



Libro bianco sulla gestione dei dati spaziali per le pubbliche amministrazioni

Authors:

Avv. Simone Aliprandi, Ph.D. – Array / Copyleft-Italia.it (www.aliprandi.org)

Avv. Carlo Piana – Array (www.piana.eu)

Release: November 2013 – original version of this document available at www.freegis.net.

License: Creative Commons Attribution – Share Alike 3.0 unported ([see the license](#))

How to cite: please cite this work as “Libro bianco sulla gestione dei dati spaziali per le pubbliche amministrazioni, by Simone Aliprandi and Carlo Piana for FreeGIS.net”

Credits: this is a derivative work of “Libro bianco per il riutilizzo dell’informazione del settore pubblico” by EVPSI Project (<http://www.evpsi.org/librobianco>)



Kanton Graubünden
Chantun Grischun
Cantone dei Grigioni

AUTONOME
PROVINZ
BOZEN
SÜDTIROL



PROVINCIA
AUTONOMA
DI BOLZANO
ALTO ADIGE



Sommario

Executive Summary.....	3
Introduzione.....	3
Target.....	3
Premessa: Open by default e open licensing.....	4
Come procedere per aprire i dati: aspetti giuridici.....	5
Step 0: Individuare uno specifico dataset da “aprire”.....	5
Step 1: Verificare se il dataset rientra nel campo d'azione della cosiddetta pubblicazione obbligatoria.....	5
Step 2: Individuare quali diritti esistano sui dati.....	5
L'open by default.....	6
Le licenze libere e il licensing dei dati pubblici.....	7
Usare con cautela “Condividi-allo-stesso-modo”.....	8
E Open Data Commons? E la Italian Open Data Licence 2.0?	9
Le licenze sono flessibili. (Ma occhio a non renderle contorte!)	10
...e se qualcosa non è protetto da alcun diritto?	10
Come procedere per aprire i dati: aspetti tecnologici.....	11
Introduzione.....	11
Mettere i dati a disposizione online.....	11
Far trovare i dati.....	11
Formati aperti, secondo la legge italiana	12
Occhio al PDF!	12
Metadati: i dati sui dati	13
No alla discriminazione tra umani e macchine!	13

Executive Summary

Per pubblicare dati geografici nel rispetto delle direttive europee e delle norme vigenti a livello locale e nazionale, vanno tenuti nella dovuta considerazione i seguenti aspetti:

- proprietà dei dati spaziali (*rectius*, titolarità dei diritti sull'insieme dei dati spaziali raccolti);
- organizzazione e gestione dei dati spaziali (raccolta e organizzazione in un dataset, formattazione secondo standard aperti e interoperabili);
- modalità di messa a disposizione del pubblico e di condivisione dei dati spaziali (tra cui la questione del licenziamento con approccio “open”);
- eventuale aggregazione, commistione o sovrapposizione con altri dataset già pubblicati, effettuando una pregressa verifica di compatibilità sia a livello tecnologico che a livello giuridico.

Introduzione

Linee guida sintetiche, semplificate e rivolte a tutti gli interlocutori interessati per operare correttamente nel rilascio di dati spaziali da parte della pubblica amministrazione.

Target

I responsabili per la gestione di dati geografici, che vogliono pubblicare i dati e hanno bisogno di linee guida sul come e perché.

Premessa: Open by default e open licensing

Innanzitutto, a titolo di fondamentale premessa, bisogna tenere presente il principio comunemente detto “open by default” secondo cui dati e documenti pubblicati da una pubblica amministrazione sono considerati come dati di tipo aperto, a meno che la PA applichi al dataset una licenza più restrittiva. Tale principio è fissato dall'art. 52 del CAD¹, che a sua volta fa riferimento alla definizione di “dati di tipo aperto” contenuta nell'art. 68 CAD.

Ne deriva che, nel momento in cui la PA titolare di un dataset lo pubblica in qualsivoglia forma e senza specificare particolari condizioni d'uso, esso ricade in uno status di “openness” per effetto dell'art. 52 CAD.

Scopo di questo meccanismo è sfruttare l'eventuale inerzia da parte della PA in merito all'applicazione di una licenza, sia a vantaggio dei cittadini, sia in favore di una generale libertà di riutilizzo e redistribuzione dei dati pubblici.

Il principio “open by default” si applica a quei dati che sono già pubblicati dalla Pubblica Amministrazione, ma la relativa normativa non impone nessun obbligo di pubblicazione. A tal proposito interviene il cosiddetto “Decreto trasparenza” del marzo 2013², il quale ha invece indicato chiaramente quali sono le categorie di dati che le PA italiane hanno *l'obbligo* formale di pubblicare, rafforzando di fatto l'idea della messa a disposizione dei dati pubblici in modo che siano accessibili a tutti e il più possibile riutilizzabili. Tale decreto fornisce anche qualche indicazione più specifica sul tipo di licenza utilizzabile nel caso di “pubblicazione obbligatoria”.

Dunque, ci si può legittimamente chiedere: se è arrivato l'open by default, perché si insiste ancora sul rilascio dei dati da parte delle PA? Perché, come detto, le PA hanno ancora un certo margine di discrezionalità nella scelta di una licenza, dunque è importante che, qualora decidano di non pubblicare in semplice modalità “open by default”, operino questa scelta nel modo più consapevole possibile.

E in generale, l'indicazione esplicita di una licenza, quand'anche molto simile allo status già attribuito “by default” dalla legge, risulta sempre cosa opportuna ai fini di maggior chiarezza sui termini d'uso del dataset; specie in un contesto di interconnessione telematica necessariamente internazionale e privo di frontiere; ciò a maggior ragione in uno scenario in cui le informazioni sui dataset siano incluse in metadati interrogati in modo automatico dagli aggregatori, senza la necessità di un intervento umano. Infatti, non tutti gli utenti delle varie nazionalità sono debitamente informati dei meccanismi della legge italiana (ancor più se basati su riforme molte recenti come quelle in questione), mentre più facilmente possono riconoscere e comprendere il senso delle più diffuse licenze di libera distribuzione attualmente disponibili.

1 Codice dell'Amministrazione Digitale, D.lgs. 7 marzo 2005 n. 82.

2 D.lgs. 14 marzo 2013 n. 33

Questo documento mira a realizzare una guida operativa di primo livello; per l'approfondimento teorico dei vari aspetti giuridici si invita a leggere gli altri documenti prodotti nell'ambito del Progetto FreeGIS.net: Wp 5-1 Analisi delle policies di condivisione dei dati spaziali, Wp 5-2 FreeGIS data license 1, Wp 5-4 FreeGIS data license 2.

COME PROCEDERE PER APRIRE I DATI: ASPETTI GIURIDICI

Step 0: Individuare uno specifico dataset da “aprire”

Analizzare i punti successivi in astratto è difficile o impossibile, per cui è utile discutere sempre di dati sufficientemente ben determinati, di cui sia possibile mostrare un campione ai vari possibili interlocutori (giuristi, dirigenti, amministratori, potenziali riutilizzatori).

È importante (poter) discutere di dati reali, visualizzarli, parlare con le persone responsabili della loro raccolta, ecc. È quindi indispensabile individuare alcuni dati. Non cadete tuttavia nell'errore di voler mappare tutti i dati disponibili prima di cominciare ad aprirne qualcuno: salvo che siate così fortunati da avere già un catalogo ben organizzato, l'opera rischierebbe di essere faraonica e vi impedirebbe di conseguire risultati concreti nel breve periodo, facendo così perdere entusiasmo a tutti.

Step 1: Verificare se il dataset rientra nel campo d'azione della cosiddetta pubblicazione obbligatoria

Come spiegato nel documento Wp 5-1, con il decreto legislativo n. 33 del 14 marzo 2013 (anche noto come “Decreto Trasparenza”) il legislatore pone a carico delle PA un espresso ed inequivocabile obbligo di messa a disposizione del pubblico “delle informazioni e dei dati concernenti l'organizzazione e l'attività delle pubbliche amministrazioni”, stabilendo all'art. 3 che:

Tutti i documenti, le informazioni e i dati oggetto di pubblicazione obbligatoria ai sensi della normativa vigente sono pubblici e chiunque ha diritto di conoscerli, di fruirne gratuitamente, e di utilizzarli e riutilizzarli ai sensi dell'articolo 7.

Dunque, prima di interrogarsi se sia il caso o meno di rilasciare liberamente un dataset, è opportuno verificare se quel tipo di informazioni rientrano nel campo d'azione del “Decreto trasparenza” e se vige a carico della PA un obbligo esplicito alla loro messa a disposizione del pubblico.

Nel caso di risposta positiva, la PA non può fare altro che attivarsi per la pubblicazione dei dati e deve farlo seguendo i principi dello stesso decreto nonché quelli già presenti nel CAD (anch'essi presentati in dettaglio nel documento Wp 5-1). In caso diverso, la scelta è discrezionale.

Step 2: Individuare quali diritti esistano sui dati

Può avvenire tuttavia che non sempre sia chiaro ad una PA quale sia la provenienza di un dataset e

se essa abbia piena titolarità su di esso.

Nella maggior parte dei casi, anche se nessuno ha mai esplicitamente reclamato alcun diritto su una banca di dati, è regola di prudenza formulare l'ipotesi di lavoro secondo cui esistano alcuni diritti di privativa su beni immateriali (*in primis* il diritto d'autore e il diritto "sui generis") e che sia necessario chiedere il permesso del titolare dei diritti per compiere la maggior parte delle operazioni di riutilizzo che possano venire in mente (estrazione di dati, riproduzione, aggiornamento, creazione di nuove banche dati che ne combinino di vecchie, etc.). Di regola, il fatto che tali atti siano compiuti per fini non commerciali non legittima l'utilizzazione (il che vale in generale per tutti i materiali protetti diritto d'autore e dunque anche nel nostro caso). In altre parole, il modello in cui ricadono automaticamente le banche di dati è quello "Tutti i diritti riservati".

È dunque necessario porsi due questioni:

- 1) può l'ente pubblico disporre legalmente dei dati in questione, usandoli liberamente ed eventualmente anche ri-licenziandoli come ritiene opportuno?
- 2) Se sì, com'è opportuno che l'ente si comporti nel gestire i propri diritti?

La prima domanda, che riguarda la cosiddetta *rights clearance*, rientra tra le questioni tecnico-giuridiche che non è possibile trattare in dettaglio in questo lavoro e che necessita quasi sempre un'analisi ad hoc (è anche per questa ragione che è importante avere un giurista nella squadra operativa dietro ogni progetto open data).

Per analizzare lo status giuridico di un database, sarà necessario rispondere a quesiti del tipo: Da dove vengono i dati? Chi li ha raccolti? Con quali risorse? Ci sono (stati) contratti di qualche tipo che regola(va)no l'attività di raccolta e/o l'uso dei dati stessi. Inoltre, sarà opportuno chiedersi: Parte di questi dati sono dati personali? O addirittura dati sensibili? Ci possono essere al loro interno segreti aziendali o parti sulle quali terzi possano vantare diritti esclusivi?

Le risposte a tali domande saranno utili ad un giurista per stabilire se e come i dati possano essere legittimamente rilasciati e licenziati dall'ente titolare.

L'open by default

Una volta appurata la questione della titolarità dei dati, si passa alla fase della scelta delle modalità di rilascio del dataset e quindi anche della licenza.

Alla regola generale descritta poco fa, nel dicembre 2012 è stata posta una sostanziale eccezione con l'introduzione del principio "open by default" già illustrato nei paragrafi precedenti e meglio dettagliato nel documento di inquadramento giuridico (Wp 5-1), cui si rimanda.

Ne consegue che, giunti a questo punto, la pubblica amministrazione che volesse rilasciare liberamente i suoi dataset può anche semplicemente pubblicarli senza particolari indicazioni relative ai termini d'uso e al copyright, lasciando agire così l'open by default. Tuttavia la buona prassi

suggerisce che comunque una pubblica amministrazione applichi una licenza al dataset se vuole aumentare in modo significativo l'uso delle risorse informative che possiede, è sempre bene associare ai dati un'esplicita licenza. **L'incertezza è nemica del riutilizzo.**

Notate che anche la presenza di termini per il riutilizzo poco chiari o contraddittori potrebbe avere effetti scoraggianti. Se per capire cosa si possa davvero fare, serve un avvocato, probabilmente nessuno riutilizzerà i vostri dati. È dunque utile adottare strumenti per ridurre i costi di transazione, ovvero i costi monetari, ma anche le perdite di tempo e l'incertezza, sia per i detentori dei dati, sia per i riutilizzatori.

Le licenze libere e il licensing dei dati pubblici

Una licenza è il documento che descrive quali, tra le attività normalmente riservate al detentore dei diritti (il licenziante), possano essere compiute da chi riceve i dati sotto i termini della licenza stessa (il licenziatario). La licenza descrive anche le condizioni per godere dei diritti concessi, se ve ne sono. Come abbiamo già discusso, definendo a che condizioni i dati siano davvero “aperti”, la scelta ottimale è tra “nessuna condizione” e “a condizione che si riconosca l'attribuzione, ovvero si menzioni l'origine dei dati”. Combinando l'opportunità di utilizzare gli strumenti più standard e diffusi con le condizioni di cui sopra, è lecito raccomandare di restringere la scelta delle licenze al seguente elenco: la licenza/liberatoria Creative Commons Zero (CC0)³, per chi voglia scegliere l'approccio “nessuna condizione”; la licenza Creative Commons Attribuzione (CC BY) per chi voglia richiedere l'attribuzione.

Sono ormai numerosi i paesi che hanno adottato – come unica soluzione per le loro informazioni pubbliche o, quantomeno, come soluzione standard – una licenza aperta.

Negli USA, le informazioni e i contenuti generati dal settore pubblico (a livello federale) ricadono automaticamente nel pubblico dominio e sono dunque liberamente riutilizzabili senza alcun vincolo e per legge. Questa è chiaramente la soluzione aperta per eccellenza, perché non richiede scelte ulteriori e si applica automaticamente a tutta la PSI.

Nel Regno Unito, è stata sviluppata una licenza *ad hoc* per l'informazione del settore pubblico, la Open Government Licence. Si tratta di una licenza che richiede la sola citazione dell'origine dei dati e che è esplicitamente compatibile con le licenze CC BY ed Open Data Commons Attribution (ODC BY). Con una scelta simile, in Nuova Zelanda, la licenza Creative Commons Attribuzione (CC BY) è stata scelta come soluzione di default. In pratica – in assenza di buone ragioni per fare altrimenti – la scelta ricade automaticamente su quella licenza.

In Italia, la Regione Piemonte ha scelto CC0 come propria licenza di default. Ad oggi, l'unica alternativa praticata è stata CC BY, prescelta per alcuni dati cartografici.

Noi, per il progetto FreeGIS.net, abbiamo raccomandato di evitare la proliferazione di licenze

3 Le licenze Creative Commons sono elencate in <http://creativecommons.it/Licenze>

consigliando di non procedere alla stesura di una nuova licenza (come si era ipotizzato in prima battuta); e abbiamo optato per l'indicazione di CC0 come lo strumento più adeguato al rilascio pubblico di dati spaziali da parte delle PA.

Il requisito di attribuzione non richiede particolari commenti e consiste nell'obbligo di citazione dell'origine/fonte dei dati (ovviamente, con modalità che non traggano in inganno l'utilizzatore – ad esempio, un dataset modificato dovrà sempre essere identificato come tale e mai spacciato per l'originale).

Al contrario, il requisito di condivisione allo stesso modo è un tema complesso e delicato. Questa clausola (detta anche share-alike, copyleft o “propagativa”) consiste nell'obbligare il riutilizzatore a licenziare qualsiasi lavoro realizzato a partire dall'originale con la medesima licenza adottata per l'originale (sicché chiunque altro resta a sua volta libero di apportare ulteriori modifiche e la catena di condivisione si prolunga, da cui il nome “clausola propagativa”).

Usare con cautela “Condividi-allo-stesso-modo”

Spesso, i dati generati dalle community online sono rilasciati sotto licenze del tipo condividi-allo-stesso-modo. Questa è una scelta di auto-difesa della comunità contro soggetti esterni che tentino di appropriarsi del lavoro della comunità stessa con la creazione di una versione modificata (e migliorata) del lavoro della community senza però restituire alla comunità alcun apporto migliorativo. Per una comunità di utenti, che si unisce intorno all'idea di un progetto comune che rappresenta anche l'obiettivo attorno al quale si aggrega la comunità stessa, questa scelta può essere appropriata. È invece dubbio che lo stesso valga per un'amministrazione pubblica, che si ritrova a possedere dati generati come “effetto collaterale” di un suo compito istituzionale. In particolare, mentre lo scopo principale di una community online è far progredire il progetto attorno al quale si è costituita, i dati pubblici sono messi a disposizione per rendere possibili riusi inattesi, spesso in combinazione con altri insiemi di dati. Per questo motivo, l'inserimento della clausola di condivisione-nello-stesso-modo è un'idea criticabile. Ogni licenza share-alike, infatti, è compatibile con se stessa e tendenzialmente incompatibile con altre licenze share-alike (ad esempio, CC BY-SA e ODC ODbL sono mutualmente incompatibili). Il risultato rischia di essere una sorta di “Babele delle licenze”, in cui numerosi insiemi di dati non possono essere uniti e mescolati. Inoltre non va sottovalutato l'effetto di proliferazione delle licenze e la barriera al riutilizzo transfrontaliero dei dati che l'utilizzo di licenze “nazionali” comporta. Per finire, il rilascio di dati pubblici dovrebbe essere neutro rispetto ai modelli industriali adottati dagli utilizzatori, mentre l'utilizzo di licenze “share alike” almeno in teoria limiterebbe lo sfruttamento di tipo proprietario delle estensioni di un determinato dataset, il che è contraddittorio rispetto al principio della sfruttabilità commerciale dei dati contenuto nei recenti provvedimenti.

Se comunque si decidesse di adottare una licenza di tipo “condividi-allo-stesso-modo”, CC BY-SA avrebbe il chiaro vantaggio di essere lo standard dominante (ad esempio, perché utilizzata da

Wikipedia e progetti collegati, come Wikidata). Tra le altre principali licenze contenenti una clausola condividi-allo-stesso-modo, la ODC ODbL presenterebbe il vantaggio di essere più robusta in termini di possibilità di far valere con successo la clausola condividi-allo-stesso-modo in alcune situazioni (poiché è stata scritta appositamente per le banche di dati, mentre le licenze CC sono più generiche). La Italian Open Data License (IODL) versione 1.0 presenterebbe invece il vantaggio di contenere un'esplicita clausola di compatibilità, che permette di mescolare dati ottenuti sotto la IODL con dati licenziati con una licenza compatibile (CC BY-SA o ODC OdbL), ri-licenziandoli con la seconda licenza. Forse, se proprio si vuole usare una licenza con effetto “propagativo” (appunto la clausola condividi-allo-stesso-modo), la sua clausola di compatibilità rende la IODL 1.0 la scelta preferibile (almeno in Italia).

E Open Data Commons? E la Italian Open Data Licence 2.0?

Esistono due alternative a CC0 e CC BY, realizzate da Open Data Commons (<http://opendatacommons.org/>) che risultano tecnicamente del tutto adeguate (e storicamente anche più “vicine” al movimento open data). La ragione per cui raccomandiamo gli strumenti realizzati da Creative Commons è semplicemente che questi sono più diffusi, meglio documentati sul web, e forse leggermente più comprensibili, oltre che – e questo non è un aspetto da trascurare – tradotti in italiano.

Per quanto riguarda la versione 2.0 della Italian Open Data Licence sviluppata da Formez PA (<http://www.dati.gov.it/i odl/2.0/>), sebbene presenti la sola clausola di attribuzione, noi non la raccomandiamo esplicitamente, per ragioni di opportunità legate ad aspetti di standardizzazione tra licenze, anche in relazione al contesto internazionale. Così come non raccomandiamo le licenze di tipo condividi-allo-stesso-modo (o share-alike) sviluppate da Open Data Commons (Open Database License, ODC ODbL) o da Creative Commons (Attribuzione Condividi- allo-stesso-modo, CC BY-SA), per le ragioni, legate all'interoperabilità tra licenze, descritte nei precedenti paragrafi.

Proprio rispetto alla necessità di compatibilità tra licenze, vanno citate iniziative di communities che spingono verso l'adozione di una singola licenza per i dati pubblici aperti in Europa (si veda, ad esempio, la petizione promossa dalla comunità Open Data spagnola – <http://chn.ge/ILUJ42>). Questo approccio pare compatibile con i principi espressi nell'Agenda Digitale europea, specie in relazione alla creazione di un singolo mercato digitale. Nel prossimo futuro, è dunque possibile che i decisori europei intraprendano il percorso verso un unico modello standard di licenza.

Per applicare una licenza ad una base di dati distribuita online, la prassi è che basti indicare la licenza stessa nella pagina dalla quale si effettua il download. Inoltre, esistono opportuni accorgimenti tecnici, per far sì che l'associazione tra file e licenza sia chiara anche alle macchine (p. es. ai motori di ricerca). E' importante che tale collegamento tra dataset e licenza sia effettuato in modo chiaro, diretto e non equivoco, indicando il nome della licenza integralmente e correttamente e aggiungendo il link ufficiale al suo testo.

Le licenze sono flessibili. (Ma occhio a non renderle contorte!)

Il detentore dei diritti può offrire diverse licenze allo stesso tempo, ad esempio garantendo una licenza standard a chiunque sia interessato a rispettarne i termini, ma affiancandovi la possibilità di accordi speciali in casi particolari. Ovviamente, a seconda dei termini di ciascuna delle licenze scelte, questa possibilità può o meno avere senso. Per esempio, un ente cartografico potrebbe offrire gratuitamente una licenza di tipo condividi-allo-stesso-modo per le proprie mappe, ma potrebbe anche offrire, dietro compenso, un'altra licenza che non includa tali limitazioni.

Questa seconda licenza, per esempio, potrebbe essere scelta da operatori commerciali, che vogliono mescolare propri dati geografici con quelli dell'ente cartografico, utilizzando poi una licenza proprietaria per il risultato finale. (Notate che questa non è necessariamente una soluzione raccomandabile: è solo un esempio di combinazione possibile di licenze).

Un'altra possibilità – forse più raccomandabile della precedente – passa per l'uso di “licenze a pagamento”, che oltre a includere il puro permesso di utilizzo, offrano servizi aggiuntivi (un call center, la possibilità di download in larga banda, una garanzia o certificazione aggiuntiva di qualche tipo sui dati). Qui, il *caveat* è contro stratagemmi che deteriorino la qualità dell'offerta di base rispetto ai principi discussi in questo libro bianco, al fine di creare artificialmente domanda per la versione premium”.

Il principio cardine è che i dati, grezzi ed aggiornati, siano disponibili a tutti, gratuitamente, sotto una licenza aperta e a condizioni non restrittive. A questa condizione, può essere sensato valutare ulteriori licenze per casi particolari.

...e se qualcosa non è protetto da alcun diritto?

Per semplicità, abbiamo ipotizzato che tutti i dati siano protetti da diritti di privativa, ma ciò non è vero. Alcuni database sono liberi e riutilizzabili da parte di chiunque. Pensiamo ad esempio ai dataset provenienti da paesi in cui sia previsto un regime di pubblico dominio automatico (come ad esempio gli USA) o semplicemente a dataset di cui siano scaduti i diritti di tutela. Il diritto sui generis (il particolare diritto che tutela le banche dati) ha infatti una durata di 15 anni dalla pubblicazione del dataset; dunque trascorso quel periodo il dataset è di pubblico dominio.

Se ci si trova in uno di questi casi, è possibile etichettare i dataset come liberi utilizzando il Creative Commons Public Domain Mark (<http://creativecommons.org/publicdomain/mark/1.0>). Questa soluzione, per esempio, è proposta anche dalle guidelines governative neozelandesi (si veda www.e.govt.nz/policy/nzgoal).

COME PROCEDERE PER APRIRE I DATI: ASPETTI TECNOLOGICI

Introduzione

I dati aperti hanno bisogno di essere tecnicamente riutilizzabili, oltre che legalmente tali. Abbiamo già detto dell'importanza di rendere possibile un download, possibilmente gratuito, via Internet. E abbiamo sottolineato quanto sia importante che i dati possano essere scaricati in blocco, ad esempio come un unico archivio compresso. A scampo di equivoci, rendere i dati disponibili solo tramite servizi web, ad esempio in pagine con menù a discesa, che restituiscono alcuni dati, sulla base delle scelte dell'utente, non è una modalità adeguata a facilitare il riutilizzo (anche se il servizio può essere utilissimo in quanto tale). Come già immaginerete è anche opportuno che i formati dei file siano aperti e conformi a standard riconosciuti e utilizzati.

Mettere i dati a disposizione online

Il modo più banale, ma anche uno dei più efficaci, per mettere a disposizione dei dati nell'era di Internet, è la semplice pubblicazione online, su un sito Web, di un file statico (eventualmente sotto forma di archivio compresso), il cui link sia chiaramente disponibile secondo lo standard HTTP.

Un altro modo è rappresentato da una Application Programming Interface (API), che renda le informazioni disponibili a richiesta, in maniera più o meno elaborata (ad esempio, permettendo di scaricare solo un certo sottoinsieme dei dati, che interessano all'utente in quel momento, o solo gli aggiornamenti degli ultimi 5 minuti, ecc.). Rendere i dati disponibili tramite un'API, laddove possibile, è utile poiché può facilitare la vita a chi voglia realizzare delle app o mashup di vario tipo (specie se girano su un telefono cellulare).

Malgrado i possibili vantaggi di un'API, il rilascio dei dati (anche) in blocco, in modo che siano facili da scaricare con un semplice link è sempre fondamentale, poiché assicura che tutti possano facilmente ottenere una copia dei dati e redistribuirla; rende facile ad altri sviluppare i propri servizi (ed eventualmente le loro API specifiche, oppure rappresentazioni degli stessi dati in altro formato, ecc.).

Far trovare i dati

Non nascondete i vostri dati: possono avere grandi potenzialità, ma se nessuno sa che esistono e dove sono, resteranno inutilizzati. Spendete dunque qualche energia anche per rendere i vostri dati facilmente individuabili sul Web. Volendo, potete creare un vostro apposito sito/catalogo o portale (a volte detto anche Information Asset Register). Oppure potete verificare se ne esiste uno, ad esempio, nella vostra regione (dati.piemonte.it è un caso di portale regionale aperto a varie pubbliche amministrazioni e istituzioni piemontesi). In Italia, il portale nazionale <http://www.dati.gov.it/> cataloga i dati aperti messi a disposizione da numerose amministrazioni pubbliche. Esiste anche un portale gestito da una comunità di utenti: IT.CKAN.NET (sezione italiana

dell'internazionale CKAN.NET). Chiunque può inserire su CKAN le informazioni relative a database aperti, rendendoli così più facili da scoprire.

Formati aperti, secondo la legge italiana

“Le informazioni sui formati aperti sono consultabili all’indirizzo: www.digitpa.gov.it/formati-aperti. Secondo le Linee guida per i siti web della PA 2010, previste ai sensi della Direttiva 8/2009 del Ministro per la pubblica amministrazione e l’innovazione, è raccomandato l’uso dei seguenti formati aperti e standardizzati: HTML/XHTML per la pubblicazione di informazioni pubbliche su internet; PDF con marcatura (secondo standard ISO/IEC 32000-1:2008); XML per la realizzazione di database di pubblico accesso ai dati; ODF e OOXML per documenti di testo; PNG per le immagini; OGG per file audio; Theora per file video”.

I formati aperti sono una condizione necessaria (e comunque assai raccomandabile), ma non sempre sufficiente. La condizione sufficiente è l’uso di un formato aperto & machine-readable (leggibile in modo automatico). Un formato aperto, lo ricordiamo, deve essere anche sufficientemente documentato perché possa essere interpretato e implementato da terzi senza perdita di informazioni e di interoperabilità. Dunque una semplice tabella di dati CSV o XML senza nessun dato semantico, dei valori di scala, delle interrelazioni tra i vari campi, senza insomma le parti che rendono un dato un'informazione e un dataset tale, non è un formato aperto con cui si possano rendere disponibili i dati. Di questo ci occupiamo più oltre.

Alcuni esempi di ottimi formati machine readable (ed aperti) sono l’XML e il CSV (Comma Separated Value). Per quanto riguarda l’XML, tuttavia, occorre precisare che si tratta di uno standard di alto livello, o di un meta-standard con cui si possono creare linguaggi e formati perfettamente aderenti a XML, ma di fatto non aperti.

Occhio al PDF!

A rigore, il formato PDF è aperto, nel senso che si tratta di uno standard ben documentato ed utilizzabile da parte di chiunque.

Questo formato va dunque benissimo per distribuire documenti, che vadano letti così come sono stati scritti (diciamo che va bene per l’accesso). Ma è pessimo per estrarre dati strutturati. Per chi è interessato al riutilizzo, ricevere una tabella in formato PDF è quasi equivalente ad averla solo su carta (perché in entrambi i casi si rischia di doversela riscrivere a mano in un foglio di calcolo o database opportuno).

In caso di incertezza, la parola d’ordine resta “DATI GREZZI, ORA!”. Buttate fuori quel che c’è, nel formato che anche a voi è comodo utilizzare. Inoltre, pubblicate possibilmente una conversione in un formato aperto, la quale può spesso essere realizzata automaticamente e a costi sostanzialmente nulli. (Inoltre, mentre è importante non pubblicare dati che legalmente non si

sarebbero potuti pubblicare, sbagliare il formato o lo standard non è un problema altrettanto grave: basta segnalare che si tratta di un esperimento, di una prova, di una release alpha... insomma, basta chiarire che in futuro la tecnica di rilascio dei dati potrebbe cambiare, per essere più aperta e più standard.)

Metadati: i dati sui dati

Quando pubblicate un dato, considerate che è spesso necessario spiegare il significato delle colonne e dei campi delle vostre tabelle a terzi che non conosco nulla dei vostri sistemi informativi interni (per esempio, se ci sono codici è necessaria una tabella apposita, che funga da riferimento per espandere i codici stessi). Ovviamente, questo è un aspetto che sarà ben chiaro a chi gestisce i vostri database. Tra le informazioni che può essere utile fornire ci sono quelle sulla qualità dei dati (qual è il livello di precisione, la frequenza di campionamento, ecc.), nonché qualsiasi informazione relativa ai valori che i dati stessi possono assumere (p. es., se le osservazioni al disotto di una certa soglia quantitativa non vengono registrate, ecc.).

Fornire tutte queste informazioni in una modalità che sia comprensibile ad un soggetto esterno alla PA che ha raccolto dei dati può essere faticoso e costoso. La soluzione? Per il futuro, pensate un po' di più all'eventualità che i dati possano essere riusati da terzi. Per il passato, chiarite che mancano dei metadati e che ci state lavorando, ma non fatevi paralizzare dall'assenza di una metadatazione perfetta: come al solito, procedere passo passo! (E considerate che, qualche volta, anche un approccio quick & dirty ha la sua utilità).

No alla discriminazione tra umani e macchine!

Non si dovrebbe mai discriminare l'accesso ai dati tra umani e macchine. Questo principio può far sorridere e più di un lettore avrà pensato a qualche romanzo di fantascienza, ma la questione è rilevante ed è una questione di libertà (di impresa, ma anche di azione democratica). La necessità di registrarsi su un portale, per esempio, oppure la necessità di accettare esplicitamente una licenza e simili sono piccole noie per un essere umano, ma possono diventare ostacoli insormontabili per una macchina.

E questi ostacoli possono rendere i vostri dati difficili da trovare (perché i motori di ricerca non li vedono), oppure difficili da scaricare e aggiornare in automatico, nell'ambito di applicazioni e servizi di terzi. Proprio per questo è anche importante cercare di utilizzare formati di rappresentazione standard e riconosciuti, per quanto possibile.