

OPEN DATA E PUBBLICA AMMINISTRAZIONE

MAURO LENZI

Ingegnere

SOMMARIO

L'Open data esprime l'idea che i dati dovrebbero essere liberamente disponibili a tutti per essere utilizzati, senza restrizioni legate a copyright, licenze od altri meccanismi di controllo. Il dibattito sugli Open Data è in costante evoluzione. In una ottica di "Open Government", un crescente numero di enti pubblici sta mettendo a disposizione dei cittadini i propri dati, per favorire la circolazione delle informazioni, la partecipazione, le attività commerciali o creare valore con altre possibili modalità.

SUMMARY

Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. The debate on Open Data is still evolving. The best open government applications seek to empower consumers, to help small businesses, or to create value in some other positive, constructive way.

PREMESSA

Nella società dell'informazione, caratterizzata da internet e da sistemi di calcolo e di gestione di basi dati sempre più potenti ed interconnessi, esiste una quantità vastissima di informazioni contenute in archivi elettronici, rappresentate in forma di database e riferite alla tematiche più disparate, quali ad esempio: cartografia, genetica, composti chimici, formule matematiche e scientifiche, medicina, dati anagrafici, dati governativi, ecc. Queste informazioni, possedute da Enti Pubblici ed Aziende Private, sia in forma aggregata che dettagliata, possono avere un valore commerciale, e quindi essere rese disponibili solo acquistandole, oppure essere inaccessibili, in quanto considerate riservate. E' sentimento diffuso che esponendo all'esterno informazioni, si possa avvantaggiare la concorrenza o possa esporre a critiche o contestazioni.

La maggior parte del patrimonio informativo esistente viene perciò protetto dalle organizzazioni che li detengono, sia pubbliche che private, limitandone l'accesso agli utenti, definendo licenze che ne li-

mitano l'uso e la diffusione o con specifici copyright. Di fronte a queste forme di controllo sui dati, e più in generale sulla conoscenza, i sostenitori dell'Open Data affermano che tali restrizioni siano un limite al bene della comunità e che i dati dovrebbero essere resi disponibili senza alcuna restrizione o forma di pagamento. Inoltre, è importante che i dati, dopo essere stati pubblicati, siano riutilizzabili senza necessità di ulteriori autorizzazioni, anche se determinate forme di riutilizzo (come la creazione di opere derivate) possono essere controllate attraverso specifiche licenze (ad esempio Creative Commons, GFDL).

D'altra parte, accade spesso che gli stessi creatori di dati sottovalutino l'importanza delle informazioni di cui dispongono e non considerino la necessità di precisare le condizioni della proprietà intellettuale, delle licenze e del loro riutilizzo. Ad esempio, molti enti (siano essi di natura scientifica o governativa) per mancanza di consapevolezza dell'importanza dei propri dati, non prendono in considerazione l'ipotesi di rilasciarli con licenze aperte. La mancanza di una determinata licenza che certifichi la possibilità di riutilizzare i dati rende difficile determinare lo stato di un insieme di dati e ne limita l'uso.

Una dichiarazione di *John Wilbanks*, direttore esecutivo dello *Science Commons* evidenzia come:

«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.»

DEFINIZIONE DI OPEN DATA

Nonostante il concetto di dato aperto non sia nuovo, al momento non si riscontra un accordo generale e condiviso su di una definizione puntuale del termine, a differenza, ad esempio, di quanto già avviene con il software libero o l'open source dove diverse dichiarazioni formali sono state comunemente accettate e condivise a livello internazionale. In generale si potrebbe dire che con Dati Aperti, comunemente chiamati con il termine inglese

Open Data anche nel contesto italiano, 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.

I dati aperti, comunemente chiamati con il termine inglese open data anche nel contesto italiano, sono alcune tipologie di dati liberamente accessibili a tutti, privi di brevetti o altre forme di controllo che ne limitino la riproduzione e le cui restrizioni di copyright eventualmente si limitano ad obbligare di citare la fonte o al rilascio delle modifiche allo stesso modo. L'open data si richiama alla più ampia disciplina dell'open government, cioè una dottrina in base alla quale la pubblica amministrazione dovrebbe essere aperta ai cittadini, tanto in termini di trasparenza quanto di partecipazione diretta al processo decisionale, anche attraverso il ricorso alle nuove tecnologie dell'informazione e della comunicazione; e ha alla base un'etica simile ad altri movimenti e comunità di sviluppo "open", come l'open source, l'open access e l'open content. Nonostante la pratica e l'ideologia che caratterizzano i dati aperti siano da anni ben consolidate, con la locuzione "open data" si identifica una nuova accezione piuttosto recente e maggiormente legata a Internet come canale principale di diffusione dei dati stessi.

Figura 1 - La definizione di Dati Aperti (Open Data), estratta da Wikipedia

Nella Figura 1 si è riportato una definizione di Dati Aperti estratta da Wikipedia; un'altra definizione comunemente accettata di Open Data è quella fornita dall'Open Data Manual, redatto dall'Open Knowledge Foundation (OKF), che è una fondazione non profit, fondata nel 2004 con l'obiettivo di promuovere la conoscenza aperta:

«dati che possono essere liberamente utilizzati, riutilizzati e redistribuiti, con la sola limitazione – al massimo – della richiesta di attribuzione dell'autore e della redistribuzione allo stesso modo (ossia senza che vengano effettuate modifiche)»

Secondo i sostenitori del movimento Open Data, i dati andrebbero trattati come beni comuni; di seguito alcune delle argomentazioni utilizzate a sostegno di questa tesi:

- I dati appartengono al genere umano. Esempi tipici sono i genomi, i dati sugli organismi per la scienza medica, dati ambientali e meteorologici, ecc.
- Restrizioni sui dati e sul loro riutilizzo limitano lo sviluppo della comunità
- I dati sono necessari per agevolare l'esecuzione di comuni attività umane (ad esempio i dati cartografici, le istituzioni pubbliche, ecc.)
- In campo scientifico il tasso di scoperta è accelerato da un migliore accesso ai dati

- È essenziale che i dati scientifici siano resi aperti per fare in modo che la scienza sia più efficace e la società ottenga il massimo beneficio dalle ricerche scientifiche

Ulteriori e più forti motivazioni vengono rivolte dai sostenitori degli Open Data verso gli Enti Pubblici, in quanto si sostiene che:

- I dati prodotti dalla pubblica amministrazione, in quanto finanziati da denaro pubblico, devono ritornare ai contribuenti, e alla comunità in generale, sotto forma di dati aperti e universalmente disponibili
- Molte informazioni e dati prodotti dagli enti pubblici (delibere, bilanci, curricula, stipendi, etc.) debbono essere disponibili e facilmente accessibili per motivi di trasparenza e per consentire ai contribuenti di verificare azioni e comportamenti degli enti e degli amministratori dei beni pubblici

OPEN DATA NEL MONDO, IN ITALIA E NELLA NOSTRA REGIONE

Sono molteplici le iniziative d'apertura del patrimonio informativo avviate nel mondo. Il primo e più celebre, che ha sostanzialmente fatto da modello ai successivi, è stato *data.gov* americano, lanciato dal governo Obama a seguito della Direttiva sull'Open government nel dicembre 2009; di lì a pochi mesi anche il Regno Unito ha aperto il suo *data.gov.uk*, fortemente voluto e sponsorizzato da Tim Berners-Lee, l'inventore del World Wide Web. In pochi anni la pratica degli open data e dei data store governativi si è estesa fino in Australia con *data.gov.au*, Canada con *data.gc.ca*, Norvegia con *data.norge.no*, Francia con *data.gouv.fr*, ect.

Il primo data store italiano è stato quello della Regione Piemonte, *dati.piemonte.it*, all'interno del quale sono catalogati dati aperti riconducibili ai vari enti regionali (comuni, province,...). Dopo circa un anno dalla nascita del data store piemontese anche la Regione Emilia-Romagna, ad ottobre 2011, pubblica online il suo catalogo di dataset *dati.emilia-romagna.it*.

Con il lancio del portale *dati.gov.it*, avvenuto il 18 ottobre 2011, si è aperta una nuova stagione per l'innovazione e la trasparenza nella PA, una strada verso l'Open Data italiano. Il processo di gestione delle informazioni prodotte sta progressivamente evolvendo, verso modelli più aperti. Per rimanere aggiornati su tutte le iniziative

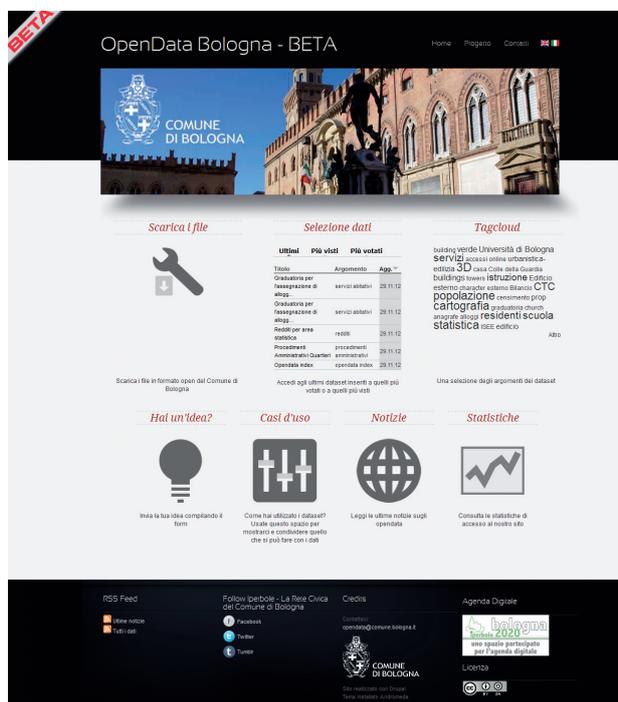


Figura 2 - Il portale sugli Open Data in fase di rilascio del Comune di Bologna

di Open Data avviate in Italia, la redazione di *dati.gov.it* ha creato un dataset che descrive i dati catalogati sul sito e li rende maggiormente interpretabili attraverso un'infografica. Il dataset alla base dell'infografica viene periodicamente aggiornato.

Attualmente anche moltissime amministrazioni pubbliche locali, quali comuni e provincie, hanno una sezione del proprio portale dedicata alla pubblicazione di Open Data. Ad esempio, il Comune di Bologna ha reso disponibili moltissimi dati ed ha in corso di ultimazione un portale dedicato a questo tema, *dati.comune.bologna.it* (vedi Figura 2): in esso vi si trovano tantissimi dati inerenti svariati argomenti: bilancio, cartografia, cultura, finanze, mobilità, popolazione, servizi amministrativi, scuole, urbanistica, edilizia, etc.

RIFERIMENTI NORMATIVI

I riferimenti normativi che interessano il tema degli Open Data sono assai numerosi, considerando che occorre considerare problematiche inerenti il diritto d'autore, la privacy, la sicurezza, la trasparenza, l'accessibilità, il riuso, e altri ancora.

Riferimento Normativo	Contenuto
Legge n. 633/1941	Legge a protezione del Diritto d'Autore e di altri diritti connessi al suo esercizio.
Legge n. 241/1990	Nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti amministrativi. Gli articoli 22 eseguenti disciplinano le specifiche fattispecie del diritto di accesso ai documenti amministrativi.
D. Lgs. n. 196/2003	Codice in materia di protezione dei dati personali; si occupa di problemi inerenti la privacy e la sicurezza.
Legge n. 4/2004	Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici; definisce regole per l'accessibilità alle informazioni
D. Lgs. n. 82/2005	Codice dell'Amministrazione; contiene normative inerenti il Diritto dei cittadini all'uso delle tecnologie, l'E-democracy, l'E-partecipation, la Titolarità dei dati pubblici, i Formati aperti. Introduce l'importante principio di "disponibilità dei dati pubblici" (enunciato all'art. 2, comma 1, e declinato dall'art. 50, comma 1, dello stesso Codice) che consiste nella possibilità, per soggetti pubblici e privati, "di accedere ai dati senza restrizioni non riconducibili a esplicite norme di legge" (art.1, lett. o)
D. Lgs. n. 163/2006	Codice dei contratti pubblici in materia di lavori, servizi e forniture; tratta di Contratti e della Titolarità dei contenuti acquisiti dalle Amministrazioni
D. Lgs. n. 150/2009	Introduce il concetto di trasparenza intesa come "accessibilità totale (...) delle informazioni concernenti ogni aspetto dell'organizzazione, degli indicatori relativi agli andamenti gestionali e all'utilizzo delle risorse per il perseguimento delle funzioni istituzionali, dei risultati dell'attività di misurazione e valutazione"
D. Lgs. n. 32/2010 in attuazione della direttiva 2007/2/CE	Attribuisce a ciascuna Amministrazione la possibilità di autorizzare il riutilizzo delle informazioni che vengono raccolte, prodotte, e diffuse nell'ambito del perseguimento dei propri compiti istituzionali.
D. Lgs. n. 235/2010	Riforma del Codice dell'Amministrazione Digitale. Recepisce espressamente la dottrina dell'Open Data, sollecitando le Amministrazioni ad aprire il proprio patrimonio informativo; nella sua attuale formulazione, infatti, l'art. 52, comma 1-bis, D. Lgs. n. 82/2005, prevede espressamente che "le pubbliche amministrazioni, al fine di valorizzare e rendere fruibili i dati pubblici di cui sono titolari, promuovono progetti di elaborazione e di diffusione degli stessi anche attraverso l'uso di strumenti di finanza di progetto", utilizzando formati aperti che ne consentano il riutilizzo.
D. Lgs. n.235/2010	La Riforma del Codice dell'Amministrazione Digitale recepisce espressamente la dottrina dell'Open Data, sollecitando le Amministrazioni ad aprire il proprio patrimonio informativo; nella sua attuale formulazione, infatti, l'art. 52, comma 1-bis, D. Lgs. n. 82/2005, prevede espressamente che "le pubbliche amministrazioni, al fine di valorizzare e rendere fruibili i dati pubblici di cui sono titolari, promuovono progetti di elaborazione e di diffusione degli stessi anche attraverso l'uso di strumenti di finanza di progetto", utilizzando formati aperti che ne consentano il riutilizzo.
D. Lgs n. 179/2012	Vedi quanto riportato nel capitolo corrispondente

Figura 3 - Tabella con il quadro dei riferimenti normativi su tematiche inerenti i Dati Aperti

La maggior parte delle norme di interesse sono state emanate a partire dal 1990 e le ultime norme sono recentissime, essendo contenute nel D. Lgs n. 179/2012, pubblicato in gazzetta il 27 ottobre 2012; il complesso di questi provvedimenti evidenzia un rapido percorso evolutivo verso un'amministrazione più aperta e trasparente.

Descrivere nel dettaglio le norme inerenti questo tema esula dallo scopo di questo articolo; si è cercato comunque di riportare nella tabella di Figura 3 un breve riepilogo di alcuni tra i riferimenti più significativi.

Un discorso a parte meritano le su citate recentissime "Ulteriori misure urgenti per la crescita del Paese" contenute nel decreto legge 18 ottobre 2012, n. 179 e pubblicate Gazzetta Ufficiale il 19 ottobre 2012. In tale decreto, ed in particolare nell'articolo 9, vi sono infatti alcuni importantissimi aggiornamenti sulla materia. Non essendo l'Autore di questo articolo particolarmente esperto su tematiche di diritto pubblico, si è preferito riportare integralmente nel seguito la Relazione Illustrativa dell'Articolo 9:

ARTICOLO 9

La norma intende razionalizzare il

processo di valorizzazione del patrimonio informativo pubblico nazionale, introducendo l'obbligo per le pubbliche amministrazioni di rendere disponibili i dati pubblici in formato aperto e identificando un organismo responsabile che vigili sulla corretta attuazione delle disposizioni di legge. Si modifica così (comma 1, lett. a) il comma 1 dell'art. 52 CAD, chiarendo che la disciplina ivi contenuta non si riferisce soltanto all'accesso ma anche al riutilizzo dei dati pubblici. In questo modo si focalizza l'attenzione sul processo di utilizzo ovvero sulla possibilità di impiegare i dati per realizzare nuove analisi, servizi e soluzioni che hanno il potenziale di generare crescita economica e sociale. Viene introdotto l'obbligo per le amministrazioni di pubblicare l'elenco dei dati e dei metadati in loro possesso al fine di costituire un indice nazionale del patrimonio informativo pubblico. Infine è indicato che le amministrazioni devono, entro 90 giorni, pubblicare un regolamento che disciplini il processo di riutilizzo dei propri dati; nel caso in cui non siano interessati a produrne uno specifico, esse adottano le linee guida nazionali come indicato dal comma 8.

La modifica al comma 2 del cit. art. 52 prevede che i soggetti pubblici debbano rendere disponibili i dati in loro possesso in formato aperto, salvo specifiche restrizioni che saranno previste dalle Linee guida nazionali, pubblicate dall'Agenzia per l'Italia digitale. Pertanto, salva indicazione esplicita, i dati pubblicati dalle amministrazioni saranno da considerarsi rilasciati come dati aperti.

La modifica al comma 3 dell'art. cit. precisa che la regola enunciata nel comma precedente si applica automaticamente ai dati già pubblicati

dalle amministrazioni che, salve diverse indicazioni da prodursi entro 90 giorni dall'approvazione del decreto, saranno da considerarsi disponibili in formato aperto e saranno perciò liberamente riutilizzabili. Con la modifica al comma 4 si prevede che le amministrazioni quando stipulano contratti di appalto per la realizzazione di prodotti e/o servizi che hanno ad oggetto o producono dati pubblici, debbano prevedere clausole specifiche per la gestione di tali dati, che, salvo precise restrizioni normative e/o di carattere industriale, sono accessibili come dati di tipo aperto.

Nel comma 5 del cit. art. viene precisato che le attività volte a garantire la pubblicazione, l'accessibilità e il riutilizzo dei dati delle amministrazioni rientrano tra i parametri di valutazione della performance dei dirigenti pubblici. Nei commi 6, 7 e 8 vengono assegnati all'Agenzia per l'Italia digitale, in collaborazione con le amministrazioni locali, la responsabilità della gestione e l'attuazione del processo di valorizzazione del patrimonio informativo pubblico nazionale. Ad essa sono affidati i seguenti compiti specifici: la redazione di un'agenda nazionale che definisca obiettivi e strumenti operativi del processo di valorizzazione dei dati nazionali; la realizzazione di un rapporto consuntivo delle attività annuali; l'elaborazione delle linee guida che definiscano le procedure che gli organismi pubblici devono seguire per pubblicare i propri dati.

Nel comma 9 del cit. art. è previsto che il Ministro delegato per l'innovazione tecnologica riferisca annualmente al Parlamento attraverso la presentazione del rapporto annuale sulla valorizzazione del patrimonio informativo pubblico nazionale redatto dall'Agenzia. La norma modifica inoltre (comma 1, lett. b) l'art. 68 del CAD, introducendo una definizione esplicita di dato aperto secondo tre dimensioni: tecnologica, economica e giuridica.

Un dato aperto deve essere riutilizzabile senza restrizioni, deve cioè essere distribuito dalle amministrazioni con una licenza – aperta e standard – che ne permetta l'utilizzo da parte di persone fisiche e giuridiche per qualunque scopo, anche a fini commerciali. Esso deve essere messo a disposizione in formati tabellari (machine readable) che permettano il suo utilizzo automatico da parte di programmi per elaboratori e, infine, il suo utilizzo non deve essere a titolo oneroso né per i cittadini né per le imprese, o al massimo, il prezzo in capo agli utilizzatore deve allinearsi al costo marginale di distribuzione.

Il comma 2 riporta un'ulteriore modifica al CAD e in particolare all'art. 1, introducendo la definizione del termine "riutilizzo", di cui al d.lgs. n. 36 del 2006. Il comma 3 inten-

de recepire la Raccomandazione della Commissione del 17 luglio 2012 (2012/417/UE) in materia di accesso aperto all'informazione scientifica che ha l'obiettivo di rinnovare e migliorare i processi di diffusione del sapere scientifico prodotto e diffuso in formato digitale.

Il comma 4 è norma transitoria. La norma fornisce poi (commi da 4 a 7) alcune definizioni non ancora previste nella normativa vigente in materia di "inclusione digitale". Fino a oggi si è sempre parlato genericamente di "accessibilità", con una definizione contenuta nell'art. 2 comma 1 l. n. 4/2004 (c.d. legge "Stanca") limitata ai prodotti hardware e software (compresi i siti Web) delle P.A. Nell'ambito dell'inclusione digitale è necessario che quest'ultima sia garantita a tutti, indipendentemente dal settore pubblico o privato, ed indipendentemente dal tipo di "periferica" (segnalatica informatizzata, TV interattiva, ecc.). Per la stessa motivazione si è ritenuto di definire il termine "smart inclusion", sottoinsieme dell'inclusione digitale, per particolare attenzione alla tematica delle "città intelligenti". Nell'ambito della definizione degli obiettivi delle P.A., la norma obbliga le amministrazioni pubbliche a rendere disponibile la consultazione degli obiettivi di accessibilità (già previsti dalla l. n. 4/2004 ma "orfani" nella medesima legge di un obbligo di pubblicazione dei medesimi) e di inclusione digitale. Esso prevede inoltre una responsabilità dirigenziale e disciplinare per i dirigenti che non applichino quanto previsto ai commi precedenti. Infine l'articolo contiene quanto previsto dall'art. 26 del disegno di legge congiunto "Disposizioni per lo sviluppo dei servizi elettronici e digitali (C. 4891 Gentiloni Silveri e C. 5093 Palmieri)" e inserisce chiarimenti sulla necessità del rispetto dei requisiti di accessibilità all'interno della c.d. "PA Digitale".

LE LINEE GUIDA SUGLI OPEN DATA PER LE P.A.

Prodotto nell'ambito del più ampio progetto delle Linee guida per i siti web della PA, il Vademecum è destinato ad amministratori, dirigenti e dipendenti pubblici, nonché a fornitori e consulenti delle pubbliche amministrazioni che vogliono approfondire il tema dei dati aperti e avviare un processo di apertura dei dati del settore pubblico.

Sin dalle premesse, il documento evidenzia l'importanza della pubblicazione dei dati prodotti e detenuti dalla Pubblica Amministrazione come strumento per tradurre il concetto di Open Government, basato sui principi

di trasparenza, partecipazione e collaborazione, in un vero e proprio modello sostenibile, e dell'importanza che tale patrimonio può avere in termini di efficienza del settore pubblico, trasparenza dell'azione amministrativa e quale elemento per favorire la crescita economica e la competitività. Si individuano così alcuni principi basilari, derivati da standard internazionali, che caratterizzano gli Open Data:

- **disponibilità e accesso** - i dati devono essere disponibili nel loro complesso, per un prezzo non superiore a un ragionevole costo di riproduzione, preferibilmente mediante scaricamento da Internet; i dati devono essere inoltre disponibili in un formato utile e modificabile
- **riutilizzo e redistribuzione** - i dati devono essere forniti a condizioni tali da permetterne il riutilizzo e la redistribuzione; ciò comprende la possibilità di combinarli con altre basi di dati
- **partecipazione universale** - tutti devono essere in grado di usare, riutilizzare e redistribuire i dati; non devono essere poste discriminazioni di ambiti di iniziativa in riferimento a soggetti o gruppi; per esempio, il divieto di utilizzare i dati per scopi commerciali o le restrizioni che permettono l'uso solo per determinati fini (quale quello educativo) non sono contemplabili.

CARATTERISTICHE DEGLI OPEN DATA

Dai suddetti principi, derivano una serie di importanti caratteristiche che gli Open Data dovrebbero soddisfare; considerando un data set, ovvero un insieme congruente di dati pubblicati, come ad esempio l'elenco delle scuole di Bologna, i dati dovrebbero essere:

- **Documentati** - i dati devono comprendere tutte le componenti (metadati) che consentano di comprenderne il significato, esportarli, utilizzarli on line e off line, integrarli e aggregarli con altre risorse e diffonderli in rete; con riferimento all'esempio dell'elenco delle scuole di Bologna, occorre specificare il significato di ogni campo contenuto nella tabella (Nome della scuola, numero di studenti, indirizzo, etc.)
- **Completi** - deve essere fornito l'intero dataset di informazioni, in relazione all'universo che si intende descrivere; con riferimento all'esempio dell'elenco delle scuole di Bologna, il dataset deve contenere l'elenco di tutte le scuole presenti in tutti i quartieri
- **Granulari e Dettagliati** - le risorse digitali devono es-

sere strutturate in modo tale che i dati siano presentati in maniera sufficientemente granulare, così che possano essere utilizzate dagli utenti per integrarle e aggregarle con altri dati e contenuti in formato digitale; con riferimento all'esempio dell'elenco delle scuole di Bologna, diverso è fornire tutto l'elenco delle scuole o solo il numero di scuole presenti per quartiere; inoltre per ciascuna scuola è possibile fornire più o meno informazioni interessanti (numero di studenti iscritti, rapporto insegnanti studenti, presenza o meno di laboratori, metri quadri, etc.)

- **Tempestivi** - i dati devono essere resi disponibili il più rapidamente possibile, per consentire agli Utenti di accedere e utilizzare le informazioni presenti in rete in modo rapido e immediato, massimizzando il valore e l'utilità derivanti da accesso e uso di queste risorse: una informazione "vecchia" è meno utile e interessante di una più recente
- **Accessibili** - i dati devono essere resi disponibili al maggior numero possibile di utenti senza barriere all'utilizzo, quindi preferibilmente attraverso il solo protocollo Hypertext Transfer Protocol (HTTP) e senza il ricorso a piattaforme proprietarie. Devono essere inoltre resi disponibili senza alcuna sottoscrizione di contratto, pagamento, registrazione o richiesta; con riferimento all'esempio dell'elenco delle scuole di Bologna, diverso è pubblicare l'elenco su una pagina web, rispetto a dovere compilare un modulo e ricevere l'elenco via mail o addirittura stampato su carta e spedito per posta.
- **Leggibili da computer** - per garantire agli utenti la piena libertà di accesso e soprattutto di utilizzo e integrazione dei contenuti digitali, è necessario che i dati siano "machine-readable", ovvero processabili in automatico dal computer; con riferimento all'esempio dell'elenco delle scuole di Bologna, diverso è pubblicare l'elenco in un file scaricabile sul computer o in una tabella html, rispetto a una immagine .png o un documento html non correttamente formattato.
- **In formati non proprietari** - I dati 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; con riferimento all'esempio dell'elenco delle scuole, diverso è rendere disponibile i dati in formati aperti e facilmente leggibili da programmi, come XML o CSV, rispetto a pubblicarli in .pdf o .xls

- **Liberi da licenze** che ne limitino l'uso - I dati aperti devono essere caratterizzati da licenze che non ne limitino l'uso, la diffusione o la redistribuzione.
- **Riutilizzabili** - Affinché i dati siano effettivamente aperti, gli utenti devono essere messi in condizione di riutilizzarli e integrarli, fino a creare nuove risorse, applicazioni e servizi di pubblica utilità.
- **Ricercabili** - I dati devono essere facilmente identificabili in rete, grazie a cataloghi e archivi facilmente indicizzabili dai motori di ricerca.
- **Permanenti** - Le peculiarità fino ad ora descritte devono caratterizzare i dati nel corso del loro intero ciclo di vita.

DAL ROW DATA AI LINKED DATA

Per mettere a disposizione del pubblico i dati di un'Amministrazione può essere utilizzata una grande varietà di formati. Per distinguere i diversi formati utilizzabili nella codifica dei set di dati, è stato proposto in seno al W3C un modello di catalogazione che li classifica in base alle loro caratteristiche su una scala di valori da 1 a 5 stelle. La Figura 4 descrive questa classificazione, che parte dalla pubblicazione di un dataset grezzo (1 stella), fino ad un dataset in grado di esprimere al meglio il proprio contenuto informativo.

Su questo tema occorre sottolineare per prima cosa che vale il principio che "è meglio pubblicare il dato anche in formato grezzo (row data) piuttosto che non pubblicarlo". Tuttavia, la modalità con cui i

★ UNA STELLA

È il livello base, costituito da dati grezzi (ROW Data) e non strutturati/classificati: ad esempio un'immagine in formato (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.

★★★ 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. Tre stelle indicano, oltre alle possibilità offerte dai dati contraddistinti da due sole stelle, la possibilità di effettuare elaborazioni sui dati senza esser costretti ad utilizzare software proprietario.

★★★★ QUATTRO STELLE

Questo livello indica dati strutturati e codificati in un formato non proprietario che sono dotati di un URI che li rende indirizzabili sulla rete internet e quindi utilizzabili direttamente on-line, 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.

★★★★★ 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 (classificazione a quattro stelle) 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.

Figura 4 - La classificazione proposta da Tim Berners Lee, Presidente del W3C e ideatore del World Wide Web e grande sostenitore degli Open Data

dati aperti vengono resi disponibili può influenzarne fortemente la loro usabilità.

Se i dati vengono pubblicati in modo che un computer possa:

- accedervi e leggerli facilmente
- riconoscerne la semantica
- correlazarli con altri dati mediante ad esempio ontologie standard

si può parlare di Open Linked Data (OLD). I dati in formato OLD possono essere incrociati dinamicamente, anche se provenienti da fonti diverse, eventualmente gestite da diverse amministrazioni.

Come esempio, si può pensare ad un file con dati di tipo demografico relativo ad esempio al Comune di Bologna; se il file è fornito in modo da essere relazionabile con un file analogo fornito da altri Comuni, è possibile realizzare applicativi software in grado di interrogare e correlare tra loro le varie fonti.

I Linked Open Data, consentono di combinare i contenuti di dataset diversi grazie ad esempio a costrutti formali formulati secondo il modello RDF in uno dei diversi formati esistenti (XML/RDF, N3, ecc...) e all'utilizzo di ontologie standard. Ciò aumenta esponenzialmente il valo-

re dei dataset reciprocamente correlati, consentendo il passaggio dal livello dei dati a quello dell'informazione e quindi a quello della conoscenza e fornendo così un quadro di contesto strutturato a partire dalla correlazione di informazioni provenienti da fonti diverse.

Produrre ed utilizzare Open Linked Data di diverse fonti è una interessante sfida, di cui anche l'Autore di questo articolo si sta occupando con il supporto dell'Università di Bologna.

COME AVVIARE UN PROGETTO PER L'APERTURA DEI DATI IN UNA PUBBLICA AMMINISTRAZIONE

Sotto il profilo tecnico-organizzativo, è auspicabile che l'organo di vertice dell'Amministrazione adotti un atto in cui delibere di pubblicare on-line i propri dati in formato aperto, con le dovute eccezioni e mediante l'uso di licenze aperte che consentano la massima riutilizzo. L'adozione di un regolamento che fornisca un quadro chiaro del procedimento da seguire all'interno dell'Ente, uniformando gli accorgimenti utilizzati (ad esempio in materia di formazione dei dataset, di metadati e di licenze) è la procedura preferibile, tuttavia, la scelta di pubblicare i dati può essere compiuta anche in altri atti normativi e amministrativi.

A valle della decisione di aprire un insieme di dati in possesso dell'Amministrazione per metterli a disposizione del cittadino e della comunità è possibile identificare un percorso strutturato nelle seguenti fasi: censimento dei dati disponibili, identificazione dei dati di interesse per la Comunità, analisi di dettaglio dei dataset selezionati, pubblicazione dei dataset, diffusione dei dataset pubblicati.

IL CENSIMENTO DEI DATI DISPONIBILI

L'Amministrazione Pubblica, come qualsiasi organizzazione complessa, gestisce quotidianamente una grande quantità di dati che servono per garantire l'operatività e l'erogazione dei servizi al cittadino. Con lo sviluppo dell'Information & Communication Technology e la diffusione dei sistemi di e-Government tali dati sono ormai disponibili in formato digitale e possono quindi essere condivisi e diffusi. Per farlo è però necessario identi-

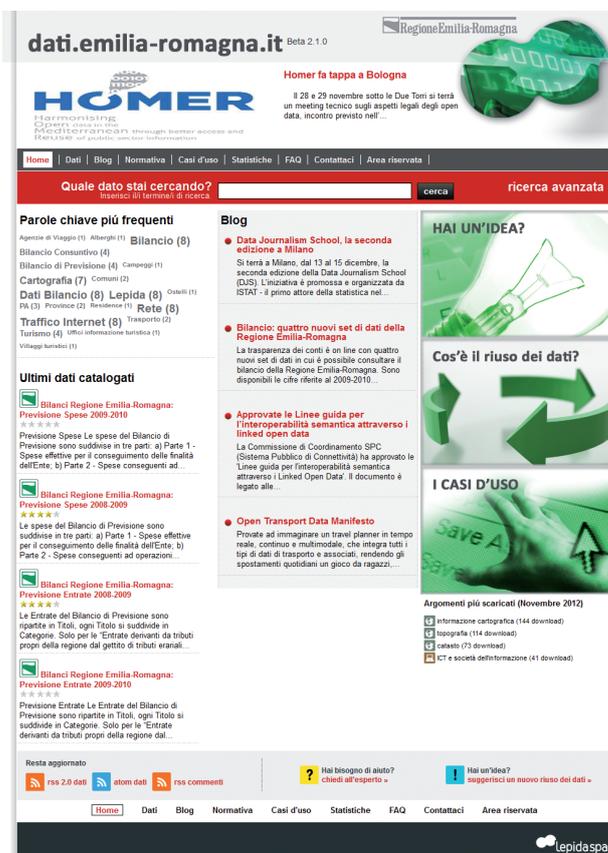


Figura 5 - Il portale, in fase di rilascio, per gli Open Data della Regione Emilia Romagna

ficarli nell'ambito della struttura burocratica dell'Ente. Le organizzazioni complesse non sempre sono a completa conoscenza dell'intero corpus di dati che producono o semplicemente che gestiscono. In molti casi, infatti, il dato è il risultato di un processo, o il sottoprodotto di un'elaborazione funzionale ad altri scopi. Decidere quali sono i dati di interesse per la Comunità detenuti dalla propria Amministrazione richiede quindi – come primo passaggio – una vera e propria mappatura dei dati complessivamente disponibili. In primo luogo, occorre identificare quali sono le strutture funzionali dell'Amministrazione che detengono dati potenzialmente utili. Con riferimento alla Pubblica Amministrazione Centrale, ad esempio, essi saranno distribuiti nell'ambito delle diverse Direzioni. Con riferimento alla Pubblica Amministrazione Locale, invece, essi saranno gestiti dai diversi Uffici; in secondo luogo, identificate le strutture funzionali, è necessario comprendere – nel loro ambito – quali sono le fonti dati di rilievo. In altri termini, è necessario identificare archivi strutturati, elenchi, basi dati prodotte da software in uso presso l'amministrazione.

Questa fase avrà come output una mappatura dei dati gestiti dall'amministrazione; questa attività, oltre ad essere utile alla definizione di una strategia di apertura dei dati, lo è anche per determinare e comprendere meglio il patrimonio informativo detenuto dalla propria Amministrazione.

L'IDENTIFICAZIONE DEI DATI DI INTERESSE PER LA COMUNITÀ

Una volta completato il censimento, si tratta di comprendere quali dati possano essere pubblicati e siano di interesse della comunità. Nel momento in cui si avvia una strategia di apertura dei dati, non è necessario identificare immediatamente tutti gli usi che verranno fatti di un dato.

Spesso è la comunità degli utenti a identificare nei dati un valore che l'Amministrazione stessa non vi ha visto o non vi ravvede. In linea generale, quindi, quando un dato è disponibile senza che la sua estrazione rappresenti un costo significativo, è opportuno provvedere alla sua apertura anche nel momento in cui non se ne ravveda un'utilità immediata.

Spesso, infatti, indicatori di interesse per la Comunità sono desumibili da elaborazioni su dati di base che non erano state previste da chi tale dato ha provveduto ad aprirlo. Nella definizione di una strategia di Open Data è comunque bene identificare priorità relative all'apertura di dati che siano funzionali ai reali interessi della Comunità di riferimento.

L'ANALISI DI DETTAGLIO DEI DATASET SELEZIONATI

Una volta realizzata la mappatura dei dati presenti nell'Amministrazione e definite delle priorità in base agli interessi della Comunità di riferimento, è necessario passare ad una fase di analisi dei dati identificati, per valutare la fattibilità della loro apertura e le eventuali modalità per attuarla.

La fase di analisi dei dati deve rispondere principalmente a tre domande:

- Chi detiene i dati quali diritti ha sugli stessi?
- Quali sono le qualità dei dati?
- In che formato sono disponibili i dati?

Dal punto di vista giuridico, la prima operazione da compiere è relativa alla verifica della titolarità del dato che si intende pubblicare. In proposito, l'art. 58, comma 1, Codice dell'Amministrazione Digitale detta un'importantissima disposizione in materia di fruibilità, affermando che "il trasferimento di un dato da un sistema informativo a un altro non modifica la titolarità del dato". Questo significa che in capo all'Amministrazione titolare del dato rimanga la responsabilità sulla gestione del dato stesso e sulla sua esattezza e veridicità. L'Amministrazione titolare del dato è quella che lo ha creato o comunque lo gestisce per fini istituzionali, mentre altre eventuali Amministrazioni che utilizzino tale dato non diventano titolari del dato medesimo. Ad esempio, la titolarità del dato anagrafico è del Comune, così come la titolarità del codice fiscale è dell'Amministrazione finanziaria.

Ulteriore disposizione che viene in rilievo è quella dettata dall'art. 11 Legge n. 633/1941 che prevede che "alle amministrazioni dello stato, alle provincie ed ai comuni spetta il diritto di autore sulle opere create e pubblicate sotto il loro nome ed a loro conto e spese". Di conseguenza, è possibile affermare che l'Amministrazione può ritenersi titolare del dato solo quando lo abbia creato direttamente oppure lo abbia commissionato ad un altro soggetto.

Verificato il fatto che i dati sotto il profilo giuridico siano distribuibili è utile valutarne le caratteristiche in termini qualitativi per definire l'opportunità della loro apertura verso l'utenza. Per farlo è opportuno prendere in considerazione una serie di fattori che attengono la dimensione qualitativa del singolo dato e l'incidenza di tali fattori rispetto all'insieme dei dati che compongono il dataset. Occorre perciò analizzarne l'accuratezza sintattica e semantica, la completezza, la consistenza, etc. Occorre inoltre identificarne il processo di aggiornamento al fine di analizzare il livello di profondità storica e i modi in cui rendere disponibili gli aggiornamenti.

Verificato il fatto che i dati che si intende aprire siano effettivamente distribuibili sotto il profilo giuridico e di qualità adeguata, è necessario effettuare una analisi del formato nel quale i dati sono disponibili; se il dato non è già contenuto e gestito da un sistema informativo adeguato, è opportuno valutare l'opportunità di effettuare azioni volte a migliorare il processo che li gestisce, al fine di garantire una migliore accessibilità delle informazioni.

LA PUBBLICAZIONE DEI DATI

La fase di pubblicazione del dato è essenzialmente tecnica. Riferendosi alla classificazione per stelle del W3C, tutte le volte in cui è possibile si dovrebbe considerare i formati che vengono classificati con tre o più stelle. I dataset classificati con quattro o cinque stelle richiedono la costruzione di dataset basati su Linked Open Data (LOD) e quindi competenze tecniche specifiche. Di norma, comunque, i dataset in formato LOD sono prodotti da database relazionali o software specifico. In fase di pubblicazione dei dati raccolti in un dataset, è opportuno descrivere il dataset corredandolo di informazioni utili a comprenderne il contenuto. Tali informazioni prendono il nome di metadati descrittivi. Con il termine metadato si intende l'informazione che descrive un insieme di dati. Nel caso specifico dei metadati descrittivi essi costituiscono un livello di informazioni descrittive che corredano il dataset, ne esplicitano alcune caratteristiche e ne rendono più semplice l'identificazione. Un quadro completo dei metadati che descrivono un dataset è quello utilizzato dal Portale italiano dei dati aperti (www.dati.gov.it)

Nel caso di dati non strutturati, di dati strutturati in formato proprietario e di dati strutturati in formato aperto la soluzione più semplice consiste nel caricare direttamente il dataset o i dataset sul Web Server della propria Amministrazione. L'operazione può essere compiuta inserendo direttamente il file contenente il dataset nella cartella di destinazione del server.

Nel caso di Linked Open Data la

soluzione per la pubblicazione può presentare un livello di complessità maggiore. I dataset in formato LOD sono di norma prodotti attraverso database relazionali o per mezzo di applicazioni software che espongono delle API appositamente concepite per produrre tali dataset.

I dataset in formato Linked Open Data possono quindi essere dei file (ad esempio in formato RDF/XML) da pubblicare in una directory del proprio sito, analogamente a quanto descritto al punto precedente, ma possono anche essere costruiti dinamicamente ad ogni richiesta da un database relazionale o da un'applicazione che esponga una API specifica.

In questa situazione non esiste un file da pubblicare in una directory, ma il dataset viene ricostruito dinamicamente ad ogni richiesta effettuata dall'utente. In tal caso sarà necessario sviluppare un'architettura software specifica, da definire sulla base del tipo di dataset e, soprattutto, dei sistemi software a disposizione dell'Amministrazione. Invece di dovere di volta in volta sviluppare una specifica interfaccia per pubblicare Open Data provenienti da diverse basi dati, è anche possibile utilizzare dei tools

appositamente studiati per l'interoperabilità applicativa. Ad esempio SOAP Dispenser (www.e-soft.it/joomla/index.php?Itemid=83), è un tool in grado di esporre servizi SOAP e Open Data di Livello 5; Soap dispenser è utilizzato anche dalla Regione Emilia per esporre servizi di geocodifica di indirizzi.

LA DIFFUSIONE DEI DATI

Dopo aver pubblicato i dataset aperti è bene attivarsi perché altre Amministrazioni, cittadini e comunità di sviluppatori vengano a conoscenza dell'esistenza di dati aperti riferiti alla propria Amministrazione. Per permettere un'adeguata diffusione dei dati aperti a livello centrale è stato sviluppato il Portale nazionale dei dati aperti raggiungibile all'indirizzo <http://www.dati.gov.it/>, verso il quale far confluire i riferimenti dei dati aperti da ogni Amministrazione. Alcune Regioni dispongono o stanno sviluppando portali regionali dei dati. Tali portali, in alcuni casi, consentono alle Amministrazioni del territorio di pubblicare anche i dati di queste ultime; quando disponibile, quindi,

Le Linee Guida per l'Open Data curato dalla Associazione per l'Open Government http://www.dati.gov.it/sites/default/files/VademecumOpenData.pdf
Il sito dell'Open Knowledge Foundation http://okfn.org
Il manuale degli Open Data curato dall'Open Knowledge Foundation Italia http://opendatahandbook.org/it/
Gruppo su Innovatori PA http://www.innovatoripa.it/groups/dati-aperti-open-data
Il portale per gli Open Data del Comune di Bologna http://dati.comune.bologna.it/
Il portale per gli Open Data della Regione Emilia-Romagna http://dati.emilia-romagna.it/

Figura 6 - Link e riferimenti utili

L'Amministrazione può pubblicare i dati, oltre che nel suo sito, anche nel portale regionale secondo modalità e funzionalità previste da ciascun portale.

LA LICENZA ITALIANA PER I DATI APERTI

Quale licenza scegliere per la pubblicazione? Come innanzi accennato, in base alla Open Definition, un dato è aperto se tutti sono in grado di usarlo, riutilizzarlo e redistribuirlo liberamente, con l'unico vincolo di citare la fonte e redistribuirlo allo stesso modo; di conseguenza, le Amministrazioni devono scegliere ed utilizzare una delle licenze di tipo aperto, optando – preferibilmente – per una di quelle già sviluppate e disponibili.

Tra di esse si segnala la Italian Open Data License (IODL), la licenza "aperta" creata allo scopo precipuo di dare a tutte le Amministrazioni uno strumento chiaro e certificato, in grado di facilitare la diffusione e il riutilizzo dell'informazione del settore pubblico. La licenza ODL è stata sviluppata da Formez PA e ha lo scopo di promuovere la "liberazione" e valorizzazione dei dati pubblici secondo la linea già tracciata dal Ministero per la pubblica amministrazione e l'innovazione. La licenza, il cui testo completo è scaricabile da www.formez.it/iodl,

prevede che l'utente possa liberamente:

- consultare, estrarre, scaricare, copiare, pubblicare, distribuire e trasmettere le informazioni;
- creare un lavoro derivato, per esempio attraverso la combinazione con altre informazioni (cosiddetto mashup), includendole in un prodotto o sviluppando un'applicazione informatica che le utilizzi come base dati.

In cambio, all'utente è chiesto di indicare la fonte delle informazioni e il nome del soggetto che fornisce il dato, includendo, se possibile, un link alla licenza; inoltre è richiesto all'utente di pubblicare e condividere gli eventuali lavori derivati con la stessa licenza o con altra licenza aperta, ritenuta compatibile.

DISCUSSIONE ONLINE

Per alimentare la discussione online sul Portale Dati.gov.it, sulla Licenza IODL e più in generale sul tema open data è stato creato un gruppo su *Innovatori PA*, community promossa da Formez PA e Forum PA che raccoglie più di 4000 operatori del settore pubblico. Su Innovatori PA vengono inoltre segnalate iniziative, convegni, seminari online dedicati ai temi dell'open data e dell'open government.