblicare dati non completamente rispondenti ai requisiti indicati o non pubblicarli affatto, allora la logica dell'Open Data indirizza la

risposta verso la prima soluzione (per questo si usa l'espressione Raw Data Now, a significare che nella peggiore delle ipotesi, quella di disporre di dati non aperti, è comunque preferibile distribuirli, anche se – appunto – in formato raw, cioè grezzo)"

Tim Berners Lee, l'inventore del Web, ha suggerito uno metodo di sviluppo a 5 stelle per gli Open Data, per ciascuna fase del quale egli esplicita costi e benefici (<http://5stardata.info/> e <http://www.w3.org/DesignIssues/LinkedData.html> che consiglio di leggere a chi voglia approfondire l'argomento), che qui riporto in sintesi:

★ make your stuff available on the Web (whatever format) under an open license example

★★ make it available as structured data (e.g., Excel instead of image scan of a table)

★★★ use non-proprietary formats (e.g., CSV instead of Excel)

★★★★ use URIs to denote things, so that people can point at your stuff

★★★★★ link your data to other data to provide context

Sempre nel Vademecum italiano, p. 22 ss., se ne dà una esplicitazione:

**"Una Stella.** È il livello base, costituito da file non strutturati: ad esempio un'immagine in formato grezzo (formati come .gif, .jpg, .png), un documento in formato Microsoft Word, un file in formato Adobe Pdf. Una sola stella indica la semplice disponibilità di una informazione e di un dato on line, in un formato qualsiasi, purché distribuito con licenza aperta. I dati distribuiti in questo formato sono leggibili e stampabili dagli utenti, possono essere conservati localmente su un PC e sono semplici da pubblicare. Tuttavia non sono un formato aperto in quanto non è possibile effettuare su di essi alcuna elaborazione.

**Due Stelle.** Questo livello indica dati strutturati ma codificati con un formato proprietario. Ad esempio un documento in formato Microsoft Excel. Due stelle indicano, oltre alle possibilità offerte dai dati contraddistinti da una sola stella, la possibilità di effettuare elaborazioni sui dati, a patto di disporre del software necessario a gestire un file codificato con un formato proprietario. I dati caratterizzati dalle due stelle non sono un formato aperto in quanto per elaborarli è necessario un software proprietario, tuttavia di norma possono essere convertiti – essendo dati strutturati – in dati aperti.

**Tre Stelle.** Questo livello indica dati strutturati e codificati in un formato non proprietario. Ad esempio il formato .csv (Comma Separated Values) al posto – ad esempio – del formato Microsoft Excel utilizzato nel caso precedente. (...) Quello caratterizzato dalle tre stelle è il formato più semplice di dati aperti.

**Quattro Stelle.** Questo livello indica dati strutturati e codificati in un formato non proprietario che sono dotati di un URI [Uniform Resource Identifier - Identificatore Univoco di Risorsa] che li rende indirizzabili sulla rete e quindi utilizzabili direttamente online, attraverso l'inclusione in una struttura basata sul modello RDF (Resource Description Framework). Quattro stelle indicano quindi il fatto che il singolo dato di un dataset, disponibile on line in un formato aperto (tipicamente XML/RDF) può essere richiamato attraverso un'URL (Uniform Resource Locator) specifico. Ciò consente di puntare al dato o ad un insieme di dati da una applicazione o accedervi dall'interno di un programma che può poi elaborarlo in vari modi. Si pensi, ad esempio, a un dataset contenente gli

indirizzi dei monumenti di una città opportunamente codificati: da qualsiasi software – anche dal browser – è possibile collegarsi all'URL che indica il singolo monumento, potendolo ad esempio georeferenziare su una mappa.

**Cinque Stelle.** Questo livello indica quelli che vengono definiti *Linked Open Data (LOD)*. Quei dati aperti, cioè, che – dal punto di vista del formato – oltre a rispondere alle caratteristiche indicate al punto precedente presentano anche, nella struttura del dataset, collegamenti ad altri dataset. In altri termini, grazie al ricorso al già citato modello di descrizione dei dati RDF, è possibile collegare dinamicamente tra loro più dataset, incrociando così informazioni provenienti da fonti diverse, eventualmente gestite da diverse Amministrazioni (...)"

Vi segnalo a questo proposito come esempio il portale della Regione Umbria, che ha pubblicato le proprie [Linee guida per la pubblicazione di Open Data](#), nelle quali "sono descritti i processi operativi per gestire tutte le fasi del ciclo di vita di un dataset dal momento della sua individuazione fino alla pubblicazione nel catalogo e vengono individuati i soggetti che intervengono, ciascuno nel proprio ambito di competenza, per la risoluzione di eventuali problematiche (come privacy, segreto statistico, diritto d'autore, comunicazione) connesse alla gestione del contenuto informativo presente nel dataset." (<http://agendadigitale.regione.umbria.it/category/openness/open-data/> e <http://dati.umbria.it/dataset/linee-guida-per-la-pubblicazione-di-open-data>).

Per quanto attiene alla questione delle licenze con cui i dati sono distribuiti, vi richiamo intanto a quanto già detto in questo corso riguardo al Copyleft e alle Creative Commons Licenses. I dati aperti devono essere rilasciati corredati da licenze che non ne limitino l'uso, la diffusione o la redistribuzione. Di fatto licenze aperte. Non mi addentro oltre nella questione, ma vi rimando al testo a cura di Simone Alibrandi, in particolare al cap. 4 "Il licensing dei dati in modalità open"

Vi suggerisco brevemente di dare un'occhiata alla Italian Open Data License, indicata come IODL v2.0 rilasciata dal FormezPA, priva di clausole del tipo "condividi-allo-stesso-modo" e con la sola richiesta di attribuzione della fonte per il riutilizzo dei dati. La licenza è progettata per i dati delle pubbliche amministrazioni italiane. ([IODL v.1.0](#) e [IODL v.2.0](#)).

Ma andiamo avanti...

*b. che cosa significa metadato? riflettiamo sull'importanza di saper descrivere bene una risorsa in rete*

Per metadato si intende un dato che descrive un altro dato, in altre parole i metadati costituiscono le informazioni descrittive a corredo di un dataset. Maggiore è il grado di strutturazione sintattica e semantica dei metadati, più facile è per noi reperire il dataset di cui abbiamo bisogno e darne una prima valutazione.

Tutti noi abbiamo esperienze di metadati, probabilmente per tutti la prima frequentazione "seria" che ci è capitata è quella con le schede catalografiche di una biblioteca, senza le quali sarebbe impossibile trovare la risorsa che stiamo cercando, come anche quella che non stiamo cercando.

Riporto dalle Linee guida siti web PA.

[Capitolo 4 - Criteri di indirizzo e strumenti per garantire la qualità dei siti web delle pubbliche](#)

## [amministrazioni. Contenuti minimi dei siti istituzionali delle PA](#)

"Al fine di favorire la ricerca delle informazioni essenziali contenute nei siti istituzionali [...] si consiglia di aggiungere al testo delle pagine web informazioni aggiuntive che ne descrivono determinate caratteristiche e i principali contenuti. Infatti l'HTML permette di inserire nei documenti delle parole-chiave, meta-dati, che ne descrivano il contenuto per consentire ai motori di ricerca di ottimizzare il recupero dell'informazione.

I metadati, definiti secondo lo standard [Dublin Core](#) recepito come norma [ISO 15836:200](#), hanno infatti due funzioni principali: individuare e identificare un documento e informare sulle sue caratteristiche."

e dal [Capitolo 5 - Criteri d'indirizzo e strumenti per il trattamento dei dati, della documentazione pubblica e per la loro reperibilità. Classificazione e semantica](#)

"I sistemi di classificazione utilizzati per le risorse dei siti web della pubblica amministrazione devono consentire l'interoperabilità semantica, ovvero la possibilità di individuare in modo omogeneo gli attributi che caratterizzano una risorsa (metadati) e i valori che gli attributi possono assumere (vocabolari) quando si descrivono i contenuti." A tal fine si raccomanda "[...] di utilizzare lo standard di metadati Dublin Core (DC). Accompagnare le risorse web (informazioni, documenti, pagine web, banche dati) con metadati che le descrivano ne consente l'identificazione univoca e stabile, ne agevola la classificazione anche nei repository istituzionali, ne facilita la ricerca"

Per chi volesse approfondire, rimando alle *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico*, paragrafo 4.2.

Per chi volesse invece provare direttamente che cosa significa metadato, consiglio un'esplorazione diretta su uno dei portali che vi ho indicato sopra.

Proviamo insieme:

Vogliamo effettuare una ricerca tra i dataset di Dati.gov.it, ci interessano, per esempio, i dati relativi a Occupazione e lavoro e seleziono il tema "Lavoro e politiche sociali" e ci vengono restituiti 6 dataset., ciascuno dei quali viene presentato, come possiamo vedere, con una breve scheda di sintesi che comprende titolo, autore, data di aggiornamento, abstract, formati ecc. Quindi di fatto ci viene data un'immediata descrizione sintetica e standardizzata del dataset che stiamo andando a consultare e/o scaricare.

Diamo un occhio ai formati in cui i dati sono disponibili: csv, rdf, xls, xml, kmz, html, shape, json (questi ultimi tre sono formati per i dati geografici).

Decidiamo di visualizzare il dataset Osservatorio sulle imprese - Numeri indice delle retribuzioni medie annue del settore artigianato, pubblicato da INPS, e scelgo il formato xml.

Quando la pagina si apre (<http://www.inps.it/docallegati/Mig/OpenData/PQ2MJJ87.xml>), proviamo a leggere cosa c'è scritto. Vi siete già confrontati con il linguaggio HTML, quindi non vi spaventate, non è necessario che comprendiamo tutto, ma che riusciamo a scorgere quello che ci interessa:

```
<metadata>
```

```
<dc:publisher>INPS</dc:publisher>
```



<dc:title>

Osservatorio sulle imprese - Numeri indice delle retribuzioni medie annue del settore artigianato

</dc:title>

<dcterms:abstract>

Osservatorio sulle imprese, occupati dipendenti del settore privato non agricolo e retribuzioni medie annue di operai e impiegati - Dati sui numeri indice delle retribuzioni del settore artigiano, divisi per dimensione d'impresa e anno

</dcterms:abstract>

<dcterms:issued>17/01/2012</dcterms:issued>

<dcterms:license>IODL 2.0</dcterms:license>

<dcterms:LicenseDocument><http://www.dati.gov.it/iodl/2.0/></dcterms:LicenseDocument>

<dc:identifier>

<http://www.inps.it/docallegati//Mig/OpenData/PQ2MJJ87.xls>

</dc:identifier>

<facets>

<argomento>

<dc:subject>Lavoro.LavoroArtigiano</dc:subject>

<dc:subject>Lavoro.Azienda</dc:subject>

</argomento>

<fonte>

<dcterms:source>Osservatori statistici</dcterms:source>

</fonte>

<periodo>

<dcterms:temporal>2001-2005</dcterms:temporal>

<dcterms:temporal>2006-2011</dcterms:temporal>

</periodo>

</facets>

</metadata>

I tag <metadata> e </metadata> ci riportano lo schema dei metadati descrittivi della risorsa. La sigla 'dc' aiuta a capire a quale standard si rifà lo schema.

Non tutti i casi sono fortunati come questo. Inoltre se decidiamo di scaricare lo stesso dataset per esempio in formato xls (quindi un formato non aperto, ma comunque strutturato, quindi a 2 stelle) potremo leggere il dato, ma non è contenuto in esso alcun riferimento al metadato.

*c. con quali applicazioni posso gestire e utilizzare i dati a cui ho accesso?*

In parte avrete già capito, anche dal punto precedente, con quali applicazioni possiamo da utenti finali utilizzare questi dati. Mi limiterei qui a considerare i dati pubblicati in formati che vanno da 1 a 3 stelle, per tornare alla classificazione di Tim Berners Lee e in particolare i dati rilasciati in formato csv. L'acronimo significa comma-separated values e indica un formato basato su file di testo mediante il quale è possibile importare ed esportare una tabella di dati, i cui campi vengono separati da un carattere, che va esplicitato al momento dell'importazione o esportazione e può essere una virgola, un punto e virgola, una tabulazione o altro.

Provate ora a ricercare un dataset di vostro interesse e scaricarlo in formato csv. Quando lo avrete scaricato dovrete scegliere con quale applicazione aprirlo. Trattandosi, come si è detto, di un file in cui sono contenuti dati strutturati per campi, la cosa più utile da fare è quella di scegliere un programma che gestisca dati strutturati in tabella. Nei primi passi di questo corso avete scaricato e cominciato ad utilizzare programmi aperti come OpenOffice e LibreOffice, ciascuno dei quali contiene un'applicazione utile al caso: Calc. Definitela come applicazione per aprire il file. Ora fate attenzione ai passaggi per l'importazione.

In modo analogo vi comporterete se decidete di aprire il file con un programma proprietario come Excel, o di importarlo in un database, come ad esempio Access, con la differenza che dovrete prima lanciare il programma e poi caricare i dati da una sorgente esterna (il file csv) e quindi importarli facendo sempre attenzione alla strutturazione dei campi e a quale separatore è stato definito.

Per chi volesse provare a giocare un pochino con i dati geografici, possiamo pensare a un approfondimento ad hoc, per ora ho pensato di non gravare troppo sulla vostra pazienza, ma sono sicura che vi divertireste.

*d. come valuto i dati a cui accedo?*

Intanto direi che il primo grado di valutazione è meramente soggettivo: mi sono utili i dati che ho scaricato? rispondono al mio bisogno? sono esaustivi per i miei scopi? correttamente documentati? effettivamente riusabili? ecc.

Per un secondo livello di valutazione della qualità che risponda a criteri di oggettività, vi ricordo che secondo la Norma ISO/IEC 25012 – Data Quality Model, esistono dei parametri utili anche per i dataset aperti, e in particolare:

**Accuratezza:** grado in cui gli attributi del dato rappresentano correttamente il valore reale di concetti o eventi in uno specifico contesto d'uso. Può essere di natura sintattica e/o semantica.

**Attualità:** adeguatezza del valore del dato rispetto alle esigenze temporali del contesto d'uso.

**Completezza:** l'estensione con cui i valori dei dati coprono l'universo cui sono riferiti.

**Consistenza:** grado in cui gli attributi del dato sono non contraddittori e coerenti con altri dati in uno specifico contesto d'uso. Può essere **interna** quando indica il grado di coerenza dei dati presenti in un dataset riferiti ad una stessa entità; **esterna** quando indica il grado di coerenza tra dati diversi ma correlati presenti tra elementi di un dataset.

**Credibilità:** il grado in cui gli attributi del dato sono considerati veri, riconoscibili (autentici, certificati) dagli utenti in uno specifico contesto d'uso (Cfr. Domenico Natale, Data quality e Open data, Convegno AICA, Torino, 15-17 novembre 2011).

In conclusione...

Ho cercato di darvi alcuni spunti su un tema che come vi ho detto nelle premesse è in continuo divenire e nei prossimi anni credo ci dispiegherà scenari di sviluppo e di utilizzo interessanti. Quello che ci piaceva è vedere quale può essere, se c'è ovviamente, la risposta della scuola. Per questo ora... lascio la parola a voi :-)

**Riferimenti utili:**

[Digital Agenda for Europe. A Europe 2020 Initiative,](#)

[Agenda Digitale Italiana,](#)

[Open Knowledge Foundation, Open Data Handbook,](#) disponibile in versione italiana

[Agenzia per l'Italia Digitale, Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico, 2014](#) (PDF 3.8MB)

[Linee guida per i siti web della PA. Vademecum Open Data Come rendere aperti i dati delle pubbliche amministrazioni](#) (PDF 269KB)

Simone Aliprandi, Il fenomeno open data. Indicazioni e norme per un mondo di dati aperti

<http://www.aliprandi.org/fenomeno-opendata/>

Versione 1, febbraio 2014

ISBN: 9788867050116

G8 Open Data Charter,  
[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/207772/Open\\_Da](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Da)  
(PDF 307KB)

Open Government Partnership Action Plan italiano,  
<http://www.funzionepubblica.gov.it/media/968937/piano%2011%20aprile%20%20opengovpartners>  
(PDF 184KB)