



Agenzia per l'Italia Digitale
Presidenza del Consiglio dei Ministri

LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO (SECONDO SEMESTRE 2013)

VERSIONE 1.0 DEL 31/07/2013



INDICE

EXECUTIVE SUMMARY	6
1. PREFERENZA.....	8
1.1. Componenti del gruppo di lavoro.....	8
1.2. Modifiche documento	9
1.3. Acronimi.....	10
1.4. Glossario.....	11
1.5. Licenza.....	14
2. SCOPO, DESTINATARI E STRUTTURA DEL DOCUMENTO	15
3. INTRODUZIONE	17
3.1. Quadro normativo.....	19
3.1.1. <i>Regolamenti delle pubbliche amministrazioni e direttive locali.....</i>	<i>21</i>
3.2. Aspetti di sussidiarietà	27
3.3. Definizioni: dati delle pubbliche amministrazioni e documenti	27
4. MODELLI DI RIFERIMENTO PER I DATI E I METADATI.....	31
4.1. Modello per i dati	31
4.1.1. <i>Linked Open Data</i>	<i>33</i>
4.2. Modello per i metadati	34
4.2.1. <i>Metadati obbligatori e obbligatori condizionatamente.....</i>	<i>35</i>
5. MODELLO OPERATIVO PER LA PRODUZIONE E GESTIONE DEI DATI DI TIPO APERTO.....	38
5.1. Ruoli e responsabilità.....	38
5.2. Modello operativo	40
5.2.1. <i>Linea 1: Dati Nativi</i>	<i>41</i>
5.2.2. <i>Linea 2: Dati Mashup</i>	<i>46</i>
5.2.3. <i>Linea 3: Linked Open Data</i>	<i>46</i>
5.2.4. <i>Linea 4: Engagement.....</i>	<i>47</i>
6. STANDARD TECNICI E ONTOLOGIE.....	49
6.1. Standard di base.....	49
6.2. Formati di tipo aperto.....	51
6.2.1. <i>Formati per i dati di tipo aperto.....</i>	<i>51</i>
6.2.2. <i>Formati per i documenti.....</i>	<i>52</i>
6.3. Ontologie e vocabolari	53
6.3.1. <i>Ontologie e vocabolari per dati trasversali.....</i>	<i>53</i>
6.3.2. <i>Ontologie e vocabolari per dati verticali</i>	<i>55</i>
6.3.2.1. <i>Dati Territoriali e Ambientali</i>	<i>56</i>
6.3.2.2. <i>Dati Statistici</i>	<i>56</i>
6.3.2.3. <i>Dati culturali.....</i>	<i>56</i>
6.3.3. <i>Ontologie e vocabolari per i metadati.....</i>	<i>57</i>



6.4.	Alcuni dataset di riferimento.....	58
6.4.1.	<i>Dati trasversali</i>	58
6.4.2.	<i>Dati territoriali e ambientali</i>	59
6.4.3.	<i>Dati sanitari</i>	59
7.	LICENZE ASSOCIATE AL DATO.....	60
8.	ASPETTI DI COSTO DEL DATO	63
8.1.	Modalità di rilascio dei dati	63
8.2.	Modelli di business.....	64
9.	ALCUNE INDICAZIONI OPERATIVE PER I PORTALI OPEN DATA	66
9.1.	Requisiti per la pubblicazione di dati di livello 4 e 5.....	66
9.2.	Soluzioni Open Data per i portali Web	67
9.3.	Requisiti per i portali Open Data	68
9.4.	Recupero e accesso ai dati	68
10.	INDICAZIONI OPERATIVE PER I DATI DI TIPO APERTO NEI CAPITOLATI DI GARA.....	70
11.	BIBLIOGRAFIA	71



INDICE DELLE FIGURE

Figura 1: Processo di valorizzazione del patrimonio informativo pubblico	18
Figura 2: Tipi di dato delle pubbliche amministrazioni.....	29
Figura 3: Modello a cinque livelli per la produzione e il rilascio di dati di tipo aperto	31
Figura 4: Modello per la definizione dei metadati.....	34
Figura 5: Modello operativo per la produzione e gestione dei dati aperti nella PA.....	41
Figura 6: Strumenti legali per l'Open Data.....	61



INDICE DELLE TABELLE

Tabella 1: Comparazione di sintesi dei piani di azione delle regioni esaminate	26
Tabella 2: Metadati obbligatori	36
Tabella 3: Metadati obbligatori al verificarsi di determinate condizioni.....	37
Tabella 4: Comparazione tra licenze	60



EXECUTIVE SUMMARY

Le pubbliche amministrazioni raccolgono, organizzano e gestiscono una vasta quantità di dati pubblici. Tuttavia, negli anni, l'autonomia delle singole amministrazioni e le modalità di gestione dei dati hanno contribuito a creare isole di informazioni, con scarsa visione sistemica, nonostante alcuni significativi interventi normativi tendessero a favorire un maggior livello di integrazione e condivisione dei dati tra le amministrazioni.

L'attuale situazione socio-economica e l'opportunità di sviluppo e di competitività che un utilizzo sistemico delle tecnologie digitali possono portare, impongono di non considerare più tali tecnologie beni strumentali ma leve di innovazione e di politica economica. In tal senso la Digital Agenda for Europe [1] e l'Agenda Digitale Italiana [2] considerano l'uso pervasivo delle tecnologie ICT come opportunità per fornire sviluppo, occupazione e benessere sociale.

In tale ottica, i dati pubblici delle amministrazioni, che solo fino a poco tempo fa avevano un ruolo funzionale al perseguimento dei compiti istituzionali delle amministrazioni, assumono una differente valenza in termini di stimolo dell'economia digitale, sviluppo dell'innovazione e trasparenza amministrativa. Essi rappresentano un patrimonio cui corrisponde un valore economico che deve essere necessariamente considerato come volano per lo sviluppo del mercato, per la nascita di nuove figure professionali e per il supporto alle decisioni di ogni organo politico, anche locale, così come accade in altri Paesi.

In tale scenario, risulta cruciale avviare opportune politiche di valorizzazione dei dati pubblici attraverso la definizione di una strategia nazionale e di linee guida che implementino tale strategia e indirizzino tutte le amministrazioni verso un processo di produzione e rilascio dei dati **standardizzato e interoperabile su scala nazionale**.

L'obiettivo principale del presente documento, redatto e revisionato con cadenza annuale da un gruppo di lavoro composto da amministrazioni centrali e locali, è quello di fornire indicazioni operative per l'implementazione della strategia nazionale di valorizzazione del patrimonio informativo pubblico, descritta nell'agenda nazionale. **Si fa presente che al momento della pubblicazione delle presenti linee guida, l'agenda nazionale è in corso di approvazione da parte della Presidenza del Consiglio dei Ministri. Una sua versione bozza è resa disponibile sul sito dell'Agenzia per l'Italia Digitale in [94].**

Il lavoro del gruppo è partito da un'analisi del quadro normativo nazionale sui dati di tipo aperto e dell'attuale stato delle disposizioni a livello locale (regionale e comunale) in tale ambito, con riferimento ad aspetti di sussidiarietà da applicare per piccole realtà locali.

In seguito, anche sulla base di tale analisi, il gruppo di lavoro ha predisposto un modello comune di riferimento per i dati di tipo aperto. Il modello segue la classificazione a stelle per i dati, ampiamente riconosciuta e nella quale si evidenziano le relative tipologie di servizi e le modalità di accesso ai dati che possono essere abilitate. Tale modello è stato altresì accompagnato dalla definizione di un modello per i metadati e dall'individuazione di un insieme di metadati obbligatori da specificare.

Per l'implementazione effettiva dei due modelli introdotti, il gruppo di lavoro ha riconosciuto all'unanimità l'importanza di affiancare al contributo tecnico anche un indirizzo in termini organizzativi, proponendo un modello operativo per il processo di valorizzazione del patrimonio informativo



pubblico nelle singole amministrazioni che va dalla produzione “nativa” di dati pubblici in formato di tipo aperto, al rilascio di tali dati, fino al coinvolgimento di attori esterni al fine di incoraggiare il massimo riuso dei dati stessi. Nel modello operativo, il gruppo di lavoro ha individuato i ruoli e le responsabilità degli attori che partecipano al processo, e le diverse interazioni tra essi necessarie per portare a compimento le fasi del modello.

Le linee guida si completano specificando (i) gli standard tecnici di riferimento, le ontologie e i vocabolari per la modellazione dei dati pubblici da rilasciare, individuando sia quelli per specifiche tipologie di dati (e.g., beni culturali), sia quelli per dati trasversali alle diverse pubbliche amministrazioni (e.g., dati sulla struttura organizzativa delle amministrazioni); (ii) le licenze d’uso da associare ai dati pubblicati, guidando le amministrazioni nella scelta più appropriata secondo le peculiarità dei dati rilasciati; (iii) gli eventuali aspetti di costo per talune tipologie di dati; e (iv) alcune raccomandazioni per la pubblicazione dei dati nei portali dei dati di tipo aperto.

Infine, in aggiunta all’analisi tecnico-organizzativa per la produzione e per il rilascio di dati di tipo aperto, il gruppo ha evidenziato l’importanza di fornire linee guida specifiche separate dalle presenti sulle indicazioni operative in merito a clausole per il trattamento di tali dati nei capitolati di gare d’appalto.



1. PREFAZIONE

1.1. Componenti del gruppo di lavoro

Questo documento è stato redatto e curato dal “Gruppo di lavoro dell’Agenzia per l’Italia Digitale per le linee guida sulla valorizzazione del patrimonio informativo pubblico” composto dalla Dott.ssa G. Lodi, Ing. A. Maccioni, dal Dott. Ciasullo e coordinato dal Dott. Tortorelli.

Al gruppo hanno partecipato numerose pubbliche amministrazioni centrali tra cui INPS, ISTAT, Ministero dei Beni e delle Attività Culturali, Ministero della Salute, Ministero degli Affari Esteri, Presidenza del Consiglio dei Ministri – Protezione Civile, Ministero della Difesa, ISPRA, e un folto numero di pubbliche amministrazioni locali con diverse Regioni (e.g., Piemonte, Liguria, Emilia Romagna, Lombardia, Umbria, Veneto, Marche, Puglia) e Comuni (e.g., Firenze, Bologna, Ravenna, Milano e Palermo) e le associazioni in rappresentanza di queste CISIS e ANCI nonché l’UNCCEM.

Si ringrazia Maurizio Napolitano della Fondazione Bruno Kessler e rappresentante dell’Open Knowledge Foundation in Italia per i commenti puntuali atti a migliorare parti delle linee guida.



1.2. Modifiche documento

Descrizione Modifica	Edizione	Data
Prima bozza indice	v 0.1	20-03-2013
Seconda bozza indice, inserimento di alcune figure e primi assignment per i contributi	v.02	19-04-2013
Inserimento primi contributi alle sezioni e revisione dell'indice	v. 03	20-05-2013
Inserimento sezioni mancanti	v. 04	25-06-2013
Inserimento alcuni contributi pervenuti e stesura bozza intero documento	v. 05	01-07-2013
Recepimento commenti emersi da alcuni partecipanti e revisione delle sezioni 3, 4, 6, 8, 9 e 10	v. 06	11-07-2013
Revisione della sezione 7 e aggiunta bibliografia	v. 07	15-07-2013
Revisione della sezione 5, degli acronimi e del glossario aggiunta di ulteriori riferimenti bibliografici	v. 08	16-07-2013
Revisione generale documento	v. 09	26-07-2013
Versione finale documento	v.1.0	31-07-2013



1.3. Acronimi

API – Application Programming Interface
CAD – Codice dell’Amministrazione Digitale
CC – Creative Commons
CMS – Content Management System
DC – Dublin Core
DCAT – Data Catalog Vocabulary
DCMI – Dublin Core Metadata Initiative
DGR – Deliberazione Giunta Regionale
DL – Decreto Legge
EE.LL – Enti Locali
GIS – Geographic Information 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
LP – Legge Provinciale
LR – Legge Regionale
OD – Open Data
OKF – Open Knowledge Foundation
PA – Pubblica Amministrazione
PSI – Public Sector Information
RNDT – Repertorio Nazionale Dati Territoriali
SDMX – Statistical Data and Metadata eXchange
SPC – Sistema Pubblico di Connettività e Cooperazione
URI – Uniform Resource Identifier
WSDL – Web Service Description Language
XML – eXtensible Markup Language



Agenzia per l'Italia Digitale

1.4. Glossario

APPS: termine utilizzato per indicare le applicazioni software sviluppate per operare su terminali mobili.

CLOUD COMPUTING: Dalla definizione fornita dal NIST (l'istituto nazionale statunitense per gli standard e le tecnologie) il cloud computing è un modello che abilita in rete l'accesso pratico e su richiesta (on demand) a un pool condiviso di risorse computazionali configurabili (e.g., reti, server, storage, applicazioni e servizi) che possono essere ottenute ed erogate rapidamente con il minimo sforzo di gestione e con un'interazione limitata con il fornitore.

DATASET: una collezione di dati, generalmente riguardanti una stessa organizzazione, che vengono erogati e gestiti congiuntamente.

DATI: rappresentazione fisica di informazioni atta alla comunicazione, interpretazione ed elaborazione da parte di essere umani o mezzi automatici.

E-GOVERNMENT: gestione digitalizzata dei processi e delle comunicazioni che riguardano la funziona pubblica e la funzione amministrativa.

HTTP (HYPER-TEXT TRANSFER PROTOCOL): protocollo standard per la trasmissione di informazione su Web. Tipicamente e nativamente usato per il trasferimento di iper-testi e iper-media.

INFERENZA: il processo con il quale un'informazione viene dedotta da altre informazioni.

INTEROPERABILITÀ: in ambito informatico, la capacità di sistemi differenti e autonomi di cooperare e di scambiare informazioni in maniera automatica, sulla base di regole comunemente condivise.

INTEROPERABILITÀ SEMANTICA: la capacità di elaborare informazioni da fonti esterne o secondarie senza perdere il reale significato delle informazioni stesse nel processo di elaborazione.

LICENZA: in ambito informatico, lo strumento con il quale si governano legalmente le condizioni d'uso e di distribuzione del software e dei dati.



MASHUP: un processo informatico in cui si integrano contenuti, dati e informazioni provenienti da fonti differenti.

ONTOLOGIA: in ambito informatico, una rappresentazione formale e condivisa dei concetti e delle mutue relazioni che caratterizzano un certo dominio di conoscenza.

OPEN SOURCE: una modalità con cui il software viene fornito. Si realizza attraverso la concessione a terzi del diritto di accedere liberamente al codice sorgente, talvolta consentendo la possibilità di effettuare modifiche e prender parte alle decisioni progettuali sul software in questione.

RIUSO DI DATI: qualsiasi uso dei dati diverso da quello per il quale sono stati prodotti o raccolti originariamente.

PSI (Public Sector Information): nella direttiva Europea 2013/37/EU è definita come l'informazione della Pubblica Amministrazione. L'informazione pubblica si caratterizza per essere di tipo statico o di tipo dinamico. L'informazione statica è rappresentata dal contenuto informativo in possesso della Pubblica Amministrazione (ad esempio gli archivi dei beni culturali); l'informazione dinamica è invece prodotta dalle istituzioni pubbliche nello svolgimento dei propri compiti istituzionali (ad esempio i dati di bilancio di un ente).

SEGRETO DI STATO: un vincolo giuridico che determina l'esclusione dalla divulgazione *“di atti, documenti, notizie, attività e ogni altra cosa la cui diffusione sia idonea a recare danno all'integrità della Repubblica, anche in relazione ad accordi internazionali, alla difesa delle istituzioni poste dalla Costituzione a suo fondamento, all'indipendenza dello Stato rispetto ad altri Stati e alle relazioni con essi, alla preparazione e alla difesa militare dello Stato”*, ponendo delle sanzioni nei confronti di chi violi l'obbligo di non divulgazione.

SEGRETO STATISTICO: i dati raccolti nell'ambito di rilevazioni statistiche comprese nel programma statistico nazionale da parte degli uffici di statistica non possono essere esternati se non in forma aggregata, in modo che non se ne possa trarre alcun riferimento relativamente a persone identificabili e possono essere utilizzati solo per scopi statistici.

SPC: l'insieme di infrastrutture tecnologiche e di regole tecniche, per lo sviluppo, la condivisione, l'integrazione e la diffusione del patrimonio informativo e dei dati della Pubblica Amministrazione, necessarie per assicurare l'interoperabilità di base ed evoluta e la cooperazione applicativa dei sistemi informatici e dei flussi informativi, garantendo la sicurezza, la riservatezza delle informazioni, nonché la salvaguardia e l'autonomia del patrimonio informativo di ciascuna Pubblica Amministrazione. SPC è definito nel D. Lgs 7 marzo 2005 n. 82, negli artt. dal 72 all'87.

TASSONOMIA: una classificazione gerarchica di concetti ed elementi che consente di rappresentare



un dominio di conoscenza.

TESAURO: collezione di termini privi di definizioni. Nella pratica, tesaurus e vocabolario (si veda sotto) sono utilizzati in maniera pressoché equivalente.

URI (Uniform Resource Identifier): stringa di caratteri che identifica univocamente una risorsa (pagina web, documento, immagine, file, ecc).

URL (Uniform Resource Locator): stringa di caratteri che identifica una risorsa su Internet, ne specifica formalmente la collocazione e indica il protocollo da utilizzare per accedervi. È un tipo specifico di URI.

VOCABOLARIO: l'insieme dei possibili valori che le entità di una classe possono assumere all'interno di un dominio di conoscenza o di una ontologia.

W3C (WORLD WIDE WEB CONSORTIUM) : il consorzio internazionale che ha lo scopo di definire gli standard aperti per il Web.

WEB SEMANTICO: insieme di modelli e standard Web in cui le risorse vengono descritte e correlate fra loro in modo formale attraverso l'uso opportuno di metadati. In questo modo si abilitano gli agenti automatici a comprendere il significato dei dati e delle informazioni.



1.5. Licenza

Il presente documento è soggetto alla licenza Creative Commons – Attribuzione – Condividi allo stesso modo 3.0 (CC-BY-SA).



Agenzia per l'Italia Digitale

2. SCOPO, DESTINATARI E STRUTTURA DEL DOCUMENTO

Scopo. Il presente elaborato rappresenta un documento di LINEE GUIDA per “*l’individuazione degli standard tecnici, compresa la determinazione delle ontologie dei servizi e dei dati, le procedure e le modalità di attuazione delle disposizioni del Capo V del Codice dell’Amministrazione Digitale con l’obiettivo di rendere il processo omogeneo a livello nazionale, efficiente ed efficace.*”, secondo quanto stabilito dall’articolo 52 del Codice dell’Amministrazione Digitale (CAD), recentemente modificato dall’art. 9 del DL n. 179/2012.

Per l’anno corrente, tenuto conto dell’entrata in vigore delle disposizioni di cui al suddetto articolo 52, le presenti linee guida si riferiscono al solo secondo semestre 2013. Esse sono incrementalmente aggiornate all’inizio di ogni anno sulla base (i) dei possibili cambiamenti normativi, (ii) dei principali contenuti dell’agenda nazionale per la valorizzazione del patrimonio informativo pubblico, anch’essa da produrre e pubblicare annualmente, (iii) dei cambiamenti in materia di standard e ontologie, (iv) degli approfondimenti che l’Agenzia avvierà per alcune tematiche affrontate nelle linee guida, (v) dei riscontri delle pubbliche amministrazioni, e, infine, (vi) delle politiche e degli obiettivi definiti di anno in anno dall’Agenzia in materia di gestione dei dati pubblici.

Il documento ha l’obiettivo di supportare le pubbliche amministrazioni nel processo di valorizzazione del proprio patrimonio informativo pubblico, proponendo da un lato un modello per i dati pubblici, corredato di un modello per la relativa metadattazione, e dall’altro uno schema operativo per attuare il modello proposto. Lo schema operativo ha il compito di identificare le fasi del processo di valorizzazione dei dati pubblici e i ruoli delle varie figure professionali delle pubbliche amministrazioni coinvolte, indicando per ogni figura le principali azioni che essa deve intraprendere. Il documento, in linea con gli obiettivi indicati nella norma suddetta, approfondisce l’uso di un insieme di standard di base e di ontologie e vocabolari specifici per categorie trasversali e verticali di dati delle pubbliche amministrazioni, al fine di guidare nell’effettiva implementazione dei modelli proposti, i.e., dei dati e dei metadati, e operativo.

In tale scenario, il documento mira infine dare supporto alle pubbliche amministrazioni (i) nella scelta della licenza da associare ai dati, (ii) nell’analisi di eventuali aspetti di costo dei dati di tipo aperto, (iii) nella pubblicazione dei dati nei portali Open Data per una maggiore standardizzazione di tale processo, e (iv) nell’individuazione di possibili clausole da inserire nei capitolati di gara per il trattamento dei dati pubblici.

Destinatari. Il presente documento è destinato a tutte le Amministrazioni, così come definite all’art. 2 comma 2 del CAD, che a esso si uniformano (art. 9 DL n. 179/2012). In virtù della duplice valenza tecnico-organizzativa del documento, le linee guida 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). In particolare, solo una stretta interazione tra figure professionali meno tecniche, ma con conoscenze specifiche di determinate tipologie di dati, e figure più tecniche può portare alla piena attuazione delle linee guida descritte dal presente documento.



Struttura del documento. Le sezioni del documento sono così strutturate. La sezione 3 introduce le linee guida, presenta il quadro normativo, le delibere locali sul tema dei dati di tipo aperto e fornisce una serie di definizioni di base sui dati pubblici. La sezione 4 introduce i modelli di riferimento per i dati di tipo aperto e per i metadati indicando, in quest'ultimo caso, l'insieme di metadati obbligatori e metadati obbligatori condizionatamente che devono essere associati ad ogni dataset pubblicato. La sezione 5 propone un modello operativo per la produzione e gestione dei dati pubblici individuando ruoli, responsabilità e azioni da intraprendere. La sezione 6 individua gli standard di base e alcune ontologie/vocabolari e dataset di riferimento da utilizzare per la produzione e pubblicazione di dati di tipo aperto, sia trasversali (comuni a tutte le amministrazioni), sia verticali (specifici per ambiti tematici); la sezione 7 descrive alcune licenze disponibili per i dati di tipo aperto, guidando nella scelta di quella più opportuna in base agli obiettivi di riuso e accesso ai dati. La sezione 8 analizza alcuni aspetti legati ai costi e alla tariffazione per i dati di tipo aperto. La sezione 9 fornisce indicazioni operative per la pubblicazione dei dati di tipo aperto nei portali Open Data e infine la sezione 10 conclude il documento discutendo della definizione delle clausole dei capitolati di gara relative al trattamento di dati di tipo aperto.

Il documento è altresì corredato di un'appendice tecnica che illustra alcuni esempi di utilizzo di formati per i dati e ontologie per dati considerabili comuni a tutte le pubbliche amministrazioni (e.g., contratti).



3. INTRODUZIONE

L'utilizzo sempre più pervasivo delle tecnologie dell'informazione e della comunicazione e dei dispositivi mobili sta contribuendo al forte sviluppo della cittadinanza digitale che, grazie anche a nuovi canali di comunicazione sociali, diventa sempre più interattiva ed esigente nei confronti delle numerose istituzioni pubbliche. Questo si traduce in una crescente domanda di dati e di servizi sviluppati a partire dagli stessi dati. Tale domanda, se opportunamente e tempestivamente intercettata e, per quanto possibile, soddisfatta può portare a un miglioramento sia del rapporto tra istituzioni e società civile, sia dei processi interni alle stesse amministrazioni. Tuttavia, si osserva che nei processi di produzione di rapporti pubblici persistono ancora lunghe e costose fasi di raccolta dei dati, di normalizzazione e di incapsulamento in documenti informatici, fasi e modalità che rendono talvolta le informazioni più facilmente obsolete e difficili da riutilizzare. Le migliaia di processi che presiedono oggi alla produzione dei dati delle amministrazioni sono stati concepiti e (ove questo è accaduto) ingegnerizzati in un periodo storico antecedente, non solo alla logica di rendere tali dati disponibili a chiunque per il riutilizzo (dati di tipo aperto), ma addirittura alla stessa Internet.

In questa cornice, il paradigma dei dati di tipo aperto (Open Data) si sta comunque affermando. Si parla spesso di Open Data per connotare la semplice esposizione di dati della pubblica amministrazione in formato aperto su una qualsiasi piattaforma Web, tipicamente per scopi di trasparenza amministrativa. Pensare tuttavia ai dati di tipo aperto in termini esclusivi di trasparenza è limitante e ne riduce le potenzialità.

L'obiettivo è quello di far sì che i dati delle amministrazioni siano visti come elemento infrastrutturale e costituiscano una ricchezza per il Paese, un'opportunità di sviluppo economico, di crescita occupazionale, di riduzione degli sprechi e di aumento dell'efficienza.

Per raggiungere questo obiettivo si ritiene indispensabile definire opportune politiche che, grazie anche alle nuove tecnologie digitali che fungono da leve di innovazione e di politica economica, possano portare complessivamente a una valorizzazione del vasto patrimonio informativo pubblico in possesso delle amministrazioni. Tali politiche richiamano la forte attenzione sui modelli di realizzazione dei sistemi informativi e di raccolta di dati e informazioni da parte delle amministrazioni.

Affinché i dati pubblici costituiscano una ricchezza e un'opportunità per le stesse amministrazioni è necessario che il processo di apertura di tali dati sia esplicitamente previsto nella realizzazione dei sistemi informativi.

Con la recente definizione dell'Agenda Digitale (DL n. 179/2012), l'Italia si è dotata di una normativa all'avanguardia relativamente alle modalità di gestione e accesso ai dati pubblici. Tale normativa non si limita a prevedere il principio dell'”Open Data by default”, ma prevede politiche di valorizzazione, strumenti e indirizzi che possono facilitare la produzione di dati aperti di qualità, interoperabili ed utilizzabili da umani e applicazioni informatiche.





Figura 1: Processo di valorizzazione del patrimonio informativo pubblico

In coerenza con il ruolo che le norme hanno attribuito all’Agenzia per l’Italia Digitale, intesa come un abilitatore dell’economia digitale e non solo come ente di regolazione o di indirizzo, la succitata normativa pone in capo all’Agenzia il compito di promuovere le politiche nazionali per la valorizzazione del patrimonio informativo pubblico. A tal riguardo, la Figura 1 illustra l’intero processo di valorizzazione dell’informazione pubblica che l’Agenzia per l’Italia Digitale mette in pratica ogni anno. In generale, il patrimonio informativo delle Pubbliche Amministrazioni può essere visto come un sistema controllato: le amministrazioni ricevono in input una strategia di valorizzazione del loro patrimonio basata su principi fondamentali del paradigma Open Data, e delle linee di azione per l’attuazione della strategia; esse forniscono poi in output il patrimonio pubblico in formato aperto; quest’ultimo è monitorato sulla base di metriche di qualità attraverso cui derivare indicatori atti a verificare l’aderenza ai principi della strategia e alle linee d’azione suddette e, conseguentemente, a guidare possibili futuri interventi (si veda il ciclo mostrato in Figura 1). Per realizzare tale ciclo, gli elementi principali da redigere sono: (i) l’agenda nazionale, (ii) le linee guida e (iii) il rapporto annuale. L’agenda nazionale (una versione bozza dell’agenda è disponibile in [94]) include la strategia di valorizzazione del patrimonio pubblico. Essa, tenuto conto dell’agenda digitale italiana ed europea e del mercato in termini di domanda e offerta, individua i principi, gli obiettivi e il piano di implementazione per la produzione e il rilascio di dati pubblici di tipo aperto.

Le linee guida, alle quali le pubbliche amministrazioni si uniformano, definiscono gli interventi principali da compiere per l’attuazione della strategia specificata dall’agenda nazionale. Esse propongono quindi un modello operativo per la produzione e la gestione dei dati pubblici, e identificano gli standard tecnici e le “best practice” di riferimento per la rappresentazione e la descrizione dei dati da aprire secondo il piano di implementazione definito nell’agenzia nazionale.

Infine, il rapporto annuale sullo stato di avanzamento del processo di valorizzazione del patrimonio pubblico rappresenta lo stato di quanto attuato dalle pubbliche amministrazioni rispetto all’agenda nazionale e alle linee guida, sulla base di indicatori qualitativi e quantitativi. Il rapporto annuale, redatto dall’Agenzia per l’Italia Digitale, aiuta a identificare le strategie e a calibrare opportunamente gli obiettivi e le azioni dell’agenda nazionale e delle linee guida dell’anno successivo.



Il presente documento di linee guida si propone quindi come supporto alle amministrazioni consentendo loro di identificare quell'insieme di azioni da compiere al proprio interno per attuare le disposizioni sulle modalità di accesso ai dati pubblici contenute nell'agenda digitale italiana e nella recente direttiva europea per il trattamento dell'informazione del settore pubblico [3]. Le linee guida approfondiscono quindi diversi aspetti organizzativi, tecnici, legali ed economici legati alla catena di produzione e rilascio di dati di tipo aperto.

3.1. Quadro normativo

In ambito nazionale, la nuova normativa per i dati di tipo aperto e le norme, ancor più recenti, in materia di trasparenza dell'azione amministrativa e di pubblicazione tracciano un quadro prescrittivo per le amministrazioni chiaramente orientato alla pubblicazione di dati in formato di tipo aperto e liberamente utilizzabili, salvo casi eccezionali da individuare sulla base di criteri oggettivi e trasparenti.

In particolare, l'art. 9 del DL n. 179/2012, convertito in Legge n. 221/2012, ha interamente riscritto l'art. 52 del CAD sull'accesso telematico e riutilizzo dei dati delle pubbliche amministrazioni. Esso stabilisce che le amministrazioni, i gestori di pubblici servizi e le società partecipate a maggioranza pubblica, inserite nel conto economico consolidato della pubblica amministrazione, *“pubblicano nel proprio sito web, all'interno della sezione Trasparenza, valutazione e merito (ora "Amministrazione trasparente" secondo quanto previsto dal Dlgs n. 33 del 14 marzo 2013), il catalogo dei dati, dei metadati e delle relative banche dati in loro possesso ed i regolamenti che ne disciplinano l'esercizio della facoltà di accesso telematico e il riutilizzo, fatti salvi i dati presenti in Anagrafe tributaria.* Lo stesso articolo stabilisce che *“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”* (principio dell'Open Data by default). *L'eventuale adozione di una licenza è motivata ai sensi delle linee guida nazionali”.* Particolarmente innovativa è la previsione, sempre definita all'art. 9 del DL suddetto, relativa alle clausole da inserire nei capitolati e negli schemi di contratto per i servizi ICT che prevedono la raccolta e la gestione di dati pubblici. In questi casi, le pubbliche amministrazioni devono prevedere *“clausole idonee a consentire l'accesso telematico e il riutilizzo, da parte di persone fisiche e giuridiche, di tali dati, dei metadati, degli schemi delle strutture di dati e delle relative banche dati”.* La norma stabilisce inoltre che *“le attività volte a garantire l'accesso telematico e il riutilizzo dei dati delle pubbliche amministrazioni rientrano tra i parametri di valutazione della performance dirigenziale”.* Infine, lo stesso art. 9 riscrive l'art. 68, comma 3, del CAD sulla definizione dei dati di tipo aperto identificandone le principali caratteristiche, come riportato nella Sezione 3.3, e pone in capo all'Agenzia per l'Italia Digitale il compito di individuare i casi eccezionali di dati aperti rilasciati a tariffe superiori ai costi marginali per la loro riproduzione e divulgazione.

I dati di tipo aperto sono stati in seguito ripresi nell'ambito della trasparenza amministrativa nel Dlgs n. 33 del 14 marzo 2013 recante *“riordino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni”.* Tale Dlgs nasce dall'applicazione delle disposizioni di cui all'art. 1 comma 35 della Legge n. 190 del 6 novembre 2012 in materia di anticorruzione dove già erano incluse alcune indicazioni su un insieme di dati da pubblicare in *“formato digitale standard di tipo aperto”* (e.g., dati relativi alle gare d'appalto, dati relativi alle consulenze e agli incarichi, ecc.). Il Dlgs n. 33 riprende tali indicazioni e fornisce un elenco puntuale di dati pubblici da rendere disponibili nei siti istituzionali delle amministrazioni nella sezione *“Amministrazione trasparente”*, sezione che *sostituisce* la precedente sezione *“Trasparenza, valutazione e merito”* avendo il Dlgs riformulato interamente l'art. 54 del CAD sul contenuto dei siti delle pubbliche amministrazioni. I dati da pubblicare inclusi nel Dlgs coinvolgono diverse aree amministrative e, secondo quanto previsto



all'art. 7, devono essere pubblicati in formato di tipo aperto ai sensi dell'art. 68 del CAD.

Se per i dati pubblici il CAD e le nuove opportunità aperte dagli Open Data hanno portato alla formulazione di nuove regole di accesso e diffusione, in ambito statistico l'attenzione alla fornitura, diffusione e accesso alle informazioni da parte degli utenti risulta antecedente rispetto alla nascita del paradigma dell'Open Data, pur se in origine con specifico riferimento al concetto di dato aggregato.

Nell'ambito dei dati statistici esistono, da tempo, specifiche normative di riferimento quali ad esempio il Dlgs. n. 322 del 6 settembre 1989 - Norme sul Sistema statistico nazionale e sulla riorganizzazione dell'Istituto nazionale di statistica - nel quale viene evidenziata e ripresa più volte la necessità di interconnettere, a fini statistici, i sistemi informativi delle pubbliche amministrazioni e degli enti facenti parte del Sistema statistico nazionale, garantendo l'accesso alle informazioni prodotte dalle singole amministrazioni. Tale riferimento è presente nell'art. 6, del Dlgs. n. 322/1989 che disciplina i compiti degli uffici di statistica che: *“a) promuovono e realizzano la rilevazione, l'elaborazione, la diffusione e l'archiviazione dei dati statistici che interessano l'amministrazione di appartenenza, nell'ambito del programma statistico nazionale; b) forniscono al Sistema statistico nazionale i dati informativi previsti del programma statistico nazionale relativi all'amministrazione di appartenenza, anche in forma individuale ma non nominativa ai fini della successiva elaborazione statistica; c) collaborano con le altre amministrazioni per l'esecuzione delle rilevazioni previste dal programma statistico nazionale; d) contribuiscono alla promozione e allo sviluppo informatico a fini statistici degli archivi gestionali e delle raccolte di dati amministrativi.”*

Per i compiti suddetti, *“gli uffici di statistica hanno accesso a tutti i dati statistici in possesso dell'amministrazione di appartenenza, salvo eccezioni relative a categorie di dati di particolare riservatezza espressamente previste dalla legge. Essi possono richiedere all'amministrazione di appartenenza elaborazioni di dati necessarie alle esigenze statistiche previste dal programma statistico nazionale”*. Inoltre, in base poi all'art. 10 che regola l'accesso ai dati statistici, il decreto prevede che: *“I dati elaborati nell'ambito delle rilevazioni statistiche comprese nel programma statistico nazionale sono patrimonio della collettività e vengono distribuiti per fini di studio e di ricerca a coloro che li richiedono [...]”*.

Infine, è importante sottolineare come il concetto del riuso venga più volte evidenziato nella normativa nazionale sia per ciò che concerne i dati, come prima riportato, sia per ciò che riguarda i programmi informatici dove *“le pubbliche amministrazioni [...] hanno obbligo di darli in formato sorgente, completi della documentazione disponibile, in uso gratuito ad altre pubbliche amministrazioni che li richiedono”* ai sensi dell'art. 69 del CAD.

In ambito europeo, si rammentano alcune indicazioni fornite dall'agenda digitale europea: la prima indicazione è quella dell'azione n. 3 del pilone relativo allo sviluppo del mercato unico digitale europeo. Tale azione individua alcune linee strategiche per gli stati membri relativamente all'apertura di dati pubblici per il loro riutilizzo. La seconda indicazione è quella dell'azione n. 85 relativa alla revisione dell'accesso pubblico alla direttiva sulle informazioni ambientali.

Sempre in ambito europeo è importante richiamare le disposizioni contenute nella direttiva sull'informazione del settore pubblico (PSI). La prima versione della direttiva risale al 2003; essa è stata recepita in Italia dal Dlgs n. 36 del 2006. Nata diversi anni prima dall'affermazione dell'Open Data, la direttiva è stata recentemente emendata accogliendo alcuni dei principi basilari dell'Open Data. In particolare, tra le novità di rilievo si segnala che, secondo la nuova versione della direttiva, pubblicata in gazzetta ufficiale lo scorso 26 giugno 2013, la diffusione della PSI non è più una raccomandazione ma diventa obbligatoria; essa estende il suo ambito di applicazione anche ai musei, biblioteche e archivi; il principio del riuso è rivisto chiarendo che tutti i documenti pubblici possono essere riutilizzati sia per fini commerciali sia per scopi non commerciali; le amministrazioni rendono disponibili i loro dati e



documenti, ove possibile, in formati aperti e “machine-readable”, accompagnandoli dai relativi metadati; sia i formati che i metadati dovrebbero essere per quanto possibile conformi a formati aperti standard; le licenze da associare ai dati non dovrebbero limitare le possibilità di riutilizzo dei dati e documenti; i costi per il riuso dei documenti dovrebbero essere limitati ai soli costi marginali sostenuti per la produzione, riproduzione e divulgazione dei dati.

Infine, il recente G8 ha posto particolare attenzione al tema dell’Open Data intervenendo su tre direttrici: la stesura di una Carta degli Open Data [4] che sancisce un insieme di principi di base che gli Stati recepiscono al fine di rendere disponibili e riutilizzabili i dati raccolti e gestiti dalle pubbliche amministrazioni, l’impegno da parte degli Stati di far propri i principi di tale Carta, e un piano di azione per il rilascio di talune tipologie di dati.

Alla luce di tali direttrici, emerge come sia l’agenda nazionale, sia le presenti linee guida sulla valorizzazione del patrimonio informativo pubblico italiano siano già allineati con quanto stabilito in sede G8, estendendo alcuni degli obiettivi sanciti al fine di cogliere le peculiarità proprie del nostro Paese.

3.1.1. Regolamenti delle pubbliche amministrazioni e direttive locali

Diverse pubbliche amministrazioni centrali (ad esempio, Presidenza del Consiglio dei Ministri, Ministero dei Beni Culturali, INPS, e ISTAT per citarne alcune) e locali hanno affrontato internamente il tema dei dati di tipo aperto dotandosi, in taluni casi, di veri e propri regolamenti oppure definendo degli atti di indirizzo per l’individuazione delle basi di dati pubbliche in loro possesso e per le relative modalità di apertura. Ai sensi dell’art. 9 del succitato DL n. 179/2012, in sede di prima applicazione della normativa sui dati aperti, i regolamenti già prodotti e approvati sono pubblicati in formato aperto sul sito istituzionale nella sezione “Amministrazione Trasparente”.

Tuttavia, al fine di rendere sistemico e omogeneo su scala nazionale il processo di valorizzazione dei dati pubblici, è auspicabile che i suddetti regolamenti, inclusi eventuali futuri di cui le pubbliche amministrazioni vorranno dotarsi, si uniformino ai principi e alle linee d’azione dell’agenda e delle linee guida nazionali sulla valorizzazione del patrimonio informativo pubblico, così come definite ogni anno dall’Agenzia per l’Italia Digitale.

A tal fine, è utile analizzare quanto già messo in campo da alcune amministrazioni. La sezione pertanto riassume i principali provvedimenti adottati soprattutto a livello locale.

Sul piano normativo regionale, il fenomeno degli Open Data è leggibile sotto un duplice profilo, in relazione all’evoluzione del concetto di trasparenza o legato al concetto della società della conoscenza, nella quale i dati pubblici, soprattutto quando di titolarità della pubblica amministrazione, diventano materia prima per l’elaborazione di ricerche, prodotti e servizi. Un’articolata infrastruttura normativa, composta di un insieme di provvedimenti e linee di indirizzo, fornisce la cornice per inquadrare il tema dei dati pubblici e del loro riutilizzo nell’ambito delle regioni italiane.



La sensibilità delle regioni nei confronti degli OD si dimostra, in alcuni casi, già a partire dal 2004 (Emilia Romagna, Toscana)¹, quando ancora deve cominciare l'intenso percorso normativo nazionale costellato da tappe intermedie e rappresentato in prima battuta dal Dlgs del 2006², che recepisce la Direttiva europea 2003/98/CE relativa al riutilizzo dell'informazione del settore pubblico³. Si individuano i prodromi dei principi di valorizzazione e condivisione del patrimonio informativo, nonché di accessibilità e disponibilità dei dati che caratterizzano l'attuale Open Government Data.

Le Regioni "trend setters", che rappresentano nell'ambito nazionale realtà trainanti per l'adozione di iniziative sui dati pubblici, dimostrano percorsi simili nella delineazione di un contesto normativo che prevede:

- approvazione di programmi o piani strategici per l'e-government e l'ICT;
- leggi specifiche sul tema Open Data;
- Linee Guida per la governance del sistema dei dati pubblici come strumento attuativo delle strategie in materia.

Questo dunque lo sviluppo normativo per Emilia Romagna⁴, Lombardia⁵, Piemonte⁶, Puglia⁷, Toscana⁸ e Umbria⁹ accomunate dal simile criterio di impostazione della politica di gestione della più ampia materia relativa alla società dell'informazione e della conoscenza.

I principi condivisi nel contesto normativo delle regioni in materia di Open Data, che sono denominatori comuni alla base dei percorsi normativi delle singole regioni, si possono riassumere come segue:

¹ LR 11/2004 dell'Emilia Romagna sullo *Sviluppo regionale della società dell'informazione*; Leggi Regionali della Toscana 1/2004 *Promozione dell'amministrazione elettronica e della società dell'informazione e della conoscenza nel sistema regionale. Disciplina della "Rete telematica regionale toscana"* e 40/2009 *Norme sul procedimento amministrativo, per la semplificazione e la trasparenza dell'attività amministrativa* (come successiv. modificata, da ultimo da l.r. 9/2013).

² Decreto Legislativo del 24 gennaio 2006 n. 36, "Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico", come successivamente modificato ed integrato dalla L. 96 del 2010.

³ DIRETTIVA 95/46/CE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 24 ottobre 1995 relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati.

⁴ Regione Emilia: LR 11/2004 sullo *"Sviluppo regionale della società dell'informazione"*, successiva DGR 52/2011 di *"approvazione delle nuove Linee Guida al Piano Telematico dell'Emilia-Romagna 2011-2013 (PiTER)"* e DGR 1587/2011 di approvazione del Programma Operativo 2011 del Piano Telematico, e DGR 2080/2012 *"LINEE GUIDA RELATIVE AL RIUTILIZZO E MESSA A DISPOSIZIONE IN OPEN DATA DEI DATI PUBBLICI DELL'AMMINISTRAZIONE REGIONALE"*.

⁵ Regione Lombardia: DGR 2585/2011 *"Approvazione Agenda Digitale Lombarda 2012-2015"*, successiva DGR 2904/2012 *"Approvazione criteri generali per l'Open Data"*, Decreto DG Semplificazione e Digitalizzazione 6115/2012 *"approvazione «PIANO DI GOVERNANCE OPEN DATA»"* e DRG 4324/2012 *"Approvazione delle Linee guida Open Data per gli Enti Locali"* e delle *"Modalità di adesione alla piattaforma regionale Open Data"*.

⁶ Regione Piemonte: D.G.R. 964/2010 di *"Approvazione linee guida del Piano pluriennale per la competitività"*, successiva LR 24/2011 recante *"Disposizioni in materia di pubblicazione tramite la rete internet e di riutilizzo dei documenti e dei dati pubblici dell'amministrazione regionale"* e DGR n. 4687/2012 di *"Approvazione del documento Linee Guida relative al riutilizzo e alla diffusione tramite la rete internet dei documenti e dei dati pubblici dell'Amministrazione regionale"* e relativi allegati⁷.

⁷ Regione Puglia: L.R. 20/2012 *"Norme sul software libero, accessibilità di dati e documenti e hardware documentato"* e successiva DGR 2183/2012 *"Legge regionale 24 luglio 2012, n. 20 - "Norme sul software libero, accessibilità di dati e documenti e hardware documentato" - Primi indirizzi."*

⁸ Regione Toscana: LR 54/2009 recante *"Istituzione del sistema informativo e del sistema statistico regionale. Misure per il coordinamento delle infrastrutture e dei servizi per la società dell'informazione e della conoscenza"*, successivo *"Programma regionale per la promozione e lo sviluppo dell'amministrazione elettronica e della società dell'informazione e della conoscenza nel sistema regionale 2012-2015"* e DGR 23/2013 *"Realizzazione piattaforma open data e approvazione linee guida recanti criteri generali per gli open data in Regione Toscana"*.

⁹ Regione Umbria LR 8/2011 sullo sviluppo dell'Amministrazione digitale che prevede all'art.15 un repertorio regionale dei dati pubblici aperti, D.G.R. 397/2012 sulle Linee guida dell'Agenda digitale dell'Umbria e D.G.R. 1551/2012 di approvazione del Disciplinare dei criteri generali per l'open data in Umbria.



- applicazione del principio che i dati degli enti pubblici sono un bene che appartiene alla comunità;
- impegno di trasparenza da parte dell'ente pubblico nei confronti della cittadinanza a garanzia di condizioni eque e non discriminatorie dell'accesso alle informazioni pubbliche;
- miglioramento della qualità della vita dei cittadini grazie all'opportunità di una maggiore conoscenza e partecipazione;
- avanzamento sul piano dell'attuazione delle agende digitali regionali e nazionale e dell'e-government;
- valorizzazione del patrimonio informativo degli enti pubblici;
- stimolo alla "creatività collettiva" e acceleratore di opportunità di impresa e di lavoro, con applicazione di licenze d'uso aperte al riuso anche a fini commerciali;
- promozione di una rivoluzione culturale volta ad avvicinare pubblica amministrazione, cittadini, mondo dell'impresa, e tutti gli attori della società civile;
- indirizzo evolutivo verso il linked open data come dato arricchito e "machine-readable", in ottica di sviluppo di servizi a valore aggiunto sempre maggiore.

Il riuso e la messa a disposizione delle basi di dati pubbliche sono elementi che le Linee Guida del Piano della competitività 2011-2015, approvate dalla Regione Piemonte, cita come fattori abilitanti in termini di competitività e sviluppo economico regionale. Anche nel caso della Lombardia è enfatizzato il legame tra Open Data e competitività e sviluppo economico come testimonia l'inserimento del tema OD quale componente importante della LR n. 7/2012 - Misure per la crescita, lo sviluppo e l'occupazione.

Carattere di uniformità rispetto al citato contesto di norme e provvedimenti è l'adozione finale di Linee Guida (pur con modalità differenti come nei casi della Regione Puglia che approva i cosiddetti primi indirizzi, e della Lombardia che emana due DGR rispettivamente sull'approvazione di criteri generali e del piano di governance per Open Data e una ulteriore DGR di approvazione di linee guida per adesione e partecipazione al progetto da parte degli EE.LL) nelle quali vengono definiti:

- principi e finalità della strategia;
- strumenti e indirizzi di governance del processo di identificazione, analisi e pubblicazione dei dati nonché di collaborazione interna all'Ente;
- criteri per definire una corretta ed uniforme comunicazione, la promozione di metodologie tecniche in grado di garantire l'interoperabilità dei dati, e l'utilizzo di opportune licenze;
- criteri per la definizione della partecipazione alla piattaforma dati da parte di altre amministrazioni del territorio, dei cittadini e delle imprese.

L'esigenza da parte delle Regioni di dare attuazione pratica al sistema degli OD si è realizzata, in alcuni casi, attraverso l'evoluzione parallela tanto delle piattaforme tecnologiche quanto del contesto normativo ed organizzativo (Emilia Romagna, Lombardia e Piemonte); in altri casi (Toscana e Puglia) l'attivazione della piattaforma avviene a seguito della definizione del modello di governance, come strumento attuativo della politica in materia.



La Tabella 4 compara i piani di azione di alcune regioni evidenziando le principali scelte operate da ognuna (per maggiori dettagli, i lettori interessati possono riferirsi a [5]).

Lo scopo di tale comparazione è di condividere esperienze di successo, evidenziandone in particolare gli aspetti comuni (ad esempio, si nota come la maggior parte delle regioni analizzate in queste linee guida abbia deciso di adottare licenze internazionali della famiglia Creative Commons – CC [6]). Questo, in ottica e spirito di riuso, può guidare le scelte interne delle amministrazioni, centrali e locali, che si trovano all’inizio o non hanno ancora intrapreso il percorso Open Data.

Regione Provincia autonoma	Legge Regionale	Linee guida	Portale Open Data	Condivisione altri enti	Licenza	Organizzazione processi di liberazione dati
Piemonte	LR n. 24 del 20/12/2011	DGR n.22- 4687 del 08/10/2012	Attivo da 06/2010 il portale dati.piemonte.it	Portale aperto a tutti gli enti del territorio. 79 enti del territorio pubblicano dati	CC0 e CC-BY	Processo interno di coinvolgimento e responsabilizzazio ne delle singole strutture dell’ente - processo di richiesta formale dati dall’esterno, vincolante per l’ente
Emilia Romagna	LR n. 11 del 2004	DGR n. 2080 del 12/2012. Inoltre, vademecum per specifici aspetti	Attivo da 10/2011 il portale dati.emilia- romagna.it	La piattaforma è predisposta per indicizzare e memorizzare dati di titolarità delle PA, organismi di diritto pubblico, imprese pubbliche	CC0 e CC-BY	Le Linee Guida tendono anche a chiarire le responsabilità in capo ai vari “attori” coinvolti nel processo di individuazione e messa a disposizione del dato
Toscana	LR n.54/2009	DGR n. 23 del 21/01/2013	Attivo il portale dati.toscana.it	Portale aperto agli enti del territorio. 2 enti pubblicano dati	CC-BY	Processo interno di sensibilizzazione responsabilizzazio ne che coinvolge tutte le strutture



Regione Provincia autonoma	Legge Regionale	Linee guida	Portale Open Data	Condivisione altri enti	Licenza	Organizzazione processi di liberazione dati
						interne. Costituzione di un gruppo di lavoro
Puglia	LR n. 20/2012	DGR n. 2183 del 31/10/ 2010	Attivo il portale www.datipuglia .it			E' assicurato il coordinamento delle azioni per ottenere la disponibilità dei dati generati da tutti le strutture regionali e dagli enti, agenzie e società indicate nell'art. 2 della LR n. 20/2012
Provincia autonoma di Trento	LP n. 16/2012	DGR n. 2858 del 27/12/2012	Attivo da 03/ 2013 dati.trentino.it Nel febbraio 2012 aperti 160 tematismi geografici in modalità linked data	Punto unico di accesso per il territorio. Già collegato ai dati aperti del Comune di Trento	CC0 e CC-BY	Prevista programmazione annuale da parte di ogni dipartimento allineata alla pianificazione ict e innovazione. Avviato un G.L. Pat trasversale su tematiche statistiche, legali, di semplificazione. Gruppo di progetto dedicato adiuvato dal sistema della ricerca locale TrentoRise – Fbk
Lombardia	LR n. 7/2012 art. 52 (Accessibilità e	DGR n. 2904 del 11/01/2012	Attivo da 3/2012 il portale	Portale aperto agli enti del territorio, come definito in	IODL 2.0	Decreto DG Semplificazione e Digitalizzazione n.



Regione Provincia autonoma	Legge Regionale	Linee guida	Portale Open Data	Condivisione altri enti	Licenza	Organizzazione processi di liberazione dati
	valorizzazione del patrimonio informativo pubblico)		dati.lombardia.it	DRG n. 4324 del 26/10/2012		6115 del 10/07/2012 - approvazione «piano di governance Open Data»
Veneto		DGR n. 2301 del 29/12/2011	Attivo da 12/2011 dati.veneto.it	Il portale è aperto a tutti gli enti del territorio. 17 amministrazioni pubblicano dati	IODL 2.0 CC- BY	Costituzione di un gruppo di lavoro (CDG-OD)
Liguria	LR n. 42/06 "Istituzione del Sistema informativo integrato regionale per lo sviluppo della Società dell'informazione in Liguria"	DGR n. 889/2012	Attivo dal 24/10/2012 dati.liguria.it	Previsto da DGR n. 889/2012	CC-BY	Secondo le modalità previste in DGR n. 889/2012
Umbria	LR n. 8/2011 sviluppo amministrazione digitale; prevede all'art.15 un repertorio regionale dei dati pubblici aperti	DGR n. 1551/2012	Attivo dal 15/04/2013 dati.umbria.it	Il repertorio è aperto a tutti gli enti del territorio, 3 enti pubblicano dati		Secondo le modalità previste in DGR 1551/2012

Tabella 1: Comparazione di sintesi dei piani di azione delle regioni esaminate

In ambito comunale, i Comuni che hanno già intrapreso il percorso di regolamentazione del paradigma degli Open Data si sono avvalsi di uno o più atti di indirizzo nei quali hanno rappresentato la volontà dell'amministrazione di garantire la trasparenza della propria azione e la finalità di utilizzare gli Open Data come volano dell'economia digitale. Gli atti sono quasi sempre accompagnati da un progetto



tecnico, spesso dettagliato, delle azioni programmate per realizzare l'apertura del proprio patrimonio informativo.

In molti casi la deliberazione di Giunta ha riguardato il tema più ampio dell'attuazione dell'agenda digitale a livello locale, individuando l'Open Data come uno tra gli ambiti prioritari di intervento dell'amministrazione, sia come risposta ai recenti dettami normativi in termini di trasparenza e accessibilità/disponibilità del dato, sia come opportunità di sviluppo locale attraverso la realizzazione di applicazioni basate sulle informazioni pubbliche.

Si riscontra una maggiore proattività nelle città e nei Comuni medio/grandi, spesso ad alta vocazione turistica e quindi interessati da picchi di presenze ben al di sopra della popolazione residente, a cui destinare una serie di informazioni utili sulla città, nonostante sussistano alcuni casi isolati di dimensioni demografiche più piccole.

Il processo di regolamentazione è ancora in una fase iniziale, pertanto sarà più efficace effettuare una analisi in una fase più avanzata, quando, a prescindere dalla coerenza normativa e dalla imprescindibile necessità di sintesi tra i due aspetti - trasparenza e opportunità di sviluppo - che caratterizzano il fenomeno degli open data, si avrà avuto modo di confrontare un panorama di esperienze più ampio.

3.2. Aspetti di sussidiarietà

Al fine di valorizzare il patrimonio informativo pubblico disponibile presso i piccoli comuni, in coerenza anche con il DL n. 95/2012, bisognerà, per i comuni con popolazione pari o inferiore ai 5.000 abitanti, procedere obbligatoriamente, nelle forme e modalità più consone (Convenzioni o Unioni di Comuni), alle gestioni associate nelle attività necessarie alla divulgazione dei dati da ciascuna singola amministrazione comunale. Inoltre, con l'obiettivo precipuo di raggiungere capillarmente e coinvolgere in questo processo di innovazione e apertura dei dati tutti i piccoli comuni, anche montani e delle "aree interne", potrà essere avviata un'azione congiunta CISIS, ANCI e Agenzia per l'Italia Digitale finalizzata a definire, attuare e dispiegare politiche tese all'informazione/formazione dei dipendenti comunali e delle Unioni di Comuni sul tema degli Open Data. Pertanto, le gestioni associate dei comuni, come intermediari tra le Regioni e i piccoli comuni, consentiranno di valorizzare i dati disponibili presso le piccole amministrazioni pubbliche che dispongono di apparati tecnici assai ridotti sia in termini di personale che di dotazioni strumentali.

3.3. Definizioni: dati delle pubbliche amministrazioni e documenti

La genericità, la complessità e l'ampiezza del contesto dei dati della pubblica amministrazione, dovuto anche alle molteplici contrapposizioni o sovrapposizioni tematiche e/o territoriali connesse alle competenze delle singole amministrazioni, rende necessario circoscrivere l'ambito di applicazione delle presenti linee guida. A tal fine è opportuno ripartire dalle definizioni contenute nel Codice dell'Amministrazione Digitale (art. 1):

"m) dato delle pubbliche amministrazioni: il dato formato, o comunque trattato da una pubblica amministrazione;

l) dato a conoscibilità limitata: il dato la cui conoscibilità è riservata per legge o regolamento a specifici soggetti o categorie di soggetti;



n) dato pubblico: il dato conoscibile da chiunque;

o) disponibilità: la possibilità di accedere ai dati senza restrizioni non riconducibili a esplicite norme di legge;"

Il concetto di dato pubblico esclude, in linea generale, i dati personali per i quali trovano applicazione le norme del "Codice in materia di protezione dei dati personali" (i.e., Dlgs n. 196/2003 e deliberazione del 88/2011 dell'Autorità Garante per la protezione dei dati personali). Laddove, in un contesto informativo, il dato pubblico contiene riferimenti o è collegato a dati personali trova applicazione il comma 5 dell'articolo 2 del CAD "Le disposizioni del presente codice si applicano nel rispetto della disciplina rilevante in materia di trattamento dei dati personali e, in particolare, delle disposizioni del codice in materia di protezione dei dati personali approvato con decreto legislativo 30 giugno 2003, n. 196. I cittadini e le imprese hanno, comunque, diritto ad ottenere che il trattamento dei dati effettuato mediante l'uso di tecnologie telematiche sia conformato al rispetto dei diritti e delle libertà fondamentali, nonché della dignità dell'interessato." o altre norme specifiche che consentono la pubblicazione di certe tipologie di informazioni, come ad esempio le norme sulla trasparenza come prima descritto (Dlgs n. 33/2013).

Escludendo i dati a conoscibilità limitata (come ad esempio i dati coperti da segreto di stato o le opere d'ingegno coperte dal diritto d'autore), oggetto delle presenti linee guida è il dato pubblico, cioè il dato della pubblica amministrazione conoscibile da chiunque.

Nelle enunciazioni contenute nell'art. 1 del CAD, la definizione di disponibilità conferma l'accezione di dato pubblico, e consente di distinguere tra dato pubblico già disponibile (ad esempio già pubblicato sul Web) e dato pubblico non ancora disponibile, ma che potrà esserlo successivamente.

Occorre inoltre fare riferimento alle definizioni riportate nell'art. 68, comma 3, del CAD, come modificato dalle norme contenute nell'art. 9 del DL n. 179/2012, convertito con la Legge n. 221/2012, recante "Ulteriori misure urgenti per la crescita del Paese":

a) formato dei dati di tipo aperto, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

b) dati di tipo aperto, i dati che presentano le seguenti caratteristiche:

1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato;

2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi metadati;

3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione. L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. In ogni caso, l'Agenzia, nel trattamento dei casi eccezionali individuati, si attiene alle indicazioni fornite dalla direttiva 2003/98/CE del Parlamento europeo e del Consiglio, del 17 novembre 2003, sul riutilizzo dell'informazione del settore pubblico, recepita con il decreto legislativo 24 gennaio 2006, n. 36.

Come si nota, la definizione di "formato dei dati di tipo aperto" fa riferimento agli aspetti tecnologici e



risulta finalizzato ad assicurare un adeguato livello di interoperabilità. E' invece direttamente collegata alle caratteristiche del dato la definizione di "dati di tipo aperto" che contempla tre requisiti essenziali:

- la *disponibilità* del dato, che questa nuova accezione qualifica con maggiore precisione consentendo di escludere dalla definizione di dato pubblico disponibile il dato pubblico (quindi potenzialmente conoscibile) cui è associata una tipologia di licenza che non consente il libero uso del dato stesso;
- l'*accessibilità* del dato, che richiama l'aspetto tecnologico del formato¹⁰ in cui il dato viene reso disponibile e richiede l'effettiva presenza dei metadati relativi al dato medesimo;
- la *gratuità* del dato o la sua disponibilità a costi marginali, salvo casi eccezionali da definire con atto dell'Agenzia per l'Italia Digitale.

Il *dato di tipo aperto* è quindi un dato della pubblica amministrazione, conoscibile (pubblico), a cui è associata una licenza che ne consente il libero utilizzo (disponibile) e che abbia le caratteristiche di accessibilità e gratuità come sopra definite.

La Figura 2 posiziona graficamente i dati di tipo aperto nell'insieme di dati individuabili dal contesto prima descritto.

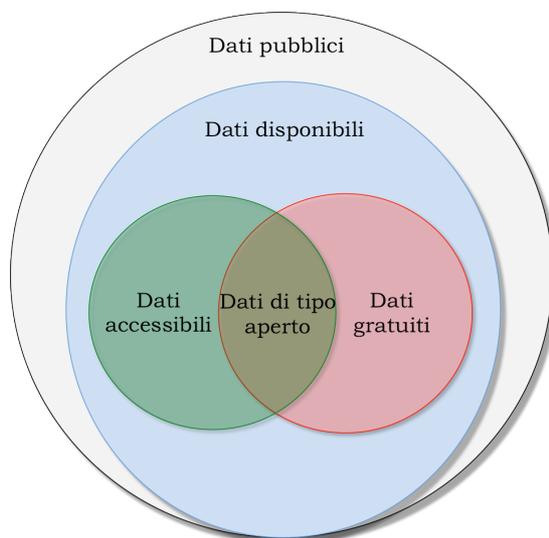


Figura 2: Tipi di dato delle pubbliche amministrazioni

Un ultimo tema di rilievo è la distinzione tra dati, atti e documenti. Anche in questo caso possiamo fare riferimento alle definizioni di documento informatico e documento analogico contenute nell'articolo 1 del CAD:

p) documento informatico: 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.

¹⁰ Con riferimento a tale aspetto, dovranno poi tenersi in considerazione le definizioni di "formato leggibile meccanicamente", "formato aperto" e "standard formale aperto" introdotti con la recente direttiva 2013/98/CE del 26 giugno 2013, che modifica la direttiva 2003/98/CE relativa al riutilizzo dell'informazione nel settore pubblico.



Anche la direttiva 2003/98/CE del 17 novembre 2003, relativa al riutilizzo dell'informazione del settore pubblico, prevede una definizione generica del termine "documento" comprendente "qualsiasi rappresentazione di atti, fatti o informazioni e qualsiasi raccolta dei medesimi".

Alla luce di tali definizioni è possibile convenire che il documento, in generale, rappresenta una collezione di informazioni (e.g., dati, contenuti multimediali, atti e narrazione di eventi) pensato principalmente per essere direttamente leggibile in forma scritta e visuale da parte di persone.

Tenuto conto che l'oggetto principale delle presenti linee guida è il dato pubblico, il rilascio di documenti è trattato nelle presenti linee guida solo nell'aspetto concernente il formato con cui renderli disponibili (sezione 6.2), e nel rispetto dei requisiti di accessibilità [7], secondo quanto anche indicato dall'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico (ad esempio, si ricorda che la scansione di un documento non è considerata accessibile).



4. MODELLI DI RIFERIMENTO PER I DATI E I METADATI

Questa sezione presenta i modelli di riferimento per la categorizzazione dei dati e dei metadati da produrre e rilasciare.

4.1. Modello per i dati

Tim Berners-Lee, considerato il padre fondatore del Web, ha proposto una serie di principi da seguire nell'apertura di dati, definendo un modello qualitativo per i dati aperti su Web, noto come modello delle cinque stelle [8].

Il modello è ampiamente riconosciuto e accettato a livello internazionale ed è quello adottato dall'agenda nazionale al fine di classificare i dati di tipo aperto delle pubbliche amministrazioni italiane.

Esso consente di individuare il livello di leggibilità dei dati pubblicati e la loro facilità d'accesso. Il modello si compone di cinque livelli come mostrato in Figura 3.

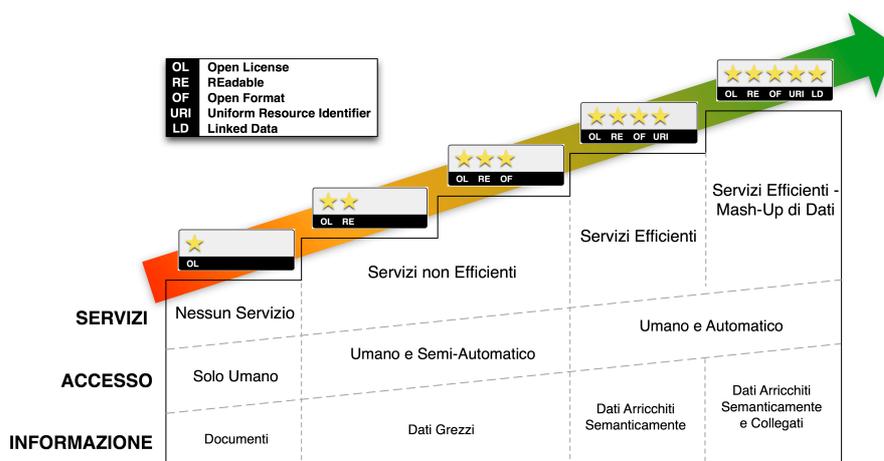


Figura 3: Modello a cinque livelli per la produzione e il rilascio di dati di tipo aperto

Più nel dettaglio, i cinque livelli sono:

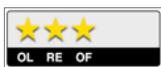


Dati disponibili in qualunque formato, ma con una licenza aperta.



Dati disponibili in un formato leggibile da un agente automatico. Tipicamente, rientrano in questo livello dati in formati proprietari (e.g., excel).





Dati con caratteristiche del livello precedente ma con un formato non proprietario.



Dati con caratteristiche del livello precedente ma esposti usando gli standard W3C RDF e SPARQL (si veda la sezione 6).



Dati con caratteristiche del livello precedente ma collegati a dati esposti da altre persone e organizzazioni. Tale livello è analizzato in dettaglio nella seguente sezione 4.1.1.

La Figura 3 arricchisce la figura nota e disponibile sul Web del modello delle cinque stelle [90] con l'aggiunta di tre differenti dimensioni: *informazione*, *accesso* e *servizi implementabili*. In particolare, a quest'ultima è data maggiore enfasi in quanto considerata la dimensione che sottende l'obiettivo da raggiungere con l'apertura dei dati. Le tre dimensioni introdotte in Figura 3, con i diversi gradi della scala, sono descritti come segue.

INFORMAZIONE

Questa dimensione descrive la qualità dell'informazione fornita insieme ai dati aperti. In questo senso si possono avere:

- *documenti*: i dati sono incorporati all'interno di documenti senza struttura e quindi leggibili e interpretabili solo da umani (livello 1);
- *dati grezzi*: i dati sono leggibili anche da un programma ma l'intervento umano è fortemente necessario per una qualche elaborazione degli stessi (livelli 2 e 3);
- *dati arricchiti semanticamente*: i dati sono descritti semanticamente tramite metadati e ontologie (livello 4);
- *dati arricchiti semanticamente e collegati*: i dati sono descritti semanticamente tramite metadati e ontologie (livello 5). L'intervento umano si può ridurre al minimo e talvolta addirittura eliminare.

ACCESSO

Questa dimensione descrive la facilità con cui utenti umani e agenti automatici riescono ad accedere ai dati e considera quindi anche lo sforzo di comprensione della struttura dei dati al fine di poterli interrogare e utilizzare in modo corretto. I gradi individuati sono:

- *solo umano*: solo gli umani sono in grado di leggere i documenti senza struttura e quindi dare un senso ai dati in esso presenti (livello 1);
- *umano e semi-automatico*: gli agenti automatici possono elaborare i dati ma non sono in grado di interpretarli; pertanto si rendere necessario un intervento umano al fine di scrivere programmi ad-hoc per il loro utilizzo (livelli 2 e 3);
- *umano e automatico*: gli agenti automatici che conoscono l'ontologia di riferimento possono elaborare



i dati senza un ulteriore intervento umano (livelli 4 e 5).

SERVIZI

Questa dimensione descrive la tipologia di servizi che possono essere progettati e implementati con i dati aperti. Dalla tipologia ne deriva il grado di efficienza e capacità con cui un servizio riesce a sfruttare informazioni anche provenienti da sorgenti diverse. In particolare, si distinguono i seguenti gradi:

- *nessun servizio*: nessun servizio può essere abilitato a partire dai dati contenuti nei documenti, a meno di significativi interventi umani di estrazione ed elaborazione dei possibili dati (livello 1);
- *servizi non efficienti*: applicazioni ad-hoc che usano i dati. Queste applicazioni devono incorporare al loro interno i dati (livelli 2 e 3);
- *servizi e apps efficienti*: applicazioni, anche per dispositivi mobili, che sfruttano accessi diretti a Web per reperire i dati di interesse (livello 4);
- *servizi efficienti e con mashup di dati*: applicazioni, anche per dispositivi mobili, che sfruttano sia accessi diretti a Web sia l'informazione ulteriore catturata attraverso i "link" dei dati di interesse (livello 5).

4.1.1. Linked Open Data

I Linked Data sono dati in formato RDF (vedi Sezione 6.1) che rispettano i seguenti quattro principi base proposti dallo stesso Tim Berners-Lee nel 2006 [8]:

- usare indirizzi Web (URI) come nomi per le "cose";
- usare URI utili al protocollo HTTP in modo che sia possibile cercarli e risolverli;
- quando qualcuno cerca una URI, fornire un'informazione utile;
- includere link ad altre URI, così da permettere a chi cerca di scoprire nuovi collegamenti.

In pratica, i Linked Data sono dati che seguono il paradigma RDF delle triple concatenate, in cui alle "cose" (o entità) è assegnata un URI univoca su 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 riferirsi (i.e., "collegarsi") tra loro. Nel riferirsi, si usano delle relazioni ("link") che hanno un significato ben definito e che spiegano quindi il tipo di legame che intercorre tra le due entità coinvolte. In questo modo, i Linked Data sono un metodo elegante ed efficace per risolvere problemi di identità e provenienza, semantica, integrazione e interoperabilità.

Il beneficio più evidente nell'utilizzo dei Linked Data è il poter sfruttare la semantica dei dati e le infrastrutture Web per abilitare la scoperta e l'inferenza di nuove informazioni, con conseguenze positive e dirette sia per gli utenti che per gli sviluppatori.

Per un approfondimento sui Linked Data nella pubblica amministrazione, i lettori interessati possono riferirsi alle linee guida pubblicate dalla Commissione di Coordinamento SPC [9].



4.2. Modello per i metadati

La metadattazione ricopre un ruolo essenziale laddove i dati sono esposti a utenti terzi e ad agenti automatici. I metadati, infatti, consentono una maggiore comprensione e aiutano pertanto la ricerca e la scoperta dei dati stessi. Alcuni dei modelli esistenti per metadati si focalizzano principalmente sugli aspetti di schema (ad esempio Dublin Core) o sul formato dei metadati stessi [10], ad esempio riproponendo una scala simile a quella vista nella sezione precedente per i dati.

Considerando il fine ultimo di promuovere la pubblicazione di dati di alta qualità, così da facilitare lo sviluppo di servizi e applicazioni, si propone un modello focalizzato sugli aspetti qualitativi dei metadati. Il modello è indipendente dal particolare schema proposto e, in parte, anche dal formato fisico di rappresentazione.

La classificazione qualitativa del modello è basata su due fattori: il *legame tra dato-metadato* e il *livello di dettaglio dei metadati*. Le combinazioni ammissibili dei gradi di questi fattori individuano quattro livelli, così come mostrato in Figura 4.

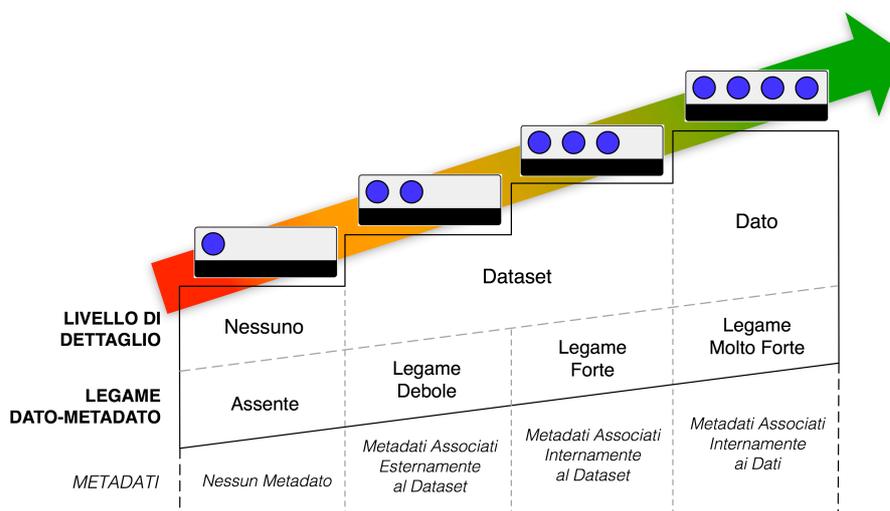


Figura 4: Modello per la definizione dei metadati

In particolare le due dimensioni sono definite come segue.

LEGAME DATO-METADATO

Questa dimensione indica quanto i metadati riescono a essere, fisicamente e logicamente, legati ai dati anche dopo un ipotetico processo di trasformazione e utilizzo:

- *assente*: i dati non sono accompagnati da un'opportuna metadattazione (livello 1);
- *legame debole*: i dati sono accompagnati da metadati esterni, ad esempio con schede nella pagina di download o in file separati (livello 2);
- *legame forte*: i dati incorporano al loro interno i metadati del dataset che li contiene (livello 3);



- *legame molto forte*: i dati incorporano al loro interno i metadati che li descrivono puntualmente (livello 4).

LIVELLO DI DETTAGLIO

Questa dimensione rappresenta l'ambito ("scope") dei metadati, ovvero la granularità della loro descrizione:

- *nessuno*: i dati non sono accompagnati da un'opportuna metadattazione (livello 1);
- *dataset*: i metadati forniscono informazioni relativamente ad un dataset, quindi sono informazioni condivise dall'insieme di dati interni a quel dataset (livelli 2 e 3);
- *dato*: i metadati forniscono informazioni relativamente al singolo dato, quindi col massimo grado di dettaglio possibile (livello 4).

In generale il livello 4 risulta essere il livello ottimale per la metadattazione. Tuttavia, per la natura dei dati stessi (ad esempio, dati con livello di dettaglio non disponibile), esso potrebbe non essere sempre perseguibile. In questi casi, il livello 3 può comunque ritenersi quello ottimale. E' opportuno evidenziare che i livelli 3 e 4 di questo modello presuppongono che i dati a cui si riferiscono siano almeno a livello 4 del relativo modello dei dati, ad eccezione di alcuni casi come, ad esempio, i dati rappresentati nello standard SDMX (si veda la sezione 6.1), che sono infatti di livello 3 del modello dei dati e di livello 4 del modello dei metadati.

4.2.1. Metadati obbligatori e obbligatori condizionatamente

Allo scopo di migliorare la qualità della metadattazione, spesso ci si riferisce a ontologie e vocabolari noti e comuni per la definizione di un insieme minimo di metadati per la descrizione dei dataset. In questo modo si facilita lo scambio di informazioni e l'interoperabilità (anche "cross-border"), favorendo di fatto il riuso e la valorizzazione dell'informazione. Tutto ciò diventa ancora più agevole se gli stessi metadati sono espressi in un formato "machine-readable", così come affermato dal modello descritto precedentemente.

Lo scopo di questa sezione è quello di indicare un insieme minimo ("core") di metadati che le pubbliche amministrazioni italiane dovranno utilizzare per accompagnare i dati che pubblicano. I metadati di quest'insieme sono stati scelti a partire da metadati ben noti a livello internazionale.

Prendendo spunto dalla recente iniziativa della Casa Bianca [11], si identificano due categorie di metadati: quelli *obbligatori* e quelli *obbligatori limitatamente al verificarsi di alcune condizioni (obbligatori condizionatamente)*.

La Tabella 2 elenca i metadati obbligatori mentre la Tabella 3 elenca quelli obbligatori al verificarsi di determinate condizioni indicate nella tabella stessa. La colonna provenienza identifica lo schema di metadati (i.e., CC – Creative Commons, DCAT – Data Catalog, DC- Dublin Core, DCMI – Dublin Core Metadata Initiative) da dove il campo è ricavabile. Per una spiegazione più approfondita degli schemi di metadati da utilizzare si può far riferimento alla Sezione 6.3.



Campo	Provenienza	Descrizione e note
license	CC	Indica la licenza utilizzata.
publisher	DC	Indica il soggetto che pubblica il dataset. Spesso coincide con creator.
creator	DC	Indica il soggetto che ha prodotto il dataset. Spesso coincide con publisher.
title	DC	Indica il titolo del dataset.
description	DC	Indica la descrizione in linguaggio naturale del dataset.
modified	DCMI	Indica la data di ultimo aggiornamento.
keyword	DCAT	Indica le parole chiave, separate da virgole, che descrivono il dataset

Tabella 2: Metadati obbligatori

Campo	Provenienza	Condizione	Descrizione e note
identifier	DC	Il dataset deve essere conforme almeno al livello 4	Indica l'URI identificativa (non obbligatoriamente l'URI fisica) per il dataset.
spatial	DCMI	Se i dati hanno significato solo all'interno di una determinata copertura spaziale.	Indica la caratterizzazione spaziale dei dati.
temporal	DCMI	Se i dati hanno significato solo all'interno di una determinata copertura temporale.	Indica la caratterizzazione temporale dei dati.
language	DC	Se la comprensione dei dati richiede la conoscenza di una determinata lingua.	Indica la lingua con cui sono espressi i dati. Si consiglia di far riferimento al vocabolario controllato RFC 4646.
accrualPeriodicity	DCMI	Se i dati sono soggetti ad aggiornamento.	Indica la frequenza di aggiornamento dei dati.



Campo	Provenienza	Condizione	Descrizione e note
byteSize	DCAT	Se la dimensione del dataset supera i 200 MB.	Indica la dimensione del dataset.
accessURL	DCAT	Se il dataset ha un endpoint di accesso.	Indica l'indirizzo del endpoint (e.g., SPARQL endpoint) a cui possiamo sottoporre le query sul dataset.
downloadURL	DCAT	Se il dataset risiede in un file scaricabile.	Indica la posizione fisica del file contenente il dataset.

Tabella 3: Metadati obbligatori al verificarsi di determinate condizioni



5. MODELLO OPERATIVO PER LA PRODUZIONE E GESTIONE DEI DATI DI TIPO APERTO

In una fase più matura della società della conoscenza come quella che stiamo attraversando, in cui si osserva la crescita dei volumi delle informazioni e delle possibilità di acquisirle anche da molteplici fonti, il problema centrale non è più come reperire l'informazione ma la capacità di intervenire sulle altre 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. Alla luce di tale contesto, rispetto al passato, oggi si rende necessario rivedere i processi e i modelli dei sistemi informativi delle pubbliche amministrazioni, organizzandoli in maniera organica e sinottica. Infatti, un dato della PA destinato alla pubblicazione è frutto di una catena di processi, più o meno complessa, nel corso della quale si generano ulteriori prodotti intermedi. Comprendere e governare la struttura di questo workflow costituisce probabilmente l'elemento chiave dell'intera attività Open Data. Tuttavia, 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 interno 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 istituzionali relativi al capitale informativo verso risorse esterne all'amministrazione.

Lo scopo di questa sezione è quindi quello di fornire un modello operativo per la produzione e gestione dei dati aperti nella pubblica amministrazione. Il modello ha come obiettivo quello di garantire la produzione e pubblicazione dei dati aperti attraverso un processo 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 nella loro applicazione alle diverse realtà amministrative. L'integrazione del modello, di conseguenza, deve essere verticale e parallela rispetto ai procedimenti interni già consolidati, e al contempo orizzontale rispetto alle necessità delle diverse amministrazioni che, nel rispetto di ogni autonomia organizzativa, devono però essere allineate sulle normative nazionali in materia di dati aperti e sulla reciproca gestione del patrimonio informativo dei dati pubblici.

Per la realizzazione del modello operativo si prevede che all'interno dell'amministrazione siano individuate delle figure strategiche per una corretta implementazione dello stesso.

5.1. 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. E' evidente che, dipendentemente dalle dimensioni delle



amministrazioni, alcune delle figure professionali sotto indicate possono coincidere o possono essere ulteriormente distinte (ad esempio il responsabile open data e il data manager potrebbero essere anche due figure separate).

Team Open Data. E' il gruppo che promuove l'uso e la diffusione degli Open Data. Riporta all'interno dell'amministrazione le novità inerenti il mondo dell'Open Government, media e valuta le esigenze di pubblicazione dati relative alla normativa sulla trasparenza, 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. Inoltre, si può occupare della formazione tecnica e concettuale all'interno dell'amministrazione riguardo i temi del 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.

Responsabile Open Data (o Data Manager). All'interno del team Open Data viene 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 paralleli già esistenti.

Il responsabile Open Data, 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, il responsabile Open Data collabora con il responsabile della trasparenza (quest'ultimo istituito ai sensi del Dlgs. n.33/2013) fornendo le indicazioni sulle risorse informative a disposizione dell'amministrazione, la qualità della loro interoperabilità, le indicazioni tecnico-operative per il rilascio dei dati della trasparenza in formato Open Data e l'eventuale riscontro in merito a possibili difficoltà nel reperimento i dati.

Titolare della banca dati. E' colui che 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. E' tipicamente un Dirigente o un Quadro che coordina un gruppo di persone che svolgono il loro lavoro quotidiano attorno alla fonte del dato in esame. 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 Titolare della banca dati; esso ha conoscenze informatiche almeno di base 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 dei cittadini sul dataset di propria competenza, e le smista eventualmente al Referente tematico per valutarne il contenuto, prima di chiedere al Titolare 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 Titolare 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 degli Open Data, sia nel promuovere nuove tipologie di dataset da esporre, sia nel validare dal punto di vista metodologico e statistico i dati pubblicati ed anche le visualizzazioni sui dati.

Ufficio giuridico-amministrativo. Può assumere le più svariate forme in base all'organizzazione interna dell'amministrazione, ma in generale rappresenta anche in estrema sintesi una singola figura che fornisce consulenza 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 ad esempio alla necessità di aprire dati prodotti da terze parti o addirittura da cittadini), e su tutte quelle problematiche di tipo giuridico o amministrativo che possono essere in qualche modo correlate all'apertura dei dati.

Team comunicazione. Anche questo componente può assumere varie forme in base all'organizzazione interna dell'amministrazione, ma in ogni caso si vogliono qui indicare le figure con competenze di comunicazione istituzionale e non solo, in grado di curare il dialogo con i cittadini e la comunicazione verso l'esterno e verso l'interno. Ha la responsabilità di pianificare e coordinare le azioni di coinvolgimento verso l'esterno e di validare il risultato delle iniziative di data visualization relative agli Open Data, prima che queste vengano pubblicate. Cura, inoltre, la segnalazione verso l'esterno di tutte le novità relative agli Open Data dell'amministrazione (e.g., nuovi dataset aperti, aggiornamenti significativi, ecc.).

5.2. Modello operativo

La Figura 5 mostra un possibile modello operativo. Utilizzando la metafora delle linee di una metropolitana, la Figura 5 illustra le principali tappe (fasi) del processo di produzione e gestione dei dati aperti da parte di una PA.



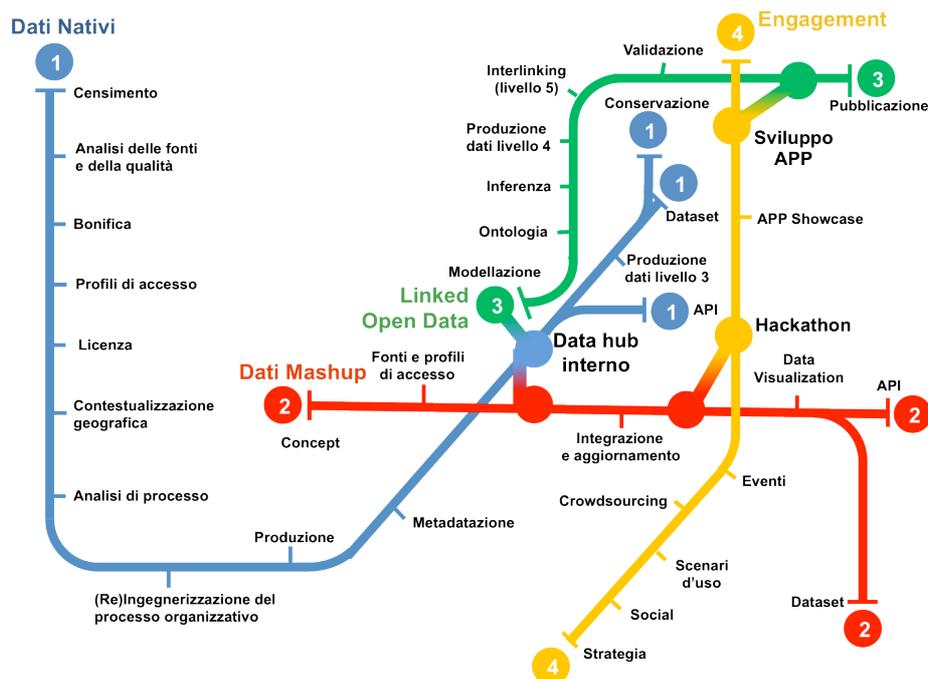


Figura 5: Modello operativo per la produzione e gestione dei dati aperti nella PA

In particolare, il modello consiste di tre linee di produzione: “*dati nativi*” che riguarda la produzione di dati nativi attraverso un’organizzazione e razionalizzazione dell’esistente, “*dati mashup*” che riguarda invece lo sviluppo di dataset complessi, prodotti a sostegno di processi comunicativi, sociali o economici, e “*Linked Open Data*” che riguarda la produzione di dati aperti di tipo Linked (livelli 4 e 5 del modello presentato in sezione 4.1) e che, come tendenza generale, è da adottare per tutti i dati pubblici che dovranno essere resi disponibili.

A queste tre linee si aggiunge quella “ortogonale” dell’engagement”. La cultura dei dati aperti deve essere sostenuta e sviluppata nel rapporto con i cittadini e il mondo delle imprese. Se è vero che gli Open Data costituiscono un potente motore di sviluppo economico e di coinvolgimento attivo della popolazione, è altrettanto vero che occorre individuare le forme e gli strumenti attraverso i quali tali processi possano effettivamente attivarsi. Per questo è necessaria una specifica filiera di attività che condivida il lavoro delle amministrazioni con tutti i possibili “stakeholder”.

5.2.1. Linea 1: Dati Nativi

Censimento. Come detto precedentemente la linea 1 dei dati nativi tratta la filiera di esposizione dell’esistente. Non c’è PA italiana, centrale o locale, che non disponga di insiemi di dati pubblicabili in formato aperto. Il loro rilascio costituisce un’includibile fonte potenziale di sviluppo cui occorre dar seguito tempestivamente, attraverso uno specifico processo produttivo.



Il percorso dei “dati nativi” parte quindi dal censimento dei dati esistenti attraverso, ove possibile, un obiettivo aziendale nell’ambito della programmazione strategica, condiviso da tutta l’amministrazione. Si raccomanda pertanto al responsabile Open Data di effettuare una ricognizione interna, in collaborazione con i titolari delle basi di dati, al fine di individuare l’insieme di dati esistenti pubblicabili in formato aperto.

A tal fine, ciascun soggetto preposto alla gestione di una particolare base di dati dovrà indicare al responsabile Open Data, tra le altre cose, le caratteristiche descrittive del dato, definire i tracciati record, precisare il tasso temporale di aggiornamento, e ogni altra informazione utile a far comprendere le caratteristiche peculiari dei dati.

E’ importante notare che nelle realtà in cui il processo di apertura dei dati ha raggiunto una fase matura, lo stesso concetto di dato nativo sta cambiando volto. Se finora i dati nativi per eccellenza erano quelli legati all’attività amministrativa di una pubblica amministrazione, per cui nella fase di censimento dell’esistente era sufficiente “semplicemente” elencare le basi dati a disposizione dell’amministrazione correlate ai diversi procedimenti amministrativi, il processo degli Open Data fa sì che si debbano raccogliere in maniera strutturata nella fase di rilevazione dei dati “nativi” anche tutte quelle informazioni che, una volta aperte, possano abilitare nuove forme di utilizzo dell’informazione e del servizio pubblico.

Ad esempio, se finora per un ufficio non era prioritario raccogliere in maniera strutturata un certo tipo di dato (ad esempio 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 verrà 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 “classici”, ma anche come espositore di patrimonio informativo che abilita nuove forme di business sul mercato.

Gli Open Data, quindi, non solo creano nuove forme di business a partire dai dati istituzionali aperti (come meglio spiegato nel seguito), ma modificano il concetto stesso di utilità del dato inserendo nella categoria dei dati “nativi” della PA delle informazioni che prima non erano ritenute tali dalla PA stessa, ma che risultano invece molto utili all’esterno.

A tal riguardo, emerge l’opportunità dell’adozione di un approccio di tipo “demand-driven” che tenga quindi conto del livello di interesse degli utilizzatori suddivisi opportunamente per categorie (e.g., cittadini, imprese, altre pubbliche amministrazioni), dei loro requisiti e delle loro necessità.

Tale processo virtuoso può portare addirittura all’eventuale modifica dei software gestionali, in modo da renderli in grado di rilevare e gestire anche queste tipologie di dati utili alla comunità una volta aperti.

Esempio: In corrispondenza di una serie di ondate di calore, nel corso della stagione estiva, il Comune di Firenze, sul modello di Chicago, ha pubblicato il dataset “luoghi freschi”. Si tratta della lista di luoghi pubblici (come biblioteche, centri anziani, centri



giovani) dove è possibile sostare liberamente in un ambiente confortevole e climatizzato in diverse fasce orarie. Il dato è stato rilasciato in forma tabellare e come mappa, nel quale ogni sito è corredato dagli orari e dalle ulteriori note caratteristiche. L'informazione è stata immediatamente rilanciata dalla stampa e il dataset è risultato primo per numero di download per diverso tempo. Questo a testimoniare il forte gradimento da parte dei cittadini. Si tratta di un dato ottenuto come sintesi di diverse fonti istituzionali, tutte già disponibili in formato aperto ma del tutto disperse rispetto all'uso specifico.

Analisi delle fonti e della qualità. Alla fase di censimento fa seguito l'analisi delle fonti e della qualità del dato. Le fonti dei dati costituiscono probabilmente l'elemento chiave del "workflow". L'analisi delle fonti consente il primo passo per la determinazione dei diritti e dei termini di licenza. In sostanza, tale tappa del modello operativo risponde alle domande "Di chi è il dato?" e "Quali sono i diritti che ne governano l'esposizione?". Le fonti si distinguono in "interne", derivanti da processi specifici tutti interni all'amministrazione, ed "esterne", quando, del tutto o in parte, provengono da soggetti esterni.

Altresì conviene ripartire la pubblicazione tra dati "grezzi", cioè come derivano direttamente dal processo produttivo, e "lavorati" e cioè sottoposti a ulteriori elaborazioni/aggregazioni o integrazioni.

In generale non è opportuno che l'esposizione del dato lavorato avvenga senza che sia stato pubblicato prioritariamente il dato grezzo. La questione va comunque definita in via di ragionevolezza, valutando caso per caso, anche sulla base delle esigenze.

Esempio: Consideriamo una città d'arte ove siano ubicati musei gestiti da diversi enti. Le presenze museali costituiscono un dato importante, anche per comprendere, da parte delle imprese turistiche, i flussi sul territorio. Alcuni di questi dati, se l'amministrazione come spesso accade gestisce direttamente alcuni musei, sono governati direttamente dall'amministrazione (fonti interne). Altri possono derivare da diversi soggetti gestori (fonti esterne). Questi ultimi possono essere esposti direttamente come Open Data o forniti per via istituzionale con adeguata licenza.

All'analisi delle fonti segue l'analisi della qualità dei dati. La qualità dei dati (o data quality) è l'insieme degli aspetti che qualificano la facilità con cui è possibile derivare l'informazione incorporata all'interno dei dati, ovvero quelli che concernono la presenza di errori, inefficienze nel processo di gestione dei dati, incompletezze, ecc. La qualità dei dati è spesso ricondotta ad un insieme di dimensioni come ad esempio l'accuratezza, la completezza o l'attualità. Non esiste comunque un accordo su quali siano le dimensioni caratterizzanti la qualità dei dati, né sull'esatta definizione di quelle comunemente proposte. Ad esempio, la tempestività (in inglese "timeliness") è definita in termini di "disponibilità di dati in tempo", oppure di "grado di aggiornamento dei dati", quindi secondo modalità evidentemente differenti.

Lo standard ISO/IEC 25012:2008 [52] definisce un insieme di dimensioni specifiche per la caratterizzazione della qualità dei dati, in particolare: accuratezza, aggiornamento, completezza,



consistenza, credibilità, accessibilità, comprensibilità, conformità, efficienza, precisione, riservatezza, tracciabilità, disponibilità, portabilità e ripristinabilità. Come evidenziato, pur assumendo come riferimento lo standard ISO, le definizioni operative che comportano un risultato costituito dalla misura della dimensione possono essere molteplici. E' importante sottolineare che il processo di analisi della qualità dei dati è un processo continuo che coinvolge tutte le tappe del percorso dei dati nativi.

Bonifica. Il risultato dell'analisi della qualità del dato può portare a dover affrontare una prima fase di bonifica dello stesso. Infatti, i dati all'interno dei sistemi informativi o degli archivi di un'amministrazione sono spesso "sporchi" e non rispondenti ai requisiti di qualità prima menzionati.

Tale tappa risulta quindi fondamentale, sebbene potenzialmente onerosa e lunga, in quanto la mancata qualità può rendere inefficienti, o addirittura impraticabili, alcune operazioni di confronto, di similitudine e di aggregazione sui dati.

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, come mostrato in Figura 5 (per maggiori dettagli sulle licenze, si rimanda alla sezione 7 del presente documento). Sebbene sia sconsigliato restringere l'accesso ai dati o procedere con la pubblicazione di aggregazioni degli stessi, esistono tuttavia casi in cui i dati possono essere diffusi (pensiamo ad esempio ai redditi) ad un livello di aggregazione sufficiente a impedire di risalire al singolo dettaglio. A tal fine, in questa fase devono essere definite delle adeguate 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.

Contestualizzazione geografica. Un ulteriore elemento da tenere in considerazione, oltre a quelli previsti dalle precedenti fasi, è la contestualizzazione geografica del dato. Ogni volta che ciò sia possibile, è bene fornire i dati con una referenziazione geografica. Tale aspetto, oltre a costituire un fondamentale supporto per l'analisi geo-statistica quantitativa, rende possibile una quantità di letture sinottiche del territorio e incrementa sensibilmente il valore del dato. E' importante notare che tale aspetto deve trovare un adeguato riscontro nella metadattazione dopo descritta, secondo le direttive in materia.

Analisi di processo, (re)ingegnerizzazione dei processi organizzativi e produzione dei dati. Completate le fasi precedenti, un aspetto basilare del modello operativo per la linea dei dati nativi è costituito dall'analisi di processo e dalla (re)ingegnerizzazione dei processi organizzativi. Ogni dato ha infatti un proprio ciclo di vita, caratterizzato da uno specifico tasso di aggiornamento o manutenzione. E' quindi necessario analizzare il processo organizzativo che produce e gestisce il dato per individuare quali attori concorrono alla prima produzione del dato, distinguendo chi è responsabile e titolare dello stesso e chi invece aggiunge altri elementi informativi nel processo produttivo. Anche in fase manutentiva diversi soggetti possono contribuire alla gestione del dato, si pensi ad esempio a chi ne cura gli aspetti di dominio (per esempio un geometra di una direzione tecnica), a chi ne manutene gli aspetti di geo-referenziazione, a chi infine ne aggiorna i metadati.

Individuare tali passaggi organizzativi è fondamentale per intraprendere un corretto percorso di produzione di dati aperti. Infatti, quello che accade sovente nelle amministrazioni è che i dati sono sviluppati 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, tuttora, che elementi



conoscitivi importanti quali ad esempio la consistenza per specie del verde pubblico, il dispiegamento di alcune reti di servizi, gli stessi dati del patrimonio immobiliare, 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 difatti a replicare informazioni funzionali alla propria attività, ciascuno a suo modo, con un incremento del rumore di fondo attorno al patrimonio informativo dell'amministrazione.

Esempio: la gestione del patrimonio immobiliare di un Comune è affidato a un applicativo client server, installato localmente. Il funzionario titolare del servizio alimenta il database con estrema attenzione per i dettagli che riguardano la propria missione, ad esempio le “coordinate catastali”, annotando diligentemente ogni aggiornamento. Informazioni apparentemente non critiche in tale contesto, per esempio la denominazione della strada, vengono annotate in maniera mnemonica, senza curarne la indicizzazione o verificarne la coerenza con il toponimo del repertorio viario. Il dato così ottenuto, più che sufficiente per la specifica finalità (essendo l'immobile comunque univocamente identificato), non risulta integrabile con altre basi di conoscenza, laddove l'unità immobiliare debba essere invece indirizzata per via, civico interno ed esterno. Risulta così un valore sostanzialmente non pubblicabile poiché non utilizzabile al di fuori dello specifico contesto.

Alla luce di tali osservazioni il passaggio verso la risorsa federata (fase data hub interno), ove adeguatamente assistito da tecnici competenti e motivati, consente di superare progressivamente le suddette criticità. L'impegno politico e il relativo sostegno da parte dei livelli manageriali più alti costituisce comunque il prerequisito fondamentale senza il quale ogni sforzo è condannato ad essere vano.

Metadattazione. Il risultato delle precedenti tappe del modello operativo si traduce nella produzione dei metadati che, in buona sostanza, certificano le caratteristiche del dato. La metadattazione è una fase cruciale del percorso previsto per la produzione e il rilascio di dati di tipo aperto: 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.

A tal proposito, si raccomanda a tutte le figure tecniche e non individuate in sezione 5.1, di perseguire quanto indicato dalle presenti linee guida per quel che riguarda il modello e la pubblicazione dei metadati obbligatori e obbligatori condizionatamente da accompagnare ai dati aperti.

Data hub interno. La risorsa federata è rappresentata nel modello operativo proposto dal cosiddetto data hub interno. Esso è una piattaforma dove far confluire tutti i dati prodotti dai diversi dipartimenti della 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”).



E' importante notare che il data hub interno 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 (il livello 3 è il livello minimo richiesto a tutte le amministrazioni per rilascio di dati pubblici, secondo quanto stabilito dall'agenda nazionale per la valorizzazione del patrimonio informativo pubblico), ma per tutte le altre direttrici indicate.

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.

5.2.2. Linea 2: Dati Mashup

Concept. Oltre ai dati strettamente legati all'attività istituzionale, è necessario che la pubblica amministrazione operi in maniera da raccogliere stimoli e proposte dei cittadini, oltreché compiere una sua autonoma linea di produzione e sviluppo. Ogni nuovo dato, in questa linea, nasce da uno specifico "concept". Il "concept", in campo industriale, è la proposta necessaria a definire gli elementi fondamentali di un progetto e ne fornisce le basi per la realizzazione. 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.

Fonti, profili d'accesso, aggiornamento e "data visualization". 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 ed aggiornamento dei dati. Questo tipo di dati, nati a seguito di particolari esigenze o di determinati disegni strategici, sono in sostanza creati in funzione dell'esposizione al pubblico e del conseguente coinvolgimento. Proprio per questo si prestano a forme di engagement e visualizzazione ("data visualization") particolarmente innovative che spesso sono definite già a livello di "concept".

API e dataset. Il risultato ultimo di questa linea è la produzione di API e/o la pubblicazione di altri dataset.

5.2.3. Linea 3: Linked Open Data

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



Tuttavia, l'intenzione delle presenti linee guida e della strategia nazionale per la valorizzazione del patrimonio informativo pubblico definita dall'agenda, è quella di governare una transizione nel medio-lungo periodo verso la produzione nativa dei Linked Open Data.

Non a caso all'interno dell'agenda nazionale sono stati indicati una serie di dataset chiave da rilasciare, nel secondo semestre 2013, secondo le caratteristiche previste dai livelli 4 e 5 del modello per i dati di tipo aperto introdotto in sezione 4.1.

Alla luce di tali osservazioni, nel modello operativo proposto, 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 altresì rafforzata dal fatto che alcune delle fasi attraversate dalla linea dei dati nativi sono necessarie per avviare anche il percorso sulla linea dei Linked Open Data. In ogni caso, considerando che la metodologia da seguire per abilitare l'interoperabilità semantica attraverso i Linked Open Data è stata ampiamente descritta nelle linee guida della Commissione di Coordinamento SPC [9], si rimanda a tali linee guida per un approfondimento sul tema. In questa sede, è importante sottolineare che alcune fasi appartenenti alla linea dei dati nativi (i.e., censimento, analisi della qualità, bonifica e metadattazione) e alla linea dei Linked Open Data (i.e., **modellazione, ontologia, inferenza, interlinking, validazione e pubblicazione**) richiamano integralmente le sette fasi dell'approccio metodologico proposto in [9], che risulta quindi pienamente valido e adottabile per una produzione e rilascio di dati aperti di livello 4 e 5.

Nel contesto di tali linee guida, si faccia riferimento alla sezione 4.1.1 per una breve introduzione sui Linked Open Data, alla sezione 6 per gli standard, i formati e le ontologie di riferimento, e alle sezioni 9.1 e 9.4 per l'identificazione degli ulteriori requisiti da soddisfare per l'impiego di una soluzione Linked Open Data.

5.2.4. Linea 4: Engagement

Il processo di apertura dei dati è sicuramente più efficace se accompagnato da un'adeguata linea di azione dedicata al coinvolgimento ("engagement") degli "stakeholder". Come è naturale, l'engagement può essere di tipo interno e di tipo esterno.

Il coinvolgimento interno all'amministrazione può essere affrontato in diverse forme, ad esempio con circolari, lettere interne, momenti di incontro; tuttavia si ritiene che la forma più efficace possa essere quella di far sì che la cultura dei dati aperti si diffonda all'interno dell'amministrazione in modo naturale attraverso un meccanismo virtuoso di emulazione e di passa-parola basato su fatti concreti. Non c'è circolare interna che possa avere la stessa efficacia di cittadino che richieda l'apertura di un nuovo dato per la propria utilità, o di un ufficio che acceda direttamente a un dato aperto senza richiederlo internamente.

Diverso invece è il coinvolgimento esterno. I fronti esterni sono infatti numerosi, e sono vaste le possibilità di attrarre "stakeholder" esterni; pertanto, è opportuno pianificare attentamente una adeguata strategia di "engagement".

Strategia. Innanzitutto occorre identificare i soggetti che si vogliono coinvolgere nelle attività relative all'Open Data. Esistono alcune categorie "speciali" che, alla luce dell'esperienza maturata da alcune



amministrazioni, si può affermare con certezza siano in prima linea nell'utilizzo degli Open data. Si tratta degli studenti universitari, interessati sia in termini di lavori di ricerca e di analisi statistica o economica, sia per lo sviluppo di “apps” basate sugli Open Data, o ancora, si tratta delle startup, o di imprese di qualunque tipo, che sfruttano i dati per aprire nuove attività (si pensi ad un'agenzia immobiliare che vuole capire dove aprire una nuova sede sulla base dei dati delle quotazioni immobiliari). Si tratta, infine, di quei giornalisti che intravedano negli Open Data un valore anche per la propria attività (“data journalism”).

Social. Individuati i soggetti, è necessario definire una forma di engagement. Quella più semplice, e a costi pressoché nulli, è rappresentata dalla comunicazione, anche interattiva, attraverso l'uso dei social network.

Scenari d'uso. Sebbene sia complesso prevedere gli usi che derivano dalla pubblicazione di dati aperti, è utile affiancare agli strumenti suddetti la definizione di scenari d'uso degli Open Data da parte dei cittadini, startup/imprese e altre amministrazioni; ciò al fine di comprendere meglio gli utilizzatori esterni e le loro possibili azioni sui dati. Utilizzando strumenti di “data visualization” (e.g., visualizzazioni su mappa) di dati aperti e sovrapponendo i vari strati informativi, si individuano diversi incroci di dati che possono essere “offerti” all'esterno, su cui poter poi instaurare un dialogo con diversi stakeholder.

Eventi, Hackaton, App Showcase. Altre forme di engagement che un'amministrazione può adottare sono l'organizzazione di eventi specifici per promuovere alcune tipologie di dataset, o per analizzare casi d'uso. Tale attività risulta più strutturata rispetto all'organizzazione dei cosiddetti Hackathon, che tipicamente sono intesi come eventi che richiedono l'istituzione di un meccanismo regolamentato di engagement e di valutazione delle proposte da parte di giurie predefinite.

L'organizzazione di tali eventi può poi portare alla produzione di nuove “apps” basate sul data hub interno mantenuto dall'amministrazione, così come a nuove forme di pubblicazione di dati.

Infine, l'“App Showcase”, ovvero una vetrina delle “apps” basate sugli Open Data, è una forma ulteriore di “engagement” verso l'esterno che può essere adottata per veicolare e concentrare ulteriore interesse da parte di cittadini, startup e altri possibili fruitori di dati aperti.



6. STANDARD TECNICI E ONTOLOGIE

In questa sezione sono illustrati gli standard, i formati tecnici e le ontologie di riferimento per implementare i modelli di dati e metadati definiti nella strategia dell'agenda nazionale e nelle precedenti sezioni del presente documento.

6.1. Standard di base

Gli standard necessari all'abilitazione dei livelli 4 e 5 del modello dei dati e dei 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. Ogni standard è introdotto brevemente in questa sezione, unitamente agli standard utilizzati per la rappresentazione e lo scambio di dati statistici.

RDF (Resource Description Framework) [12]. RDF nasce come strumento per rappresentare i metadati delle risorse Web; esso poi si è affermato come modello di rappresentazione flessibile delle informazioni, conservando però la possibilità di definire vincoli strutturali fra i dati. Per questo motivo RDF è diventato popolare come modello per i dati aperti pubblicati su Web e si è affermato come uno degli standard W3C alla base del Web Semantico. Esso consente di catturare la semantica dei dati, quindi la loro comprensibilità, e di facilitarne 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, può essere considerata una risorsa. Associando a ogni risorsa un identificativo univoco, nello specifico un URI, si possono poi rappresentare su Web le informazioni relative a tali risorse, rendendole quindi accessibili a tutti.

Tecnicamente, RDF è un framework concettuale che consente di identificare le risorse con la modalità appena citata consentendo inoltre di descrivere le risorse mediante il suo costrutto informativo di base, la cosiddetta tripla $\langle \text{soggetto} \rangle \langle \text{predicato} \rangle \langle \text{oggetto} \rangle$. Un *soggetto* è sempre una risorsa (i.e., il suo URI), un *oggetto* è una risorsa o un valore (cioè un'espressione puramente simbolica come un numero, una stringa, ecc.), un *predicato* è una relazione tipizzata tra due risorse o una proprietà di una risorsa. Si noti che anche i predicati sono rappresentati con un URI. In questo modo, oltre che identificare univocamente risorse e predicati, si possono anche descrivere con dati e metadati mettendoli in relazione con predicati significativi.

Le triple RDF sono strutture ricorsive, analogamente alle frasi del linguaggio naturale, Soggetto-Verbo-Oggetto. La concatenazione di triple genera un “grafo RDF”; un dataset RDF è un quindi grafo. Lo spazio Web in cui questi dati sono localizzati è il cosiddetto Web dei Dati (“Web of Data”), mentre la sua prospettiva più semantica viene chiamata Web Semantico.

RDF è un framework concettuale e può essere implementato tramite diverse forme sintattiche, quali RDF/XML, Notation3, N-Triple e Turtle (si veda la sezione 6.2). La scelta tra queste diverse soluzioni sintattiche, anche dette serializzazioni di RDF, deve essere compiuta sulla base di requisiti richiesti quali



compattezza, spazio fisico utilizzato, leggibilità, ecc. Le serializzazioni sono comunque fra loro intertraducibili.

RDFS (RDF Schema) [13]. RDF ha un'estensione, chiamata RDF Schema (RDFS), che permette di definire semplici schemi per i dati. RDFS introduce alcuni costrutti come le classi (`rdfs:Class`), le collezioni (ad esempio, `rdfs:List`) e una serie di proprietà per poter definire tassonomie 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 come ad esempio i vincoli di dominio e di codominio. Anche gli schemi definiti con RDFS, oltre a quelli spiegati nella sezione 6.3.1, sono comunemente chiamati ontologie.

OWL (Web Ontology Language) [14]. 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 di fatto la comprensione e il riuso di schemi e metadati, e di conseguenza abilita 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 diventa possibile l'uso di cosiddetti ragionatori automatici per le logiche descrittive, che consentono di inferire, sui dati conformi all'ontologia, nuove triple e quindi informazione addizionale. Ad esempio, si possono generare triple inverse, triple simmetriche e triple da predicati transitivi.

SPARQL (Sparql Protocol And Rdf Query Language) [15]. 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 query SPARQL cerca tra i dati le concatenazioni di triple "conformi" a quelle della query, assegnando (i.e., istanziando) degli URI o dei valori alle variabili che possono anche essere restituiti in output. Inoltre, in un'interrogazione è possibile specificare operatori più complessi come disgiunzioni, triple opzionali, filtri, espressioni regolari, ecc.

SPARQL è anche un protocollo che consente di convogliare via Web (appoggiandosi al protocollo HTTP) le interrogazioni verso servizi Web detti SPARQL endpoint.

SDMX (Statistical Data and Metadata eXchange) [16]. E' 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, e dunque 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 associate code-list (o classificazioni). Si nota che sebbene SDMX sia nato come modello per lo scambio di dati, esso viene anche usato per la rappresentazione dei dati.



6.2. Formati di tipo aperto

Lo scopo di questa sezione è di elencare e descrivere brevemente i principali formati di tipo aperto utilizzabili per i dati e per i documenti.

6.2.1. Formati per i dati di tipo aperto

Per i dati di tipo aperto sono ampiamente utilizzati i seguenti formati:

XML (eXtensible Markup Language) [53]. XML è un linguaggio di marcatura standardizzato dal W3C [53] usato per l'annotazione di documenti e per la costruzione di altri linguaggi per l'annotazione di documenti. Esistono strumenti che possono essere utilizzati per definire la struttura dell'XML come DTD [93] XML schema [91] [92]. Il mondo legato all'XML è enormemente ampio e la sua trattazione non rientra tra gli obiettivi del presente documento. In questo contesto, ci si limita a far presente che una delle serializzazioni (formati fisici) di RDF fa uso di documenti e sintassi XML [54].

Notation3 e Turtle. Notation3 (o N3) [55] è una serializzazione testuale di RDF pensata per essere più compatta rispetto a quella ottenuta utilizzando la sintassi XML. Essa risulta più leggibile da parte degli umani e possiede delle caratteristiche che vanno anche oltre l'uso di RDF (e.g., rappresentazione di formule logiche).

Turtle (anche noto come Terse RDF Triple Language) [56] è una versione semplificata (un sottoinsieme di funzionalità) di N3 che include solo aspetti relativi a RDF.

JSON (JavaScript Object Notation) [57]. E' un formato aperto per lo scambio di dati che siano leggibili anche dagli umani. Nasce dalla rappresentazione di strutture dati semplici nel linguaggio di programmazione JavaScript, ma rimane comunque indipendente dai linguaggi di programmazione. Ha avuto ampia diffusione per il fatto di essere flessibile e compatto. Esistono delle serializzazioni di RDF con JSON, ma nessuna di queste è ancora stata standardizzata.

N-Triples [58]. N-Triples è un formato testuale di serializzazione per RDF in cui ogni tripla è espressa interamente (i.e., non in modo sintetico) e indipendentemente dalle altre.

CSV (Comma Separated Values). E' un formato di file testuale usato 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, in genere la virgola. Il CSV non è uno standard vero e proprio ma la sua modalità d'uso è descritta nell'RFC 4180 [17].



Shapefile [59]. E' il formato standard de-facto per la rappresentazione dei dati dei sistemi informativi geografici (GIS). Esso è creato dalla società ESRI che rende pubbliche le sue specifiche. Questo ha consentito lo sviluppo di strumenti in grado di gestire e creare tale formato. I dati sono di tipo vettoriale.

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. Molto spesso, a questi tre si accompagna anche un file *.prj* che contiene le impostazioni del sistema di riferimento.

6.2.2. Formati per i documenti

Per quanto riguarda i documenti (e.g., delibere, atti amministrativi) da rilasciare, si consiglia di considerare i seguenti formati aperti. Si noti che se devono essere pubblicati dati solo in forma tabellare senza elementi aggiuntivi (e.g., grafici, formule) si può utilizzare il formato CSV prima descritto (come ad esempio nel caso dei tassi di assenza del personale delle pubbliche amministrazioni). Negli altri casi di dati in forma tabellare con elementi aggiuntivi, si raccomanda l'uso di ODS.

ODT (Open Document Text). E' uno standard aperto per documenti testuali basato su XML. Fa parte dello standard OASIS Open Document Format for Office Applications [18]. E' stato adottato come formato principale per i testi in alcune suite per l'automazione d'ufficio come OpenOffice.org e LibreOffice ed è supportato da altre come Microsoft Office, Google Drive e IBM Lotus.

ODS (Open Document Spreadsheet). E' uno standard aperto per fogli di calcolo basato su XML. Fa parte dello standard OASIS Open Document Format for Office Applications [18]. 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 ed è supportato da altre come Microsoft Office, Google Drive e IBM Lotus.

ODP (Open Document Presentation). E' uno standard aperto per documenti di presentazione basato su XML. Fa parte dello standard OASIS Open Document Format for Office Applications [18]. E' stato adottato come formato principale per i documenti di presentazione in alcune suite per l'automazione d'ufficio come OpenOffice.org e LibreOffice ed è supportato da altre come Microsoft Office, Google Drive e IBM Lotus.

PDF (Portable Document Format). E' 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). E' stato successivamente 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.).



6.3. Ontologie e vocabolari

Questa sezione presenta una prima selezione di ontologie e vocabolari che possono essere utilizzati per descrivere alcune tipologie di dati di tipo aperto (una lista piuttosto ricca di ontologie e vocabolari è comunque messa a disposizione dall'OKFN¹¹). L'uso consolidato di tali ontologie e vocabolari ha l'obiettivo di far sì che, all'interno della pubblica amministrazione, dataset differenti ma di una stessa tipologia abbiano uno schema comune. In questo caso l'integrazione di dati e il riuso di software risultano enormemente facilitati in quanto, ad esempio, le stesse interrogazioni possono essere effettuate su dataset differenti (quindi potenzialmente è possibile usare uno stesso software).

Seguendo quanto indicato nell'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico, possiamo distinguere tra dati *trasversali* ovvero comuni a tutte le amministrazioni, centrali e locali, (e.g., dati amministrativi) e dati *verticali* gestiti da o per conto di specifiche amministrazioni riferibili a determinati settori di competenza (e.g., dati sanitari, dati sulla cultura). Alla luce di tale distinzione, le seguenti sotto sezioni si propongono di individuare alcune ontologie e vocabolari utilizzabili rispettivamente per la descrizione di dati comuni a tutte le amministrazioni e per la descrizione di domini di interesse specifici che sono identificati per macro aree tematiche.

6.3.1. Ontologie e vocabolari per dati trasversali

Esistono ontologie e vocabolari che riguardano dati di interesse generale e che quindi possono essere utilizzati in diversi contesti e da amministrazioni di diverso tipo. Di seguito si descrivono brevemente tali ontologie e vocabolari, scelti sulla base della loro diffusione, del livello di standardizzazione e del contesto in cui si sono sviluppati (come nel caso dei vocabolari “core” [88] definiti nell'ambito di specifici programmi della Commissione Europea nati anche al fine di implementare determinate azioni previste dall'Agenda Digitale Europea).

SKOS (Simple Knowledge Organization System) [19]. E' un'ontologia che consente di rappresentare vocabolari controllati, tassonomie e tesauri. E' nato per facilitare la pubblicazione e l'uso di vocabolari nel contesto dei Linked Data ed è stato standardizzato dal W3C.

GOODRELATIONS [60]. E' un'ontologia che consente di rappresentare dati relativi al commercio quali prodotti, prezzi, negozi, ecc. I dati rappresentati con Goodrelations possono anche essere incorporati all'interno delle pagine Web per essere letti dai classici motori di ricerca e da altri software. Tale ontologia può essere utilizzata per descrivere i servizi acquisiti tramite contratti dalle pubbliche amministrazioni (un esempio di utilizzo è presentato in appendice).

PUBLIC CONTRACTS ONTOLOGY [61]. E' un'ontologia che consente di descrivere dati RDF relativi ai contratti pubblici (e.g., contraente, data di validità) e all'intero processo di aggiudicazione di una gara d'appalto (e.g., tipologia di gara, massimale, partecipanti, aggiudicatari, ecc.).

¹¹ <http://lov.okfn.org/dataset/lov/>



PAYMENT ONTOLOGY [62]. E' un'ontologia che consente di rappresentare dati relativi ai pagamenti e alle informazioni di spesa.

FOAF (Friend of a Friend) [63]. E' un'ontologia che consente di rappresentare dati riguardanti le persone e le loro relazioni. Esso può quindi essere utilizzato per la rappresentazione di alcuni dati della basi di dati di interesse nazionale "Anagrafe della Popolazione Residente – ANPR".

WAI (Who Am I) [64]. E' un'ontologia che estende FOAF con i concetti di ruolo e profilo che possono essere associati alle persone.

SIOC (Semantically Interlinked Online Communities) [65]. E' un'ontologia che consente di rappresentare informazioni riguardanti le comunità online (ad esempio, blog, wiki, forum, mailing list, ecc.) e di relazionarle tra loro.

CPSV (Core Public Service Vocabulary) [20]. E' un modello dati per la rappresentazione dei servizi pubblici offerti dalle pubbliche amministrazioni. Il suo scopo è quello di facilitare (i) lo scambio di informazioni sui servizi nel settore pubblico, (ii) la scoperta di servizi esistenti e delle normative che li riguardano, e (iii) lo sviluppo dei portali della pubblica amministrazione. Inoltre, grazie all'uso del CPSV è possibile confrontare servizi simili erogati da differenti organizzazioni. Di tale vocabolario esiste la versione per i Linked Data. Infine, è importante sottolineare che esso fa parte dei vocabolari "core" della commissione europea standardizzati o in fase di standardizzazione da parte del W3C, recentemente utilizzato in attività pilota nel contesto del progetto europeo SPOCS - Building the next generation Points of Single Contact [21].

CLV (Core Location Vocabulary) [66]. E' un modello dati per la rappresentazione delle locazioni. Esso quindi consente di descrivere indirizzi, struttura geometrica, nome geografico, ecc. E' conforme alle specifiche INSPIRE [22] ed esiste una versione ufficiale che può essere utilizzata con i Linked Data. Anch'esso, come nel caso precedente, fa parte dei vocabolari "core" della commissione europea standardizzati o in fase di standardizzazione da parte del W3C.

CPV (Core Person Vocabulary) [67]. E' un modello dati per la rappresentazione delle informazioni sulle persone. Nasce per essere ben integrato agli altri vocabolari "core" della commissione europea e fa largo riuso di schema.org e FOAF. Come nel caso di FOAF, tale vocabolario può essere utilizzato per rappresentare certe informazioni presenti nella base di dati di interesse nazionale "Anagrafe della Popolazione Residente – ANPR". Attualmente il vocabolario è utilizzato all'interno del progetto europeo e-CODEX – eJustice Communication via Online Data eXchange [51].

ORG (Organization Ontology) [68]. E' un'ontologia che consente di rappresentare strutture organizzative, in particolare di pubbliche amministrazioni. Essendo rivolta al mondo dei Linked Data, ben si presta a estensioni e integrazioni con altre ontologie in modo da considerare classificazioni delle



organizzazioni, l'appartenenza del personale con i relativi ruoli. Fa parte dei vocabolari “core” della commissione europea standardizzati o in fase di standardizzazione da parte del W3C ed è utilizzata nella versione Linked Open Data dell'Indice nazionale della Pubblica Amministrazione italiana (IPA).

RegOrg [69]. E' un vocabolario, in corso di standardizzazione W3C, che estende la precedente ontologia ORG per la specifica delle organizzazioni che sono ufficialmente riconosciute attraverso l'iscrizione a registri nazionali o regionali. Ad esempio, nel caso italiano potrebbe agevolmente essere usato per rappresentare la base dati di interesse nazionale “Registro delle imprese” (art. 60 del CAD).

OWL-S (Semantic Markup for Web Service) [70]. E' un'ontologia che consente di descrivere i servizi Web e il loro ciclo di vita. Questo consente di trovare, invocare e comporre servizi Web in modo automatico attraverso il Web Semantico. Si compone di tre parti principali: *profile* per la profilazione base e il discovery dei servizi, il *process model* per la descrizione delle operazioni del servizio e, infine, il *grounding* per le informazioni di interoperabilità e scambio di messaggi con altri servizi. Si noti che molte delle specifiche di OWL-S corrispondono a quelle del linguaggio WSDL.

DOAP (Description Of A Project) [71]. E' un'ontologia che consente di descrivere progetti. Nata principalmente per descrivere progetti di produzione del software, esiste anche una versione pronta per essere utilizzata in ambito Linked Data.

VCARD ONTOLOGY [72]. E' un'ontologia che nasce dal rispettivo standard IETF¹² (RFC 2426) per rappresentare i biglietti da visita elettronici. Consente quindi di rappresentare informazioni su persone e le loro informazioni di affiliazione.

EUROVOC [73]. E' un tesoro multilingua (22 lingue) contenente la terminologia usata nell'ambito delle attività dell'unione europea ovvero il parlamento europeo, gli uffici delle pubblicazioni ufficiali dell'unione europea, ecc. Eurovoc è reso disponibile anche con una versione ufficiale in RDF che fa uso di SKOS¹³. Il download di Eurovoc è regolato da accesso tramite credenziali.

6.3.2. Ontologie e vocabolari per dati verticali

Le seguenti ontologie e vocabolari sono riferibili ad ambiti specifici. Tali ambiti sono stati suddivisi per macro-aree tematiche scelte, in questa prima versione delle linee guida, principalmente sulla base dei settori di competenza delle amministrazioni coinvolte nel rilascio di dataset chiave individuati nell'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico. In aggiunta a tali aree, il gruppo di lavoro ha deciso di avviare un'analisi del settore dei dati territoriali e ambientali in quanto settore di elevato interesse, soprattutto in ottica Open Data.

¹² <http://www.ietf.org/rfc/rfc2426.txt>

¹³ <http://eurovoc.europa.eu/drupal/?q=it/legalnoticeredirect>



6.3.2.1. Dati Territoriali e Ambientali

INSPIRE (Infrastructure for Spatial Information in Europe) [22]. Nasce come direttiva della commissione europea per mettere a disposizione in modo armonico i dati geo-spaziali e i servizi associati tramite geo-portali e altri punti di accesso. Di fatto, definisce un'ontologia nata proprio con l'obiettivo di migliorare l'interoperabilità tra i sistemi geo-spaziali.

AGROVOC [74]. E' un vocabolario controllato creato dalla FAO per rappresentare i concetti delle sue aree di interesse (cibo, nutrizione, agricoltura, ambiente, ecc.). Esiste una versione SKOS già pronta per essere utilizzata in ambito Linked Data.

6.3.2.2. Dati Statistici

RDF Data Cube [23]. L'esigenza di rappresentare i dati multidimensionali sul Web ha portato alla proposta da parte del W3C di RDF Data Cube Vocabulary (RDF QB) [23]. RDF QB prevede un modello multidimensionale e consente dunque di rappresentare "cubi" (data cube) in termini di misure e dimensioni. Il modello dati di RDF QB è basato su SDMX, ma sfrutta le caratteristiche di RDF, consentendo, dunque, di introdurre i dati multidimensionali nel Web semantico ed eventualmente di collegarli anche a dati elementari.

XKOS (eXtended Knowledge Organization System) [75]. E' un'estensione di SKOS promossa dall'UNECE, dall'Eurostat e dall'OECD per gestire i requisiti specifici per la pubblicazione, da parte degli Istituti Nazionali di Statistica, dei dati della produzione di statistica ufficiale. Ad esempio XKOS introduce i "livelli" delle classificazioni statistiche, delle classificazioni statistiche; tali livelli, identificando i livelli di aggregazione nelle statistiche pubblicate, sono fondamentali sia per la consistenza dei dati prodotti sia per controllare i rischi di violazione della riservatezza.

6.3.2.3. Dati culturali

Tesauro PICO [76]. E' il vocabolario controllato utilizzato nel portale della cultura italiana¹⁴ ma è sufficientemente generale per essere usato nella descrizione di dati relativi al dominio dei dati culturali.

Tesauro UNESCO [77]. E' un tesauro per i documenti e le pubblicazioni sulle attività dell'UNESCO. In particolare contiene i termini usati negli ambiti dell'istruzione, cultura, scienze naturali, scienze sociali, comunicazione e informazione. Ne esiste una versione modellata attraverso SKOS¹⁵.

CIDOC [78]. E' un modello dati di riferimento per rappresentare le informazioni su concetti e

¹⁴ <http://www.culturaitalia.it/>

¹⁵ <http://skos.um.es/unescothes/>



relazioni nell'ambito della cultura (e.g., documenti di musei, archivi, biblioteche). Esiste già una versione in OWL¹⁶.

EDM (Europeana Data Model) [79]. E' il modello dati (utilizzabile anche per dati in RDF) utilizzato all'interno del progetto europeo Europeana che ha lo scopo di favorire la condivisione, l'interoperabilità e l'accesso alle informazioni delle opere culturali.

6.3.3. Ontologie e vocabolari per i metadati

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

VoID (Vocabulary of Interlinked Datasets) [80]. E' un vocabolario che consente di rappresentare metadati sui dataset RDF, in particolare, per caratterizzare le informazioni riguardanti le connessioni tra diversi dataset, come ad esempio i linkset.

DCAT (Data Catalog) [24]. E' un vocabolario utilizzato per metadattare i cataloghi di dati, i dataset contenuti in un catalogo e le diverse modalità di reperimento e accesso ai dati. DCAT è uno standard del W3C ed è attualmente usato per caratterizzare alcuni profili applicativi.

Dublin Core [81]. La versione "core" consiste di un vocabolario minimo di metadati (15 elementi) per descrivere generiche risorse. Tra questi vi sono, ad esempio, autore, data e titolo. Tale insieme minimo è stato esteso con un insieme più ampio di elementi definiti dalla iniziativa DCMI (Dublin Core Metadata Initiative) [82].

ADMS (Asset Description Metadata Schema) [39]. E' un modello utilizzato per rappresentare asset di interoperabilità semantica. Possibili asset sono ad esempio le classificazioni, i modelli, gli standard, ecc. che possono essere utilizzati per abilitare l'interoperabilità semantica. Questa ontologia ha l'obiettivo di favorire la scoperta e la federazione di diversi "repository" di asset semantici.

PROV [83]. E' un modello di rappresentazione della provenienza di artefatti, recentemente standardizzato dal W3C. Consente di modellare anche il processo di generazione di un artefatto in maniera quasi analoga ai sistemi di controllo versione. Le specifiche dello standard sono definite attraverso quattro diversi componenti: PROV-DM (il modello dei dati), PROV-N (notazione per uso da parte di esseri umani), PROV-O (l'ontologia OWL) e PROV-CONSTRAINTS (i vincoli della modellazione). Per rappresentare i metadati della provenienza, in particolare nel caso di Linked Data, si consiglia l'uso di PROV-O.

¹⁶ <http://erlangen-crm.org/>.



CREATIVE COMMONS [37]. E' un sistema di licenze per le opere soggette a diritto d'autore. E' possibile utilizzare una versione RDF del vocabolario soprattutto quando si gestiscono dati di tipo Linked.

SCHEMA.ORG [84]. E' un insieme di vocabolari che contengono i metadati che possono essere usati all'interno delle pagine Web per strutturare meglio le informazioni presenti nelle pagine. Questi servono per facilitare la comprensione degli elementi delle pagine da parte dei "crawler" dei motori di ricerca.

VANN (Vocabulary for ANNotating vocabulary descriptions) [85]. E' un vocabolario che consente di rappresentare metadati su vocabolari. In particolare, esso è utilizzato per arricchirli con esempi e note di utilizzo.

6.4. Alcuni dataset di riferimento

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

6.4.1. Dati trasversali

DBPedia [25]. E' un grande dataset in formato Linked Data contenente informazioni estratte da Wikipedia. Essendo Wikipedia un enorme e variegato contenitore informativo, DBPedia è il dataset più collegato del Web dei Dati e per questo considerato il suo hub. Oltre alla presenza di una grossa mole di informazioni, DBPedia contiene collegamenti in uscita ad altri dataset esistenti.

DBPedia Italia [26]. E' la controparte italiana di DBPedia contenente i dati estratti da Wikipedia Italia.

SPCData [27]. E' lo spazio dati del Sistema Pubblico di Connettività e Cooperazione. E' composto da diversi dataset in formato Linked Data come quello dell'IPA (Indice della Pubblica Amministrazione) e quello dei dati sui contratti SPC relativi ai servizi Web. La presenza al suo interno della base di dati ufficiale prevista dal CAD (art. 57-bis) per l'identificazione univoca delle pubbliche amministrazioni italiane, candida SPCData a diventare l'hub dei dati di tipo Linked della pubblica amministrazione italiana. Attualmente, SPCData è collegato ad altri Linked Data di alcune pubbliche amministrazioni come CNR, Comune di Firenze e Regione Piemonte.



6.4.2. Dati territoriali e ambientali

GEONAMES [28]. E' un database che contiene informazioni geografiche provenienti da ogni parte del mondo. E' accessibile sia tramite Web services, sia tramite SPARQL e per questo è referenziato da tanti altri dataset che intendono geolocalizzare le proprie informazioni.

RNDT [29]. E' una base dati di interesse nazionale che costituisce il catalogo nazionale dei metadati territoriali riguardanti i dati territoriali e i servizi ad essi relativi disponibili presso le Pubbliche Amministrazioni. E' pienamente conforme alla direttiva INSPIRE.

SINAnet [30]. E' la rete del Sistema Informativo Ambientale Nazionale (SINA) che raccoglie dati e informazioni necessari a descrivere e comprendere i fenomeni ambientali, al fine di fornire supporto dell'azione di governo e di produrre con continuità prodotti e servizi informativi basati su indicatori e indici.

6.4.3. Dati sanitari

MeSH (Medical Subject Headings) [86]. E' un dataset, strutturato come vocabolario controllato, che tiene traccia della letteratura scientifica biomedica. Non nasce per il contesto del Web Semantico, ma diversi sforzi hanno prodotto degli adattamenti di MeSH in Linked Data. Una di queste versioni può essere reperita all'indirizzo <http://bioportal.bioontology.org/ontologies/3019>.

SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) [87]. E' un dataset, strutturato come vocabolario controllato, di termini medici e clinici in diverse lingue. E' mantenuto dall'associazione no-profit IHTSDO. Esistono varie conversioni dell'ontologia per il Web Semantico; tuttavia nessuna di queste è attualmente ufficiale.



7. LICENZE ASSOCIATE AL DATO

Un passo importante del modello operativo introdotto dalle presenti linee guida riguarda la scelta della licenza da associare al dato. Tale scelta risulta cruciale in quanto i termini stabiliti dalla licenza definiscono l'uso che gli utenti possono effettivamente fare dei dati, la paternità dei dati e le modalità di creazione di eventuali lavori derivati. Essa rientra tra i metadati obbligatori minimi da fornire.

In linea generale, le licenze già descritte nelle precedenti linee guida sull'interoperabilità semantica attraverso Linked Open Data della Commissione di Coordinamento SPC [9] sono valide anche ai fini del presente documento.

A tali licenze si aggiunge, nel contesto di questo documento, la licenza **ODbL (Open Data Commons Open Database License)** [32], una licenza che si applica nello specifico ai dati e che è adottata dal progetto OpenStreetMap [31] e dal Comune di Parigi. Essa consente di condividere, creare e modificare i dati ma obbliga a preservare la paternità (attribuzione), a condividere allo stesso modo e a mantenere sempre nei possibili lavori derivati almeno una versione aperta dei dati.

Le licenze della famiglia Creative Commons, quelle italiane ed europee, e la licenza ODbL sono comparate sulla base di alcuni parametri che si ritengono importanti per guidare la giusta scelta della licenza stessa. Tali parametri sono quelli già identificati nel contesto delle linee guida della Commissione di Coordinamento SPC (per una descrizione di tali parametri più dettagliata ci si può riferire alla sezione 8.1 di [9]) e sono riportati in Tabella 4.

	CC0	CC-BY	CC-BY-SA	CC-BY-SA-NC	CC-BY-ND	IODL 1.0	IODL 2.0	ODbL	ISA Open Metadata 1.1
Portabilità inter-lingua	X	X	X	X	X			X	
Riconoscibilità internazionale del logo	X	X	X	X	X			X	
Uso per fini commerciali	X	X	X				X	X	X
Uso gratuito del lavoro soggetto alla licenza	X	X	X	X	X	X	X	X	X
Possibilità di modificare i dati	X	X	X	X		X	X	X	X
Possibilità di preservare la paternità		X	X	X	X	X	X	X	X
Compatibilità inter-licenza per lavori derivati	X	X					X		

Tabella 4: Comparazione tra licenze



Facendo riferimento alla definizione Open Data fornita dall’Open Knowledge Foundation (OKF) per cui un dato aperto è un dato riutilizzabile da chiunque anche per finalità commerciali, “*soggetto al massimo alla richiesta di attribuzione e condivisione allo stesso modo*”, le precedenti licenze sono classificate in Figura 6¹⁷ evidenziando quali di loro supportano tali caratteristiche.

La figura mostra quindi come le licenze che non consentono lavori derivati non rientrano tra gli strumenti legali per l’Open Data; inoltre, si nota la chiara distinzione tra le licenze suddette e la CC0. La CC0 infatti non è una vera e propria licenza; piuttosto, essa rappresenta uno strumento legale utilizzato per attribuire un’opera al pubblico dominio, attraverso il quale il dichiarante “*apertamente, pienamente, permanentemente, irrevocabilmente ed incondizionatamente rinuncia, abbandona e cede ogni proprio Diritto d’autore e Connesso, ogni relativa pretesa, rivendicazione, causa ed azione, sia al momento nota o ignota (inclusando espressamente le pretese presenti come quelle future) relativa all’Opera*” (i.e., il cosiddetto “waiver”).

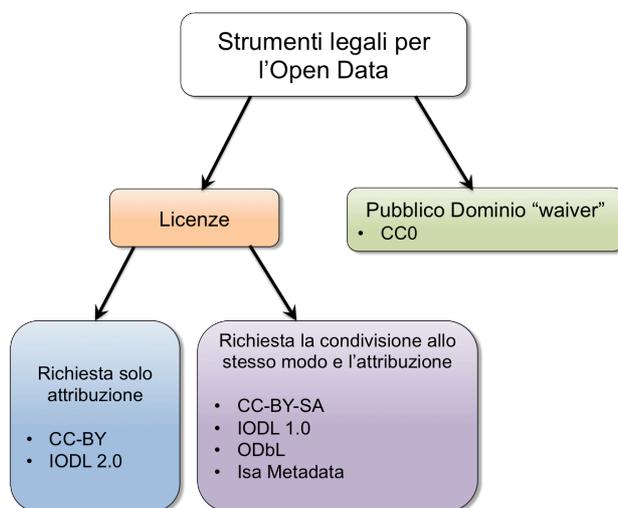


Figura 6: Strumenti legali per l'Open Data

E’ importante sottolineare che nel recente Dgls n. 33/2013 all’art. 7 si stabilisce che i dati pubblici che le amministrazioni devono rendere disponibili online nella sezione “Amministrazione Trasparente” dei siti istituzionali devono essere in formato di tipo aperto e riutilizzabili “*senza ulteriori restrizioni diverse dall’obbligo di citare la fonte e di rispettarne l’integrità*”. Tale disposizione sembra quindi indicare una tipologia di licenza che, sulla base dei parametri indicati in Tabella 4 consente di preservare la paternità (attribuzione) e non consente di modificare i dati (no lavori derivati).

¹⁷ La Figura 6 è una rivisitazione anche grafica dello schema proposto in [40].



In tale contesto, il principio generale che anche questo gruppo di lavoro ritiene di sostenere è quello di associare ai dati pubblici una licenza aperta (secondo quanto indicato in Figura 6) che consenta di rispettare i requisiti di (i) interoperabilità, anche transfrontaliera, (non limitandosi ai soli confini nazionali) e, per quanto possibile, (ii) di massimo riutilizzo dei dati. Quest'ultimo principio è espressamente indicato anche nella nuova direttiva PSI all'art. 8.

Le amministrazioni possono prevedere casi di applicazione di licenze che limitino il riutilizzo dei dati se e solo se è necessario il rispetto di altre norme (e.g., norme sul diritto d'autore). Tale scelta dovrà comunque essere opportunamente motivata dall'amministrazione.

In ogni caso il rilascio dei dati con licenze che ne consentono il massimo riuso non modifica la titolarità dei dati stessi. Conseguentemente l'amministrazione titolare resta comunque responsabile del contenuto informativo originario del dato reso disponibile e del suo aggiornamento.

Infine, è importante ribadire che, ai sensi dell'art. 9 del DL n.179/2012 la mancata indicazione di una licenza associata ai dati già pubblicati implica che gli stessi si ritengano di tipo aperto secondo le caratteristiche principali sancite dall'art. 68 del CAD, già richiamato nell'introduzione delle presenti linee guida (principio dell'Open Data by default).

Secondo quanto anche definito dall'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico, nei casi di mancata indicazione della licenza, i dati si ritengono rilasciati secondo i termini stabiliti dalla licenza CC-BY (attribuzione), ossia con il solo obbligo di citare la fonte. L'attribuzione della fonte può essere fatta in maniera semplice indicando il nome dell'organizzazione unitamente all'URL della pagina Web dove si trovano i dati/contenuti da licenziare. In generale, si consiglia l'uso della CC-BY nella sua versione 2.5.



8. ASPETTI DI COSTO DEL DATO

Come già ribadito in Sezione 3.1, la direttiva europea sulla valorizzazione del patrimonio informativo pubblico [3] emessa nel 2003 non contemplava il paradigma dell'Open Data e tutti gli scenari da esso sviluppati. Essa non ammetteva che soggetti privati potessero beneficiare di profitti attraverso la creazione di servizi derivanti dai dati pubblici.

I recenti sviluppi relativi all'Open Data hanno quasi imposto la revisione di tale normativa in modo da tenere in considerazione i processi e i meccanismi di apertura e utilizzo di dati pubblici. Uno degli aspetti su cui la nuova direttiva PSI è stata quasi radicalmente aggiornata è quello relativo ai costi di rilascio dei dati pubblici.

L'art. 9 del DL n.179/2012 è conforme alle indicazioni della nuova direttiva PSI in quanto afferma che i dati di tipo aperto *“sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione. L'Agenzia per l'Italia digitale può stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali.”*

8.1. Modalità di rilascio dei dati

Esistono tre modalità di rilascio dei dati dal punto di vista dei costi ad essi associati: uso a titolo gratuito, rilascio con tariffa non superiore ai costi marginali e rilascio con tariffa superiore ai costi marginali. Questi sono discussi separatamente nel seguito.

Uso a titolo gratuito (nessuna tariffa). I costi di produzione e di rilascio dei dati sono interamente a carico dell'amministrazione. Questi potrebbero essere ridotti laddove i flussi informativi dell'amministrazione già incorporassero un processo di apertura nativa dei dati secondo il paradigma Open Data. Esistono studi economici [46] che dimostrano come l'applicazione di tariffe superiori ai costi marginali non sia una buona scelta in termini di benefici economici. Infatti, nella maggior parte dei casi, rendere pubblici i dati già posseduti ha un costo irrisorio per un'amministrazione, e il beneficio che l'uso di quei dati può avere è sicuramente maggiore al costo. Ad esempio, facendo riferimento al rapporto contenuto in [47] si può citare il caso del concorso *Apps for Democracy* per applicazioni facenti uso di dati gratuiti. Il concorso aveva un montepremi di 50 mila dollari e ha ottenuto 47 sottomissioni di applicazioni che avrebbero avuto un costo di sviluppo complessivo di 2.6 milioni di dollari. Per fornire i dati in questa modalità si raccomanda di utilizzare licenze per uso a titolo gratuito, come spiegato in Sezione 7.

Tariffe non superiori ai costi marginali. Esistono casi in cui un'amministrazione può avere necessità di coprire parte dei costi di raccolta, produzione e pubblicazione per alcuni tipi di dato. In tali casi, essa può avvalersi della possibilità di fornire i dati a una tariffa che non deve superare i costi marginali prima menzionati. A tal riguardo, si raccomanda di accompagnare questa modalità con delle clausole all'interno della licenza d'uso.



Infine, seppur non previsto da alcuna norma, è preferibile fornire e pubblicare le motivazioni, i criteri e la determinazione dei costi per i dati forniti secondo questa modalità.

Tariffe superiori ai costi marginali. L'amministrazione può fornire i dati a una tariffa superiore ai costi marginali in due casi eccezionali:

1. se è un ente pubblico che deve coprire i costi di supporto alla propria attività amministrativa;
2. se è un istituto della cultura come ad esempio una biblioteca, un'università, un museo o un archivio.

Entrambi i casi recepiscono le indicazioni della nuova direttiva PSI [3] che apre alla possibilità per gli istituti culturali di generare delle entrate per far fronte, almeno in parte, ai costi di raccolta, produzione, conservazione e acquisto dei diritti di sfruttamento dei

In entrambi i precedenti casi, l'amministrazione deve avvalersi di metodi di analisi dei costi (ad esempio, Activity Based Costing) che siano oggettivi, trasparenti e verificabili. A seguito dell'analisi dei costi l'amministrazione deve considerare un opportuno modello di business (si veda sotto) per la determinazione delle tariffe. In questo senso, la direttiva PSI non vincola l'uso di particolari metodologie di determinazione di costi e tariffe, tuttavia i criteri e il calcolo di determinazione dei costi devono essere pubblicati prima che una richiesta di fornitura di dati sia espletata.

A tal fine, può essere utile consultare lo studio della commissione europea [48] sui modelli di tariffazione del settore informativo pubblico.

8.2. Modelli di business

La generazione di valore a seguito della pubblicazione degli Open Data è evidente, ma non è sempre immediato quantificare in termini economici l'ammontare di questo valore. Di conseguenza, non è affatto semplice scegliere un modello di business e delle opportune tariffe da applicare ai dati che ricadono nei casi eccezionali precedentemente individuati.

L'applicazione di modelli di business è stata analizzata nelle "Linee guida per l'interoperabilità semantica attraverso i Linked Open Data" [9]. Le linee guida riprendevano lo studio dell'Istituto Superiore Mario Boella [49] fornendo una panoramica generale su come l'Open Data genera valore e trattando i benefici ottenibili tra gli stakeholder dell'ecosistema.

In questa sezione si elencano brevemente alcuni dei modelli di business che l'amministrazione può considerare qualora decidesse di fornire i propri dati in base a una tariffa. Come detto precedentemente, si raccomanda di fornire comunque le motivazioni e i criteri per la determinazione dei costi e del modello di business scelto.

Si noti che l'elenco non intende essere esaustivo ma rappresenta uno spunto di riflessione del gruppo di lavoro sul tema dei costi associati ai dati.



Tariffa fissa. La tariffa è fissata a priori. Lo svantaggio di questo modello è la poca flessibilità e un'attenta analisi necessaria per la determinazione della tariffa.

Freemium. La tariffa è applicata ad una sola parte dei dati, lasciando comunque la possibilità di sfruttarne alcuni in modo gratuito.

Dual licensing. Gli stessi dati possono essere soggetti a diverse licenze: una licenza che consente un uso gratuito e un'altra licenza che impone il pagamento di una tariffa. Ad esempio, una amministrazione potrebbe scegliere di dare i dati a titolo gratuito per diffonderli a una potenziale vasta utenza che tipicamente non li riutilizza per fini commerciali, ma allo stesso tempo potrebbe utilizzare una licenza a pagamento per coprire eventuali costi, laddove ammessi come precedentemente discusso.

Sponsorizzazione. L'amministrazione può fornire i dati a chi li riusa per la creazione di altri dati o servizi. In questo caso, si vincola l'utilizzatore a mettere in evidenza il marchio del produttore o la provenienza dei dati. Il guadagno dell'amministrazione è quindi determinato dal ritorno d'immagine.

Donazioni e crowdfunding. I dati sono gratuiti ma vengono accettate donazioni e forme di finanziamento libero da parte di utenti che vogliono contribuire al lavoro di apertura dei dati.

Servizi di supporto. I dati possono essere forniti in modo gratuito e il ritorno economico è ottenuto dall'erogazione di servizi di supporto all'uso dei dati come analisi e consulenze, visualizzazioni, elaborazioni, "mashup" di dati provenienti da sorgenti differenti, ecc.

Equity sui servizi. L'amministrazione fornisce i dati gratuitamente a chi li usa per la creazione di un qualche tipo di servizio che a sua volta deve generare introiti. L'amministrazione si riserva la possibilità di trattenere una percentuale prestabilita di quei guadagni.

E' importante sottolineare che nell'applicare le tariffe è necessario considerare il ritorno dell'investimento non dovuto ai pagamenti delle tariffe. Una lista di linee guida sul calcolo del ritorno economico in ambito PSI è fornita in [50].



9. ALCUNE INDICAZIONI OPERATIVE PER I PORTALI OPEN DATA

Inevitabilmente ogni amministrazione si scontra con la scelta di dover decidere le modalità con cui esporre i propri dati all'esterno, come contemplato dal modello operativo introdotto in Sezione 5. Questa sezione intende illustrare alcune caratteristiche e possibili modalità di pubblicazione dei dati da parte delle pubbliche amministrazioni. Si noti che il presente documento non intende entrare nel merito di come organizzare i contenuti e il “layout” grafico delle pagine dei siti tematici Open Data e dei siti istituzionali per la parte “Amministrazione Trasparente”; lo scopo di questa sezione è piuttosto indirizzare le amministrazioni tra le diverse opportunità tecnologiche adottabili per implementare soluzioni Open Data. Pertanto si illustrano i principali requisiti che le infrastrutture di esposizione devono prevedere per rendere disponibili dati con caratteristiche dei livelli 4 e 5 del modello proposto in sezione 4.1. Verranno infine discusse alcune modalità generiche di pubblicazione dei dati all'interno di portali Web.

9.1. Requisiti per la pubblicazione di dati di livello 4 e 5

Per la pubblicazione degli Open Data secondo i livelli 4 e 5 del modello introdotto nel presente documento, è necessario implementare sulla piattaforma di esposizione dei dati le seguenti tre caratteristiche.

Presenza di uno SPARQL endpoint. Pubblicare solo dataset in formato RDF non è sufficiente. La maggior parte dei database per dati RDF fornisce questo tipo di servizio.

Affinché sia possibile interrogare i dati RDF da Web, e quindi navigare i collegamenti tra i dati stessi, è necessaria la presenza di un servizio Web che accetti interrogazioni SPARQL, le risolva e restituisca i risultati in output.

URI persistenti. Come detto precedentemente, i Linked Data utilizzano URI per risolvere il problema dell'identità.

A tal riguardo, è opportuno che l'identità non muti nel tempo: (i) una stessa entità deve mantenere nel tempo la sua URI e (ii) qualora un'entità cessi di esistere è opportuno che la sua URI non venga riutilizzata da un'entità differente.

Se il punto (i) non venisse rispettato, un utente non potrebbe accedere nuovamente alle informazioni di un'entità già acceduta in passato e dovrebbe effettuare nuovamente tutto il processo di “discovery”. Se il punto (ii) non venisse rispettato, un utente accedrebbe alle informazioni della nuova entità pensando di essere in possesso di quelle dell'entità precedente, generando così potenziali errori.

I lettori interessati possono riferirsi alle raccomandazioni per assegnare URI in modo persistente, redatte da gruppi di lavoro istituiti dalla Commissione Europea nell'ambito del programma ISA [89].



Dereferenziazione delle URI e “content negotiation”. Facendo uso di URI HTTP per identificare le risorse RDF, si potrebbe incorrere in URI ambigue, ossia URI che rappresentano sia entità del Web Semantico, sia risorse Web (ad esempio, pagine Web, file, ecc.).

Per ovviare a eventuali problemi di identificazione è opportuno implementare a livello di server Web una corretta gestione della risoluzione delle URI.

Pertanto, 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”). Tale processo viene anche detto “content negotiation”.

Inoltre, 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 pensare di 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, con la quale si intende l'atto necessario per ottenere una rappresentazione Web di una risorsa identificata da un URI [33].

Il W3C ha pubblicato un rapporto tecnico dettagliato [45] sulla dereferenziazione delle URI e sulla “content negotiation” al quale si consiglia di far riferimento.

9.2. Soluzioni Open Data per i portali Web

Nei casi in cui i dati sono aperti secondo i requisiti del livello 3 del modello proposto per i dati o nei casi in cui si vuole accompagnare ai dati di livello 4 e 5 una serie di altri contenuti utili (ad esempio moduli per i feedback, app showcase, ecc.), è utile disporre di un portale Web. Esistono fondamentalmente tre diverse soluzioni, di seguito riassunte.

Soluzione nativa. La creazione avviene “from scratch”, ovvero 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. Si raccomanda tuttavia di mettere in evidenza eventuali strumenti di interrogazione dei dati (e.g., SPARQL endpoint) che possano essere utilizzati sia da esseri umani ma anche e soprattutto da macchine. Si segnala per esempio il portale SPCData [27].

Estensione soluzione CMS esistente. Molto spesso l'amministrazione è già in possesso di 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, risulta molto utile il lavoro della Comunità del Web Semantico IKS¹⁸ che mira a fornire un'abilitazione alle tecnologie semantiche all'interno dei CMS. In particolare, tra i progetti della comunità merita una menzione il progetto Apache Stanbol¹⁹ che fornisce una serie di strumenti utili per le integrazioni con i CMS esistenti. Si segnalano per esempio i portali del

¹⁸ <http://www.iks-project.eu/>

¹⁹ <http://stanbol.apache.org/index.html>



Comune di Firenze e di Cesena e la piattaforma del CSI Piemonte che prevede una personalizzazione del CMS Joomla! e che è utilizzata dalla regione Piemonte e dalla regione Emilia Romagna.

Utilizzo di strumenti esterni. Esistono casi in cui, per la pubblicazione dei dati, si fa affidamento a strumenti e piattaforme esterne, in modo analogo al dispiegamento di servizi in modalità Cloud Computing. I vantaggi in questo caso sono quelli di esternalizzare il processo di gestione della pubblicazione e in alcuni casi anche della memorizzazione e storicizzazione dei dati. Questi strumenti possono essere facilmente integrati con portali già esistenti. Gli strumenti di questo tipo più utilizzati sono CKAN [34] (per esempio la provincia autonoma di Trento e la regione Toscana utilizzano CKAN), Socrata [35] (per esempio la regione Lombardia utilizza Socrata) e OGD Data Lab [36] (per esempio il Ministero della Salute ha avviato una sperimentazione sull'uso di OGD Data Lab).

9.3. Requisiti per i portali Open Data

Pur lasciando alle amministrazioni piena libertà nella progettazione dei portali Open Data, è auspicabile che le stesse seguano alcune indicazioni operative utili in fase di progettazione. Tali indicazioni sono:

1. assegnare ai dataset nomi autoesplicativi che consentano di comprendere il principale contenuto;
2. pubblicare i responsabili del dataset con i relativi contatti;
3. fornire delle interrogazioni (query) d'esempio per facilitare l'utente nel comprendere l'ontologia in uso;
4. produrre e mostrare statistiche di uso, accesso e produzione;
5. mettere in evidenza la presenza di un eventuale SPARQL endpoint, pubblicandone il link di accesso;
6. mettere in evidenza la licenza in uso in forma "human-readable";
7. nel caso di uso di molteplici licenze diverse con obbligo di attribuzione (si veda Figura 6), se lo spazio a video è limitato (come nel caso di applicazioni mobili) si consiglia di creare una pagina Web dedicata per dare indicazioni sulla licenza e la relativa fonte.
8. fare, ove possibile, descrizioni testuali esaustive sui dataset.

9.4. Recupero e accesso ai dati

Il recupero e l'accesso puntuale ai dati è necessario per garantire una buona fruibilità dell'informazione. Il problema assume una rilevanza maggiore se l'utente non conosce la struttura della sorgente dati, la sua collocazione fisica o addirittura la sua esistenza.

A tal fine, occorre far presente che se i dati sono esposti a livello 3 (il livello minimo previsto dall'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico), si limita la possibilità di recuperare i dati in maniera semplice. Il modello qualitativo dei dati suggerisce quindi che le amministrazioni facciano uno sforzo maggiore per esporre dati di alta qualità proprio per rendere anche più semplici le procedure di accesso alle informazioni.



Accesso a dati di livello 3. Con i dati di livello 3 si individuano quattro diverse modalità, anche considerabili in modo misto, di recupero informazioni:

1. si demanda all'utente il “download” dei dataset e l'uso delle funzionalità di ricerca offerte dai classici editor (e.g., editor di testo e di fogli di calcolo);
2. si possono costruire sistemi informativi ad-hoc (e.g., motori di ricerca verticali) per il recupero di informazioni all'interno dei singoli dataset;
3. si predispongono una serie di API che interrogano i dati aperti;
4. si sfruttano i metadati per facilitare l'individuazione dei dataset che potenzialmente contengono i dati di interesse attraverso l'uso di piattaforme per l'Open Data (e.g., CKAN).

Accesso a dati di livello 4 e 5. L'accesso puntuale ai dati di livello 4 e 5 è notevolmente facilitato dagli standard e dalle funzionalità dell'infrastruttura Web esistente. Ad esempio, all'interno della nuvola dei dati SPC è molto facile fornire uno strumento di ricerca basato su parole chiave [27]. Tuttavia è importante evidenziare in questa sezione alcuni prodotti e strumenti esistenti che, sfruttando dati di alta qualità, riescono a supportare l'utente nell'accesso ai dati.

SINDICE [41]. E' probabilmente il più grande indice del Web Semantico. Esso sfrutta tecniche efficienti di indicizzazione inversa e di crawling per il Web dei Dati. Esso può essere interrogato sia attraverso ricerche basate su parole chiave sia con modalità più avanzate.

SQUIN [42]. E' uno strumento che consente di fare interrogazioni SPARQL sul Web dei Dati. Esso sfrutta la presenza di uno SPARQL endpoint e la dereferenziazione delle URI, come precedentemente descritto, per risolvere le interrogazioni con l'attraversamento dei collegamenti dei dati di livello 5.

LODLIVE [43]. E' uno strumento che, se opportunamente configurato, consente di navigare i dati di livello 4 e 5 in maniera efficace (attraverso il proprio browser Web e mediante un'interfaccia grafica accattivante) ed efficiente (funziona lato client). E' stato utilizzato nell'iniziativa Open Data della Camera dei Deputati [44] e in SPCData [27]. In maniera dimostrativa, si veda il suo funzionamento a partire dalla seguente risorsa Roma Capitale <http://spcdata.digitpa.gov.it/lodlive/live.html?http://spcdata.digitpa.gov.it/Comune/H501>.



10. INDICAZIONI OPERATIVE PER I DATI DI TIPO APERTO NEI CAPITOLATI DI GARA

Nella definizione dei capitolati o degli schemi dei contratti di appalto relativi a prodotti e servizi che comportino la raccolta e la gestione di dati pubblici, le pubbliche amministrazioni, ai sensi dell'art. 9 del DL n. 179/2012, devono prevedere clausole idonee che permettano collateralmente alla gestione amministrativa dei procedimenti anche la produzione dei dati pubblici che da tali procedimenti discendano, prestando attenzione a possibili requisiti di privacy e individuando prioritariamente all'interno del procedimento il momento nel quale il dato disaggregato viene prodotto e il soggetto deputato ad effettuare le opportune operazioni di trattamento del dato stesso con particolare riguardo alla metadattazione.

Le clausole, in conformità con i modelli per i dati e metadati introdotti dalle presenti linee guida, devono garantire (i) una chiara definizione della proprietà dei dati, (ii) che le pubbliche amministrazioni possano accedere telematicamente alla totalità dei dati e dei relativi metadati, (iii) che il dato presenti il maggior grado di fruibilità possibile in modo da consentire alle persone fisiche e giuridiche, oltre all'accesso telematico, anche il riutilizzo.

Tenuto conto della complessità del tema, che inevitabilmente coinvolge diversi interlocutori, il gruppo di lavoro ha ritenuto più appropriato approfondirlo nei prossimi mesi. A tal fine, saranno dedicate linee guida specifiche distinte dalle presenti allo scopo di porre particolare attenzione su una leva importante e delicata da considerare che è il procurement dei sistemi informativi in un'ottica focalizzata sui dati.



11. BIBLIOGRAFIA

- [1] European Commission, “Digital Agenda For Europe”, <http://ec.europa.eu/digital-agenda/>, 2013.
- [2] “Agenda Digitale Italiana”, http://www.agenda-digitale.it/agenda_digitale/, 2013.
- [3] “Directive 2013/37/EU of the European Parliament and of the council – amending directive 2003/98/EC on the reuse of public sector information, in *Official Journal of the European Union*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>, 26 giugno 2013.
- [4] Cabinet Office, “G8 Open Data Charter and Technical Annex”, <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>, 18 giugno 2013.
- [5] Regioni – CISIS, “Regolamenti e direttive locali”, <http://www.agid.gov.it/dati-pubblici/open-data>, 2013.
- [6] Creative Commons Italia, <http://www.creativecommons.it/Licenze>, 2013.
- [7] Agenzia per l'Italia Digitale, “Obiettivi di accessibilità per l'anno 2013”, <http://www.digitpa.gov.it/sites/default/files/Obiettivi%20accessibilit%C3%A0%20AGID.pdf>, Marzo 2013.
- [8] Tim Berners Lee, “Linked Data”, <http://www.w3.org/DesignIssues/LinkedData.html>, 2013.
- [9] Commissione di Coordinamento SPC, “Linee guida per l'interoperabilità semantica attraverso i Linked Open Data”, Novembre 2012.
- [10] ISA programme, “Towards open government metadata”, Project Open Data, “Common Core Metadata Schema”, <http://project-open-data.github.io/schema/>, settembre 2011.
- [11] Project Open Data, “Common Core Metadata Schema”, <http://project-open-data.github.io/schema/>, 2013.
- [12] W3C, RDF Working Group, http://www.w3.org/2011/rdf-wg/wiki/Main_Page, 2013.
- [13] W3C, RDF Vocabulary Description Language 1.0: RDF Schema, <http://www.w3.org/TR/rdf-schema/>, 2013.
- [14] W3C, OWL Working Group, http://www.w3.org/2007/OWL/wiki/OWL_Working_Group, 2013.
- [15] W3C, SPARQL Working Group, http://www.w3.org/2009/sparql/wiki/Main_Page, 2013.
- [16] SDMX - Statistical Data and Metadata Exchange, <http://sdmx.org/>, 2013.
- [17] IETF, “RFC 4180 – Common Format and MIME Type for Comma Separated Values (CSV) Files”, <http://tools.ietf.org/html/rfc4180>, 2013.
- [18] OASIS, “Open Document Format for Office Applications”, https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office, 2013.
- [19] W3C, SKOS Simple Knowledge Organization System - Home Page, <http://www.w3.org/2004/02/skos/>, 2013.



- [20] European Commission, “Core public service vocabulary”, http://joinup.ec.europa.eu/asset/core_public_service/news/core-public-service-vocabulary-released-public-review, 2013.
- [21] SPOCS project, <http://www.eu-spocs.eu/>, 2013.
- [22] European Commission, “INSPIRE – Infrastructure for Spatial Information in the European Community”, <http://inspire.jrc.ec.europa.eu/>, 2013.
- [23] W3C, RDF Data Cube, <http://www.w3.org/TR/vocab-data-cube/>, 2013.
- [24] W3C, Data Catalog Vocabulary (DCAT), <http://www.w3.org/TR/vocab-dcat/>, 2013.
- [25] DBpedia, <http://dbpedia.org/About>, 2013.
- [26] DBpedia Italiana, <http://it.dbpedia.org/>, 2013.
- [27] SPCData, <http://spcdata.digitpa.gov.it>, 2013.
- [28] GeoNames, <http://www.geonames.org/>, 2013.
- [29] Agenzia per l'Italia Digitale, “Repertorio nazionale per i dati territoriali – RNDT”, <http://www.rndt.gov.it/RNDT/home/index.php>, 2013.
- [30] ISPRA, SINAnet, <http://www.mais.sinanet.isprambiente.it/ost/>, 2013.
- [31] OpenStreetMap – The free Wiki World Map, <http://www.openstreetmap.org/>, 2013.
- [32] Open Data Commons, Open Database License (ODbL), <http://opendatacommons.org/licenses/odbl/>, 2013.
- [33] W3C Draft Tag Finding, Deferencing http URI, <http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14>, 2013.
- [34] CKAN – The Data Hub, “Linking Open Data Cloud”, <http://thedatahub.org/group/locloud>, 2012.
- [35] Socrata, <http://www.socrata.com/>, 2013.
- [36] OpenLab, Microsoft, <https://github.com/openlab>, 2013.
- [37] Creative Commons, “Describing Copyright in RDF”, <http://creativecommons.org/ns>, 2013.
- [38] Dublin Core Metadata Initiative, <http://dublincore.org/documents/dcmes-xml/>, 2013.
- [39] W3C, Asset Description Metadata Schema (ADMS), <http://www.w3.org/ns/adms>, 2013.
- [40] Simone Aliprandi, Legal tools for Open Data, http://www.ifosslr.org/public/opendata_graph.pdf, 2013.
- [41] Sindice – The Semantic Web Index, <http://sindice.com/>, 2013.
- [42] Squin – Query the Web of Linked Data, <http://squin.sourceforge.net/index.shtml>, 2013.
- [43] Lodlive – browsing the Web of Data, <http://en.lodlive.it/>, 2013.
- [44] Camera dei deputati, <http://dati.camera.it/it/>, 2013.
- [45] W3C, Cool URIs for the Semantic Web –<http://www.w3.org/TR/cooluris/>, 3 Dicembre 2008.
- [46] Paul Uhrlir, “The Socioeconomic effects of public information on digital networks”, 2009.



- [47] United Nations, “E-Government Survey 2010”, 2010.
- [48] European Commission, “Pricing Of Public Sector Information Study - Models of Supply and Charging for Public Sector Information (ABC)”, 2011.
- [49] E. Ferro, M. Osella, “Modelli di Business nel Riuso dell'Informazione Pubblica”, Osservatorio ICT della Regione Piemonte,
<http://www.osservatorioict.piemonte.it/it/images/phocadownload/modelli%20di%20business%20nel%20riuso%20dellinformazione%20pubblica.pdf>, 2011.
- [50] C. Corbin, Public Sector Information Economic Indicators & Economic case study on charging models, 2010.
- [51] e-Codex project, <http://www.e-codex.eu/home.html>, 2013.
- [52] ISO/IEC 25012 “Data Quality Model”. 2008.
- [53] W3C, “Recommendation - Extensible tensible markup language (XML) 1.0”, Novembre 2008, <http://www.w3.org/TR/xml/>.
- [54] W3C, “Recommendation - RDF/XML Syntax Specification”, <http://www.w3.org/TR/rdf-syntax-grammar/>.
- [55] W3C, “Notation3 (N3): A readable RDF syntax”, <http://www.w3.org/TeamSubmission/n3/>, 2013.
- [56] W3C, “Turtle - Terse RDF Triple Language”, <http://www.w3.org/TeamSubmission/turtle/>, 2013.
- [57] IETF, RFC 4627 - The application/json Media Type for JavaScript Object Notation (JSON), <http://www.ietf.org/rfc/rfc4627.txt>, 2013.
- [58] W3C, “N-Triples”, <http://www.w3.org/TR/n-triples/>, 2013.
- [59] ESRI, “Shapefile Technical Description”, <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, 2013.
- [60] GoodRelations Language Reference, <http://www.heppnetz.de/ontologies/goodrelations/v1.html>, 2013.
- [61] Public Contracts Ontology, <https://code.google.com/p/public-contracts-ontology/>, 2013.
- [62] Guide to the Payments Ontology, <http://data.gov.uk/resources/payments>, 2013.
- [63] FOAF Vocabulary Specification, <http://xmlns.com/foaf/spec/>, 2013.
- [64] Who Am I! Vocabulary Specification, <http://vocab.ctic.es/wai/wai.html>, 2013.
- [65] SIOC Core Ontology Specification, <http://rdfs.org/sioc/spec/>, 2013.
- [66] European Commission, Core Location Vocabulary, https://joinup.ec.europa.eu/asset/core_location/description, 2013.
- [67] European Commission, Core Person Vocabulary, http://joinup.ec.europa.eu/asset/core_person/description, 2013.
- [68] W3C, “Candidate Recommendation. The Organization Ontology”, <http://www.w3.org/TR/vocab-org/>, 2013.
- [69] W3C, “Working Group Note. Registered Organization Vocabulary”,



- <https://dvcs.w3.org/hg/gld/raw-file/default/legal/index.html>, 2013.
- [70] W3C, “OWL-S: Semantic Markup for Web Services”, <http://www.w3.org/Submission/OWL-S/>, 2013.
- [71] Description of a Project, <https://github.com/edumbill/doap/wiki>, 2013.
- [72] W3C, “vCard Ontology”, <http://www.w3.org/TR/vcard-rdf/>, 2013.
- [73] Eurovoc, <http://eurovoc.europa.eu/>, 2013.
- [74] FAO, Agrovoc, <http://aims.fao.org/standards/agrovoc/about>, 2013.
- [75] UNICE, EUROSTAT, OECD, “Extended Knowledge Organization System (XKOS)”, <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP10.pdf>, 2013.
- [76] Thesaurus PICO, http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml, 2013.
- [77] UNESCO Thesaurus, <http://databases.unesco.org/thesaurus/>, 2013.
- [78] The CIDOC Conceptual Reference Model, http://www.cidoc-crm.org/official_release_cidoc.html, 2013.
- [79] Europeana Data Model (EDM) Documentation, <http://pro.europeana.eu/edm-documentation>, 2013.
- [80] W3C Interest Group Note, “Describing Linked Datasets with the VOID Vocabulary”, <http://www.w3.org/TR/void/>, 2013.
- [81] Dublin Core Metadata Element Set, Version 1.1, <http://dublincore.org/documents/dces/>, 2013.
- [82] Dublin Core Metadata Initiative, <http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#>, 2013.
- [83] W3C Working Group Note, “PROV-Overview An Overview of the PROV Family of Documents”, <http://www.w3.org/TR/prov-overview/>, 2013.
- [84] Schema.org, <http://schema.org/>, 2013.
- [85] VANN: A vocabulary for annotating vocabulary descriptions, <http://vocab.org/vann/.html>, 2013.
- [86] U.S. National Library of Medicine, “Medical Subject Headings”, <http://www.nlm.nih.gov/mesh/>, 2013.
- [87] SNOMED CT, <http://www.ihtsdo.org/snomed-ct/>, 2013.
- [88] European Commission, “e-Government Core Vocabularies”, https://joinup.ec.europa.eu/community/core_vocabularies/description, 2013.
- [89] European Commission, “10 Rules for Persistent URIs”, <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>, 2013.
- [90] 5 star open data, <http://5stardata.info/>, 2013.
- [91] W3C, "Recommendation - XML Schema Definition Language (XSD) 1.1 Part 1: Structures", <http://www.w3.org/TR/xmlschema11-1/>, 5 April 2012.



- [92] W3C, "Recommendation - XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes", <http://www.w3.org/TR/xmlschema11-2/>, 5 April 2012.
- [93] DTD, <http://www.w3schools.com/dtd/default.asp>, 2013.
- [94] Agenzia per l'Italia Digitale, "Agenda Nazionale per la valorizzazione del patrimonio informativo pubblico (secondo semestre 2013)", <http://www.agid.gov.it/dati-pubblici/open-data>, 2013.



Agenzia per l'Italia Digitale

APPENDICE I

Lo scopo di questa appendice è quello di fornire esempi concreti di utilizzo degli standard e delle ontologie di riferimento.

Considerando il contenuto tecnico della stessa, l'appendice si rivolge a tecnici delle pubbliche amministrazioni.

ESEMPIO SULLE SERIALIZZAZIONI RDF

Come riportato in sezione 6.1, esistono diversi formati di rappresentazione dei dati RDF. Nel seguito si prendono in esame i dati che descrivono l'Agenzia per l'Italia Digitale e la sua locazione fisica rappresentandoli in N3, N-Triple e RDF/XML.

RDF/XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:locn="http://www.w3.org/ns/locn#"
xmlns:org="http://www.w3.org/ns/org#" xmlns:spcdata="http://spcdata.digitpa.gov.it/">
  <spcdata:Amministrazione rdf:about="http://spcdata.digitpa.gov.it/Amministrazione/agid">
    <rdf:label>Agenzia per L'Italia Digitale</rdf:label>
    <org:identifier>97103420580</org:identifier>
    <org:classification rdf:resource="http://spcdata.digitpa.gov.it/CategoriaAmministrazione/33" />
    <locn:address>
      <locn:Address rdf:about="http://spcdata.digitpa.gov.it/Indirizzo/agid">
        <locn:postCode>00137</locn:postCode>
        <locn:fullAddress>Viale Marx 43, 00137, Roma</locn:fullAddress>
      </locn:Address>
    </locn:address>
  </spcdata:Amministrazione>
</rdf:RDF>
```

NOTATION 3

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix org: <http://www.w3.org/ns/org#> .
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix spcdata: <http://spcdata.digitpa.gov.it/> .
```



Agenzia per l'Italia Digitale

<http://spcdata.digitpa.gov.it/Amministrazione/agid> rdf:type spcdata:Amministrazione .
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> rdf:label "Agenzia per L'Italia Digitale" .
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> org:identifier "97103420580" .
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> org:classification
 <http://spcdata.digitpa.gov.it/CategoriaAmministrazione/33> .
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> locn:address <http://spcdata.digitpa.gov.it/Indirizzo/agid> .
 <http://spcdata.digitpa.gov.it/Indirizzo/agid> rdf:type locn:Address .
 <http://spcdata.digitpa.gov.it/Indirizzo/agid> locn:postCode "00137" .
 <http://spcdata.digitpa.gov.it/Indirizzo/agid> locn:fullAddress "Viale Marx 43, 00137, Roma" .

N-TRIPLES

<http://spcdata.digitpa.gov.it/Amministrazione/agid> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
 <http://spcdata.digitpa.gov.it/Amministrazione> .
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> <http://www.w3.org/1999/02/22-rdf-syntax-ns#label> "Agenzia per
 L'Italia Digitale".
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> <http://www.w3.org/ns/org#identifier> "97103420580".
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> <http://www.w3.org/ns/org#classification>
 <http://spcdata.digitpa.gov.it/CategoriaAmministrazione/33> .
 <http://spcdata.digitpa.gov.it/Amministrazione/agid> <http://www.w3.org/ns/locn#address>
 <http://spcdata.digitpa.gov.it/Indirizzo/agid> .
 <http://spcdata.digitpa.gov.it/Indirizzo/agid> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
 <http://www.w3.org/ns/locn#Address> .
 <http://spcdata.digitpa.gov.it/Indirizzo/agid> <http://www.w3.org/ns/locn#postCode> "00137".
 <http://spcdata.digitpa.gov.it/Indirizzo/agid> <http://www.w3.org/ns/locn#fullAddress> "Viale Marx 43, 00137, Roma".

ESEMPIO DI UTILIZZO DI ONTOLOGIE E VOCABOLARI PER DATI TRASVERSALI

Prendendo spunto da SPCData e dalle classificazioni in esso utilizzate, l'esempio descrive un'amministrazione del Lazio che ha dichiarato di avere un responsabile e che ha sottoscritto un contratto per l'acquisizione di un dato servizio.

L'esempio non corrisponde a un caso reale, sebbene vengano utilizzate alcune risorse esistenti. Esso si basa sull'uso di alcune delle ontologie della sezione 6.3.1.

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
 @prefix geonames: <http://www.geonames.org/ontology#> .
 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
 @prefix owl: <http://www.w3.org/2002/07/owl#> .
 @prefix pc: <http://purl.org/procurement/public-contracts#> .



Agenzia per l'Italia Digitale

@prefix gr: <http://purl.org/goodrelations/v1#> .
 @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
 @prefix org: <http://www.w3.org/ns/org#> .
 @prefix adms: <http://www.w3.org/ns/adms#> .
 @prefix spcdata: <http://spcdata.digitpa.gov.it/> .

<http://spcdata.digitpa.gov.it/Regione/12> rdf:type spcdata:Regione .
 <http://spcdata.digitpa.gov.it/Regione/12> rdfs:label "Lazio" .
 <http://spcdata.digitpa.gov.it/Regione/12> owl:sameAs <http://dbpedia.org/resource/Lazio> .
 _:b1 geonames:locatedIn <http://spcdata.digitpa.gov.it/Regione/12> .
 _:b1 rdf:type spcdata:Amministrazione .
 spcdata:Amministrazione rdfs:subClassOf org:FormalOrganization .
 <http://spcdata.digitpa.gov.it/GruppoAmministrazioni/reg12-catL6> foaf:member _:b1 .
 <http://spcdata.digitpa.gov.it/GruppoAmministrazioni/reg12-catL6> rdf:type spcdata:GruppoAmministrazioni .
 <http://spcdata.digitpa.gov.it/GruppoAmministrazioni/reg12-catL6> rdfs:member
 <http://spcdata.digitpa.gov.it/CategoriaAmministrazione/68> .
 <http://spcdata.digitpa.gov.it/CategoriaAmministrazione/68> rdf:type spcdata:CategoriaAmministrazione .
 <http://spcdata.digitpa.gov.it/CategoriaAmministrazione/68> skos:inScheme spcdata:ClassificazionePA-S13 .
 spcdata:ClassificazionePA-S13 rdf:type skos:ConceptScheme .
 spcdata:CategoriaAmministrazione rdf:type skos:Concept .
 <http://spcdata.digitpa.gov.it/Contratto/12-L6-7> pc:contractingAuthority _:b1 .
 <http://spcdata.digitpa.gov.it/Contratto/12-L6-7> rdf:type pc:Contract .
 <http://spcdata.digitpa.gov.it/Contratto/12-L6-7> pc:startDate "2008-02-25" .
 <http://spcdata.digitpa.gov.it/Contratto/12-L6-7> pc:actualEndDate "2012-06-20" .
 <http://spcdata.digitpa.gov.it/Contratto/12-L6-7> pc:item <http://spcdata.digitpa.gov.it/VoceContratto/12-L6-7-SRV_32> .
 <http://spcdata.digitpa.gov.it/VoceContratto/12-L6-7-SRV_32> gr:includes
 <http://spcdata.digitpa.gov.it/ServizioContratto/SYM> .
 <http://spcdata.digitpa.gov.it/VoceContratto/21-C1-1-SYM_3> rdf:type gr:Offering .
 <http://spcdata.digitpa.gov.it/ServizioContratto/SYM> rdf:type gr:ProductOrService .
 spcdata:Responsabile rdfs:subClassOf foaf:Person .
 <http://spcdata.digitpa.gov.it/Responsabile/resp-abnfa_tn-abnfa_tn_aoo1> org:headOf
 <http://spcdata.digitpa.gov.it/Amministrazione/abftrm> .

ESEMPIO DI UTILIZZO DI ONTOLOGIE PER I METADATI

Nel seguito si mostra l'uso delle ontologie per i metadati spiegate nella sezione 6.3.3. L'esempio è reale e basato sulla metadattazione del dataset Linked Open IPA di SPCData dove si utilizzano sia alcuni metadati minimi che altri supplementari.

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .



Agenzia per l'Italia Digitale

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

@prefix adms: <http://www.w3.org/ns/adms#> .

@prefix dcat: <http://www.w3.org/ns/dcat#> .

@prefix cc: <http://creativecommons.org/ns#> .

@prefix dcmi: <http://purl.org/dc/terms/> .

@prefix spcdata: <http://spcdata.digitpa.gov.it/> .

spcdata:ClassificazionePA-S13 rdf:type adms:SemanticAsset .

spcdata:ClassificazionePA-S13 adms:interoperabilityLevel <http://purl.org/adms/interoperabilitylevel/Semantic> .

spcdata:ClassificazionePA-S13 adms:representationTechnique
<http://purl.org/adms/representationtechnique/SKOS> .

spcdata:ClassificazionePA-S13 dcmi:type <http://purl.org/adms/assettype/Taxonomy> .

spcdata:ClassificazionePA-S13 rdfs:label "Classificazione IPA delle PA" .

spcdata:ClassificazionePA-S13 <http://www.w3.org/ns/rad#keyword> "Classificazione PA, S13" .

spcdata:ClassificazionePA-S13 dcmi:publisher <http://spcdata.digitpa.gov.it/Amministrazione/agid> .

spcdata:lodspc rdf:type dcat:Catalog .

spcdata:lodspc dcmi:title "Linked Open Data SPC" .

spcdata:lodspc dcmi:description "Web dei Dati SPC" .

spcdata:lodspc dcmi:creator <http://spcdata.digitpa.gov.it/Amministrazione/agid> .

spcdata:lodspc dcmi:publisher <http://spcdata.digitpa.gov.it/Amministrazione/agid> .

spcdata:lodspc dcat:dataset spcdata:loipa .

spcdata:loipa rdf:type dcat:Dataset .

spcdata:loipa dcmi:description "Versione Linked Data dell'Indice della Pubblica Amministrazione" .

spcdata:loipa dcat:keyword "PEC, IPA, IndicePA" .

spcdata:loipa dcmi:creator <http://spcdata.digitpa.gov.it/Amministrazione/agid> .

spcdata:loipa dcmi:publisher <http://spcdata.digitpa.gov.it/Amministrazione/agid> .

spcdata:loipadump rdf:type dcat:Distribution .

spcdata:loipadump rdf:type dcat:Distribution .

spcdata:loipa dcat:distribution spcdata:loipadump .

spcdata:loipa dcat:distribution spcdata:loipawod .

spcdata:loipadump dcat:accessURL "http://spcdata.digitpa.gov.it/data/ipa.nt" .

spcdata:loipawod dcat:accessURL "http://spcdata.digitpa.gov.it:8899/sparql" .

spcdata:loipadump dcmi:format spcdata:dump .

spcdata:dump rdf:type dcmi:IMT .

spcdata:dump rdf:value "text/n3" .

spcdata:dump rdfs:label "NTRIPLE" .

spcdata:loipawod dcmi:format spcdata:endpoint .

spcdata:endpoint rdf:type dcmi:IMT .

spcdata:endpoint rdf:value "application/sparql-query" .



spcdata:endpoint rdfs:label "SPARQL Endpoint" .
spcdata:License rdf:type cc:License .
spcdata:loipa cc:license spcdata:License .
spcdata:License cc:permits cc:Reproduction .
spcdata:License cc:permits cc:Distribution .
spcdata:License cc:permits cc:Sharing .
spcdata:License cc:permits cc:DerivativeWorks .

