



Linked Open Data

Linee Guida per la pubblicazione
degli Open Data della Regione
Puglia

Istruzioni Operative

Versione 1.0



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

Maggio 2014

Codice.: [RP1201_D.7.3.Q1_v0.2]

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.

Sommario

1. Introduzione	4
1. Quadro normativo di riferimento	5
2. Pubblicazione degli Open Data sulla piattaforma della Regione Puglia	6
2.1 Il processo di pubblicazione degli Open Data	6
2.1.1 I ruoli	6
2.1.2 L'iter di pubblicazione	6
2.2 Il processo di produzione dei Linked Open Data	8
2.3 Pubblicazione dei Linked Open Data	11
2.4 Pubblicazione dei dati non strutturati	11
2.4.1 Pubblicazione delle FAQ	12
2.5 Licenze d'uso dei dati della Regione Puglia	13
2.6 Promozione dei dati aperti e monitoraggio della loro diffusione	14
3. Risorse e link utili	15

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].

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 giugnom2009, 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.
- dall’Atto Integrativo alla Convenzione di cui alla deliberazione della Giunta Regionale n.2485 del 27 novembre 2012 (n.rep.014453 del 13/12/2012).
- dalla Deliberazione della Giunta Regionale 7 maggio 2009, n.751 - Convenzione tra la Regione Puglia e la Società in house InnovaPuglia Spa per la disciplina di fornitura dei servizi nell’ambito dei programmi comunitari 2007-2013
- 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.

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;
- 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.

2.2 *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;
- 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. STARCISETTE - EUROCISETTE - 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; a responsabilit&#224; limitata</FORMA_GIURIDICA>

<PROV_SEDE_LEGALE>BA</PROV_SEDE_LEGALE>

<PROV_SEDE_OPERATIVA>BA</PROV_SEDE_OPERATIVA>

<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>
<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>

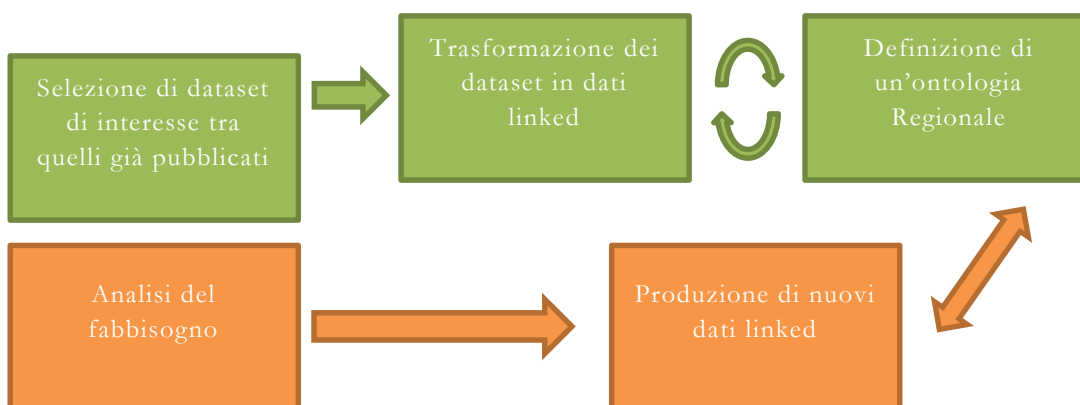
```

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

2.3 *Publicazione 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.

2.4 *Publicazione 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.

2.4.1 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 amministrazione "Linea Amica"¹ pubblichi 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>
```

¹<http://www.lineaamica.gov.it/>

```

<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;ahref="/aree-tematiche/lavoro" typeof="skos:Concept" property="rdfs:labelskos:prefLabel" datatype=""&gt;Lavoro&lt;/a&gt;</Aree-tematiche>

<Tag-liberi>&lt;ahref="/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:

- 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.

2.5 *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/iodl/2.0/>

- 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.

2.6 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/>).
- [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.

===== *PAGINA DI FINE DOCUMENTO* =====