

# LINKED OPEN TRENTINO

*Pavel Shvaiko, Lorenzino Vaccari,  
Maurizio Napolitano, Michele Barbera*

## *1. Introduzione*

La Pubblica Amministrazione (PA) è un dominio applicativo complesso con una varietà di vincoli e interessi ed un ampio spettro di utenti diversi, come ad esempio cittadini e imprese. Di solito le informazioni vengono create da enti diversi, sono rappresentate in formati diversi e possono essere di diversa qualità. Tuttavia, le fonti dati devono essere interoperabili, permettendone la condivisione e lo scambio di dati tra diverse applicazioni. Le tecnologie semantiche in generale e gli *open linked data* nello specifico hanno un grande potenziale per supportare la PA a gestire gli aspetti di eterogeneità menzionati sopra.

Con il termine *open data* si intende in generale il processo di apertura dei dati a qualsiasi utente che può utilizzarli per qualsiasi scopo con il limite, eventuale, del vincolo di citare la fonte o di proporre lavori derivati che rimangano comunque aperti. Il concetto di apertura, in tal senso, implica che non vi debbano essere restrizioni di alcun tipo nella diffusione dei dati. Le restrizioni possono essere eventualmente di tipo tecnologico o giuridico. Nel primo caso si fa riferimento a formati di file che obbligano l'utilizzo di *software* i cui costi di licenza potrebbero essere elevati o a servizi di distribuzione che, oltre al vincolo analogo ai formati, pongono alcune restrizioni di accesso a livello utente. Nel secondo caso invece si fa riferimento alle norme giuridiche che limitano permessi o introducono restrizioni nell'uso dei dati distribuiti. L'apertura dei dati deve comunque essere soggetta al sistema normativo di riferimento e, pertanto, deve includere la garanzia di aspetti quali la *privacy*, il segreto statistico, ecc.

Il movimento *open data* è iniziato con l'amministrazione statunitense sotto il presidente Obama, diffondendosi poi rapidamente in ogni

parte del mondo. Diversi sono gli esempi (in particolare nel Regno Unito) dove questa strategia è stata identificata essere un ottimo sistema per creare valore economico e sociale. Dalla parte dei produttori dei dati, gli *open data*, sono divenuti una sorta di moltiplicatore del valore prodotto, aumentando l'utilizzazione dei dati creati e pubblicati, ed il loro impatto, senza aumentare il carico di lavoro per la gestione degli stessi, questo anche in virtù del fatto che, i dati, una volta disponibili online, hanno permesso una forte riduzione delle richieste dirette agli stessi fornitori.

Con il termine *open government data* ci si riferisce al caso in cui i dati vengano rilasciati dalla pubblica amministrazione. Quando invece gli strumenti di distribuzione ne semplificano la ricerca e l'integrazione con altri dati creando dei *link* (collegamenti), ci si riferisce agli stessi dati con *open linked data*. Con il termine *linked open government data* si esprime la combinazione del concetto di dati aperti e linkati della PA. I principali vantaggi che derivano dagli *open government data* sono i seguenti:

- aumento della trasparenza e dell'efficienza della PA,
- crescita del coinvolgimento e della partecipazione dei cittadini nella PA,
- riutilizzo dei dati e quindi generazione di nuove opportunità di *business*, nonché stimolazione dello sviluppo economico in generale.

Si fa notare che, secondo il recente studio di Vickery del 2011, i vantaggi economici complessivi di apertura delle informazioni del settore pubblico con un accesso libero e facile o con costi di distribuzione marginali potrebbero arrivare fino a 40 miliardi di euro per l'UE27<sup>1</sup>. Invece, l'aggregato degli impatti economici diretti e indiretti provenienti dalle applicazioni basate sulle informazioni del settore pubblico e il loro utilizzo nell'intera economia dell'UE-27 sono nell'ordine di 140 miliardi di euro annui. Tali numeri mostrano tangibili benefici economici derivanti da un accesso ai dati più facile e da un maggiore uso delle informazioni del settore pubblico.

---

<sup>1</sup> G. VICKERY, *Review of Recent Studies on PSI Re-Use and Related Market Developments*, Paris, 2011.

## 2. Stato dell'arte

In questa sezione vengono riportati: un quadro sintetico dello stato dell'arte relativo alle esperienze di pubblicazione dei dati aperti a livello internazionale e nazionale (§2.1), il riferimento al modello utile per la valutazione del rilascio dei dati aperti in termini della loro usabilità tecnica (§2.2).

### 2.1. Esperienze internazionali e nazionali

Segnaliamo come prime e principali esperienze di pubblicazione di dati aperti, due iniziative internazionali di rilievo relative al Regno Unito ([data.gov.uk](http://data.gov.uk)) e agli Stati Uniti ([data.gov](http://data.gov)) a cui si sono aggiunte successivamente quelle di altri paesi (si veda a questo proposito Schellong and Stepanets<sup>2</sup> per una *review* dello stato dell'arte dell'*open data* in Europa). L'idea comune di tutte le iniziative di pubblicazione dei dati aperti è quella di rendere disponibili al pubblico la maggior parte dei dati non sensibili (pubblici) sui quali sono basati i servizi (on-line) della PA. Alla pubblicazione si accompagna normalmente la spinta verso le aziende e i cittadini ad utilizzare tali dati e a costruire le proprie applicazioni, garantendo in tal modo all'utente rilevanza sui dati esposti e sulla possibilità di fornire servizi innovativi (non-previsti) dalla PA. Si sostiene infatti che il riuso non-previsto sarà il driver principale per il successo degli *open government data*, come è già successo in passato per il *web* dei documenti. Non a caso, sin dal 2009 da quando ha iniziato il suo incarico di presidente degli Stati Uniti, Barack Obama ha emesso una serie di misure di stimolo economico volte anche a costruire e mantenere la fiducia degli elettori, in parte compromessa dalla crisi finanziaria che ha avuto inizio nel 2008. In questo modo, la trasparenza è diventata lo stimolo principale per gli sviluppi futuri del paese. A partire da questa esperienza, l'approccio al tema dei dati aperti è diventato sempre più strutturato come è successo, ad esempio, nel caso degli UK dove è stato creato l'*Open Data Institute*<sup>3</sup> (ODI), che mira a catalizzare

---

<sup>2</sup> A. SCHELLONG, E. STEPANETS, *Unchartered Waters: The State of Open Data in Europe*, in *CSC, Public Sector Study Series*, 2011.

<sup>3</sup> <http://www.theodi.org/>.

l'evoluzione della cultura degli *open data* finalizzata alla creazione di valore economico, ambientale e sociale.

Sul fronte invece delle istituzioni che producono dati si citano i casi di World Bank, Eurostat e ISTAT che, per la pubblicazione dei propri dati aperti e analogamente alle istituzioni governative nazionali, hanno stabilito di rilasciare i propri dati senza restrizioni tecnologiche e legali. Normalmente, la tecnologia utilizzata si basa su formati e standard statistici aperti e su piattaforme online e/o su *plugin* integrabili sul proprio PC per la distribuzione dei dati aperti. A titolo di esempio si cita ISTAT, che ha rilasciato uno specifico *plugin* per Wordpress che permette all'utente di integrare alcuni dati di tipo demografico pubblicati sul *web*. Le scelte relative alle restrizioni giuridiche sulla diffusione dei dati favoriscono il più possibile l'utilizzo (ed anche la commercializzazione) dei dati prodotti. Per esempio, Word Bank distribuisce i propri dati senza alcun tipo di vincolo, adottando di fatto una licenza Creative Commons Zero (CC0), sinonimo del concetto di dominio pubblico. Eurostat (così come per ISTAT), invece, adotta una licenza di attribuzione (ovvero si pone il vincolo di citare sempre la fonte) ma, trattando dati provenienti da diverse nazioni europee, pone ulteriori vincoli minimi sul segreto statistico relativi ai sistemi normativi di ciascun paese di provenienza dei dati.

A seguito di tutte queste iniziative internazionali e anche sulla spinta delle normative internazionali e nazionali in proposito di pubblicazione del patrimonio informativo pubblico, anche in Italia molte regioni ed enti pubblici hanno già iniziato ormai da qualche anno il percorso di pubblicazione del proprio patrimonio informativo pubblico<sup>4</sup>. Si citano qui, a titolo di esempio il Piemonte ([dati.piemonte.it](http://dati.piemonte.it)), l'Emilia-Romagna ([dati.emilia-romagna.it](http://dati.emilia-romagna.it)), la Lombardia ([dati.lombardia.it](http://dati.lombardia.it)), il Veneto ([dati.veneto.it](http://dati.veneto.it)), e la regione Toscana ([dati.toscana.it](http://dati.toscana.it)).

## 2.2. Valutazione dell'usabilità tecnica dei dati rilasciati

La pubblicazione dei dati aperti sul *web* da parte delle diverse organizzazioni, può essere attuata secondo diversi criteri, adottando formati,

---

<sup>4</sup> <http://www.dati.gov.it>.

sistemi e dati di qualità molto diversi tra loro. Sono diversi ad esempio i casi di rilascio di una tabella in formato proprietario piuttosto che in un formato aperto. Uno dei riferimenti per la valutazione dell'usabilità tecnica dei dati rilasciati sul *web* è lo star rating system di Tim Berners-Lee<sup>5</sup>, costituito da cinque livelli, o stelle, di valutazione. Normalmente, le prime 3 stelle sono facili da raggiungere, ma hanno il limite di permettere il riuso di dati solo attraverso un operatore umano che ne deve curare gli aspetti semantici di pulizia, adattamento alle proprie esigenze ed integrazione con gli altri dati. Per ottenere dati più facili da trovare e da integrare è necessario adottare le procedure descritte per i livelli a 4 e a 5 stelle. Il modello tecnico è illustrato sinteticamente nella tabella esposta di seguito:

★	Rendere i dati leggibili da umani via <i>web</i> in qualsiasi formato sotto forma di cataloghi o portali, ad es. anche come immagine acquisita con uno scanner. In questa maniera gli utenti possono visualizzare, stampare e scaricare i documenti. Tuttavia, i dati esposti in questo modo non facilitano la loro integrazione.
★★★	Rendere i dati disponibili in un formato strutturato caricabile direttamente su un computer, es. fogli excel, esportazioni di database, normalmente in formato proprietari. In questo caso i formati di rappresentazione del dato in PDF o le immagini non sono ammessi. Gli utenti hanno accesso ai dati grezzi (se dispongono delle conoscenze sui <i>software</i> proprietari). In questo caso, inoltre, oltre ad adottare <i>software</i> proprietario, l'utente necessita di interpretare il contenuto informativo dei dati (schema concettuale), e quindi occorre fornire anche la documentazione sui attributi e sugli oggetti rappresentati dai dati.

---

<sup>5</sup> <http://inkdroid.org/journal/2010/06/04/the-5-stars-of-open-linked-data/>.

★★★	Utilizzare uno standard strutturato, aperto, non proprietario, es. CSV, XML, RTF. In questa maniera gli utenti non necessitano di <i>software</i> proprietario per accedere direttamente ai dati grezzi e si utilizzano formati che, molto probabilmente, rimarranno gli stessi anche in futuro (formati aperti stabili). Tuttavia, l'interoperabilità è garantita ancora solo a livello sintattico, in quanto non viene specificato il contenuto informativo dei dati.
★★★★	Usare un formato universale e standard sia per la sintassi che per la semantica, ovvero RDF: una lingua franca, dove i termini sono "ben definiti" e puntano a dei vocabolari/ontologie pubblicati sul <i>web</i> (es. news aggregation: RSS; social networking: FOAF). Usare un linguaggio di interrogazione semantico che, nel caso di RDF, è lo SPARQL (Protocol and RDF Query Language).
★★★★★	Linkare i propri dati a dati di altri per fornire il contesto semantico. Si possono realizzare appositi RDF per l'integrazione automatica tra fonti di dati eterogenee. Per ogni nodo RDF usare una URI (identificatore) esterna ben nota: in questo modo, se altri usano la stessa URI, se ne conosce il riferimento comune. Elementi di semantica possono essere incapsulati anche nella rappresentazione grafica (HTML).

Si fa notare che, nonostante sia, ovviamente, auspicabile il rilascio dei dati nella forma descritta dai livelli a 4 o a 5 stelle, che consentono il più potente e il più facile riutilizzo dei dati, la raccomandazione è comunque di non ritardare la pubblicazione dei dati, rilasciandoli "as is" (come sono) al più presto possibile, e poi successivamente elaborarli per migliorarne la qualità a livello superiore.

### 3. *Linked open Trentino*

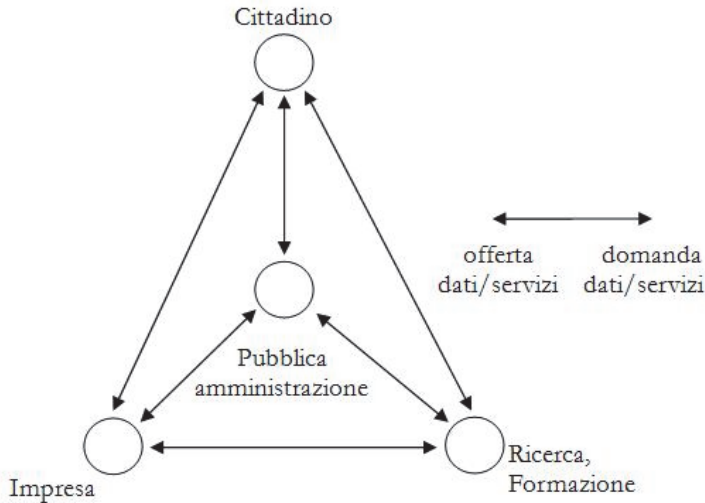
In questa sezione viene riportato l'approccio adottato dal Trentino per il rilascio dei dati aperti. Vengono in particolare descritti: l'ecosistema

stema abilitante al rilascio dei dati aperti (§3.1), gli aspetti istituzionali (§3.2), le licenze per il riutilizzo (§3.3), gli aspetti tecnici (§3.4) ed alcuni esempi progettuali che attuano il paradigma dei *linked open government data* in Trentino (§3.5).

### 3.1.1 L'ecosistema

Di seguito viene illustrato l'ecosistema degli *open government data* ritenuto abilitante allo sviluppo del paradigma di apertura di dati.

La Figura 1 illustra sinteticamente gli attori principali dell'ecosistema che sono: (i) la pubblica amministrazione, (ii) le imprese, (iii) i cittadini, (iv) la ricerca e la formazione. Ognuno di questi attori può produrre ed utilizzare i dati come un servizio interno, come per esempio avviene nello scambio dati tra uffici della PA ai fini di un'autorizzazione amministrativa. Tuttavia, l'attività principale abilitante è quella di produzione e di consumo dati tra gli attori come è nel caso in cui la pubblica amministrazione pubblica i dati aperti che vengono utilizzati in maniera innovativa da imprese ai fini della creazione di nuovi servizi per i cittadini. Per questo motivo, la comunicazione tra gli attori è bidirezionale: ogni attore produce e consuma dati e servizi degli altri attori.



La relazione di ogni attore deve esistere con tutti gli altri attori per aumentare il riutilizzo del dato e dei servizi basati sui dati rilasciati, nonché per abilitare usi impreveduti di dati (eterogenei) provenienti da più attori. Una delle particolarità del Trentino è la forte presenza di centri di ricerca e alta formazione riconosciuti a livello nazionale e internazionale che quindi giocano un ruolo importante nell'ecosistema fornendo conoscenze su soluzioni innovative e nuovi problemi. Un'altra particolarità consiste nel fatto che la PA trentina, centrale nel modello illustrato, è sensibile all'innovazione e all'erogazione dei servizi ai cittadini e alle imprese in maniera efficiente ed efficace.

Nel caso specifico del Trentino, si è osservato inoltre come, oltre alle questioni tecnologiche e giuridiche, sia necessario attuare azioni di comunicazione internamente ed esternamente agli attori coinvolti che stimolano il riutilizzo dei dati rilasciati. Le azioni di divulgazione dell'*open data* non devono essere rivolte solo a chi consuma dati ma anche verso chi li produce e quindi a tutti gli attori coinvolti nell'ecosistema.

Nel caso specifico dei dati prodotti dalla PA ci si scontra prima di tutto in quella che, Brigitte Lutz – responsabile dell'ufficio di E-Government della città di Vienna, definisce la “dura noce dell'open government”, ovvero in tutta quella serie di problematiche, per la maggior parte culturali, che una pubblica amministrazione affronta quando avvia un processo innovativo come il paradigma *open data*. Si tratta di una serie di resistenze che in alcuni casi sono legate a timori, in altri ad effettive problematiche nel sistema organizzativo della produzione. I principali sono:

- insufficienza sulla qualità dei dati,
- mancanza di risorse: *budget* insufficiente o costi di distribuzione o risorse umane,
- paura degli impatti: troppo *feedback* o richiesta di nuovi requisiti.

Ciascuno di questi elementi deve essere trattato con attenzione in ogni situazione da affrontare, facendo presente un chiaro programma di comunicazione e di impostazione del sistema di rilascio dei dati (licenze, formati, procedure, protezione dei dati) in cui, gli *input* derivati dall'esterno devono essere veicolati come ottimi strumenti di *feedback* per migliorare la produzione dei dati, mentre i problemi interni possono diventare una occasione per rinnovare il sistema organizzativo di ciascu-



na struttura (dipartimento) che rilascia pubblicamente il proprio patrimonio informativo.

È importante che la comunicazione verso i fruitori dei dati sia accompagnata dal giusto messaggio e che si offrano, nel contempo, anche gli strumenti necessari ad interpretare correttamente i dati. Gestire dati, interpretarli, estrarre da loro valore è sicuramente una operazione complessa che deve essere accompagnata da strumenti quali informazioni grafiche (infografiche), metadati (documentazione su come i dati sono strutturati e creati), strumenti di *feedback* in modo da favorire la partecipazione, diversi formati aperti per la rappresentazione degli stessi dati (che favoriscono il riuso). In questo modo le distanze tra chi pubblica il dato e chi lo utilizza si riducono e l'aumento di interesse e di utilizzo attorno all'iniziativa di apertura dei dati rappresenta il vero salto di qualità che deve essere comunque accompagnato da un piano di comunicazione di rilascio dei dati.

In mancanza di tale piano, infatti, si rischia di creare un effetto di *wasted (food) data*, ovvero materiale inutilizzato (nel senso di non consumabile). Per superare questo ostacolo, la metodologia da utilizzare è rappresentata dai cinque stadi attraverso cui elaborare una strategia di coinvolgimento nel rilascio dei dati aperti<sup>6</sup>: (i) essere guidati dalla domanda, (ii) inserire i dati nel contesto, (iii) supportare le conversazioni intorno ai dati, (iv) creare capacità, competenze e reti, (v) collaborare su dati come una risorsa comune. La strategia nel rilascio dei dati, in primis, deve creare interesse fra chi produce i dati garantendo pertanto qualità nella pubblicazione degli stessi. Il passaggio successivo coinvolge gli esperti di dominio: un gruppo di persone altamente specializzate, che gestiscono ed utilizzano dati in maniera continua, che ne conoscono il contenuto informativo, e che possono indicare formati e tipo di rappresentazione affinché i dati siano più utilizzabili ed "appetibili". Alle due categorie di utenti indicati, che offrono ed indicano gli aspetti di qualità e tecnologici per la produzione e la pubblicazione del dato, si sommano gli utenti finali, ma anche quegli utenti che hanno dimostrato di essere in grado di usare i dati in maniera più brillante e, magari, totalmente diversa dallo scopo iniziale per cui erano stati creati. Sono

---

<sup>6</sup> <http://www.opendataimpacts.net/engagement/>.

soprattutto questi gli utenti che possono dare un contributo innovativo per creare una cultura efficace ed efficiente intorno al dato rilasciato e, sicuramente, possono rappresentare un volano per la nascita di nuovi scenari imprenditoriali.

### 3.2. *Aspetti istituzionali*

A partire dall'inizio del 2012, con l'apertura dei primi dati geografici<sup>7</sup>, la Provincia Autonoma di Trento (PAT) ha avviato il processo di apertura del proprio patrimonio informativo pubblico (PSI). Con l'apertura dei suoi dati, la PAT ha inteso abilitare efficienza, crescita e partecipazione, favorire la valorizzazione e la creazione di una cultura del dato sul territorio in modo governato sul piano istituzionale, legislativo, organizzativo e tecnologico. Obiettivi specifici della PAT sono<sup>8</sup>:

- accelerare lo sviluppo di servizi innovativi per e con i cittadini come sviluppo naturale del processo di apertura del patrimonio informativo pubblico;
- agire sulla *catena lunga* dei dati dalla loro produzione al loro riutilizzo grazie a servizi che li valorizzano trasformandoli in informazione, quindi in conoscenza, per creare nuovi servizi in grado di rispondere a domande concrete di efficienza e innovazione, derivate dai bisogni che emergono dall'azione amministrativa;
- incentivare l'innovazione delle imprese con modalità di collaborazione centrate alla partecipazione alla community interessata a monte e a valle, alla valorizzazione del patrimonio informativo del territorio nella soluzione di problemi reali con strumenti derivati dal mondo della ricerca;
- costruire un sistema innovativo territoriale che rappresenti potenzialmente uno scenario di riferimento anche per il contesto europeo.

Al fine di attuare gli obiettivi elencati ed avviare formalmente il processo di apertura dei dati alla fine del 2012 la PAT ha approvato con Delibera Provinciale nr. 2858/2012 del 27 Dicembre 2012 le "Linee

---

<sup>7</sup> [http://www.territorio.provincia.tn.it/portal/server.pt/community/sgc\\_-\\_geocatalogo/862/sgc\\_-\\_geocatalogo/32157](http://www.territorio.provincia.tn.it/portal/server.pt/community/sgc_-_geocatalogo/862/sgc_-_geocatalogo/32157).

<sup>8</sup> <http://www.innovazione.provincia.tn.it/opendata>.

Guida per la diffusione e il riutilizzo dei dati pubblici”<sup>9</sup> che contengono gli elementi utili che caratterizzano i dati aperti della PAT. Inoltre, in ottemperanza di quanto previsto dall’art. 9 della L. P. n. 16/2012, la PAT mette a disposizione delle proprie strutture una piattaforma Open Source, per la creazione di un catalogo documentato come unico punto di accesso per facilitare la fruibilità e riutilizzabilità in modalità digitale dei dati resi disponibili secondo le modalità e gli standard definiti nelle Linee, ed in modo conforme a quanto stabilito dalla legislazione provinciale e nazionale. A tale fine le Linee Guida definiscono:

- a) i dati pubblici che possono essere oggetto di riutilizzo;
- b) i criteri e le modalità di individuazione di dati e documenti contenenti dati che potranno essere oggetto di riutilizzo;
- c) le modalità di pubblicazione e le modalità di gestione e aggiornamento della piattaforma di accesso ai dati;
- d) le licenze per il riutilizzo dei dati e dei documenti contenenti dati pubblici di cui l’amministrazione provinciale è titolare, o di cui abbia disponibilità.
- e) i formati di dati e i metadati individuati per la pubblicazione degli stessi.

Di particolare interesse, anche al fine di superare gli aspetti della “dura noce dell’open government” illustrati nella sezione precedente, sono gli ultimi due punti d) e), che verranno sinteticamente descritti nel seguito.

### *3.3. Licenze per il riutilizzo*

Le licenze per il riutilizzo definiscono le condizioni e le modalità di riutilizzo dei dati pubblici e dei documenti contenenti dati pubblici di cui la Provincia Autonoma di Trento abbia la titolarità o la piena disponibilità, consentendone la più ampia e libera utilizzazione gratuita, anche per fini commerciali, in conformità all’art. 8 del d.lgs. 36/2006, e all’art. 9 della L. P. 27 luglio 2012, n. 16, comma 5, nel rispetto dei principi di diffusione del patrimonio informativo provinciale.

---

<sup>9</sup> <http://www.innovazione.provincia.tn.it/notizie/pagina482.html>.

I dati pubblici concessi per il riutilizzo sono licenziati con Licenza Creative Commons Zero (CC0) o, alternativamente, con Licenza Creative Commons Attribuzione (CC-BY). Le licenze Creative Commons Zero e Creative Commons Attribuzione sono state individuate in ragione dell'ampiezza del riutilizzo concesso dalla licenza, anche per fini commerciali, del loro alto livello di interoperabilità con altri modelli di licenze standard, della loro facilità di comprensione e diffusione nel pubblico. Le licenze Creative Commons posseggono un linguaggio semplice e facilmente comprensibile da parte degli utenti e garantiscono la redistribuibilità del dato, nonché un livello di diffusione e di conoscenza a livello nazionale, europeo ed internazionale. Inoltre, grazie alla loro diffusione sulla rete Internet (anche nell'ambito di iniziative volte alla messa a disposizione di documenti e dati pubblici di altre amministrazioni in Italia, in Europa e nel mondo), rappresentano ad oggi lo standard di fatto per la licenza di diritti di proprietà intellettuale online. La loro diffusione minimizza le barriere al riuso, riducendo i costi per i riutilizzatori e massimizzando la probabilità che i dati dell'amministrazione provinciale possano essere combinati con altri dati pubblici e con dati generati dagli utenti della rete o da operatori privati.

Creative Commons Zero (CC0) è una dichiarazione che consente "la più ampia e libera utilizzazione gratuita, anche per fini commerciali" di quanto viene ad essa associato. Si tratta di una licenza in grado di adempiere a quanto previsto d.lgs. 36/2006, art. 8, dall'art. 68, c. 3, lett. b) del Codice dell'Amministrazione Digitale e dalla L. P. 16/2012, art. 9, c. 5. Non si tratta, infatti, di una licenza in senso stretto, intesa quale concessione da parte del titolare di un determinato utilizzo dell'opera, delle informazioni ovvero della banca dati, ma di una rinuncia totale ed incondizionata a qualunque diritto su di essi. Apponendo su un documento la dichiarazione CC0 si rinuncia, infatti, a tutti i diritti sul documento e sui suoi contenuti, dati inclusi, nella misura massima possibile prevista dalla legge, in piena sintonia con le politiche Open Government Data.

Licenza Creative Commons "Attribuzione o equivalente". Una possibile alternativa alla licenza CC0 come sopra individuata è rappresentata dalla licenza Creative Commons "Attribuzione o equivalente" (CC-BY). Anche questa licenza, al pari della standard CC0, è espressione

del principio della “più ampia e libera utilizzazione gratuita anche per fini commerciali” espresso dall’art. 9 comma 5 della Legge Provinciale 16/2012.

Unico obbligo imposto al licenziatario è quello di citare l’autore della banca dati o del documento, oggetto di riutilizzo, nel rispetto delle modalità indicate dall’autore stesso nella licenza o a corredo della stessa, come di seguito meglio specificato (“Attribuzione”). Tale licenza è adottabile per le banche dati che risultano chiaramente tutelate dal diritto d’autore, ai sensi dell’art. 64 quinquies L. 633/41 e s.a.m.i., e/o dal diritto sui generis sulle banche dati, secondo quanto disposto dall’art. 102 bis L. 633/41 e s.m.i..

Al fine di prevenire qualsiasi incertezza interpretativa da parte del licenziatario ed incoraggiare il riutilizzo dei dati, è opportuno chiarire che la licenza stessa si applica sia agli eventuali diritti d’autore relativi alla banca dati licenziata, sia ai cosiddetti diritti sui generis sulla banca dati stessa. Va cioè chiarito che la licenza disciplina tutti i diritti di cui alla L. 633/41 e s.m.i., con esplicita inclusione dei Diritti del costituente di una banca di dati, di cui al Titolo II-bis della legge stessa.

### *3.4. Aspetti tecnici*

Nelle linee guida, la PAT definisce anche cosa si intende per formato aperto ed elenca i formati aperti più comuni ed i metadati per la pubblicazione dei dati aperti del suo patrimonio informativo Pubblico. I formati individuati sono stati selezionati con riguardo agli standard internazionali e alle prescrizioni normative previste dal Codice dall’Amministrazione Digitale (CAD).

#### *3.4.1. Formati aperti*

Come indicato nelle Linee Guida, i dati di tipo aperto devono essere “accessibili attraverso le tecnologie dell’informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti”, cioè resi pubblici, documentati esaustivamente e senza vincoli all’implementazione; “sono adatti all’utilizzo automatico da parte di programmi per elaboratori” (machine readable) e sono “provvisi dei

relativi metadati”. L’accessibilità da parte di altri programmi di elaborazione automatica è ottenuta attraverso l’uso di formati aperti tali da permettere di massimizzare le loro possibilità di riutilizzo e interoperabilità.

A titolo di esempio, la Tabella 1 fornisce un primo elenco dei formati aperti maggiormente diffusi nell’ambito dell’amministrazione provinciale. La tabella, nella prima colonna, elenca il nome, l’acronimo e una breve descrizione del formato. La seconda colonna contiene il tipo di dato pubblicato dal formato aperto. La terza colonna indica l’estensione del file normalmente utilizzata per questo formato.

Tabella 1. Formati aperti adatti alla pubblicazione dei dati aperti.

Nome (Acronimo) <i>Descrizione</i>	Tipo di Dato	Estensione del file
Comma Separated Value (CSV) <i>Formato testuale per l’interscambio di tabelle, le cui righe corrispondono a record e i cui valori delle singole colonne sono separati da una virgola (o punto e virgola)</i>	Dato tabellare	.csv
Geographic Markup Language (GML) <i>Formato XML utile allo scambio di dati territoriali di tipo vettoriale</i>	Dato geografico vettoriale	.gml
Keyhole Markup Language (KML) <i>Formato basato su XML creato per gestire dati territoriali in tre dimensioni nei programmi Google Earth, Google Maps</i>	Dato geografico vettoriale	.kml
Open Document Format per dati tabellari (ODS) <i>Formato per l’archiviazione e lo scambio di fogli di calcolo</i>	Dato tabellare	.ods
Resource Description Framework (RDF) <i>Basato su XML, è lo strumento base proposto da World Wide Web Consortium (W3C) per la codifica, lo scambio e il riutilizzo di metadati strutturati e consente l’interoperabilità tra applicazioni che si scambiano informazioni sul Web</i>	Dato strutturato	.rdf
ESRI Shapefile (SHP) <i>Lo Shapefile ESRI è un popolare formato</i>	Dato geografico vettoriale	.shp, .shx, .dbf, .prj

<i>vettoriale per sistemi informativi geografici. Il dato geografico viene distribuito normalmente attraverso tre o quattro files (se indicato il sistema di riferimento delle coordinate). Il formato è stato rilasciato da ESRI come formato (quasi) aperto</i>		
Tab Separated Value (TSV) <i>Formato testuale per l'interscambio di tabelle, le cui righe corrispondono a record e i cui valori delle singole colonne sono separati da un carattere di tabulazione</i>	Dato tabellare	.tsv
Extensible Markup Language (XML) <i>È un formato di markup, ovvero basato su un meccanismo che consente di definire e controllare il significato degli elementi contenuti in un documento o in un testo attraverso delle etichette (markup)</i>	Dato strutturato	.xml

### 3.4.2. Metadati

Nello specifico ambito dei dati aperti, i metadati sono intesi come i “dati associati ai dati” tali da informare gli utilizzatori del significato del dato stesso. La definizione stessa di metadato evidenzia l’ampiezza dei possibili metadati associabili ad un insieme di dati (dataset), ed è per questo che, nelle Linee Guida, sono stati identificati un insieme minimo di informazioni di base che possano essere utilizzate ai fini della ricerca, dell’accesso e dell’utilizzo del dato pubblicato.

Si fa notare che alcuni dati di settore includevano il metadato già codificato in formato standard, ma nella maggior parte dei casi è stato necessario definire una documentazione di minima non essendo disponibili standard precisi in tale senso. Per tale motivo si riporta di seguito il set minimo di metadati considerati indispensabili per la pubblicazione dei dati aperti da parte della PAT. Per i dati territoriali lo standard utilizzato è quello adottato dal geo-catalogo della PAT (§3.5.1). La Tabella 2 definisce il set minimo obbligatorio di metadati richiesto per la pubblicazione dei dati aperti della PAT.

Tabella 2. Metadati: set minimo.

Metadato	Descrizione
Titolo	Nome del dataset comprensibile da un utente generico. È il nome che viene mostrato all'utente quando consulta il dataset (es. "Limiti amministrativi comunali del Trentino")
Titolare	Ente titolare del dataset
Referente	Struttura o Unità organizzativa referente del dataset
Contatto	Indirizzo email per la richiesta di informazioni
Descrizione	Descrizione testuale del contenuto informativo del dataset
Categorie	Categoria o categorie alle quali il dataset appartiene
Tag/Parole chiave	Lista di termini associati al dataset che lo descrivono, e rendono possibile la sua classificazione e la sua ricerca
Documentazione tecnica	Indirizzo o indirizzi delle pagine <i>web</i> che contengono informazioni utili alla comprensione del contenuto del dataset
Descrizione campi	<p>Se non specificato nella descrizione del dataset o nella "Documentazione tecnica", è necessario descrivere in questo metadato i campi (attributi) delle informazioni contenute nel dataset. Nel caso di dato tabellare, gli attributi corrispondono alle colonne della tabella. Le informazioni da specificare includono:</p> <p>Nome completo. È il nome che viene mostrato all'utente quando consulta il dataset (es. "Descrizione")</p> <p>Nome breve. È il nome dell'attributo utilizzato nel dataset, normalmente un'abbreviazione del nome completo (es. "Desc")</p> <p>Descrizione. Illustra il significato del campo, eventuali relazioni con altri dataset, come il campo è stato generato, etc. (es. "Codice ISTAT del comune amministrativo")</p> <p>Tipologia dell'attributo (testo, numero, data, etc.)</p> <p>Formato dell'attributo (es. unità di misura o struttura della data)</p>
Copertura Geografica	Territorio amministrativo di riferimento (es. "Trento", "Provincia autonoma di Trento")
Copertura Temporale (Data di inizio)	Data di inizio validità
Copertura Temporale (Data di fine)	Data di fine validità



Aggiornamento	Frequenza di aggiornamento (es. “mensile”, “annuale”, “giornaliera”)
Data di pubblicazione	Data in cui il dataset è reso disponibile on-line
Data di aggiornamento	Data di ultima modifica del dataset
Licenza	Tipo di licenza applicata (vedi Allegato A delle linee guida)
Formato	Formato di pubblicazione (standard) del dataset
Codifica Caratteri	Codifica dei caratteri utilizzata (es. “latin-1”, “PC-850”)
Dimensione	Dimensione in byte (o multipli di byte) del dataset
URI permanente	Identificatore univoco del dataset

### 3.5. Alcuni esempi progettuali

Di seguito si illustrano due iniziative progettuali che istanziano l’approccio sopra descritto, ovvero il geo-catalogo semantico (§3.5.1) e il catalogo dei dati aperti del Trentino (§3.5.2).

#### 3.5.1. Il geo-catalogo semantico

Il geo-catalogo mette a disposizione un meccanismo standard per classificare, descrivere e ricercare (anche a livello concettuale) informazioni relative a geo-dati e geo-servizi pubblicati in formato aperto. Il progetto rappresenta un esempio concreto di sperimentazione di una collaborazione tripolare tra aziende, centri di ricerca e utente pubblico. Tecnicamente è stato realizzato uno strumento *web*, il cosiddetto geo-catalogo, per ampliare il portafoglio di servizi del Portale Geo-cartografico Trentino ([www.territorio.provincia.tn.it](http://www.territorio.provincia.tn.it)).

Nell’ambito del progetto, con la delibera n. 195 del 17/02/2012 sono stati rilasciati come dati aperti e linkati alcuni dataset (c.a. 40%) del Sistema Informativo Ambiente e Territorio (SIAT) del Trentino. Per il rilascio dei dati, si è scelto di utilizzare la licenza Creative Commons Zero. Invece, l’approccio tecnico alla pubblicazione dei dataset era costituito da quattro fasi: (i) conversione di dati e di metadati in RDF, (ii) linking dei dataset ad alcuni altri dataset, come ad es. DBPedia e

Freebase, (iii) pubblicazione dei dataset sul *web*, (iv) valutazione dell'utilizzo dei dataset rilasciati<sup>10</sup>.

La Figura 2 illustra i risultati di una ricerca utente attraverso il geocatalogo. La ricerca permette di ricercare il testo digitato, come ad esempio “piste ciclabili”, nell'apposito campo di ricerca (1). L'elenco dei risultati della ricerca viene visualizzato sotto il modulo di ricerca. I dati sono accompagnati dai metadati titolo, punto di contatto (es. “Dipartimento Risorse Forestali e Montane”), parole chiave e descrizione. L'icona sulla sinistra (3) indica il tipo di licenza con cui il dataset è stato rilasciato (licenza d'uso Creative Commons Zero). Per ogni dataset è inoltre possibile: visualizzare il metadato (4), scaricare il metadato in formato XML (5); scaricare il dato grezzo (6); scaricare l'RDF relativo al metadato e al dato (7).

I dataset rilasciati hanno un valore importante per gli utenti finali, perché l'informazione geografica costituisce il livello base su cui costruire servizi di tipo location-based. Statisticamente inoltre, i tre dataset più scaricati sono: comuni amministrativi, aree di pertinenza dei fiumi principali e fiumi. Il progetto, il primo esempio di pubblicazione verticale sui dati geografici per la PAT, ha rappresentato la base per avviare l'iniziativa di pubblicazione di tutto il patrimonio informativo della amministrazione pubblica, descritta al punto §3.5.2.

---

<sup>10</sup> P. SHVAIKO, F. FARAZI, V. MALTESE, A. IVANYUKOVICH, V. RIZZI, D. FERRARI, G. UCELLI, *Trentino Government Linked Open Geo-Data: A Case Study*, in *Proc. 11th International Semantic Web Conference (ISWC)*, volume 7650 of Lecture notes in computer science, pages 196-211, Boston (MA US), 2012.

The screenshot shows the 'Portale Geocartografico Trentino' interface. On the left is a vertical menu with categories like 'Cartografia di base', 'Interoperabilità - Servizi WMS', and 'Ricerca nel Geo-catalogo'. The main area has a search bar containing 'piste ciclabili' (1) and an 'Inizia la ricerca' button (2). Below the search bar, there is a description of the geo-catalog and a 'Risultati: 1-1 (1)' section. The search results for 'PISTE CICLABILI' include a search icon (3), a 'Metadato' button (4), and an 'Scarica XML' button (5). The description section contains a 'Scarica dati' button (6) and a 'Scarica RDF' button (7).

Figura 2. Geo-catalogo: risultati di ricerca.

### 3.5.2. Il catalogo dei dati aperti del Trentino

In attuazione delle linee guida, il 19 Marzo 2013, la PAT ha pubblicato il proprio catalogo dei dati aperti, disponibile all'indirizzo <http://dati.trentino.it>, la cui pagina iniziale è rappresentata in Figura 3.

La piattaforma di pubblicazione del catalogo è basata sul *software open source CKAN* ed è mantenuta dalla *Open Knowledge Foundation*, una organizzazione no profit con sede a Londra che, dal 2004, si fa promotrice di una serie di attività volte alla diffusione dei concetti di apertura (open) nell'era digitale. Le funzionalità di base del catalogo, disponibili a tutti gli attori del territorio provinciale che intendano pubblicare i propri dati come dati aperti, permettono agli utenti di condividere un unico punto di accesso per pubblicare, cercare, accedere ed utilizzare i dati aperti del Trentino. La pubblicazione dei dati può essere effettuata in modo distribuito tramite uno o più account creati per ciascun attore (organizzazione) partecipante alla pubblicazione che provvede all'inserimento dei metadati e della licenza per ciascun dataset come descritti nelle Linee Guida e al caricamento del file o del *link* alle relative risorse associate. L'utente finale ha a disposizione vari meccanismi per la ricerca dei dati (per *keyword*, per nuvola di tag e per categorizzazione) e, una volta identificati i dataset di suo interesse, può eseguire il down-

load dei dati in locale. La piattaforma inoltre mette a disposizione anche una serie di API secondo lo standard DCAT che permettono l'accesso automatico alle risorse del catalogo dei dati aperti.



Figura 3. L'home page del catalogo dei dati aperti della PAT.

#### 4. Discussione

Il contesto più ampio nel quale sono stati creati geo-catalogo, linee guida e catalogo dei dati aperti del Trentino include anche aspetti tecnologici e legislativi di base oltre che aspetti organizzativi, di comunità *open data*, e relativi alla semantica. In particolare, per quest'ultimo aspetto, è importante considerare che il modello *linked data* a cinque stelle sia universalmente riconosciuto e sempre più adottato dalle amministrazioni di tutto il mondo come *best practice* per la pubblicazione di *open data*, tuttavia, esso presenta anche alcune sfide da affrontare, sia dal lato della produzione dei dati, sia dal lato del consumo degli stessi. Se da una parte i *linked data* massimizzano l'interoperabilità aprendo la strada a scenari di riutilizzo dei dati più avanzati, dall'altra

parte, essi si basano su un modello dati a grafo distribuito, che richiede al produttore e al consumatore di comprenderne a fondo la natura e la complessità per essere sfruttato al meglio. Tale radicale cambiamento di approccio rispetto ai noti modelli tabulari e relazionali che hanno dominato i sistemi informativi negli ultimi 30 anni, implica alcune sfide sia sul piano tecnologico, sia sul piano della comunicazione.

Il *web* dei dati si distingue da altri sistemi di organizzazione e condivisione della conoscenza per la sua natura universale, per la dimensione virtualmente infinita e per l'eterogeneità dei dati in esso contenuti. Si tratta di una evoluzione del *web* tradizionale, che si trasforma in una rete di documenti multimediali collegati tra loro da *link* "muti" (in cui non è esplicita la semantica del *link*), a una rete universale di dati strutturati e semi-strutturati machine readable, collegati tra loro da *link* etichettati in modo tale che la semantica del collegamento sia anch'essa comprensibile per un elaboratore. Gli elementi atomici che costituiscono il *web* dei dati sono le entità (luoghi, persone, organizzazioni, eventi, osservazioni statistiche, etc.) e i predicati (*link*) che collegano tra loro le entità secondo delle triple nella forma soggetto - predicato - oggetto. L'unione delle triple, prodotte da individui e organizzazioni diverse e mantenute in luoghi differenti, danno luogo a un grafo globale e universale, a cui un agente *software* è in grado di accedere come un tutt'uno.

Alcune di queste caratteristiche distintive sono già alla base dei principi ispiratori dell'architettura del *web* e influenzano profondamente l'architettura delle applicazioni che producono e consumano i contenuti del *web* dei dati. Un'altra delle caratteristiche peculiari dell'approccio *linked data* consiste nello sfruttare la natura eterogenea dei dati, in modo tale che le applicazioni siano in grado di inferire nuova conoscenza manipolando dati il cui modello e la cui dimensione non è nota a priori.

La flessibilità e le enormi potenzialità di questo modello presentano delle sfide di natura tecnologica e sociale che la comunità internazionale sta in questo momento affrontando. Dal punto di vista tecnologico due delle sfide più pressanti hanno a che fare con la capacità di gestire grandi volumi di dati complessi e densamente interconnessi in modo efficiente e con la capacità di manipolare dati fortemente eterogenei di cui non sempre è nota a priori la struttura (*big data*). Dal punto di vista

sociale e organizzativo le sfide sono non meno ardue, a partire dal radicale cambiamento di approccio necessario a gestire la natura distribuita dei dati in cui i confini tra produttore e consumatore dei dati sono più sfumati e dove il modello permette a chiunque di “alimentare”, arricchire e ripubblicare una base di conoscenza prodotta da terze parti. Nel prosieguo ci soffermeremo sulle sfide di carattere sociale tralasciando volutamente le problematiche tecnologiche, tenendo però a mente che nel *web* dei dati c’è un’mutua influenza ancor più profonda che nei sistemi informativi tradizionali, tra gli aspetti tecnici e quelli socio-organizzativi, che non deve essere trascurata in fase di pianificazione strategica.

*Thinking in the graph.* Perché i *linked data* rappresentano oggi soltanto una piccola porzione degli *open government data*? Uno dei motivi è senz’altro da ricercarsi nella relativa immaturità delle tecnologie e degli strumenti per produrli e consumarli, ma crediamo che l’ostacolo più importante sia legato alla difficoltà di “pensare a grafo”.

Come già accennato il passaggio da un modello tabulare e bidimensionale di rappresentazione della conoscenza, che insieme al modello relazionale sono gli approcci oggi dominanti all’interno delle organizzazioni pubbliche e private, al modello a grafo distribuito, proprio dei *linked data*, ha alcune profonde implicazioni sociali e di architettura dell’informazione. Una prima difficoltà è data dalla scarsa familiarità con la struttura a grafo, sia da parte dei produttori, sia da parte dei consumatori di dati. Manager privati e pubblici, sviluppatori, architetti dell’informazione si sono educati studiando per anni database relazionali e fogli di calcolo, confrontandosi raramente con strutture dati a grafo. Questa abitudine a ragionare su modelli mentali tabulari dei dati rende difficoltoso ai produttori modellare correttamente l’informazione, che di conseguenza diviene difficile da interpretare e da utilizzare da parte dei consumatori. Il problema è esacerbato dal fatto che la larga maggioranza dei *government linked data* oggi esistenti sono prodotti a partire da basi dati relazionali o tabulari pre-esistenti, che spesso non vengono modellate correttamente nella fase di esportazione in formato *linked data*. Infine, i limiti tecnologici dovuti alla novità e all’immaturità delle tecnologie e degli strumenti per la produzione e il consumo dei *linked data*, contribuiscono ad acuire il problema.

Queste difficoltà fanno sì che molte delle basi di conoscenza *linked government data* esistenti siano modellate in modi inconsistenti o difficili da comprendere per il consumatore e che alcune delle applicazioni che utilizzano questi dati non siano in grado di interpretarli in modo corretto a causa di alcuni tipici errori di progettazione.

Il principio AAA e la natura distribuita dei *linked data*. Uno dei principi fondamentali del *web* dei dati è noto come AAA, un acronimo che sta per “Anyone can say Anything, Anywhere”. Letteralmente: chiunque può dire qualsiasi cosa, in ogni luogo (digitale). Questo principio, che è alla base della natura distribuita e universale dei *linked data*, ha un impatto profondo sui paradigmi di produzione e consumo di dati e sui modelli di design delle architetture informative. Nel mondo del *web* dei dati, i dati nascono per essere collegati, estesi e ricombinati con altre informazioni provenienti da terze parti. Nei sistemi informativi tradizionali i dati venivano strutturati in modo tale da facilitarne la comprensione e l’uso da parte degli utenti in uno specifico contesto. Al contrario, nel *web* dei dati è impossibile (e perfino sbagliato) tentare di prevedere come, da chi e in quale contesto i dati verranno riutilizzati e men che meno con quali altri dati essi verranno ricombinati. Dal lato della produzione, il fatto di non conoscere in quale modo i dati verranno utilizzati, dovrebbe spingere i produttori a modellare i dati secondo schemi il più possibile flessibili e universali. Questa flessibilità ha un costo: comporta l’abbandono di ogni tentativo di ottimizzazione del design dell’informazione in funzione dell’usabilità o delle performance e inoltre comporta un radicale cambiamento di paradigma (e una perdita di controllo) per i responsabili dei sistemi informativi. Infine la natura intrinsecamente distribuita del *web* dei dati rende difficoltoso utilizzare gli strumenti tradizionali di gestione della proprietà intellettuale. Si pensi ad esempio alle difficoltà di gestione che possono generarsi utilizzando licenze con clausole di tipo “Attribution” – molto comuni nel mondo degli *open data* – quando una base di conoscenza viene ricombinata più volte da soggetti diversi con una moltitudine di altre sorgenti a loro volta frutto di ricombinazione di dati fatti in tempi e da attori differenti.

Queste sono solo alcune delle sfide che dovranno essere affrontate nel prossimo futuro da tutti i produttori di dati pubblici e privati. La

soluzione a questi e altri complessi problemi deriverà dalla continua sperimentazione di modelli tecnologici, sociali, legali, economici e organizzativi.

## 5. Conclusioni

Il tema dei dati aperti e linkati della PA sta suscitando sempre più interesse nei vari paesi del mondo in quanto favorisce la trasparenza della PA e lo sviluppo economico in generale. Anche in Trentino sono già state avviate alcune iniziative progettuali e un apposito gruppo di lavoro, che promuovono i dati aperti e linkati della PA. Lo scopo di queste iniziative è quello di dare evidenze concrete e inconfutabili su dove i *linked open government data* possono fare la differenza<sup>11</sup>.

---

<sup>11</sup> Si ringraziano per i diversi contributi forniti tutti coloro che hanno collaborato al presente tema, in particolare: Eleonora Bassi, Vincenzo Bertozzi, Roberto Bona, Isabella Bressan, Marco Combetto, Giulio De Petra, Daniela Ferrari, Fausto Giunchiglia, Juan Pane, Luca Senter, Giovanni Tummarello, Giuliana Ucelli.