

Cap. 113040	Missione	Programma	Titolo	Macroaggregato.	Cod. Livello III	Cod. Livello IV
ATTUALE	16	1	1	4	3	99
NUOVO	16	1	1	4	1	2

(trasferimenti correnti a Amministrazioni Locali);

- di comunicare la presente Deliberazione al Consiglio regionale entro dieci giorni dalla data di adozione, con le modalità di cui all'articolo 5 della legge regionale 19 luglio 2013, n. 19;
- di disporre la pubblicazione del presente atto sul Bollettino Ufficiale della Regione ai sensi dell'art.42, comma 7 L.R. n. 28/01.

Il Segretario della Giunta
Dott.ssa Antonella Bisceglia

Il Presidente della Giunta
Dott. Nichi Vendola

DELIBERAZIONE DELLA GIUNTA REGIONALE 26 maggio 2015, n. 1122

Open Data - Approvazione Linee Guida regionali e Piano Annuale Open Data 2015.

Assente l'Assessore allo sviluppo economico, Loredana Capone, sulla base dell'istruttoria espletata dall'AP "Partecipazione e Innovazione" e dalle dirigenti dell'Ufficio Servizi e-government e ICT e del Servizio Ricerca industriale e Innovazione, confermata dalla Direttrice dell'Area di coordinamento Politiche per lo Sviluppo economico, il Lavoro e l'Innovazione, riferisce quanto segue l'Ass. Sasso.

Premesso che

la legge regionale 24 luglio 2012, n. 20 - "Norme sul software libero, accessibilità di dati e documenti e hardware documentato" agisce in coerenza e in continuità con le norme regionali in materia di trasparenza e favorisce il progresso sociale, il miglioramento della qualità della vita e lo sviluppo delle iniziative economiche private legate al riutilizzo delle informazioni del settore pubblico;

con Deliberazione n. 2183 del 24/07/2012 la Giunta Regionale ha approvato i primi indirizzi per l'attuazione della citata L.R. 20/2012 e l'avvio di un processo di diffusione dei dati liberamente accessibili per tutti (open data) per una Pubblica Amministrazione aperta ai cittadini in termini di partecipazione al percorso decisionale (open government);

l'**Agenda Digitale Puglia 2020**, approvata con DGR 1732/2014, individua come una delle sfide principali per le Pubbliche Amministrazioni quella degli "Open data", non solo come strumento indispensabile per una Amministrazione trasparente ma anche come una opportunità per farsi "promotrice dello sviluppo a livello territoriale" ponendo a disposizione della collettività il vastissimo patrimonio di banche dati di proprietà;

lo stesso documento riconosce alla valorizzazione dell'enorme patrimonio informativo regionale la capacità di sostenere la generazione di nuovi strumenti di Business Intelligence a supporto delle decisioni e delle politiche di governo e la definizione di un modello democratico di produzione di "dati aperti";

con DGR 2342/2014 è stato dato avvio alla "**Comunità di pratica**", promossa dalla legge regionale 20/2012 all'art.17, per garantire un confronto aperto costante su questi temi, con particolare riguardo all'Agenda Digitale della Puglia, e sostenere il massimo coinvolgimento dei sistemi regionali produttivi, dell'innovazione, della ricerca e del governo dei territori per la condivisione delle scelte che saranno operate.

In attuazione agli indirizzi dettati con la citata DGR 2183/2012:

sono state avviate iniziative volte a promuovere e diffondere la cultura degli open data e sostenerne e diffonderne un utilizzo attivo e valorizzarne buone prassi;

a febbraio 2013 è stata pubblicata la prima versione del portale www.dati.puglia.it rendendo disponibili i primi Dataset sui temi della Ricerca e dell'Innovazione;

sono state avviate azioni coordinate per garantire la disponibilità dei dati generati da tutte le strutture regionali e dagli enti, agenzie e società indicate nell'art. 2 della L.R. 20/2012.

Il Percorso

Le attività promosse dal Servizio Ricerca industriale e Innovazione ed InnovaPuglia spa per il coordinamento affidatogli con la citata DGR 2183/2012, hanno coinvolto, in diverse fasi, sia le strutture regionali insieme al sistema di agenzie e società partecipate, che gli Stakeholders del sistema pubblico e privato pugliese. In sintesi:

- l'**azione pilota** posta in essere con l'apertura del portale, la pubblicazione dei primi Dataset e la disponibilità di un motore di ricerca semantico, ha restituito primi elementi per testare la qualità, la capacità di valorizzazione del patrimonio informativo liberato e, quindi, verificare il livello di interesse registrato dai dati resi disponibili;
- il primo confronto pubblico sugli Open Data connessi all'ambito della Salute è stato promosso all'interno delle iniziative regionali dedicate al benessere, nell'edizione della Fiera del Levante 2013; alla sperimentazione avviata sui temi della Ricerca e dell'Innovazione si sono così aggiunti i primi **Data set della Sanità**, tema particolarmente sensibile e che necessita particolare cura nel trattamento delle informazioni, ampliando le sinergie con i contenuti del Piano regionale per la sanità elettronica;
- sul fronte interno, in parallelo, il tema degli Open Data è stato oggetto di un **percorso integrato di formazione** per le Regioni Obiettivo Convergenza, promosso dal ForomezPA nell'ambito del progetto "Etica", che ha coinvolto referenti di tutte le strutture interne regionali, insieme agli enti locali pugliesi, e ha permesso il confronto diretto con esperienze e buone prassi di altre amministrazioni pubbliche italiane;
- nell'ambito del percorso partecipativo di Capacità istituzionale Smart Puglia 2020, il Laboratorio del 9 e 10 aprile 2013 dedicato alla **Cittadinanza digitale** ha permesso un efficace confronto con gli Stakeholders regionali, che ha confermato la cen-

tralità degli Open Data come strumento per la piena accessibilità delle informazioni, per la conoscenza condivisa dell'azione amministrativa e per lavorare sulla reputazione e la condivisione delle responsabilità in qualità di cittadinanza attiva nella promozione e nel co-design della comunità sociale (contenuti ripresi in Agenda Digitale Puglia2020).

Le Linee Guida

La prima fase di sperimentazione di www.dati.puglia.it ed i momenti di formazione e confronto descritti, promossi o partecipati dalla Regione Puglia, hanno quindi offerto elementi utili all'elaborazione di una proposta di **Linee Guida per gli Open Data**, che si focalizza su Principi generali, scelte per le modalità di pubblicazione degli open data ed un tutorial operativo, con l'obiettivo di realizzare un modello di riferimento per governare il processo di apertura dei dati pubblici nelle amministrazioni locali regionali - basato sul paradigma dei Linked Open Data;

Le Linee Guida, quindi, partendo dal percorso della Regione Puglia e dal contesto delineato a livello nazionale ed europeo, descrivono i principi generali posti a base del modello operativo che traggnerà i **Linked Open Data** e tracciano obiettivi di medio periodo, che impegnano l'amministrazione regionale a sostenere il processo di cambiamento culturale per la realizzazione di un modello di **Open Government**.

In particolare, il modello operativo individuato intende predisporre tutti gli strumenti coordinati e funzionali al potenziamento di iniziative già attive di apertura dei dati pubblici delle amministrazioni pugliesi, realizzando sinergie con gli attori del territorio - soggetti pubblici, sistema della formazione e ricerca, sistema delle imprese e privati - che dimostrino interesse nello sviluppo di servizi e di contenuti Open Data per la comunità di cittadini, operatori economici e Amministrazioni locali.

La consultazione pubblica

La proposta delle Linee Guida è stata posta in **consultazione pubblica** nel portale dati.puglia.it dall'ottobre 2014, attivando un forum di discussione.

La fase di consultazione sugli indirizzi regionali per gli Open Data è stata arricchita da altri due Laboratori partecipativi, organizzati con i **Comuni**,

sempre nell'ambito del percorso di Capacità istituzionale. Il primo **"Open Data in Open Space"** si è svolto il 25 marzo 2015 a Brindisi ed ha coinvolto il Comune e la comunità degli innovatori attiva nel laboratorio Brindisi SmartLab. Il secondo è stato organizzato con il Comune di Bari e ha coinvolto giovani start up che utilizzano i dati aperti per generare servizi digitali e la comunità di innovatori impegnata nella promozione dei dati aperti. A questi due Laboratori, caratterizzati dall'attivazione della Comunità di pratiche, sono stati invitati anche i Comuni capofila delle Aree Vaste e quelli attivi nei Patti per le Città.

Dalla consultazione è emersa una complessiva condivisione della proposta regionale e un particolare apprezzamento per le scelte proposte e per la previsione di dotarsi di **Piani annuali** degli Open Data. Particolare interesse ha suscitato l'ipotesi di un **Centro di competenze**, luogo aperto dove potranno confluire sinergicamente le molteplici competenze presenti nel territorio per sostenere la diffusione della cultura degli Open Data, la loro valorizzazione, la promozione della domanda pubblica di informazione e innovazione. Questo presidio plurale potrà essere attore di esperienze pilota in materia di Open Data, il cui perimetro tematico di riferimento è già stato disegnato negli obiettivi della programmazione regionale 2014-2020.

Il Piano annuale 2015

Per il 2015, la proposta di Piano prevede la pubblicazione di data set organizzati per aree tematiche, le cui priorità sono state individuate valutando, insieme alle strutture regionali titolari dei dati, sia la loro pronta disponibilità che il potenziale interesse pubblico per i temi trattati.

In questo senso, l'attività ha coinvolto attivamente, oltre al Servizio Ricerca Industriale e innovazione e alla stessa InnovaPuglia spa, le seguenti strutture regionali:

- Servizio Programmazione sociale ed integrazione socio-sanitaria
- Servizio Politiche di Benessere sociale e Pari Opportunità
- Servizio Sistemi informativi e investimenti in sanità
- Servizio Turismo
- Servizio Beni culturali
- Servizio Sport per tutti

- Servizio Assetto del Territorio
- Ufficio Statistico Regione Puglia

Infine, in questa fase, il coinvolgimento della Responsabile della Trasparenza ha permesso di accelerare la previsione di rilascio massivo in formato aperto di una parte dei dati già pubblicati nella sezione **"Amministrazione Trasparente"** del portale regionale, per il quale dovranno essere coinvolte tutte le strutture titolari dei Dati, come descritte nelle tabelle del Piano stesso.

Il Piano Open Data 2015 sarà oggetto di monitoraggio per dare luogo, nella logica di co-progettazione, ad un modello di gestione del paradigma Open Data condiviso, in grado di abilitare la creazione di servizi per il territorio a supporto della catena del valore connessa alla disponibilità di dati pubblici in formato aperto, riutilizzabili.

Pertanto, considerato che

- le **Linee Guida** regionali per gli Open Data offrono alla Regione Puglia e all'intero sistema regionale, pubblico e privato, metodologie e strumenti coordinati e funzionali al potenziamento di iniziative avviate o già attive di apertura dei dati pubblici della Puglia, realizzando sinergie con gli attori del territorio interessati allo sviluppo di servizi e di contenuti Open Data per la comunità di cittadini, operatori economici e Amministrazioni locali;
- che il **Piano annuale** per gli Open Data è uno strumento utile a valorizzare il patrimonio informativo regionale e ad orientare l'attività delle strutture verso una costante e metodica organizzazione dei propri dati ed alla loro pubblicazione in formato aperto;
- che il Piano annuale è, altresì, un riferimento per tutti i soggetti della comunità di innovatori e delle imprese che utilizzano gli Open Data per sviluppare soluzioni innovative e generare nuova economia.

Preso atto

- in relazione al Programma Triennale della Trasparenza e Integrazione 2014-2016, con la recente DGR 773/2015 la Giunta regionale ha approvato un aggiornamento che richiama e conferma l'impegno dell'amministrazione regionale a raggiungere i massimi livelli di accessibilità e conoscibilità dell'azione pubblica, facendo leva sui principi e

sulla portata del processo d'innovazione offerto dagli Open Data nonché dagli indirizzi normativi disegnati dalla L.R.20/2012;

- che le Linee Guida e lo stesso Piano sono, quindi, lo strumento utile per dare attuazione agli indirizzi assunti dalla Giunta regionale con il Programma Triennale della Trasparenza e Integrazione 2014-2016.

Valutato che

- le attività connesse ai Linked Open Data rientrano tra le iniziative ammesse a finanziamento per la realizzazione ed implementazione del Sistema Informativo regionale della Regione Puglia - (SIR-RP), di cui alla DGR 2730 del 18/12/2014, e che il relativo Piano Operativo è già stato approvato secondo quanto previsto dalla Convenzione RP/InnovaPuglia spa, sottoscritta l'11 luglio 2014.

Propone

- di approvare le Linee Guida regionali per i Linked Open Data, composte dei tre documenti allegati e parti integranti del presente provvedimento: "Principi generali per lo sviluppo degli Open Data della Regione Puglia" (Allegato A); "Linee Guida per la pubblicazione degli Open Data - La scelta della Regione Puglia" (Allegato B); "Linee Guida per la pubblicazione degli Open Data - Tutorial" (Allegato C);
- di approvare il Piano annuale degli Open Data 2015, Allegato D, parte integrante del presente provvedimento;
- di confermare l'affidamento del coordinamento delle attività in argomento al Servizio Ricerca industriale e Innovazione, che si avvarrà della competente collaborazione della società in house InnovaPuglia spa.

Copertura Finanziaria di cui alla legge regionale n. 28/2001 e smi

La presente deliberazione non comporta implicazioni di natura finanziaria sia in entrata che di spesa e dalla stessa non deriva alcun onere a carico del Bilancio regionale.

Tutto ciò premesso, l'Assessore relatore, sulla base delle risultanze istruttorie come innanzi illustrate, propone alla Giunta Regionale l'adozione del

conseguente atto finale ai sensi del comma 4, lett. K) dell'art. 4 della L.R. n. 7/97.

LA GIUNTA REGIONALE

udita la relazione istruttoria e la conseguente proposta dell'Assessore relatore;

viste le sottoscrizioni poste in calce al presente provvedimento, che attestano la conformità alla legislazione vigente;

a voti unanimi espressi ai sensi di legge

DELIBERA

di prendere atto di quanto esposto in premessa, parte integrante del presente provvedimento;

- di approvare le Linee Guida regionali per i Linked Open Data, composte dei tre documenti allegati e parti integranti del presente provvedimento: "Principi generali per lo sviluppo degli Open Data della Regione Puglia" (Allegato A); "Linee Guida per la pubblicazione degli Open Data - La scelta della Regione Puglia" (Allegato B); "Linee Guida per la pubblicazione degli Open Data - Tutorial" (Allegato C);
- di approvare il Piano annuale degli Open Data 2015, Allegato D, parte integrante del presente provvedimento;
- di affidare il coordinamento delle attività connesse all'attuazione degli Open Data e dei Linked Open Data al Servizio Ricerca industriale e Innovazione, che si avvarrà della competente collaborazione della società in house InnovaPuglia spa, in continuità con quanto deciso con la DGR 2183/2012;
- di disporre la pubblicazione del presente provvedimento nel Bollettino Ufficiale della Regione Puglia.

Il Segretario della Giunta
Dott.ssa Antonella Bisceglia

Il Presidente della Giunta
Dott. Nichi Vendola

ALLEGATO A

Regione Puglia
Area Politiche per lo Sviluppo Economico,
il Lavoro e l'Innovazione



Regione Puglia

Servizio Ricerca Industriale e Innovazione

Linked Open Data

**Principi generali per lo sviluppo degli
Open Data della Regione Puglia**



InnovaPuglia SpA
Strada prov. per
Casamassima, Km. 3
70100 Valenzano –
BARI
Italia

Maggio 2015

Linee Guida Regionali per la pubblicazione di Open Data – Principi Generali

Sommario

- 1. Introduzione.....**
- 1. Il contesto**
- 2. Il percorso della Regione Puglia.....**
- 3. Risorse e link utili.....**

Linee Guida Regionali per la pubblicazione di Open Data – Principi Generali

1. Introduzione

Questo documento riporta in sintesi un insieme di principi guida per la creazione, sviluppo, diffusione e utilizzo dei dati dell'amministrazione regionale in formato aperto e accessibile alla comunità di stakeholder, imprese, ricercatori e cittadini interessati per qualsiasi scopo al loro riutilizzo.

L'Unione Europea attribuisce al riutilizzo delle informazioni del settore pubblico un ruolo fondamentale, sia per lo sviluppo economico e sociale del territorio (corretto funzionamento dei mercati, libera circolazione di merci, di servizi ed individui, miglioramento della competitività, superamento del divario fra nazioni e cittadini), sia per la diffusione delle nuove tecnologie digitali fra enti pubblici, imprese e cittadini, attraverso l'utilizzo di strumenti informatici per diffondere e memorizzare le informazioni medesime.

In questo contesto, la Regione Puglia intende attivare la "Comunità di pratica" promossa dalla legge regionale 20/2012 (art.17) per garantire un confronto aperto e costante su questi temi, con particolare riguardo all'Agenda Digitale della Puglia, e sostenere il massimo coinvolgimento dei sistemi regionali produttivi, dell'innovazione, della ricerca e del governo dei territori per la condivisione delle scelte che saranno operate.

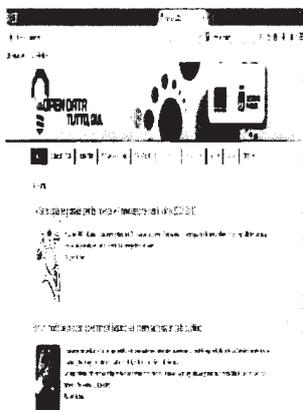
In particolare, la **Comunità di pratica** sarà attiva sul tema della valorizzazione dei dati della PA al fine di:

- a. abilitare l'evoluzione delle modalità in cui usare, condividere e consumare i dati della pubblica amministrazione locale (regione, province, comuni, enti locali, società partecipate), sia all'interno dei propri processi istituzionali che nelle interazioni tra queste amministrazioni con i cittadini e le imprese;
- b. promuovere la trasformazione dei dati in "servizi";
- c. favorire e incoraggiare un percorso di continuo "miglioramento" dei processi decisionali e di creazione di valore dei dati pubblici.

E' auspicabile che, attraverso la condivisione del percorso di evoluzione e crescita dei dati in formato aperto da parte di tutte le amministrazioni locali pugliesi, si possa creare un "ecosistema pugliese" dei dati pubblici che sia in grado di supportare un ampio spettro di risultati positivi per l'intero territorio regionale: servizi ai cittadini più efficienti, migliore livello decisionale delle amministrazioni pubbliche, nuovi modelli di business per le imprese e, quindi, migliore qualità della vita.

Linee Guida Regionali per la pubblicazione di Open Data – Principi Generali

Il Portale degli Open Data regionale abilita ...



... un ecosistema basato sui dati aperti, che supporta



Ricerche e analisi



Monitoraggio performance



Verifica e aggiornamento



Sviluppo nuove applicazioni

... la realizzazione di un'ampia gamma di risultati positivi



Incremento della



Servizi ai cittadini più efficienti



Migliori decisioni



Nuovi modelli di business

Le informazioni di base in merito a cos'è l'Open data e alle attività necessarie per pubblicare dati aperti utilizzando gli strumenti offerti dalla Regione Puglia sono riportate nel documento "Linee Guida per la pubblicazione degli Open Data della Regione Puglia- Tutorial" [LOD_LG_LineeGuida_Tutorial_OpenData_v2.0].

Le "istruzioni per l'uso" per la raccolta, archiviazione, pubblicazione e accesso ai dataset delle strutture operative regionali, nonché le modalità di utilizzo da parte degli Enti locali interessati a condividere la piattaforma regionale, sono riportate nel documento "Linee Guida per la pubblicazione degli Open Data della Regione Puglia - La scelta della Regione Puglia" [LOD_LG_LineeGuida_Operativo_OpenData_v2.0].

Linee Guida Regionali per la pubblicazione di Open Data – Principi Generali

1. Il contesto

Le nuove tecnologie ICT aprono nuove opportunità agli enti pubblici, dal punto di vista dello svecchiamento dei processi decisionali e degli schemi burocratici. La digitalizzazione, inoltre, permette alle istituzioni di adottare un nuovo modello amministrativo, basato su “apertura” e “trasparenza” nei confronti dei cittadini: l’Open Government. Il dibattito sul rinnovamento delle amministrazioni in ottica “open”, dal punto di vista culturale, tecnico e giuridico, ha posto l’accento su un nuovo approccio alla gestione dei dati e delle informazioni in ambito pubblico, anche in questo caso reso possibile dalle tecnologie telematiche.

I dati detenuti dalle Pubbliche Amministrazioni rappresentano un enorme patrimonio e stanno acquisendo un’importanza sempre crescente; infatti, grazie all’uso delle tecnologie info-telematiche, è possibile sia utilizzarli per rendere l’Amministrazione più trasparente ed erogare servizi ancor più efficienti sia riutilizzarli in ambiti differenti da quelli per i quali sono stati raccolti.

In passato, questi dati rimanevano relegati nell’ambito dei procedimenti amministrativi per i quali erano stati formati; in tempi più recenti, invece, si è affermata a livello comunitario la tendenza a rendere queste informazioni conoscibili anche ad altri soggetti.

L’Agenda digitale e, in particolar modo, l’**Agenda Digitale Puglia 2020**, individuano come una delle sfide principali per le Pubbliche Amministrazioni quella degli “Open data” non solo come momento ineludibile di una Amministrazione trasparente ma anche come una opportunità per l’Amministrazione di divenire “promotrice dello sviluppo a livello territoriale” ponendo a disposizione della collettività il vastissimo patrimonio di banche dati di proprietà. L’AD Puglia2020 individua questa scelta di valorizzazione dell’enorme patrimonio informativo regionale utile anche all’attivazione di strumenti di Business Intelligence a supporto delle decisioni e delle politiche di governo e per la definizione di un modello democratico di produzione di “dati aperti”.

Gli Open Data rappresentano quindi l’opportunità da cogliere per promuovere progetti d’innovazione dei servizi al territorio coinvolgendo fin dalla progettazione gli utenti finali (cittadini/imprese), le Pubbliche Amministrazioni, le Università e le PMI ottenendo vantaggi socio-economici che siano sostenibili e basati prevalentemente sul mercato digitale della banda larga e delle applicazioni mobili interoperabili.

L’Agenda digitale costituisce pertanto l’opportunità di ridefinire e dare attuazione a nuovi modelli di rapporto tra Soggetti pubblici e privati ed alla evoluzione dei servizi e delle organizzazioni.

Peraltro, elementi di cambiamento erano già stati introdotti dal decreto sulla semplificazione e la trasparenza della PA (D. Lgs. 150/2009) che ha avviato un percorso di ridefinizione dell’approccio alle tematiche di rinnovamento della macchina pubblica attraverso il passaggio in rete dell’E-gov, l’utilizzo dei servizi digitali già oggi in rete da parte di una platea sempre più vasta, sfruttando il ricambio generazionale e spingendo

investimenti verso le nuove frontiere delle tecnologie applicative attraverso le Smart communities.

Stante i principi ispiratori del contesto in cui si è inteso muoversi, il Progetto ha tenuto in considerazione quanto avviato dalle Regioni Umbria e Puglia che, nel tempo, hanno sviluppato un percorso di innovazione attraverso la partecipazione alle diverse iniziative proposte dai Piani sia nazionali sia di livello europeo.

Principi generali

L'universalità dell'accesso ai saperi e alla conoscenza costituiscono obiettivo essenziale e prioritario del processo di modernizzazione tecnologica della Regione Puglia. Ogni cittadino ha diritto ad accedere, in condizioni di eguaglianza, a tutte le informazioni e ai servizi forniti dalla Pubblica Amministrazione mediante sistemi informatici.

La recente legge regionale del Luglio 2012 agisce in coerenza e in continuità con le norme regionali in materia di trasparenza e avvia un processo di diffusione dei dati liberamente accessibili per tutti (Open Data) per una Pubblica Amministrazione aperta ai cittadini in termini di partecipazione al percorso decisionale (Open Government). In particolare, essa favorisce il progresso sociale, il miglioramento della qualità della vita e lo sviluppo delle iniziative economiche private legate al riutilizzo delle informazioni del settore pubblico.

La Regione Puglia utilizza le tecnologie dell'informazione e della comunicazione per rendere fruibili i documenti e i dati pubblici di cui è titolare, assicurandone la pubblicazione tramite la rete Internet in formati aperti secondo gli standard internazionali. I documenti e i dati pubblici, salvo i casi eccezionali individuati da appositi provvedimenti della Giunta regionale, sono accessibili e riutilizzabili a titolo gratuito nel rispetto della normativa vigente in materia di digitalizzazione della Pubblica Amministrazione, di accesso agli atti amministrativi, di protezione dei dati personali, di riutilizzo delle informazioni del settore pubblico, di diritto della proprietà intellettuale e industriale. Le licenze per il riutilizzo dei documenti e dei dati pubblici predisposte in ottemperanza al decreto legislativo 24 gennaio 2006, n. 36 (Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico), consentono la più ampia e libera utilizzazione gratuita, anche per fini commerciali e con finalità di lucro.

Il riuso degli archivi e l'utilizzo dei dati dagli stessi estratti è consentito solo in osservanza alla normativa prevista dal Codice della privacy di cui al decreto legislativo 30 giugno 2003, n. 196 (Codice in materia di protezione dei dati personali) e successive modifiche e integrazioni.

2. Il percorso della Regione Puglia

Partendo dal contesto delle azioni previste nell'Agenda digitale italiana e dalle iniziative in essere per la sua realizzazione, nonché tenendo conto di tutte le iniziative in materia di sviluppo dell'Agenda italiana ed europea, la Regione Puglia intende realizzare un modello di riferimento per governare il processo di apertura dei dati pubblici nelle amministrazioni locali regionali – basato sul paradigma dei Linked Open Data - e la realizzazione e pubblicazione di Open Data secondo quanto previsto dall'Agenda Digitale italiana e da quanto regolamentato dalla normativa regionale. Il modello si basa sul riconoscimento dei seguenti principi:

- a) definizione di un modello democratico di produzione di “dati aperti” dove i singoli gestori/responsabili dell'informazione sono anche i decisori della pubblicazione, all'interno di una organizzazione strutturata di raccolta e diffusione dei dati;
- b) ampliamento del modello di pubblicazione dei Dataset in formato Open a tutte le Amministrazioni locali interessate attraverso un unico Portale regionale dell'Open Data (www.dati.puglia.it);
- c) utilizzo dell'approccio Linked Open Data a tutto il patrimonio informativo pubblico generato e gestito dalle amministrazioni locali, anche in funzione dello sviluppo di applicazioni di nuova generazione da parte del sistema produttivo locale;
- d) creazione di sinergie operative (in termini di risorse tecniche, strumentali ed economiche) con i progetti ICT di interesse già finanziati dal Sistema della PA pugliese, oltre che di carattere europeo (HOMER: progetto open data del mediterraneo), ottimizzando gli investimenti ed utilizzando i contenuti già disponibili sul territorio;
- e) valorizzazione degli investimenti effettuati, gestione del processo di liberalizzazione dei dati in modalità Open Data, supportando uffici ed Amministrazioni locali in tema di pubblicazione dei dati, promuovendo progetti pilota per l'uso dei dati pubblici, anche favorendo i processi di sviluppo e consolidamento di Living Lab sul tema delle applicazioni di nuova generazione basate sugli OD;
- f) coordinamento di attività per gli uffici regionali e per tutte le Amministrazioni locali interessate alla creazione della “nuvola” di dati del territorio pugliese (Cloud open data Puglia)
- g) ampliamento e adeguamento della piattaforma ICT (www.dati.puglia.it) in grado di mettere insieme le esperienze di interesse già realizzate con gli investimenti pubblici, rendere interoperante la piattaforma realizzata con sistemi analoghi già esistenti in altre regioni, personalizzare strumenti e funzioni al fine di una gestione complessiva del processo di produzione e pubblicazione degli Open Data
- h) realizzazione del framework di riferimento – basato sulla piattaforma ICT Linked Open Data realizzata – su cui abilitare lo sviluppo di nuovi progetti LOD su scala interregionale e nazionale

In particolare, il modello operativo proposto intende predisporre tutti gli strumenti coordinati e funzionali al potenziamento di iniziative già attive di apertura dei dati pubblici delle amministrazioni pugliesi realizzando sinergie con gli attori del territorio - soggetti pubblici, sistema della formazione e ricerca, sistema delle imprese e privati - che dimostrino interesse nello sviluppo di servizi e di contenuti Open Data per la comunità di cittadini, operatori economici e Amministrazioni locali.

Punti chiave del percorso regionale verso la valorizzazione delle informazioni e dei dati di carattere pubblico sono:

1. costruire i contenuti partendo dalla domanda di servizi innovativi
2. stimolare la realizzazione di progetti per l'utilizzo esterno dell'OD attraverso l'ascolto come approccio metodologico di partenza, passaggio mancante in quasi tutte le esperienze attivate in campo nazionale [*costruzione di Community regionali*]
3. definire un'organizzazione per garantire il valore degli OD e per supportare i gestori interni e gli utenti esterni [*ruoli interni di analista di dati, metodologie, procedure e tecnologie a supporto*]
4. utilizzare investimenti realizzati e organizzazione per concentrare le risorse sugli aspetti veri di innovazione e sui processi operativi di gestione ed evoluzione degli OD e del patrimonio informativo pubblico [*evitare le cattedrali nel deserto*]
5. strutturare un modello forte di interrelazione tra le amministrazioni locali pugliesi per non disperdere le iniziative e non sprecare gli sforzi in un contesto di informazione ridondata [*Integrazione come chiave di trasparenza e semplificazione*]
6. individuare gli steps possibili e realizzabili, prendendo in considerazione i risultati raggiungibili, senza usare questa logica per limitare l'impegno o per resistenza al cambiamento! [*osare per ottenere*]

Gli obiettivi di medio periodo che la Regione Puglia intende perseguire sono sintetizzati nei seguenti punti:

- OM.1** sostenere una progettazione regionale che eviti l'improvvisazione delle singole strutture e che non risponda alle sole esigenze burocratiche. Lo strumento primario affinché si realizzi questo obiettivo è rappresentato dal **"Piano annuale degli Open Data regionali"** che stabilisce le priorità di intervento per l'apertura dei dati, le categorie di dati da pubblicare, il formato dei dataset, le licenze di riuso, ecc.
- OM.2** definire la **filiera di produzione degli OD** a partire dagli uffici gestori del dato in un percorso di partecipazione e selezione dei dati sulla base della disponibilità degli stessi in formato digitale e facilmente trasformabili in ottica open
- OM.3** organizzazione specifica per affrontare e gestire i problemi, formare tavoli per la progettualità intersettoriale, inter Amministrativa e verso i Soggetti interessati, anche come incubatore di living lab o di Progetti innovativi
- OM.4** apertura a Soggetti esterni per stimolare forme ed opportunità di sviluppo economico dove il digitale sia fattore di sviluppo
- OM.5** ingegnerizzazione degli Open Data verso i **Linked Open Data** con coniugazione del web semantico nella declinazione dei Dataset e negli strumenti di supporto al processo di pubblicazione
- OM.6** riutilizzo di soluzioni ICT sviluppate in questi anni nell'ambito delle infrastrutture e dei servizi

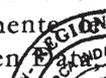
Nello specifico, gli obiettivi prevedono lo svolgimento delle seguenti attività che impegnano tutte le strutture regionali e saranno coordinate dal Servizio Ricerca industriale e Innovazione, avvalendosi del supporto tecnico-specialistico di InnovaPuglia spa:

- a) effettuare una ricognizione dei dati e dei servizi individuando quegli insiemi di dati che possono essere candidati alla pubblicazione in modalità Open Data;
- b) diffondere la cultura dell'Open Data in tutte le strutture organizzative regionali e di Enti e strutture pubbliche con cui ci si propone di costruire un modello comune e

- coordinato di raccolta e di integrazione nell'ottica dei "linked data" (Comuni, Provincie, ARPA, ASL, ecc..) con divulgazione concordata e gestita dei servizi disponibili;
- c) offrire Know How, assistenza e supporto alla pubblicazione a tutti gli attori locali, con l'obiettivo di abbattere per quanto possibile il digital divide in materia all'interno delle Amministrazioni pubbliche, con riferimento soprattutto ai piccoli Enti e Comuni. In questo contesto l'attenzione sarà focalizzata sulla diffusione del modello Open Data e Open Government in tutte le iniziative promosse dalla Regione;
 - d) individuare soluzioni organizzative e una roadmap costituita da singoli interventi concreti e circoscrivibili per arrivare a "Liberare i dati" della Regione Puglia in più ambiti: informazione geografica, dati sanitari, mobilità e trasporti, turismo, conoscenza, statistica, trasparenza...;
 - e) realizzare servizi tecnologici per la pubblicazione di Linked Open Data e renderli disponibili a tutta la comunità locale, sia di natura pubblica che del sistema economico e accademico;
 - f) individuare contenuti informativi di interesse per la collettività, gestiti da Soggetti terzi istituzionali centrali e locali che possono essere pubblicati anche come Linked Open Data;
 - g) intercettare e promuovere la domanda di informazione pubblica – anche in connessione con i Living Lab territoriali - con servizi ed iniziative di promozione e realizzazione di servizi rivolti a soggetti privati e realizzati da soggetti privati;
 - h) individuare ed avviare la realizzazione di sperimentazioni, in collaborazione con Università, Associazioni e Organizzazioni del territorio, soggetti privati che abbiano predisposto Progetti di interesse collettivo e di sviluppo economico basati sull'informazione, per temi quali:
 - "Big Data analysis", collegando i dati pubblici a disposizione con i contenuti informativi raccolti dai grandi contenitori di informazione come ad esempio i Social network, caratterizzando reciprocamente l'informazione istituzionale con gli elementi e gli umori di valutazione degli stessi da parte delle comunità dei cittadini che li "subiscono" e/o "li usano" e di cui molto spesso sono la fonte del contenuto stesso;
 - Crowd Sourcing dell'Open Data nell'ambito del web, per favorire i processi di partecipazione e decentralizzazione dell'attività di creazione e predisposizione dei Servizi. A riguardo il Centro di Competenze ha l'obiettivo di consentire, favorire e supportare questo modello di business dove, attraverso iniziative di privati o di organizzazione come il Living Lab, si affida la progettazione e la realizzazione di un servizio/progetto a un insieme indefinito di persone non organizzate precedentemente.

In quest'ottica la Regione Puglia potrà conseguire finalità di trasparenza e apertura ed avrà ricadute positive in termini di razionalizzazione dell'informazione, adeguamento tecnologico e semplificazione dei servizi erogati. In particolare, un processo che porti all'interoperabilità dei dati pubblici basato sugli standard del web semantico realizza di fatto "il cloud" dei dati pubblici ad uso anche interno della Pubblica Amministrazione.

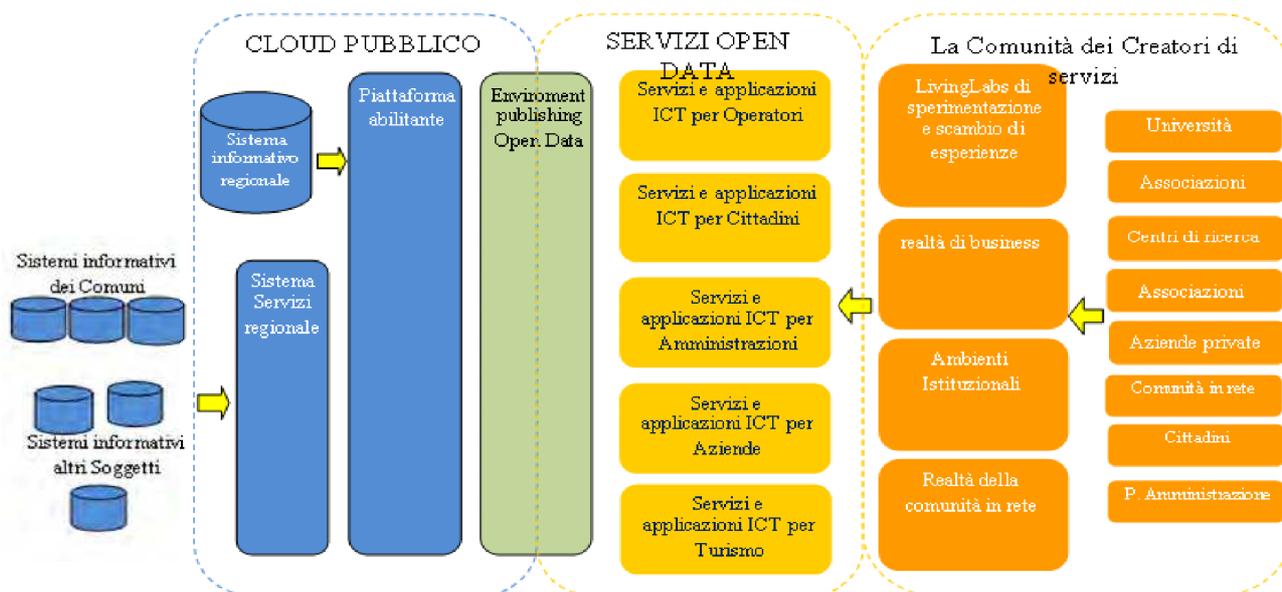
Per assicurare la piena valorizzazione del vasto patrimonio informativo regionale e di tutti i soggetti pubblici del territorio, traggere i macro obiettivi posti dal nuovo paradigma degli Open Data e creare condizioni di mercato più favorevoli e competitive, anche attraverso politiche e modalità di condivisione con gli operatori economici che consentano di creare nuovi servizi dell'economia digitale, è prevista l'attivazione di un **Centro di competenze**. Si tratta di un luogo aperto dove potranno confluire sinergicamente competenze diffuse del territorio per sostenere la diffusione della cultura degli Open



la loro valorizzazione, la promozione della domanda pubblica di informazione e innovazione. Questo **presidio plurale** potrà essere attore di **esperienze pilota** in materia di Open Data, il cui perimetro tematico di riferimento è già stato disegnato negli obiettivi della programmazione regionale 2014-2020.

Il Centro di competenze e tutta la larga gamma dei soggetti che potrà aggregare, dal mondo accademico a quello delle imprese e dell'associazionismo attivo, potrà consolidare l'approccio collaborativo del "fare rete" e divenire punto di riferimento per la generazione di nuove competenze per sviluppare inediti processi di analisi, razionalizzazione e condivisione dei dati e/o realizzazione di nuovi servizi di business e/o innovazione e ricerca. Un laboratorio regionale di innovazione sul tema del riuso dei dati pubblici in modalità Open, aperto alle continue evoluzioni delle tecnologie digitali ed al dinamico mutare dei fabbisogni pubblici, flessibile per potersi adeguare a molteplici necessità organizzative.

La figura seguente rappresenta l'ecosistema dei dati aperti su scala regionale in cui la Regione Puglia intende rafforzare il processo di apertura ampio e sistematico dei dati secondo standard quanto più possibile interoperabili. La comunità dei fruitori della PSI comprende anche i Creatori dei servizi a valore aggiunto basati sugli Open Data.



3. Risorse e link utili

- [1] Ahern, R. Turner, A. Clouse. CMMI Distilled: A Practical Introduction to Integrated Process Improvement. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA 2003, ISBN:0321186133
- [2] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. Redmond, WA, 2009. ISBN: 978-0-9825442-0-4.
- [3] E. Almirall, M. Lee, and J. Wareham. Mapping Living Labs in the Landscape of Innovation Methodologies. *Technology Innovation Management Review*, 12-18, September 2012.
- [4] T. Di Noia, R. De Virgilio, E. Di Sciascio, F. M. Donini. *Semantic Web: Tra ontologie e Open Data*. Apogeo. Aprile 2012.
- [5] <http://www.w3.org/TR/owl2-overview/>.
- [6] Vademecum sui dati aperti
(<http://www.funzionepubblica.gov.it/media/982175/vademecumopendata.pdf>).
- [7] Libro bianco per il riutilizzo dell'informazione del settore pubblico
(http://www.evpsi.org/evpsifiles/bianco_beta.pdf).
- [8] Open data handbook (<http://opendatahandbook.org/it/>).
- [9] Come si fa Open Data
([http://www.reteforum.it/suite/allegati/Come si fa Open Data per la PA.pdf](http://www.reteforum.it/suite/allegati/Come_si_fa_Open_Data_per_la_PA.pdf)).
- [10] Lo stato trasparente – Linked Open Data e cittadinanza attiva
(<http://www.linkedopendata.it/wp-content/uploads/statotrasparente.pdf>).
- [11] Linee Guida per l'interoperabilità semantica attraverso i Linked Open Data
(http://www.digitpa.gov.it/sites/default/files/CdC-SPC-GdL6-InteroperabilitaSemOpenData_0.pdf).
- [12] Agenda Digitale Europea (http://ec.europa.eu/information_society/digital-agenda/index_en.htm).
- [13] European Public Sector Information Platform, “Review of Recent PSI Re-Use Studies Published” (<http://epsiplatform.eu/content/review-recent-psi-re-use-studies-published>).
- [14] Open definition, “Defining the Open in Open Data, Open Content and Open Services” (<http://opendefinition.org/okd/>).
- [15] Open Government Data, “8 Principles of Open Government Data”
(<http://www.opengovdata.org/home/8principles>), Sebastopol, California USA, 2007.
- [16] W3C, Linking Open Data
(<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>).
- [17] W3C, DataSetRDFDumps (<http://www.w3.org/wiki/DataSetRDFDumps>).

- [18] W3C, RDF Working Group (http://www.w3.org/2011/rdf-wg/wiki/Main_Page).
- [19] W3C, OWL Working Group
(http://www.w3.org/2007/OWL/wiki/OWL_Working_Group).
- [20] W3C, SPARQL Working Group (http://www.w3.org/2009/sparql/wiki/Main_Page).
- [21] W3C, SKOS Simple Knowledge Organization System
(<http://www.w3.org/2004/02/skos/>).
- [22] Linked PA – Portale Semantico della Pubblica Amministrazione
(<http://www.ontologiapa.it/>).
- [23] W3C, Semantic Web Deployment Working Group
(<http://www.w3.org/2006/07/SWD/>).
- [24] W3C, Government Linked Data Working Group
(<http://www.w3.org/2011/gld/charter>).
- [25] Creative Commons, “Describing Copyright in RDF” (<http://creativecommons.org/ns>).
- [26] Creative Commons Italia (<http://www.creativecommons.it/Licenze>).
- [27] Creative Commons, CC0 1.0 Universal
(<http://creativecommons.org/publicdomain/zero/1.0/legalcode>).
- [28] Creative Commons, CC-BY Attribution 2.5
(<http://creativecommons.org/licenses/by/2.5/legalcode>).
- [29] Italian Open Data License v.2.0 (<http://www.dati.gov.it/iodl/2.0/>).
- [30] Specifiche SPAQL 1.0(<http://www.w3.org/TR/rdf-sparql-query/>).
- [31] Microformati (<http://schema.org/>).
- [32] Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico (anno 2014), Agenzia per l'Italia Digitale, Presidenza del Consiglio dei Ministri. Ver 0.60 (bozza consultazione):
http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.6.pdf

ALEGATO B

Regione Puglia
Area Politiche per lo Sviluppo Economico,
il Lavoro e l'Innovazione



Regione Puglia

Servizio Ricerca Industriale e Innovazione

Linked Open Data

Linee Guida per la pubblicazione
degli Open Data della Regione
Puglia

La scelta della Regione Puglia



InnovaPuglia SpA
Strada prov. per
Casamassima, Km. 3

70100 Valenzano –
BARI

Italia

Maggio 2015

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

Note

Alcune parti di questo documento sono tratte o rappresentano rielaborazioni dei seguenti documenti:

- **"LINEE GUIDA PER L'INTEROPERABILITÀ SEMANTICA ATTRAVERSO I LINKED OPEN DATA"** pubblicato dalla **COMMISSIONE DI COORDINAMENTO SPC** dell'Agenzia per l'Italia Digitale con licenza Creative Commons-BY-SA.
- **"LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO (anno 2014)"** pubblicato dall' Agenzia per l'Italia Digitale, Presidenza del Consiglio dei Ministri con licenza Creative Commons-BY-SA.
Ver 0.60 (bozza consultazione):
http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicog2014_v0.6.pdf.

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

Sommario

1. Introduzione	
1. Quadro normativo di riferimento	
2. Pubblicazione degli Open Data sulla piattaforma della Regione Puglia	
2.1 Il processo di pubblicazione degli Open Data	
2.1.1 I ruoli	
2.1.2 L'iter di pubblicazione	
2.2 Il processo di produzione dei Linked Open Data	
2.3 Pubblicazione dei Linked Open Data	
2.4 Pubblicazione dei dati non strutturati	
2.4.1 Pubblicazione delle FAQ	
2.5 Licenze d'uso dei dati della Regione Puglia	
2.6 Promozione dei dati aperti e monitoraggio della loro diffusione	
3. Risorse e link utili	

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

1. Introduzione

Questo documento rappresenta un supporto per le strutture operative (Servizi, Uffici, Progetti, ...) della Regione Puglia, oltre che per gli Enti Locali interessati, nella pubblicazione dei dati di loro proprietà nella forma di Open Data (OD). Tale attività di pubblicazione e diffusione dei dati con modalità che ne facilitino il riuso è espressamente prevista dal Codice dell'Amministrazione Digitale.

L'Unione Europea attribuisce al riutilizzo delle informazioni del settore pubblico un ruolo fondamentale, sia per lo sviluppo economico e sociale del territorio (corretto funzionamento dei mercati, libera circolazione di merci, di servizi ed individui, miglioramento della competitività, superamento del divario fra nazioni e cittadini), sia per la diffusione delle nuove tecnologie digitali fra enti pubblici, imprese e cittadini, attraverso l'utilizzo di strumenti informatici per diffondere e memorizzare le informazioni medesime.

Al fine di agevolare il riutilizzo delle informazioni in possesso degli enti pubblici degli Stati membri, l'Unione Europea ha adottato la Direttiva 2003/98/CE del 17 novembre 2003 (oltre all'avvio di varie iniziative tra cui il "Libro Verde sull'informazione Pubblica sulla Società dell'informazione", il "Programma e-Content", la Comunicazione "e-Europe 2002").

La Direttiva in oggetto è stata recepita nell'ordinamento italiano con il Decreto Legislativo 24 gennaio 2006 n. 36, "Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico".

La Direttiva in questione attribuisce agli Stati membri - o all'ente pubblico interessato - la decisione di autorizzare il riutilizzo dei documenti che vengono raccolti, prodotti, riprodotti e diffusi nell'ambito del perseguimento dei propri compiti istituzionali.

Tale facoltà viene altresì riconosciuta nell'ordinamento italiano dall'art. 1 comma 2 del Decreto Legislativo 24 gennaio 2006 n. 36, secondo cui: "*La decisione di consentire o meno tale riutilizzo spetta all'amministrazione o all'organismo interessato, salvo diversa previsione di legge o di regolamento*". La norma precisa, inoltre, che oggetto del "riutilizzo" sono i documenti (rappresentazione di atti, fatti e dati) contenenti dati pubblici, ossia dati conoscibili da chiunque.

Le informazioni di base in merito a cos'è l'Open data e alle attività necessarie per pubblicare dati aperti utilizzando gli strumenti offerti dalla Regione Puglia sono riportate nel documento "Linee Guida per la pubblicazione degli Open Data della Regione Puglia- Tutorial" [LOD_LG_LineeGuida_Tutorial_OpenData_v2.0].

Linee Guida Regionali per la pubblicazione di Open Data .- Operativo

1. Quadro normativo di riferimento

Le presenti Linee Guida sono state predisposte nel rispetto di quanto stabilito in materia di “Open Data” dalla vigente normativa regionale, nazionale ed europea ed in particolare:

- dalla Legge Regionale 24 luglio 2012, n. 20 - Norme sul software libero, accessibilità di dati e documenti e hardware documentato
- dalla Deliberazione della Giunta Regionale 31 Ottobre 2012 n. 2183, – Legge regionale 24 Luglio 2012, n. 20 “Norme sul software libero , accessibilità di dati e documenti e hardware documentato” – Primi indirizzi
- dalla Delibera CIVIT n. 2/2012 - Linee guida per il miglioramento della predisposizione e dell’aggiornamento del Programma triennale per la trasparenza e l’integrità
- dalla Delibera CIVIT n. 105/2010 - Linee guida per la predisposizione del Programma triennale per la trasparenza e l’integrità (articolo 13, comma 6, lettera e, del decreto legislativo 27 ottobre 2009, n. 150) - ALLEGATO – Esempio di struttura della sezione del sito “Trasparenza, valutazione e merito” approvata nella seduta del 14 ottobre 2010
- dal Vademecum Open Data - Come rendere aperti i dati delle pubbliche amministrazioni – Versione Beta 2011
- dalle linee guida per i siti web della PA - art. 4 della Direttiva 8/09 del Ministro per la pubblica amministrazione e l’innovazione
- dal Decreto Legislativo 7 Marzo 2005, n. 82 – CAD: Codice della Pubblica Amministrazione Digitale
- dal Decreto Legislativo 30 dicembre 2010, n. 235 - Modifiche ed integrazioni al decreto legislativo 7 Marzo 2005, n. 82, recante Codice dell’amministrazione digitale, a norma dell’articolo 33 della legge 18 giugno 2009, n. 69.
- dalla Deliberazione della Giunta Regionale 27 novembre 2012, n.2485 – PO FESR 2007-2013 – ASSE I – Linea 1.5.3 “Sviluppo di servizi, contenuti e portali regionali” – Approvazione degli interventi ammessi a finanziamento.
- dal D. Lgs. n. 163/2006 “Il Codice degli appalti pubblici”;
- dal D. Lgs. n. 36/2006 relativo all’accesso all’informazione nel settore pubblico, attuativo della Direttiva 2003/98 relativa alla “public sector information”;
- dalla Legge n. 96/2010 “Disposizioni per l’adempimento di obblighi derivanti dall’appartenenza dell’Italia alle Comunità Europee. Legge Comunitaria 2009”;
- dalla L. 241/90 “Nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti amministrativi”;
- dal D. Lgs. 196/03 “Codice in materia di protezione dei dati personali”;
- dalla Legge 22 aprile 1941 n. 633 “Protezione del diritto d’autore e di altri diritti connessi al suo esercizio” e s.m.i.;
- dalla Direttiva 2007/2/CE che istituisce un’infrastruttura per l’informazione territoriale nella Comunità europea (INSPIRE) nonché dal Decreto Legislativo 27 gennaio 2010 di recepimento come s.m.i. e dalla “Guidance on the ‘Regulation on access to spatial data sets and services of the Member States by Community institutions and bodies under harmonised conditions”.

2. Pubblicazione degli Open Data sulla piattaforma della Regione Puglia

2.1 Il processo di pubblicazione degli Open Data

Nel presente paragrafo si definisce l'organizzazione e il processo che le strutture operative regionali – Servizi, Uffici, Progetti – e gli Enti locali che aderiscono al portale regionale sugli OD (chiamate d'ora in avanti “strutture”) dovranno adottare per pubblicare dataset fino alla terza stella sulla piattaforma regionale.

È importante che tali strutture, siano esse interne o esterne all'amministrazione regionale, siano dotate di una chiara organizzazione interna che identifichi i ruoli e i dati da pubblicare e che individui a chi sono assegnate le responsabilità.

2.1.1 I ruoli

È necessario che all'interno delle strutture siano identificati gli attori che svolgono i seguenti ruoli:

1. Analista dei dati
2. Gestore dei dati aperti

Di seguito sono riportati i compiti assegnati agli attori con i ruoli suddetti.

Analista dei dati:

- Identificare ed analizzare i dati e valutare se esistono i presupposti per la loro pubblicazione
- Proporre i dati da pubblicare in modalità open data

Gestore dei dati aperti:

- Pianificare le attività di pubblicazione dei dati aperti
- Promuovere la pubblicazione dei dati in modalità open data
- Definire le aree tematiche da esplorare per l'identificazione dei dati da pubblicare
- Supervisionare le attività di pubblicazione dei dati aperti
- Approvare le proposte dei dati da pubblicare

2.1.2 L'iter di pubblicazione

Il processo di apertura e di diffusione dei dati da parte delle strutture prevede le fasi riportate di seguito.

A) Identificazione dei dati

Occorre innanzitutto individuare i dati che potrebbero essere pubblicati. La stima del potenziale interesse che i dati possono suscitare non può essere condotta in modo oggettivo. Per questo motivo, se un dato può essere pubblicato senza affrontare costi significativi, si suggerisce di propendere comunque per la loro pubblicazione.

Il lavoro di ricerca dei dati non può prescindere da una attività di censimento dei dati disponibili alle strutture. Per iniziare questo censimento si suggerisce di consultare i sistemi informativi gestiti dalle strutture e in particolare quelli già esposti sul web a beneficio dei cittadini.

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

Durante questa attività sarà utile confrontarsi con i referenti interni alle strutture per le varie aree tematiche. Essi potranno aiutare a definire le strategie di apertura dei dati.

B) Analisi dei dati e scelta dei dati da pubblicare

I dati selezionati nel passo precedente devono essere analizzati per comprendere se esistono ostacoli alla loro apertura. Gli ostacoli più comuni sono i seguenti:

- proprietà dei dati;
- qualità dei dati insufficiente;
- veridicità non appurabile;
- obsolescenza dei dati;
- presenza di dati sensibili;
- costi di estrazione dalla fonte originale non sostenibili;
- possibili conseguenze negative per la pubblica sicurezza;
- altri problemi di natura giuridica.

C) Estrazione dei dati da pubblicare

Se l'estrazione dei dati risulta complessa e soprattutto se i dati da pubblicare necessitano di aggiornamenti successivi, è consigliabile utilizzare o implementare una procedura automatizzata di estrazione e trasformazione dei dati.

In questa fase occorre decidere quale formato utilizzare per i dati da pubblicare e occorre individuare tutte le modifiche necessarie come ad esempio la rimozione dei dati sensibili.

D) Organizzazione dei dataset da proporre come dati aperti

I dati pubblicabili come dati aperti devono essere suddivisi in raggruppamenti coerenti.

In altre parole, occorre una attività di aggregazione dei singoli file contenenti i dati - che vengono definiti "**risorse**" - in raggruppamenti denominati **dataset** in cui coesistono dati relativi allo stesso argomento e da distribuire con la stessa licenza.

E) Definizione dei dati aggiuntivi

A corredo dei dati aperti è utile fornire delle informazioni aggiuntive (**metadati**) che occorre definire accuratamente:

- una descrizione informale dei dati forniti;
- eventuali tag associabili al dataset per categorizzarlo e facilitarne la ricerca;
- la licenza con cui i dati vengono distribuiti;
- l'origine dei dati distribuiti;
- la versione dei dati se applicabile;
- l'autore originale dei dati;
- il responsabile della manutenzione dei dati;

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

- altre informazioni.

F) Pubblicazione dei dati aperti

I dataset individuati durante le fasi precedenti insieme ai metadati aggiuntivi e le risorse che li compongono sono oggetto di una *proposta di pubblicazione*.

Ogni eventuale inserimento di nuovi dati o di versioni aggiornate nello stesso dataset pubblicato deve essere oggetto di una nuova proposta da parte dell'analista dei dati. Le modalità operative di utilizzo sono esplicitate nel documento [*Allegato A - Manuale operativo LOD*].

Le attività più rilevanti di questa fase sono le seguenti:

a. Redazione e candidatura Dataset/risorse

La proposta viene avanzata dall'analista dei dati che, dopo aver reperito tutte le informazioni necessarie alla pubblicazione (attraverso le precedenti fasi), procede inviando la proposta di candidatura al gestore responsabile dei dati aperti della stessa struttura.

b. Approvazione e pubblicazione Dataset/risorse

Il gestore responsabile dei dati, dopo la valutazione delle proposte (candidature) di pubblicazione può:

- approvare determinando l'immediata pubblicazione dei dataset e delle risorse oggetto della proposta;
- sospendere la proposta con eventuale richiesta di integrazioni e modifiche;
- rifiutare la proposta lasciando una motivazione.

2.1.3 Il processo di produzione dei Linked Open Data

La produzione dei Linked Data richiede uno sforzo aggiuntivo rispetto a quello degli Open Data classificabili da uno a tre stelle. Per rendere più efficiente questo processo sarebbe necessario lavorare su più fronti:

- definire un **vocabolario comune regionale** contenente concetti e relazioni che riflettano l'organizzazione della Regione Puglia, degli Enti Locali e delle relazioni tra questi enti. La costruzione di questo vocabolario dovrebbe riutilizzare concetti già presenti in risorse allo stato dell'arte come SKOS e DBpedia.

Il vocabolario dovrà essere definito da esperti del settore che conoscono a fondo il modello organizzativo della Regione e da esperti di modellazione della conoscenza come i centri di ricerca universitari che si occupano di tali tematiche. La pubblicazione di tale vocabolario regionale dovrà prendere in considerazione i vocabolari già esistenti e suggeriti dall'Agenda Digitale Italiana;

- limitare la produzione di dati in formato da uno a tre stelle in favore della produzione di dati direttamente in formato linked. La produzione di tali dati prescinde dalla definizione del vocabolario comune. L'utilizzo del vocabolario comune permetterebbe ai dati di essere direttamente linked;

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

- trasformare i dati di interesse presenti sulla piattaforma attuale in formato linked e utilizzare questo primo processo di trasformazione per avviare la fase di definizione del vocabolario comune.

La definizione del vocabolario comune permetterebbe di trasformare in maniera molto semplice i dati attualmente presenti in formato linked. Generalmente, qualsiasi dato sepolto nelle strutture operative dell'amministrazione regionale (e anche degli Enti Locali) può essere esportato in una forma tabellare come può essere il classico formato CSV, oppure in una forma più complessa che preveda relazioni gerarchiche tra gli elementi come XML.

Come abbiamo visto il modello RDF basato su grafo permette in maniera semplice di poter rappresentare sia dati tabellari che gerarchici.

Quindi, la difficoltà non risiede nella trasformazione del dato in formato RDF ma nella definizione di concetti e relazioni che rappresentino una **visione condivisa e formale** dell'organizzazione regionale.

Facciamo un esempio pratico, prendendo come riferimento due dataset presenti attualmente sulla piattaforma regionale che riportano i dettagli sui progetti ammessi al bando TIC 2010 e 2011.

Di seguito si riporta un record relativo al dataset del bando TIC 2010:

```
<DATA_RECORD>
<PROGRAMMA>PO FESR 2007-2013</PROGRAMMA>
<AZIONE>1.4.1</AZIONE>
<BANDO_AVVISO_PROC_NEGOZIALE>Avviso Pubblico Bando TIC 2010 &#34;Aiuti
alla diffusione delle tecnologie dell&#39;informazione e comunicazione nelle reti di
PMI&#34;</BANDO_AVVISO_PROC_NEGOZIALE>
<SETTORE>Tecnologie dell&#39;informazione e della comunicazione</SETTORE>
<DENOMINAZIONE_RAGGRUPPAMENTO>COSTITUENDA A.T.I. STARCISSETTE -
EUROCISSETTE - LENOCI CED - I.MAR. - COSTANTIELLO
SUPERMERCATI</DENOMINAZIONE_RAGGRUPPAMENTO>
<TITOLO_DEL_PROGETTO>FIDELITY NETWORK</TITOLO_DEL_PROGETTO>
<COSTO_COMPLESSIVO_RICHiesto>800090</COSTO_COMPLESSIVO_RICHiesto
>
<RAGIONE_SOCIALE>LENOCI CED</RAGIONE_SOCIALE>
<DATA_DI_AVVIO>03/01/2011</DATA_DI_AVVIO>
<DURATA_DEL_PROGETTO>15</DURATA_DEL_PROGETTO>
<FORMA_GIURIDICA>societ&#224;
limitata</FORMA_GIURIDICA>
a
responsabilit&#224;
<PROV_SEDE_LEGALE>BA</PROV_SEDE_LEGALE>
<PROV_SEDE_OPERATIVA>BA</PROV_SEDE_OPERATIVA>
```

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

```
<CODICE_ISTAT>63.11.11 - Elaborazione elettronica di dati contabili &#40;esclusi i Centri di assistenza fiscale - Caf&#41;</CODICE_ISTAT>
```

```
<REQUISITO_DI_IMPRESA_FEMMINILE>S</REQUISITO_DI_IMPRESA_FEMMINILE>
```

```
<REQUISITO_DI_IMPRESA_GIOVANILE>S</REQUISITO_DI_IMPRESA_GIOVANILE>
```

```
<COSTO_RICHIESTO>131760,4</COSTO_RICHIESTO>
```

```
<COSTO_AMMESSO>131760,4</COSTO_AMMESSO>
```

```
<CONTRIBUTO_AMMESSO>65880,2</CONTRIBUTO_AMMESSO>
```

```
</DATA_RECORD>
```

Di seguito invece un record estratto dal bando TIC 2011:

```
<DATA_RECORD>
```

```
<PROGRAMMA>PO FESR 2007-2013</PROGRAMMA>
```

```
<AZIONE>1.4.1</AZIONE>
```

```
<BANDO_AVVISO_PROC_NEGOZIALE>Avviso Pubblico Bando TIC 2011 &#34;Aiuti alla diffusione delle tecnologie dell&#39;informazione e comunicazione nelle reti di PMI&#34;</BANDO_AVVISO_PROC_NEGOZIALE>
```

```
<RAGIONE_SOCIALE>Fonderie Orazio E Fortunato De Riccardis S.R.L.</RAGIONE_SOCIALE>
```

```
<NUM_PMI>1</NUM_PMI>
```

```
<NUM_PMI_BENEFIC>1</NUM_PMI_BENEFIC>
```

```
<DENOMINAZ_RAGGRUPPAMENTO>Fonderie Orazio E Fortunato De Riccardis S.R.L.</DENOMINAZ_RAGGRUPPAMENTO>
```

```
<TITOLO_DEL_PROGETTO>Sistema innovativo per la gestione in real-time della catena di produzione</TITOLO_DEL_PROGETTO>
```

```
<COSTO_COMPLESSIVO_RICHIESTO>149320,5</COSTO_COMPLESSIVO_RICHIESTO>
```

```
<DATA_DI_AVVIO>01/12/2011</DATA_DI_AVVIO>
```

```
<DURATA_DEL_PROGETTO>12</DURATA_DEL_PROGETTO>
```

```
<FORMA_GIURIDICA>Societ&#224; a responsabilit&#224; limitata</FORMA_GIURIDICA>
```

```
<PROV_SEDE_LEGALE>LE</PROV_SEDE_LEGALE>
```

```
<PMI_CON_ISO14001>1</PMI_CON_ISO14001>
```

```
<PMI_CON_EMAS>0</PMI_CON_EMAS>
```

```
<PMI_CON_SA_8000>0</PMI_CON_SA_8000>
```

```
<CODICE_ISTAT>24.51.00 - Fusione di ghisa e produzione di tubi e raccordi in ghisa</CODICE_ISTAT>
```

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

```

<PMI_IMPRESA_FEMMINILE>0</PMI_IMPRESA_FEMMINILE>
<PMI_IMPRESA_GIOVANILE>0</PMI_IMPRESA_GIOVANILE>
<COSTO_RICHIESTO>149320,5</COSTO_RICHIESTO>
<COSTO_AMMISSIBILE>106989,02</COSTO_AMMISSIBILE>
<CONTRIBUTO_AMMISSIBILE>53494,51</CONTRIBUTO_AMMISSIBILE>
</DATA_RECORD>

```

La prima cosa che possiamo notare è che ci sono campi che fanno riferimento allo stesso “concetto” ma con denominazione diversa ad esempio: “REQUISITO_DI_IMPRESA_GIOVANILE” e “PMI_IMPRESA_GIOVANILE” che ovviamente vanno ricondotti ad un unico concetto. Inoltre, possiamo notare degli elementi che sono comuni a tutti i bandi: un titolo, il programma e l’azione a cui fanno riferimento. Oppure il “CODICE_ISTAT” sarà sicuramente prelevato da una gerarchia di codici ISTAT esistente che potrebbe essere facilmente definita in un’ontologia e riutilizzata da tutti i dataset che hanno bisogno di utilizzare i codici ISTAT.

L’amministrazione regionale intende dotarsi di propria ontologia per la definizione dei **programmi** e delle **linee di intervento** che potrebbe essere riutilizzata da tutti i dataset che necessitano di quel tipo di informazioni. Allo stesso modo, si potrebbe definire il concetto di **bando** e la sua descrizione, ovvero un titolo, il nome dell’avviso della procedura negoziale, ecc... Tale ontologia potrebbe fare riferimento a vocabolari già realizzati in ambito europeo o nazione.

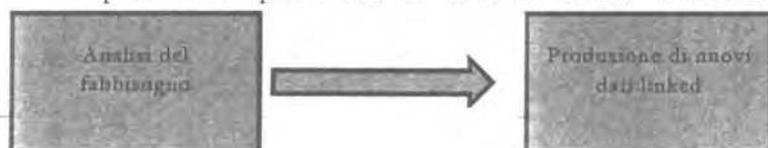
Da questo semplice esempio si evince come il processo di **produzione dei linked data** sia un processo complesso che richiede una serie di interventi e competenze. Affinché la Regione possa dotarsi di un ecosistema di dati linked è necessaria la definizione di un vocabolario comune. La realizzazione di tale vocabolario richiede uno sforzo iniziale che però garantisce nel futuro l’interoperabilità tra i diversi dataset. Il vocabolario comune garantirebbe alla Regione anche un’azione di controllo e verifica dei dataset prodotti da Enti Locali differenti.

Il processo che l’amministrazione regionale intende attuare è riassunto nel



seguente schema a blocchi:

- In una prima fase si selezionano tra i dataset già pubblicati quelli di maggior interesse e si procede alla loro trasformazione in dati linked producendo nello stesso tempo una prima versione di un’ontologia Regionale.
- Successivamente si applica un processo costante in cui si analizza il fabbisogno per capire quali sono i dati di maggior interesse, si producono questi dati in formato linked utilizzando le risorse già



Linee Guida Regionali per la pubblicazione di Open Data - Operativo

definite nell'ontologia Regionale.

Durante questo processo continuo potranno emergere nuovi concetti e relazioni che andranno ad arricchire in maniera incrementale l'ontologia Regionale.

2.2 *Pubblicazione dei Linked Open Data*

La pubblicazione dei Linked Open Data (LOD) è un **processo incrementale** che potenzialmente segue il corso del percorso di apertura dei dati pubblici (percorso OD) rilasciando progressivamente dati sempre più raffinati. È infatti ragionevole pensare che subito a valle del processo di reingegnerizzazione si possano già rilasciare i primi LOD, e che successivamente gli stessi siano prima affiancati e poi sostituiti da versioni concettualmente più adeguate, inter-linked, e arricchite.

È fondamentale che la piattaforma tecnologica metta a disposizione delle funzionalità che facilitino il riutilizzo e l'interoperabilità: per questo è importante mettere a disposizione (direttamente o tramite "hosting") una funzionalità di ricerca (SPARQL endpoint) che consenta di interrogare agevolmente i dataset pubblicati. Inoltre, è altrettanto importante individuare una soluzione che consenta un'integrazione leggera e flessibile con il sistema informativo e organizzativo istituzionale.

In questo modo è possibile garantire la sostenibilità del percorso OD nel tempo e soprattutto l'aggiornamento costante dei dataset pubblicati. Se è infatti ragionevole pensare che un'amministrazione compia un sforzo iniziale importante per il rilascio dei primi dati, è altrettanto evidente che tale impegno non possa essere mantenuto a lungo nel tempo. Per questo motivo è irrinunciabile dotarsi di strumenti tecnologici e organizzativi tali da rendere sostenibile nel tempo lo sforzo richiesto dal processo di pubblicazione e aggiornamento dei dati.

Di seguito quattro "best practice" che possono essere adottate in fase di pubblicazione di linked open data:

- È preferibile pubblicare pochi dati ma di buona qualità e in modalità Linked, anziché pubblicare grosse quantità di dati non interoperabili
- Fare attenzione a non pubblicare dati con URI inconsistenti, soprattutto quando i dati sono di tipo dinamico. Le URI di una stessa entità non devono variare al cambiare della versione del dataset
- Individuare una soluzione tecnologica per la pubblicazione che permetta un'integrazione leggera e flessibile con il sistema informativo e organizzativo dell'amministrazione
- Non limitare il servizio di pubblicazione al semplice download dei Linked Open Data, ma consentire accessi puntuali ai dati con standard come SPARQL

Al fine di permettere il riuso agile dei LOD, è opportuno esporre uno SPARQL endpoint; in questo modo si sfrutta pienamente la strutturazione dei dati e si consente l'accesso puntuale al dato di interesse.

2.3 Pubblicazione dei dati non strutturati

Per quanto riguarda i dati non strutturati l'esistenza di standard ormai consolidati permette di semplificare il processo di pubblicazione. L'importante è che qualsiasi piattaforma regionale (sito web o portale dell'amministrazione locale) esporti il suo contenuto attraverso uno più **feed RSS** aderenti allo standard 2.0. L'utilizzo dei feed RSS permette a qualsiasi agente software di connettersi e prelevare in maniera semplice le informazioni. I feed RSS potranno fare riferimento a pagine pubblicate sui siti delle amministrazioni locali attraverso le URL, è importante che queste pagine siano scritte in maniera aderente alle seguenti linee guida:

1. tra i metadata della pagina sia sempre dichiarato in maniera corretta il set di caratteri utilizzato (ISO, UTF, ecc...);
2. che la pagina contenga i metadati minimi relativi a titolo, autore, data e keyword;
3. che la sezione della pagina contenente il contenuto informativo più importante sia facilmente individuabile rispetto alle altre sezioni, questo si può ottenere utilizzando degli attributi specifici nel codice HTML della pagina;
4. evitare, ove possibile, che la URL punti ad una pagina dinamica che viene generata in base ai parametri della richiesta HTTP;
5. evitare assolutamente l'utilizzo di formati proprietari per la pubblicazione dei dati, in generale si dovrebbe evitare di incapsulare contenuti testuali all'interno di formati binari, sarebbe auspicabile avere a disposizione direttamente il testo all'interno del feed RSS o in formato HTML;
6. evitare la pubblicazione dei contenuti in PDF e soprattutto di includere il testo come immagine scansionata in un PDF o altro formato.

In generale, come linea di principio, tutti i contenuti testuali dovrebbero essere forniti come semplice testo o codice HTML evitando l'utilizzo di formati binari.

Ad esempio, se una determina è contenuta in file PDF, all'interno del feed RSS si dovrebbe inserire il contenuto testuale negli appropriati tag XML e l'URL potrebbe puntare al documento PDF. In questo modo il campo testo può essere utilizzato dall'agente software per indicizzare il contenuto, invece all'utente viene mostrato il PDF.

2.3.1 Pubblicazione delle FAQ

Le FAQ rappresentano un'importante e interessante sorgente informativa. Ad esempio, per il portale Sistema Puglia sono fondamentali durante l'iter procedurale di un bando per mettere in contatto la cittadinanza con gli esperti regionali.

Le FAQ rappresentano un canale di comunicazione semplice e privilegiato per supportare i cittadini e le imprese nelle loro interazioni con l'amministrazione locale. Non è un caso che a livello nazionale il call center della pubblica

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

amministrazione "Linea Amica"¹ pubblici le FAQ di maggiore interesse in formato Open Data. Le FAQ oltre ad essere pubblicate sul portale sono disponibili in formato Open Data con licenza IODL 2.0.

Il formato adottato da "Linea Amica" sarà utilizzato dalla Regione Puglia per pubblicare tutte le sue FAQ. A titolo di esempio si riporta un record relativo ad una singola FAQ:

```
<node>
<Titolo>&lt;a href="/risposte/collocamento-mirato-sospensione-obblighi-assunzione-disabili"&gt; Collocamento mirato: sospensione degli obblighi assunzione disabili&lt;/a&gt;</Titolo>
<Domanda>&lt;p&gt;Presso quale ufficio l'azienda deve presentare richiesta di sospensione degli obblighi di assunzione dei lavoratori disabili?&lt;/p&gt;
&lt;p&gt; &lt;/p&gt;
</Domanda>
<Risposta>&lt;p&gt;In base a quanto previsto dal &lt;a href="http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:2000-10-10;333{vig}"&gt;Decreto del Presidente della Repubblica del 10 ottobre 2000, n. 333 &lt;/a&gt;articolo 4 commi 1 e 3, la richiesta di sospensione degli obblighi occupazionali di lavoratori disabili per le imprese interessate da interventi di integrazione salariale, deve essere presentata al Centro per l'impiego della provincia presso cui è ubicata la sede dell'impresa.&lt;br /&gt;
&lt;/p&gt;
</Risposta>
<Fonte>FormezPA</Fonte>
<Data-di-creazione>&lt;spanclass="date-display-single" property="dc:date"
datatype="xsd:dateTime" content="2012-04-24T18:16:44+02:00"&gt;Martedì, 24 Aprile,
2012&lt;/span&gt;</Data-di-creazione>
<Data-di-aggiornamento>&lt;spanclass="date-display-single" property="dc:date"
datatype="xsd:dateTime" content="2014-05-02T15:34:12+02:00"&gt;Venerdì, 2 Maggio,
2014&lt;/span&gt;</Data-di-aggiornamento>
<Aree-tematiche>&lt;a href="/aree-tematiche/lavoro" typeof="skos:Concept"
property="rdfs:labelskos:prefLabel" datatype=""&gt;Lavoro&lt;/a&gt;</Aree-tematiche>
<Tag-liberi>&lt;a href="/tags/lavoro-e-disabilita" typeof="skos:Concept"
property="rdfs:labelskos:prefLabel" datatype=""&gt;Lavoro e disabilità&lt;/a&gt;</Tag-liberi>
</node>
```

Il formato XML è molto semplice, ogni FAQ è racchiusa in un tagnode, che contiene i seguenti elementi:

¹<http://www.lineaamica.gov.it/>

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

- Titolo: il titolo della FAQ. Come si può notare l'elemento contiene direttamente codice HTML
- Domanda: la domanda collegata alla FAQ
- Risposta: la relativa risposta
- Fonte: la fonte che ha fornito la risposta
- Data-di-creazione: la data in cui la FAQ è stata creata
- Data-di-aggiornamento: la data in cui la FAQ è stata eventualmente modificata
- Aree-tematiche: l'area tematica a cui fa riferimento la FAQ
- Tag-liberi: una serie di tag che descrivono la FAQ

Caratteristica interessante è quella che alcuni tag come Data-di-creazione, Data-di-aggiornamento, Aree-tematiche, Tag-liberi, contengono del particolare codice HTML che permette di definire predicati RDF all'interno dell'HTML. Questa particolare tecnica rientra in quelli che sono chiamati **microformati**. Per una più approfondita discussione si faccia riferimento al seguente link: <http://schema.org/>.

I microformati permettono di aggiungere metadati ad elementi di una pagina HTML, questi metadati possono essere utilizzati dai motori di ricerca per aggiungere semantica alla descrizione di una determinata pagina. L'utilizzo dei metadati è fortemente consigliato per poter permettere agli agenti software di estrarre automaticamente la semantica di una pagina.

2.4 Licenze d'uso dei dati della Regione Puglia

Considerato quanto deliberato dalla Giunta della Regione Puglia con deliberazione del 31 ottobre 2012 n. 2183:

“scegliere l'ultima versione rilasciata, attualmente la 2.0, della Italian Open Data License (IODL), referenziata nel portale nazionale www.dati.gov.it, come licenza di distribuzione dei dati aperti della Regione Puglia”

si consiglia la pubblicazione degli Open Data con licenza IODL 2.02. Tale licenza prevede la libertà di:

- riprodurre, distribuire al pubblico, concedere in locazione, presentare e dimostrare in pubblico, comunicare al pubblico, messa a disposizione del pubblico inclusa, trasmettere e ritrasmettere in qualunque modo, eseguire, recitare, rappresentare, includere in opere collettive e/o composte pubblicare, estrarre e reimpiegare le Informazioni;
- creare un Lavoro derivato ed esercitare sul Lavoro derivato i diritti di cui al punto precedente, per esempio attraverso la combinazione con altre informazioni (mashup).

A condizione di:

²<http://www.dati.gov.it/i odl/2.0/>

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

- indicare la fonte delle Informazioni e il nome del Licenziante, includendo, se possibile, una copia di questa licenza o un collegamento (link) ad essa.
- non riutilizzare le Informazioni in un modo che suggerisca che abbiano carattere di ufficialità o che il Licenziante approvi l'uso che fai delle Informazioni;
- prendere ogni misura ragionevole affinché gli usi innanzi consentiti non traggano in inganno altri soggetti e le Informazioni medesime non vengano travisate.

La licenza, inoltre, consente i seguenti utilizzi:

Tu puoi esercitare i diritti concessi con la presente licenza in modo libero e gratuito, anche qualora la finalità da Te perseguita sia di tipo commerciale. La presente licenza non intende in alcun modo creare ulteriori diritti in capo al Licenziante rispetto a quelli previsti dalla legge sul diritto d'autore o ridurre, limitare o restringere alcun diritto di libera utilizzazione o l'operare della regola dell'esaurimento del diritto od altre limitazioni dei diritti sulle Informazioni derivanti dalle leggi applicabili.

2.5 Promozione dei dati aperti e monitoraggio della loro diffusione

I dati aperti pubblicati necessitano di un'opportuna attività di comunicazione affinché i potenziali utilizzatori vengano a conoscenza della loro disponibilità. Si suggerisce quindi di associare la pubblicazione dei dati ad attività di promozione come la segnalazione dei dataset pubblicati ad altri cataloghi come dati.gov.it.

Per valutare il grado di accoglimento dei dataset e l'interesse che gli utenti hanno espresso è possibile monitorare il numero di visite alle pagine ed il numero dei download nonché le attività social legate ai dataset come l'aggiunta di commenti ed i voti espressi dagli utenti.

Per promuovere la realizzazione e la pubblicazione di nuovi dati aperti, seguendo le direttive della Giunta Regionale, si potranno attivare percorsi partecipati dedicati alle imprese innovative e alla start-up giovanili per la sperimentazione di applicazioni inedite di riuso dei dati aperti, generando interesse attivo all'interno dell'intero sistema produttivo pugliese e creando condizioni di replicabilità di buone prassi consolidate. Tali percorsi possono includere l'organizzazione di Hackathon specifici per la realizzazione di Open Data e Linked Data. Gli Hackathon potrebbero focalizzarsi su alcuni ambiti specifici, ad esempio: mobilità, sanità, turismo e ambiente.

3. Risorse e link utili

- [1] Ahern, R. Turner, A. Clouse. *CMMI Distilled: A Practical Introduction to Integrated Process Improvement*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA 2003, ISBN:0321186133
- [2] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. Redmond, WA, 2009. ISBN: 978-0-9825442-0-4.
- [3] E. Almirall, M. Lee, and J. Wareham. Mapping Living Labs in the Landscape of Innovation Methodologies. *Technology Innovation Management Review*, 12-18, September 2012.
- [4] T. Di Noia, R. De Virgilio, E. Di Sciascio, F. M. Donini. *Semantic Web: Tra ontologie e Open Data*. Apogeo. Aprile 2012.
- [5] <http://www.w3.org/TR/owl2-overview/>.
- [6] Vademecum sui dati aperti (<http://www.funzionepubblica.gov.it/media/982175/vademecumopendata.pdf>).
- [7] Libro bianco per il riutilizzo dell'informazione del settore pubblico (http://www.evpsi.org/evpsifiles/bianco_beta.pdf).
- [8] Open data handbook (<http://opendatahandbook.org/it/>).
- [9] Come si fa Open Data ([http://www.reteforum.it/suite/allegati/Come si fa Open Data per la PA.pdf](http://www.reteforum.it/suite/allegati/Come_si_fa_Open_Data_per_la_PA.pdf)).
- [10] Lo stato trasparente – Linked Open Data e cittadinanza attiva (<http://www.linkedopendata.it/wp-content/uploads/statotrasparente.pdf>).
- [11] Linee Guida per l'interoperabilità semantica attraverso i Linked Open Data (http://www.digitpa.gov.it/sites/default/files/CdC-SPC-GdL6-InteroperabilitaSemOpenData_0.pdf).
- [12] Agenda Digitale Europea (http://ec.europa.eu/information_society/digital-agenda/index_en.htm).
- [13] European Public Sector Information Platform, "Review of Recent PSI Re-Use Studies Published" (<http://epsiplatform.eu/content/review-recent-psi-re-use-studies-published>).
- [14] Open definition, "Defining the Open in Open Data, Open Content and Open Services" (<http://opendefinition.org/okd/>).
- [15] Open Government Data, "8 Principles of Open Government Data" (<http://www.opengovdata.org/home/8principles>), Sebastopol, California USA, 2007.
- [16] W3C, Linking Open Data (<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>).
- [17] W3C, DataSetRDFDumps (<http://www.w3.org/wiki/DataSetRDFDumps>).
- [18] W3C, RDF Working Group (http://www.w3.org/2011/rdf-wg/wiki/Main_Page).
- [19] W3C, OWL Working Group (http://www.w3.org/2007/OWL/wiki/OWL_Working_Group).
- [20] W3C, SPARQL Working Group (http://www.w3.org/2009/sparql/wiki/Main_Page).
- [21] W3C, SKOS Simple Knowledge Organization System (<http://www.w3.org/2004/02/skos/>).

Linee Guida Regionali per la pubblicazione di Open Data - Operativo

-
- [22] Linked PA – Portale Semantico della Pubblica Amministrazione (<http://www.ontologiapa.it/>).
- [23] W3C, Semantic Web Deployment Working Group (<http://www.w3.org/2006/07/SWD/>).
- [24] W3C, Government Linked Data Working Group (<http://www.w3.org/2011/gld/charter>).
- [25] Creative Commons, “Describing Copyright in RDF” (<http://creativecommons.org/ns>).
- [26] Creative Commons Italia (<http://www.creativecommons.it/Licenze>).
- [27] Creative Commons, CC0 1.0 Universal (<http://creativecommons.org/publicdomain/zero/1.0/legalcode>).
- [28] Creative Commons, CC-BY Attribution 2.5 (<http://creativecommons.org/licenses/by/2.5/legalcode>).
- [29] Italian Open Data License v.2.0 (<http://www.dati.gov.it/iodl/2.0/>).
- [30] Specifiche SPAQL 1.0(<http://www.w3.org/TR/rdf-sparql-query/>).
- [31] Microformati (<http://schema.org/>).
- [32] LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO (anno 2014), Agenzia per l’Italia Digitale, Presidenza del Consiglio dei Ministri. Ver 0.60 (bozza consultazione): http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicol2014_v0.6.pdf.

ALLEGATO C

Regione Puglia
Area Politiche per lo Sviluppo Economico,
il Lavoro e l'Innovazione



Regione Puglia

Servizio Ricerca Industriale e Innovazione

Linked Open Data

Linee Guida per la pubblicazione
degli Open Data della Regione
Puglia

TUTORIAL



InnovaPuglia SpA
Strada prov. per
Casamassima, Km. 3

70100 Valenzano –
BARI

Italia

Maggio 2015

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

Note

Alcune parti di questo documento sono tratte o rappresentano rielaborazioni dei seguenti documenti:

- **“LINEE GUIDA PER L’INTEROPERABILITÀ SEMANTICA ATTRAVERSO I LINKED OPEN DATA”** pubblicato dalla **COMMISSIONE DI COORDINAMENTO SPC** dell’Agenzia per l’Italia Digitale con licenza Creative Commons-BY-SA.
- **“LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO (anno 2014)”** pubblicato dall’Agenzia per l’Italia Digitale, Presidenza del Consiglio dei Ministri con licenza Creative Commons-BY-SA. Ver 0.60 (bozza consultazione):
http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblico2014_v0.6.pdf.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

Sommario

1.	Introduzione	
1.	Quadro normativo di riferimento	
2.	Open Data	
2.1	La dimensione economica e l'impatto sul mercato del lavoro	
2.2	Modelli e servizi basati sugli OD	
2.3	Modelli di business per gli OD	
2.3.1	Modello cd. di 'co-work'	
2.3.2	Modello cd. di 'refinement'	
2.3.3	Modello cd. di 'services'	
2.3.4	Quali dati?	
2.4	Organizzazione commerciale per gli OD	
3.	Linked Open Data	
3.1	Dagli Open Data ai Linked Open Data	
3.2	RDF	
3.3	OWL	
3.4	SPARQL	
3.5	Strumenti per la produzione di Linked Data	
3.5.1	Strumenti per la pulizia dei dati	
3.5.2	Strumenti per l'analisi e la modellazione dei dati	
3.5.3	Strumenti per l'arricchimento dei dati	
3.5.4	Strumenti per l'interlinking dei dati	
3.5.5	Strumenti per la pubblicazione dei dati	
3.6	Ontologie e vocabolari	
3.6.1	Ontologie e vocabolari per dati trasversali	
3.6.2	Ontologie e vocabolari per dati verticali	
3.6.3	Ontologie e vocabolari per i metadati	
3.7	Dataset di riferimento	
4.	Pubblicazioni di dati non strutturati	
4.1	Specifiche RSS 2.0	
4.2	Commenti sullo standard RSS	
5.	Licenze d'uso dei dati	
5.1	Analisi delle licenze	
6.	Risorse e link utili	

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

1. Introduzione

Questo documento rappresenta un supporto per le Aree ed i Servizi della Regione Puglia, oltre che per gli Enti Locali interessati, nella pubblicazione dei dati di loro proprietà nella forma di Open Data (OD). Tale attività di pubblicazione e diffusione dei dati con modalità che ne facilitino il riuso è espressamente prevista dal Codice dell'Amministrazione Digitale.

L'Unione Europea attribuisce al riutilizzo delle informazioni del settore pubblico un ruolo fondamentale, sia per lo sviluppo economico e sociale del territorio (corretto funzionamento dei mercati, libera circolazione di merci, di servizi ed individui, miglioramento della competitività, superamento del divario fra nazioni e cittadini), sia per la diffusione delle nuove tecnologie digitali fra enti pubblici, imprese e cittadini, attraverso l'utilizzo di strumenti informatici per diffondere e memorizzare le informazioni medesime.

Al fine di agevolare il riutilizzo delle informazioni in possesso degli enti pubblici degli Stati membri, l'Unione Europea ha adottato la Direttiva 2003/98/CE del 17 novembre 2003 (oltre all'avvio di varie iniziative tra cui il "Libro Verde sull'informazione Pubblica sulla Società dell'informazione", il "Programma e-Content", la Comunicazione "e-Europe 2002").

La Direttiva in oggetto è stata recepita nell'ordinamento italiano con il Decreto Legislativo 24 gennaio 2006 n. 36, "Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico".

La Direttiva in questione attribuisce agli Stati membri - o all'ente pubblico interessato - la decisione di autorizzare il riutilizzo dei documenti che vengono raccolti, prodotti, riprodotti e diffusi nell'ambito del perseguimento dei propri compiti istituzionali.

Tale facoltà viene altresì riconosciuta nell'ordinamento italiano dall'art. 1 comma 2 del Decreto Legislativo 24 gennaio 2006 n. 36, secondo cui: *"La decisione di consentire o meno tale riutilizzo spetta all'amministrazione o all'organismo interessato, salvo diversa previsione di legge o di regolamento"*. La norma precisa, inoltre, che oggetto del "riutilizzo" sono i documenti (rappresentazione di atti, fatti e dati) contenenti dati pubblici, ossia dati conoscibili da chiunque.

Nel seguito sono fornite le informazioni di base in merito a cos'è l'open data e alle differenti tematiche che caratterizzano il percorso di pubblicazione degli OD in un contesto di valorizzazione dei dati pubblici.

Linee Guida Regionali per la pubblicazione di Open Data - Tutorial

1. Quadro normativo di riferimento

Le presenti Linee Guida sono state predisposte nel rispetto di quanto stabilito in materia di “open data” dalla vigente normativa regionale, nazionale ed europea ed in particolare:

- dalla Legge Regionale 24 luglio 2012, n. 20 - Norme sul software libero, accessibilità di dati e documenti e hardware documentato
- dalla Deliberazione della Giunta Regionale 31 Ottobre 2012 n. 2183, – Legge regionale 24 Luglio 2012, n. 20 “Norme sul software libero , accessibilità di dati e documenti e hardware documentato” – Primi indirizzi
- dalla Delibera CIVIT n. 2/2012 - Linee guida per il miglioramento della predisposizione e dell’aggiornamento del Programma triennale per la trasparenza e l’integrità
- dalla Delibera CIVIT n. 105/2010 - Linee guida per la predisposizione del Programma triennale per la trasparenza e l’integrità (articolo 13, comma 6, lettera e, del decreto legislativo 27 ottobre 2009, n. 150) - ALLEGATO – Esempio di struttura della sezione del sito “Trasparenza, valutazione e merito” approvata nella seduta del 14 ottobre 2010
- dal Vademecum Open Data - Come rendere aperti i dati delle pubbliche amministrazioni – Versione Beta 2011
- dalle linee guida per i siti web della PA - art. 4 della Direttiva 8/09 del Ministro per la pubblica amministrazione e l’innovazione
- dal Decreto Legislativo 7 Marzo 2005, n. 82 – CAD: Codice della Pubblica Amministrazione Digitale
- dal Decreto Legislativo 30 dicembre 2010, n. 235 - Modifiche ed integrazioni al decreto legislativo 7 Marzo 2005, n. 82, recante Codice dell’amministrazione digitale, a norma dell’articolo 33 della legge 18 giugno 2009, n. 69.
- dal D. Lgs. n. 163/2006 “Il Codice degli appalti pubblici”;
- dal D. Lgs. n. 36/2006 relativo all’accesso all’informazione nel settore pubblico, attuativo della Direttiva 2003/98 relativa alla “public sector information”;
- dalla Legge n. 96/2010 “Disposizioni per l’adempimento di obblighi derivanti dall’appartenenza dell’Italia alle Comunità Europee. Legge Comunitaria 2009”;
- dalla L. 241/90 “Nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti amministrativi”;
- dal D. Lgs. 196/03 “Codice in materia di protezione dei dati personali”;
- dalla Legge 22 aprile 1941 n. 633 “Protezione del diritto d’autore e di altri diritti connessi al suo esercizio” e s.m.i.;
- dalla Direttiva 2007/2/CE che istituisce un’infrastruttura per l’informazione territoriale nella Comunità europea (INSPIRE) nonché dal Decreto Legislativo 27 gennaio 2010 di recepimento come s.m.i. e dalla “Guidance on the ‘Regulation on access to spatial data sets and services of the Member States by Community institutions and bodies under harmonised conditions”.

2. Open Data

Al giorno d’oggi, esiste una vasta quantità – in continuo aumento – di dati scientifici, industriali, provenienti dalle pubbliche amministrazioni o prodotti da

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

processi di crowd-sourcing che sono disponibili online e pubblicati per un accesso libero (open access). Secondo la visione del pioniere dell'informatica Jim Gray, il quarto paradigma della scienza è proprio l'analisi di dataset di vaste dimensioni. Questo punto di vista viene anche ripreso da Microsoft nel libro "*The Fourth Paradigm – Data-Intensive Discovery*".

Nel prossimo futuro, ci si aspetta che gli Open Data giochino un ruolo catalizzatore nel modo in cui l'informazione strutturata viene sfruttata su larga scala e offrono un enorme potenziale in termini di creazione e fornitura di prodotti e servizi innovativi che creano valore aggiunto rispetto a quello già presente nei dati collezionati. Gli Open Data vengono visti, a ragion veduta, come uno strumento per favorire e promuovere la cittadinanza attiva attraverso, ad esempio, iniziative di giornalismo data-driven, iniziative per la green energy, supply-chain per prodotti a chilometro zero, monitoraggio di emissioni di gas ad effetto serra, mobilità intelligente, etc..

L'attenzione verso gli Open Data è molto cresciuta in seguito all'elezione di Barack Obama alla presidenza degli Stati Uniti. Uno dei primi provvedimenti dell'attuale presidente è stata infatti la Open Government Directive, che ha imposto la pubblicazione sul Web di tutti i dati pubblici delle amministrazioni governative. Sulla base del modello statunitense vari Paesi, tra cui Inghilterra, Australia, Canada, hanno avviato iniziative simili. In Italia la Regione Piemonte si è mossa per prima e sono già realtà le esperienze di Linked Open Camera e Senato e di OpenParlamento¹. Altre regioni italiane, tra cui il Lazio, si stanno attivando con specifici progetti di legge regionali. La Regione Puglia ha pubblicato nel B.U.R.Puglia del 24 luglio 2012 n.109 la legge denominata 'Norme sul software libero, accessibilità di dati e documenti e hardware documentato' che "agisce in coerenza e in continuità con le norme regionali in materia di trasparenza e avvia un processo di diffusione dei dati liberamente accessibili per tutti (Open Data) per una Pubblica Amministrazione aperta ai cittadini in termini di partecipazione al percorso decisionale (open government)."

Il vantaggio più rilevante associato agli Open Data è che gli attori principali sono i dati e non i documenti, cosa che pone le basi per la scoperta di nuova conoscenza, non racchiusa all'interno delle singole fonti, e per la scoperta di soluzioni di problemi che non erano stati originariamente previsti dai creatori dei dati stessi.

La visione di un processo di supporto alle decisioni che sia guidato dai dati è diventato oramai una realtà ben assestata. Nello specifico, gli Open Data pubblicati secondo il paradigma dei Linked Open Data trasformano il Web da un semplice archivio di documenti e servizi in un ecosistema di informazioni in cui gli utenti, sino a ieri visti semplicemente come lettori passivi, diventano aggregatori di dati e creatori essi stessi del cosiddetto Web dei Dati (Web of Data).

Esistono diverse definizioni di cosa significa Open Data. Di seguito ne riportiamo alcune ritenute rilevanti.

¹ <http://parlamento17.openpolis.it/>

Linee Guida Regionali per la pubblicazione di Open Data --Tutorial

- un contenuto o un dato si definisce aperto se chiunque è in grado di utilizzarlo, ri-utilizzarlo e ridistribuirlo, soggetto, al massimo, alla richiesta di attribuzione e condivisione allo stesso modo (def. della Open Knowledge Foundation)
- “Open Government Data” sono dati che 1) “possono essere liberamente usati, riusati e distribuiti da chiunque 2) “sono prodotti o commissionati dalla PA o da entità controllate”
- [...] “i dati devono essere pubblicati come sono stati raccolti alla fonte, con il livello di granularità più fine possibile, non in forme aggregate o modificate”
- [...] “i dati devono essere resi disponibili il prima possibile per preservarne il valore ”.

Anche sulla base delle precedenti definizioni, il gruppo di lavoro su Open Government Data – riunitosi per la prima volta a Sebastopoli (CA) nel 2007 – ha definito i seguenti 8 principi degli Open Government Data che possono essere generalizzati per qualsiasi tipo di Open Data:

1. **I dati devono essere completi.** I dati sono delle informazioni o registrazioni memorizzate in maniera elettronica che includono, ma non sono limitati a: documenti, database, trascrizioni e registrazioni audio/video. I dati pubblici sono dati liberi da vincoli (tecnologici e legali) che ne limitano, di fatto, la più ampia diffusione.
2. **I dati devono essere pubblicati senza nessuna manipolazione precedente.** I dati sono pubblicati e collezionati così come provengono dalla sorgente, con il livello di granularità più fine possibile, e non in forma aggregata e/o modificata.
3. **I dati devono essere tempestivi.** I dati sono resi disponibili il più velocemente possibile per preservare il valore dei dati stessi.
4. **I dati devono essere accessibili.** I dati sono resi disponibili alla più ampia gamma possibile di utenti e per il maggior numero possibile di usi.
5. **I dati devono essere *machine processable*.** I dati sono ragionevolmente strutturati al fine di permettere una elaborazione automatica degli stessi.
6. **L'accesso deve essere non discriminatorio.** I dati sono disponibili per tutti senza nessuna richiesta di registrazione.
7. **I formati dei dati devono essere non-proprietari.** I dati sono disponibili in formati aperti e pubblici sui quali non vi siano entità (aziende e/o organizzazioni) che ne abbiano il controllo esclusivo.
8. **I dati devono essere liberi da licenza.** I dati devono essere caratterizzati da licenze che non ne limitino l'uso, la diffusione o la redistribuzione.

I principi enunciati in precedenza, così come le definizioni date, non pongono vincoli tecnologici sui formati di dati da esporre come dataset Open Data. Per superare questa ambiguità, lo stesso Tim Berners Lee definisce un modello di catalogazione che li classifica in base alle loro caratteristiche su una scala di valori che vanno da 1 (una stella ★) a 5 (cinque stelle ★★★★★).

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

(★) Una Stella. È il livello più semplice. I dati sono pubblicati come file non strutturati: ad esempio un'immagine in formato grezzo (formati come .gif, .jpg, .png), un documento in formato Microsoft Word, un file in formato Adobe Pdf. Una stella indica semplicemente la disponibilità del dato on line distribuito con licenza aperta. Non essendo in formato aperto non è possibile tuttavia effettuare su di essi alcuna elaborazione.

(★★) Due Stelle. Indica dati strutturati e codificati con un formato proprietario. Ad esempio un documento in formato Microsoft Excel. Questa classificazione indica la possibilità di effettuare elaborazioni sui dati, a patto di disporre del software necessario a gestire un file codificato con un formato proprietario. I dati caratterizzati dalle due stelle di norma possono essere convertiti in dati aperti.

(★★★) Tre Stelle. Indica dati strutturati e codificati in un formato non proprietario. In questo caso, il formato maggiormente utilizzato è .csv (Comma Separated Values). Questa classificazione indica la possibilità di effettuare elaborazioni sui dati senza esser costretti ad utilizzare software proprietario. Quello caratterizzato dalle tre stelle è il formato più semplice di dati aperti ed è quello attualmente più diffuso soprattutto presso le Pubbliche Amministrazioni.

(★★★★) Quattro Stelle. Indica dati strutturati e codificati in un formato non proprietario che sono dotati di un URI (Uniform Resource Identifier) che li rende indirizzabili sulla rete e quindi utilizzabili direttamente online, attraverso l'inclusione in una struttura basata sul modello RDF (Resource Description Framework). Questa classificazione indica che il singolo dato di un dataset, disponibile on line in un formato aperto (tipicamente RDF/XML), può essere richiamato attraverso un URL (Uniform Resource Locator) specifico. Questo rende possibile il riferimento al dato o ad un insieme di dati dall'interno di un'applicazione che può poi rielaborarlo.

(★★★★★) Cinque Stelle. È il livello più completo di classificazione che consente una completa interoperabilità tra dataset diversi. Indica quelli che vengono definiti Linked Open Data (LOD); ossia quei dati aperti che oltre ad essere classificabili con 4 stelle presentano anche collegamenti ad altri dataset. Sfruttando il modello di descrizione dei dati RDF si possono collegare tra loro più dataset, incrociando così informazioni provenienti da fonti diverse, eventualmente gestite da diverse Amministrazioni. Tutto questo non fa che aumentare il valore dei dataset reciprocamente correlati. Di fondamentale importanza per la diffusione di Open Data a 5 stelle è il suo puntare sin dall'inizio sulla apertura pubblica dei dataset (Linked Open Data) e la disponibilità di dataset sempre più grandi ed interessanti. In questo senso va certamente osservata la rilevanza – a fungere da volano – di DBpedia, l'equivalente *machine understandable* e interrogabile via SPARQL di Wikipedia e di altri rilevanti dataset disponibili come SPARQL endpoint. Una importante caratteristica della iniziativa Linked Open Data è inoltre il suo crescente riconoscimento di canale preferenziale di divulgazione dei dati in ambito e-government. Il trend si è sviluppato negli USA a partire dalla Open Government Directive della Presidenza degli Stati Uniti del 2009 ha condotto alla creazione del portale <http://www.data.gov/> che mette a disposizione i dati dei

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

governo federale. Esistono peraltro numerosi altri stati europei che già forniscono servizi simili. In Figura 1 viene mostrato lo stato della cosiddetta nuvola dei Linked Open Data (LOD cloud) aggiornato a Settembre 2011.

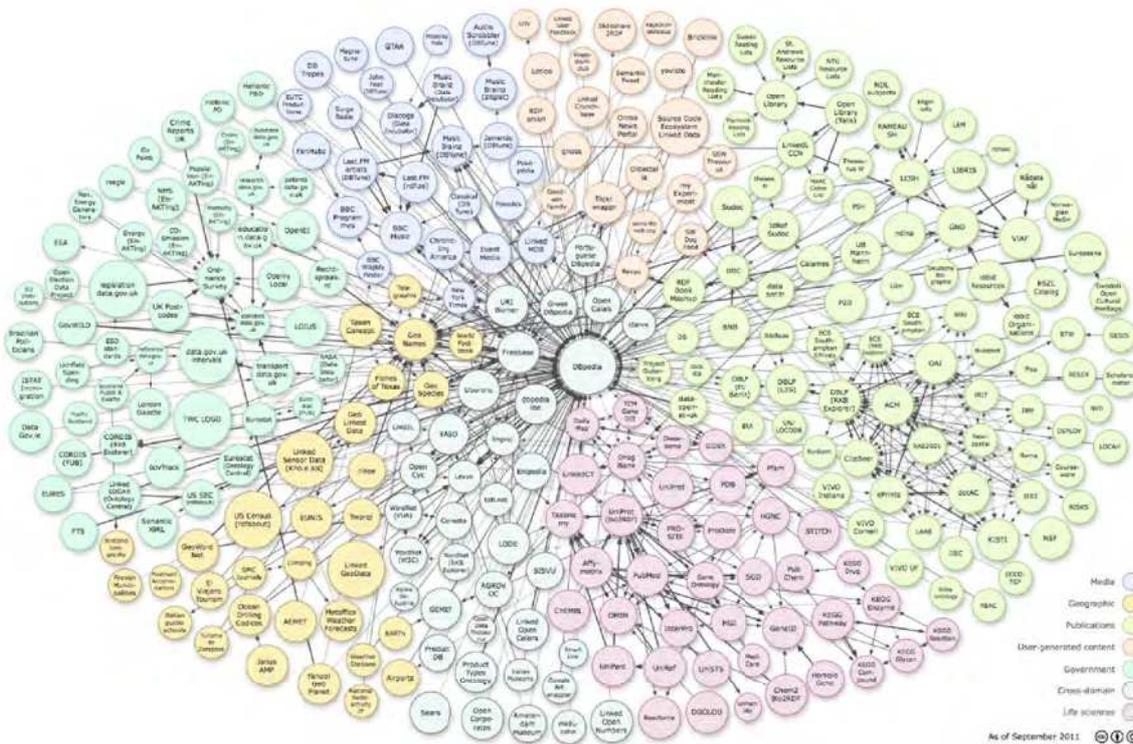


Figura . La Linked Open Data cloud fotografata nel Settembre 2011

2.1 La dimensione economica e l'impatto sul mercato del lavoro

L'Open Data nel nostro paese è da anni oggetto di attenzione da parte di una comunità eterogenea di attivisti provenienti da ambiti tecnologici, di economia e di giurisprudenza². Tale comunità anima iniziative di divulgazione e sensibilizzazione al tema nell'ambito di un più ampio dibattito dell'Open Government, nella comune volontà di innovazione e sulla scorta delle esperienze già intraprese a livello internazionale (il primo portale www.data.gov è americano, segue il britannico www.data.gov.uk, e le analoghe iniziative Francia e Austria). Dall'esperienza americana provengono i primi numeri sulla creazione di servizi a valore aggiunto: negli USA, infatti la disponibilità di informazione del settore pubblico (o Public Sector Information - PSI) ha assunto una non trascurabile importanza macroeconomica (più del 7% del PNL). Ulteriore stimolo proviene dall'Agenda Digitale della Commissione Europea che indica

²International Open Data Day, <http://opendataday.it/>, Open Knowledge Foundation sezione italiana - <http://it.okfn.org/>, Datagov.it - <http://www.datagov.it/>, Spaghetiopendata - <http://www.spaghetiopendata.org/>, DatiOpen.it - <http://www.datiopen.it/>, Linked Open Data - <http://www.linkedopendata.it/>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

esplicitamente l'obiettivo strategico di "mettere a disposizione le informazioni relative al settore pubblico in modo trasparente, efficace e non discriminatorio, come fonte importante di crescita potenziale di servizi online innovativi". La Commissione ritiene, infatti, che una strategia sui dati aperti per l'intera Europa potrebbe dare un contributo all'economia europea quantificabile in 40 miliardi di euro all'anno³.

Su questo tema si stanno, inoltre, consolidando strumenti ed esperienze per passare dalle indagini accademiche e dal dibattito degli ambienti dell'attivismo ad un piano operativo di iniziative, progetti tecnologici, manuali⁴ e collezioni di idee, suggerimenti e best practice che possono aiutare cittadini e imprenditori a "dar vita" ai dati pubblici creando opportunità di crescita economica e in termini di partecipazione democratica (i driver che insieme concorrono alla spinta di innovazione prodotta dall'Open Data sono, infatti, la possibilità di azione sociale e partecipazione democratica, la potenziale maggiore efficienza della pubblica amministrazione e la vera e propria nuova filiera economica e produttiva abilitata dalla possibilità di ri-uso dei dati da parte di privati ed aziende).

L'esperienza già maturata a livello internazionale e negli esperimenti pilota in Italia, indicano poi, che l'approccio più efficiente all'Open Data è quello in cui vi è la collaborazione di un gruppo eterogeneo di professionalità e competenze (dagli esperti di diritto d'autore, di privacy, di diritto amministrativo agli economisti, fino, inevitabilmente, agli ingegneri informatici ed esperti di tecnologie semantiche), perché eterogenea è la materia del riuso dell'informazione del settore pubblico. Facendo proprio tale approccio multidisciplinare, come sottoprodotto dell'interessamento al tema, si ottiene anche un'interlocuzione con i diversi stakeholder di un sistema produttivo (quale per esempio in Regione Puglia quello rappresentato dal Distretto Produttivo dell'Informatica) e la possibilità di raccordo per iniziative congiunte. Inoltre, la sinergica interazione porta ad implementare una comunità di pratica, come indicato dalla regione nella *legge regionale sul software libero*⁵.

Primo passo nell'esplorazione della dimensione economica e di impatto sul mercato del lavoro dell'IT è la consapevolezza che l'Open Data riguarda la messa a disposizione dei dati, perché altri possano riusarli. Open data, quindi, non significa spingere la PA a sviluppare nuovi servizi a partire dai dati. Il servizio che il proprietario dei dati deve sviluppare è "semplicemente" un modo pratico ed efficace per pubblicare i dati in proprio possesso. Da questo concetto emerge la natura di "filiera del dato" e cioè un articolato sistema di estrazione, trasporto, raffinamento e utilizzo come materia prima in nuovi processi produttivi. È necessario individuare e rappresentare i diversi stadi e possibilità di business primario e di indotto di quello che è opportuno concepire come un sistema articolato. Analizzando i processi interni della pubblica amministrazione, si ha che le informazioni possono essere generate direttamente come risultato

³ EUROPEAN COMMISSION, *Digital Agenda: Turning government data into gold*, Press release, 2012, http://europa.eu/rapid/press-release_IP-11-1524_en.htm?locale=en.

⁴ Alcuni esempi sono il manuale dell'*Associazione DataGov.it* o l'*Open Data Handbook* della *Open Knowledge Foundation*

⁵ LEGGE REGIONALE 24 luglio 2012, n. 20 - Art. 17 (BURP - n. 109 del 24-07-2012)

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

primario dell'attività di una PA (es. dati cartografici, catasto, informazioni meteorologiche, ecc.). In altri casi, invece, i dati sono acquisiti, organizzati e impiegati dall'amministrazione nell'adempimento dei propri obiettivi istituzionali, ma rappresentano un mezzo rispetto all'output principale (prodotto o servizio erogato dalla PA). In ogni caso il costo per la produzione di questi dati è già stato sostenuto tramite il costo per il mantenimento della PA che li ha generati. Da ciò l'intelligenza economica in generale, e l'intelligenza industriale di progetti e iniziative di ri-uso: i costi relativi all'apertura dei dati, per quanto importanti, sono solo una minima parte dei costi che altrimenti bisognerebbe sostenere per generarli ex-novo.

I dati generati dalla PA hanno anche una qualità particolare, che manca ai dati raccolti da operatori privati: si tratta di dati raccolti nell'arco di molti anni e sottoposti a verifiche (ad esempio nel corso dell'attività che essi supportano). Sono quindi dotati di uno specifico alto margine di affidabilità. E sono alla fonte privi di interessi commerciali diretti (neutrali).

In ottica di filiera produttiva, è poi possibile guardare all'industria di trasformazione da PSI (semplice dato residente nei sistemi informativi delle diverse PA) a Open Data (dato reso disponibile in formati adatti e mirati al ri-uso⁶, e corredato di licenze/liberatorie che chiariscono i termini di ri-uso e redistribuzione). L'espressione Open Data (dati aperti) o la più specifica open government data (dati pubblici aperti) vengono spesso utilizzate come un sinonimo più accattivante di "informazione pubblica" o PSI. PSI finisce così per essere il termine tecnico usato dai giuristi e dai legislatori europei, mentre Open Data diventa il termine per la comunicazione informale. Per non ingenerare confusione è opportuno che Open Data sia usato come sinonimo di "informazione pubblica liberamente riutilizzabile" (con i vincoli di formato e di licenza suindicati) e non di "informazione pubblica" in generale.

La PSI può essere riutilizzata in molte prospettive:

1. *Nell'ambito di mercati già esistenti o nuovi, per creare servizi a valore aggiunto, con benefici di tipo economico.* Ad esempio, guardando dati meteorologici pubblici, non si pensi solo alle classiche previsioni del tempo, ma anche a servizi specializzati, ad esempio per l'agricoltura o il turismo invernale. In alcuni casi, gli introiti fiscali indiretti generati da tali attività possono essere pari o superiori a quelli che la PA otterrebbe facendo direttamente pagare per l'utilizzo dei dati;
2. *Per contribuire ad un miglioramento delle scelte pubbliche, in un quadro di e-governance e trasparenza.* Ad esempio, il semplice fatto di mettere a disposizione e permettere la rielaborazione di dati quali il numero, il salario o i giorni di assenza degli amministratori e dei dipendenti pubblici può stimolare l'efficienza della PA. In altre parole, condividendo informazioni su sé stessa, la PA è spinta a migliorarsi;

⁶ Formati, standardizzati, interoperabili e machine-readable, al fine di rendere possibile la massima integrazione tra fonti diverse di dati o "silos informativi".

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

3. *Per consentire forme più sofisticate di partecipazione dei cittadini alle scelte pubbliche (e-democracy).* Ad esempio, la combinazione di dati sanitari e ambientali può permettere a gruppi di persone di valutare l'impatto di diverse scelte politiche e può avere ricadute sulla partecipazione dei cittadini alla vita pubblica. Inoltre, la partecipazione della popolazione può anche passare per la creazione di servizi, simili a quelli che potrebbe offrire lo stato o il mercato, ma su base volontaria e/o associativa.
4. *Per la coesistenza ed integrazione di un servizio pubblico e di più servizi privati.* Si pensi ad esempio ai servizi di trasporto e calcolo di percorsi urbani con mezzi pubblici.

Trattandosi poi, di una dinamica tipicamente legata all'informatica, vale la pena di indicare la specifica possibilità di produzione che scaturisce dal fatto che spesso la singola informazione ha un'utilità limitata, ma la raccolta organizzata e la combinazione di più informazioni con fonti differenti può, a sua volta, generare un valore aggiunto notevole.

Una visione di prospettiva sul tema dell'Open Data può anche essere acquisita guardando alle "traiettorie evolutive della PA" tracciate dalla progressiva migrazione da una visione monolitica ad un modello di governance estesa⁷ indotto dalla circostanza storica che oggi le PA faticano a svolgere l'intero ventaglio di mansioni assegnato a loro, a causa della crescente complessità dei sistemi socio-economici, del rapido passo dell'innovazione tecnologica e delle crescenti limitazioni di budget cui sono sottoposte da alcuni anni.

Una progressiva apertura della galassia pubblica diventa pertanto necessaria non solo per incrementare la trasparenza della "macchina governativa" ma, soprattutto, per spianare la strada a nuovi modelli di gestione capaci di combinare in maniera efficace ed efficiente asset pubblici e risorse messe a disposizione dalla società.

La visione monolitica della Pubblica Amministrazione prevede un archetipo di organizzazione pubblica rigida (anche definibile come "Governo 1.0"), che svolge all'interno del suo perimetro un corpus di attività che avvengono sia a livello di front-office che di back-office:

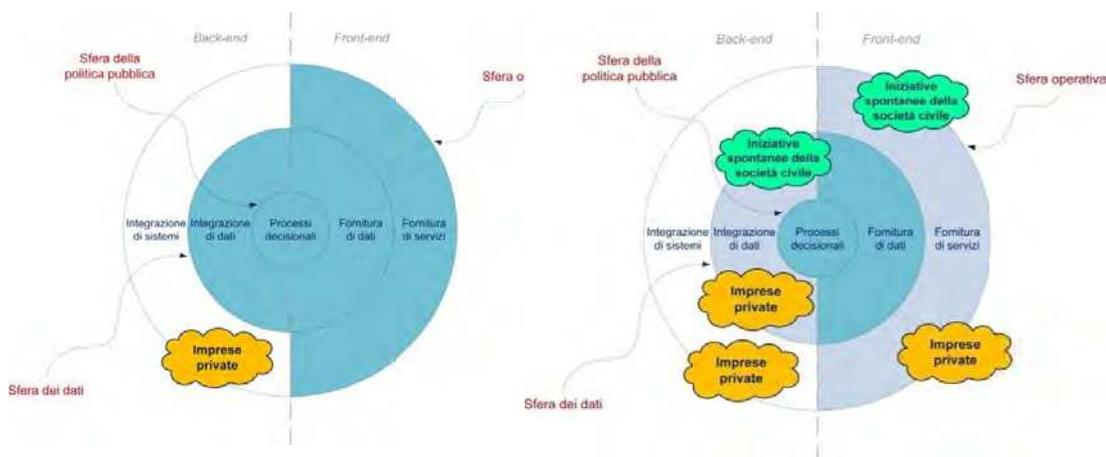
- Formulazione dei processi decisionali, riconducibile sostanzialmente al potere esecutivo ed a quello legislativo.
- Integrazione di dati, che riguarda l'organizzazione, la catalogazione e l'armonizzazione di risorse informative al fine di trasformare i dati operativi in informazione provvista di un senso in funzione del contesto, della cultura, degli obiettivi e del sistema di relazioni in essere.
- Fornitura di dati, che prevede la diffusione dell'informazione prodotta da (o per conto di) organismi pubblici, sia quando la suddetta informazione è consumata direttamente dagli enti pubblici sia quando essa viene liberata "pubblicamente".

⁷Raguseo, E., Ferro, E., *eGovernment and organizational changes: towards an extended governance mode* Lecture Notes in Computer Science, 6846, 418-430, 2011

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

- Fornitura di servizi pubblici, in cui ricadono le prestazioni erogate dagli enti pubblici ai fini del soddisfacimento diretto di fabbisogni collocabili in ambiti in cui l'insufficienza o l'inesistenza del mercato richiedono un intervento pubblico.

[..] Il graduale processo di apertura dei confini propri della Pubblica Amministrazione conduce ad un modello di "governance estesa" ("Governo 2.0") dove si registra un consistente ampliamento del potenziale spazio di intervento per attori esterni alla "macchina pubblica"⁸(Figura 2 - evoluzione della PA).



Visione monolitica della PA

Modello di "governance estesa"

Figura - evoluzione della PA

Nella ricognizione delle ragioni di opportunità legate agli Open Data, è opportuno mettere in luce alcune specificità del territorio pugliese che ne fanno contesto fecondo per l'avvio di iniziative e progetti:

- In generale, da parte delle istituzioni regionali e locali, vi è grande attenzione al tema. In particolare, la Puglia è uno dei territori pilota e cioè una delle 4 regioni italiane ad aver una legge regionale in materia di Open Data⁹.
- Le università pugliesi (in particolare l'Università di Bari e il Politecnico di Bari) presidiano da tempo le materie, le discipline e gli standard tecnologici alla base degli Open Data e dispongono ampio knowhow delle tecnologie necessarie (in primis il Semantic Web) e garantiscono la disponibilità delle competenze tecniche necessarie a sviluppare l'argomento (Figura 3 - W3C Semantic Web Layer Cake: la struttura e le tecnologie del Web semantico).

⁸Ferro E., Osella M., *Modelli di Business nel Riutilizzo dell'Informazione Pubblica*, Studio Esplorativo 2011, Osservatorio ICT – PIEMONTE.

⁹<http://blog.ernestobelisario.eu/2012/03/09/mappa-delle-leggi-regionali-in-materia-di-open-data/>

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

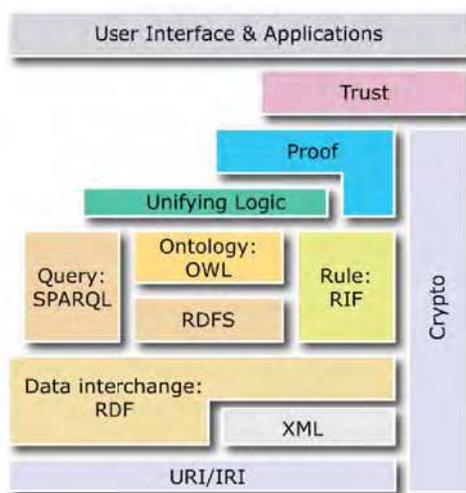


Figura 3 - W3C Semantic Web Layer Cake: la struttura e le tecnologie del Web semantico

- Dal 2012 la Regione Puglia, nell'ambito del P O- FESR 2007-2013 e attraverso il progetto *Apulian ICT Living Labs*, ha messo a disposizione delle PMI pugliesi importanti strumenti finanziari. Grazie a tale iniziativa di investimento e alla successiva intensa attività di divulgazione del paradigma Living Labs in tutto il territorio pugliese, è possibile ritenere tale paradigma noto e conosciuto presso numerose Pubbliche Amministrazioni, operatori economici e componenti della società civile della regione. Come si vedrà nel seguito tale fattore rappresenta un vantaggio competitivo poiché proprio il modello Living Labs si presta ad implementare la necessaria collaborazione tra gli stakeholders della filiera del dato provenienti dai diversi settori della società.

2.2 Modelli e servizi basati sugli OD

L'approccio metodologico che un'organizzazione dovrebbe seguire per rendere aperti i dati di cui dispone è il cosiddetto "approccio per processi", noto anche come ciclo di Deming o modello PDCA (Plan, Do, Check, Act). L'approccio per processi è proprio dei sistemi di gestione della qualità e parte integrante delle norme UNI EN ISO 9001:2000. Il modello PDCA, ideato da W. Edwards Deming in Giappone negli anni '50, è un modello studiato per promuovere una cultura della qualità tesa al miglioramento continuo dei processi ed all'utilizzo ottimale delle risorse perseguite attraverso una costante interazione tra ricerca, progettazione, test, produzione e vendita. In tal senso, consente di orientare processi complessi verso un modello che prevede momenti di analisi, verifica e cambiamento che ne consentono un'ottimizzazione continua.

Applicare il modello PDCA al fine di "aprire" i dati di un'organizzazione significa adattare "la ruota di Deming" come segue (Figura 4):

P - Plan. Pianificazione = *Determinare gli obiettivi + Determinare i metodi e pianificare azioni*

Determinare gli obiettivi: Gli obiettivi della politica di Open Data vanno indicati in modo esplicito, così che sia più semplice procedere nelle fasi successive di identificazione dei data set da aprire.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

Determinare i metodi e pianificare le azioni: Definiti gli obiettivi da raggiungere, è necessario identificare le procedure, razionali e facili da seguire, che consentano di perseguirli. Determinare un metodo vuol dire standardizzarlo e renderlo utile e accessibile; di fatto l'apertura dei dati di un'organizzazione non può essere concepita come un'operazione una tantum, piuttosto va progettata come un'operazione da reiterare in quanto i dati mutano nel tempo e necessitano di essere aggiornati. È opportuno quindi standardizzare le procedure di apertura identificando sin da questa fase i momenti di confronto con le comunità degli utilizzatori e degli sviluppatori, secondo il modello dei Living Labs, per definire assieme le priorità.

D - Do. Esecuzione = *Svolgere il lavoro.* Attuare le procedure identificate al passo precedente. Ciò comporta comprovarne l'efficacia ed identificarne i punti critici al fine di migliorarle progressivamente.

C - Check. Test e controllo, raccolta ed analisi dei risultati e dei riscontri = *Controllare gli effetti dell'apertura dei dati.* Per controllare l'efficacia di un'attività di "apertura" dei dati è necessario monitorarne l'uso; in particolare è fondamentale verificare se le comunità degli utilizzatori e degli sviluppatori, identificate al passo P, facciano uso dei dati, di quali dataset e con quale frequenza. Altri indicatori utili sono: il numero di accessi ai singoli dataset, il numero di servizi sviluppati su di esse, il numero di attori che a vario titolo e con scopi differenti ne fa uso. Tale passo è fondamentale per comprendere se i dataset oggetto dell'azione di "apertura" siano effettivamente utili per le comunità di riferimento ed eventualmente se sia opportuno identificare altri dataset.

A - Act. Intraprendere azioni adeguate = *Ottimizzare il sistema + Far evolvere il sistema.*

Ottimizzare il sistema. Individuazione ed eventuale correzione delle criticità connesse con le fasi precedenti.

Far evolvere il sistema. La strategia per la "apertura" dei dati deve essere concepita come un vero e proprio strumento per la costruzione di un nuovo rapporto, basato sul dialogo e la partecipazione, con le comunità di riferimento per cui deve evolversi sulla base delle indicazioni acquisite nelle fasi precedenti.

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

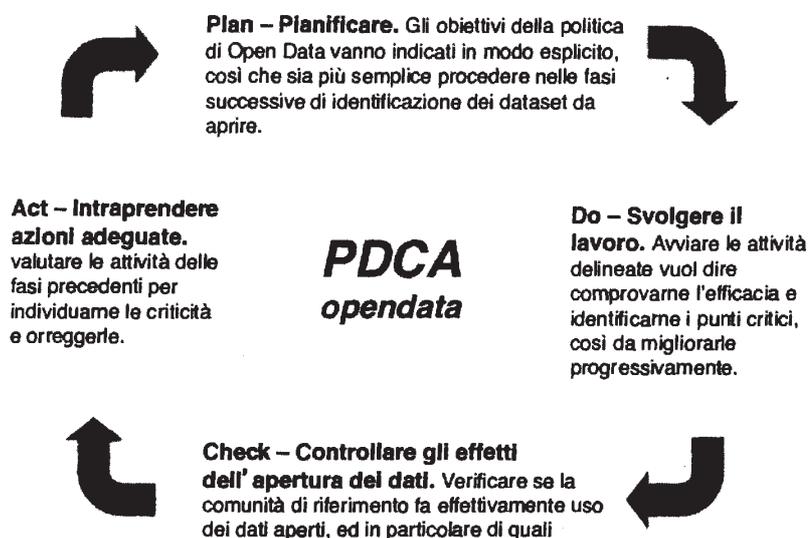


Figura 4. Il modello PDCA applicato all'“apertura” dei dati.

2.3 Modelli di business per gli OD

Al fine di coinvolgere quanto più aziende possibili nella logica dell'openness connessa all'utilizzo degli Open Data, è necessario individuare, preliminarmente, le opportunità che possono presentarsi alle aziende locali (ICT in primis ma non solo), e che guardano al mondo degli Open Data come una nuova possibilità di business.

Partiamo dal presupposto che gli Open Data sono essenzialmente i dati prodotti dalle pubbliche amministrazioni nella loro ordinaria attività amministrative, e che vengono da queste resi pubblici: l'idea che ne è alla base, e che costituisce intima essenza della struttura stessa di un dato aperto, è quella di consentire a chiunque il riutilizzo del dato stesso, per qualsiasi scopo, quindi anche per scopi commerciali.

L'attuale contesto giuridico è fortemente favorevole ad iniziative in questa direzione: con il definitivo varo dell'agenda digitale, le applicazioni, applicate al mondo dei dati aperti diventano il contesto ideale per trasformare delle semplici idee in attività di impresa. Per questo, è opportuno comprendere ed individuare quali possono essere i modelli di business di riferimento, tali da rendere gli Open Data di interesse per le aziende.

2.3.1 Modello cd. di 'co-work'

Un primo possibile modello di business, ed è quello tipico del rapporto tra impresa e PA. In esso, infatti, l'azienda affianca e supporta la pubblica amministrazione, intenzionata - o obbligata ex lege - a rilasciare i propri dati in formato aperto, in tutte le fasi del procedimento che culmina con la pubblicazione dei dati.

Solitamente, il processo viene distinto in cinque fasi:

1. scelta del tipo di dato;
2. scelta della licenza;
3. scelta del formato di output;
4. pulizia del dato e creazione del dataset;

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

5. pubblicazione del dataset.

Questo comporterà, d'altra parte, l'ulteriore necessità di affrontare gli aspetti più prettamente tecnologici legati alla scelta e alla implementazione della piattaforma che servirà per la pubblicazione dei dataset.

Le operazioni descritte possono essere realizzate in vari modi ed a vari livelli di raffinatezza: una iniziativa Open Data, infatti, può essere semplicemente vista come la pubblicazione di un elenco di file scaricabili, ma può anche essere realizzata utilizzando modalità raffinate di modellazione dei dati. In quest'ultimo caso, pertanto, parliamo non solo di dati e metadati strutturati - o eventualmente di dati 'linkati' - ma anche di modalità IT avanzate per accedere direttamente ai dataset in modalità "machine to machine", come Webservices, API RESTful e endpoint SPARQL.

È altrettanto chiaro, però, che non tutte le aziende hanno la possibilità di accedere ad un modello di business così strutturato: in questo caso, infatti, la tipologia dei servizi tende a favorire le stesse aziende già fornitrici di servizi IT alle pubbliche amministrazioni, ossia le stesse che di fatto hanno la possibilità di avviare con la PA un confronto diretto relativo ai servizi di Open Data, di modellazione dei dati, di predisposizione dei relativi applicativi, etc.

2.3.2 Modello cd. di 'refinement'

È il modello caratterizzato dall'inesistenza di rapporti privilegiati con la PA, e da una attività di business che guarda all'intero mercato. In un modello così strutturato, infatti, sono le aziende stesse ad effettuare un investimento per 'armonizzare' i dati pubblici creando una base di conoscenza che contenga non più dati "sparpagliati", ma dati che sono già modellati, integrati tra loro e fruibili da parte di soggetti terzi - cittadini, altre aziende, giornalisti, stessa pubblica amministrazione.

Un siffatto modello di business prevede di lavorare e raffinare i dati pubblicati da diverse pubbliche amministrazioni, facendo data integration e restituendo alla rete e al mercato dei dati consolidati e armonizzati, fruibili anche attraverso modalità avanzate "machine to machine".

2.3.3 Modello cd. di 'services'

Costituisce una terza via al business sui dati, probabilmente anche la via finora più utilizzata, anche se in forme leggermente differenti dalla presente teorizzazione.

In tale modello, infatti, l'azienda non solo interviene andando a recuperare i dati dalle varie pubbliche amministrazioni armonizzandoli, integrandoli e consolidandoli, ma costruisce con essi dei servizi, che possono essere concepiti secondo un preciso modello di business e che possano essere collocati sul mercato stesso o anche "rivenduti" alla pubblica amministrazione.

Possiamo, in questo caso, immaginare servizi erogati attraverso applicazioni Web o applicazioni mobili - si vedano, ad esempio, i prodotti realizzati dalle società che hanno partecipato ai vari contesti banditi finora - : in tali casi, infatti, è l'azienda che realizza un prodotto con i dati pubblici, effettua un investimento per poter creare valore aggiunto sui dati, immaginando (o realizzando), in tal modo, dei servizi che possono essere generati a partire da quei dati pub

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

collocarli sul mercato direttamente o tramite la pubblica amministrazione di riferimento.

2.3.4 Quali dati?

Uno degli interrogativi più frequenti e importanti riguarda la tipologia dei dati che possono essere utilizzati per la realizzazione dei modelli di business che abbiamo descritto.

L'indagine sull'utilizzo degli stessi ha finora evidenziato come i dati a maggiore potenzialità siano i dati ambientali, i dati relativi ai trasporti e, comunque, tutti quei dati che provengono in tempo reale o in pseudo tempo reale.

Immaginiamo, ad esempio, l'utilizzo che si potrebbe fare dei dati delle linee degli autobus urbani del trasporto pubblico se georeferenziati, o dei dati delle centraline per la rilevazione delle polveri sottili, e così via.

Finora, anche per predisposizione di un favorevole contesto normativo, è sempre stata la PA a decidere quali dati rendere pubblici, salvo poi sondare le reazioni del mercato. Un passo avanti è stato compiuto attraverso l'attività di 'mappatura dei bisogni', effettuata da alcune PA – ma soprattutto dalle community – per individuare le specifiche esigenze della collettività sui dati.

La strada, per le aziende pugliesi, potrebbe essere proprio questa: coinvolgere le aziende al fine di censire i potenziali progetti, le idee e quindi, indirettamente, identificare le tipologie di dati necessari per sviluppare dei progetti di investimento su di essi.

2.4 Organizzazione commerciale per gli OD

L'esplorazione del livello commerciale, nel caso specifico dell'Open Data è possibile solo in parte poiché al termine del "processo estrattivo" le possibilità di business abilitate dalla disponibilità di Open Data riguardano nuovi prodotti e nuovi servizi non necessariamente legati al medesimo modello commerciale e per i quali gli OD rappresentano una "materia prima".

Dal punto di vista dei processi estrattivi, è già stato evidenziato come esistano modelli di business di interesse primario per aziende già fornitrici di servizi IT alle pubbliche amministrazioni. Tali aziende sono le stesse che "di fatto hanno la possibilità di avviare con la PA un confronto diretto relativo ai servizi di open data, di modellazione dei dati, di predisposizione dei relativi applicativi, etc". L'apertura di prospettive commerciali legate alla produzione di open data a partire dalle informazioni del settore pubblico (PSI), oltre che incentivare l'attenzione di tali aziende al tema (anche in termini di acquisizione di knowhow specifici quali, ad esempio, quelli sulle tecnologie semantiche), può avere un importante effetto di catalizzazione della disponibilità di "materia prima" per i successivi processi di ri-uso. Iniziative isolate presso singole PA non sono infatti sufficienti a generare una massa critica di open data tale da poter avviare iniziative commerciali di ri-uso.

A livello generale, per lo sviluppo del tema nel suo complesso al fine di porre le basi di iniziative di "completamento" della filiera del dato, l'approccio proposto parte da una ricognizione della filiera basata su settori, e a partire dai potenziali stadi finali della filiera stessa e legati al ri-uso dei dati. In altre parole a partire dal "bisogno di dati" in ottica di ri-uso.

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

In quest'ottica, settori di interesse sono identificati dall'ambito nel quale determinate Pubbliche Amministrazioni forniscono servizi (es. sanità, trasporto, turismo, ...) e, nel contempo, individuano potenziali ambiti entro i quali collocare i prodotti derivati dal riuso dei dati generati da quelle PA.

In questo ambito la Regione Puglia favorisce un'azione coordinata e top-down a livello regionale, che punti a creare dei tavoli di lavoro per settore (sanità, turismo, trasporto, ...) con tutti gli attori del territorio (stakeholders), pubblici e privati, allo scopo di:

- Dare la definizione "user-centered" dei bisogni di Open Data specifici del settore;
- Creare degli standard di pubblicazione in merito;
- Individuare i vincoli per le PA e le aziende interessate, sia per quanto già esistente che per le nuove gare da pubblicare.

Dai tavoli di settore, eventualmente modellabili come diverse "Innovation Arena" secondo una metodologia Living Lab, potrebbero essere create delle aggregazioni con fini commerciali, per Settore, per tutte le attività della Filiera degli Open Data. Nella metafora della filiera del petrolio, queste filiere di settore potranno organizzarsi secondo i diversi settori, per le varie fasi di estrazione, trasporto, trasformazione, servizio, e portare nuovo business e servizi non solo alle Pubbliche Amministrazioni ma anche a privati (B2C) o aziende (B2B). Si potranno quindi costituire queste aggregazioni di Settore, sostenute da una progettualità di ricerca e sviluppo pluriennale, per sviluppare i nuovi business legati all'open data.

3. Linked Open Data

I Linked Data definiscono un metodo per la pubblicazione di dati strutturati in modo tale che essi possano essere inter-connessi tra loro al fine di aumentarne l'utilità. Le specifiche Linked Data sono costruite su tecnologie Web ormai consolidate: il protocollo HTTP, il linguaggio RDF e le URIs. I Linked Data utilizzano le tecnologie Web non per produrre delle pagine fruibili da essere umani, ma estendono quest'ultime in modo tale che possano essere interpretate automaticamente da un calcolatore. Questo permette di connettere e interrogare dati provenienti da diverse sorgenti.

Ritornando alle tecnologie alla base dei linked data possiamo elencare i seguenti principi:

1. Tutti gli oggetti presenti nei dati sono identificati da URIs. L'URI (Uniform Resource Identifier) è una sequenza di caratteri che identifica in maniera univoca una risorsa sul Web. Le URI non vanno confuse con le URL o le URN. Le URN identificano in maniera univoca il nome di una risorsa, invece una URL identifica anche il protocollo da utilizzare per recuperare la risorsa identifica dal nome. Ad esempio gli indirizzi delle pagine Web che noi quotidianamente utilizziamo sono delle URL con il prefisso http: che sta ad identificare il protocollo da utilizzare (http) per recuperare la risorsa. Sia le URN che le URL sono classificabili come URI;
2. Utilizzare il protocollo http per recuperare le risorse (dati). Il protocollo http permette di deferenziare il nome di un oggetto e recuperarlo dal Web;

Linee Guida Regionali per la pubblicazione di Open Data --Tutorial

3. Utilizzare RDF per creare collegamenti tra URI. L'RDF permette di mettere in relazione URI e quindi i relativi oggetti denotati dalle URI;
4. Utilizzare SPARQL per l'interrogazione dei dati RDF.

Queste tecnologie verranno ampiamente discusse nelle successive sotto-sezioni.

3.1 *Dagli Open Data ai Linked Open Data*

Gli Open Data permettono la pubblicazione di dati in formato aperto riutilizzabili per la produzione di nuovi servizi. Quello che gli Open Data non definiscono è come sia possibile rendere questi dati interoperabili, ovvero come mettere in contatto dati provenienti da diverse sorgenti ma logicamente connessi tra di loro.

Ad esempio, supponiamo che un'amministrazione metta a disposizione i dati relativi al trasporto pubblico come le fermate e i percorsi dei mezzi di trasporto pubblico di una determinata città. Supponiamo che un altro ente metta a disposizione un elenco del patrimonio artistico della stessa città con le informazioni relative alla localizzazione dei vari beni.

Quello che potremmo fare e voler collegare, ad esempio, le varie fermate dei mezzi pubblici con i luoghi da visitare in modo da realizzare un servizio utile ai turisti. Tutto ciò implica un'operazione di integrazione dei dati per poter ad esempio produrre una nuova base dati utilizzabile dal servizio che vogliamo realizzare.

Tutto sarebbe più semplice se le due risorse fossero costruite su un vocabolario comune e tutti i dati fossero identificati in maniera univoca da una URI. Ad esempio, prendiamo un ipotetico indirizzo "Via Verdi, 14, Bari", se questo indirizzo fosse identificato da una URI univoca ad esempio <http://comune.bari.it/via-verdi-14> potremmo utilizzare questa URI per descrivere lo stesso dato nei due diversi dataset. Questo già ci permetterebbe senza alcun costo di integrazione aggiuntivo di recuperare l'indirizzo di una fermata dal dataset dei trasporti e cercare con la stessa URI nel dataset dei beni culturali. In questo caso l'utilizzo di un vocabolario comune ci ha permesso a costo zero l'interoperabilità tra due set di dati differenti. Nelle sezioni successive vedremo come utilizzare e definire vocabolari di risorse comuni e come mettere in relazione i dati (le URI che li identificano) attraverso il linguaggio RDF.

3.2 *RDF*

RDF (Resource Description Framework) è un modello per la descrizione delle risorse ed è alla base per rappresentare dati e metadati nel Semantic Web. RDF è il modello alla base per la costruzione dei Linked Data.

Il modello RDF nasce dalla necessità di rappresentare i metadati, ovvero dati che descrivono altri dati. L'utilizzo dei metadati nasce dal bisogno di organizzare sistemi complessi come le biblioteche. Ad esempio per accedere in maniera efficiente ad un enorme catalogo di pubblicazioni sono stati definiti una serie di metadati come: l'autore, il genere, il titolo, la data di pubblicazione. Tutti questi dati sono appunto dei metadati che noi utilizziamo per descrivere il dato "libro" a cui i metadati fanno riferimento. Proprio grazie alla realizzazione di un sistema di strutturazione e standardizzazione delle descrizioni bibliografiche che nasce il progetto Dublin Core (DC). Il DC prevede la definizione di un insieme

elementi che rappresentano i metadati per descrivere opportunamente risorse in rete. Il DC prevede l'utilizzo di quindici elementi¹⁰ che possono essere utilizzati per aggiungere metadati. Ad esempio, il concetto di autore è identificato dalla classe *Creator* che ha URI <http://purl.org/dc/elements/1.1/creator>. Ciò sta ad indicare che ogni volta che vogliamo aggiungere ad un dato il metadato di autore possiamo utilizzare la URI del concetto *Creator* per identificarlo. Questo ci permette di avere un vocabolario condiviso per la definizione di metadati, in questo modo non sarà necessario ridefinire il concetto di autore in ogni dataset ma si potrà far riferimento a quello stabilito dal DC.

Maggiori dettagli sul linguaggio RDF sono riportati nell'Allegato Tecnico.

3.3 OWL

OWL (*Web Ontology Language*) è un linguaggio di markup progettato per definire e istanziare *ontologie Web*. Una *ontologia OWL* può includere le descrizioni delle *classi*, *proprietà* e delle loro istanze. Dato questo tipo di Ontologia, la *semantica formale* di OWL specifica come derivare le sue conseguenze logiche, ovvero i fatti che non sono presenti letteralmente nell'ontologia, ma *derivati logicamente* dalla semantica. Tali derivazioni logiche possono essere basate su un solo documento o su più documenti distribuiti che sono stati combinati fra loro usando dei meccanismi OWL predefiniti.

Maggiori dettagli sul linguaggio OWL sono riportati nell'Allegato Tecnico.

3.4 SPARQL

SPARQL è un linguaggio di interrogazione di dati RDF divenuto standard W3C nel 2008, nel 2013 è stata standardizzata anche la versione 1.1 che aggiunge alcune caratteristiche alla precedente versione.

Come detto in precedenza RDF è sostanzialmente un grafo diretto ed etichettato per rappresentare in formazioni sul Web. SPARQL può essere utilizzato per interrogare dati RDF provenienti da sorgenti differenti e sostanzialmente è un linguaggio per imporre pattern sui i grafi e vincoli sui valori. Il risultato di una query SPARQL può essere un insieme di risorse o valori, oppure un grafo RDF.

Maggiori dettagli sul linguaggio SPARQL sono riportati nell'Allegato Tecnico.

3.5 Strumenti per la produzione di Linked Data

Esiste un'ampia gamma di tecnologie utilizzabili per la produzione di Open Data e Linked Data. La moltiplicazione di queste tecnologie è direttamente riconducibile ai numerosi standard che, in tempi recenti, si stanno affermando. A seconda degli input al processo di generazione dei LOD, degli output che da esso si desiderano ottenere e delle risorse disponibili, le PA e i vari produttori di dati sono chiamati a svolgere operazioni diverse mediante ricorso a strumenti e tecnologie, a loro volta, differenti. Questa sezione illustra le principali tecnologie e strumenti oggi disponibili, suddividendoli in macro-categorie corrispondenti alle fasi proposte nell'approccio metodologico della sezione precedente.

¹⁰<http://dublincore.org/>

3.5.1 Strumenti per la pulizia dei dati

E' necessario che i dati siano conformi a dei criteri e che rispettino alcuni dei vincoli di qualità e integrità. Questa fase di pulizia ("data cleaning") può essere affrontata con varie tecnologie e strumenti, alcuni dei quali riportati nel seguito.

ETL

Questi strumenti permettono di trattare fonti dati eterogenee, fare operazioni di estrazione, normalizzazione, denormalizzazione, riconciliazione e "data cleansing" in maniera semplice, intuitiva e scalabile. Un processo di ETL, una volta sviluppato, può diventare di fatto il componente che si occupa dell'aggiornamento del dataset pubblicato. Mentre in passato queste operazioni e questi strumenti erano impiegati prevalentemente per la popolazione di data warehouse, oggi, vista la loro flessibilità d'uso, si qualificano anche come strumenti per la messa a sistema di operazioni semplici e ripetitive (nella fattispecie l'aggiornamento di dataset) in un ambiente facilmente mantenibile. Esempi di strumenti ETL Open Source sono Kettle¹¹ e Talend Open Studio¹². Entrambi molto utilizzati anche in ambiti enterprise, sono ricchi di molte componenti. Kettle è molto intuitivo e facile da usare. La conferma della facilità di utilizzo si nota immediatamente nel modo in cui vengono realizzati i processi di ETL e di trasformazione. Rispetto a Talend, riduce i passaggi in cui è necessaria la generazione di codice Java per effettuare la trasformazione. La piattaforma Talend, di contro, risulta più aperta e orientata al mondo Java.

SISTEMI DI GESTIONE DEI DATI

Anche i Sistemi di gestione dei dati (DBMS), possono essere utilizzati per compiere operazioni di pulizia di dati, specialmente quando i dati da aprire risiedono proprio all'interno di un DBMS. I database relazionali open source più noti sono PostgreSQL e MySQL. Ultimamente, per contrastare la crescita esponenziale della quantità di dati e la loro multidimensionalità e incompletezza, si stanno sempre più affermando i DBMS non-relazionali (NoSQL).

GOOGLE REFINE¹³ è uno strumento open source per la pulizia, l'analisi e l'elaborazione di dati. Esso può operare su diversi tipi di dati in ingresso e offre un piccolo supporto alla funzione di "record linkage" grazie a funzionalità di espansione semantica.

3.5.2 Strumenti per l'analisi e la modellazione dei dati

In questa fase, è utile individuare strumenti che facilitino la reingegnerizzazione del dataset secondo il nuovo modello logico e concettuale su cui si basano i linked data.

PROTEGE¹⁴ è una soluzione open-source scritta in Java sviluppata dall'Università di Stanford (e in seguito con l'aiuto dell'Università di Manchester) per la definizione e manutenzione di ontologie (schemi, vocabolari) in

¹¹<http://community.pentaho.com/projects/data-integration/>

¹²<http://www.talend.com/products/talend-open-studio>

¹³<http://code.google.com/p/google-refine/>

¹⁴<http://protege.stanford.edu/>

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

RDF/OWL. È un editor di ontologie la cui architettura è basata su plug-in estendibili. Ha un'interfaccia grafica intuitiva ed è supportato da una nutrita comunità. L'attenzione è posta su OWL, quindi i Linked Data in RDF puro hanno spesso qualche problema a essere gestiti in maniera adeguata. Il limite principale di Protégé è la mancanza di un plugin per utilizzare SPARQL, dovuta appunto all'approccio "OWLcentric". Protégé implementa un buon debugger logico per le ontologie.

NeOn Toolkit¹⁵ è un'altra soluzione open-source in Java, sviluppata dal progetto europeo NeOn, per la definizione e manutenzione di ontologie (schemi, vocabolari) in RDF/OWL. Anche questo editor di ontologie è basato su plug-in. Diversamente da Protégé, è pienamente compatibile con RDF e gestisce SPARQL, associandolo ai ragionatori automatici per OWL. Essendo molto più recente, la comunità di supporto è limitata. NeOn Toolkit implementa un buon debugger logico, ma anche un debugger basato su "best practices" di modellazione. Attraverso i suoi plug-in, ci sono anche funzionalità di progettazione avanzata che utilizzano i cosiddetti "ontology design patterns", e funzionalità semplici e intuitive di visualizzazione.

D2R (Database to RDF)¹⁶ è un framework che permette l'accesso al contenuto di un database relazionale come se si trattasse di un dataset RDF. Tramite un linguaggio dichiarativo è possibile definire clausole che, in modo flessibile, descrivono le relazioni tra le ontologie di riferimento e la struttura tabellare della base di dati. L'applicazione delle clausole consente, quindi, la produzione di un dataset RDF (anche chiamato "RDF dump") secondo le specifiche desiderate. Il suo limite principale è la mancanza di una semplice e intuitiva interfaccia utente.

Oltre agli strumenti per la modellazione di ontologie risultano di particolare interesse il riuso di vocabolari ed ontologie già esistenti. Il riuso di basi di conoscenza già esistenti e riconosciute a livello internazionale permette di realizzare in maniera automatica l'interlinking dei dati, in quanto riutilizzando concetti già presenti e utilizzati in altre basi di conoscenza sarà possibile collegare varie sorgenti.

SKOS¹⁷ definisce e fornisce degli standard per la rappresentazione della conoscenza in organizzazioni complesse attraverso l'utilizzo di RDF. In pratica, definisce una serie di concetti e relazioni (predicati) che possono essere riutilizzati per descrivere sistemi organizzativi complessi. La rappresentazione RDF garantisce l'interoperabilità tra diverse sorgenti informative che utilizzano le risorse fornite da SKOS.

dbpedia¹⁸ potremmo definire dbpedia la rappresentazione RDF della conoscenza racchiusa all'interno della più grande enciclopedia esistente, ovvero Wikipedia. L'idea alla base di dbpedia è quella di estrarre automaticamente informazioni da

¹⁵<http://neon-toolkit.org>

¹⁶<http://d2rq.org/>

¹⁷<http://www.w3.org/2004/02/skos/>

¹⁸<http://dbpedia.org/>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

Wikipedia e di strutturarle in RDF. Per fare ciò, dbpedia sfrutta la conoscenza contenuta nelle tabelle di Wikipedia che sono definite utilizzando particolari template che permettono facilmente l'estrazione automatica dei dati. La conoscenza contenuta di dbpedia è quella di senso comune e quindi non settoriale, questo ha permesso a dbpedia di diventare l'hub della Linked Open Data (LOD) cloud in quanto altre sorgenti possono riutilizzare e linkare i concetti di senso in essa contenuti. L'utilizzo delle risorse definite in dbpedia consente automaticamente di linkare il proprio dataset ad una miriade di dati già presenti nella LOD cloud.

3.5.3 Strumenti per l'arricchimento dei dati

Le tecnologie che rientrano in questa categoria permettono di "verificare", "inferire", o "popolare" automaticamente i dati. La verifica automatica di dati e ontologie serve a controllare che non siano presenti incoerenze nello schema dei dati (per esempio che qualcosa possa essere sia una città sia una persona, se si è indicato esplicitamente nello schema OWL che la classe Città è disgiunta dalla classe Persona), né inconsistenze nei dati stessi (per esempio che i dati contengano triple per cui una certa entità è sia una città sia una persona). La verifica è importante per l'arricchimento perché serve a evitare che si introducano automaticamente affermazioni non valide, e serve a garantire le prestazioni di un ragionatore automatico, che non funzionerebbe, o sarebbe limitato, in caso di incoerenze o inconsistenze. Il ragionamento automatico su dati e ontologie permette l'arricchimento mediante inferenze logiche; in pratica, mediante la materializzazione della conoscenza logicamente implicita negli schemi e nei dati esistenti. La materializzazione permette di costruire il cosiddetto "modello inferito" dei dati, cioè l'insieme di triple che si possono dedurre logicamente dalla struttura dei dati esistenti. Il ragionamento automatico è tecnicamente limitato a dati consistenti e a schemi coerenti: per questomotivo gli strumenti di ragionamento automatico svolgono anche il ruolo di "consistencychecker". I migliori ragionatori automatici sono, in teoria, quelli che permettono inferenze complete su dati e schemi (in particolare su ontologie in OWL), ossia quelli che non trascurano nulla. A questo insieme appartengono attualmente Hermit¹⁹ dell'Università di Oxford, dotato di licenza aperta LGPL, FacT++²⁰ dell'Università di Manchester (anche questo aperto, ma come GPL), Pellet²¹ di Clark&Parsia, che ha una licenza aperta accademica, ma commerciale per le aziende, RacerPro²² di Racer Systems, che ha solo licenze commerciali. Esistono poi ragionatori automatici che lavorano solo su sottoinsiemi del linguaggio OWL per ottimizzare l'efficienza computazionale. Oracle 11g25 (commerciale), che lavora su un sottoinsieme di OWL più efficientemente computabile ed è integrato con il DBMS di Oracle. Ci sono poi ragionatori automatici che si possono definire incompleti, che si preoccupano più

¹⁹<http://hermit-reasoner.com/>

²⁰<http://owl.man.ac.uk/factplusplus/>

²¹<http://clarkparsia.com/pellet/>

²²<http://racer.sts.tuhh.de>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

dell'efficienza che della completezza logica e della verifica di consistenza e coerenza. Questo tipo di approccio comprende molte diverse soluzioni ed è raccomandabile soprattutto quando non si lavora con OWL, ma solo con dati RDF(S). Tra questi si può menzionare OWLIM²³. Il popolamento automatico di dati e ontologie è infine un tipo di arricchimento basato su tecnologie di estrazione della conoscenza. In questo caso, la conoscenza implicita che si vuole inferire è quella linguistica e l'accuratezza non è quasi mai del 100%. Questi strumenti sono quindi utilissimi quando i dati contengono molto testo, o quando è importante far emergere la conoscenza "sepolta" dentro documenti testuali, accettando in compenso un certo margine di imprecisione. Le tecnologie disponibili per l'arricchimento automatico sono moltissime ed è difficile darne una valutazione obiettiva. Si va dai riconoscitori di entità (cioè nomi propri come "Mario Rossi" o "Berlin") ai risolutori di entità, che permettono di dare un'identità specifica alle entità riconosciute (per esempio <http://dbpedia.org/resource/Berlin>), agli estrattori di relazioni (per esempio "BerlinisTheCapitalCityOfGermany"). Esistono molti strumenti, sia open source (e.g., GATE²⁴ dell'Università di Sheffield, DBpediaSpotlight²⁵ dell'Università di Lipsia, Text2Onto²⁶ dell'Università di Mannheim, ReVerb²⁷ dell'Università di Washington) sia commerciali (e.g., Alchemy²⁸), che implementano queste funzionalità.

3.5.4 Strumenti per l'interlinking dei dati

Come è ovvio immaginare, il linking è una funzionalità molto importante per i Linked Data e di fatto può essere considerata una forma particolare di arricchimento. La particolarità consiste nel fatto che l'arricchimento avviene grazie all'interlinking fra dataset di origine diversa, tipicamente fra amministrazioni o istituzioni diverse, ma anche, al limite, all'interno di una stessa amministrazione. SILK²⁹ è uno degli strumenti più utilizzati per fare "record linkage". È in grado di scoprire somiglianze (similarity) fra entità appartenenti a diverse sorgenti e materializzare i collegamenti in nuovi dataset RDF.

3.5.5 Strumenti per la pubblicazione dei dati

Probabilmente la più vasta categoria di tecnologie di LOD è quella sul processo di pubblicazione e accesso ai dati. Al fine di permettere il riuso agile dei LOD, è

²³<https://www.ontotext.com/owlim>

²⁴<https://gate.ac.uk/>

²⁵<https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki>

²⁶<http://code.google.com/p/text2onto/>

²⁷<http://reverb.cs.washington.edu/>

²⁸<http://www.alchemyapi.com/>

²⁹<https://www.assembla.com/spaces/silk/wiki/Home>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

opportuno esporre uno SPARQL endpoint. In generale, è importante evidenziare come tali sistemi, vista anche la loro recente adozione, siano ancora affetti da problematiche relative alla scalabilità in presenza di grandi dataset sui quali eseguire un elevato numero di interrogazioni. Pur se tale circostanza è indice di successo, essa pone evidentemente problemi prestazionali da governare. A tal riguardo, opportuni meccanismi forniti dai sistemi descritti possono essere sfruttati per contrastare le suddette problematiche. Per esempio, si può efficacemente limitare il numero di risultati da visualizzare in una singola interrogazione attraverso l'uso di clausole dell'linguaggio di interrogazione, oppure, in maniera più sofisticata, si possono configurare i sistemi in base all'uso specifico che l'amministrazione ne deve fare (e.g., uso di indici, tecniche di caching, ecc.). In ogni caso, qualora i predetti accorgimenti non risultassero sufficienti, è necessario agire a livello di risorse, dispiegando i sistemi in maniera potenziata e più flessibile così da bilanciare il carico di lavoro atteso. Di seguito si riportano alcuni dei sistemi attualmente utilizzati per il processo di pubblicazione e accesso ai dati.

Openlink Virtuoso Universal Server è un knowledgestore in quanto offre funzionalità di RDBMS, ORDBMS, XML Database, RDF Store, Web Service Server. È utilizzato principalmente come RDF Store poiché rappresenta una soluzione completa e offre un supporto unitario in grado di gestire quasi tutte le principali questioni relative alla gestione dei Linked Data. Raggiunge buone performance grazie a una rappresentazione fisica dei dati RDF in un "QuadStore" e consente di esporre e configurare molto semplicemente uno SPARQL endpoint. Il sistema è disponibile in diverse versioni: commerciale, "cloud" (modalità PaaS) e ridotta rilasciata sotto licenza Open Source.

D2RQ SERVER la piattaforma D2RQ già sopra introdotta offre, oltre alla capacità di trasformazione "bulk" del contenuto di una base di dati relazionale in RDF, anche la capacità di gestire varie modalità di accesso ai dati, fornendone la navigazione Web (servizio di deferenza) e mediante hyper-data browser. In particolare, offre anche un punto di accesso SPARQL. Tale risultato è possibile attraverso uno strato intermedio di interpretazione delle richieste in ingresso, che sfrutta un "mapping" definito nel linguaggio D2RQ. Esso è quindi in grado di trasformare un'interrogazione SPARQL in un'interrogazione SQL tramite un confezionamento dell'informazione che maschera la natura relazionale dello strato fisico sottostante. Sia che la pubblicazione dei dati avvenga mettendo a disposizione i file contenenti le triple del dataset sia che essa sia realizzata scegliendo di non mettere a disposizione dei dati interoperabili, occorre predisporre un portale o sito Web. Si può far riferimento a una serie di opzioni tecnologiche che facilitano questo tipo di pubblicazione.

PORTALE CKAN Il sistema CKAN, sviluppato da Open Knowledge Foundation, è un prodotto nato per la catalogazione di risorse aventi la natura di file accessibili tramite URL. Si tratta di una piattaforma onnicomprensiva, ben integrata e altamente personalizzabile, con cui si possono realizzare tutti gli elementi di un sistema di gestione di Open Data, dalla loro memorizzazione

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

fisica, organizzazione logica, metadattazione e, infine, esposizione su un sito Web. Essa copre i seguenti aspetti del processo di pubblicazione di un dataset:

- sistema di redazione delle schede dei metadati;
- storicizzazione automatica;
- sistema di memorizzazione dei file;
- pubblicazione delle schede all'interno di un portale Web personalizzabile;
- funzioni di ricerca per chiave;
- sistema di esposizione dei dati come servizio mediante API di tipo REST;
- funzioni di anteprima per i tipi di formati più comuni;
- funzioni basilari di analisi statistica degli accessi;
- sistema minimale di qualityassurance e feedback dagli utenti.

Il linguaggio con cui è stato programmato CKAN è Python, che per la realizzazione dei servizi sopra elencati si integra con un database PostgreSQL e un motore di ricerca Solr su application server Jetty (un Apache Web Server). Grazie alla potenza di queste API, quindi, è possibile integrare le funzionalità del sistema di catalogazione all'interno di altri software. Ad esempio, è possibile inserire qualunque funzione di CKAN all'interno di un CMS moderno, svincolando in tal modo la fase di pubblicazione da quella di redazione.

CMS Molti dei portali esistenti sono basati, per la realizzazione delle funzioni di gestione di contenuti, su CMS Open Source. Questi sistemi sono utili qualora i fruitori dei dati siano solamente utenti umani. I tre principali e più utilizzati prodotti CMS sono Wordpress, Drupal e Joomla. Sono strumenti sufficientemente flessibili e personalizzabili da adeguarsi allo scopo di pubblicazione di (Linked) Open Data. In particolare in Drupal sono presenti in maniera nativa le tecnologie legate al Web Semantico, come RDF e RDFa; inoltre esso possiede un modulo per CKAN che integra le funzionalità del catalogo, di un motore di ricerca e di wiki.

Geoportali Con il termine geoportale si intende un sito Web realizzato in modo tale da costituire un punto di accesso unico (gateway) ai servizi relativi a dati e risorse spaziali, che non devono necessariamente risiedere all'interno del sito stesso ma che possono invece essere distribuiti. Soggetti, siano essi pubblici, privati o generiche comunità di utenti che realizzano un geoportale, permettono l'accesso a informazioni territoriali per mezzo di un'interfaccia Web e attraverso l'utilizzo di web services. Dall'altro lato, gli utenti Web che hanno la necessità di utilizzare informazioni territoriali, siano essi utenti generici, professionisti o pubbliche amministrazioni, utilizzano i geoportali per ricercare, accedere e utilizzare i dati. I geoportali possono rappresentare inoltre un elemento fondamentale delle infrastrutture di dati spaziali (SDI - Spatial Data Infrastructure), di cui l'esempio più significativo a livello nazionale e europeo è la SDI definita e regolata dalla direttiva INSPIRE (Direttiva 2007/2/CE). Per garantire la ricerca e l'accesso alle informazioni territoriali, un geoportale offre all'utente una serie di servizi che, pur essendo spesso fruibili tramite interfaccia grafica per mezzo di un comune browser Web, devono essere implementati, all'interno di un'architettura di tipo Service Oriented (SOA), come veri e propri web-services realizzati in modo conforme a una serie di standard che ne

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

garantiscono l'interoperabilità. Per quanto riguarda il livello nazionale ed europeo la direttiva INSPIRE prevede che i servizi implementati siano resi disponibili attraverso il geoportale comunitario ed eventualmente attraverso punti di accesso nazionali. I servizi previsti dalla Direttiva sono:

1. discoveryservices, che permettono la ricerca dei dati territoriali e relativi servizi attraverso i
2. metadati e di visualizzare i metadati stessi;
3. viewservices, che consentono la visualizzazione dei dati territoriali;
4. download services, che consentono lo scaricamento dei dati;
5. transformationservices, che consentono la trasformazione dei dati allo scopo di conseguire
6. l'interoperabilità;
7. invokeservices, che consentono di richiamare altri servizi sui dati territoriali.

Per l'implementazione di tali servizi, sono state predisposte apposite linee guida tecniche INSPIRE che fanno riferimento ad una serie di standard internazionali già esistenti; in particolare, per quanto riguarda il contenuto dei metadati, agli standard ISO della serie 19100 e a quelli OGC (Open Geospatial Consortium), per quanto riguarda i dettagli tecnici e architetturali. Allo stato attuale, INSPIRE ha predisposto e reso disponibili le linee guida relative ai primi due servizi; per gli altri, invece, le linee guida sono ancora in versione di bozza. Per quanto riguarda i servizi di discovery (1), i servizi di view (2) e i servizi di download (3), gli standard cui fare riferimento sono quelli del OGC, relativamente ai CS-W (Catalogue Services for the Web), ai WMS (Web Map Services) e ai WFS (Web Feature Services).

Attualmente, sono disponibili diverse soluzioni software applicative sia open-source che proprietarie, che permettono di pubblicare all'interno di un geoportale servizi web conformi agli standard OGC (CSW, WMS e WFS). Per quanto riguarda il panorama open-source, le soluzioni software più utilizzate per la realizzazione di geoportali, sia in ambito pubblico che privato, possono essere suddivise in tre categorie principali:

- applicazioni server che implementano e consentono la pubblicazione di web services per la
- visualizzazione e il download di dati territoriali;
- applicazioni server che implementano e consentono la pubblicazione di web services per la
- gestione dei metadati relativi ai dati territoriali;
- librerie software API per la fruizione tramite browser web di dati territoriali.

Tra le applicazioni open-source per la pubblicazione e il download di dati territoriali (a) si citano Geoserver, UMN Mapserver e Degree, tra quelli per la gestione dei metadati (b) Geonetwork, Degree e Geoportal, mentre per quanto riguarda le API per la fruizione di dati territoriali (c) OpenLayers. Un'ulteriore modalità di pubblicazione dei dati è quella dell'opendata-as-a-service dove

Linee Guida Regionali per la pubblicazione di Open Data --Tutorial

anziché mettere a punto una piattaforma in-house si sfruttano piattaforme di cloud (pubblico o privato).

SOCRATA³⁰ è una soluzione proprietaria Open Data as a service composta da: un sistema semantico di archiviazione dell'informazione; un'interfaccia web dinamica, per l'accesso ai dati mediante maschere parametriche; una Socrata Open data API, per esporre interfacce applicative; una serie di strumenti di indagine statistica e di visualizzazione grafica con cui realizzare semplici attività di data mining; un sistema di social networking integrato, con cui gestire il feedback degli utenti; un sistema integrato di metadattazione e classificazione dell'informazione. Offre la possibilità di caricare dataset su di un sistema esterno, e di utilizzare una serie di funzionalità avanzate che permettono ad un ente anche sprovvisto di un proprio asset IT interno di avere esposti i propri Open Data.

OpenLab - già OGDI (Open Government Data Initiative)³¹: è un progetto di Microsoft Corp. Nato per fornire ad enti governativi e pubbliche amministrazioni un servizio opendata-as-a-service. I dati sono messi a disposizione mediante interfacce standard (API), pensate espressamente per la realizzazione di applicazioni Web. La tecnologia di memorizzazione dei dati è la piattaforma Microsoft Azure, ovvero un sistema cloud orientato alle applicazioni, già integrato con il framework di sviluppo .NET. Su questa base si innestano inoltre una serie di strumenti che realizzano un sistema integrato di interfacce per l'accesso e la gestione dei dati presenti in Azure. OpenLab è gratuito e contiene componenti Open Source. Come tutte le soluzioni di hosting, i vantaggi di OpenLab stanno soprattutto nella quantità di strumenti e di interfacce già a disposizione, e nella totale mancanza di carico per quanto riguarda la parte di archiviazione e gestione dei dataset. Analogamente alla piattaforma Socrata, con OpenLab è necessario comunque aggiungere l'impegno di risorse umane interne per l'associazione della componente semantica ai dataset caricati.

Esistono dei casi in cui i dati sono forniti attraverso API, servizi Web oppure attraverso sistemi di ricerca e di navigazione non convenzionali.

WEB SERVICES: nel caso in cui si volesse sviluppare un sistema informativo interno che si basi sugli Open Data si può pensare anche alle classiche tecnologie di "remote call". In questo caso il consumatore degli Open Data è rappresentato da un programma o da un sistema informativo, anziché una persona, e, l'aspetto cruciale passa quindi dalla presentazione dei dati alle funzionalità che possono essere richiamate e ai loro metodi di invocazione. In questo campo la tecnologia più diffusa è rappresentata dai web service, che si dividono sostanzialmente in due tipologie:

- **WS SOAP:** usa il protocollo HTTP solo per il trasporto dell'informazione e ricostruisce completamente lo strato di servizio dedicato alla comunicazione e l'interoperabilità tra soggetti;

³⁰<https://opendata.socrata.com/>

³¹<http://www.msopenlab.com/>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

- WS REST: poggia esclusivamente sui servizi del protocollo http per eseguire operazioni su un insieme di dati che viene opportunamente elaborato e restituito a seconda della URL chiamata, come se già esistesse in quella forma a quell'indirizzo. Nel caso in cui i dati sono fruiti come servizi, è importante che i primi siano restituiti in un formato facilmente elaborabile da un'applicazione software.
- API (Application Program Interface): una via di interrogazione percorribile è quella resa possibile dalle API, interfacce realizzate tramite applicazioni che permettono la comunicazione con altre applicazioni. Di solito invisibili all'utente, consentono l'interazione di applicazioni Web o mobile, con un determinato servizio o con repository di dati.

3.6 *Ontologie e vocabolari*

Questa sezione presenta una prima selezione di ontologie e vocabolari che possono essere utilizzati per descrivere alcune tipologie di dati di tipo aperto (una lista piuttosto ricca di ontologie e vocabolari è comunque messa a disposizione dall'OKFN³²). L'uso consolidato di tali ontologie e vocabolari ha l'obiettivo di far sì che, all'interno della pubblica amministrazione, dataset differenti ma di una stessa tipologia abbiano uno schema comune. In questo caso l'integrazione di dati e il riuso di software risultano enormemente facilitati in quanto, ad esempio, le stesse interrogazioni possono essere effettuate su dataset differenti (quindi potenzialmente è possibile usare uno stesso software) e dataset differenti possono essere inseriti all'interno di uno stesso database senza operazioni di conversione intermedie. Si noti che le stesse ontologie vengono fornite in modalità Open Data. Seguendo quanto indicato nell'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico, si distinguono dati trasversali ovvero comuni a tutte le amministrazioni, centrali e locali, (e.g., dati amministrativi) e dati verticali gestiti da o per conto di specifiche amministrazioni riferibili a determinati settori di competenza (e.g., dati sulla medicina, dati sulla cultura). Alla luce di tale distinzione, le seguenti sotto sezioni si propongono di individuare alcune ontologie e vocabolari utilizzabili rispettivamente per la descrizione di dati comuni a tutte le amministrazioni e per la descrizione di domini di interesse specifici che sono identificati per macro aree tematiche.

3.6.1 *Ontologie e vocabolari per dati trasversali*

Esistono ontologie e vocabolari che riguardano dati di interesse generale e che quindi possono essere utilizzati in diversi contesti e da amministrazioni di diverso tipo. Di seguito si descrivono brevemente tali ontologie e vocabolari, scelti sulla base della loro diffusione, del livello di standardizzazione e del contesto in cui si sono sviluppati (come nel caso dei vocabolari "core"³³ definiti nell'ambito di specifici programmi della Commissione Europea nati anche al fine di implementare determinate azioni previste dall'Agenda Digitale Europea).

SKOS (Simple Knowledge Organization System)³⁴. E' un'ontologia che consente di rappresentare vocabolari controllati, tassonomie e tesauri. E' nato per facilitare la pubblicazione e l'uso di vocabolari nel contesto dei Linked Data ed è stato standardizzato dal W3C.

GOODRELATIONS³⁵. E' un'ontologia che consente di rappresentare dati relativi al commercio quali prodotti, prezzi, negozi, ecc. I dati rappresentati con Goodrelations possono anche essere incorporati all'interno delle pagine Web per

³²<http://lov.okfn.org/dataset/lov/>

³³ European Commission, "e-Government Core Vocabularies",

https://joinup.ec.europa.eu/community/core_vocabularies/description, 2014.

³⁴ W3C, SKOS Simple Knowledge Organization System - Home Page, <http://www.w3.org/2004/02/skos/>, 2014.

³⁵ GoodRelations Language Reference, <http://www.heppnetz.de/ontologies/goodrelations/v1.html>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

essere letti dai classici motori di ricerca e da altri software. Tale ontologia può essere utilizzata per descrivere i servizi acquisiti tramite contratti dalle pubbliche amministrazioni.

PUBLIC CONTRACTS ONTOLOGY³⁶. E' un'ontologia che consente di descrivere dati RDF relativi ai contratti pubblici (e.g., contraente, data di validità) e all'intero processo di aggiudicazione di una gara d'appalto (e.g., tipologia di gara, massimale, partecipanti, aggiudicatari, ecc.). Tale ontologia si presta quindi a essere utilizzata nel caso, per esempio, di apertura di parte della base di dati di interesse nazionale sui contratti pubblici.

PAYMENT ONTOLOGY³⁷. E' un'ontologia sviluppata per rappresentare informazioni di spesa relative ad organizzazioni (non necessariamente pubbliche amministrazioni). Il nucleo dell'ontologia è rappresentato dal concetto di Pagamento che è associato a un soggetto che paga (e.g., un'autorità locale che deve effettuare un pagamento), uno che riceve (e.g., il fornitore che riceve il pagamento eseguito dell'erogazione di una prestazione) e la data. Il pagamento può inoltre ulteriormente descritto attraverso una serie di proprietà opzionali che includono, la specifica unità organizzativa che eroga il pagamento, l'insieme di prodotti e/o servizi oggetto del pagamento e eventuali link a risorse che documentano il processo. Tale ontologia si presta quindi a essere utilizzata nel caso di rappresentazione di informazioni che riguardano le spese e i pagamenti delle pubbliche amministrazioni per esempio.

FOAF (Friend of a Friend)³⁸. E' un'ontologia che consente di rappresentare dati riguardanti le persone e le loro relazioni. Essa può quindi essere utilizzata per la rappresentazione di alcuni dati inclusi in basi di dati di interesse nazionale che riguardano le persone quali "Anagrafe della Popolazione Residente – ANPR", l'Anagrafe Nazionale dell'Assistiti (ANA), cataloghi della piattaforma delle comunità intelligenti.

WAI (WhoAm I)³⁹. E' un'ontologia che estende FOAF con i concetti di ruolo e profilo che possono essere associati alle persone.

SIOC (Semantically Interlinked Online Communities)⁴⁰. E' un'ontologia che consente di rappresentare informazioni riguardanti le comunità online (ad esempio, blog, wiki, forum, mailing list, ecc.) e di relazionarle tra loro.

CPSV (Core Public Service Vocabulary)⁴¹. E' un modello dati per la rappresentazione dei servizi pubblici offerti dalle pubbliche amministrazioni. Il suo scopo è quello di facilitare (i) lo scambio di informazioni sui servizi nel settore pubblico, (ii) la scoperta di servizi esistenti e delle normative che li riguardano, e (iii) lo sviluppo dei portali della pubblica amministrazione. Inoltre,

³⁶ Public Contracts Ontology, <https://code.google.com/p/public-contracts-ontology/>, 2014.

³⁷ Guide to the Payments Ontology, <http://data.gov.uk/resources/payments>, 2014.

³⁸ FOAF Vocabulary Specification, <http://xmlns.com/foaf/spec/>, 2014.

³⁹ Who Am I! Vocabulary Specification, <http://vocab.ctic.es/wai/wai.html>, 2014.

⁴⁰ SIOC Core Ontology Specification, <http://rdfs.org/sioc/spec/>, 2014.

⁴¹ European Commission, "Core public service vocabulary", 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

grazie all'uso del CPSV è possibile confrontare servizi simili erogati da differenti organizzazioni. Di tale vocabolario esiste la versione per i Linked Data. Infine, è importante sottolineare che esso fa parte dei vocabolari "core" della commissione europea standardizzati o in fase di standardizzazione da parte del W3C, recentemente utilizzato in attività pilota nel contesto del progetto europeo SPOCS -Building the next generation Points of Single Contact⁴². E' stato altresì proposto al Comitato Tecnico delle comunità intelligenti per la rappresentazione delle informazioni relative al catalogo dei servizi della piattaforma nazionale delle comunità intelligenti.

CLV (Core Location Vocabulary)⁴³. E' un modello dati per la rappresentazione delle locazioni. Esso quindi consente di descrivere indirizzi, struttura geometrica, nome geografico, ecc. E' conforme alle specifiche INSPIRE ed esiste una versione ufficiale che può essere utilizzata con i Linked Data. Anch'esso, come nel caso precedente, fa parte dei vocabolari "core" della commissione europea standardizzati o in fase di standardizzazione da parte del W3C.

CPV (Core Person Vocabulary)⁴⁴. E' un modello dati per la rappresentazione delle informazioni sulle persone. Nasce per essere ben integrato con gli altri vocabolari "core" della commissione europea e fa largo riuso di schema.org e FOAF. Come nel caso di FOAF, tale vocabolario può essere utilizzato per rappresentare alcune informazioni presenti nella base di dati di interesse nazionale "Anagrafe della Popolazione Residente - ANPR" e "Anagrafe Nazionale Assistiti - ANA". Attualmente il vocabolario è utilizzato all'interno del progetto europeo e-CODEX - eJusticeCommunication via Online Data eXchange.

ORG (Organization Ontology)⁴⁵. E' un'ontologia che consente di rappresentare strutture organizzative, in particolare di pubbliche amministrazioni. Essendo rivolta al mondo dei Linked Data, ben si presta a estensioni e integrazioni con altre ontologie in modo da considerare classificazioni delle organizzazioni, l'appartenenza del personale con i relativi ruoli. Fa parte dei vocabolari "core" della commissione europea standardizzati o in fase di standardizzazione da parte del W3C ed è ampiamente utilizzata nella versione Linked Open Data dell'Indice nazionale della Pubblica Amministrazione italiana (IPA). Il vocabolario è multilingua ed è stato recentemente tradotto anche in italiano dall'Agenzia per l'Italia Digitale.

RegOrg⁴⁶. E' un vocabolario, in corso di standardizzazione W3C, che estende la precedente ontologia ORG per la specifica delle organizzazioni che sono

⁴² SPOCS project, <http://www.eu-spocs.eu/>, 2014.

⁴³ European Commission, Core Location Vocabulary, https://joinup.ec.europa.eu/asset/core_location/description, 2014.

⁴⁴ European Commission, Core Person Vocabulary, http://joinup.ec.europa.eu/asset/core_person/description, 2013.

⁴⁵ W3C, "Candidate Recommendation. The Organization Ontology", <http://www.w3.org/TR/vocab-org/>, 2014.

⁴⁶ W3C, "Working Group Note. Registered Organization Vocabulary", <https://dvcs.w3.org/hg/gld/raw-file/default/legal/index.html>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

ufficialmente riconosciute attraverso l'iscrizione a registri nazionali o regionali. Ad esempio, nel caso italiano potrebbe agevolmente essere usato come vocabolario base per la rappresentazione dei dati della base dati di interesse nazionale "Registro delle imprese" (art. 60 del CAD).

OWL-S (Semantic Markup for Web Service)⁴⁷. E' un'ontologia che consente di descrivere i servizi Web e il loro ciclo di vita. Questo consente di trovare, invocare e comporre servizi Web in modo automatico attraverso il Web Semantico. Si compone di tre parti principali: profile per la profilazione base e il discovery dei servizi, il process model per la descrizione delle operazioni del servizio e, infine, il grounding per le informazioni di interoperabilità e scambio di messaggi con altri servizi. Si noti che molte delle specifiche di OWL-S corrispondono a quelle del linguaggio WSDL.

DOAP (Description Of A Project)⁴⁸. E' un'ontologia che consente di descrivere progetti. Nata principalmente per descrivere progetti di produzione del software, esiste anche una versione pronta per essere utilizzata in ambito Linked Data.

VCARD ONTOLOGY⁴⁹. E' un'ontologia che nasce dal rispettivo standard IETF⁵⁰(RFC 2426)per rappresentare i biglietti da visita elettronici. Consente quindi di rappresentare informazioni supersone e le loro informazioni di affiliazione.

EUROVOC⁵¹. E' un tesoro multilingua (22 lingue) contenente la terminologia usata nell'ambito delle attività dell'unione europea ovvero il parlamento europeo, gli uffici delle pubblicazioni ufficiali dell'unione europea, ecc. Eurovoc è reso disponibile anche con una versione ufficiale in RDF che fa uso di SKOS.

3.6.2 Ontologie e vocabolari per dati verticali

Le seguenti ontologie e vocabolari sono riferibili ad ambiti specifici. Tali ambiti sono stati suddivisi per acro-aree tematiche scelte principalmente sulla base di alcuni dei settori di competenza delle amministrazioni coinvolte nel rilascio di dataset chiave individuati nell'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico.

Dati territoriali e ambientali

INSPIRE (Infrastructure for Spatial Information in Europe)⁵². Nasce come direttiva della commissione europea per mettere a disposizione in modo armonico i dati geo-spaziali e i servizi associati tramite geo-portali e altri punti di

⁴⁷ W3C, "OWL-S: Semantic Markup for Web Services", <http://www.w3.org/Submission/OWL-S/>, 2014.

⁴⁸ Description of a Project, <https://github.com/edumbill/doap/wiki>, 2014.

⁴⁹ W3C, "vCard Ontology", <http://www.w3.org/TR/vcard-rdf/>, 2014.

⁵⁰ <http://www.ietf.org/rfc/rfc2426.txt>

⁵¹ Eurovoc, <http://eurovoc.europa.eu/>, 2014.

⁵² European Commission, "INSPIRE – Infrastructure for Spatial Information in the European Community", <http://inspire.jrc.ec.europa.eu/>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

accesso. Di fatto, definisce un'ontologia nata proprio con l'obiettivo di migliorare l'interoperabilità tra i sistemi geo-spaziali.

AGROVOC⁵³. E' un vocabolario controllato creato dalla FAO per rappresentare i concetti delle sue aree di interesse (cibo, nutrizione, agricoltura, ambiente, ecc.). Esiste una versione SKOS già pronta per essere utilizzata in ambito Linked Data.

GEMET (General Multilingual Environmental Thesaurus)⁵⁴. E' un dizionario terminologico multilingue realizzato a livello europeo dall'agenzia europea per l'ambiente "EEA" che si avvale della rete "EIONET" per rendere fruibile il contenuto dei documenti relativi alle attività ambientali. GEMET nasce con la funzione principale di strumento di gestione informativa del Centro Tematico Europeo per il Catalogo delle Fonti Dati) e ha origine dalla fusione del contenuto dei migliori tesauri realizzati in ambito europeo al fine di ottenere una base condivisa per la descrizione delle informazioni. Si tratta di un database terminologico specificatamente orientato ai temi del territorio e dell'ambiente. È dotato di struttura semantica interna ed è multilingue (ad oggi 33 lingue). Le tavole di definizione (e i dati) sono scaricabili: <http://www.eu-spocs.eu/>.

ThIST (Thesaurus Italiano di Scienze della Terra)⁵⁵. Il tesoro è il risultato di una pressoché totale revisione del tesoro nato dall'integrazione tra la banca dati terminologica dell'allora Biblioteca del Servizio geologico d'Italia (oggi Biblioteca APAT) e il tesoro pubblicato dal CNR nel '1997. È strettamente connesso al Multilingual Thesaurus of Geosciences (MTG), curato dal Gruppo di lavoro internazionale MTG, all'interno del quale l'Italia è rappresentata dall'APAT. Viene già utilizzato per l'indicizzazione del materiale bibliografico (monografie, articoli di periodici, carte) posseduto dalla Biblioteca e il recupero delle informazioni relative. Allo stato attuale il ThIST presenta 10.350 descrittori, 14 collegati tra loro da 103.850 relazioni di tipo gerarchico, associativo e di equivalenza. Ad ogni descrittore è affiancata la corrispondente traduzione in inglese così da rendere il ThIST un tesoro bilingue. I descrittori sono generalmente sostantivi o sintagmi con valore di sostantivo; gli aggettivi, i verbi e gli avverbi sono esclusi dal corpo del tesoro, in ottemperanza alla norma ISO 278815. Si è preferita la forma singolare, a eccezione di alcuni casi specifici in cui è utilizzato il plurale. Al contrario, per i descrittori in lingua inglese si è utilizzata la forma plurale in conformità al modello GeoRef TH, da cui molti di essi provengono. Quanto alla struttura, si è mantenuta al momento quella del tesoro CNR di provenienza, basata su una classificazione di tipo tematico-disciplinare comprendente 56 categorie.

Dati statistici

⁵³ FAO, Agrovoc, <http://aims.fao.org/standards/agrovoc/about>, 2014.

⁵⁴ EIONET, GEMET Thesaurus, <http://www.eionet.europa.eu/gemet/index.html?>, 2014.

⁵⁵ ISPRA, ThIST Thesaurus, <http://www.isprambiente.gov.it/contentfiles/00003600/3687-thist.pdf/view>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data -Tutorial

RDF Data Cube⁵⁶. L'esigenza di rappresentare i dati multidimensionali sul Web ha portato alla proposta da parte del W3C di RDF Data Cube Vocabulary (RDF QB). RDF QB prevede un modello multidimensionale e consente dunque di rappresentare "cubi" (data cube) in termini di misure e dimensioni. Il modello dati di RDF QB è basato su SDMX, ma sfrutta le caratteristiche di RDF, consentendo, dunque, di introdurre i dati multidimensionali nel Web semantico ed eventualmente di collegarli anche a dati elementari.

XKOS (eXtended Knowledge Organization System)⁵⁷. E' un'estensione di SKOS per gestire i requisiti specifici per la pubblicazione, da parte degli Istituti Nazionali di Statistica, dei dati della produzione di statistica ufficiale. Ad esempio, XKOS introduce i "livelli" delle classificazioni statistiche: tali livelli, identificando i livelli di aggregazione nelle statistiche pubblicate, sono fondamentali sia per garantire la consistenza dei dati prodotti sia per controllare i rischi di violazione della riservatezza.

Dati culturali

Tesaurus PICO⁵⁸. E' il vocabolario controllato utilizzato nel portale della cultura italiana²² ma è sufficientemente generale per essere usato nella descrizione di dati relativi al dominio dei dati culturali. **Tesaurus UNESCO**⁵⁹. E' un tesaurus per i documenti e le pubblicazioni sulle attività dell'UNESCO. In particolare contiene i termini usati negli ambiti dell'istruzione, cultura, scienze naturali, scienze sociali, comunicazione e informazione. Ne esiste una versione modellata attraverso SKOS.

CIDOC-CRM⁶⁰. E' un modello dati di riferimento per rappresentare le informazioni su concetti e relazioni nell'ambito della cultura (e.g., documenti di musei, archivi, biblioteche). Esiste già una versione in OWL.

EDM (Europeana Data Model)⁶¹. E' il modello dati (utilizzabile anche per dati in RDF) utilizzato all'interno del progetto europeo Europeana che ha lo scopo di favorire la condivisione, l'interoperabilità e l'accesso alle informazioni delle opere culturali.

Dati legali e contabili

⁵⁶ W3C, RDF Data Cube, <http://www.w3.org/TR/vocab-data-cube/>, 2014.

⁵⁷ UNICE, EUROSTAT, OECD, "Extended Knowledge Organization System (XKOS)", <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP10.pdf>, 2014.

⁵⁸ Thesaurus PICO, http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml, 2014.

⁵⁹ UNESCO Thesaurus, <http://databases.unesco.org/thesaurus/>, 2014.

⁶⁰ The CIDOC Conceptual Reference Model, http://www.cidoc-crm.org/official_release_cidoc.html, 2014.

⁶¹ Europeana Data Model (EDM) Documentation, <http://pro.europeana.eu/edm-documentation>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

AkomaNtoso⁶². E' un linguaggio basato su XML per la rappresentazione di documenti giuridici. Attualmente è in fase di approvazione presso il consorzio OASIS. È utilizzato dal Parlamento Europeo e dalla Commissione Europea come standard documentale per i documenti legislativi, giuridici e allegati tecnici.

XBRL (eXtensible Business Reporting Language)⁶³. E' un linguaggio basato su XML per la rappresentazione e lo scambio di informazioni contabili e finanziarie.

3.6.3 Ontologie e vocabolari per i metadati

Questa sezione descrive brevemente una selezione di ontologie e vocabolari utilizzati per metadattare cataloghi, dataset e dati di interesse.

VoID (Vocabulary of Interlinked Datasets)⁶⁴. E' un vocabolario che consente di rappresentare metadati sui dataset RDF, in particolare, per caratterizzare le informazioni riguardanti le connessioni tra diversi dataset, come ad esempio i linkset.

DCAT (Data Catalog)⁶⁵. E' un vocabolario standardizzato dal W3C utilizzato per metadattare icataloghi di dati, i dataset contenuti in un catalogo e le diverse modalità di reperimento e accesso ai dati. E' stato impiegato per caratterizzare i profili applicativi dei portali di dati (DCAT-AP).Attualmente è il vocabolario utilizzato dal catalogo nazionale dati.gov.it.

Dublin Core⁶⁶. La versione "core" consiste di un vocabolario minimo di metadati (15 elementi) per descrivere generiche risorse. Tra questi vi sono, ad esempio, autore, data e titolo. Tale insieme minimo è stato esteso con un insieme più ampio di elementi definiti dalla iniziativa DCMI (Dublin Core Metadata Initiative).

ADMS (Asset Description Metadata Schema)⁶⁷. E' un modello utilizzato per rappresentare asset di interoperabilità semantica. Possibili asset sono ad esempio le classificazioni, i modelli, gli standard, ecc. che possono essere utilizzati per abilitare l'interoperabilità semantica. Questa ontologia ha l'obiettivo di favorire la scoperta e la federazione di diversi "repository" di asset semantici.

PROV⁶⁸. E' un framework standardizzato dal W3C per la rappresentazione della provenienza di artefatti. Consente di modellare anche il processo di generazione di un artefatto in maniera quasi analoga ai sistemi di controllo versione.

CREATIVE COMMONS⁶⁹. E' un sistema di licenze per le opere soggette a diritto d'autore. E' possibile utilizzare una versione RDF del vocabolario soprattutto quando si gestiscono dati di tipo Linked.

⁶²AkomaNtoso, "XML for parliamentary, legislative & judiciary documents", <http://www.akomantoso.org/>, 2014.

⁶³ XBRL, <http://www.xbrl.org/>, 2014.

⁶⁴W3C Interest Group Note, "Describing Linked Datasets with the VoID Vocabulary", <http://www.w3.org/TR/void/>, 2014.

⁶⁵ W3C, Data Catalog Vocabulary (DCAT), <http://www.w3.org/TR/vocab-dcat/>, 2014.

⁶⁶ Dublin Core Metadata Element Set, Version 1.1, <http://dublincore.org/documents/dces/>, 2014.

⁶⁷ W3C, Asset Description Metadata Schema (ADMS), <http://www.w3.org/ns/adms>, 2014.

⁶⁸W3C Working Group Note, "PROV-Overview An Overview of the PROV Family of Documents", <http://www.w3.org/TR/prov-overview/>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

SCHEMA.ORG⁷⁰. E' un insieme di vocabolari che contengono i metadati che possono essere usati all'interno delle pagine Web per strutturare meglio le informazioni presenti nelle pagine. Questi servono per facilitare la comprensione degli elementi delle pagine da parte dei "crawler" dei motori di ricerca.

VANN (Vocabulary for ANNotating vocabulary descriptions)⁷¹. E' un vocabolario che consente di rappresentare metadati e annotazioni su vocabolari. In particolare, esso è utilizzato per arricchirli con esempi e note di utilizzo.

VOAF (Vocabulary of a Friend)⁷². E' un vocabolario che consente di descrivere i vocabolari usati coi Linked Data. Ad esempio, esso consente di mettere in relazione diversi vocabolari o di indicare quali dataset ne fanno uso.

3.7 Dataset di riferimento

Questa sezione introduce alcuni dataset di riferimento per dati di tipo trasversale e per dati verticali, come precedentemente descritto. Alcuni di questi dataset sono noti e molto utilizzati, in particolare, nel mondo dei Linked Open Data. Essi sono considerati dati di riferimento in quanto affidabili e completi; pertanto *si consiglia di valutare la possibilità di collegamento dei propri dati con i questi dataset di riferimento.*

Dati trasversali

DBPedia⁷³. E' un grande dataset in formato Linked Data contenente informazioni estratte da Wikipedia. Data la sua natura enciclopedica, viene "linkato" da molteplici dataset all'interno del Web dei dati tanto da essere considerato l'hub principale.

DBPedia Italia⁷⁴. E' la controparte italiana di DBPedia contenente i dati estratti da Wikipedia Italia.

PCData⁷⁵. E' lo spazio dati del Sistema Pubblico di Connettività e Cooperazione. E' composta da diversi dataset in formato Linked Data come quello dell'IPA (Indice della Pubblica Amministrazione) e quello dei dati sui contratti SPC relativi ai servizi Web. La presenza al suo interno della base di dati ufficiale rivista dal CAD (art. 57-bis) per l'identificazione univoca delle pubbliche amministrazioni italiane, candida *SPCData a diventare l'hub dei dati di tipo Linked della pubblica amministrazione italiana.* Attualmente, SPCData è collegato ad altri Linked Data di alcune pubbliche amministrazioni come CNR, Comune di Firenze e Regione Piemonte.

Dati territoriali e ambientali

⁶⁹ Creative Commons, "Describing Copyright in RDF", <http://creativecommons.org/ns>, 2014.

⁷⁰ Schema.org, <http://schema.org/>, 2013.

⁷¹ VANN: A vocabulary for annotating vocabulary descriptions, <http://vocab.org/vann/.html>, 2014.

⁷² VOAF, <http://lov.okfn.org/vocab/voaf/v2.3/index.html>, 2014.

⁷³ DBPedia, <http://dbpedia.org/About>, 2014.

⁷⁴ DBPedia Italiana, <http://it.dbpedia.org/>, 2014.

⁷⁵ SPCData, <http://spcdata.digitpa.gov.it>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

GEONAMES⁷⁶. E' un database che contiene informazioni geografiche provenienti da ogni parte del mondo. E' accessibile sia tramite Web services, sia tramite SPARQL e per questo è referenziato da tanti altri dataset che intendono geolocalizzare le proprie informazioni.

RNDT⁷⁷. E' una base dati di interesse nazionale che costituisce il catalogo nazionale dei metadati territoriali riguardanti i dati territoriali e i servizi ad essi relativi disponibili presso le Pubbliche Amministrazioni. E' pienamente conforme alla direttiva INSPIRE.

SINAnet⁷⁸. E' la rete del Sistema Informativo Ambientale Nazionale (SINA) che raccoglie dati e informazioni necessari a descrivere e comprendere i fenomeni ambientali, al fine di fornire supporto dell'azione di governo e di produrre con continuità prodotti e servizi informativi basati su indicatori e indici.

OpenStreetMap – OSM⁷⁹. E' un progetto collaborativo, ispirato a Wikipedia, che ha il fine di creare una base dati georeferenziata. Gli output di OSM sono molteplici: mappe digitali online, oggetti classici del mondo della cartografia stampata, servizi di routing, geocodifica, mappe tematiche (es. ciclismo, escursionismo, accessibilità motoria, archeologia), ecc. Il progetto tutela la propria banca dati attraverso la licenza ODbL che prevede l'obbligo di continuare a tenere aperta la banca dati citando la fonte e usando la frase "Copyright OpenStreetMap and contributors" nei prodotti derivati. I dati sono disponibili attraverso varie modalità: download diretto dell'intera banca dati in formato XML, download di alcune classi convertite nei formati GIS più noti, download tramite API attraverso interrogazioni per area o categoria. La categorizzazione dei dati avviene tramite un processo bottom-up governato da una pagina wiki dove gli attributi sono definiti con più tag. Maggiori dettagli su applicazioni, tassonomie, api, si trovano nella sezione del wiki dedicato al progetto.

Dati sulla medicina

MeSH (Medical Subject Headings)⁸⁰. E' un dataset, strutturato come vocabolario controllato, che tiene traccia della letteratura scientifica biomedica. Non nasce per il contesto del Web Semantico, ma diversi sforzi hanno prodotto degli adattamenti di MeSH in Linked Data. Una di queste versioni può essere reperita all'indirizzo <http://bioportal.bioontology.org/ontologies/3019>.

⁷⁶GeoNames, <http://www.geonames.org/>, 2014.

⁷⁷Agenzia per l'Italia Digitale, "Repertorio nazionale per i dati territoriali - RNDT", <http://www.rndt.gov.it/RNDT/home/index.php>, 2014.

⁷⁸ISPR, SINAnet, <http://www.mais.sinanet.isprambiente.it/ost/>, 2014.

⁷⁹OpenStreetMap, <http://www.openstreetmap.org/>.

⁸⁰U.S. National Library of Medicine, "Medical Subject Headings", <http://www.nlm.nih.gov/mesh/>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms)⁸¹. E' un dataset, strutturato come vocabolario controllato, di termini medici e clinici in diverse lingue. E' mantenuto dall'associazione no-profit IHTSDO. Esistono varie conversioni dell'ontologia per il Web Semantico; tuttavia nessuna di queste è attualmente ufficiale.

⁸¹ SNOMED CT, <http://www.ihtsdo.org/snomed-ct/>, 2014.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

4. Pubblicazioni di dati non strutturati

Interesse della Regione Puglia è anche la pubblicazione e diffusione dei dati non strutturati. Per dati non strutturati si intende dati privi di una qualsiasi struttura, tipicamente sono dei dati testuali che per loro natura non è possibile struttura secondo schemi ben precisi. In questa categoria ricadono i dati Regionali come: news, delibere, atti dirigenziali, bandi, FAQ, bollettini, ecc... Questi dati vengono prodotti e pubblicati da differenti portali, ma esistono delle tecnologie che permettono la pubblicazione secondo specifici standard facilmente interpretabili dai calcolatori. La pubblicazione dei dati secondo questi formati permette di alimentare automaticamente sistemi di archiviazione, browsing e ricerca.

Il formato più utilizzato per la pubblicazione di contenuti sul Web è RSS. Il suo nome è acronimo di Really Simple Syndication. L'RSS è un dialetto dell'XML. Ogni file RSS deve essere conforme alle specifiche XML 1.0, così come pubblicato sul sito del World Wide Web Consortium (W3C).

4.1 Specifiche RSS 2.0

Il livello più alto in un documento RSS è costituito da un elemento <rss> con un attributo obbligatorio chiamato version, che specifica la versione RSS con la quale il file risulta conforme. Per essere conforme a queste specifiche l'attributo version deve avere come valore 2.0.

Subordinato all'elemento <rss> è l'elemento <channel>, che contiene informazioni relative al canale (metadati) e il suo contenuto.

Questa è una lista degli elementi obbligatori per <channel>, ciascuno con una breve descrizione e un esempio.

title	Il nome del canale. Per gli utenti rappresenta il riferimento a chi fornisce il servizio. Se è disponibile un sito che presenta le stesse informazioni del file RSS, il titolo del canale (channel) dovrebbe essere lo stesso titolo del sito web.	GoUpstate.com News Headlines
link	L'URL del sito web HTML che corrisponde al canale.	http://www.goupstate.com/
description	Una frase o un riassunto per descrivere il canale.	The latest news from GoUpstate.com, a Spartanburg Herald-Journal Web site.

Questa è una lista degli elementi opzionali per <channel>.

language	La lingua nella quale è scritto il canale. Questo tag consente agli aggregatori di raggruppare tutti i siti italiani, per esempio, in una singola pagina. Una lista di	en-us
-----------------	--	-------

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

	valori consentiti per questo elemento, così come fornita da Netscape. E' possibile usare i valori definiti ⁸² dal W3C.	
copyright	Informativa sul copyright per il contenuto del canale.	Copyright 2002, SpartanburgHerald-Journal
managingEditor	Indirizzo email della persona responsabile dei contenuti editoriali.	geo@herald.com (George Matesky)
webMaster	Indirizzo email della persona responsabile dei problemi tecnici relativi al canale.	betty@herald.com (Betty Guernsey)
pubDate		Sat, 07 Sep 2002 00:00:01 GMT
lastBuildDate	La data dell'ultima modifica dei contenuti del canale.	Sat, 07 Sep 2002 09:42:31 GMT
category	Specifica una o più categorie alle quali fa parte il canale. Segue le stesse regole dell'elementocategory del livello <item>.	<category>Newspapers</category>
generator	Una stringa indicante il programma usato per generare il canale.	MightyInHouse Content System v2.3
docs	Un URL che punta alla documentazione del formato usato per il file RSS. Probabilmente si tratterà di un collegamento alla pagina delle specifiche RSS. L'utilità è per gli utenti che potranno incrociare un file RSS in un web server tra 25 anni che si chiederanno di cosa si tratti.	http://www.rssboard.org/rss-specification
cloud	Consente di registrare processi con un cloud (Servizio web che supporti l'interfaccia rssCloud) per essere notificati degli aggiornamenti del canale, implementando un protocollo leggero di pubblicazione-iscrizione per feed RSS.	<cloud domain="rpc.sys.com" port="80" path="/RPC2" registerProcedure="pingMe" protocol="soap"/>
ttl	ttlsta per time to live. E' il numero di minuti che indica per quanto tempo un canale può essere mantenuto in cache prima	<ttl>60</ttl>

⁸²<http://www.w3.org/TR/REC-html40/struct/dirlang.html#langcodes>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

	di rieseguire un aggiornamento dalla sorgente.
image	Specifica una immagine GIF, JPEG o PNG che può essere mostrata con il canale.
rating	Il PICS ⁸³ per la votazione del canale.
textInput	Specifica un campo input text che può essere mostrato con il canale.
skipHours	Un suggerimento per gli aggregatori indicante in quali orari possono saltare la lettura.
skipDays	Un suggerimento per gli aggregatori indicante in quali giorni possono saltare la lettura.

<image> sub-element of <channel>

<image> è un sotto elemento opzionale di <channel>, che contiene a sua volta tre sotto elementi obbligatori e tre sotto elementi facoltativi.

<url> è l'URL di una immagine GIF, JPEG o PNG che rappresenta il canale.

<title> descrive l'immagine, è usato come contenuto dell'attributo ALT del tag HTML quando il canale è rappresentato come HTML.

<link> è l'URL del sito, quando il canale è rappresentato come HTML l'immagine è un collegamento al sito. (Nota. in pratica i tag<title> e <link> dell'immagine dovrebbero avere gli stessi valori di <title> and <link> appartenenti a channel).

Gli elementi opzionali includono <width> e <height>, numeri indicanti la larghezza e l'altezza in pixel dell'immagine. <description> contiene il testo da includere come contenuto dell'attributo title del link creato attorno all'immagine in fase di rappresentazione HTML.

Il valore massimo per width è 144, quello predefinito è 88.

Il valore massimo per height è 400, quello predefinito è 31.

<cloud> sub-element of <channel>

<cloud> è un sotto elemento opzionale di <channel>.

Specifica un servizio web che supporti l'interfaccia rssCloud che può essere implementata in HTTP-POST, XML-RPC o SOAP 1.1. Il suo scopo è consentire la registrazione di processi per essere notificati degli aggiornamenti del canale, implementando un protocollo leggero di pubblicazione-iscrizione per feed RSS.

⁸³<http://www.w3.org/PICS/>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

```
<cloud domain="rpc.sys.com" port="80" path="/RPC2"
registerProcedure="myCloud.rssPleaseNotify" protocol="xml-rpc" />
```

In questo esempio, per richiedere la notifica del canale riportato, è necessario inviare un messaggio XML-RPC alla porta 80 del server rpc.sys.com, con il percorso /RPC2. La procedura da richiamare è denominata myCloud.rssPleaseNotify.

Una spiegazione completa di questo elemento e dell'interfaccia rssCloud è al seguente link <http://www.rssboard.org/rsscloud-interface>.

<ttl> sub-element of <channel>

<ttl> è un sotto elemento opzionale di <channel>.

ttlsta per time to live. E' un numero di minuti che indica per quanto tempo un canale può essere mantenuto in cache prima di rieseguire un aggiornamento dalla sorgente. Questo rende possibile la gestione delle sorgenti RSS tramite un network di file sharing.

Esempio: <ttl>60</ttl>

<textInput> sub-element of <channel>

Un channel può contenere, facoltativamente, un sotto elemento <textInput>, che contiene a sua volta 4 sotto elementi obbligatori.

<title> -- L'etichetta del bottone di invio del modulo input text.

<description> -- Spiega il modulo input text.

<name> -- Il nome della casella di testo nel campo input text.

<link> -- L'URL dello script CGI predisposto per processare le richieste dell'input text.

Lo scopo dell'elemento <textInput> è ancora qualcosa di misterioso. E' possibile fare uso di questo tag per specificare il campo di un motore di ricerca interno. In alternativa lo si può adottare per consentire ai lettori di inviare un feedback. La maggior parte degli aggregatori ignora questo tag.

Un channel può contenere un qualsiasi numero di <item>. Un item può rappresentare una "storia", come ad esempio una storia in un giornale o una rivista; in questo caso la sua descrizione è una sintesi della storia ed il collegamento indirizza alla storia completa.

Un item può anche contenere informazioni sufficienti per essere autonomo; in questo caso la descrizione contiene il testo (è ammesso un testo HTML **con le entità codificate**) ed i tag<link> e <title> possono essere omessi.

Tutti gli elementi di un <item> sono opzionali, tuttavia deve essere almeno presente un <title> o una <description>.

title	Il titolo di item.	Venice Film Festival Tries to Quit Sinking
link	L'URL di item.	http://nytimes.com/2004/12/07FEST.html
description	Una sintesi di item.	Some of the most heated chatter at the Venice Film Festival this week was about the way that the arrival of the

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

	stars at the Palazzo del Cinema was being staged.
author	Indirizzo email dell'autore di item.
category	Include item in una o più categorie.
comments	URL di una pagina contenente i commenti correlati a item.
enclosure	Descrive un oggetto multimediale allegato a item.
guid	Una stringa che identifica univocamente item.
pubDate	Indica quando item è stato pubblicato.
source	Il canale RSS dal quale proviene item.

<source> sub-element of <item>

<source> è un sotto elemento opzionale di <item>.

Il suo valore è il nome del canale RSS da quale proviene l'item, derivato dal suo <title>. Ha un attributo obbligatorio, url, che collega alla rappresentazione XML della sorgente.

```
<source url="http://www.tomalak.org/links2.xml">Tomalak's Realm</source>
```

Lo scopo di questo elemento è di diffondere i crediti per link e di pubblicizzare le sorgenti delle news / degli elementi del feed. Può essere utilizzato nel comando Post di un aggregatore.

Dovrebbe essere generato automaticamente quando si invia un elemento da un aggregatore ad uno strumento di scrittura per blog.

<enclosure> sub-element of <item>

<enclosure> è un sotto elemento opzionale di <item>.

Ha tre attributi obbligatori. url indica dove si trova l'allegato, length indica la dimensione espressa in byte e type indica il tipo di file corrispondente ad uno dei MIME type standard.

L'url deve essere un url http.

```
<enclosure url="http://www.scripting.com/mp3s/weatherReportSuite.mp3" length="12216320" type="audio/mpeg" />
```

<category> sub-element of <item>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

<category> è un sotto elemento opzionale di <item>.

Ha un attributo opzionale, domain, una stringa che identifica una classificazione tassonomica per categorie.

Il valore dell'elemento è una stringa separata da slash (/) che identifica una collocazione gerarchica all'interno della tassonomia indicata. Gli interpreti possono stabilire convenzioni per l'interpretazione delle categorie. Due esempi sono forniti di seguito:

```
<category>Grateful Dead</category>
```

```
<category domain="http://www.fool.com/cusips">MSFT</category>
```

E' possibile includere tanti elementi category quanti siano necessari, per domini differenti, ed avere un elemento con riferimenti incrociati in diverse parti dello stesso dominio.

<pubDate> sub-element of <item>

<pubDate> è un sotto elemento opzionale di <item>.

Il suo valore è una data⁸⁴, corrispondente alla pubblicazione di item. Se la data fa riferimento ad un periodo futuro gli aggregatori possono scegliere di non visualizzare item fino a quella data.

```
<pubDate>Sun, 19 May 2002 15:21:36 GMT</pubDate>
```

<guid> sub-element of <item>

<guid> è un sotto elemento opzionale di <item>.

guid per Globally Inique Identifier. E' una stringa che identifica in modo univoco item.

Quando presente, l'aggregatore può scegliere di usare questa stringa per determinare se un item è nuovo.

```
<guid>http://some.server.com/weblogItem3207</guid>
```

Non ci sono regole particolari per la sintassi dell'elemento guid. Gli aggregatori devono considerarlo come stringa. Spetta alla sorgente del feed garantire l'unicità del valore della stringa.

Se l'elemento guid ha un attributo chiamato isPermalink impostato con valore *true*, il lettore può assumere che si tratti di un permalink all'elemento, ovvero un URL che può essere aperto in un web browser e che punta all'intero contenuto descritto dall'elemento <item>. Un esempio:

```
<guid
```

```
isPermaLink="true">http://inessential.com/2002/09/01.php#a2</guid>
```

isPermalink è opzionale ed il suo valore predefinito è *true*. Se il suo valore è *false*, l'elemento guid non può essere considerato un URL o un indirizzo a qualcosa in particolare.

<comments> sub-element of <item>

<comments> è un sotto elemento opzionale di <item>.

Se presente, corrisponde all'URL della pagina contenente i commenti a item.

⁸⁴<http://asg.web.cmu.edu/rfc/rfc822.html>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

<comments>http://ekzemplo.com/entry/4403/comments</comments>

<author> sub-element of <item>

<author> è un sotto elemento opzionale di <item>.

Corrisponde all'indirizzo email dell'autore di item. Per i giornali e le riviste che distribuiscono un feed RSS, l'autore è la persona che ha scritto l'articolo che item descrive. For blog composti da più collaboratori, l'autore di item potrebbe essere differente dall'editore responsabile (managin editor) o dal webmaster. Per un blog gestito da un singolo individuo avrebbe senso omettere l'elemento <author>.

4.2 *Commenti sullo standard RSS*

L'RSS imposta delle restrizioni al primo carattere non spazio e non bianco che compone il valore degli elementi <link> e <url>. Il valore di questi elementi deve cominciare con uno schema URI IANA-registrato⁸⁵, come ad esempio http://, https://, news://, mailto:// e ftp://. Prima del formato RSS 2.0 le specifiche consentivano solo http:// e ftp://, tuttavia, in pratica gli altri schemi URI erano ugualmente usati dagli sviluppatori e supportati dagli aggregatori. Gli aggregatori possono avere dei limiti per quanto riguarda il supporto a schemi URI. Gli sviluppatori di contenuti non dovrebbero dare per scontato che ogni lettore per feed supporti tutti gli schemi.

Nell'RSS 0.91, molti elementi sono limitati a 500 o 100 caratteri. Non ci possono essere più di 15 <items> in un <channel> nel formato 0.91. Non ci sono limiti di lunghezza per le stringhe o limiti a livelli XML nell'RSS 0.92 e superiori. Gli interpreti possono imporre limiti specifici e i produttori di contenuti possono avere preferenze sul numero massimo di <item> che desiderano inserire in un channel, o quali stringhe devono avere limiti di lunghezza.

Nell'RSS 2.0 c'è un modo per collegare un canale al suo identificatore in un sistema di catalogazione usando la funzionalità dell'elemento category a livello channel, come descritto in precedenza. Per esempio, per collegare un canale al suo identificatore in Syndic8, basta includere un elemento category come sub-elemento di <channel>, con dominio "Syndic8" e come valore l'identificativo per il canale nel database di Syndic8. L'elemento category appropriato per Scripting News dovrebbe essere <category domain="Syndic8">1765</category>.

Una domanda frequente per quanto riguarda gli elementi <guid> è se/come possano essere comparati agli elementi <link>. Non sono la stessa cosa? Sì, in alcuni sistemi di contenuti, e no in altri. In alcuni sistemi, <link> è un permalink ad un post di un blog. Tuttavia, in altri sistemi, ogni <item> è un riassunto di un articolo più lungo, <link> punta all'articolo e <guid> è il permalink al post del blog.

In ogni caso, è consigliato fornire un guid e se possibile renderlo un permalink. Questo consente agli aggregatori di evitare la ripetizione di item, anche se è stato modificato o sono stati apportati cambiamenti.

Se avete domande sul formato RSS 2.0, siete invitati a postarle nella mailing list pubblica RSS-Public. La mailing list, gestita dall'RSS Advisory Board, serve come

⁸⁵<http://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

risorsa di supporto per utenti, autori e sviluppatori alle prese con la creazione e l'uso di contenuti nel formato RSS 2.0.

5. Licenze d'uso dei dati

L'aspetto relativo alle licenze dei dati è cruciale per quanto riguarda l'uso che gli utenti possono effettivamente fare dei dati, i vincoli di copyright da applicare a lavori derivati da quei dati, il mantenimento della paternità sui dati, ecc. Il mondo delle licenze è vasto e intricato. Questa sezione analizza e confronta le licenze maggiormente utilizzate nel contesto Open Data. Le licenze più diffuse sono quelle basate su Creative Commons⁸⁶(CC). CC è un'organizzazione non a fini di lucro che nasce con l'intenzione di armonizzare l'articolato mondo del diritto d'autore (in Italia regolato dalla legge n. 633 del 22 aprile 1941) e del copyright. Nel 2002, CC ha pubblicato un primo insieme di licenze che si sono affermate come standard de facto a livello internazionale. In linea generale, la scelta di adottare un modello di licenze basate su CC deriva principalmente dall'esigenza di armonizzare il rilascio di dati aperti con analoghe iniziative di carattere internazionale, semplificando e promuovendo il riuso dei dati stessi. Le licenze CC, infatti, facilitano la comprensione dei dati e consentono un loro ampio riuso grazie a un buon grado di permessi.

In funzione delle specificità dei diversi insiemi di dati la scelta può ricadere su diversi tipi di licenze, anche non necessariamente CC. Nella maggior parte dei casi, è opportuno seguire i criteri derivanti dalla definizione di Open Data della OKF “[...] 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)”. Di seguito si analizzano alcune licenze utilizzate dalle amministrazioni rappresentate all'interno del gruppo di lavoro incaricato della produzione delle presenti Linee guida; per ogni licenza, se ne evidenziano le caratteristiche e si forniscono alcune indicazioni pratiche per il loro corretto uso. In Italia, licenze CC sono state utilizzate e promosse da alcune regioni quali, ad esempio, Emilia-Romagna e Piemonte, e da comuni come quello di Genova; a livello di PAC lo scenario è, invece, più variegato con amministrazioni che adottano licenze CC (e.g., MIUR) e altre (e.g., INPS) che si orientano per l'adozione di licenze italiane.

LICENZA CREATIVE COMMONS ZERO (CC0)

La Creative Commons Zero esprime “la più ampia e libera utilizzazione gratuita, anche per fini commerciali e con finalità di lucro”. Apponendo su un documento la dichiarazione CC0 si rinuncia a tutti i diritti sul documento e sui suoi contenuti, dati inclusi, nella misura massima possibile prevista dalla legge. La Creative Commons Zero deve essere preceduta da una dichiarazione relativa all'origine del documento.

⁸⁶<http://www.creativecommons.it/>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

A titolo di esempio si riporta un passaggio che riguarda l'applicazione della suddetta licenza per il riutilizzo delle Banche dati della rilevazione scolastica della regione Piemonte.

Il riutilizzo della "Banca dati della rilevazione scolastica" è stato concesso da Regione Piemonte ai sensi della Legge regionale n. 24/2011 e s.m.i.. Regione Piemonte autorizza, pertanto, la libera e gratuita consultazione, estrazione, riproduzione, modifica e riutilizzo del documento e dei dati in esso contenuti da parte di chiunque vi abbia interesse per qualunque fine secondo i termini della Dichiarazione Creative Commons - CC0 1.0 Universal.

LICENZA CREATIVE COMMONS ATTRIBUZIONE (CC-BY)

Un'alternativa alla licenza CC0 è la licenza Creative Commons "Attribuzione o equivalente". Questa permette al soggetto utilizzatore di riprodurre, distribuire, comunicare, esporre, rappresentare, nonché di modificare e usare un insieme di dati anche a fini commerciali con il solo obbligo di attribuire la paternità dell'opera. Anche questa licenza risulta essere, al pari della standard, espressionedel principio della "più ampia e libera utilizzazione gratuita anche per fini commerciali e con finalità di lucro". In caso di uso della CC-BY però, l'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, o a corredo, della licenza, come di seguito meglio specificato ("Attribuzione"). In generale, questa licenza è adottabile per le banche dati che risultano chiaramente tutelate dal diritto d'autore e/o dal diritto sui generis⁸⁷.

Al fine di prevenire qualsiasi incertezza interpretativa da parte del licenziatario e 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 diritti cosiddetti sui generis a tutela dei contenuti della banca dati stessa. Vado è chiarito che la licenza disciplina tutti i diritti di cui alla L. 633/41 e s.m.i., con esplicita inclusione dei Diritti del titolare di una banca di dati, di cui al Titolo II-bis della legge stessa. Un esempio di applicazione di questa licenza a una generica banca dati è riportata qui di seguito:

La titolarità piena ed esclusiva del documento "[DENOMINAZIONE E DESCRIZIONE SINTETICA

DEL DOCUMENTO]" è di Regione Piemonte, ai sensi della L. 633/41 e s.m.i.

(Licenziante).

Regione Piemonte autorizza la libera e gratuita consultazione, estrazione, riproduzione e modifica dei dati in essa contenuti da parte di chiunque (Licenziatario) vi abbia interesse per qualunque fine, purché nel rispetto dei termini della licenza Creative Commons – Attribuzione 2.5 Italia.

Si precisa esplicitamente che con la presente licenza il Licenziante intende autorizzare il Licenziatario ad esercitare, ferme restando le restrizioni della licenza di cui sopra, anche i

⁸⁷La tutela giuridica del diritto sui generis è stata introdotta dalla Direttiva 96/9/CE. La tutela sui generis si riferisce alla protezione garantita all'insieme delle informazioni contenute all'interno di una raccolta di dati per distinguerla dalla tutela, riconosciuta dal diritto d'autore, che può interessare invece la struttura o architettura della banca dati.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

diritti disciplinati dall'art. 102-bis e ss., L. 633/41 e s.m.i. (c.d. diritto sui generis del costituente di una banca di dati).

L'attribuzione prevista dalla licenza dovrà avvenire nella seguente forma: [INSERIRE NOTA PER

L'ATTRIBUZIONE]

Per quanto concerne l'attribuzione, caratteristica propria delle licenze CC-BY, il licenziatario dovrà provvedere alla menzione, rispetto al mezzo di comunicazione o supporto utilizzato, di:

- l'autore originale e/o titolare dei diritti;
- le terze parti designate, se esistenti;
- la descrizione/titolo del documento;
- nella misura in cui ciò sia ragionevolmente possibile, l'Uniform Resource Identifier (URI) che
- il Licenziante specifichi dover essere associato con il documento oggetto di riutilizzo;
- in caso di documenti rielaborati o opere derivate di vario genere, l'attribuzione dovrà essere
- effettuata in modo tale da non ingenerare confusione rispetto all'origine del documento stesso,
- ad esempio: "carta topografica basata su ...".

Alle licenze suddette potranno essere altresì allegati l'invito a segnalare errori o imprecisioni, l'invito a inviare alla Direzione competente per materia eventuali versioni aggiornate/rielaborate del documento reso disponibile al riuso. A differenza della licenza CC-BY sopra descritta, versione 2.5, successive versioni della CC-BY Italia producono sul diritto sui generis della banca dati i medesimi effetti della Dichiarazione CC0, ovvero una rinuncia totale e incondizionata ai diritti.

ALTRE LICENZE CREATIVE COMMONS

La licenza CC-BY può essere estesa mediante alcuni attributi, quali:

Share Alike (SA): obbliga i lavori derivati a essere licenziati con la stessa licenza del lavoro originale

Non Commercial (NC): consente la copia, la distribuzione e l'uso del lavoro (o dati) solo per scopi non commerciali

No Derivative Works (ND): consente la copia, distribuzione e l'uso del lavoro, impedendola creazione di lavori derivati

La combinazione di questi attributi genera di fatto altre licenze. Naturalmente, considerato che alcuni attributi sono mutuamente esclusivi, non tutte le combinazioni hanno senso. Nella pratica si considerano altre cinque licenze oltre alla CC0 e alla CC-BY, che sono:

CC-BY-ND: Attribution No Derivatives;

CC-BY-NC: Attribution Non-Commercial;

CC-BY-NC-SA: Attribution Non-Commercial Share Alike;

CC-BY-NC-ND: Attribution Non-Commercial No Derivatives.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

È evidente che le licenze che non consentono il riutilizzo dei dati non sono compatibili con lo spirito e gli scopi degli Open Data.**LICENZA IODL 2.0**

La licenza IODL 2.0⁸⁸ (Italian Open Data License) prevede che l'utente possa liberamente (i) consultare, estrarre, copiare e pubblicare i dati; e (ii) creare un lavoro derivato integrando diversidataset.

LICENZA IODL 1.0

Questa prima versione della IODL⁸⁹ è simile alla IODL 2.0 ma con l'obbligo dell'utente di pubblicare o condividere i lavori derivati con la stessa licenza.

5.1 Analisi delle licenze

La tabella seguente analizza e confronta le diverse licenze prima descritte sulla base di alcune caratteristiche identificate, in funzione dei livelli di interoperabilità e riusabilità dei dati che possono essere garantiti. Alcune di queste caratteristiche come ad esempio "Uso per fini commerciali", "Uso gratuito del lavoro soggetto alla licenza", ecc, sono ben note e auto-esplicative. Altre, invece, nascono proprio dall'analisi prodotta dal gruppo di lavoro; per esempio, si è voluto analizzare le licenze sotto l'aspetto della "Portabilità inter-lingua", ossia la possibilità che essa sia supportata da una traduzione in diverse lingue. Tale peculiarità è particolarmente importante in un'ottica di riuso transfrontaliero dei dati. Anche la "Riconoscibilità internazionale del logo", che consente di capire se la licenza gode di sufficiente fama internazionale riconducibile al logo ad essa associato, è una peculiarità individuata importante ancora una volta nell'ottica di riuso dei dati per un'utenza non necessariamente confinata all'Italia. L'analisi delle licenze qui proposta si sofferma anche sulla "Compatibilità inter-licenza per lavori derivati" intesa come la possibilità di applicare licenze differenti a lavori derivati dall'uso dei dati.

	CC0	CC-BY	CC-BY-SA	CC-BYSA-NC	IODL 1.0	IODL 2.0
Portabilità inter-lingua	X	X	X	X		
Riconoscibilità internazionale del logo	X	X	X	X		
Uso per fini commerciali	X	X	X			X
Uso gratuito del lavoro soggetto alla licenza	X	X	X	X	X	X
Possibilità di modificare i dati	X	X	X	X	X	X
Possibilità di preservare la paternità		X	X	X	X	X

⁸⁸<http://www.dati.gov.it/iodl/2.0/>

⁸⁹<http://www.formez.it/iodl/>

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

Compatibilità inter-licenza per lavori derivati	X	X				X
---	---	---	--	--	--	---

Dalla tabella si può notare come la maggior parte delle licenze CC soddisfino la quasi totalità delle caratteristiche considerate, risultando quindi quelle che meglio si prestano a rispondere ai requisiti d'interoperabilità e massimo riuso dei dati. Ai fini della scelta del tipo di licenza da associare ai dati da pubblicare, infine, è importante sottolineare anche la relazione bidirezionale che intercorre tra le licenze e i modelli di business che possono essere abilitati dagli Open Data. Se da un lato è possibile sostenere che le licenze con caratteristiche di maggior apertura consentono lo sviluppo di maggiori opportunità di business per il mondo delle imprese, è anche vero, dall'altro, che l'aumento delle opportunità di business facilitano la diffusione e l'affermazione delle stesse licenze, con innegabili benefici per l'espansione del movimento Open Data nel mondo.

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

6. Risorse e link utili

- [1] Ahern, R. Turner, A. Clouse. *CMMI Distilled: A Practical Introduction to Integrated Process Improvement*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA 2003, ISBN:0321186133
- [2] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. Redmond, WA, 2009. ISBN: 978-0-9825442-0-4.
- [3] E. Almirall, M. Lee, and J. Wareham. Mapping Living Labs in the Landscape of Innovation Methodologies. *Technology Innovation Management Review*, 12-18, September 2012.
- [4] T. Di Noia, R. De Virgilio, E. Di Sciascio, F. M. Donini. *Semantic Web: Tra ontologie e Open Data*. Apogeo. Aprile 2012.
- [5] <http://www.w3.org/TR/owl2-overview/>.
- [6] Vademecum sui dati aperti (<http://www.funzionepubblica.gov.it/media/982175/vademecumopendata.pdf>).
- [7] Libro bianco per il riutilizzo dell'informazione del settore pubblico (http://www.evpsi.org/evpsifiles/bianco_beta.pdf).
- [8] Open data handbook (<http://opendatahandbook.org/it/>).
- [9] Come si fa Open Data ([http://www.reteforum.it/suite/allegati/Come si fa Open Data per la PA.pdf](http://www.reteforum.it/suite/allegati/Come%20si%20fa%20Open%20Data%20per%20la%20PA.pdf)).
- [10] Lo stato trasparente – Linked Open Data e cittadinanza attiva (<http://www.linkedopendata.it/wp-content/uploads/statotrasparente.pdf>).
- [11] Linee Guida per l'interoperabilità semantica attraverso i Linked Open Data (http://www.digitpa.gov.it/sites/default/files/CdC-SPC-GdL6-InteroperabilitaSemOpenData_0.pdf).
- [12] Agenda Digitale Europea (http://ec.europa.eu/information_society/digital-agenda/index_en.htm).
- [13] European Public Sector Information Platform, "Review of Recent PSI Re-Use Studies Published" (<http://epsiplatform.eu/content/review-recent-psi-re-use-studies-published>).
- [14] Open definition, "Defining the Open in Open Data, Open Content and Open Services" (<http://opendefinition.org/okd/>).
- [15] Open Government Data, "8 Principles of Open Government Data" (<http://www.opengovdata.org/home/8principles>), Sebastopol, California USA, 2007.
- [16] W3C, Linking Open Data (<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>).
- [17] W3C, DataSetRDFDumps (<http://www.w3.org/wiki/DataSetRDFDumps>).
- [18] W3C, RDF Working Group (http://www.w3.org/2011/rdf-wg/wiki/Main_Page).
- [19] W3C, OWL Working Group (http://www.w3.org/2007/OWL/wiki/OWL_Working_Group).
- [20] W3C, SPARQL Working Group (http://www.w3.org/2009/sparql/wiki/Main_Page).

Linee Guida Regionali per la pubblicazione di Open Data .-Tutorial

-
- [21] W3C, SKOS Simple Knowledge Organization System (<http://www.w3.org/2004/02/skos/>).
- [22] Linked PA – Portale Semantico della Pubblica Amministrazione (<http://www.ontologiapa.it/>).
- [23] W3C, Semantic Web Deployment Working Group (<http://www.w3.org/2006/07/SWD/>).
- [24] W3C, Government Linked Data Working Group (<http://www.w3.org/2011/gld/charter>).
- [25] Creative Commons, “Describing Copyright in RDF” (<http://creativecommons.org/ns>).
- [26] Creative Commons Italia (<http://www.creativecommons.it/Licenze>).
- [27] Creative Commons, CC0 1.0 Universal (<http://creativecommons.org/publicdomain/zero/1.0/legalcode>).
- [28] Creative Commons, CC-BY Attribution 2.5 (<http://creativecommons.org/licenses/by/2.5/legalcode>).
- [29] Italian Open Data License v.2.0 (<http://www.dati.gov.it/iodl/2.0/>).
- [30] Specifiche SPAQL 1.0(<http://www.w3.org/TR/rdf-sparql-query/>).
- [31] Microformati (<http://schema.org/>).
- [32] LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO (anno 2014), Agenzia per l’Italia Digitale, Presidenza del Consiglio dei Ministri. Ver 0.60 (bozza consultazione):
http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.6.pdf.

ALLEGATO D

Piano Annuale degli Open Data regionali

Regione Puglia
Area Politiche per lo Sviluppo Economico,
il Lavoro e l'Innovazione



**Regione Puglia
Servizio Ricerca Industriale e Innovazione**

**Piano annuale degli Open Data regionali
Anno 2015**



*InnovaPuglia SpA
Strada prov. per Casamassima,
Km. 3
70100 Valenzano - BARI
Italia*

Maggio 2015

INDICE DEI CONTENUTI

1	Introduzione	
2	Quadro normativo di riferimento.....	
3	Glossario e termini di riferimento.....	
4	Obiettivi del Piano	
5	La piattaforma regionale e il formato dati	
6	La licenza d'uso dei dati	
7	Il percorso del 2015.....	
	Obiettivo 1: Migliorare la qualità, la quantità e la disponibilità dei dati aperti attraverso la piattaforma regionale dati.puglia.it	
	Obiettivo 2: Promozione e diffusione di un utilizzo attivo dei dati aperti presenti sulla piattaforma regionale dati.puglia.it 19	
	Allegato 1	
	Elenco dei Dataset in pubblicazione sul portale dati.puglia.it.....	
	Area Tematica: Ricerca e Sperimentazione	
	Area Tematica: Puglia Sociale.....	
	Area Tematica: Promozione della Salute	
	Area Tematica: Cultura e Turismo	
	Area Tematica: Statistica ufficiale e Sistema statistico nazionale.....	
	Area Tematica: Territorio e Aree Protette.....	
	Amministrazione Trasparente	

INDICE DELLE TABELLE

Tabella 1 - Formati di dato aperto.....
Tabella 2 - Licenze Creative Commons.....
Tabella 3 - Dataset Ricerca e Sperimentazione.....
Tabella 4 - Dataset Puglia Sociale
Tabella 5 - Dataset Promozione della Salute
Tabella 6 - Dataset Cultura e Turismo.....
Tabella 7 - Dataset Statistica ufficiale e Sistema Statistico Nazionale.....
Tabella 8 - Dataset Territorio e Aree Protette
Tabella 9 - Dataset Amministrazione Trasparente.....

1 Introduzione

In linea con quanto previsto dalla documentazione relativa ai *Linked Open Data* della Regione Puglia (*Principi generali per lo sviluppo degli Open Data della Regione Puglia, Linee Guida per la pubblicazione degli Open Data della Regione Puglia - Istruzioni Operative, Linee Guida per la pubblicazione degli Open Data della Regione Puglia – Tutorial*) aperti alla consultazione pubblica¹ e a cui si rimanda per gli approfondimenti, col presente documento si tracciano gli sviluppi per l'anno 2015 del percorso intrapreso dalla Regione Puglia per la valorizzazione del patrimonio informativo pubblico, trasformato e reso disponibile in formato aperto. Considerando le opportunità di sviluppo e di competitività indotte dalle tecnologie digitali, i dati pubblici delle amministrazioni, che solo fino a poco tempo fa avevano un ruolo funzionale al perseguimento dei compiti istituzionali delle amministrazioni, assumono in quest'ottica una differente valenza in termini di stimolo dell'economia digitale, sviluppo dell'innovazione e trasparenza amministrativa².

Nell'ambito della **Agenda Digitale Puglia 2020**³ il patrimonio informativo pubblico insieme alle competenze digitali sono individuati quali substrato di crescita dell'imprenditoria digitale. In particolare, inoltre, si individua attraverso l'*open data* il passaggio culturale necessario per il rinnovamento delle istituzioni nella direzione di apertura e trasparenza proprie dell'Open Government, perseguito con l'attuazione del progetto *Linked Open Data Regionale*, a cui fa riferimento la **piattaforma *dati.puglia.it***, che punta a rendere possibile il collegamento e l'utilizzo di dati provenienti da diverse sorgenti. Tutto ciò nella prospettiva di promuovere progetti di innovazione che, partendo da una revisione tecnologica e organizzativa degli attuali strumenti in uso nelle differenti aree regionali, consentano il libero accesso all'informazione regionale e traducano il concetto di Open Government in un vero e proprio modello sostenibile, al fine di rendere l'amministrazione regionale trasparente, aperta, partecipativa, collaborativa. In tal senso, l'intervento sugli open data non può che correlarsi con azioni atte a migliorare interoperabilità ed efficienza della digitalizzazione dei processi organizzativi, sfruttando al contempo i feedback dagli utilizzatori dei dati per migliorarne la qualità.

Secondo tale ottica, il presente documento, *aggiornato con cadenza annuale*, definisce il programma di dettaglio dei dataset "da liberare" nell'anno di riferimento, "stabilisce le priorità di intervento per l'apertura dei dati, le categorie di dati da pubblicare, il formato dei dataset, le licenze di riuso"⁴. Tutto ciò in linea con quanto previsto dalla Legge Regionale n.20 del 24 luglio 2012 "Norme sul software libero, accessibilità di dati e documenti e hardware documentato" e dalla relativa Deliberazione della Giunta Regionale contenente i primi indirizzi⁵, in termini di valorizzazione del patrimonio informativo pubblico "al fine di avviare prime sperimentazioni che puntino a qualificare l'utilizzo dei dati aperti come strumento di open government, attraverso percorsi partecipati".

¹ <http://www.dati.puglia.it/forum>

² vedi "Linee Guida Nazionali per la valorizzazione del Patrimonio Informativo Pubblico (Anno 2014) - Agenzia per l'Italia Digitale" (http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicog2014_v0.7finale.pdf)

³ <http://www.regione.puglia.it/index.php?page=documenti&opz=getdoc&id=1119>

⁴ vedi "Linked Open Data - Principi generali per lo sviluppo degli Open Data della Regione Puglia, Settembre 2014" (http://www.dati.puglia.it/documents/portlet_file_entry/10184/LOD_LG_PrincipiGenerali_OpenData_v1.2.pdf/269fab92-dfe5-4d89-b9cf-91aaa2763eba)

⁵ Deliberazione della Giunta Regionale n. 2183 del 31 ottobre 2012 "Legge regionale 24 luglio 2012, n. 20 - "Norme sul software libero, accessibilità di dati e documenti e hardware documentato" - Primi indirizzi" (Bollettino Ufficiale della Regione Puglia - n. 164 del 14-11-2012)

2 Quadro normativo di riferimento

Riferimento	Descrizione
Legge Regionale 24 luglio 2012, n. 20	Norme sul software libero, accessibilità di dati e documenti e hardware documentato
Deliberazione della Giunta Regionale 31 Ottobre 2012 n. 2183	Legge regionale 24 Luglio 2012, n. 20 "Norme sul software libero , accessibilità di dati e documenti e hardware documentato" – Primi indirizzi
Delibera CIVIT n. 2/2012	Linee guida per il miglioramento della predisposizione e dell'aggiornamento del Programma triennale per la trasparenza e l'integrità
Delibera CIVIT n. 105/2010	Linee guida per la predisposizione del Programma triennale per la trasparenza e l'integrità (articolo 13, comma 6, lettera e, del decreto legislativo 27 ottobre 2009, n. 150) - ALLEGATO – Esempio di struttura della sezione del sito "Trasparenza, valutazione e merito" approvata nella seduta del 14 ottobre 2010
Vademecum Open Data	Come rendere aperti i dati delle pubbliche amministrazioni – Versione Beta 2011
Decreto legislativo 7 marzo 2005, n. 82	Codice dell'Amministrazione Digitale
Direttiva 2003/98/CE del Parlamento europeo e del Consiglio del 17 novembre 2003	Riutilizzo dell'informazione del settore pubblico
Decreto legislativo 24 gennaio 2006, n. 36	Attuazione della Direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico
Direttiva 2003/98/CE - Bruxelles, 7.5.2009, COM(2009) 212	Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, Riutilizzo dell'informazione del settore pubblico: riesame della Direttiva 2003/98/CE - Bruxelles, 7.5.2009, COM(2009) 212 definitivo
Decreto Legislativo 30 dicembre 2010, n. 235	Modifiche ed integrazioni al decreto legislativo 7 Marzo 2005, n. 82, recante Codice dell'amministrazione digitale, a norma dell'articolo 33 della legge 18 giugnom2009, n. 69
Decreto Legislativo n. 36/2006	Accesso all'informazione nel settore pubblico, attuativo della Direttiva 2003/98 relativa alla "Public Sector Information - PSI"
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
D. Lgs. n. 196/2003	Codice in materia di protezione dei dati personali

Piano Annuale degli Open Data regionali

Riferimento	Descrizione
Legge n. 4/2004	Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici
Direttiva 2007/2/CE	Istituisce un'infrastruttura per l'informazione territoriale nella Comunità europea (INSPIRE)
Decreto Legislativo n. 32 del 27 gennaio 2010	Attuazione della direttiva 2007/2/CE, che istituisce un'infrastruttura per l'informazione territoriale nella Comunità europea (INSPIRE)
D. Lgs. n. 163/2006	Codice dei contratti pubblici in materia di lavori, servizi e forniture
D. Lgs. n. 150/2009	Ottimizzazione della produttività del lavoro pubblico e di efficienza e trasparenza delle pubbliche amministrazioni

3 Glossario e termini di riferimento

Riferimento	Descrizione
Attribution License (Licenza di Attribuzione)	Una licenza che richiede che venga citata la fonte originale del materiale distribuito (attribuzione)
Copyright	Diritto per i realizzatori di opere creative a circoscrivere l'utilizzo di quei lavori da parte di altri. Chi possiede il copyright ha titolo per decidere come altri possono utilizzare quel lavoro
Data protection legislation	La legislazione per la protezione dei dati non riguarda il proteggere i dati, ma il proteggere il diritto dei cittadini a vivere senza il timore che informazioni sulle loro vite private possano diventare pubbliche. La legge protegge la privacy (si pensi alle informazioni sulla posizione economica, la salute e le posizioni politiche di una persona) ed altri diritti quali il diritto alla libertà di movimento e di riunione
EU PSI Directive	La direttiva sul riutilizzo dell'informazione del settore pubblico, 2003/98/CE "si occupa di come gli enti del settore pubblico massimizzino il riuso delle proprie risorse informative"
Information Asset Register (Registro dei Dati Esistenti) - IAR	Gli IAR sono registri appositamente istituiti per immagazzinare e organizzare meta-dati sull'enorme quantità di informazioni detenute da amministrazioni ed enti pubblici. Uno IAR completo include set di dati, vecchi insiemi di file, file elettronici recenti, raccolte di statistiche, ricerche e così via. La Direttiva Europea sulla PSI riconosce l'importanza dei registri per i potenziali utilizzatori delle informazioni pubbliche. Richiede che gli Stati membri forniscano liste, portali, o qualcosa di analogo. Gli IAR possono essere sviluppati in diversi modi. I dipartimenti governativi possono sviluppare i propri cataloghi e collegarli a IAR nazionali. Gli IAR possono includere informazioni detenute da enti pubblici, ma che non sono ancora state pubblicate - e forse non lo saranno - in modo proattivo. Di conseguenza consentono al pubblico di individuare informazioni esistenti e che possono essere richieste. Perché il pubblico faccia uso degli IAR è importante che questi siano il più possibile completi in modo da fornire la garanzia che i documenti possano essere rinvenuti. La mancata completezza di alcuni registri costituisce un importante problema in quanto genera una

Piano Annuale degli Open Data regionali

Riferimento	Descrizione
	percezione di inaffidabilità che potrebbe scoraggiare dall'uso dei registri per la ricerca di informazioni. È essenziale che i meta-dati negli IAR siano completi in modo che i motori di ricerca siano grado di funzionare efficacemente. Nello spirito degli open Government data, gli enti pubblici dovrebbero rendere disponibili al pubblico i propri IAR offrendo dati grezzi e con una licenza aperta, in modo che i "civic hackers" possano utilizzare i dati, per creare - ad esempio - motori di ricerca e interfacce utente
Machine-readable	Le informazioni che sono leggibili a macchina - machine readable - sono informazioni che possono essere estratte da programmi informatici. I documenti PDF non sono machine readable. I computer sono in grado di visualizzare il testo, ma hanno grande difficoltà a comprendere il contesto che circonda il testo
Open Data	I dati aperti (Open Data) possono essere utilizzati per qualsiasi scopo. Maggiori informazioni sono reperibili presso opendefinition.org
Open Government Data	Open data prodotti dal governo. È comunemente accettato riferirsi ai dati raccolti nel corso dell'attività amministrativa, che non identificano soggetti specifici e che non violano la "commercial sensitivity". Il concetto di "dati governativi aperti" è un sottoinsieme della nozione di Public Sector Information, che ha una portata più ampia. Cfr. per ulteriori dettagli http://opengovernmentdata.org
Open standard	Generalmente intesi come norme tecniche svincolati da restrizioni di licenza. Possono anche essere interpretati nel senso che gli standard sono sviluppati in modo da essere neutrali rispetto al produttore
Public domain	L'opera creativa non è protetta dal diritto d'autore
Public Sector Information	Le informazioni raccolte o controllate dal settore pubblico
Re-use	Utilizzo di contenuti al di fuori della loro finalità originaria
Share-alike License	Una licenza che richiede che un'opera derivata sia protetta alle stesse condizioni o termini simile a quelle dell'opera originale
Formato di dati aperti	Formati di memorizzazione e rappresentazione di dati informatici le cui specifiche sono note e liberamente utilizzabili. I formati di dati aperti sono documentati in modo adeguato a consentire, senza restrizioni, la scrittura di programmi per elaboratore in grado di leggere e scrivere dati in tali formati sfruttando tutte le strutture e le specifiche descritte nella documentazione

4 Obiettivi del Piano

La legge regionale del Luglio 2012 agisce in coerenza e in continuità con le norme regionali in materia di trasparenza e avvia un processo di diffusione dei dati liberamente accessibili per tutti (Open Data) per una Pubblica Amministrazione aperta ai cittadini in termini di partecipazione al percorso decisionale (Open Government).

In particolare, essa favorisce il progresso sociale, il miglioramento della qualità della vita e lo sviluppo delle iniziative economiche private legate al riutilizzo delle informazioni del settore pubblico.

La Regione Puglia utilizza le tecnologie dell'informazione e della comunicazione per rendere fruibili i documenti e i dati pubblici di cui è titolare, assicurandone la pubblicazione tramite la rete Internet in formati aperti secondo gli standard internazionali.

I documenti e i dati pubblici, salvo i casi eccezionali individuati da appositi provvedimenti della Giunta regionale, sono accessibili e riutilizzabili a titolo gratuito nel rispetto della normativa vigente in materia di digitalizzazione della Pubblica Amministrazione, di accesso agli atti amministrativi, di protezione dei dati personali, di riutilizzo delle informazioni del settore pubblico, di diritto della proprietà intellettuale e industriale.

Le licenze per il riutilizzo dei documenti e dei dati pubblici predisposte in ottemperanza al decreto legislativo 24 gennaio 2006, n. 36 (Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico), consentono la più ampia e libera utilizzazione gratuita, anche per fini commerciali e con finalità di lucro.

Il riuso degli archivi e l'utilizzo dei dati dagli stessi estratti è consentito solo in osservanza alla normativa prevista dal Codice della privacy di cui al decreto legislativo 30 giugno 2003, n. 196 (Codice in materia di protezione dei dati personali) e successive modifiche e integrazioni.

Partendo dal contesto delle azioni previste nell'Agenda digitale italiana e dalle iniziative in essere per la sua realizzazione, nonché tenendo conto di tutte le iniziative in materia di sviluppo dell'Agenda italiana ed europea, la Regione Puglia intende realizzare un modello di riferimento per governare il processo di apertura dei dati pubblici nelle amministrazioni locali regionali – basato sul paradigma dei Linked Open Data - e la realizzazione e pubblicazione di Open Data secondo quanto previsto dall'Agenda Digitale italiana e da quanto regolamentato dalla normativa regionale.

Il modello si basa sul riconoscimento dei seguenti principi e azioni:

- a) definizione di un processo di produzione di “dati aperti” dove i singoli gestori/responsabili dell'informazione sono anche i decisori della pubblicazione, all'interno di una organizzazione strutturata di raccolta e diffusione dei dati
- b) ampliamento del modello di pubblicazione dei Dataset in formato Open a tutte le Amministrazioni locali interessate attraverso un unico Portale regionale dell'Open Data (www.dati.puglia.it)
- c) utilizzo dell'approccio Linked Open Data a tutto il patrimonio informativo pubblico generato e gestito dalle amministrazioni locali, anche in funzione dello sviluppo di applicazioni di nuova generazione da parte del sistema produttivo locale
- d) creazione di sinergie operative (in termini di risorse tecniche, strumentali ed economiche) con i progetti ICT di interesse già finanziati dal Sistema della PA pugliese, oltre che di carattere europeo (HOMER: progetto open data del mediterraneo, OpenDIA), ottimizzando gli investimenti ed utilizzando i contenuti già disponibili sul territorio

Piano Annuale degli Open Data regionali

- e) valorizzazione degli investimenti effettuati, gestione del processo di liberalizzazione dei dati in modalità Open Data, supportando uffici ed Amministrazioni locali in tema di pubblicazione dei dati, promuovendo progetti pilota per l'uso dei dati pubblici, anche favorendo i processi di sviluppo e consolidamento di Living Lab sul tema delle applicazioni di nuova generazione basate sugli OD;
- f) coordinamento di attività per gli uffici regionali e per tutte le Amministrazioni locali interessate alla creazione della "nuvola" di dati del territorio pugliese (Cloud open data Puglia)
- g) ampliamento e adeguamento della piattaforma ICT (www.dati.puglia.it) in grado di mettere insieme le esperienze di interesse già realizzate con gli investimenti pubblici, rendere interoperante la piattaforma realizzata con sistemi analoghi già esistenti in altre regioni, personalizzare strumenti e funzioni al fine di una gestione complessiva del processo di produzione e pubblicazione degli Open Data
- h) realizzazione del framework di riferimento – basato sulla piattaforma ICT Linked Open Data realizzata – su cui abilitare lo sviluppo di nuovi progetti LOD su scala interregionale e nazionale

In attuazione del modello di governo del processo di pubblicazione dei dati aperti il Piano Annuale degli Open Data della Regione Puglia ha l'obiettivo di:

- a. stabilire le priorità di intervento per l'apertura dei dati
- b. individuare nel tempo le differenti categorie di dati da pubblicare
- c. indicare il formato di rappresentazione più idoneo alla pubblicazione dei Dataset
- d. indicare le licenze di riuso applicabili
- e. indicare le strutture regionali di servizio responsabili della produzione, aggiornamento e pubblicazione dei Dataset

Per dati di tipo aperto, ci si riferisce ai dati che presentano le seguenti caratteristiche:

1. **Completi:** I dati devono comprendere tutte le componenti (metadati) che consentano di esportarli, utilizzarli on line e off line, integrarli e aggregarli con altre risorse e diffonderli in rete.
2. **Primari:** Le risorse digitali devono essere 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;
3. **Tempestivi:** Gli utenti devono essere messi in condizione di accedere e utilizzare i dati presenti in rete in modo rapido e immediato, massimizzando il valore e l'utilità derivanti da accesso e uso di queste risorse;
4. **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
5. **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 processabili in automatico dal computer
6. **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

Piano Annuale degli Open Data regionali

esclusivo. Sono preferibili i formati con le codifiche più semplici e maggiormente supportati

7. **Liberi da licenze** che ne limitino l'uso: sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, e la diffusione o la redistribuzione
8. **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
9. **Ricercabili**: I dati devono essere facilmente identificabili in rete, grazie a cataloghi e archivi facilmente indicizzabili dai motori di ricerca
10. **Permanenti**: Le peculiarità fino ad ora descritte devono caratterizzare i dati nel corso del loro intero ciclo di vita
11. **Aggiornati**: a cura della struttura che li produce
12. **Soggetti a monitoraggio**: sulla pubblicazione dei dati, effettuato da parte della struttura che gestisce i Dataset

Il Piano sarà aggiornato su base annuale **attraverso un processo partecipativo** in cui sono coinvolte sia le strutture di Servizio interne all'amministrazione sia i diversi soggetti esterni interessati alla creazione di un **"ecosistema pugliese"** dei dati pubblici che sia in grado di supportare un ampio spettro di risultati positivi per l'intero territorio regionale: servizi ai cittadini più efficienti, migliore livello decisionale delle amministrazioni pubbliche, nuovi modelli di business per le imprese e, quindi, migliore qualità della vita. I dati saranno resi disponibili accompagnati dai relativi metadati, salvo specifiche e motivate eccezioni, indicate per ciascun Dataset da ciascun Servizio regionale nell'ambito dell'individuazione periodica dei dati che saranno rilasciati in formato aperto, in una logica di decentramento e distribuzione della produzione del dato e responsabilità dell'aggiornamento.

L'attività di individuazione dei dati oggetto di pubblicazione in formato aperto dovrà essere, in ogni caso, condotta in modo tale da escludere quelli che, per il tipo di riutilizzo o per le modalità con cui si intende realizzarlo, potrebbero violare:

- a. la sicurezza pubblica, la difesa nazionale, lo svolgimento di indagini penali o disciplinari
- b. il diritto di terzi al segreto industriale, statistico e commerciale, o altri vincoli di segretezza fissati in obblighi di legge
- c. i diritti di proprietà intellettuale
- d. il diritto alla protezione dei dati personali

In ogni caso, per assicurare la trasparenza amministrativa garantendo, al contempo, la protezione dei dati personali o coperti da segreto, l'amministrazione regionale procederà, quando necessario, alla pubblicazione di dati aggregati o resi anonimi in modo da non consentire alcuna identificazione, nemmeno indiretta, dei soggetti a cui tali dati si riferiscono, coerentemente con la normativa vigente in materia.

5 La piattaforma regionale e il formato dati

La Legge Regionale n. 20 del 2012 “*Norme sul software libero, accessibilità di dati e documenti e hardware documentato*” tratta i temi dell’Open Data specificatamente all’art. 6 “*Riutilizzo dei documenti e dei dati pubblici*”.

Le prime Linee di indirizzo in materia di Open Data sono state oggetto della DGR n. 2183 del 2012. Con la Legge Regionale n. 20 del 2012 in particolare è stato assicurato il coordinamento delle azioni volte a garantire la disponibilità dei dati generati da tutte le strutture regionali e dagli enti, agenzie e società indicate nell’art. 2 della stessa. A complemento dei passi normativi compiuti in materia di Open Data, in particolare, ai sensi della succitata Deliberazione della Giunta Regionale, si è provveduto alla creazione del portale <http://dati.puglia.it>, referenziato al portale nazionale *dati.gov.it*, nel rispetto di tutte le condizioni generali dettate dalla legge regionale in parola.

Si avvale della piattaforma CKAN. Il sistema CKAN, sviluppato da Open Knowledge Foundation, è un prodotto nato per la catalogazione di risorse aventi la natura di file accessibili tramite URL. Si tratta di una piattaforma onnicomprensiva, ben integrata e altamente personalizzabile, con cui si possono realizzare tutti gli elementi di un sistema di gestione di Open Data, dalla loro memorizzazione fisica, organizzazione logica, metadattazione e, infine, esposizione su un sito Web.

Essa copre i seguenti aspetti del processo di pubblicazione di un Dataset:

- a. sistema di redazione delle schede dei metadati;
- b. storicizzazione automatica;
- c. sistema di memorizzazione dei file;
- d. pubblicazione delle schede all'interno di un portale Web personalizzabile;
- e. funzioni di ricerca per chiave;
- f. sistema di esposizione dei dati come servizio mediante API di tipo REST;
- g. funzioni di anteprima per i tipi di formati più comuni;
- h. funzioni basilari di analisi statistica degli accessi;
- i. sistema minimale di quality assurance e feedback dagli utenti.

Il linguaggio con cui è stato programmato CKAN è Python, che per la realizzazione dei servizi sopra elencati si integra con un database PostgreSQL e un motore di ricerca Solr su application server Jetty (un Apache Web Server).

Grazie alla potenza di queste API, quindi, è possibile integrare le funzionalità del sistema di catalogazione all'interno di altri software. Ad esempio, è possibile inserire qualunque funzione di CKAN all'interno di un CMS moderno, svincolando in tal modo la fase di pubblicazione da quella di redazione.

Oltre a contenere il catalogo dei dataset - accessibile dall'area pubblica, con possibilità di ricerca e download dei dati secondo le diverse chiavi (organizzazioni, gruppi, tag, formati, licenze) - e le molteplici funzionalità operative, il portale regionale <http://dati.puglia.it> costituisce luogo di approfondimento, confronto e partecipazione attiva anche per richieste, segnalazioni e candidature di nuovi Dataset proposti da soggetti esterni all'amministrazione regionale.

Sulla piattaforma regionale confluirà in formato aperto tutto il notevole patrimonio informativo prodotto nel tempo dalla Regione Puglia, dalle sue Agenzie e Società in house e da tutti gli altri soggetti pubblici e privati controllati, vigilati o partecipati dalla Regione.

Piano Annuale degli Open Data regionali

L'avvio del portale è messo in relazione alla realizzazione di un modello di *open government* basato sul libero accesso ai dati pubblici come acceleratore economico per il territorio, alla connessione degli open data con il ciclo delle politiche pubbliche e all'incentivazione delle aziende al riutilizzo, anche a scopi commerciali, delle informazioni pubbliche e alla creazione di applicazioni innovative basate su questi dati.

A partire dall'agenda delle priorità individuate e attraverso una fase di accompagnamento (soprattutto negli stadi iniziali) atta a garantire la confluenza dei dati, tutte le strutture regionali, gli enti, agenzie e società indicate nell'art. 2 della L.R. n. 20/2012 potranno avviare il processo di liberazione dei dati in loro possesso, secondo le indicazioni del processo di pubblicazione contenute nella documentazione relativa alle Linee Guida *Linked Open Data* della Regione Puglia⁶.

Questa piattaforma rappresenta l'*hub* degli Open Data della Regione Puglia, oltre che delle Amministrazioni locali e di tutti i soggetti pubblici del territorio pugliese pronti ad essere coinvolti anche attraverso accordi che facciano riferimento alle modalità di adesione previste dalle Linee Guida Regionali.

Tale operazione permetterà, nel medio periodo, di far convergere questa gran mole di informazioni, secondo le diverse modalità previste in base alle scelte e al grado di maturazione sul tema presente sul territorio, unificarne il linguaggio, puntando alla definizione di un vocabolario comune regionale e rendere tali dati disponibili a tutti perché vengano utilizzati secondo le differenti finalità degli utenti, compresi quelli di carattere commerciale.

Nel corso dell'anno saranno avviate le attività di ampliamento e adeguamento della piattaforma ICT (www.dati.puglia.it) in grado di:

- a. mettere insieme le esperienze di interesse già realizzate con gli investimenti pubblici sul territorio pugliese (Aree Vaste, Patti per le Città, Living Labs, ecc.)
- b. renderla interoperante con sistemi analoghi già esistenti in altre regioni (Homer, OpenDAI)
- c. personalizzare strumenti e funzioni al fine di una gestione complessiva del processo di produzione e pubblicazione degli Open Data.

In linea con quanto previsto dalle Linee Guida nazionali per la valorizzazione del Patrimonio Informativo Pubblico e dalle Linee Guida Regionali "*Linked Open Data*" si sono già avviati una serie di interventi finalizzati al raggiungimento dell'obiettivo di definizione e modellizzazione del processo e della metodologia per la pubblicazione dei dati pubblici regionali in formato aperto, secondo i principi condivisi a livello internazionale e ripresi dalle Linee Guida.

Il formato aperto dei dati tuttavia è una condizione necessaria ma non sufficiente per poter definire un dato di tipo aperto; infatti quest'ultimo è definito sulla base di ulteriori elementi caratteristici quali i metadati, le licenze che ne consentono il riutilizzo anche per finalità commerciali, le tecnologie neutre che consentono di riutilizzare realmente tali documenti/dati⁷.

⁶ vedi Capitolo 2 "Pubblicazione degli Open Data sulla piattaforma della Regione Puglia" di "Linked Open Data - Linee Guida per la pubblicazione degli Open Data della Regione Puglia - Istruzioni Operative - Versione 1.0" Open Data - Regione Puglia, Maggio 2014 (http://www.dati.puglia.it/documents/portlet_file_entry/10184/LOD_LG_PrincipiGenerali_OpenData_v1.2.pdf/269fab92-dfe5-4d89-b9cf-91aaa2763eba)

⁷ vedi "Linee Guida Nazionali per la valorizzazione del Patrimonio Informativo Pubblico (Anno 2014) - Agenzia per l'Italia Digitale" (http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblico2014_v0.7finale.pdf)

A complemento della fase di analisi dei dati da compiersi nell'iter di produzione e pubblicazione degli stessi, secondo i criteri di individuazione e selezione indicati nelle Linee Guida Regionali, una volta verificato che i dati che si intende aprire siano effettivamente distribuibili sotto il profilo giuridico e di qualità adeguata - in riferimento al modello "a cinque stelle" di Tim Berners Lee⁸ - si punterà all'obiettivo di limitare la produzione di dati in formato da *uno a tre stelle* in favore della produzione di dati verso il formato **Linked**.

In questa fase di avvio, compatibilmente con il formato di partenza dei dati, si stanno già mettendo in campo azioni per realizzare l'obiettivo di pubblicazione sulla piattaforma regionale di **dataset almeno a tre stelle**, il formato più semplice di dati aperti, che offre la possibilità di effettuare elaborazioni sui dati senza esser costretti ad utilizzare software proprietario, fino alla definizione e utilizzo di ontologie e del vocabolario comune.

I dati proposti per la pubblicazione in formato aperto sulla piattaforma <http://dati.puglia.it> dovranno essere corredati da una serie di informazioni aggiuntive (di carattere descrittivo, tecnico, tematico con *tag*, amministrativo, ...) - i.c.d. *metadati* -. In particolare, in merito al processo di metadattazione di carattere tematico al fine di agevolare la massima integrazione e interoperabilità, tale aspetto è fortemente connesso alle aree tematiche individuate nel portale regionale dei servizi al cittadino ed è in corso, di concerto con le diverse strutture regionali coinvolte, un processo di definizione dell'universo dei *tag* possibili per ciascuna area tematica, al fine di semplificare la corretta classificazione dei data set. Questo permette, inoltre, ai motori di ricerca di rintracciare in maniera adeguata i data set collegati al tematismo richiesto.

Una delle caratteristiche imprescindibili nella rilevazione e diffusione dei dati aperti è la licenza con cui i dati stessi vengono distribuiti.

I dati pubblici saranno resi disponibili, ove possibile, in modalità elettronica e nei seguenti formati aperti al fine di abilitarne l'interoperabilità ed il riuso:

Nome	Tipo	Estensione
Comma Separated Value (CSV) 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)	Dato tabellare	.csv
Geographic Markup Language (GML) Formato XML utile allo scambio di dati territoriali di tipo vettoriale	Dato geografico vettoriale	.gml
GeoJSON E' un formato di testo aperto, per la codifica di oggetti geografici e dei correlati attributi non spaziali, scritto in JSON (JavaScript Object Notation)	Dato geografico vettoriale	.geojson, .topojson, o .json
Keyhole Markup Language (KML) Formato basato su XML creato per gestire dati territoriali in tre dimensioni.	Dato geografico vettoriale	.kml
Open Document Format per dati tabellari (ODS) Formato per l'archiviazione e lo scambio di fogli di calcolo	Dato tabellare	.ods

⁸ definisce un modello di catalogazione in base alle caratteristiche su una scala di valori che va da 1 (una stella ★) a 5 (cinque stelle ★★★★★) permettendo di individuare il livello di leggibilità dei dati pubblicati e la loro facilità d'accesso, vedi: Paragrafo 2 "Open Data" di "Linked Open Data - Linee Guida per la pubblicazione degli Open Data della Regione Puglia - Tutorial- Versione 1.0" Open Data - Regione Puglia, Maggio 2014

Piano Annuale degli Open Data regionali

Nome	Tipo	Estensione
Resource Description Framework (RDF) Basato su XML, e' 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	Dato strutturato	.rdf
ESRI Shapefile (SHP) Lo Shapefile ESRI è un popolare formato 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	Dato geografico vettoriale	.shp, .shx, .dbf, .prj
Extensible Markup Language (XML) E' 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)	Dato strutturato	.xml
Tab Separated Value (TSV) 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	Dato tabellare	.tsv

Tabella 1 - Formati di dato aperto

6 La licenza d'uso dei dati

Attenendosi al principio fondamentale condiviso di adottare licenze che permettano il riutilizzo del dato anche a fini commerciali nell'ottica del più ampio e libero riuso dell'informazione pubblica, considerato la scelta fatta dalla Regione Puglia con la Deliberazione della Giunta Regionale n. 2183 del 31 ottobre 2012, a ciascun Dataset è abbinata una licenza con cui viene pubblicato. Sarà associato al Dataset una licenza che garantisca la massima possibilità di riutilizzo dei dati, a meno di specificità che richiedano l'adozione di licenze meno aperte.

Saranno adottate quindi Licenze riconosciute a livello nazionale ed internazionale al fine di promuovere un ampio uso e riuso dei dati del sistema regionale, che siano compatibili con l'esigenza di permettere la libera circolazione anche del materiale creativo protetto dal diritto d'autore, dal momento che sono ispirate al modello "alcuni diritti riservati".

I Dataset presenti nel portale regionale <http://dati.puglia.it> saranno rilasciati per default con **licenza di distribuzione IODL 2.0⁹** (*Italian Open Data License*), che concede all'utente la possibilità di riprodurre, distribuire, trasmettere e adattare liberamente i dati, anche a scopi commerciali, a condizione che venga citata la fonte, tranne nei casi in cui sia esplicitamente dichiarata altra Licenza specifica.

Questa licenza è rivolta principalmente alle Pubbliche Amministrazioni ed è una licenza *Share-Alike*, cioè che impone di mantenere le opere derivate sotto la stessa licenza o una delle licenze compatibili. La licenza IODL è compatibile sia con ODbL sia con CC-BY-SA 3.0 (Creative Commons).

La licenza **ODbL** è parte delle licenze Open Data Commons (progetto di Open Knowledge Foundation) che, inizialmente definita per i dati cartografici OpenStreetMap, è diventata ben presto un punto di riferimento anche in ambito pubblico. La ODbL offre all'utente tre diritti fondamentali:

- a) **To share:** tale diritto consiste nell'offrire la possibilità di copiare, distribuire ed utilizzare il database
- b) **To create:** l'utilizzo della licenza ODbL garantisce la possibilità di lavorare e creare nuove opere a partire dal database fornito
- c) **To adapt:** è possibile modificare, trasformare e costruire opere derivate a partire dall'iniziale database

Le licenze **Creative Commons¹⁰** possono essere adottate in funzione della specificità e della natura dei dati da rendere disponibili in formato aperto, per cui è essenziale che ogni Dataset proposto dai Servizi regionali sia accompagnato dalla licenza che ne regola il riuso e la distribuzione (ad esempio, contenuti multimediali creati dall'amministrazione per i propri siti).

Le licenze Creative Commons utilizzabili per l'Open Data regionale sono:

Logo	Nome	Descrizione	Riferimento
------	------	-------------	-------------

⁹ vedi: Paragrafo 2.5 "Licenze d'uso dei dati della Regione Puglia" di "Linked Open Data - Linee Guida per la pubblicazione degli Open Data della Regione Puglia - Istruzioni Operative - Versione 1.0" Open Data - Regione Puglia, Maggio 2014; Capitolo 5 "Licenze d'uso dei dati" di "Linked Open Data - Linee Guida per la pubblicazione degli Open Data della Regione Puglia - Tutorial- Versione 1.0" Open Data - Regione Puglia, Maggio 2014

¹⁰ Creative Commons: <http://www.creativecommons.it/>

Piano Annuale degli Open Data regionali

	Attribuzione CC BY	Questa licenza permette a terzi di distribuire, modificare, ottimizzare ed utilizzare la tua opera come base, anche commercialmente, fino a che ti diano il credito per la creazione originale. Questa è la più accomodante delle licenze offerte. È raccomandata per la diffusione e l'uso massimo di materiali coperti da licenza.	Vedi il testo legale
	Attribuzione - Condividi allo stesso modo CC BY-SA	Questa licenza permette a terzi di modificare, ottimizzare ed utilizzare la tua opera come base, anche commercialmente, fino a che ti diano il credito per la creazione originale e autorizza le loro nuove creazioni con i medesimi termini. Questa licenza è spesso comparata con le licenze usate dai software open source e gratuite "copyleft". Tutte le opere basate sulla tua porteranno la stessa licenza, quindi tutte le derivate permetteranno anche un uso commerciale. Questa è la licenza usata da Wikipedia, ed è consigliata per materiali che potrebbero beneficiare dell'incorporazione di contenuti da progetti come Wikipedia e similari.	Vedi il testo legale
	Attribuzione - Non opere derivate CC BY-ND	Questa licenza permette la redistribuzione, commerciale e non, fintanto che viene trasmessa intera ed invariata, dandoti credito.	Vedi il testo legale
	Attribuzione - Non commerciale CC BY-NC	Questa licenza permette a terzi di modificare, ottimizzare ed utilizzare la tua opera come base per altre non commerciali, e benché le loro nuove opere dovranno accreditarti ed essere non commerciali, non devono licenziare le loro opere derivate con i medesimi termini	Vedi il testo legale
	Attribuzione - Non commerciale - Condividi allo stesso modo CC BY-NC-SA	Questa licenza permette a terzi di modificare, redistribuire, ottimizzare ed utilizzare la tua opera come base non commerciale, fino a che ti diano il credito e licenzino le loro nuove creazioni mediante i medesimi termini	Vedi il testo legale
	Attribuzione - Non commerciale - Non opere derivate CC BY-NC-ND	Questa licenza è la più restrittiva delle sei licenze principali, permettendo a terzi soltanto di scaricare le tue opere e condividerle ad altri fino a che ti diano il giusto credito, ma non possono cambiarle in nessun modo od utilizzarle commercialmente	Vedi il testo legale

Tabella 2 – Licenze Creative Commons

7 Il percorso del 2015

Nel corso del 2015 sarà adottata la metodologia di pubblicazione dei Dataset in formato aperto sviluppata dalla Regione Puglia e basata sostanzialmente sulla collaborazione attiva degli attori delle strutture coinvolte, in relazione al ruolo ricoperto secondo quanto previsto dalle istruzioni operative delle Linee Guida Regionali (analista e responsabile dei dati).

Tutto ciò al fine di assicurare le condizioni organizzative e i processi funzionali finalizzati a ottimizzare l'apertura dei Dataset individuati secondo i tempi e le frequenze di aggiornamento prevista, in esecuzione del presente piano.

Tale azione si avvale del coordinamento del Servizio Ricerca Industriale e Innovazione e del supporto della società in-house InnoVaPuglia.

Il percorso per l'anno 2015 si articola in una serie di Obiettivi e Attività specifiche per la realizzazione di tali obiettivi, secondo la finalità previste dalle indicazioni normative nazionali e regionali:

Obiettivo 1: Migliorare la qualità, la quantità e la disponibilità dei dati aperti attraverso la piattaforma regionale dati.puglia.it

Attività 1.1 – Revisione e ricognizione dei dati disponibili

Le azioni previste riguardano principalmente, oltre al consolidamento dei Dataset già pubblicati sulla piattaforma, anche la realizzazione di una ricognizione continuativa finalizzata all'individuare quegli insiemi di dati che possono essere candidati alla pubblicazione in modalità Open Data, sia già esposti sul sito istituzionale regionale a beneficio dei cittadini¹¹ sia su proposta dei *referenti interni* alle strutture coinvolte secondo un modello comune e coordinato di raccolta e di integrazione nell'ottica dei "linked data". Per quanto riguarda il primo caso, nell'Allegato 1 è presente un primo elenco di dataset relativi al rilascio in formato aperto dei dati pubblicati in ottemperanza agli obblighi di Trasparenza.

Attività 1.2 – Apertura dei dati relativi a priorità individuate

Il target posto in relazione ai dati e ai domini che costituiscono le priorità di intervento ravvisate riguarda:

- **Il coinvolgimento attivo di alcune Strutture regionali**
Pur garantendo la possibilità di rendere pubblico il proprio patrimonio informativo a qualunque ufficio che volesse candidare propri dataset, i target prioritari di intervento concordati per l'avvio della sperimentazione della metodologia regionale in avvio nel 2015 sono descritti nell'**Allegato 1** e coinvolgono le seguenti strutture:
 - a. Servizio Ricerca Industriale e innovazione
 - b. Servizio Programmazione sociale ed integrazione socio-sanitaria
 - c. Servizio Politiche di Benessere sociale e Pari Opportunità
 - d. Servizio Sistemi informativi e investimenti in sanità
 - e. Servizio Turismo
 - f. Servizio Beni culturali
 - g. Servizio Sport per tutti

¹¹ ad es. in ottemperanza al Decreto Legislativo 14 marzo 2013, n. 33 "Riordino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni" nella sezione del sito istituzionale "Amministrazione trasparente"

- h. Servizio Assetto del Territorio
- i. Ufficio Statistico Regione Puglia

Come si legge nell'Allegato 1, i Dataset aggregati in "Amministrazione Trasparente" fanno riferimento alla responsabilità di diverse strutture regionali. Per l'anno 2015, quindi, saranno rilasciati in formato aperto una prima parte dei dati già disponibili e pubblicati nel portale regionale, a cura delle singole strutture indicate in tabella nella loro qualità di responsabili/titolari del dato..

Nel corso dell'anno, l'attività ed il confronto con le strutture regionali potrà far emergere l'opportunità e la possibilità di rilasciare ulteriori Data Set rispetto a quelli elencati in Allegato 1.

Il documento "Agenda nazionale per la valorizzazione del patrimonio informativo pubblico (anno 2014)"¹² nel percorso di valorizzazione del patrimonio pubblico informativo proposto ha tra le direttrici principali i "dati di tipo aperto", riportando alcuni dataset da rilasciare indicati come "chiave", tra i quali alcuni relativi a "Dati trasversali", "Educazione", "Dati territoriali" per cui si individuano tra le amministrazioni di competenza anche le Regioni¹³. Condividendo la logica seguita dall'Agenda Nazionale, nel corso del 2015 saranno rilasciati in formato aperto – oltre quelli in elenco - i dati che fanno riferimento agli obblighi di Trasparenza, riconducibili alla tipologia "Dati trasversali". Sarà possibile, inoltre, rendere disponibili nel portale open data regionale i dati territoriali già pubblicati in formato aperto nel Sistema Informativo Territoriale (SIT), organizzandoli secondo la logica prevista per gli Open Data.

Attività 1.3 – Avvio sperimentazione di apertura dati di Enti, agenzie e società regionali¹⁴ e degli Enti locali e territoriali

Dal confronto con gli attori degli enti competenti individuati per l'avvio di tali azioni, l'intervento sperimentale si svilupperà secondo le seguenti azioni:

- a) analisi, identificazione e estrazione dei dati relativi ai sistemi informativi regionali secondo le strategie di liberazione dei dati concordate con i referenti interni delle strutture di riferimento;
- b) verifica - attraverso "progetti pilota" - della metodologia regionale a supporto del ciclo di vita degli open data con alcuni Enti locali e territoriali sugli ambiti prioritari individuati, tenendo conto anche delle tipologie di informazioni maggiormente richieste da cittadini, professionisti e imprese¹⁵, secondo quanto previsto nelle Linee Guida Regionali.

¹² rilasciato ai sensi delle disposizioni dell'art. 9 del D.L. n. 179/2012 (ha interamente riscritto l'art. 52 del Codice dell'Amministrazione Digitale), indirizza specifiche azioni riferibili a quelle disposizioni del capo V del CAD dedicate alla disponibilità, alla fruibilità, alle modalità di accesso e riutilizzo dei dati delle pubbliche amministrazioni, nonché a quelle relative alle basi dati di interesse nazionale

¹³ vedi Tabella 1 pag. 15 di "Agenda nazionale per la valorizzazione del patrimonio informativo pubblico (anno 2014)"

¹⁴ indicate nell'art. 2 della Legge Regionale n.20 del 2012 "Art. 2 - Enti interessati 1. La presente legge si applica alle azioni e agli interventi della Regione Puglia e degli enti pubblici e privati, comunque costituiti, sui quali la medesima esercita poteri di indirizzo e controllo nelle materie oggetto della presente legge, ivi inclusi – a titolo meramente esemplificativo e non esaustivo - enti, aziende, società, agenzie, istituzioni, consorzi e organismi comunque denominati, controllati e/o vigilati e/o partecipati dalla Regione Puglia.

¹⁵ Nel par. 6.3 "Linee Guida per i siti Web della PA Vademecum Open Data - Come rendere aperti i dati delle pubbliche amministrazioni – Versione Beta 2011", pubblicate ai sensi dell'art. 4 Direttiva del Ministro per la Pubblica Amministrazione e l'Innovazione del 26 novembre 2009, n. 8 (<http://www.dat.gov.it/sites/default/files/VademecumOpenData.pdf>) si riporta l'esito di un'analisi effettuata dall'Associazione Italiana per l'Open Government. L'Agenzia per l'Italia Digitale ha raccolto nei mesi scorsi diverse richieste di apertura di specifiche tipologie di dati di valore aggiunto, riportate nell'Appendice III del documento "Agenda Nazionale per la valorizzazione del patrimonio informativo pubblico(Anno 2014)" e di cui ha tenuto conto nell'individuazione delle priorità, cui si fa riferimento nell'Attività 2 del presente piano; inoltre, a febbraio 2015 l'ha reso noto gli esiti della consultazione pubblica sulle tipologie di dati da rendere disponibili secondo i principi dell'Open ~~Data~~ (http://www.agid.gov.it/sites/default/files/linee_guida/segnalazioni_dati_in_open_20150212.pdf)

Si tratta di interventi strettamente correlati alla previsione e realizzazione di iniziative mirate, tese a diffondere la cultura degli Open Data presso i Servizi regionali e di altri Enti e strutture pubbliche con cui ci si propone di costruire un modello comune e coordinato di raccolta e integrazione nell'ottica anche dei "Linked data".

Obiettivo 2: Promozione e diffusione di un utilizzo attivo dei dati aperti presenti sulla piattaforma regionale dati.puglia.it

Attività 2.1 – Rafforzamento della cultura degli open data attraverso iniziative di approfondimento e partecipazione

In linea con l'Obiettivo 3.6¹⁶ dell'“Agenda nazionale per la valorizzazione del patrimonio informativo pubblico (anno 2014)”, l'attività pianificata di supporto alla diffusione degli Open Data regionali passa attraverso iniziative di promozione informativa e di carattere tecnico (seminari, workshop, laboratori) rivolte principalmente ai dipendenti delle Pubbliche Amministrazioni, aventi finalità di accrescimento delle competenze e degli skill, ma anche di stimolo alla partecipazione a “comunità di pratiche”, anche nell'ottica di poter apprendere le procedure idonee per la produzione e pubblicazione di dati in formato aperto su *dati.puglia.it*.

Inoltre, saranno attivate iniziative mirate (mediante workshop, community on line) a rafforzare il coinvolgimento e partecipazione dei sistemi regionali dell'innovazione e della ricerca, per un confronto costante finalizzato a incentivare l'affinamento, la consultazione e l'utilizzo dei dati di maggiore interesse presenti in piattaforma.

Di seguito una possibile proposta di iniziative schedate per l'anno 2015

N° e Tipologia di eventi	Descrizione	Target	Periodo
n.1 seminario informativo	Finalità: promozione, coinvolgimento operativo, diffusione cultura open data	Strutture regionali, enti, agenzie e società indicate nell'art. 2 della Legge Regionale n.20 del 2012	Luglio 2015
n. 1 laboratorio partecipativo territoriale	Finalità: promozione, coinvolgimento operativo, diffusione cultura open data, comunità	Enti locali	Settembre 2015
n. 1 workshop	Finalità: confronto su fabbisogni, possibili applicazioni	Imprese innovative, distretti, Università, Sistema della ricerca	Ottobre 2015

Attività 2.2 – Supporto all'avvio di sperimentazioni di applicazioni innovative

¹⁶ “OBIETTIVO 3.6 – al fine di assicurare che i dati resi disponibili dalle amministrazioni siano effettivamente fruibili e largamente riutilizzati, le amministrazioni intraprendono azioni di coinvolgimento nei confronti della società civile”.

Piano Annuale degli Open Data regionali

Al fine di promuovere e sostenere la condivisione e il riuso dei dati aperti pubblicati si sostanzierà nella realizzazione di manifestazioni pubbliche di carattere partecipativo dedicate a cittadini, sviluppatori, imprese innovative, start-up giovanili e associazioni per la sperimentazione di applicazioni inedite basate sul riuso dei dati aperti, quali “*hackathon*” condotti su ambiti specifici e un contest dedicato aperto al territorio.

N° e Tipologia di eventi	Descrizione	Target	Periodo
n.1 hackathon	Finalità: realizzazione e sperimentazione di applicazioni inedite su ambiti tematici specifici	Cittadini, sviluppatori, associazioni civiche, start up	Settembre 2015
n.1 contest	Finalità: sostegno a sviluppo di soluzioni e applicazioni innovative basate sull'utilizzo di dati pubblici presenti sulla piattaforma regionale	Cittadini, imprese, pubbliche amministrazioni, Università	Novembre 2015

A completamento del supporto al potenziamento dell'ecosistema pugliese in ambito open data, in linea con l'Agenda Digitale Regionale e con quanto previsto nell'ambito della nuova programmazione, attraverso il POR PUGLIA 2014 – 2020¹⁷, si prevede anche l'avvio di **azioni pilota** riconducibili al Centro di Competenza regionale sugli Open Data.

Il Centro di competenza è presidio plurale e luogo aperto in cui potranno confluire sinergicamente le competenze diffuse del territorio per sostenere la diffusione della cultura degli Open Data, la loro valorizzazione, la promozione della domanda pubblica di informazione e innovazione.

Il Centro di competenza sugli Open Data e tutta la larga gamma dei soggetti che potrà aggregare potrà consolidare l'approccio collaborativo del “fare rete” e divenire punto di riferimento per la generazione di nuove competenze per sviluppare inediti processi di analisi, razionalizzazione e condivisione dei dati e/o realizzazione di nuovi servizi di business e/o innovazione e ricerca. Un laboratorio regionale di innovazione sul tema del riuso dei dati pubblici in modalità Open, aperto alle continue evoluzioni delle tecnologie digitali ed al dinamico mutare dei fabbisogni pubblici, flessibile per potersi adeguare a molteplici necessità organizzative.

In particolare, le azioni pilota previste sono:

- Open data e cittadino digitale
Si tratta di interventi che, anche attraverso iniziative di co-progettazione, co-design, co-working, coinvolgono i cittadini quali fornitori di feed-back sui dati, produttori di nuovi dati e contenuti, agenti di cambiamento della PAL, nell'ottica di una democrazia partecipativa.
- Living lab per lo sviluppo delle applicazioni e i servizi basati sugli open data
Si tratta di laboratori volti a promuovere nuove iniziative imprenditoriali per lo sviluppo di app e servizi, in ambiente open source, e basati sul patrimonio informativo della PA. In tal senso si privilegeranno iniziative operanti nell'ambito della più complessiva programmazione 2014-2020 con particolare riferimento alle smart communities e smart specialization regionale

¹⁷ Indicare quali tipologie di azioni 2.3 relative all'Obiettivo specifico “Potenziare la domanda ICT di cittadini e imprese in termini di utilizzo dei servizi on line, inclusione digitale e partecipazione in rete” nell'ambito dell'Asse prioritario II “Migliorare l'accesso, l'impiego e la qualità delle TIC”, DGR n.1498 del 17 luglio 2014 “Approvazione Programma Operativo Regionale Puglia 2014-2020”, pag. 29481 del BURP n.112 del 20-08-2014

Piano Annuale degli Open Data regionali

L'incremento dei dataset pubblicati sulla piattaforma regionale permetterà di poter eseguire un adeguato **monitoraggio**, operando un'osservazione costante dell'interazione registrata sulla piattaforma regionale, attraverso le statistiche di utilizzo, le candidature proposte, le segnalazioni e le attività di carattere social (quali commenti, giudizi, suggerimenti). Ciò sarà funzionale, da un lato, a migliorare la qualità e l'accessibilità dei dati pubblicati, dall'altro a valutare il grado di diffusione raggiunto dai dati aperti e gli ambiti di maggiore interesse e riuso.

Allegato 1

Elenco dei Dataset in pubblicazione sul portale dati.puglia.it

Di seguito si riporta – suddiviso per Area Tematica - un primo elenco di riferimento dei dataset previsti nel Piano 2015:

Area Tematica: Ricerca e Sperimentazione

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A) ¹⁸	Disponibilità
La ricerca nelle Grandi Imprese: i Contratti di programma	Dati dei programmi di investimento promossi da grandi imprese anche in associazione con PMI	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
La ricerca nelle Medie imprese: i Programmi Integrati di Agevolazione	Dati dei programmi di investimento promossi da medie imprese anche in associazione con PMI	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
La ricerca industriale nelle PMI: il Bando Ricerca per le PMI (2008)	Dati dell'intervento a sostegno degli investimenti delle PMI per la ricerca e lo sviluppo tecnologico	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
L'innovazione: Aiuti alle piccole imprese innovative operative e di nuova costituzione	Dati degli interventi di valorizzazione industriale di attività di ricerca nell'ambito dei «settori industriali innovativi»	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
La ricerca collaborativa: La rete dei Laboratori Pubblici	Dati sulle Reti di laboratori pubblici di ricerca, espressione della capacità di offerta tecnologica del sistema della ricerca pubblica regionale.	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
La filiera della conoscenza e le Alleanze per l'innovazione: gli Uffici ILO	Anagrafica della Rete degli uffici di trasferimento tecnologico (ILO) della Puglia	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
La ricerca	Dati su Partenariati regionali	Servizio Ricerca	A	Luglio 2015

¹⁸ si fa riferimento ai dataset già pubblicati sulla piattaforma dati.puglia.it prima dell'approvazione del presente piano.

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A) ¹⁸	Disponibilità
collaborativa: i partenariati regionali dell'innovazione	dell'innovazione per la promozione della specializzazione intelligente del sistema socio-economico regionale (Imprese, Organismi di ricerca e Università)	Industriale e Innovazione		
Il Bando per l'erogazione di servizi per le spin off ad alto contenuto di conoscenza	Dati sul bando Spin off per sostenere la nascita e il consolidamento di imprese innovative gemmate nell'ambito della ricerca pubblica "già costituite" o "in via di costituzione", con finalità di valorizzazione imprenditoriale dei risultati della ricerca pubblica pugliese.	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
L'Accordo di Programma Quadro in materia di Ricerca: i Progetti Strategici	Dati dei progetti per contribuire allo sviluppo regionale attraverso attività di ricerca precompetitiva, Ricerca industriale e formazione, in 5 ambiti tematici	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
L'Accordo di Programma Quadro in materia di Ricerca: i Progetti Esplorativi	Dati dei progetti con l'obiettivo di condurre studi di fattibilità e sperimentare innovazioni di processo e/o prodotto di interesse per le imprese	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
L'ITC: gli Aiuti per la diffusione delle tecnologie dell'informazione e comunicazione nelle PMI (bando 2010)	Dati dei progetti e delle imprese interessate all'intervento TIC per l'acquisizione di beni e servizi di consulenza specialistica, con particolare riferimento ai servizi digitali avanzati.	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
L'ITC: gli Aiuti per la diffusione delle tecnologie dell'informazione e comunicazione nelle PMI (bando 2011)	Dati dei progetti e delle imprese interessate all'intervento TIC per l'acquisizione di beni e servizi di consulenza specialistica, con particolare riferimento ai servizi digitali avanzati.	Servizio Ricerca Industriale e Innovazione	A	Luglio 2015
Aiuti ai servizi per	Esiti del bando, progetti	Servizio Ricerca	R	Settembre

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)¹⁸	Disponibilità
l'innovazione tecnologica delle PMI - Bando 2013 (Passpartout Smartpuglia)	finanziati e soggetti beneficiari	Industriale e Innovazione		2015
Apulian ICT Living Labs	Fabbisogni pubblici, Progetti finanziati e soggetti beneficiari	Servizio Ricerca Industriale e Innovazione	R	Ottobre 2015
Living Labs Smart Puglia 2020	Fabbisogni pubblici, Progetti finanziati e soggetti bebeneficiari	Servizio Ricerca Industriale e Innovazione	R	Ottobre 2015
Distretti Tecnologici	Anagrafica dei DT dei soggetti aderenti e obiettivi di sviluppo	Servizio Ricerca Industriale e Innovazione	R	Luglio 2015
Distretti Produttivi	Anagrafica dei DP e dei soggetti aderenti e dati dei Piani di sviluppo	Servizio Ricerca Industriale e Innovazione	R	Luglio 2015
La Puglia delle Key Enabling Technologies	Esiti lavoro di indagine sull'utilizzo delle KETs nel sistema imprenditoriale pugliese	ARTI	R	Ottobre 2015

Tabella 3 – Dataset Ricerca e Sperimentazione


Area Tematica: Puglia Sociale

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
Registri Regionali delle strutture e servizi socio-assistenziali pubblici e privati autorizzati	Tutte le strutture e i servizi che erogano prestazioni sociosanitarie organizzabili per territorio comunale, provinciale, distretti sociali (ATO), tipologia di servizio, tipologia di persone a cui si rivolgono	Servizio Politiche di Benessere Sociale e Pari Opportunità	R	Giugno 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
Bandi Assistenza Indiretta Personalizzata (A.I.P.) e Assegni di Cura (ASCURA)	Stato generale della popolazione affetta da gravi o gravissime patologie, che sono o meno prese in carico da strutture sanitarie o dai servizi sociali (anche su base territoriale)	Servizio Programmazione sociale ed integrazione socio-sanitaria	R	Luglio 2015
Bando Rete per l'occupazione e i Servizi di Assistenza (R.O.S.A.)	N. di nuclei familiari percettori di incentivi per l'assunzione di assistenti familiari (iscritti negli elenchi speciali del Progetto), suddivisi per territori provinciali e fasce di reddito	Servizio Politiche di benessere sociale e Pari Opportunità	R	Giugno 2015
Bando Prima Dote	Elenco dei nuclei familiari con minori a carico con bassi redditi	Servizio Politiche di benessere sociale e Pari Opportunità	R	Giugno 2015
Bandi Offerta Servizi di Conciliazione	Costituzione dei nuclei familiari con minori anziani o disabili a carico, relazione con stato economico delle famiglie e della richiesta di servizi correlati alla formazione del nucleo familiare; offerta dei servizi rivolti a categorie di persone organizzati su base territoriale	Servizio Programmazione sociale ed integrazione socio-sanitaria Servizio Politiche di benessere sociale e Pari Opportunità	R	Dicembre 2015

Tabella 4 – Dataset Puglia Sociale

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
Studi medici di base	Dati dei medici di medicina generale e dei pediatri di libera scelta con incarico di titolarità registrati da parte delle Aziende Sanitarie Locali nel sistema informativo...	Servizio Sistemi informativi e investimenti in sanità	A	Luglio 2015
Continuità Assistenziale	Dati delle sedi di continuità assistenziale registrati dalle Aziende Sanitarie Locali nell'Area Applicativa Anagrafe delle Strutture Sanitarie del sistema informativo sanitario...	Servizio Sistemi informativi e investimenti in sanità	A	Luglio 2015
Farmacie territoriali	Dati delle farmacie territoriali estratti dall'area applicativa Anagrafe delle Strutture Sanitarie del sistema informativo sanitario regionale della Regione Puglia (Edotto)	Servizio Sistemi informativi e investimenti in sanità	A	Luglio 2015
Anagrafiche Strutture Sanitarie	Dati con relative codifiche delle strutture sanitarie (ospedali, reparti, pronto soccorso, ambulatori, consultori)	Servizio Sistemi informativi e investimenti in sanità	R	Ottobre 2015
Popolazione assistita	Dati aggregati sulla popolazione assistita estratti dall'anagrafe assistiti del sistema Edotto (popolazione assistita per distretto, distribuzione esenzioni, ecc.)	Servizio Sistemi informativi e investimenti in sanità	R	Ottobre 2015
Assistenza Ospedaliera	Dati aggregati sulle attività di ricovero	Servizio Sistemi informativi e investimenti in sanità	R	Ottobre 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
Assistenza farmaceutica	Dati aggregati sulla assistenza farmaceutica	Servizio Sistemi informativi e investimenti in sanità	R	Ottobre 2015
Assistenza specialistica	Dati aggregati sulla assistenza specialistica (ospedaliera, ambulatoriale pubblica e privata accreditata)	Servizio Sistemi informativi e investimenti in sanità	R	Ottobre 2015
Mobilità sanitaria	Dati aggregati sulla mobilità sanitaria interregionale e infra-regionale	Servizio Sistemi informativi e investimenti in sanità	R	Ottobre 2015
Enti proprietari/gestori di impianti sportivi pugliesi	Dati relativi agli enti pubblici e privati proprietari o gestori degli impianti sportivi della Puglia	Servizio Sport per Tutti	R	Giugno 2015
Impianti sportivi pugliesi	Dati relativi agli impianti sportivi ubicati nel territorio pugliese	Servizio Sport per Tutti	R	Giugno 2015
Spazi di attività	Dati relativi agli spazi di attività degli impianti sportivi ubicati nel territorio pugliese	Servizio Sport per Tutti	R	Giugno 2015
Enti pubblici e Organizzazioni sportive iscritte per la richiesta di contributi economici	Dati relativi alle organizzazioni iscritte per la richiesta di contributi economici	Servizio Sport per Tutti	R	Luglio 2015

Tabella 5 – Dataset Promozione della Salute

Piano Annuale degli Open Data regionali

Area Tematica: Cultura e Turismo

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
Movimento Turistico 2014	Flussi turistici dichiarati dalle strutture ricettive, alberghiere ed extralberghiere, per mese, comune, tipologia di struttura provenienza, sesso, fasce di età	Pugliapromozione – Osservatorio del Turismo/ Servizio Turismo	R	Giugno 2015
Elenco delle strutture ricettive (circa 5.000 strutture)	Strutture ricettive alberghiere ed extralberghiere: dati anagrafici con foto, Tipologia (categoria e classificazione), Servizi, Attrezzature, Prezzi, Periodi di apertura, Coordinate geo	Pugliapromozione/ Servizio Turismo	R	Giugno 2015 Ottobre 2015
Consistenza Ricettiva Alberghiera 2014	Dettaglio dell'offerta turistica alberghiera al 31/12/2014 per comune, tipologia e categoria ricettiva	Pugliapromozione/ Servizio Turismo	R	Giugno 2015
Consistenza Ricettiva Extralberghiera 2014	Dettaglio dell'offerta turistica extralberghiera al 31/12/2014 per comune, tipologia e categoria ricettiva	Pugliapromozione/ Servizio Turismo	R	Giugno 2015
Elenco degli Stabilimenti Balneari	Stabilimenti balneari: dati anagrafici, Servizi, Prezzi, Periodi di apertura, Coordinate geo	Pugliapromozione/ Servizio Turismo	R	Giugno 2015 Ottobre 2015
Servizi per il turismo, tra cui: servizi di trasporto interno e autonoleggi, guide turistiche, servizi di interpretariato, centri sportivi, parchi divertimento, centri benessere e termali,	Dati anagrafici con foto, Servizi, Coordinate geo	Pugliapromozione/ Servizio Turismo	R	Giugno 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
centri congressi ed eventi, attrattori enogastronomici, ristoranti				
Eventi	Dati anagrafici, date, orari, luoghi, foto, testo	Pugliapromozione / Teatro Pubblico Pugliese-PugliaEvents	R	Giugno 2015
Luoghi di interesse turistico, culturale, naturalistico (es. Castelli, musei, biblioteche, Parchi, Masserie didattiche, ecc.)	Dati anagrafici, Caratteristiche, Orari di apertura, testo, Servizi/Impianti/Attrezzature	www.viaggiareinpuglia.it/ Pugliapromozione	R	Giugno 2015
Digital Library della Regione Puglia	Risorse digitali presenti nell'archivio digitale multimediale del patrimonio culturale regionale	Servizio Beni Culturali / Servizio Beni Culturali	R	Ottobre 2015

Tabella 6 – Dataset Cultura e Turismo

Area Tematica: Statistica ufficiale e Sistema statistico nazionale

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
Popolazione residente nei comuni pugliesi	Popolazione residente nei comuni pugliesi per età e sesso della popolazione dei comuni pugliesi desunti da DEMO.ISTAT (Archivio Statistico derivante dalle indagini	Ufficio Statistico	R	Giugno 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
	effettuate presso gli Uffici di Anagrafe dei Comuni) aggiornati all'ultimo anno disponibile			
Densità abitativa della popolazione residente	Dati desunti dall'archivio DEMO-ISTAT all'ultimo anno disponibile	Ufficio Statistico	R	Giugno 2015
Bilancio demografico e della popolazione a livello comunale	L'archivio contiene i dati relativi ai movimenti anagrafici dei residenti nei comuni. Le informazioni sono ricavate dall'archivio DEMO-ISTAT (nascite, morti, iscrizioni e cancellazioni, l'ammontare al 31 dicembre, della popolazione residente distinta per comune di residenza, numero famiglie, numero convivenze e numero medio componenti per famiglia all'ultimo anno disponibile).	Ufficio Statistico	R	Giugno 2015
Indicatori di invecchiamento della popolazione nei comuni pugliesi	Indicatori di invecchiamento demografico della popolazione residente nei comuni pugliesi. Indicatori: percentuale di residenti di 65 anni ed oltre sul totale dei residenti, indice di dipendenza demografica, indice di vecchiaia, indice di ricambio della popolazione in età lavorativa, classe di età mediana all'ultimo anno disponibile.	Ufficio Statistico	R	Giugno 2015
Popolazione straniera residente per sesso e cittadinanza.	L'archivio contiene l'ammontare degli stranieri, distinto per sesso, per paese, continente e area geografica di cittadinanza	Ufficio Statistico	R	Giugno 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
	per singolo comune. I dati sono ricavati dall'archivio DEMO-ISTAT			
Classificazione aggregati territoriali	Le informazioni sono ottenute dalle classificazioni territoriali Istat (Comuni, Zone altimetriche, Risorse turistiche), dalle classificazioni di Regione Puglia (ASL, Zone socio-sanitarie, Centri per l'impiego) e dalle classificazioni dell'agenzia delle entrate (Codice catastale Comuni)	Ufficio Statistico	R	Giugno 2015
Conti economici Regione Puglia	Serie storica del PIL e sue componenti per la Puglia all'ultimo anno disponibile	Ufficio Statistico	R	Giugno 2015
Imprese/unità locali, forma giuridica e addetti per settore di attività economica a livello comunale	Imprese/unità locali e relativi addetti per settore di attività forma giuridica e comune della Puglia Archivio ASIA (ultimi dati disponibili)	Ufficio Statistico	R	Giugno 2015
IX Censimento Industria e servizi: Istituzioni non profit per natura giuridica e comune	Le informazioni provengono dal 9° Censimento generale dell'industria e dei servizi - Rilevazione sulle istituzioni non profit: Il Censimento rileva le istituzioni non profit e le loro unità locali alla data del 31 dicembre 2011. L'archivio contiene i dati sulle istituzioni con sede in Puglia per natura giuridica, comune e numeri di addetti per tipologia	Ufficio Statistico	R	Dicembre 2015
IX Censimento Industria e servizi: Istituzioni non profit per attività economia prevalente e comune	Le informazioni provengono dal 9° Censimento generale dell'industria e dei servizi - Rilevazione sulle istituzioni non profit: Il	Ufficio Statistico	R	Dicembre 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
	Censimento rileva le istituzioni non profit e le loro unità locali alla data del 31 dicembre 2011. L'archivio contiene i dati sulle istituzioni con sede in Puglia per attività economica prevalente, comune e addetti			
IX Censimento "Industria e servizi": unità locali delle imprese per comune e settore di attività economica	Le informazioni provengono dal 9° Censimento generale dell'industria e dei servizi. Il Censimento rileva le imprese e le loro unità locali alla data del 31 dicembre 2011. L'archivio contiene i dati delle unità locali di impresa per comune, settore di attività economica (ateco 2007) e su addetti e dipendenti	Ufficio Statistico	R	Dicembre 2015
IX Censimento "Industria e servizi": imprese per comune, tipologia e classe di addetti. Anno 2011	Le informazioni provengono dal 9° Censimento generale dell'industria e dei servizi. Il Censimento rileva le imprese alla data del 31 dicembre 2011. L'archivio contiene i dati delle imprese, localizzate in Puglia e dei relativi addetti, per tipologia (artigiana e non artigiana), classe di addetti e comune	Ufficio Statistico	R	Dicembre 2015
IX Censimento "Industria e servizi": imprese per comune, settore di attività economica e forma giuridica. Anno 2011	Le informazioni provengono dal 9° Censimento generale dell'industria e dei servizi. Il Censimento rileva le imprese alla data del 31 dicembre 2011. L'archivio contiene i dati delle imprese, localizzate in Puglia, per settore di attività economica (ateco 2007) e forma giuridica,	Ufficio Statistico	R	Dicembre 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
	addetti e dipendenti			
IX Censimento "Industria e servizi" istituzioni pubbliche per forma giuridica e comune	Le informazioni provengono dal 9° Censimento generale dell'industria e dei servizi. Il Censimento rileva le istituzioni pubbliche e le loro unità locali alla data del 31 dicembre 2011	Ufficio Statistico	R	Dicembre 2015
6° Censimento dell'agricoltura: aziende agricole e superficie agricola utilizzata per comune di localizzazione del centro aziendale e tipo di coltivazione	L'archivio contiene i dati relativi alle aziende agricole pugliesi (numero e superficie agricola utilizzata) per comune di localizzazione del centro aziendale e tipo di coltivazione, alla data del 6° Censimento dell'agricoltura (24 ottobre 2010)	Ufficio Statistico	R	Dicembre 2015

Tabella 7 – Dataset Statistica ufficiale e Sistema Statistico Nazionale

Area Tematica: Territorio e Aree Protette

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Disponibilità
Carta Tecnica Regionale	Confini Regionali, Confini Provinciali, Confini Comunali, Porti, Aeroporti, Reticolo Ferroviario, Reticolo Stradale Reticolo Idrografico, Edifici, Curve di Livello, ..	SIT/Servizio Assetto del Territorio	R	Settembre 2015
Elementi estratti dall'uso del suolo	Superfici Artificiali, Superfici Agricole, Superfici Boscate ed altri Ambienti Naturali, Superfici Ambiente Umido e Acque, dati statistici	SIT/Servizio Assetto del Territorio	R	Settembre 2015

Piano Annuale degli Open Data regionali

Catasti	Geositi, Grotte, Poseidonieti	SIT/Servizio Assetto del Territorio	R	Dicembre 2015
Rilievi Dati di Traffico	Dati di Traffico sulle principali arterie extraurbane	SIT/Servizio Assetto del Territorio	R	Ottobre 2015
Parchi, aree naturali protette, siti di importanza rilevante	Parchi e Riserve Naturali Regionali, Parchi e Riserve Nazionali, Riserve e Aree Naturali Marine, Siti UNESCO, Siti di Interesse Nazionale (SIN), Zone IBA, Siti di Importanza Comunitaria (SIC), Siti di Importanza Comunitaria Mare (SIC-Mare), Zone a Protezione Speciale (ZPS)	SIT/Servizio Assetto del Territorio	R	Novembre 2015
Dati della pianificazione territoriale/ paesaggistica regionale	Luoghi Panoramici, Strade a Valenza Paesaggistica, Strade Panoramiche, ...	SIT/Servizio Assetto del Territorio	R	Ottobre 2015

Tabella 8 – Dataset Territorio e Aree Protette

Amministrazione Trasparente

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Tempi previsti
Elenco dei dipendenti con recapiti telefonici e di posta elettronica	Pubblicazione prevista dall'art.13 del d.lgs. n. 33/2013 relativa all'articolazione degli uffici	Servizio Personale e Organizzazione	R	Luglio 2015
Strutture regionali	Pubblicazione prevista dall'art. 13 del d.lgs. n. 33/2013 relativa all'articolazione degli uffici	Servizio Personale e Organizzazione	R	Luglio 2015
Elenco PEC	Pubblicazione prevista dall'art. 13 del d.lgs. n. 33/2013 relativamente alle caselle di Posta	Servizio Personale e Organizzazione	R	Luglio 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Tempi previsti
	Elettronica Certificata (PEC)			
Elenco dei Dirigenti	Pubblicazione prevista dall'art. all'art. 10, c. 8, lett. d), art. 15, c. 1,2,5, art. 41, c. 2,3 del d.lgs. n. 33/2013 relativa ai titolari di incarichi dirigenziali	Segretariato generale della Giunta / Servizio Personale e Organizzazione	R	Settembre 2015
Tassi di assenza	Pubblicazione prevista dall'art. 16 del d.lgs. n. 33/2013, i dati si riferiscono al personale per l'anno 2014	Servizio Personale e Organizzazione	R	Settembre 2015
Incarichi conferiti o autorizzati ai dipendenti	Pubblicazione prevista dall'art. 18 del d.lgs. n. 33/2013, i dati si riferiscono al personale per l'anno 2014	Servizio Personale e Organizzazione	R	Settembre 2015
Enti controllati	Pubblicazione prevista dall'art. 22 del d.lgs. n. 33/2013 relativa agli enti pubblici vigilati, società partecipate, enti di diritto privato controllati	Enti controllati / Servizio Controlli	R	Luglio 2015
Elenco delle Deliberazioni della Giunta Regionale	Pubblicazione prevista dall'art. 23 del d.lgs. n. 33/2013 relativa al periodo 2010 - 2015	Segretariato generale della Giunta	R	Ottobre 2015
Elenco degli Atti Dirigenziali della Giunta Regionale	Pubblicazione prevista dall'art. 23 del d.lgs. n. 33/2013 relativa al periodo luglio 2011 - 2015	Tutti i Servizi	R	Settembre 2015
Bandi di gara	Pubblicazione prevista dall'art. 1, comma 32, della legge n. 190/2012 da parte delle Pubbliche Amministrazioni di informazioni relative a lavori, forniture e servizi appaltati a terzi, relativa ai periodi 2012 - 2013 -2014	EmPULIA / Affari Generali	R	Luglio 2015
Atti di concessione	Pubblicazione prevista dall'art. 26, comma 2, del d.lgs. n. 33/2013, i dati fanno riferimento al 31 dicembre 2013	Tutti i servizi	R	Settembre 2015



Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Tempi previsti
Canoni di locazione e affitto	Pubblicazione prevista dall'art. 30 del d.lgs. n. 33/2013 concernente i beni immobili e la gestione del patrimonio, i dati si riferiscono alle locazioni attive e passive relative al 2013	Servizio Demanio e Patrimonio	R	Settembre 2015
Indicatore di tempestività dei pagamenti	Pubblicazione prevista dall'art. 33 del d.lgs. n. 33/2013 concernente i tempi di pagamento dell'amministrazione	Servizio Bilancio e Ragioneria	R	Settembre 2015
Elenco dei conti di pagamento intestati a Regione Puglia	Pubblicazione prevista dall'art. 36 del d.lgs. n. 33/2013 concernente le informazioni necessarie per l'effettuazione di pagamenti informatici	Servizio Bilancio e Ragioneria		Ottobre 2015
Elenco codici e causali di pagamento sul CC postale n.60225323 intestato a Regione Puglia	Pubblicazione prevista dall'art. 36 del d.lgs. n. 33/2013 concernente le informazioni necessarie per l'effettuazione di pagamenti informatici	Servizio Bilancio e Ragioneria	R	Ottobre 2015
Interventi Straordinari e di emergenza Protezione Civile	Pubblicazione prevista dall'art. 42 del d.lgs. n. 33/2013 concernente interventi straordinari e di emergenza che comportano deroghe alla legislazione vigente, relativi al 2013	Servizio Protezione Civile	R	Ottobre 2015
Elenco Aree, Servizi, Uffici della Regione Puglia con relativi codici IPA	Elenco dei codici univoci e delle strutture regionali destinate alla fatturazione elettronica	www.indicepa.gov.it / Servizio Personale e Organizzazione	R	Novembre 2015
Elenco beneficiari PO FESR 2007-2013	Pubblicazione prevista dall'art. 7 del Regolamento n. 1828/2006 della commissione Europea, concernente le Responsabilità dell'autorità di gestione relative agli interventi informativi e	Strutture regionali agenti su PO / Autorità di gestione FESR	R	Dicembre 2015

Piano Annuale degli Open Data regionali

Dataset	Descrizione	Struttura fonte e responsabile dei dati	Rilascio (R) / Aggiornamento (A)	Tempi previsti
	pubblicitari destinati al pubblico			

Tabella 9 – Dataset Amministrazione Trasparente