Agenzia per l'Italia Digitale

Presidenza del Consiglio dei Ministri

Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico

Anno 2016

EXECUTIVE SUMMARY

È opinione abbastanza diffusa, soprattutto tra le comunità e gli addetti ai lavori, che, nonostante gli sforzi profusi negli scorsi anni a livello centrale come a livello locale, i risultati della valorizzazione del patrimonio informativo pubblico siano ancora troppo spesso confinati a iniziative virtuose isolate di alcune amministrazioni.

Il principio dell'open data by default, introdotto nel 2012 con la modifica dell'articolo 52 del Codice dell'Amministrazione, per cui "*i dati e i documenti che le amministrazioni titolari pubblicano senza l'espressa adozione di una licenza si intendono rilasciati come dati di tipo aperto*", a oggi non può più considerarsi sufficiente. Tuttavia, da allora altri importanti cambiamenti normativi sono intervenuti soprattutto per quel che riguarda il recepimento della nuova direttiva Europea 2013/37/UE, detta PSI 2.0, che impone alle amministrazioni azioni finalizzate al riutilizzo dei dati pubblici per fini commerciali e non commerciali.

Il patrimonio informativo pubblico è comunque vasto e articolato, con diverse tipologie di dati che devono essere considerate in una strategia complessiva di valorizzazione. Il "Piano Triennale per l'ICT della Pubblica Amministrazione" pone particolare attenzione al tema delle infrastrutture immateriali e delle basi di dati prevedendo, tra le altre, azioni specifiche attinenti a: i) Basi di dati di interesse nazionale, ii) Condivisione dei dati tra pubbliche amministrazioni e iii) Rilascio di dati pubblici secondo il paradigma dell'Open Data e loro riutilizzo. La strategia suggerisce un percorso che passa dall'individuazione di basi di dati di interesse nazionale, ovvero altamente affidabili ed essenziali per un elevato numero di procedimenti amministrativi (altrimenti dette "base register" secondo la terminologia prevista nell'ambito dell'European Interoperability Framework), alla piena condivisione dei dati tra ciascuna amministrazione per finalità istituzionali, inclusi dati personali, fino ad arrivare all'apertura, secondo l'Open Data, della gran parte dei dati prodotti dalle amministrazioni, nel rispetto degli ambiti di applicazione previsti dalle norme di riferimento.

Nel contesto dei dati aperti, la strategia complessiva include la definizione di un "Paniere dinamico di dataset" (aggiornabile di anno in anno) attraverso il quale sono individuate le basi di dati, sia regionali, sia nazionali, che le amministrazioni intendono rendere disponibili dal 2016 in poi secondo i dettami dell'Open Data. Per quanto riguarda i dati delle Regioni, si prevede l'individuazione di un insieme di basi di dati comuni che possano essere rese aperte in maniera collegiale e uniforme da tutte le Regioni. Tale paniere indirizza quindi l'azione delle amministrazioni per la pianificazione e l'effettiva apertura dei dataset, tenendo altresì conto degli obiettivi e dei dataset individuati e/o concordati nel contesto delle azioni previste dal piano nazionale dell'Open Government Partnership. Conseguentemente, il Paniere costituisce anche la base di riferimento per diverse azioni di monitoraggio che devono essere intraprese per dar seguito sia agli impegni assunti nell'ambito dell'accordo di partenariato 2014-2020, per l'impiego dei fondi strutturali e di investimento europei, sia alle disposizioni dell'articolo 52 del Codice dell'Amministrazione Digitale e della suddetta Direttiva PSI 2.0.

In questo scenario, il presente documento aggiorna il precedente, già pubblicato dall'Agenzia nel corso del 2014, con alcuni elementi di novità. Esso è costruito nel tentativo di renderlo di più immediata lettura, guidando l'utente attraverso un percorso che prevede azioni specifiche, la cui mancata attuazione può vanificare il processo di valorizzazione dei dati della pubblica amministrazione. Inoltre, esso pone una rinnovata attenzione al tema dei metadati, introducendo il profilo nazionale DCAT-AP_IT anche necessario per la documentazione di quelli inclusi nel portale dati.gov.it. Quest'ultimo assume un ruolo più rilevante rispetto al passato in attuazione delle norme di recepimento della direttiva PSI 2.0. Infine, l'aggiornamento delle linee guida mira a (i) strutturare una prima versione dell'architettura dell'informazione del settore pubblico grazie alla quale identificare un insieme di dati di riferimento e "core" (o indipendenti da domini applicativi), per i quali applicare modellazioni comuni; (ii) porre attenzione su dimensioni specifiche di qualità dei dati da garantire secondo quanto previsto dagli standard ISO/IEC 25012 e 25024; e (iii) suggerire una licenza di riferimento per tutti i dati della pubblica amministrazione che sia aperta e internazionalmente riconosciuta.

INDICE


Scopo, destinatari e struttura del documento


Normativa di Riferimento


Dati della Pubblica Amministrazione


Modello per i dati aperti e per i metadati


Aspetti organizzativi e di qualità per i dati


Architettura dell'informazione del settore pubblico


Aspetti legali e di costo


Pubblicazione e dati.gov.it

ACRONIMI

ANNCSU – Anagrafe Nazionale dei Numeri Civici e delle Strade Urbane

ANPR – Anagrafe Nazionale della Popolazione Residente

API – Application Programming Interface

CAD – Codice dell'Amministrazione Digitale

CC – Creative Commons

CMS – Content Management System

CPSV – Core Public Service Vocabulary

CSV – Comma Separated Value

DCAT – Data Catalog Vocabulary

DCAT-AP – Data Catalog Vocabulary - Application Profile

DCAT-AP_IT – Data Catalog Vocabulary - Application Profile ITaliano

D.lgs – Decreto legislativo

GPS – Global Position System

HTTP – HyperText Transfer Protocol

INSPIRE – INfrastructure for SPatial InfoRmation in Europe

ICT – Information and Communication Technology

IPA – Indice della Pubblica Amministrazione

ISA – Interoperability Solutions for public Administration

LOD – Linked Open Data

JSON – JavaScript Object Notation

OD – Open Data

ODI – Open Data Institute

OSM – Open Street Map

OWL – Ontology Web Language

OKFN – Open Knowledge Foundation

PA – Pubblica Amministrazione

PSI – Public Sector Information

RDF – Resource Description Framework

RDFS – RDF Schema

RNDT – Repertorio Nazionale Dati Territoriali

SDMX – Statistical Data and Metadata eXchange

SPARQL – Sparql Protocol And Rdf Query Language

URI – Uniform Resource Identifier

XML – eXtensible Markup Language

WGS – World Geodetic System

SCOPO, DESTINATARI E STRUTTURA DEL DOCUMENTO

SCOPO

Il presente elaborato rappresenta un documento di linee guida che ha l'obiettivo di supportare le pubbliche amministrazioni nel processo di valorizzazione del proprio patrimonio informativo pubblico, proponendo una serie di azioni che devono essere necessariamente intraprese per attuare in maniera omogenea su scala nazionale questo processo. Il documento, in linea con gli obiettivi indicati nell'articolo 52 del D.lgs 7 marzo 2005, n. 82 – Codice dell'Amministrazione Digitale (CAD), approfondisce da un lato un modello e un'architettura di riferimento per l'informazione del settore pubblico, individuando standard di base, formati, vocabolari/ontologie per dati di riferimento e "core", ricorrenti e indipendenti da domini applicativi, profili di metadati descrittivi nazionali, e dall'altro gli aspetti organizzativi e di qualità dei dati necessari per individuare i ruoli e le figure professionali delle pubbliche amministrazioni nonché le fasi dei processi per la gestione e pubblicazione di dati di qualità. Inoltre, il documento mira a fornire supporto (i) nella scelta della licenza per i dati di tipo aperto, (ii) nell'analisi di eventuali aspetti di costo dei dati, e (iii) nella loro pubblicazione nei portali per una maggiore standardizzazione di questo processo.
Le presenti linee guida rappresentano un aggiornamento rispetto a quelle pubblicate nel corso del 2014. L'aggiornamento ha riguardato quasi tutte le sezioni della precedente versione tranne quella sulle indicazioni per i capitolati di gara, le cui raccomandazioni sono a oggi ancora valide.

DESTINATARI

Secondo quanto previsto dal CAD (art. 2, commi 2 e 4) per l'applicazione del Capo V, il presente documento è destinato a tutte le pubbliche amministrazioni, alle società interamente partecipate da enti pubblici o con prevalente capitale pubblico inserite nel conto economico consolidato della pubblica amministrazione, come individuate dall'ISTAT ai sensi dell'art. 1, co. 5, della L. 311/2004. Con riferimento alle disposizioni concernenti l'accesso ai documenti informatici e alla fruibilità delle informazioni digitali di cui al capo V del CAD, il documento è destinato anche ai gestori di servizi pubblici e agli organismi di diritto pubblico.
In virtù della duplice valenza tecnico-organizzativa delle linee guida, esse si rivolgono sia a figure professionali delle amministrazioni in possesso di competenze tecnico-informatiche (ad esempio, direttori dei sistemi informativi, responsabili siti Web, funzionari e consulenti tecnici), sia a figure professionali individuabili in quelle aree più amministrative preposte all'organizzazione dei dati (ad esempio, responsabili di basi di dati specifiche, responsabili amministrativi, esperti di dominio).

STRUTTURA DEL DOCUMENTO

Il documento si articola in sezioni che rappresentano passi di un'ipotetica "check list" da seguire per attuare il processo di valorizzazione del patrimonio informativo pubblico. La sezione "Normativa di Riferimento" presenta il quadro normativo e la sezione "Dati della Pubblica Amministrazione" introduce le definizioni sui dati pubblici. La sezione "Modello per i dati aperti e i metadati" descrive i modelli di riferimento per i dati di tipo aperto e per i metadati riportando, in quest'ultimo caso, il profilo di metadatazione nazionale DCAT-AP_IT. La sezione "Aspetti organizzativi e di qualità per i dati" propone un modello operativo per la produzione e gestione dei dati pubblici individuando ruoli, responsabilità e azioni da intraprendere, nonché dimensioni di qualità dei dati e una metodologia per monitorare tali dimensioni. La sezione "Architettura dell'informazione del settore pubblico" individua l'architettura generale fornendo un'indicazione sugli standard di base e formati aperti per dati e documenti; la sezione "Aspetti legali e di costo" raccomanda alcune licenze per i dati di tipo aperto e analizza aspetti legati ai costi e alla tariffazione per i dati del settore pubblico. Infine, la sezione "Pubblicazione e dati.gov.it" descrive i passi per la pubblicazione e discute del rinnovato ruolo del portale nazionale dei dati, dati.gov.it.

NORMATIVA DI RIFERIMENTO

AZIONE 1 : RISPETTA I PRINCIPI DELLE SEGUENTI NORMATIVE E LINEE GUIDA

D.lgs 7 marzo 2005 n. 82 e s.m.i- Il Codice dell'Amministrazione Digitale (CAD) 1 – in particolare articoli 50 "Disponibilità dei dati delle pubbliche amministrazioni", 52 "Accesso telematico e riutilizzo dei dati delle pubbliche amministrazioni" che introduce il principio dell'Open Data by default, e 68 comma 3 per la definizione di dato aperto.

D.lgs 24 gennaio 2006, n.36, come modificato dal D.lgs 18 maggio 2015 n. 102 - Attuazione della direttiva 2013/37/UE (che modifica la direttiva 2003/98/CE relativa al "Riutilizzo dell'informazione del settore pubblico) **2.**

Statuto internazionale degli open data 3.

Linee guida europee su licenze standard e dataset raccomandati e tariffe da applicare nel riutilizzo di dati pubblici 4.

La direttiva 2013/37/UE (direttiva PSI 2.0) ha modificato radicalmente lo scenario in materia di "Riutilizzo della Informazione del Settore pubblico", declinando il principio generale secondo il quale "… *Gli Stati membri provvedono affinché i documenti cui si applica la presente direttiva … siano riutilizzabili a fini commerciali o non commerciali …*", fermo restando l'ambito di applicazione delineato dalla direttiva medesima. Tale principio è stato naturalmente ripreso dalla norma italiana di recepimento della direttiva, diventando quindi un preciso adempimento per le amministrazioni.

Ciò stante, si evidenziano i seguenti aspetti più significativi dell'attuale normativa di riferimento in materia:

- si applica ai dati pubblici, cioè ai dati conoscibili da chiunque;
- estende l'applicabilità ai documenti i cui diritti di proprietà intellettuale sono detenuti dalle biblioteche, comprese le biblioteche universitarie, dai musei e dagli archivi, qualora il riutilizzo di questi ultimi documenti sia autorizzato in conformità alle disposizioni in materia;
- delimita l'ambito di utilizzo, specificando esclusioni e norme di salvaguardia;
- prevede la possibilità di richiedere esplicitamente dati pubblici non ancora disponibili;
- ribadisce il principio generale di disponibilità gratuita dei dati e prevede apposite modalità di tariffazione per l'applicazione dei costi marginali o, nei casi eccezionali, di costi superiori a quelli marginali
- prevede la necessità di agevolare la ricerca dei dati mediante un apposito portale gestito da AgID (individuato in dati.gov.it).

Alla luce delle succitate disposizioni, le amministrazioni terranno conto delle differenze specifiche tra Open Data, Trasparenza e Condivisione dei dati tra pubbliche amministrazioni per finalità istituzionali. Queste tre azioni mirano a soddisfare esigenze diverse e anche se su alcuni aspetti convergono, fanno sempre riferimento a obiettivi specifici senza mai veramente confluire in un "corpus" organico.

1. D.lgs 7 marzo 2005 n. 82 – Codice Amministrazione Digitale www.gazzettaufficiale.it/eli/id/2005/05/16/005G0104/sg
2. D.lgs 18 maggio 2015 n. 102 – www.gazzettaufficiale.it/eli/id/2015/07/10/15G00116/sg
3. International open data charter - http://opendatacharter.net/
4. Commissione europea "Guidelines on recommended standard licences, datasets and charging for the reuse of documents" (2014/C 240/01) - http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=6421

Ad esempio, in termini di trasparenza, alcuni documenti resi pubblici a seguito dell'applicazione del D.lgs 33/2013 e s.m.i 5. nella sezione "Amministrazione Trasparente" del sito web istituzionale di una amministrazione devono essere rimossi dopo aver svolto la loro funzione (di solito dopo 3 anni - cfr. art. 14 e 15). In questo senso, essi non possono essere propriamente considerati Open Data, per i quali tali restrizioni temporali non si applicano. Esistono poi dati delle pubbliche amministrazioni che assumono un ruolo importante nell'ecosistema degli Open Data e nella creazione di nuove forme di partecipazione (e.g. edifici, farmacie, musei, turismo, etc.) ma che non risultano nell'elenco dei dati obbligatori da pubblicare ai sensi del D.lgs n. 33/2013 e s.m.i.

In sostanza i concetti "Condivisione", "Trasparenza" e "Open Data" **svolgono ruoli informativi e funzionali diversi**; ove possibile, **si raccomanda pertanto di coordinare le attività a essi connesse**, così come anche indicato nelle presenti linee guida in "Aspetti organizzativi e di qualità per i dati".

LINEE GUIDA OPEN DATA LOCALI

Molte amministrazioni (regioni e comuni in particolare) hanno affrontato internamente il tema dei dati di tipo aperto definendo delle linee guida per l'individuazione delle basi di dati pubbliche in loro possesso e per le relative modalità di apertura. Le linee guida sono di solito approvate con atti amministrativi quali Deliberazioni di Giunta (come nel caso dei comuni per esempio). Tali deliberazioni hanno valore di indicazione operativa e di processo per l'ente pubblico che se ne dota, ma se un contenuto/obiettivo delle linee guida non viene rispettato/raggiunto, di solito nella pratica non vengono attivate penalità o sanzioni interne.
Nella difficoltà di tener conto delle diverse iniziative (e dei relativi aggiornamenti) si è quindi ritenuto di non riportare più, in questa sede, l'elenco analitico-descrittivo delle stesse.

Tuttavia, in generale, al fine di rendere sistemico e omogeneo su scala nazionale il processo di valorizzazione dei dati pubblici, **regolamenti locali o interni, inclusi quelli futuri di cui le pubbliche amministrazioni vorranno dotarsi, devono uniformarsi ai principi e alle linee d'azione delle presenti linee guida, nonché alla strategia in materia di dati aperti definita nel piano triennale per l'ICT nella pubblica amministrazione**, previsto dalle disposizioni di cui all'art.1, comma 513 e seguenti della legge 28 dicembre 2015, n.208 (Legge di stabilità 2016).
Si noti che ai sensi dell'art. 1 comma 517 della legge di stabilità 2016, la **mancata osservanza delle disposizioni dei commi 512-516 (e quindi dell'adeguamento al piano triennale), rileva ai fini della responsabilità disciplinare e per danno erariale**. Infine, l'articolo 52 comma 4 prevede che le attività volte a garantire l'accesso telematico e il riutilizzo dei dati delle pubbliche amministrazioni rientrano tra i parametri di valutazione delle performance dirigenziale ai sensi dell'articolo 11, comma 9, del d.lgs 27 ottobre 2009, n.150.

5. D.lgs 14 Marzo 2013 n. 33 – Decreto Trasparenza - www.gazzettaufficiale.it/eli/id/2013/04/05/13G00076/sg

DATI DELLA PUBBLICA AMMINISTRAZIONE

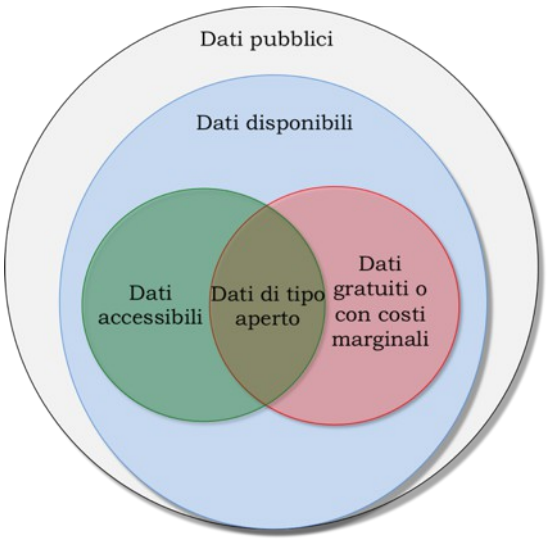


Figura 1: Tipi di dato della pubblica amministrazione

Le presenti linee guida si applicano al dato pubblico**,** ovvero al dato della pubblica amministrazione conoscibile da chiunque e non soggetto a restrizioni temporali (e.g., diritto all'oblio).

Si escludono pertanto: (i) dati a conoscibilità limitata come i dati coperti da segreto di stato o le opere d'ingegno coperte dal diritto d'autore; (ii) i dati personali, per i quali trovano applicazione le norme del "Codice in materia di protezione dei dati personali" (i.e., D.lgs n. 196/2003 e Linee guida in materia di trattamento di dati personali 6). In questo caso, si ponga anche attenzione a non esporre quasi-identificatori (e.g., data di nascita, domicilio, residenza, sesso, razza, etnia, composizione nucleo famigliare, status giuridico, ecc.) che possono facilmente re-identificare i soggetti che si intende invece tutelare o che hanno una tutela speciale perché appartenenti a fasce protette (e.g., testimoni giudiziari, profughi, rifugiati, pentiti, ecc.).

In ogni caso, si raccomanda di verificare gli artt. 3 e 4 del D. Lgs. 36/2006 per una visione approfondita circa le esclusioni e le norme di salvaguardia.

AZIONE 2: RICORDA E VERIFICA ART. 68 COMMA 3 DEL CAD…

- **Dato pubblico –** dato conoscibile da chiunque. A seguito delle recenti modifiche apportate con il D. Lgs. n.179/2016, il CAD non contempla più, tra le altre, la definizione di dato pubblico. Tuttavia, nel contesto delle presenti linee guida, si ritiene opportuno continuare a fare riferimento al concetto di dato pubblico come precedentemente definito
- **Formato dei dati di tipo aperto -** un formato reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi

- **Dato aperto (risponde a tre requisiti):**
  - *Disponibile (requisito giuridico)* secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato
  - *Accessibile (requisito tecnologico)* attraverso le tecnologie dell'informazione e della comunicazione, in formato aperto e con i relativi metadati
  - *Gratuito (requisito economico):*
    - disponibili gratuitamente oppure
    - disponibili ai costi marginali sostenuti per la loro riproduzione, messa a disposizione e divulgazione. AgID, su proposta dell'amministrazione titolare, determina le tariffe standard e le pubblica sul proprio sito istituzionale.
    - **Eccezione:** dati per i quali le pubbliche amministrazioni e gli organismi di diritto pubblico generano utili sufficienti per coprire una parte sostanziale dei costi di raccolta, produzione, riproduzione e diffusione. Con decreti dei Ministeri competenti, di concerto con il Ministero dell'economia e delle finanze, sentita AgID, si determinano le tariffe e le modalità di versamento a fronte delle suddette attività.

6. Garante per la Protezione dei Dati Personali, "Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati", http://194.242.234.211/documents/10160/0/Linee+guida+trasparenza+2014.pdf, 2014.

MODELLO PER I DATI APERTI E PER I METADATI

MODELLO PER I DATI APERTI

AZIONE 3: VERIFICA LA CONFORMITÀ AL MODELLO PER I DATI APERTI…

Si adotta il modello qualitativo per i dati aperti sul Web, noto come modello a cinque stelle, così come rappresentato in Figura 2.
In particolare, si raccomanda un percorso graduale verso la produzione nativa di Linked Open Data – LOD (livello cinque stelle), iniziando dal livello 3 di Figura 2. Produzione e pubblicazione di dati aperti *solo* di livello 1 e 2 non sono più ammessi: quest'ultimi devono essere accompagnati da quelli che rispecchiano le caratteristiche dei livelli 3 e/o superiori (per esempio, rilasciare dati strutturati solo in excel con licenza aperta non è ammesso; questi devono essere sempre affiancati da dati strutturati in formato non proprietario).

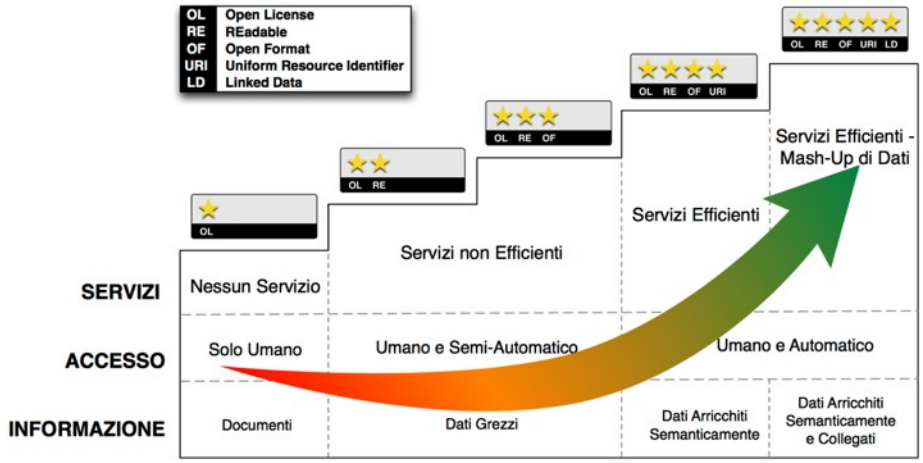


Figura 2: Modello a cinque stelle per i dati aperti sul Web (rivisitazione della figura nota e disponibile sul Web[1])

I livelli del modello per i dati aperti

Informazione:	Dati disponibili tramite una licenza aperta e inclusi in documenti leggibili e interpretabili solo grazie a un significativo intervento umano (e.g., PDF).
Accesso:	Prevalentemente umano, necessario anche per dare un senso ai dati inclusi nei documenti.
Servizi:	Solo rilevanti interventi umani di estrazione ed elaborazione dei possibili dati consentono di sviluppare servizi con l'informazione disponibile in questo livello.

Informazione:	Dati disponibili in forma strutturata e con licenza aperta. Tuttavia, i formati sono proprietari (e.g., Excel) e un intervento umano è fortemente necessario per un'elaborazione dei dati.
Accesso:	I programmi possono elaborare i dati ma non sono in grado di interpretarli; pertanto è necessario un intervento umano al fine di scrivere programmi ad-hoc per il loro utilizzo.
Servizi:	Servizi ad-hoc che devono incorporare i dati per consentire un accesso diretto via Web agli stessi.

[1] http://5stardata.info/en/

Informazione:	Dati con caratteristiche del livello precedente ma in un formato non proprietario (e.g., CSV, JSON, geoJSON). I dati sono leggibili da un programma ma l'intervento umano è necessario per una qualche elaborazione degli stessi.
Accesso:	I programmi possono elaborare i dati ma non sono in grado di interpretarli; pertanto è necessario un intervento umano al fine di scrivere programmi ad-hoc per il loro utilizzo.
Servizi:	Servizi ad-hoc che devono incorporare i dati per consentire un accesso diretto via Web agli stessi.

Informazione:	Dati con caratteristiche del livello precedente ma esposti usando standard W3C quali RDF e SPARQL I dati sono descritti semanticamente tramite metadati e ontologie.
Accesso:	I programmi sono in grado di conoscere l'ontologia di riferimento e pertanto di elaborare i dati quasi senza ulteriori interventi umani.
Servizi:	Servizi, anche per dispositivi mobili, che sfruttano accessi diretti a Web per reperire i dati di interesse.

Informazione:	Dati con caratteristiche del livello precedente ma collegati a quelli esposti da altre persone e organizzazioni (i.e., Linked Open Data[2]). I dati sono descritti semanticamente tramite metadati e ontologie. Essi seguono il paradigma RDF (si veda "Architettura dell'informazione del settore pubblico"), in cui alle "cose" (o entità) è assegnata un URI univoca sul Web. Conseguentemente tale URI può essere utilizzata per effettuare accessi diretti alle informazioni relative a quella entità. I dati sono detti "linked" per la possibilità di referenziarsi (i.e., "collegarsi") tra loro. Nel referenziarsi, si usano relazioni ("link") che hanno un preciso significato e spiegano il tipo di legame che intercorre tra le due entità coinvolte nel collegamento. I Linked (Open) Data sono quindi un metodo elegante ed efficace per risolvere problemi di identità e provenienza, semantica, integrazione e interoperabilità. *Triple RDF i cui URI non siano utilizzabili da un agente Web per recuperare le informazioni a essi associati, non possono essere considerati pienamente conformi al paradigma Linked Data.* Nei caso dei Linked Open Data l'intervento umano si può ridurre al minimo e talvolta addirittura eliminare.
Accesso:	I programmi sono in grado di conoscere l'ontologia di riferimento e pertanto di elaborare i dati quasi senza ulteriori interventi umani.
Servizi:	Servizi, anche per dispositivi mobili, che sfruttano sia accessi diretti a Web sia l'informazione ulteriore catturata attraverso i "link" dei dati di interesse, facilitando il mashup di dati.

[2] https://www.ted.com/talks/tim_berners_lee_on_the_next_web?nolanguage=en%2C, https://www.w3.org/DesignIssues/LinkedData.html, http://linkeddatabook.com/editions/1.0/ http://linkeddata.org/home

Modello per i Metadati

Azione 4: Correda i dati con i relativi metadati…

La metadatazione ricopre un ruolo essenziale laddove i dati sono esposti a utenti terzi e a software. I metadati, infatti, consentono una maggiore comprensione e rappresentano la chiave attraverso cui abilitare più agevolmente la ricerca, la scoperta, l'accesso e quindi il riuso dei dati stessi. A tale scopo, si adotta il modello per i metadati rappresentato in Figura 3. Il modello si focalizza sugli aspetti qualitativi dei metadati, è indipendente dal particolare schema proposto e, in parte, anche dal formato fisico di rappresentazione. La classificazione qualitativa dei metadati si fonda su due fattori principali: legame tra dato-metadato e livello di dettaglio.

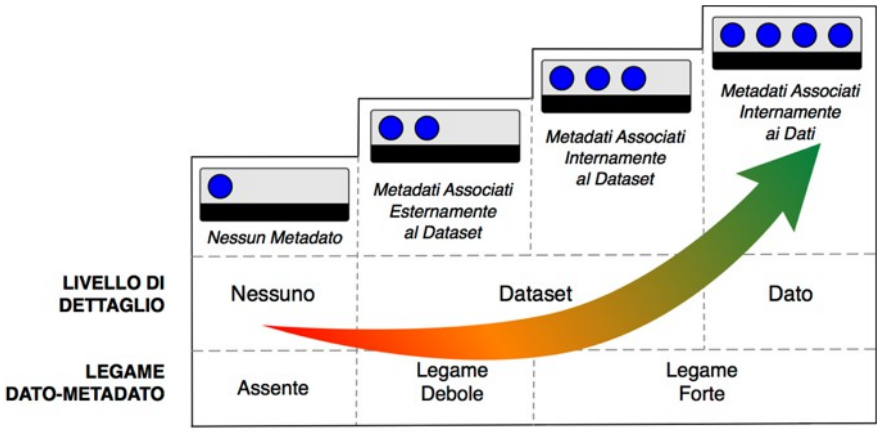


Figura 3: Modello a quattro livelli per i metadati

I livelli del modello per i metadati

Legame dato metadato:	Nessun legame in quanto i dati non sono accompagnati da un'opportuna metadatazione.
Livello di dettaglio:	Nessuno in quanto i metadati non sono presenti.

Legame dato metadato:	Il legame è debole perché i dati sono accompagnati da metadati esterni, (e.g., inclusi nella pagina di download del dataset o in file separati).
Livello di dettaglio:	I metadati forniscono informazioni relativamente a un dataset, quindi sono informazioni condivise dall'insieme di dati interni a quel dataset.

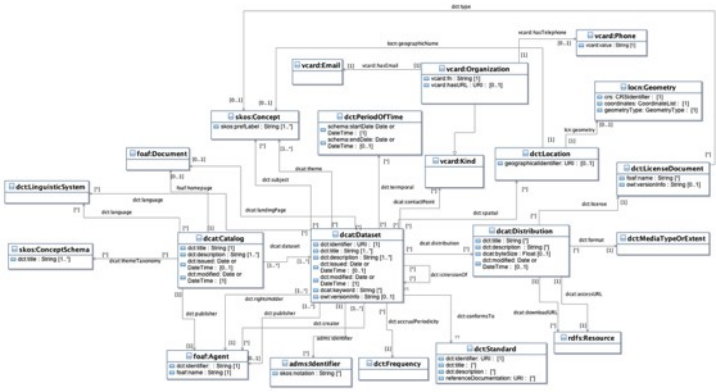
Legame dato metadato:	Il legame è forte perché i dati incorporano i metadati che li descrivono.
Livello di dettaglio:	I metadati forniscono informazioni relative a un dataset, quindi sono informazioni condivise dall'insieme di dati interni a quel dataset.

Legame dato metadato:	Il legame è forte perché i dati incorporano i metadati che li descrivono.
Livello di dettaglio:	I metadati forniscono informazioni relative al singolo dato, quindi col massimo grado di dettaglio possibile

PROFILO NAZIONALE PER I METADATI: DCAT-AP_IT

AZIONE 5: RISPETTA IL PROFILO DI METADATAZIONE DCAT-AP_IT…

Per i metadati descrittivi generali, ovvero non dipendenti da tipologie di dati, si adotta il profilo nazionale DCAT-AP_IT, rispettando le obbligatorietà, le raccomandazioni e seguendo gli esempi così come definiti nella relativa specifica 7, 8. Il profilo, disponibile secondo gli standard del Web Semantico (si veda la sezione "Architettura dell'informazione del settore pubblico"), si basa sullo standard DCAT e su vocabolari ampiamente utilizzati nel Web quali per esempio Dublin Core e schema.org. Il profilo si applica a tutti i tipi di dati pubblici, è pienamente conforme a quello europeo DCAT-AP 9, quest'ultimo nato al fine di uniformare la specifica dei metadati descrittivi per tutti gli stati membri europei, facilitando lo scambio di informazioni e l'interoperabilità anche transfrontaliera e favorendo il riutilizzo e la valorizzazione dell'informazione.



Si noti che, nel caso di dati geografici, siano essi aperti o non aperti (secondo le definizioni riportate in sezione "Dati delle pubbliche amministrazioni"), il profilo di metadatazione da adottare è quello del Repertorio Nazionale dei Dati Territoriali (RNDT), conforme alla direttiva INSPIRE, i.e., profilo RNDT/INSPIRE 10.

L'RNDT, in quanto banca dati di interesse nazionale ai sensi dell'articolo 60 del CAD e banca dati critica, è soggetta a regole di interoperabilità e gestione che prevedono, tra le altre, anche l'applicazione del principio "once only". Secondo questo principio, i dati geografici sono documentati *solo una volta* e inclusi all'interno del catalogo RNDT, secondo le regole del profilo RNDT/INSPIRE (Figura 4). Sarà lo stesso catalogo, in maniera automatizzata, a fornire l'adeguata integrazione con i metadati descrittivi definiti mediante DCAT-AP_IT, grazie a una specifica estensione per il trattamento dei dati geografici detta GeoDCAT-AP 11 che il Repertorio implementerà a tale scopo.

Lo stesso principio può trovare applicazione anche per altre tipologie di dati, come nel caso dei dati statistici per cui si raccomanda di considerare la relativa estensione StatDCAT-AP 12, sviluppata in ambito Europeo.

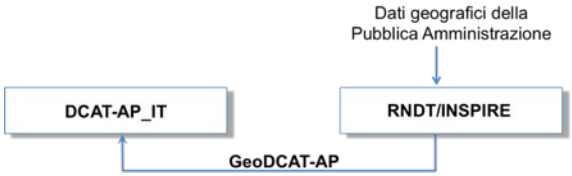


Figura 4: DCAT-AP_IT e RNDT/INSPIRE

7. Agenzia per l'Italia Digitale, "DCAT-AP_IT – profilo italiano di DCAT-AP", http://www.dati.gov.it/sites/default/files/DCAT-AP_IT_v10.pdf,
8. Agenzia per l'Italia Digitale, "Ontologia del profilo DCAT-AP_IT", http://www.dati.gov.it/onto/dcatapit
9. ISA programme, "DCAT-AP v. 1.1.", https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-ap-v11
10. Agenzia per l'Italia Digitale, Profilo RNDT/INSPIRE, http://www.rndt.gov.it/RNDT/home/index.php?option=com_content&view=article&id=53&Itemid=221
11. ISA programme, "GeoDCAT-AP v 1.0", https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/geodcat-ap-v10
12. ISA programme, "StatDCAT-AP - Draft 4", https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/asset_release/statdcat-ap-draft-4

Ulteriori metadati di provenienza (provenance)

Le pubbliche amministrazioni possono integrare i metadati previsti dal modello DCAT-AP_IT con metadati aggiuntivi, secondo le proprie necessità seppur nel pieno rispetto delle regole di conformità come definite nella specifica DCAT-AP_IT.

In particolare, come già riportato in ambito Europeo in DCAT-AP, **si raccomanda di inserire metadati sulle entità e sulla filiera di attività, che va dalla generazione alla pubblicazione del dato.** Questo consente di certificare in maniera più accurata la reale provenienza del dato e la titolarità dello stesso, fornendo garanzie di qualità per eventuali riutilizzatori.

Per documentare entità e relative attività, lo standard W3C di riferimento da utilizzare è PROV Framework 13. Attraverso PROV è possibile descrivere in maniera strutturata la provenienza di artefatti e quindi anche di dati che si intende pubblicare, nonché modellare il processo di generazione di un artefatto in maniera quasi analoga ai sistemi di controllo versione.

Il framework PROV è costituito da una famiglia di specifiche articolate in diverse componenti. Per gli scopi delle presenti linee guida, si evidenziano:

- **PROV-DM**: descrive il modello concettuale dei dati; costituisce quindi il nucleo centrale della famiglia di specifiche. Esso non fa riferimento a uno specifico dominio ma è corredato di estensioni per domini più specifici.
- **PROV-O**: anche detto PROV Ontology 14, definisce l'ontologia OWL2 del PROV-DM in modo da poter essere utilizzata direttamente nell'ambito del Web Semantico e dei Linked Data. Alla luce di queste caratteristiche, PROV-O si integra perfettamente con il modello di metadatazione nazionale di riferimento DCAT-AP_IT.
- **PROV-N**: definisce una notazione fruibile da un utente umano per i dati di provenienza creati attraverso il framework.

Metadati di qualità e di struttura del dato

Per facilitare ulteriormente i possibili fruitori del dato, e quindi favorire il più ampio riutilizzo dei dati, si raccomanda di considerare anche l'aggiunta di:

- **metadati che forniscono una descrizione dello schema del dataset da pubblicare**. Nel caso di dati espressi secondo il livello 3 del modello per i dati, lo schema rappresenta l'insieme degli attributi elencati; nel caso dei livelli 4 e 5 esso può essere rappresentato dalle ontologie che accompagnano i dati;
- **metadati che forniscono un riscontro della qualità dei dati esposti e di come tale qualità è misurata e certificata**. In quest'ultimo caso, si raccomanda di utilizzare le linee guida del W3C pubblicate dal gruppo di lavoro su "Data on the Web Best Practices: Data Quality Vocabulary" 15.

13. W3C Working Group Note, PROV-Overview, https://www.w3.org/TR/prov-overview/, 30 Aprile 2013
14. W3C Recommendation, PROV-O: The PROV Ontology, https://www.w3.org/TR/prov-o/, 30 Aprile 2013
15. W3C Working Draft, Data on the Web Best Practices: Data Quality Vocabulary, https://www.w3.org/TR/vocab-dqv/, 19 Maggio 2016

ASPETTI ORGANIZZATIVI E DI QUALITÀ PER I DATI

ASPETTI ORGANIZZATIVI

Figura 5 : Modello operativo: produzione e pubblicazione di dati aperti

L'attuale contesto, sempre più incentrato sull'uso dei dati, pone il problema di intervenire su alcune fasi della catena del valore del dato: la scelta della migliore fonte informativa, il controllo della qualità del dato, l'integrazione di fonti diverse, la tempestività nell'aggiornamento, ecc. Al riguardo, oggi si rende sempre più necessaria la revisione dei processi e dei modelli dei sistemi informativi delle pubbliche amministrazioni, organizzandoli in maniera organica, facendo in modo che il processo di apertura dei dati non sia sempre e solo parallelo a quello di gestione dei dati ma pienamente integrato.

Un dato della PA destinato alla pubblicazione è frutto di una catena di processi nel corso della quale si generano ulteriori prodotti intermedi. Comprendere e governare la struttura di questa catena diventa elemento cruciale. Affinché tale attività non sia assunta come un mero adempimento tecnologico, a essa deve corrispondere:

1) l'ottimizzazione dei processi esistenti all'interno dei quali l'Open Data deve far parte integrante;
2) la dislocazione di soluzioni interoperabili che possano contribuire all'ottimizzazione dei processi;
3) una riduzione nei costi e nei tempi di accesso al capitale informativo;
4) una riduzione della complessità dei processi interni attraverso il consolidamento delle attività derivate da 1) e 2);
5) l'ottimizzazione dei tempi e dei canali di comunicazione i verso risorse esterne all'amministrazione.

I primo passo da compiere è quello di individuare una chiara *data governance* interna con professionalità strategiche e specifiche.

AZIONE 6: INDIVIDUA UNA DATA GOVERNANCE E ASSICURATI CHE I PROCESSI INTEGRINO IL RILASCIO DI DATI APERTI E IL COINVOLGIMENTO DEGLI UTENTI…

Si adotta il modello operativo mostrato in Figura 5 . Il modello ha l'obiettivo di garantire la produzione e la pubblicazione di dati (aperti) di qualità attraverso un processo omogeneo, auto-sostenibile, coordinato tra gli organi interni dell'amministrazione, con la definizione di procedimenti condivisi che possano creare un tessuto sufficientemente robusto e stabile nei suoi punti fondamentali, e necessariamente elastico per l'applicazione alle diverse realtà amministrative.
Per attuare il modello è necessario (i) definire una chiara data governance interna con l'individuazione di ruoli e relative responsabilità; (ii) integrare le sue fasi sia verticalmente, rispetto ai processi interni già consolidati, che orizzontalmente rispetto alle necessità delle diverse amministrazioni. L'applicazione del modello deve avvenire in maniera costante: le attività non si esauriscono con la mera pubblicazione dei dati, ove questo sia possibile, ma devono prevedere un costante aggiornamento, monitoraggio e coinvolgimento con gli utenti finali.

Ruoli e responsabilità

Di seguito si elencano i componenti di un possibile gruppo di lavoro orizzontale e inter-settoriale che un'amministrazione può costituire per avviare e gestire a regime il processo di gestione dei dati in generale e, nello specifico, di apertura dei dati. Dipendentemente dalle dimensioni delle amministrazioni, alcune figure professionali possono coincidere o possono essere ulteriormente distinte.

GRUPPO DI LAVORO OPEN DATA[3]. Il gruppo che promuove l'uso e la diffusione degli Open Data. Esso riporta all'interno dell'amministrazione le novità inerenti il mondo dell'Open Government, media e valuta le esigenze di pubblicazione dati in base alle normative di riferimento, e ne cura la razionalizzazione rispetto agli altri processi di apertura del dato. Ha la responsabilità di pianificare e coordinare l'evoluzione continua dell'apertura dei dati nell'amministrazione, nonché dell'infrastruttura IT a supporto. All'interno del gruppo di lavoro è bene prevedere figure che possano fornire il necessario supporto per l'analisi della qualità dei dati, per la definizione delle interfacce d'accesso ai dati, per la promozione di applicazioni sviluppate a partire dai dati pubblicati, fornendo anche nel caso esempi di servizi dimostrativi attraverso cui incentivare il riutilizzo.
Inoltre, il gruppo di lavoro si può occupare della formazione tecnica e concettuale all'interno dell'amministrazione sui temi legati al paradigma Open Data, anche sulla base delle linee guida pubblicate dall'Agenzia per l'Italia Digitale e sullo stato dell'arte degli Open Data dell'amministrazione. Alcuni membri del team (e.g., esperti di tecnologie Web, esperti GIS, esperti di tecnologie e strumenti per i Linked Data) possono occuparsi della gestione del processo di apertura del dato dal punto di vista IT. Affinché il lavoro del Team Open Data possa essere incisivo all'interno dell'amministrazione, è importante che tale team si confronti con il livello più politico, sia per ottenere da questo le necessarie 'spinte', sia per offrire al decisore politico proposte e stimoli.

RESPONSABILE OPEN DATA (O DATA MANAGER). All'interno del team Open Data è nominato un responsabile. Tale figura permette da un lato di localizzare le competenze necessarie alla gestione delle attività Open Data entro un sistema autonomo di comunicazione e funzionamento, e dall'altro di integrare i processi relativi alle attività di trasparenza in modo parallelo e non seriale. Il responsabile Open Data deve quindi possedere sia le capacità operative di controllo di tale sistema, sia quelle amministrative di coordinamento con i processi già esistenti. Insieme al team suddetto, conosce i dati dell'amministrazione nel loro insieme, redige linee guida operative per lo scambio dati tra le diverse figure coinvolte (si veda sotto), e pianifica la strategia di apertura dei dati raccolti e analizzati e le attività di diffusione dei dati. Infine, collabora e si coordina con il Responsabile della Trasparenza (quest'ultimo istituito ai sensi del D.lgs. n. 33/2013 e s.m.i) al fine di rafforzare vicendevolmente gli obiettivi da un lato di massimo riutilizzo dei dati pubblici di tipo aperto e dall'altro di trasparenza.

RESPONSABILE DELLA BANCA DATI. All'interno dell'amministrazione è responsabile del procedimento amministrativo che popola la specifica fonte del dato, che ne cura la qualità e il relativo aggiornamento. Tipicamente un Dirigente o un Quadro, coordina un gruppo di persone che svolgono il loro lavoro quotidiano attorno alla fonte del dato. Ha anche il potere di decidere se modificare un certo dato sulla base di indicazioni pervenute ad esempio da cittadini che, vedendo il dataset, ne richiedono una versione evoluta.
REFERENTE TECNICO DELLA BANCA DATI. Si tratta tipicamente di un componente del gruppo coordinato dal responsabile della banca dati; esso deve avere conoscenze informatiche e svolge un ruolo operativo sul sistema gestionale afferente al dato. Inoltre, fornisce indicazioni circa il reperimento concreto dei dati dalla base dati, e cura il monitoraggio dei vari "connettori" che a partire dalla base dati espongono il dato come Open Data. Tipicamente riceve materialmente le segnalazioni

[3] L'art. 17 del nuovo Codice dell'Amministrazione Digitale individua un ufficio dirigenziale generale responsabile per la transizione alla modalità operativa digitale e un difensore civico per il digitale che ha il compito di ricevere segnalazioni di violazione del CAD invitando l'ufficio a porvi rimedio. Si ritiene importante che il responsabile dell'ufficio suddetto (articolo 17 comma 1-ter) faccia parte del gruppo di lavoro open data, anche come figura di raccordo con il livello più politico e che il difensore civico operi in stretta collaborazione con il gruppo open data.

dei cittadini sul dataset di propria competenza, e le smista eventualmente al Referente tematico per valutarne il contenuto, prima di chiedere al Responsabile della Banca Dati l'approvazione per eventuali azioni correttive strutturali sul dataset.
REFERENTE TEMATICO DELLA BANCA DATI. Si tratta di un esperto di dominio che conosce in modo approfondito l'ufficio e la storia dei dati su cui l'ufficio opera. Spesso propone nuovi dataset da esporre a partire dal sistema gestionale corrispondente e cura eventuali valutazioni di dominio o relative al significato dei dati. Ha anche la possibilità di compiere bonifiche e semplici adeguamenti sulla banca dati, su segnalazione di cittadini o su valutazioni proprie. Riferisce invece al Responsabile della Banca dati la necessità di eventuali variazioni strutturali al sistema gestionale che insiste sui dati.

UFFICIO STATISTICA. E' un anello importante dell'intera catena, sia nel promuovere nuove tipologie di dataset da esporre, sia nel validare dal punto di vista metodologico e statistico i dati pubblicati e le loro visualizzazioni.

UFFICIO GIURIDICO-AMMINISTRATIVO. Può assumere le più svariate forme in base all'organizzazione interna dell'amministrazione. In generale esso rappresenta una singola figura che fornisce consulenza sia su aspetti non tecnici legati agli Open Data, come la definizione delle licenze e delle note legali associate ai dati, la loro rimodulazione sulla base di esigenze specifiche (si pensi per esempio alla necessità di aprire dati prodotti da terze parti o addirittura da cittadini), sia su tutte quelle problematiche di tipo giuridico o amministrativo, comprese quelle di privacy, di finalità del dataset e di trattamento del dato personale ove presente.

GRUPPO COMUNICAZIONE. Può assumere varie forme in base all'organizzazione interna dell'amministrazione, ma in ogni caso si indicano quelle figure con competenze di comunicazione istituzionale e non solo, in grado di curare la comunicazione e il dialogo con i cittadini.

Rispetto alle linee di azione del modello operativo mostrato in Figura 5, e descritte di seguito, si individuano i Ruoli e le Responsabilità (RACI)[4] tra le diverse figure identificate.

Processo	**Responsabile Open Data**	**Responsabile banca dati**	**Referente tecnico banca dati**	**Referente tematico banca dati**	**Ufficio statistica**	**Ufficio giuridico-amministrativo**	**Team comunicazione**
Dati nativi	A/R	R	R	R	C	C	I
Dati mashup	A/R	C	R	C	C	C	I
Linked Open Data	R	A/R	R	R	C	C	I
Coinvolgimento	A	C	I	I	C	C	R

Responsible (R): Coloro che lavorano per eseguire un determinato compito. Esiste almeno un ruolo di responsabile.
Accountable (A): Il solo che può approvare il corretto completamento di un compito e che delega il lavoro ai responsabili. Può esistere un solo ruolo accountable per uno specifico compito.
Consulted (C): Coloro che possono essere consultati in quanto esperti di dominio e con i quali instaurare una comunicazione bidirezionale.
Informed (I): Coloro che devono essere tenuti aggiornati sui progressi del processo, spesso al termine dello stesso.

[4] A Guide to the Project Management Body of Knowledge (PMBOK Guide). PMI Standards Committee, Project Management Institute. 2010. ISBN 1-933890-66-5.

I processi del modello operativo

Linea 1: Dati Nativi.

La linea 1 dei *dati nativi* tratta tutta la filiera di gestione ed esposizione dei dati esistenti generati dalle amministrazioni. Questi dati sono principalmente prodotti dai vari uffici durante l'adempimento delle proprie funzioni istituzionali. La maggior parte di questi dati possono essere pubblicati come dati aperti, portando un'ineludibile fonte potenziale di sviluppo per il territorio e per l'intero sistema Paese.

CENSIMENTO. All'interno dei singoli uffici o dei vari settori dell'amministrazione vanno quindi ricercate quelle che si possono chiamare "*basi di dati primarie*" oggetto del censimento. Si tratta di individuare quegli uffici che generano, mantengono e sono responsabili delle specifiche tipologie di dati che si vogliono rendere aperti (e.g., lo sportello unico per le attività produttive (SUAP) del comune è lo strumento che va a semplificare gli adempimenti connessi alla creazione, l'avvio, la modifica e la cessazione delle imprese per la produzione di beni e servizi. L'ufficio SUAP gestisce e mantiene quindi l'archivio con i dati di tutte le imprese del territorio).

Si raccomanda al responsabile Open Data di effettuare una ricognizione interna, alla luce della normativa vigente, in collaborazione con i responsabili delle basi di dati, **al fine di individuare l'insieme di dati esistenti pubblicabili in formato aperto**. Ciascun soggetto preposto alla gestione di una particolare base di dati **indica al responsabile Open Data, tra le altre cose, le caratteristiche descrittive del dato, i tracciati record, il tasso temporale di aggiornamento, e ogni altra informazione utile** a far comprendere le caratteristiche peculiari dei dati.

In quelle realtà in cui il processo di apertura dei dati ha raggiunto una fase matura, il concetto di dato nativo può evolvere, includendo non solo i dati raccolti perché legati all'attività amministrativa, ma anche tutte quelle informazioni che, una volta aperte, possano abilitare nuove forme di riutilizzo dell'informazione. Per esempio, se finora per un ufficio non era prioritario raccogliere in maniera strutturata un certo tipo di dato (e.g., gli esercizi che vendono prodotti a km zero o i locali che hanno prodotti per celiaci), perché non strettamente correlato a qualche norma o regolamento amministrativo, il solo fatto che un dato "nativo" poi viene aperto e reso fruibile in forme strutturate al cittadino, lo rende un dato utile all'attività istituzionale nel concetto "esteso" della pubblica amministrazione, inteso non solo come soggetto erogatore di servizi pubblici, ma anche come espositore di patrimonio informativo che abilita nuove forme di business sul mercato. I dati aperti, quindi, modificano il concetto stesso di utilità del dato inserendo nella categoria dei dati "nativi" della PA informazioni che prima non erano ritenute tali dalla PA stessa, ma che risultano invece utili all'esterno.

Si raccomanda quindi **l'adozione di un approccio di tipo "demand- driven" per individuare i dati nativi che tenga conto dell'impatto economico e sociale nonché del livello di interesse degli utilizzatori** suddivisi opportunamente per categorie (e.g., cittadini, imprese, altre pubbliche amministrazioni), dei loro requisiti e delle loro necessità.

A tal riguardo si evidenzia che il titolare del dato, ai sensi dell'articolo 5 comma 2 del D.lgs 36/2006 come modificato dal D.lgs 18 maggio 2015, n. 102 e s.m.i., stabilisce le modalità di acquisizione delle richieste con proprio provvedimento, instaurando così una collaborazione con le suddette categorie che possono sfruttare tali modalità per avanzare le proprie proposte.

ANALISI GIURIDICA DELLE FONTI. Alla fase di censimento fa seguito l'analisi giuridica delle fonti del dato. Essa è fondamentale per garantire sostenibilità nel tempo del processo di produzione e pubblicazione dei dati e creare un servizio equilibrato nel rispetto della funzione pubblica e dei diritti dei singoli individui. L'analisi giuridica delle fonti mira quindi a valutare questi delicati equilibri, evidenziando limitazioni d'uso, finalità di competenza, determinazione dei diritti e dei termini di licenza.

Si riporta di seguito una breve "**check list", utile per verificare se tutti gli aspetti giuridici** sono stati valutati dal responsabile della banca dati in collaborazione con il responsabile Open Data.

Aspetto	Domanda	Sì/No
Privacy	i dati sono liberi da ogni informazione personale che possa identificare in modo diretto l'individuo (nome, cognome, indirizzo, codice fiscale, patente, telefono, email, foto, descrizione fisica, ecc.)? In caso negativo queste informazioni sono autorizzate per legge?	
	i dati sono liberi da ogni informazione indiretta che possa identificare l'individuo (caratteristiche personali che possono identificare facilmente il soggetto)? In caso negativo queste informazioni sono autorizzate per legge?	
	i dati sono liberi da ogni informazione sensibile riconducibile all'individuo? In caso negativo queste informazioni sono autorizzate per legge?	
	i dati sono liberi da ogni informazione relativa al soggetto che incrociata con dati comunemente reperibili nel web (e.g. google maps, linked data, ecc.) possa identificare l'individuo? In caso negativo queste informazioni sono autorizzate per legge?	
	i dati sono liberi da ogni riferimento a profughi, protetti di giustizia, vittime di violenze o in ogni caso categorie protette?	
	hai considerato il rischio di de-anonimizzazione del tuo dataset prima di pubblicarlo?	
	esponi dei servizi di ricerca tali da poter filtrare i dati in modo da ottenere un solo record geolocalizzato**,** che sia facilmente riconducibile ad una persona fisica?	
Proprietà intellettuale della sorgente	il dataset è stato creato da uno o più dipendenti della tua pubblica amministrazione nell'ambito della loro attività lavorativa? I singoli elementi del dataset suscettibili di autonoma protezione (es., immagini, fotografie, testi in qualche modo creativi) sono stati a loro volta prodotti da uno o più dipendenti della tua pubblica amministrazione nell'ambito della loro attività lavorativa?	
	l'amministrazione è proprietaria dei dati, anche se non sono stati creati direttamente da suoi dipendenti??	
	sei sicuro di non usare dati per i quali vi è una licenza o un brevetto di terzi?	
	se i dati non sono della tua amministrazione hai un accordo o una licenza che ti autorizzi a pubblicarli?	
Licenza di rilascio	stai rilasciando i dati di cui possiedi la proprietà accompagnati da una licenza?	
	hai incluso anche la clausola di salvaguardia "Questo dataset contiene informazioni indirettamente riferibili a persone fisiche. In ogni caso, i dati non possono essere utilizzati al fine di identificare nuovamente gli interessati."?	
Limiti alla pubblicazione	hai verificato che non vi siano impedimenti di legge o contrattuali che per la pubblicazione dei dati?	
Segretezza	hai verificato se non vi siano motivi di ordine pubblico o di sicurezza nazionale che ti impediscono la pubblicazione dei dati?	
	hai verificato se non vi siano motivi legati al segreto d'ufficio che impediscono la pubblicazione dei dati?	
	hai verificato se non vi siano motivi legati al segreto di stato che impediscono la pubblicazione dei dati?	
Temporalizzazione	i dati sono soggetti per legge a restrizioni temporali di pubblicazione?	
	i dati sono aggiornati frequentemente in modo da sanare eventuali informazioni lesive di persone o organizzazioni?	
	i dati hanno dei divieti di legge o giurisprudenziali che impediscono la loro indicizzazione da parte di motori di ricerca?	
Trasparenza	i dati rientrano nella lista dell'allegato A del d.lgs. 33/2013? Se sì come sono stati trattati dal responsabile della trasparenza nella sezione	

	"Amministrazione trasparente"?	

ANALISI DELLA QUALITÀ DEI DATI. All'analisi giuridica delle fonti segue l'analisi della qualità dei dati. Per la definizione del concetto generale di qualità si può ricorrere alla norma ISO 9000:2015, secondo cui la qualità è la totalità degli elementi e delle caratteristiche di un prodotto o servizio che concorrono alla capacità dello stesso di soddisfare esigenze espresse o implicite.

Il presente aggiornamento delle linee guida pone un'attenzione particolare alla qualità dei dati e al relativo monitoraggio con una discussione dedicata "qualità dei dati" di seguito riportata che mira a identificare alcune misure e un metodo di valutazione, considerando gli standard ISO di riferimento ISO/IEC 25012 e il recente ISO/IEC 25024.

BONIFICA. Generalmente l'analisi della qualità del dato può richiedere una fase di bonifica. Infatti, i dati all'interno dei sistemi informativi o degli archivi di un'amministrazione sono spesso "sporchi" e non rispondenti ai requisiti di qualità (e.g., accuratezza, completezza, ecc.). L'apertura dei dati può essere uno stimolo importante per la conduzione di attività mirate di bonifica. Si distinguono processi di bonifica *basati sui dati* e *basati sui processi*. I processi di bonifica *basati sui dati* prevedono che il dataset sia corretto in uno dei due seguenti modi: (i) confronto con il mondo reale (anche con attività economicamente onerose come contattare direttamente i soggetti preposti alla gestione della base dati che presenta errori per correggerli insieme loro) e (ii) confronto incrociato (matching) con altri dataset. Tali processi hanno il vantaggio di poter pervenire in termini relativamente brevi al risultato, ma lo svantaggio di non risolvere il problema alla radice. Infatti, in un arco temporale medio-lungo, i dataset potrebbero nuovamente presentare i problemi di qualità.

I processi di bonifica *basati sui processi* hanno invece la caratteristica di analizzare le cause che hanno portato alla scarsa qualità del dato e di rivedere i processi di produzione del dato per garantirne la qualità nel tempo. Per esempio, se si riscontra che la scarsa accuratezza di una base di dati deriva da un processo di "data entry" manuale, si può intervenire prevedendo una fase di acquisizione automatica dei dati che minimizzi la possibilità di errore di acquisizione. L'adozione di processi di bonifica "basati sui processi" ha dunque il consistente vantaggio di essere una strategia risolutiva.

POLITICHE DI ACCESSO E LICENZA. Altro aspetto importante da considerare sono eventuali forme di aggregazione dei dati e restrizioni di accesso, che hanno anche un impatto sulla scelta della licenza, tappa quest'ultima prevista dal modello operativo e trattata ampiamente in "Aspetti legali e di costo" a cui si rimanda.

Sebbene sia sconsigliato restringere l'accesso ai dati o procedere con la pubblicazione di aggregazioni degli stessi (in generale non è opportuno che l'esposizione del dato lavorato avvenga senza che sia stato pubblicato prioritariamente il dato grezzo), **esistono casi in cui i dati possono essere diffusi solo in forma anonima** (pensiamo ad esempio ai redditi), **ossia a un livello di aggregazione tale da impedire di identificare le persone cui i dati si riferiscon**o. A tal fine, **è bene definire delle politiche di accesso ai dati in cui sia indicato un profilo di accesso specifico per ogni dato, dettato dai diritti sull'informazione di base, dalle norme o dalle policy in atto**.

ANALISI DI PROCESSO, (RE)INGEGNERIZZAZIONE DEI PROCESSI ORGANIZZATIVI E PRODUZIONE DEI DATI. Ogni dato ha un proprio ciclo di vita, caratterizzato da uno specifico tasso di aggiornamento o manutenzione.

Risulta quindi necessario **analizzare il processo organizzativo che produce e gestisce il dato per fare in modo che la produzione di quel dato sia consolidata e diventi stabile, secondo la frequenza di aggiornamento e le modalità di rilascio adottate**.

Vanno quindi individuati non solo i dati nativi "grezzi" di partenza ma anche gli attori che concorrono alla prima produzione del dato, distinguendo chi è responsabile e titolare dello stesso e chi invece aggiunge altri elementi informativi nel processo produttivo. Quello che accade sovente nelle amministrazioni è che i dati sono gestiti da singoli funzionari, nell'ambito di processi "verticali" chiusi a livello di dipartimento e molto spesso ancorati alle conoscenze di una persona specifica. Accade così che elementi conoscitivi importanti siano delocalizzati tra i servizi di competenza, senza che tuttavia sussista una gestione federata e complessiva della risorsa dati. Questo fatto, tra i molteplici effetti

negativi, ha spesso quello della duplicazione dei dati: uffici tematicamente contigui tendono a replicare informazioni funzionali alla propria attività, con un incremento del rumore di fondo attorno al patrimonio informativo dell'amministrazione. L'utilizzo di codici condivisi a livello nazionale, di classificazioni comuni per tipologie di dato non dipendenti da specifici domini e ll passaggio verso la creazione di una risorsa federata (fase data hub interno) consentono di superare progressivamente le suddette criticità. L'impegno politico e il relativo sostegno da parte dei livelli manageriali più alti costituiscono comunque il prerequisito fondamentale senza il quale ogni sforzo può essere vano.

METADATAZIONE. Il risultato delle precedenti tappe del modello operativo si traduce nella produzione di metadati che, in buona sostanza, certificano le caratteristiche del dato. Come detto precedentemente la metadatazione è cruciale: una delle peggiori malattie che affliggono i dati della PA è la molteplicità di copie disponibili di una stessa informazione, senza che sussista la necessaria certezza sulle caratteristiche e sulla validazione di ciascun rilascio. Si ricorda a tal riguardo di seguire il modello per i metadati descritto in "Modello per i dati aperti e i metadati" e in particolare il profilo DCAT-AP_IT che consente di specificare i più importanti metadati descrittivi per i dataset (e.g., soggetti e relativi ruoli, contestualizzazione geografica e temporale, licenza, frequenza di aggiornamento, aspetti di distribuzione, punto di contatto, ecc.).

DATA HUB INTERNO, PRODUZIONE DI LIVELLO 3, E PUBBLICAZIONE. Nel modello operativo proposto, la risorsa federata è rappresentata dal cosiddetto data hub interno. Essa è una piattaforma dove far confluire tutti i dati prodotti dai diversi dipartimenti dell'amministrazione nella loro versione rilasciata ufficialmente. Questa infrastruttura, una volta attivata e messa a regime, viene a contenere lo stato dell'arte del patrimonio informativo e costituisce un potente punto di riferimento, accessibile da parte delle autorità di accesso, secondo diverse modalità (a "tag" o "query"). Essa, inoltre, costituisce lo snodo fondamentale, non solo per la linea dei dati nativi che può proseguire verso la produzione e la pubblicazione di dataset di livello 3, ma per tutte le altre direttrici indicate.

In generale, il data hub interno, presumibilmente creato anche attraverso basi di dati consolidate e mantenute costantemente aggiornate attraverso l'inserimento di dati da parte dei funzionari dell'amministrazione, può essere agevolmente utilizzato per la gestione di un processo dinamico e sostenibile nel tempo di produzione di dati aperti, periodicamente aggiornati a ogni nuova revisione del data hub stesso.

Infine, è bene notare che l'uso degli standard previsti per i livelli 4 e 5 del modello per i dati aperti (i.e., standard del Web semantico, come per esempio RDF e OWL descritti in "Architettura dell'informazione del settore pubblico") può facilitare la definizione e la gestione del data hub interno, consentendo una più agevole integrazione tra i dati del patrimonio informativo.

CONSERVAZIONE E STORICIZZAZIONE. I dataset rilasciati costituiscono non solo una risorsa per la collettività, ma un prezioso patrimonio anche per le pubbliche amministrazioni che possono in questo modo archiviare in modo alternativo i loro dati in modalità indipendente dagli applicativi software originali che li hanno prodotti. Per questo motivo è importante premunirsi di un sistema di archiviazione/conservazione che mantenga le diverse versioni dei dati nel lungo periodo.

A tal fine si raccomanda di assicurare che le versioni stesse siano accessibili a un URL stabile, che sia anche documentato unitamente alla pubblicazione del dato.

Linea 2: Dati Mashup.

Oltre alla pubblicazione dei dati nativi, attività istituzionali multidisciplinari, che coinvolgono più di una pubblica amministrazione, potrebbero rendersi necessarie. Inoltre è cruciale la sensibilità dell'amministrazione rispetto agli stimoli e alle proposte provenienti dalla società civile. A tal riguardo, ogni nuovo dato in questa linea nasce da uno specifico "concept". ovvero la proposta necessaria a definire gli elementi fondamentali di un progetto. All'interno di un "concept" si identifica l'idea generale e le linee guida del progetto che ne accompagnano la declinazione nel corso della fase esecutiva. Al "concept" fa seguito la raccolta delle informazioni dalle diverse fonti interne ed esterne che concorrono alla formazione del dato. Questa operazione di "mashup" (da cui il nome della linea) non implica soltanto la raccolta del dato da fonti diverse e la relativa definizione degli algoritmi di integrazione. La parte più importante è la definizione delle modalità di accesso a partire dalle politiche

dei singoli produttori dei dati e le relative modalità di rilascio e aggiornamento dei dati. Questo tipo di dati, nati a seguito di particolari esigenze o di determinati disegni strategici, sono creati in funzione dell'esposizione al pubblico e del conseguente coinvolgimento. Per questo, essi si prestano a forme di coinvolgimento e visualizzazione ("data visualization") particolarmente innovative che spesso sono definite già a livello di "concept". Il risultato ultimo di questa linea è la produzione di API e/o la pubblicazione di altri dataset. In generale, si raccomanda di utilizzare un approccio di pubblicazione dataset/API, pubblicando come API sicuramente i dataset che necessitano di un aggiornamento dinamico e variabile, alleviando dall'onere dell'aggiornamento manuale.
Si noti infine che i risultati attesi da questa linea possono essere anche ottenuti con l'applicazione dei principi e metodologie previste per la linea 3 dei Linked Open Data, di seguito descritta, grazie ai collegamenti possibili tra i dati.

Linea 3: Linked Open Data.

Nel modello operativo proposto in Figura 5, la linea Linked Open Data è raffigurata come una filiera di lavorazione autonoma in quanto considerata ancora per molte amministrazioni, soprattutto medio piccole, un percorso complesso da intraprendere, dove sono richieste competenze tecniche specifiche.

Tuttavia, l'intenzione delle presenti linee guida è quella di **governare una transizione graduale verso la produzione nativa dei Linked Open Data** e, le recenti iniziative significative in merito da parte dell'ISTAT, dell'ISPRA, del Ministero dell'Economia e Finanze, del Ministero dell'Agricoltura, per citarne alcune, **indicano che tale transizione può essere possibile, soprattutto se trainata da pubbliche amministrazioni centrali e regionali**.

Nel modello operativo, vi è una chiara interconnessione tra la linea dei dati nativi e quella dei Linked Open Data. La connessione tra queste due linee (seppur non illustrata graficamente in Figura 5) è anche rafforzata dal fatto che alcune delle fasi attraversate dalla linea dei dati nativi sono necessarie per avviare, analogamente, il percorso sulla linea dei Linked Open Data. E' altresì importante notare che nella pratica si ritiene a volte necessario passare da modelli di rappresentazione tradizionali come quello relazionale per la modellazione dei dati operando opportune trasformazioni poi per renderli disponibili secondo i principi dei Linked Open Data. Tuttavia tale pratica non è necessariamente quella più appropriata: esistono situazioni per cui può essere più conveniente partire da un'ontologia del dominio e che si intende modellare e dall'uso di standard del Web semantico per poter governare i processi di gestione dei dati.
Sebbene le linee guida della Commissione di Coordinamento SPC[5] sull'interoperabilità semantica attraverso i Linked Open Data siano risalenti al 2012, la metodologia ivi proposta risulta essere ancora valida e solida per una produzione ottimale di Linked Open Data. Infatti, analizzando alcune fasi appartenenti alla linea dei dati nativi (i.e., censimento, analisi della qualità, bonifica e metadatazione) e alla linea dei Linked Open Data (i.e., modellazione, ontologia, inferenza, interlinking, validazione e pubblicazione) si nota come queste richiamino integralmente le sette fasi dell'approccio metodologico delle suddette linee guida. Si incoraggiano quindi le amministrazioni a riferirsi ancora a quel lavoro per affrontare il processo di produzione di Linked Open Data.

[5] http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf

Linea 4: Coinvolgimento (Engagement).

AZIONE 7: DEFINISCI UNA CHIARA STRATEGIA DI COINVOLGIMENTO INTERNO ED ESTERNO…

Si raccomanda alle amministrazioni di accompagnare il modello operativo con azioni di coinvolgimento degli stakeholder sia interni all'amministrazione che esterni.
Il coinvolgimento interno può avvenire attraverso la diffusione della cultura dei dati di qualità e aperti, facendo comprendere l'impatto di questa diffusione anche in termini semplificativi delle procedure interne.
Il coinvolgimento esterno passa in primo luogo dall'identificazione dei soggetti da coinvolgere (e.g., studenti universitari, soggetti preposti a indagini e analisi statistiche e/o economiche, startup e aziende). In secondo luogo esso passa dalla definizione della forma di coinvolgimento, da quella più semplice della comunicazione, anche interattiva, all'individuazione di scenari d'uso affiancati da forme più strutturate di coinvolgimento quali l'organizzazione di eventi per promuovere alcune tipologie di dataset e/o per analizzare casi d'uso, hackaton e app showcase.

Tale percorso si relaziona facilmente anche con il noto modello internazionale a cinque stelle dell'engagement, proposto dal ricercatore inglese Tim Davies per attivare una strategia di rilascio di dataset aperti che sia il più possibile inclusiva. Il modello si compone dei seguenti livelli:

★Essere guidati dalla domanda – pubblicare dati che soddisfino una domanda specifica di stakeholder esterni implica cominciare a ridurre le continue richieste di dati a un ufficio.

★★Inserire dati nel contesto – accompagnare i dati con una ricca documentazione ne permette un facile riutilizzo. Porli nel corretto contesto amplifica tale possibilità. Due ottimi esempi di implementazione di strategia di engagement di livello 2 vengono dal progetto recente Open Cantieri del Ministero delle Infrastrutture e dei Trasporti e dal progetto "OpenCoesione" del Dipartimento per lo Sviluppo e la Coesione Sociale. Il portale OpenCoesione presenta una grafica, corredata da una mappa e diagrammi, che permettono di prendere visione, in maniera efficace, della distribuzione dei fondi sociali europei sul territorio italiano. L'applicazione permette inoltre di scaricare i dati sia nella loro totalità, sia nello specifico caso dei progetti presentati o nelle loro aggregazioni per categoria o amministrazione comunale/provinciale/regionale.

★★★Supportare conversazioni intorno ai dati – Molti cataloghi Open Data ospitano una sezione FAQ e offrono diversi canali di interazione quali email o social network attraverso cui dialogare con l'ente pubblico che distribuisce i dati. Nuovamente, il caso di OpenCoesione può essere visto come una buona iniziativa di coinvolgimento di questo livello in quanto offre la possibilità di usufruire di tali canali per innescare una conversazione online.

★★★★Creare capacità, competenze e reti – in questo livello rientra la fase "scenari d'uso" nel presentare i dati attraverso infografiche interattive si fornisce la possibilità di capire al meglio i dati. Rimane importante però stimolare il riutilizzo organizzando, ove possibile, incontri formativi volti a spiegare i dati e/o a mostrare strumenti di pulizia, analisi, e visualizzazione. Tra gli esempi virtuosi di tali pratiche rientrano "School of Data" dell'Open Knowledge Foundation, i datalab promossi da ISTAT e "A scuola di OpenCoesione" del Dipartimento per lo Sviluppo e la Coesione Sociale e del Ministero dell'Istruzione.

★★★★★Collaborare su dati come una risorsa comune – il rilascio dei dati prevede cicli di feedback con una comunità di riferimento (spesso quella da cui si è partiti per aprire i dati) da cui trarne delle considerazioni e produrre nuovi dati e strumenti. Nuovamente, l'esempio di OpenCoesione fornisce iniziative virtuose di coinvolgimento a cinque stelle quali hackaton organizzati con la comunità e il progetto monithon.it dove, attraverso segnalazioni partendo dai progetti presentati nel sito di OpenCoesione, chiunque può riportare informazioni aggiuntive per stimolare evoluzioni dei progetti finanziati).

Coordinamento tra livello nazionale e livello locale

AZIONE 8: FACILITA IL COORDINAMENTO TRA IL LIVELLO NAZIONALE E LOCALE ATTRAVERSO GLI OPEN DATA…

Diverse pubbliche amministrazioni centrali, al fine di adempiere a specifici obblighi normativi a loro assegnati o per dar seguito a impegni presi in iniziative internazionali (e.g., Open Government Partnership), hanno necessità di raccogliere dati provenienti dal livello di governo locale (e.g., SIOPE per la rilevazione telematica degli incassi e dei pagamenti di tutte le amministrazioni, ISTAT per le rilevazioni relative ai censimenti o ai numeri civici, Dipartimento della Protezione Civile che opera quasi esclusivamente sulla base di tale modello).
In queste situazioni, si raccomanda alle amministrazioni di coordinarsi tra loro prima di intraprendere iniziative singole isolate. In particolare, le amministrazioni centrali possono assumere un ruolo di coordinatore e di promotore di apertura dei dati secondo i livelli più alti del modello per i dati aperti proposto dalle presenti linee guida, definendo anche schemi comuni secondo quanto descritto in "Architettura dell'informazione del settore pubblico".
Si raccomanda poi di mantenere il colloquio, mediante scambio di dati, tra il livello centrale e locale attraverso l'uso dei dati aperti stessi, ove presenti, automatizzando quanto più possibile il processo di acquisizione da parte del livello centrale.

Con un eventuale supporto tecnico, su richiesta, di AgID, si consiglia inoltre di:

- identificare l'insieme minimo di dati rilasciati dal livello centrale, anche secondo quanto stabilito da disposizioni normative, e quelli che il livello locale può ulteriormente dettagliare per cogliere le specificità della propria realtà locale, abilitando ove possibile meccanismi automatici di collegamento tra i due insiemi. Questo consentirebbe di avere una vista nazionale e un unico punto di accesso centrale ai dati, e una vista locale e più specializzata offerta dal governo locale. Si noti che il paradigma dei Linked Open Data può essere particolarmente conveniente in questi casi in quanto il collegamento degli URI consentirebbe un'agevole integrazione dei dati e navigazione degli stessi da parte di programmi;
- documentare sia a livello centrale che locale i dati secondo il profilo nazionale per i metadati DCAT-AP_IT con l'aggiunta dei metadati di provenienza come precedentemente discusso, al fine di agevolare i possibili utilizzatori nel comprendere le diverse fasi di gestione del dato.

QUALITÀ DEI DATI

Il miglioramento della qualità dei dati, e la maggiore diffusione delle tecniche di misurazione, dipende da vari fattori tra cui l'adesione a modelli di qualità condivisi. Il raggiungimento della qualità non è in ogni caso frutto di un impegno sporadico di singole amministrazioni, ma il frutto di una sinergia concertata che, basata su un cambio culturale, si apra a collaborazioni orizzontali che, pur nel rispetto della privacy, consentano un maggior dialogo tra le banche dati e razionalizzazione delle informazioni.

Per determinare la bontà dei dati è necessario definire delle misure attraverso le quali quantificare la qualità dei dati. Lo standard ISO/IEC 25012:2008, divenuto norma italiana UNI ISO/IEC 25012:2014, definisce un insieme di caratteristiche specifiche per la caratterizzazione della qualità dei dati: accuratezza, aggiornamento, completezza, consistenza, credibilità, accessibilità, comprensibilità, conformità, efficienza, precisione, riservatezza, tracciabilità, disponibilità, portabilità e ripristinabilità.

Di queste caratteristiche, le presenti linee guida richiedono la garanzia di almeno quattro come elencate in azione 9, ovvero *accuratezza, coerenza, completezza e attualità (o tempestività di aggiornamento)*.

Il passo successivo è quantificarle in termini di misure, individuando delle soglie che consentano di discriminare la bontà o meno di un dato rispetto alla caratteristica in esame.

La fase di valutazione della qualità dei dati è importante in tutti i sistemi informativi indipendentemente dall'apertura dei dati. Con l'adozione di politiche di apertura dei dati, la qualità dei dati assume un ruolo ancora più rilevante in quanto elemento per la certificazione della bontà dei dati forniti e soprattutto dell'appropriatezza rispetto all'utilizzo che del dato si vuole fare.

L'ISO/IEC 25024 estende l'ISO/IEC 25012 "Data quality model" del 2008 al campo delle misurazioni, definendo 63 misure di qualità applicabili alle 15 caratteristiche di qualità dei dati, con le relative funzioni di calcolo.

AZIONE 9: GARANTISCI LE SEGUENTI DIMENSIONI DI QUALITÀ DEI DATI…

Partendo dalle quattro caratteristiche, delle 15 previste dall'ISO/IEC 25012, individuate nella Determinazione Commissariale n. 68/2013 dell'AgID per le banche dati di interesse nazionale critiche, si garantisce il loro costante rispetto in tutto il processo di gestione e pubblicazione dei dati anche aperti. Queste quattro caratteristiche sono:

- **accuratezza (sintattica e semantica) - il dato, e i suoi attributi, rappresenta correttamente il valore reale del concetto o evento cui si riferisce**
- **coerenza - il dato, e i suoi attributi, non presenta contraddittorietà rispetto ad altri dati del contesto d'uso dell'amministrazione titolare**
- **completezza – il dato risulta esaustivo per tutti i suoi valori attesi e rispetto alle entità relative (fonti) che concorrono alla definizione del procedimento.**
- **attualità (o tempestività di aggiornamento) - il dato, e i suoi attributi, è del "giusto tempo" (è aggiornato) rispetto al procedimento cui si riferisce.**

Per le quattro caratteristiche di qualità, messe in risalto dalla Determinazione Commissariale dell'Agenzia per l'Italia Digitale n. 68/2013, si riporta nella tabella seguente un insieme esemplificativo di misure, sulle 24 definite nello standard ISO per le stesse caratteristiche, a supporto delle attività di valutazione della qualità dei dati delle amministrazioni.

Caratteristiche	Descrizione	Misure e funzioni di misura principali
Completezza	Il grado per cui il dato associato a un'entità presenta valori per tutti gli attributi attesi e	Si individuano le i seguenti livelli di completezza: **1) completezza di schema:** percentuale di valori nulli per concetti e proprietà rispetto al numero totale di valori attesi **2) completezza dei record**: numero di dati elementari associati a un valore non nullo in un record, rispetto al numero di dati elementari del record per cui può essere

	le relative istanze in un certo contesto	misurata la completezza 3) **completezza di popolazione:** percentuale di valori nulli rispetto a una popolazione di riferimento Si noti che non sempre valori mancanti indicano incompletezza. Per esempio: si supponga di considerare dati relativi ai musei italiani e ai loro canali di contatto (telefono ed email). Può capitare che i musei abbiano tutti un indirizzo email ma non per tutti è presente un numero di telefono.
Accuratezza	Il grado in cui gli attributi rappresentano in maniera corretta il valore reale del dato in uno specifico contesto	Si individuano due tipi di accuratezza: 1) **sintattica:** ad esempio Girgia invece che Giorgia 2) **semantica:** ad esempio nel caso in cui si utilizzi Gloria Sani intendendo invece un'altra persona e.g., Giorgia Sani Una misura dell'accuratezza è data dalla ratio tra gli attributi dei dati che hanno valori accurati sintatticamente/semanticamente su il numero di attributi dei dati per i quali è richiesta accuratezza sintattica/semantica.
Coerenza	Il grado in cui gli attributi del dato non sono in contraddizione con altri dati in uno specifico contesto	Per poter valutare la coerenza una misura è quella che consente di identificare le violazioni di regole semantiche definite su alcuni elementi dei dati. Per esempio, se una persona è "patentata" non può essere possibile che la sua età sia "17 anni". Essa può essere calcolata come la ratio tra il numero di attributi dei dati i cui valori sono semanticamente corretti nel dataset sul numero di attributi dei dati per i quali sono state definite delle regole semantiche. Altra misura consiste nel rapporto tra il numero di valori duplicati per ogni attributo della base dati e il numero totale degli elementi della base dati.
Attualità o tempestività	Il grado in cui gli attributi del dato sono al "giusto tempo" rispetto al contesto di riferimento	La metrica è basata sull'uso dei metadati che indicano quando il dato è stato aggiornato l'ultima volta. Sulla base di questi metadati, si distinguono poi: 1) **dati con periodicità di aggiornamento nota:** in questo caso è possibile calcolare la tempestività in maniera esatta identificando se la data di ultima modifica del dato rispetto al tempo di misurazione ricade nell'intervallo della frequenza di aggiornamento; 2) **dati con periodicità di aggiornamento media:** in questo caso è possibile calcolare la tempestività media con una percentuale di errore.

A completamento della suddetta analisi, si ricorda anche un'iniziativa nota dell'Istituto Open Data inglese (ODI) sui certificati Open Data[6]. I certificati sono uno strumento utile per ottenere un'auto-certificazione sulla qualità dei dati prodotti e pubblicati. I certificati sono stati tradotti anche in italiano dal nodo dell'ODI di Trento.

Per ottenere il certificato è necessario compilare un questionario online suddiviso in cinque macro-categorie che aiutano a identificare una scala di riutilizzo di un dataset. Queste sono: informazioni descrittive (molte delle quali già richieste dalle presenti linee guida), informazioni legali (che devono aver già trovato risposte positive ed esaustive mediante la "check list" proposta nella fase di analisi giuridica delle fonti), informazioni pratiche (e.g., reperibilità, note metodologiche, ecc.), informazioni tecniche e informazioni sociali.

Le risposte alle domande producono un livello di certificazione che si distingue in: (i) "bronze", che rappresenta una base per iniziare il processo di apertura dei dati; (ii) "silver", dove il dato è documentato in un formato aperto e machine-readable e gli utilizzatori dei dati possono ricevere maggior supporto; (iii) "gold", che fornisce le garanzie del livello precedente con ulteriori riguardanti l'aggiornamento costante e un più ampio supporto, (iv) "platinum", che racchiude le garanzie gold, identificatori univoci dei dati; rappresenta quindi un'eccellente esempio di infrastruttura informativa.

[6] https://certificates.theodi.org/en/

ARCHITETTURA DELL'INFORMAZIONE DEL SETTORE PUBBLICO

AZIONE 10: RISPETTA L'ARCHITETTURA DELL'INFORMAZIONE DEL SETTORE PUBBLICO CON I RELATIVI STANDARD, FORMATI E VOCABOLARI…

Si adotta l'architettura dell'informazione del settore pubblico come mostrata in Figura 6. Per tutti i dati di riferimento e "core", si raccomanda di *non* ridefinire degli schemi o modelli per i dati ma di riutilizzare quelli dell'architettura nazionale dell'informazione del settore pubblico, in larga parte disponibili come standard aperti del Web e in formati aperti. Tali schemi possono anche essere estesi dalle amministrazioni in base alle proprie esigenze di modellazione, nel rispetto tuttavia delle regole di conformità agli schemi stessi e del principio di apertura per la loro pubblicazione e fruizione.
Questa raccomandazione si applica anche ai dati di domini verticali, dove in alcuni casi lavori consolidati per la definizione di schemi comuni sono già stati intrapresi a livello internazionale e/o nazionale.

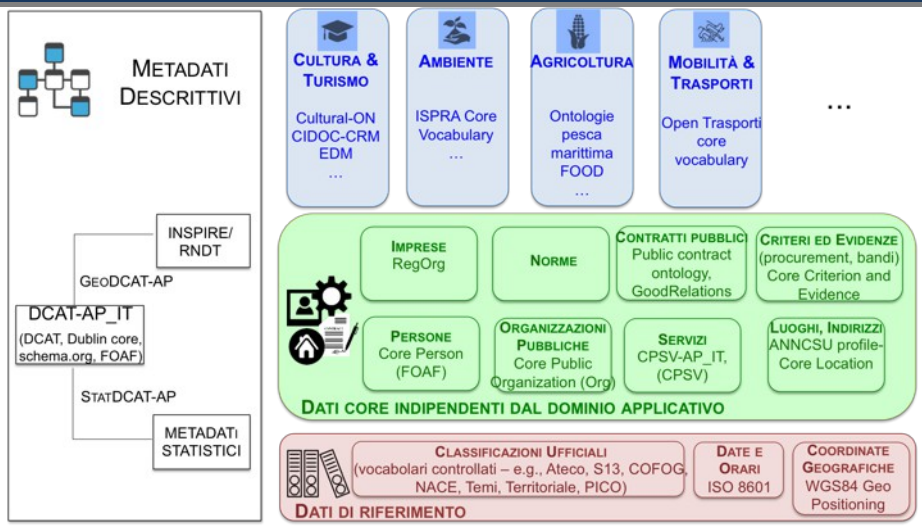


Figura 6: Architettura dell'informazione del settore pubblico

La Figura 6 rappresenta un primo tentativo di delineare l'architettura di riferimento per l'informazione del settore pubblico.

La figura non ha la pretesa di essere esaustiva rispetto a dati specifici delle pubbliche amministrazioni ma **classifica alcune tipologie di dati, indicando per ognuna vocabolari noti e condivisi a livello internazionale che definiscono modelli dati di cui si incoraggia l'adozione**.

L'obiettivo è quello di individuare, in linea generale, degli schemi da condividere tra tutte le amministrazioni al fine di rappresentare dati ricorrenti, indipendenti dallo specifico dominio applicativo, come per esempio i dati sulle persone, sulle organizzazioni pubbliche e private, sui luoghi e gli indirizzi, ecc.. Si ritiene questo possa inoltre facilitare la creazione di collegamenti tra dati (in figura sono riportati vocabolari e classificazioni già disponibili anche secondo il paradigma dei Linked (Open) Data), portando alla costruzione di una grande base di conoscenza dell'informazione del settore pubblico da utilizzare per lo sviluppo di nuovi e proattivi servizi.
Nell'architettura si identificano due livelli: *dati di riferimento* e *dati core indipendenti dal dominio applicativo*. Essi consistono di quei dati identificati univocamente e necessari per gestire e utilizzare in maniera affidabile infrastrutture di interesse nazionale e per interfacciare più agevolmente altri dati dipendenti da domini verticali. Il livello dei "dati di riferimento" consiste, in particolare, di tutte le classificazioni

ufficiali che si raccomanda di utilizzare in quanto di riferimento per svariati contesti, e di dati relativi a informazioni temporali (date e orari) e geografiche (coordinate geografiche). Nell'ambito di questo livello si evidenziano pertanto classificazioni come quella territoriale, rilasciata dall'ISTAT anche sotto forma di LOD[7], quella sui temi (o domini), applicabile sia al contesto dei dati che a quello dei servizi[8], quella relativa alle funzioni amministrative/di governo (COFOG), anch'essa disponibile secondo il paradigma LOD e già adottata nell'ambito del bilancio pubblico, per citarne alcune. Infine, per i dati temporali e geografici comuni, come indicato in Figura 6, si raccomanda, rispettivamente, l'uso dello standard ISO 8601 16 e del vocabolario del W3C per la definizione delle coordinate geografiche 17.

Il livello dei "dati core indipendenti dal dominio applicativo" (o dati core orizzontali) consiste dell'insieme di tipologie di dati riferibili principalmente a soggetti, luoghi, organizzazioni, servizi e altri asset e requisiti tipici della pubblica amministrazione. A oggi, sono state individuate otto tipologie di dati "core", come rappresentate in Figura 6: persone, organizzazioni pubbliche, servizi, luoghi e indirizzi, imprese, contratti pubblici, criteri ed evidenze, norme. Molte di queste tipologie sono direttamente collegabili a banche dati di interesse nazionale; per esempio, la tipologia "persone" si collega all'Anagrafe Nazionale della Popolazione Residente (ANPR), la tipologia "luoghi e indirizzi" è correlata all'Anagrafe Nazionale dei Numeri Civici e delle Strade Urbane, la tipologia "contratti pubblici" è connessa alla banca dati di interesse nazionale definita dall'articolo 60 del CAD sui contratti pubblici, e così via.

L'architettura dell'informazione del settore pubblico individua, per alcune di queste otto tipologie, vocabolari particolarmente diffusi nel Web e in ambito europeo che hanno il merito di proporre degli schemi di dati condivisi (entità, relazioni tra entità e proprietà/attributi) per la loro rappresentazione. In particolare, si adottano i cosiddetti "Core Vocabulary", definiti dalla commissione europea nell'ambito del programma ISA sull'interoperabilità semantica e in parte standardizzati dal W3C.

Nel dettaglio, per la rappresentazione delle ***persone*** **si raccomanda l'uso del profilo Core Person 18** che si basa sul vocabolario FOAF (Friend Of A Friend), definito per descrivere persone e relazioni sociali tra loro. Per la modellazione dei dati sulle ***organizzazioni pubbliche*** **si raccomanda l'uso del Core Public Organization Vocabulary 19, basato principalmente sullo standard del web Org 20**, quest'ultimo definito con l'obiettivo di rappresentare dati sulle organizzazioni e già utilizzato nel contesto dello sviluppo LOD dell'Indice della Pubblica Amministrazione (IPA) e in altri casi di dati aperti italiani. Per le ***imprese*** **si raccomanda invece il vocabolario RegOrg 21** che nasce come specializzazione della suddetta ontologia Org per tutte quelle organizzazioni private iscritte in registri pubblici (e.g., il registro imprese – banca dati di interesse nazionale ai sensi dell'articolo 60 del CAD).

Per quanto riguarda **i *servizi*, offerti dalle amministrazioni per il beneficio di cittadini, professionisti e imprese, si richiede di utilizzare per la loro rappresentazione il profilo di interoperabilità semantica definito a livello nazionale** come estensione del Core Public Service Vocabulary. Il profilo è detto **CPSV-AP_IT 22,** la specifica e la relativa ontologia sono pubblicate nella sezione ontologie di dati.gov.it e sono utilizzate per la modellazione del catalogo nazionale per i servizi pubblici servizi.gov.it[9].

Per quanto concerne i ***luoghi e gli indirizzi*,** si segnala che nell'ambito del piano di azione OGP (Open Government Partnership) italiano, l'ISTAT rilascerà, secondo il paradigma LOD ed entro il 2017, i dati dell'Anagrafe Nazionale dei Numeri Civici e delle Strade Urbane (ANNCSU). A tal proposito **si raccomanda la definizione di uno specifico profilo di interoperabilità che possa essere adottato da tutte le amministrazioni per la rappresentazione di questi dati. Si raccomanda di definire il profilo sulla base del vocabolario Core Location 23 che nasce per rispondere a tali esigenze, proponendo uno schema dati conforme ai requisiti dettati dalla direttiva INSPIRE** e già adottato in altri paesi europei come il Belgio per l'apertura dell'analoga base di dati. Infine, nell'ambito dei Core Vocabulary, **si raccomanda l'uso del Core Criterion and Evidence 24 per la modellazione di *criteri e di evidenze*** ovvero requisiti utilizzati per giudicare o prendere decisioni, e prove che qualcosa è avvenuto o che criteri specifici sono stati rispettati da

[7] http://datiopen.istat.it/datasetOntologie.php?call=ontologie

[8] http://publications.europa.eu/mdr/resource/authority/data-theme/skos/data-theme-skos.rdf

[9] Una versione beta del catalogo sarà disponibile online il prossimo inverno 2016.

parte di soggetti. Tale vocabolario è particolarmente **utile nei casi di modellazione di informazioni relative al procurement e a bandi e gare pubbliche**, strumenti tipicamente adottati dalle amministrazioni per lo svolgimento di alcune delle loro attività istituzionali. Questo vocabolario può essere utilizzato insieme a **quelli raccomandati per la rappresentazione dei dati sui *contratti pubblici* come l'ontologia Public Contract 25 e GoodRelations 26.**

A partire dai livelli dell'architettura sopra citati, è possibile collocare e costruire modelli per dati specifici di domini verticali. In Figura 6 sono mostrati solo alcuni domini a titolo di esempio, con l'indicazione di vocabolari in taluni casi già sviluppati da amministrazioni centrali, come il caso del Ministero dei Beni e Attività Culturali e del Turismo (MIBACT) che ha deciso di adottare l'ontologia Cultural-ON[10] per i luoghi e gli eventi culturali e dell'ISPRA che ha recentemente rilasciato una piattaforma LOD che include le ontologie per i dati sul consumo del suolo, sulla rete mareografica e ondametrica e sui sistemi di cartografia che, grazie anche ai collegamenti abilitati tramite il paradigma Linked Data, sono stati collegati con successo alla classificazione territoriale di riferimento pubblicata dall'ISTAT.

L'architettura si compone poi del livello verticale dei metadati descrittivi che coinvolge tutti i tipi di dati fin qui discussi. Punto di riferimento per i metadati descrittivi è DCAT-AP_IT con le sue estensioni per i dati geografici e statistici che consentono un raccordo con i rispettivi profili come definiti nel contesto del Repertorio Nazionale dei Dati Territoriali (RNDT) e dall'ISTAT.

L'architettura di riferimento per l'informazione del settore pubblico si completa con l'indicazione degli standard e dei formati, descritti di seguito, che possono essere utilizzati per rappresentare i dati che la compongono.

Si raccomanda in generale di rendere disponibili in forma Open Data tutti i dati di riferimento. Si raccomanda altresì di prediligere tale paradigma per i dati core indipendenti dal dominio, prestando attenzione ai dati a conoscibilità limitata e ai dati personali per i quali il paradigma non può applicarsi (si veda "Dati della Pubblica Amministrazione").

16. ISO 8601 – Date and Time format, http://www.iso.org/iso/home/standards/iso8601.htm, 2016.
17. W3C, WGS84 Geo Positioning: an RDF vocabulary, https://www.w3.org/2003/01/geo/wgs84_pos, 2016
18. ISA programme, Core Person, https://joinup.ec.europa.eu/asset/core_person/asset_release/core-person-vocabulary#download-links, 2016
19. ISA programme, Core Public Organization Vocabulary, https://joinup.ec.europa.eu/asset/cpov/asset_release/core-public-organisation-vocabulary-draft-4#download-links, 2016.
20. W3C Reccommendation, The Organization Ontology, https://www.w3.org/TR/vocab-org/, gennaio 2014
21. W3C Working Group Note, Registered Organization Vocabulary, https://www.w3.org/TR/vocab-regorg/, agosto 2013
22. Agenzia per l'Italia Digitale, CPSV-AP_IT, http://www.dati.gov.it/onto/cpsvapit, 2016
23. W3C, ISA Programme Location Core Vocabulary, https://www.w3.org/ns/locn, 2016
24. ISA Programme, Core Criterion and Core Evidence Vocabulary, https://joinup.ec.europa.eu/asset/criterion_evidence_cv/asset_release/core-criterion-and-core-evidence-vocabulary-draft-4#download-links, 2016
25. Public Contract Ontology, https://github.com/opendatacz/public-contracts-ontology, 2016
26. Good Relations, http://www.heppnetz.de/projects/goodrelations/, 2016

[10] L'ontologia non è ancora stata pubblicata ufficialmente dal Ministero, ma è stata da esso segnalata ad AgID e presentata in anteprima nel contesto del workshop "*Linked open data per i beni culturali: iniziative e prospettive*" organizzato dall'Istituto per i beni artistici culturali e naturali della Regione Emilia-Romagna in collaborazione con il Ministero dei Beni e delle Attività Culturali e del Turismo (Ferrara, Aprile 2016).

STANDARD DI RIFERIMENTO

I principali standard di riferimento per l'architettura dell'informazione del settore pubblico, necessari anche ad abilitare i livelli 4 e 5 del modello dei dati e i livelli 3 e 4 del modello dei metadati derivano dalle esperienze maturate dagli esperti nel settore del Web Semantico, con la visione di trasformare il Web in un unico spazio informativo globale. Essi sono riportati nella tabella seguente.

Standard	Descrizione
RDF (Resource Description Framework)27	È un framework per la rappresentazione dell'informazione nel Web e uno degli standard alla base del Web Semantico. Esso consente di catturare la semantica dei dati, quindi la loro comprensibilità, facilitandone l'accessibilità da parte di agenti automatici tramite l'infrastruttura e i protocolli Internet esistenti. In una concezione astratta della realtà, ogni oggetto e ogni entità (reale o virtuale) sono risorse. Associando a ogni risorsa un identificativo univoco, nello specifico un URI (Uniform Resource Identifier), si rappresentano nel Web le informazioni relative alle risorse, rendendole accessibili e riferibili da tutti. Tecnicamente, RDF è un framework concettuale che consente, sfruttando la suddetta identificazione delle risorse, di descriverle mettendole in relazione tra loro. RDF ha un solo costrutto informativo di base, la cosiddetta tripla `<soggetto> <predicato> <oggetto>.` Un soggetto è sempre una risorsa (i.e., il suo URI), un oggetto è una risorsa o un valore (in quest'ultimo caso un'espressione puramente simbolica come un numero, una stringa, una data, ecc.), un predicato è una relazione, cui è associato un tipo, tra due risorse o una proprietà di una risorsa. Si noti che anche i predicati sono rappresentati con URI. In questo modo le risorse sono descritte tramite delle relazioni aventi un significato ben preciso e inserite in un particolare contesto. Le triple RDF sono strutture ricorsive, soggetto-verbo-oggetto (come nel caso del linguaggio naturale). La concatenazione di triple genera un "grafo RDF"; pertanto, un insieme di dati rappresentati attraverso il framework RDF è un grafo. Lo spazio Web in cui dati RDF sono localizzati è il cosiddetto Web dei Dati ("Web of Data"), mentre la sua prospettiva, focalizzata maggiormente sul contenuto informativo, è detta Web Semantico. RDF può essere implementato attraverso diverse forme sintattiche, anche dette serializzazioni, quali RDF/XML, Notation3, N-Triple, Turtle e JSON-LD (si veda sotto). La scelta tra le diverse soluzioni sintattiche deve essere fatta sulla base di requisiti richiesti quali compattezza, spazio fisico utilizzato, leggibilità, ecc. Le serializzazioni sono comunque fra loro inter-traducibili. Infine, esiste la possibilità di poter includere informazioni RDF all'interno di pagine Web mediante il formalismo RDFa (RDF in Attributes) 28.
RDFS (RDF Schema)29	È un'estensione di RDF che permette di definire semplici schemi per i dati. Lo standard introduce alcuni costrutti come le classi (rdfs:Class), le collezioni (ad esempio, rdfs:List) e una serie di proprietà per poter definire tassonomie e relazioni tra classi e proprietà (ad esempio, rdfs:subClassOf, rdfs:subPropertyOf). In pratica, con RDFS si possono gestire relazioni insiemistiche, ereditarietà e vari tipi di vincoli. Gli schemi definiti con RDFS sono comunemente detti ontologie.
OWL (Ontology Web Language)30	Mentre RDFS consente di definire semplici schemi per dati RDF, schemi più evoluti possono essere definiti tramite OWL, uno standard W3C che arricchisce RDFS con ulteriori formalismi, includendo semantica formale e logica descrittiva. Un'ontologia consente in modo preciso ed efficace di modellare un dominio di interesse, quindi i suoi oggetti e le relazioni tra questi. In pratica, OWL fornisce il pieno supporto alla definizione di ontologie. Molte ontologie, nate per rappresentare le informazioni di domini ben precisi, sono note e condivise globalmente. Questa condivisione agevola la comprensione e il riutilizzo di schemi e metadati, abilitando di conseguenza l'interoperabilità semantica tra sistemi

	differenti. L'aspetto logico delle ontologie fornisce la possibilità di verificare automaticamente la correttezza logica di ciò che si rappresenta. Inoltre i cosiddetti ragionatori automatici per le logiche descrittive consentono di inferire, sui dati conformi all'ontologia, nuove triple e quindi informazione addizionale.
SPARQL (Sparql Protocol And Rdf Query Language)31	Tra le diverse proposte di linguaggi di interrogazione per dati RDF, il W3C ha standardizzato SPARQL. Una semplice interrogazione SPARQL si compone di una concatenazione di triple in cui alcuni elementi possono essere delle variabili incognite. L'esecuzione di una interrogazione SPARQL cerca tra i dati le concatenazioni di triple "conformi" a quelle dell'interrogazione, assegnando (i.e., istanziando) degli URI o dei valori alle variabili che possono anche essere restituiti in output. È anche possibile specificare operazioni di manipolazione dei dati, come ad esempio istruzioni di "insert", "update" e "delete". SPARQL non è solo un linguaggio di interrogazione ma è un protocollo completo per l'accesso ai dati in quanto definisce anche le modalità con cui le interrogazioni possono essere eseguite via Web (appoggiandosi al protocollo HTTP) e come i risultati devono essere restituiti all'utente. I servizi Web che implementano il protocollo SPARQL sono detti SPARQL endpoint.
SDMX (Statistical Data and Metadata eXchange)32	È uno standard ISO per lo scambio di dati statistici basato su sintassi XML. Esso implementa al suo interno un modello dati per la rappresentazione di dati multidimensionali. Pertanto descrive la struttura di un particolare "dataflow" attraverso un insieme di dimensioni (e.g., territorio o tempo), un insieme di attributi (e.g., unità di misura) e le classificazioni associate. Si nota che sebbene SDMX sia nato come modello per lo scambio di dati, esso viene anche usato per la loro rappresentazione.

27. W3C, RDF (Reseource Description Framewok) https://www.w3.org/standards/techs/rdf#w3c_all, 2016
28. W3C, RDFa, https://www.w3.org/standards/techs/rdfa#w3c_all, 2016.
29. W3C Recommendation, RDF Schema 1.1, https://www.w3.org/TR/rdf-schema/, 25 febbraio 2014.
30. W3C, OWL – Web Ontology Language, https://www.w3.org/OWL/, 2016.
31. W3C Reccommendation, SPARQL 1.1 Query Language, https://www.w3.org/TR/sparql11-query/, 21 marzo 2013.
32. SDMX, https://sdmx.org/, 2016.

FORMATI APERTI PER I DATI E I DOCUMENTI

AZIONE 11: SELEZIONA I FORMATI CHE MEGLIO SI ADATTANO AL CONTENUTO E AI DATI DA CONDIVIDERE E RILASCIARE…

Si adottano formati aperti senza assumere che gli utenti possano leggere formati proprietari. Nel caso inevitabile di rilascio in formati proprietari, è necessario assicurare la disponibilità anche di un'alternativa non proprietaria.
È necessario evitare di utilizzare un formato per dati non strutturati (e.g., PDF) in presenza di dati strutturati (e.g., è da evitare la pubblicazione di tabelle di tassi di assenza in PDF, privilegiando un formato come il CSV). Si raccomanda inoltre, nel rilasciare i dati secondo i formati sotto riportati, di specificare la codifica dei caratteri privilegiando, ove possibile, UTF 8 50.
Infine, nel caso di rilascio programmato di dati, è da evitare l'uso di formati per dati non strutturati, privilegiando formati "machine-readable".
Nel caso di documenti, sono da evitare scansioni di documenti cartacei in quanto non accessibili e quindi non aperti. In generale, si raccomanda di adottare, ove esistano, standard XML documentali internazionali o nazionali.

Formati aperti per i dati

Formato	Descrizione
XML (eXtensible Markup Language) 33	È un linguaggio di marcatura standardizzato dal W3C usato per l'annotazione di documenti e per la costruzione di altri linguaggi più specifici per l'annotazione di documenti (e.g., XBRL per la rappresentazione dei bilanci, Normattiva per la rappresentazione di documenti informatici in ambito giuridico, ecc.). Il mondo legato all'XML è molto ampio e la sua trattazione non rientra tra gli obiettivi del presente documento. Nell'ambito del Web Semantico è stata definita una specifica serializzazione RDF/XML.
N-Triples 36	È una serializzazione di RDF in cui ogni tripla è espressa interamente e indipendentemente dalle altre. La concatenazione delle triple di un dataset RDF secondo N-Triples avviene utilizzando il carattere punto (i.e., <soggetto1> <predicato1> <oggetto1> . <soggetto2> <predicato2> <oggetto2>).
Notation3 34	Notation3 (o N3) è una serializzazione RDF pensata per essere più compatta rispetto a quella ottenuta utilizzando la sintassi XML Essa risulta più leggibile da parte degli utenti e possiede delle caratteristiche che esulano dall'uso stretto di RDF (e.g., rappresentazione di formule logiche).
Turtle 35	È una versione semplificata (un sottoinsieme di funzionalità) di N3. Un dataset in Turtle è una rappresentazione testuale di un grafo RDF e, al contrario di RDF/XML, è di più facile lettura e gestione anche manuale.
JSON (JavaScript Object Notation) 37	È un formato aperto per la rappresentazione e lo scambio di dati semi-strutturati, leggibile anche dagli utenti e che mantiene, rispetto a formati simili come l'XML, una sintassi poco prolissa. Questo aspetto ne fa un formato flessibile e compatto. Esso nasce dalla rappresentazione di strutture dati semplici nel linguaggio di programmazione JavaScript, ma mantiene indipendenza rispetto ai linguaggi di programmazione.
JSON-LD 38	È un formato di serializzazione per RDF, standardizzato dal W3C, che fa uso di una sintassi JSON. Viene proposto come formato per Linked Data, mascherando di proposito la sua natura di serializzazione di RDF per ragioni di diffusione del formato. Il gruppo di lavoro che l'ha definito ha posto come obiettivo, oltre quello di mettere a disposizione un'ulteriore funzionalità al framework RDF, anche quello di avvicinare il mondo dello sviluppo Web e degli utilizzatori dei sistemi di gestione dati

	NoSQL (in particolare dei document store) al Web Semantico. Da un punto di vista pratico è possibile rilasciare dati RDF utilizzando questo "dialetto" JSON nelle situazioni in cui inizialmente non ci si possa dotare di tecnologie ad-hoc come triple store. Allo stesso tempo, con JSON-LD si fornisce uno strumento standard che consente il collegamento di documenti JSON che per loro natura sono unità di informazione indipendenti.
CSV (Comma Separated Values)	È un formato di file testuale utilizzato per rappresentare informazioni con struttura tabellare. Esso è spesso usato per importare ed esportare il contenuto di tabelle di database relazionali e fogli elettronici. Le righe delle tabelle corrispondono a righe nel file di testo CSV e i valori delle celle sono divisi da un carattere separatore, che di solito, come indica il nome stesso, è la virgola. Il CSV non è uno standard vero e proprio ma la sua modalità d'uso è descritta nell'RFC 4180 39.. Nel rilascio di dati secondo il formato CSV, per agevolare i riutilizzatoti, si raccomanda di dichiarare almeno 1) il separatore di campo utilizzato (e.g, virgola, punto e virgola); 2) se è stato usato un carattere per delimitare i campi di testo[11]. Nel corso del 2015, un gruppo di lavoro del W3C "CSV on the web" ha rilasciato una serie di standard del Web tra cui alcuni relativi ai meccanismi necessari a trasformare CSV in vari formati quali JSON, XML e RDF. Per gli scopi del presente aggiornamento delle linee guida, si raccomanda di considerare due standard: "Generating JSON from Tabular Data on the Web" 40 e "Generating RDF from Tabular Data on the Web" 41.

Formati aperti più diffusi per i dati geografici

Formato	Descrizione
Shapefile 42	È il formato standard de-facto per la rappresentazione dei dati dei sistemi informativi geografici (GIS). I dati sono di tipo vettoriale. Lo shapefile è stato creato dalla società privata ESRI che rende comunque pubbliche le sue specifiche. L'apertura delle specifiche ha consentito lo sviluppo di diversi strumenti in grado di gestire e creare tale formato. Seppur impropriamente ci si riferisca a *uno* shapefile, nella pratica si devono considerare almeno tre file: un .shp contenente le forme geometriche, un .dbf contenente il database degli attributi delle forme geometriche e un file .shx come indice delle forme geometriche. A questi tre si deve anche accompagnare un file .prj che contiene le impostazioni del sistema di riferimento. Si raccomanda comunque di specificare nei metadati la proiezione utilizzata. È importante notare che non risulta ancora chiaro se tale formato lo si possa considerare propriamente aperto (e quindi coerente con la definizione introdotta dall'art. 68 del CAD) di livello 3 secondo il modello per i dati proposto nel presente documento. Questo è dovuto al fatto che, per alcune comunità, esso è un formato proprietario e quindi di livello 2, mentre per altre i dati possono essere gestiti attraverso una serie di strumenti non necessariamente confinati a determinate tipologie software (grazie alle specifiche tecniche aperte e pubbliche rese disponibili da ESRI). Tenuto conto dell'ampio uso di tale formato per la rappresentazione dei dati geografici si ritiene opportuno includerlo comunque in questo elenco.
KML 43	È un formato basato su XML per rappresentare dati geografici. Nato con Google, è diventato poi uno standard OGC. Le specifiche della versione 2.2 presentano una serie di entità XML attraverso cui archiviare le coordinate geografiche che rappresentano punti, linee e poligoni espressi in coordinate WGS84 e altre utili a definire gli stili attraverso cui visualizzare i dati. Eventuali attributi delle geometrie vanno espressi invece attraverso la personalizzazione di alcune entità. Molti strumenti di conversione non si occupano tuttavia di creare questa struttura dati e delegano gli attributi delle geometrie allo stile di visualizzazione. Si consiglia pertanto di distribuire

[11] http://specs.frictionlessdata.io/csv-dialect/#specification

	questo dato prestando attenzione o, eventualmente, accompagnando il dataset assieme ad un altro formato aperto per i dati geografici (ad esempio, .shp, .geojson).
GeoJSON 44	È un formato aperto per la rappresentazione e l'interscambio dei dati territoriali in forma vettoriale, basato su JSON. Ogni dato è codificato come oggetto che può rappresentare una geometria, una caratteristica o una collezione di caratteristiche. A ogni oggetto è associato un insieme di coppie nome/valore (membri). I principali nomi di membri che rappresentano le caratteristiche dei dati geografici sono: "type" che serve a indicare il tipo di geometria (punto, linea, poligono o insieme multi-parte di questi tipi); "coordinates" attraverso cui sono indicate le coordinate dell'oggetto in un dato sistema di riferimento; "bbox" attraverso cui sono indicate le coordinate di un riquadro di delimitazione geografica; "crs" (opzionale) per l'indicazione del sistema di riferimento. Inoltre, è possibile associare all'oggetto specifici attributi, attraverso il membro con nome "properties". Si tratta di un formato molto diffuso e supportato da diversi software, ampiamente utilizzato in ambito di sviluppo web. Lo scorso agosto 2016 è stata pubblicata la relativa RFC 7946 "The GeoJSON Format" 49. La specifica raccomanda di limitare la precisione delle coordinate a 6 decimali, attraverso cui si può specificare qualsiasi posizione sulla terra con una tolleranza di 10 centimetri. La specifica inoltre richiede che i dati siano memorizzati con un sistema di riferimento di coordinate geografiche WGS 84, in latitudine e longitudine, nello stesso stile dei dati GPS.
GML (Geography Markup Language) 45	È una grammatica XML che rappresenta un formato di scambio aperto per i dati territoriali. Definita originariamente da OGC, e diventata poi lo Standard ISO 19136:2008, essa fornisce la codifica XML (schemi XSD) delle classi concettuali definite in diversi Standard ISO della serie 19100 e di classi aggiuntive quali: geometrie, oggetti topologici, unità di misura, tipi di base, riferimenti temporali, caratteristiche, sistemi di riferimento, copertura.
GeoPackage 46	È un formato aperto per la rappresentazione di dati geografici e può essere un'alternativa al suddetto formato shapefile. Esso supporta SpatiaLite ovvero un'estensione dello schema del database SQLite. Il principale vantaggio offerto da GeoPackage è quello di rappresentare in un unico file diversi dati geografici, sia di tipo vettoriale che raster, che possono essere gestiti anche tramite apposite interrogazioni SQL. Lo standard è riconosciuto dall'Open Geospatial Consortium.

Formati aperti per i documenti

L'articolo 1 del nuovo CAD definisce:

> *p) documento informatico: lil documento elettronico che contiene la rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti;*
>
> *p-bis) documento analogico: la rappresentazione non informatica di atti, fatti o dati giuridicamente rilevanti.*

Il contesto normativo del recepimento della direttiva relativa al riutilizzo dell'informazione del settore pubblico (D.lgs 2006, 24 gennaio, n. 36 – art. 2), definisce il documento come "*la rappresentazione di atti, fatti e dati a prescindere dal supporto nella disponibilità della pubblica amministrazione o dell'organismo di diritto pubblico. La definizione di documento non comprende i programmi informatici*".

Formato	Descrizione
ODF (Open Document Format) 47	È uno standard dell'OASIS che specifica le caratteristiche di un formato per documenti digitali basato su XML, indipendente dall'applicazione e dalla piattaforma utilizzata. La seguente serie di formati aperti è parte dello standard OASIS ODF: • ODT (Open Document Text). Standard aperto per documenti testuali. È stato adottato come formato principale per i testi in alcune suite per l'automazione d'ufficio come OpenOffice.org e LibreOffice; è supportato da altre come Microsoft Office, Google Drive e IBM Lotus. • ODS (Open Document Spreadsheet). Standard aperto per fogli di calcolo.

	Come nel caso precedente, è stato adottato come formato principale per i fogli di calcolo in alcune suite per l'automazione d'ufficio come OpenOffice.org e LibreOffice; è supportato da altre come Microsoft Office, Google Drive e IBM Lotus. • ODP (Open Document Presentation). Standard aperto per documenti di presentazione. È stato adottato come formato principale per i documenti di presentazione in alcune suite per l'automazione d'ufficio come OpenOffice.org e LibreOffice; è supportato da altre come Microsoft Office, Google Drive e IBM Lotus.
PDF	È un formato aperto creato da Adobe per la rappresentazione di documenti contenenti testo e immagini che sia indipendente dalla piattaforma di lettura (applicativo, sistema operativo e hardware). È stato standardizzato dall'ISO (ISO/IEC 32000-1:2008) con una serie di formati differenti, ognuno avente una propria prerogativa (e.g., PDF/UA per l'accessibilità, PDF/H per documenti sanitari, PDF/A per l'archiviazione, ecc.). Si noti che rilasciare *dati* secondo tale formato limita fortemente il riutilizzo dei dati stessi in quanto l'intervento umano richiesto per la loro elaborazione è molto elevato (dati rilasciati in formato PDF con una licenza aperta rappresentano solo il primo livello del modello dei dati aperti).
Akoma Ntoso 48	È un linguaggio basato su XML per la rappresentazione di documenti giuridici. È in fase di approvazione presso il consorzio OASIS ed è utilizzato dal Parlamento Europeo e dalla Commissione Europea come standard documentale per i documenti legislativi, giuridici e allegati tecnici.

33. W3C, Recommendation, Extensible markup language (XML) 1.0, http://www .w3.org/TR/xml/, novembre 2008
34. W3C, W3C, "Notation3 (N3): A readable RDF syntax", http://www.w3.org/TeamSubmission/n3/, 2016
35. W3C Recommendation, RDF 1.1. Turtle, https://www.w3.org/TR/2014/REC-turtle-20140225/, febbraio 2014
36. W3C Recommendation, RDF 1.1. N-Triples https://www.w3.org/TR/2014/REC-n-triples-20140225/, febbraio 2014
37. IETF, RFC 4627 - The application/json Media Type for JavaScript Object Notation (JSON), http://www.ietf.org/rfc/rfc4627.txt, 2016
38. W3C, Recommendation, "JSON-LD 1.0", https://www.w3.org/TR/json-ld/, gennaio 2014
39. IETF, RFC 4180 - Common Format and MIME Type for Comma Separated Values (CSV) Files, http://tools.ietf.org/html/rfc4180 , 2016.
40. W3C Recommendation – Generating JSON from Tabular Data on the Web https://www.w3.org/TR/csv2json/, dicembre 2015
41. W3C Recommendation – Generating RDF from Tabular Data on the Web https://www.w3.org/TR/csv2json/, dicembre 2015
42. ESRI, "Shapefile Technical Description", http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf, luglio 1998
43. OGC, KML, http://www.opengeospatial.org/standards/kml, 2016
44. geoJSON, http://geojson.org/, 2016.
45. OGC, Geography Markup Language – GML",http://www.opengeospatial.org/standards/gml, 2016
46. OGC, GeoPackage http://www.geopackage.org/, 2016.
47. OASIS, "Open Document Format for Office Applications", https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office-collab, 2016.
48. Akoma Ntoso, XML for parliamentary, legislative & judiciary documents, http://www.akomantoso.org/, 2016.
49. IETF, RFC 7946 "The GeoJSON format", https://tools.ietf.org/html/rfc7946, agosto 2016
50. UTF 8, https://tools.ietf.org/html/rfc3629, novembre 2003

ASPETTI LEGALI E DI COSTO

LICENZE

AZIONE 12: ASSICURATI DI ASSEGNARE UNA LICENZA AI DATASET…

L'informazione sul tipo di licenza è metadato indispensabile per determinare come poter riutilizzare il dataset. Deve pertanto essere *sempre* specificata indicando, il nome, la versione e fornendo il riferimento al testo della licenza. Nel contesto dei dati aperti, considerando la definizione Open Data fornita dal CAD e dall'Open Knowledge Foundation (OKFN), per cui un dato è aperto se è "*liberamente usabile, riutilizzabile e ridistribuibile da chiunque per qualsiasi scopo, soggetto al massimo alla richiesta di attribuzione e condivisione allo stesso modo,* le sole licenze ammesse per abilitare l'effettivo paradigma dell'Open Data sono classificate come mostrato in Figura 6. Come evidenziato in figura, tutte le licenze che non consentono lavori derivati, anche per finalità commerciali, i.e., licenze che riportano chiaramente clausole Non Commercial - NC e/o Non Derivative – ND e/o ogni altra clausola che limita la possibilità di riutilizzo e ridistribuzione dei dati, non possono essere ritenute valide per identificare dataset aperti.

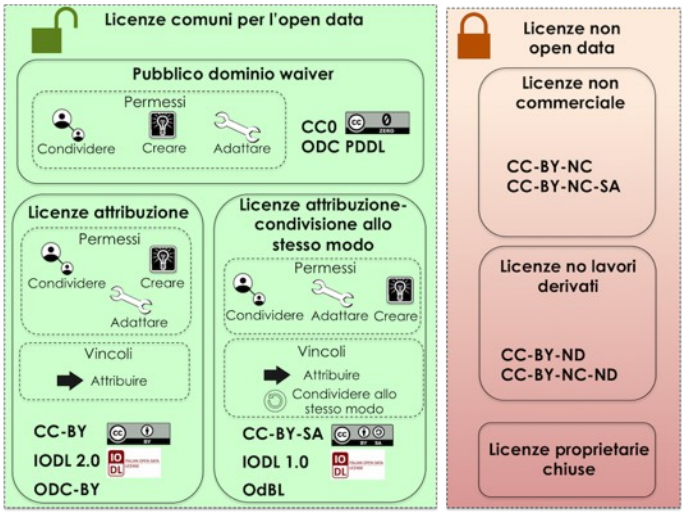


Figura 7: Licenze aperte e non aperte per i dataset

La Figura 7 mostra le licenze più usate per l'Open Data. Esse appartengono a tre categorie principali:

1) *il pubblico dominio o "waiver"* [12] dove il dichiarante "apertamente, pienamente, permanentemente, irrevocabilmente e incondizionatamente rinuncia, abbandona e cede ogni proprio diritto d'autore e connesso, ogni relativa pretesa, rivendicazione, causa e azione, sia al momento nota o ignota (includendo espressamente le pretese presenti come quelle future) relativa all'opera". Rientrano in questa categoria la CC0 della famiglia delle licenze internazionali Creative Commons 51 e la Open Data Commons – Public Domain Dedication License (ODC-PDDL) 52 per i dataset/database;
2) *le licenze per l'open data con richiesta di attribuzione,* che consentono di condividere, adattare e creare anche per finalità commerciali con il solo vincolo di attribuire la paternità del dataset. Rientrano in questa categoria la licenza CC-BY della famiglia Creative Commons 51, la IODL (Italian Open Data License) nella sua versione 2.0 53 e la Open Data Commons Attribution License (ODC-BY) per dataset/database 54.
3) *le licenze per l'open data con richiesta di attribuzione e condivisione allo stesso modo,* che consentono di condividere, adattare e creare anche per finalità commerciali nel rispetto però di due vincoli: a) attribuire la paternità del dataset; b) distribuire eventuali lavori derivati con la stessa licenza che governa il lavoro originale. Rientrano in questa categoria la licenza CC-BY-SA della famiglia Creative Commons la IODL nella sua versione 1.0 55 e la Open Data Commons Open Database License (ODbL) 56 utilizzata dal progetto OpenStreetMap (OSM).

In relazione a quanto sopra riportato, tenuto conto del contesto normativo di riferimento, delle indicazioni in tema di licenze contenute nella Comunicazione della Commissione 2014/C - 240/01 e dei principi di indisponibilità dei beni del demanio culturale espresso negli artt. 10 e 53 del Codice

[12] Essendo internazionale, è assoggettato ai vincoli imposti dal diritto nazionale, è bene quindi verificare la sua compatibilità con il contesto in cui si utilizza.

dei beni culturali (D.lgs. 22 gennaio 2004, n. 42), **si ritiene opportuno fare riferimento ad una licenza unica aperta, che garantisca libertà di riutilizzo, che sia internazionalmente riconosciuta e che consenta di attribuire la paternità dei dataset (attribuire la fonte). Pertanto, si suggerisce l'adozione generalizzata della licenza CC-BY nella sua versione 4.0, presupponendo altresì l'attribuzione automatica di tale licenza nel caso di applicazione del principio "Open Data by default", espresso nelle disposizioni contenute nell'articolo 52 del CAD.**

Si raccomanda inoltre di gestire l'attribuzione della fonte indicando il nome dell'organizzazione unitamente all'URL della pagina Web dove si trovano i dataset/contenuti da licenziare.

Infine, nell'applicazione della licenza si ricorda che non si può disporre/attribuire diritti più ampi rispetto alla licenza di partenza (e.g., non si può attribuire un pubblico dominio - o waiver - a un dataset ottenuto da una fonte a cui è associata una licenza che richiede attribuzione).

A completamento dell'argomento, si evidenzia l'opportunità di verificare gli aspetti relativi a:

- titolarità dei dati secondo la competenza amministrativa;
- elaborazione di un'opera derivata, con il conseguente onere di citazione della fonte originale del dataset e di specifica attribuzione all'opera derivata;
- finalità per i quali i dati sono stati creati che eventualmente non consentono di renderli automaticamente disponibili in open data;
- responsabilità del titolare rispetto al riutilizzo dei dati da parte di terzi

e, nel caso, *specificare una nota legale*, che integra e accompagna la licenza.

COMPATIBILITÀ TRA LICENZE

Un'indicazione di compatibilità tra le licenze Open Data indicate in Figura 7 è riportata nella seguente tabella[13]:

Licenza opera derivata / Licenza opera originaria	CC0	CC-BY	CC-BY-SA	IODL v. 2.0	IODL v. 1.0	ODbL
CC0						
CC-BY						
CC-BY-SA						
IODL v. 2.0						
IODL v. 1.0						
ODbL						

Tabella 1: Compatibilità tra licenze

La creazione di un'opera derivata e la sua pubblicazione è possibile

La creazione di un'opera derivata potrebbe essere possibile ma vi è incertezza (ad esempio sui diritti licenziati) circa l'effettiva compatibilità o altri problemi (problema di stratificazione delle attribuzioni), oppure sul tipo di prodotto derivato (e.s. per la ODbL le modifiche dei dati sono rilasciabili solo con ODbL mentre i prodotti derivati come le mappe con ogni altra licenza).

La creazione di un'opera derivata sotto la licenza proposta è impossibile

51. Creative Commons, http://creativecommons.org
52. Open Data Commons, ODC-PDDL, http://opendatacommons.org/licenses/pddl/summary/
53. Italian Open Data License (IODL) 2.0, http://www.dati.gov.it/iodl/2.0/
54. Open Data Commons, ODC-BY http://opendatacommons.org/licenses/by/summary/
55. Italian Open Data License (IODL) 1.0, http://www.formez.it/iodl/
56. Open Data Commons, ODbL, http://opendatacommons.org/licenses/odbl/

[13] Lo schema proposto in tabella è tratto principalmente da: Federico Morando, "Interoperabilità giuridica: rendere i dati (pubblici) aperti compatibili con imprese e comunità online", JLIS.it Italian Journal of Library and Information Science, Gennaio 2013, http://leo.cineca.it/index.php/jlis/article/download/5461/7928 e modificato secondo gli aggiornamenti delle licenze considerate.

ASPETTI DI COSTO DEL DATO

AZIONE 13: DEFINISCI GLI ASPETTI DI COSTO PER I DATI…

Premesse le azioni di condivisione dei dati tra pubbliche amministrazioni per finalità istituzionali (artt. 50 e 58 del CAD), che avvengono esclusivamente a titolo gratuito, nel caso dell'Open Data si suggerisce azioni volte a renderli disponibili esclusivamente a titolo gratuito. Tuttavia, è prevista la possibilità di richiedere per il riutilizzo dei dati un corrispettivo specifico, limitato ai costi sostenuti effettivamente per la riproduzione, messa a disposizione e divulgazione dei dati. In tali casi, come previsto dall'art. 7 del D.Lgs 24 gennaio 2006, n. 36, AgID determina, su proposta motivata del titolare del dato, le tariffe standard da applicare, pubblicandole sul proprio sito istituzionale.
sulla base del "Metodo dei costi marginali" esplicitato nella Comunicazione della Commissione 2014/C - 240/01 contenente, tra gli altri, gli orientamenti sulla tariffazione 4.

In linea con quanto previsto dalla direttiva comunitaria, il citato articolo 7 del D. Lgs. 36/2006 prevede inoltre casi specifici per i quali è possibile determinare *tariffe superiori ai costi marginali* in deroga al principio generale di rendere disponibili i dati gratuitamente o a costi marginali, ovvero:

1. alle biblioteche, comprese quelle universitarie, di musei e archivi;
2. alle amministrazioni e agli organismi di diritto pubblico che devono generare utili per coprire una parte sostanziale dei costi inerenti allo svolgimento dei propri compiti di servizio pubblico;
3. ai casi eccezionali relativi a documenti per i quali le pubbliche amministrazioni e gli organismi di diritto pubblico sono tenuti a generare utili sufficienti per coprire una parte sostanziale dei costi di raccolta, produzione, riproduzione e diffusione.

In tutti i tre casi, i Ministeri competenti, di concerto con il Ministero dell'economia e delle finanze, sentita AgID, determinano, con appositi decreti, i criteri generali per le tariffe e le relative modalità di versamento, mantenendo aggiornate le stesse ogni due anni. Nel primo caso, l'importo delle tariffe comprende i costi di raccolta, produzione, riproduzione, diffusione, conservazione e gestione dei diritti, maggiorati, nel caso di riutilizzo per *fini commerciali*, di un congruo utile da determinarsi in relazione alle spese per investimenti sostenute nel triennio precedente. Negli altri due casi l'importo delle tariffe comprende i costi di raccolta, produzione, riproduzione, diffusione, maggiorati di un congruo utile, da determinarsi con appositi decreti, nei casi di riutilizzo per *fini commerciali* e in relazione alle spese per investimenti sostenute nel triennio precedente. Nei tre casi di cui sopra, in presenza di riutilizzo dei dati per scopi *non commerciali* è prevista una tariffa differenziata da determinarsi con le modalità suddette secondo il criterio della copertura dei soli costi effettivi sostenuti dalle amministrazioni.
Alla data di pubblicazione delle presenti linee guida, non si riscontrano ancora casi specifici di applicazione dei suddetti principi di tariffazione.

Nei casi eccezionali di applicazione di tariffe superiori ai costi marginali, va tenuto conto delle indicazioni contenute nella Comunicazione della Commissione 2014/C - 240/01, "Metodo del recupero dei costi". Inoltre, è possibile avvalersi di metodi di analisi dei costi (e.g., Activity Based Costing, che assegna costi ai prodotti, servizi, progetti e compiti sulla base sia delle attività svolte per gli stessi sia delle risorse consumate per tali attività) **che siano oggettivi, trasparenti** e **verificabili**. A seguito di tale analisi, l'amministrazione può considerare un modello di business per la determinazione delle tariffe. Un elenco non esaustivo di possibili modelli di business è riportato nelle linee guida per la valorizzazione del patrimonio informativo pubblico (anno 2014) 57. Questi modelli sono stati presentati nel progetto europeo Share-PSI 2.0, nell'ambito del workshop "A Self Sustaining Business Model for Open Data" 58 e possono ancora essere considerati un riferimento per gli scopi del presente aggiornamento delle linee guida.

57. Agenzia Per l'Italia Digitale, "Linee guida per la valorizzazione del patrimonio informativo pubblico (anno 2014), http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.7finale.pdf (Maggio 2014)
58. https://www.w3.org/2013/share-psi/workshop/krems/report

PUBBLICAZIONE E DATI.GOV.IT

PUBBLICAZIONE DEI DATI

AZIONE 14: PUBBLICA I DATI MA SOLO DOPO AVER COMPLETATO LE AZIONI PRECEDENTI…

Prima di pubblicare i dati, assicurati di aver completato queste azioni precedenti e quindi:

- ✓ **AZIONI 1 e 2:** di aver chiari i principi delle normative in materia di dati pubblici e loro riutilizzo
- ✓ **AZIONI 3 e 4:** di aver compreso e selezionato il livello più appropriato del modello per i dati e i metadati , tenendo conto che il requisito minimo per i dati è il livello 3
- ✓ **AZIONE 6 – censimento:** di aver identificato nel censimento dei dati la domanda e l'impatto sociale ed economico che possono e riescono a generare
- ✓ **AZIONE 6 – analisi giuridica delle fonti:** di aver verificato eventuali limitazioni giuridiche (proprietà dei dati, privacy, ecc.)
- ✓ **AZIONE 5:** di aver predisposto i metadati secondo il profilo DCAT-AP_IT
- ✓ **AZIONE 9:** di aver pianificato le attività in modo da mantenere i dati costantemente aggiornati e da garantire altri aspetti di qualità (i.e.., completezza, accuratezza, coerenza)
- ✓ **AZIONE 10:** di aver descritto i dati di riferimento e "core" secondo i modelli indicati nell'architettura di riferimento per l'informazione del settore pubblico
- ✓ **AZIONE 11:** di aver predisposto i dati con almeno un formato aperto machine-readable
- ✓ **AZIONE 12:** di aver assegnato una licenza aperta, possibilmente quella raccomandata dalle presenti linee guida (CC-BY 4.0)

Durante la fase di pubblicazione è necessario garantire agli utenti la possibilità di ottenere i dati in bulk, ovvero fornirli in blocco in un file o insieme di file, senza richiedere credenziali di accesso (a meno di farlo per mere iniziative conoscitive dell'utenza che dovranno essere comunque esplicitate, dando all'utente la possibilità di rifiutare e/o rimanere anonimo), e di interrogare il dato mediante la messa a disposizione di API (Application Programming Interface) che possono essere usate anche per acquisire piccole porzioni dei dati.
Nel caso di pubblicazione di LOD è necessario garantire che gli URI dei dati siano persistenti e deferenziabili e che uno SPARQL endpoint sia presente per abilitare funzioni di interrogazione.
Infine, assicurati di documentare i dati pubblicati in dati.gov.it (AZIONE 15).

Elementi architetturali

I principali livelli architetturali che compongono una soluzione per la pubblicazione e interrogazione di dati aperti possono essere istanziati in diverso modo a seconda delle capacità economiche e tecniche delle amministrazioni, nonché della qualità del servizio che si vuole offrire agli utenti. Si distinguono due livelli: *livello di front-end* e *livello infrastrutturale*.
Il *livello di front-end* consiste di una parte di presentazione che può essere sia un sito Web, sia una sezione in un sito esistente. In questa parte rientrano tutti quegli strumenti che consentono di (i) dare massima visibilità ai dataset disponibili e (ii) di interagire in maniera "user-friendly" con gli utenti stessi, per esempio per capire quali dati sono di loro interesse, quali nuovi dati sono richiesti, quali suggerimenti vogliono dare per migliorare anche la qualità dei dati.
Il livello di presentazione si completa con l'interfaccia di accesso via Web per interrogazioni puntuali sui dati e metadati. Questa ha come obiettivo quello di aumentare l'interazione machine-to-machine attraverso il dispiegamento di una piattaforma di esposizione dati basata su API di servizio (o Open Data Service). Nel caso di dati dei livelli 4 e 5 del modello per i dati, l'interfaccia di accesso via Web è rappresentata dallo SPARQL endpoint.

In generale, si raccomanda di:

(i) **assegnare ai dataset nomi autoesplicativi** per comprenderne il principale contenuto;
(ii) **fornire**, ove possibile, **descrizioni testuali dei dataset**;
(iii) **mettere in evidenza la licenza in uso** in forma "human e machine-readable";
(iv) fornire, ove possibile, strumenti di visualizzazione e navigazione, anche georiferita, dei dati, che possano facilitare la lettura degli stessi;
(v) **fornire**, ove possibile, **statistiche di uso, accesso** e produzione;
(vi) fornire notifiche di cambiamenti nel sito web, di aggiornamenti ai dataset (e.g., RSS feed);
(vii) **fornire strumenti per rendere le interrogazioni più agevoli**, anche per utenti non del tutto esperti. Nel caso dei dati dei livelli 4 e 5 **non si può pubblicare solo dataset RDF ma è bene mettere in evidenza la presenza dello SPARQL endpoint (i.e., servizio Web che accetta interrogazioni SPARQL, le risolve e restituisce i risultati in output), pubblicando il link di accesso, fornendo altresì un ampio insieme di "query" di esempio che con pochi click possono essere eseguite** producendo risultati disponibili in diversi formati di più facile e comune utilizzo (e.g., CSV, JSON, XML, HTML)

Nei casi di amministrazioni di minori dimensioni o amministrazioni che non siano nelle condizioni di poter fornire un servizio con le caratteristiche sopra elencate, si consiglia di **implementare azioni di sussidiarietà verticale** (ad esempio, i comuni di medio-piccole dimensioni possono riferirsi alla Regione di appartenenza) o di unirsi in iniziative comuni.

Le iniziative OpenCoesione[14] e Linked Open Data della Camera dei Deputati[15] offrono buoni esempi e buone pratiche per applicare le suddette raccomandazioni.

Il *livello infrastrutturale* è rappresentato dall'infrastruttura che ospita i dati e i metadati. Nel caso di dati aperti, tenuto conto della loro natura intrinseca, ovvero dati tipicamente non riferibili a singole persone e per i quali solitamente non si richiede il soddisfacimento di specifici requisiti di protezione dei dati personali, tecnologie basate sul paradigma del cloud computing pubblico (o di comunità come il cloud del Sistema Pubblico di Connettività) possono essere facilmente impiegabili al fine di ospitare le infrastrutture per la pubblicazione di dati aperti.

Soluzioni Open Data per i portali Web

Si raccomanda di non creare tanti portali diversi per singole iniziative ma, ove possibile, di raccordarle per facilitare il reperimento e il riutilizzo dei dati da parte degli utenti finali.

Di seguito si riportano alcune possibili soluzioni per la creazione di piattaforme di pubblicazione dei dati.

SOLUZIONE NATIVA. Viene creato un portale ad-hoc o creata un'apposita sezione di un portale esistente. In questo caso, la creazione non differisce dalla creazione di un sito Web classico[16].

ESTENSIONE SOLUZIONE CMS ESISTENTE. Molto spesso l'amministrazione gestisce già un sito Web, realizzato mediante l'uso di un CMS, che vuole estendere con una sezione dedicata agli Open Data. La criticità in questo caso è data dall'aggiunta di una componente semantica all'interno della configurazione del CMS stesso. In questo ambito, merita una menzione il progetto Apache Stanbol[17] che mira a dare supporto in questo senso.

UTILIZZO DI PIATTAFORME ESTERNE. Viene utilizzata una piattaforma che integra già funzionalità per la catalogazione, visualizzazione, ricerca e interrogazione dei dati. In alcuni casi queste piattaforme sono disponibili in modalità cloud computing. Gli strumenti di questo tipo più utilizzati sono CKAN, DKAN, Socrata. Essi si prestano anche per essere facilmente integrati con portali già esistenti.

14 http://www.opencoesione.gov.it/
15 http://dati.camera.it/sparql
16 Si applicano anche le raccomandazioni delle linee guida di design per i siti web delle PA, http://designer.italia.it/
17 https://stanbol.apache.org/

Requisiti per la pubblicazione di dati di livello 4 e 5

I Linked Data utilizzano URI per risolvere il problema dell'identità; gli URI devono essere *persistenti* e *dereferenziabili*.

Una politica per garantire *URI persistenti* e fornire aspetti di naming è proposta dalla commissione europea con il documento sulle "10 regole per URI persistenti" 59. Facendo riferimento a tale documento, per la creazione di URI persistenti sono da evitare quelli che contengano:

- nome del progetto/ufficio/unità amministrativa che detiene la risorsa per evitare problemi derivanti dalla fine del progetto stesso o fusioni o chiusure di uffici nell'organizzazione;
- numeri di versione;
- identificatori esistenti che in passato sono stati utilizzati per identificare risorse differenti;
- riferimenti generati in modo automatico e incrementale a meno che non vi sia la garanzia che il processo non venga mai più ripetuto o, se ripetuto, generi sicuramente gli stessi identificatori per gli stessi dati di input;
- stringhe rappresentanti "query" a database;
- estensione del file.

Sono, invece, da ritenersi buone pratiche le seguenti:

- strutturare l'URI come segue:

  http://{dominio}/{tipo}/{concetto}/{riferimento}

  dove gli elementi che compongono la URI sono:
  - *Dominio*: il dominio Web su cui reperire la risorsa
  - *Tipo*: l'elemento che specifica il tipo di risorsa. Dovrebbe poter assumere un numero limitato di valori come "doc" se la risorsa identificata è un documento descrittivo, "set" se la risorsa è un dataset, "id" o "item" se la risorsa è un oggetto del mondo reale
  - *Concetto*: il tipo di un oggetto del mondo reale
  - *Riferimento*: lo specifico elemento, termine o concetto che rappresenta la risorsa
- costruire URI per più formati al fine di identificare al meglio la risorsa
- collegare tra loro le rappresentazioni multiple della stessa risorsa
- implementare il codice di risposta 303 per gli oggetti del mondo reale (si veda sotto "content negotiation" e "dereferenziazione" degli URI)
- utilizzare servizi dedicati

Si raccomanda di considerare anche la possibilità di mantenere URI persistenti mediante w3id.org 60, ovvero un servizio per applicazioni Web che fornisce meccanismi sicuri e permanenti di re-direzione, garantendo l'uso di URI sempre riferibili a siti web funzionanti. Il servizio è mantenuto dal W3C Permanent Identifier Community Group.

Inoltre, facendo uso di URI HTTP per identificare le risorse RDF, si potrebbe incorrere in URI ambigue, ovvero URI che rappresentano sia entità del Web Semantico, sia risorse Web (ad esempio, pagine Web, file, ecc.). A tal riguardo occorre gestire le richieste HTTP sulla base del loro tipo: queste possono richiedere dati (e.g., l'attributo "Accept" della richiesta valorizzato con "application/rdf+xml") oppure risorse Web (e.g., l'attributo "Accept" della richiesta valorizzato con "text/html"). Questo processo è anche detto "*content negotiation*". Esistono strumenti quali Pubby, ELDA, LodLive che integrano nativamente la "content-negotiation".

Infine, esistono situazioni, tipicamente con accesso da Web browser, in cui è richiesta una risorsa (non ambigua) del Web Semantico come se questa fosse una pagina HTML. In questi casi si può rispondere all'utente con una pagina Web informativa relativa alle informazioni associate all'entità identificata con quell'URI. Questa operazione è detta *dereferenziazione* degli URI.

Il W3C ha pubblicato un rapporto tecnico dettagliato 61 sulla dereferenziazione delle URI e sulla "content negotiation" al quale si consiglia di far riferimento.

59. ISA and W3C, "Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC", https://joinup.ec.europa.eu/sites/default/files/c0/7d/10/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf
60. W3C Permanent Identifier Community Group, "Permanent Identifiers for the Web", https://w3id.org/, 2016.
61. W3C, "Cool URIs for the Semantic Web, https://www.w3.org/TR/cooluris/, 3 Dicembre 2008.

DATI.GOV.IT

L'architettura complessiva del portale nazionale e il suo interfacciamento con i cataloghi dei dati delle pubbliche amministrazioni e con il portale europeo sono illustrati in Figura 8.

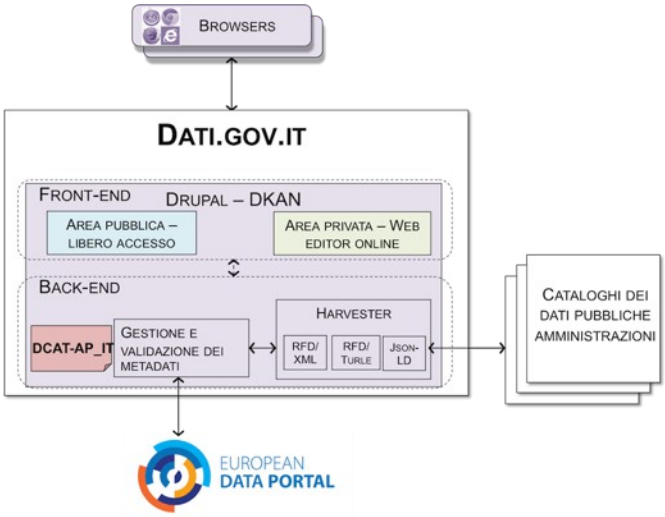


Figura 8: Architettura del portale nazionale dei dati

Come mostrato in figura, esso si basa su DKAN, uno strumento open data interamente integrato e gestito nell'ambito di Drupal, che offre funzionalità di catalogazione, pubblicazione e visualizzazione dei dati. Il front-end del portale ha una parte di libero accesso, dove chiunque può ricercare e visionare e riutilizzare i metadati inclusi nel catalogo, e una parte di area privata riservata alle pubbliche amministrazioni per alimentare il catalogo. Il back-end è interamente gestito sulla base del nuovo profilo di metadatazione nazionale DCAT-AP_IT, ha il compito di predisporre i metadati per l'harvesting da parte del portale europeo e prevede una funzionalità di harvesting periodica e automatizzata verso i cataloghi delle altre amministrazioni. A oggi quest'ultima funzione supporta tre formati principali: JSON-LD, utile per quei cataloghi basati principalmente sul formato JSON (e.g., Socrata), RDF/XML, per utenti più familiari con il formato XML, e RDF/Turtle.
La tabella seguente fornisce indicazioni di massima sulle modalità di alimentazione attualmente previste e adottabili a seconda del proprio contesto.

AZIONE 15: ASSICURATI CHE I METADATI RELATIVI AI TUOI DATASET SIANO PRESENTI NEL PORTALE NAZIONALE DEI DATI…

Ai sensi dell'articolo 1 comma 8 del D.Lgs. 18 Maggio 2015, n.102, il portale nazionale dei dati (dati.gov.it) è l'unico riferimento per la documentazione e la ricerca di tutti i dati della pubblica amministrazione. Esso, inoltre, è l'unico ad abilitare il colloquio con l'analogo portale europeo(http://www.europeandataportal.eu/).
Il portale nazionale dei dati include esclusivamente i metadati, conformi al profilo DCAT-AP_IT, che descrivono sia i database delle amministrazioni, sia i relativi dati aperti.
Le amministrazioni sono tenute pertanto a inserire e a mantenere aggiornati, attraverso le modalità di alimentazione previste dal catalogo, tali metadati. I dati primari, il cui riferimento è pubblicato sul portale nazionale, rimangono presso il titolare del dato che conserva la responsabilità della loro divulgazione a livello nazionale.
Come già precedentemente riportato, i dati geografici devono essere documentati esclusivamente presso il Repertorio Nazionale dei Dati Territoriali (RNDT) che, in maniera automatizzata, si occupa dell'allineamento con il portale nazionale dei dati.

Numero di dataset e relativo aggiornamento	Modalità di alimentazione
Pochi dataset (nell'ordine delle decine) – frequenza di aggiornamento molto ampia (e.g., annuale, semestrale) oppure anche più regolare (e.g., trimestrale).	Uso del web editor che guida nella definizione dei metadati, già predisposti per essere conformi a DCAT-AP_IT
Molti dataset (nell'ordine delle centinaia e oltre) – frequenza di aggiornamento sia ampia, sia regolare e molto breve (e.g., giornaliera).	Alimentazione e aggiornamento automatico e periodico del portale attraverso funzionalità di harvesting verso i cataloghi dei dati delle

	pubbliche amministrazioni.
Per tutti i dataset relativi a dati geografici	Strumenti di alimentazione del catalogo RNDT. Il portale nazionale dei dati sarà alimentato in maniera trasparente per le amministrazioni e automatizzata attraverso l'RNDT, grazie alla futura implementazione da parte di quest'ultimo dell'estensione GeoDCAT-AP per i dati geografici.

Ulteriori elementi di federazione

I meccanismi di alimentazione del portale nazionale abilitano, di fatto, una federazione tra portali di pubbliche amministrazioni. Si possono individuare anche ulteriori modalità di federazione e condivisione.

Per esempio, un'amministrazione può mettere a disposizione di altre la propria soluzione Open Data (e.g., un'amministrazione regionale dotata di una piattaforma Open Data la può mettere a disposizione dei comuni della regione, e si raccomanda di farlo).

Nell'ambito invece del paradigma Linked Data, si segnala che il W3C ha definito uno standard per federare SPARQL endpoint 62. Lo standard prevede una sintassi aggiuntiva per SPARQL in grado di considerare, in una stessa "query", dati provenienti da SPARQL endpoint differenti. Inoltre, lo standard prevede funzioni per cui molteplici SPARQL endpoint gestiscono, in maniera del tutto trasparente per l'utente, l'invio della "query" a più endpoint o la scomposizione della "query" e la ricomposizione dei frammenti del risultato finale. In generale, al meglio delle nostre conoscenze, la federazione di SPARQL endpoint rimane ancora confinata a soluzioni di ricerca.

62. W3C Recommendation, "SPARQL 1.1 Federated Query", https://www.w3.org/TR/sparql11-federated-query/, 21 Marzo 2013