



Agenzia per l'Italia Digitale
Presidenza del Consiglio dei Ministri

LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO (ANNO 2014)



allegato alla Determinazione Commissariale n. 95/2014 del 26-06-2014

INDICE

EXECUTIVE SUMMARY	6
1. PREFAZIONE.....	9
1.1 Componenti del gruppo di lavoro	9
1.2 Modifiche al documento	10
1.3 Acronimi	11
1.4 Glossario	13
1.5 Licenza	16
2. SCOPO, DESTINATARI E STRUTTURA DEL DOCUMENTO	17
3. INTRODUZIONE.....	19
3.1. Quadro normativo	21
3.1.1. <i>Le relazioni tra trasparenza e Open Data</i>	22
3.1.2. <i>Regolamenti delle pubbliche amministrazioni e direttive locali</i>	24
3.2. Aspetti di sussidiarietà	32
3.3. Definizioni: dati delle pubbliche amministrazioni e documenti	33
4. MODELLI DI RIFERIMENTO PER I DATI E I METADATI	36
4.1. Modello per i dati.....	36
4.1.1. <i>Linked Open Data</i>	38
4.2. Modello per i metadati	39
4.2.1. <i>Metadati obbligatori e obbligatori condizionatamente</i>	40
5. MODELLO OPERATIVO PER LA PRODUZIONE E GESTIONE DEI DATI DI TIPO APERTO	43
5.1. Ruoli e responsabilità	44
5.2. Modello operativo.....	46
5.2.1. <i>Linea 1: Dati Nativi</i>	46
5.2.2. <i>Linea 2: Dati Mashup</i>	55
5.2.3. <i>Linea 3: Linked Open Data</i>	56
5.2.4. <i>Linea 4: Engagement (Coinvolgimento)</i>	56
5.3. Coordinamento tra livello nazionale e livello locale	59
6. STANDARD TECNICI E ONTOLOGIE.....	61
6.1. Standard di base	61
6.2. Formati di tipo aperto	63
6.2.1. <i>Formati per i dati di tipo aperto</i>	63
6.2.2. <i>Formati per i dati geografici</i>	65
6.2.3. <i>Formati per i documenti</i>	66
6.3. Ontologie e vocabolari.....	67
6.3.1. <i>Ontologie e vocabolari per dati trasversali</i>	68
6.3.2. <i>Ontologie e vocabolari per dati verticali</i>	71
6.3.2.1. <i>Dati Territoriali e Ambientali</i>	71



6.3.2.2. Dati Statistici.....	72
6.3.2.3. Dati culturali	72
6.3.2.4. Dati contabili.....	73
6.3.3. <i>Ontologie e vocabolari per i metadati</i>	73
6.4. Alcuni dataset di riferimento	74
6.4.1. <i>Dati trasversali</i>	75
6.4.2. <i>Dati territoriali e ambientali</i>	75
6.4.3. <i>Dati sulla medicina</i>	76
7. TITOLARITÀ E PROVENIENZA DEI DATI	77
7.1. PROV Framework	77
7.2. Un esempio di utilizzo del framework PROV	78
8. LICENZE ASSOCIATE AL DATO	80
9. ASPETTI DI COSTO DEL DATO.....	85
9.1. Modalità di rilascio dei dati.....	85
9.2. Modelli di business.....	86
10. ALCUNE INDICAZIONI OPERATIVE PER LA PUBBLICAZIONE DI DATI APERTI.....	89
10.1. Elementi architettonici per la pubblicazione.....	89
10.2. Soluzioni Open Data per i portali Web.....	90
10.3. Requisiti per la pubblicazione di dati di livello 4 e 5	91
10.4. Recupero e accesso ai dati.....	94
10.5. Elementi di federazione.....	96
11. INDICAZIONI OPERATIVE PER I DATI DI TIPO APERTO NEI CAPITOLATI DI GARA.....	98
12. BIBLIOGRAFIA	100
APPENDICE I	106
APPENDICE II	113
APPENDICE III.....	117



INDICE DELLE FIGURE

Figura 1: Processo di valorizzazione del patrimonio informativo pubblico.....	7
Figura 2: Tipi di dato delle pubbliche amministrazioni	35
Figura 3: Modello a cinque livelli per la produzione e il rilascio di dati di tipo aperto.....	36
Figura 4: Modello per la definizione dei metadati	39
Figura 5: Modello operativo per la produzione e gestione dei dati aperti nella PA	46
Figura 6: Il framework PROV per la generazione e pubblicazione della classificazione ATECO2007 ..	79
Figura 7: Strumenti legali per l'Open Data	80
Figura 8: Provenance Linked Open Data Ateco 2007.....	113
Figura 9: Dettaglio dell'entità LinkedAteco2007	114
Figura 10: Ulteriore modello di dettaglio della classificazione LinkedAteco2007.....	115
Figura 11: istanziazione di alcuni elementi della LinkedAteco2007	116



INDICE DELLE TABELLE

Tabella 1: Comparazione di sintesi dei piani di azione delle regioni esaminate.....	31
Tabella 2: Metadati obbligatori.....	41
Tabella 3: Metadati obbligatori al verificarsi di determinate condizioni	42
Tabella 4: Matrice RACI.....	59
Tabella 5: Compatibilità tra licenze.....	83
Tabella 6: Check-list di supporto all'analisi giuridica delle fonti del modello operativo	119



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.

Le pubbliche amministrazioni devono favorire i processi di innovazione sociale in tutte le forme, creando presupposti di conoscenza, partecipazione ed innescando cicli virtuosi che possano produrre idee, servizi, modelli di gestione e di business innovativi ed alternativi. Non è possibile cambiare i servizi (e le strutture) delle amministrazioni con un processo solo endogeno senza stimoli ed interazioni con gli utenti; al tempo stesso le informazioni che le amministrazioni raccolgono, trattano e organizzano devono ritornare in forma elementare ai cittadini.

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**.



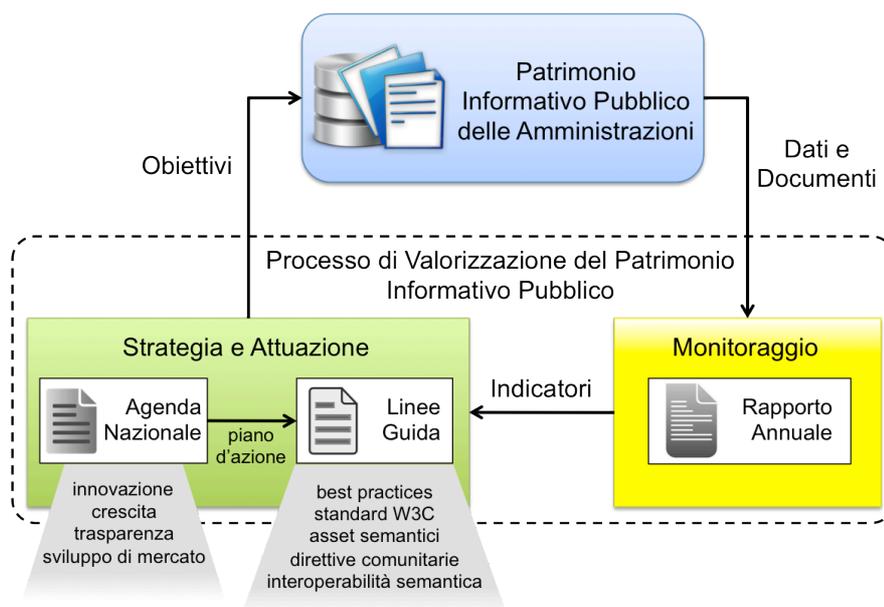


Figura 1: Processo di valorizzazione del patrimonio informativo pubblico

La Figura 1 illustra gli elementi principali del processo di valorizzazione dell'informazione pubblica che l'Agenzia per l'Italia Digitale (nel seguito Agenzia) 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 patrimonio informativo pubblico (tale strategia è definita nell'ambito della cosiddetta agenda nazionale la cui versione per l'anno 2014 è stata recentemente pubblicata dall'Agenzia [94]) e delle linee di azione (linee guida) per l'attuazione della strategia. Le amministrazioni implementano la strategia con il supporto delle linee guida producendo una serie di risultati, quest'ultimi monitorati dall'Agenzia sulla base di indicatori di qualità, quantità e impatto di valorizzazione (definiti e descritti all'interno di un rapporto annuale che l'Agenzia sottopone all'approvazione della Presidenza del Consiglio dei Ministri o del Ministero con delega all'innovazione) attraverso cui verificare l'aderenza delle amministrazioni ai principi della strategia e alle linee guida suddette e, conseguentemente, guidare possibili futuri interventi.

Il presente documento, aggiornato con cadenza annuale da un gruppo di lavoro composto da amministrazioni centrali e locali, rappresenta le linee guida del processo di valorizzazione del patrimonio informativo pubblico, redatte allo scopo di fornire indicazioni operative per l'implementazione della strategia nazionale di valorizzazione, descritta nell'agenda nazionale [94].

Il documento **aggiorna e sostituisce nel complesso il precedente**, già pubblicato dall'Agenzia lo scorso luglio 2013, introducendo altresì alcuni elementi di novità. Al fine di agevolare la lettura, si riassumono di seguito i principali interventi apportati rispetto alla versione 2013:

- viene considerato l'intero percorso di valorizzazione del patrimonio informativo pubblico così come descritto all'interno dell'agenda nazionale, coinvolgendo le tre direttrici principali e rimandando per alcune di queste alle relative linee guida e regole tecniche, ove presenti (sezione 3);
- viene evidenziata la relazione tra i due concetti di trasparenza e dati di tipo aperto (sezione



3.1.1);

- vengono aggiornati i regolamenti locali e il relativo documento esteso prodotto a cura del CISIS [5] (sezione 0);
- viene ulteriormente dettagliato il modello operativo per ciò che concerne gli aspetti di analisi della qualità del dato, bonifica e metadatozione ponendo maggiore enfasi a requisiti di aggiornamento dei dati indispensabili per poter abilitare un efficace riutilizzo dei dati stessi (sezione 5). Vengono altresì proposte raccomandazioni per un possibile coordinamento tra il livello amministrativo centrale e locale in presenza di pubblicazione e produzione di analoghe tipologie di dati (si pensi ad esempio al caso di un ministero centrale che abbia necessità di pubblicare dati in formato aperto che sono, per norma, raccolti dalle amministrazioni locali come nel caso del Ministero dell'Istruzione, Ricerca e Università per i dati riguardanti gli aspetti di sicurezza degli edifici scolastici dislocati sul territorio) (sezione 5.3).
- viene arricchito il panorama dei formati per i dati territoriali e i dataset di riferimento nel contesto geospaziale e nel contesto legale e contabile (sezioni 6.2 e 6.4.2);
- viene introdotto il concetto di titolarità del dato, evidenziando, attraverso la descrizione di una best practice, gli elementi che possano caratterizzare in maniera completa la provenienza di un dato, certificandone così la fonte (sezione 7);
- viene posta attenzione alle *sole* licenze necessarie per applicare pienamente il paradigma dei dati aperti (sezione 8);
- vengono proposte alcune raccomandazioni sui principali elementi infrastrutturali, di federazione per la pubblicazione e gestione dei dati aperti (sezioni 10.1, 10.5) e raccomandazioni per la definizione standard di URI persistenti (sezione 10.3);
- infine, vengono suggerite raccomandazioni più articolate per ciò che concerne il trattamento di dati aperti nei capitolati di gara d'appalto (sezione 11).

Il documento è stato reso disponibile in consultazione pubblica; alcuni commenti e arricchimenti pervenuti sono stati recepiti per la definizione della versione finale delle linee guida per l'anno 2014. Un report sui risultati della consultazione è disponibile in [131].



1. PREFERAZIONE

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 dal Dott. G. Ciasullo, Dott.ssa G. Lodi, Ing. A. Maccioni, Ing. A. Rotundo e coordinato dal Dott. F. Tortorelli.

Di tutte le amministrazioni invitate al gruppo di lavoro, hanno aderito le seguenti amministrazioni centrali: Agenzia Spaziale Italiana (ASI), Autorità per la Vigilanza sui Contratti Pubblici di Lavori, Servizi e Forniture (AVCP), Consiglio Nazionale delle Ricerche (CNR), INPS, INAIL, ISPRA, ISTAT, Garante per la protezione dei dati personali, Ministero dei Beni e delle Attività Culturali e del Turismo, Ministero della Salute, Ministero degli Affari Esteri, Ministero della Difesa, Ministero delle Infrastrutture e dei Trasporti, Ministero dell’Istruzione, Ricerca e Università, Ministero dello Sviluppo Economico, Ministero della Giustizia, Ministero dell’Interno e Presidenza del Consiglio dei Ministri – Protezione Civile.

Inoltre, le pubbliche amministrazioni locali hanno aderito al gruppo con diverse Regioni (e.g., Emilia Romagna, Friuli Venezia Giulia, Liguria, Lombardia, Marche, Piemonte, Provincia autonoma di Trento in collaborazione con la Fondazione Bruno Kessler, Veneto) e Comuni (e.g., Firenze, Milano, Ravenna) e le associazioni in rappresentanza di queste CISIS e ANCI.



1.2 Modifiche al documento

Descrizione Modifica	Edizione	Data
Prima revisione linee guida anno 2013	v. 0.1	15-01-2014
Seconda revisione linee guida	v. 0.2	17-02-2014
Terza revisione linee guida	v. 0.3	16-04-2014
Quarta revisione linee guida: executive summary, scopo, introduzione, capitolati di gara, licenze	v. 0.4	16-05-2014
Quinta revisione: modello per i dati e metadati, standard e ontologie, modelli di business	v. 0.5	21-05-2014
Sesta revisione: inserimento sezioni titolarità dei dati, appendici tecniche, modello operativo, indicazioni operative per la pubblicazione di open data e revisione generale	v. 0.6	26-05-2014
Settima revisione: recepimento dei contributi pervenuti durante il periodo di consultazione pubblica	v. 0.7	26-06-2014



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
DPS – Dipartimento per lo Sviluppo e la Coesione Sociale
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
MIUR – Ministero Istruzione Università e Ricerca
OD – Open Data
ODI – Open Data Institute
OKF – Open Knowledge Foundation
OSM – OpenStreetMap
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



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 fatti, eventi atomici, fenomeni oggettivi, informazioni atte 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 funzione 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.

INFORMAZIONE: lo scambio di conoscenza all'interno di una comunità o nella società nonché il significato che viene attribuito a tale conoscenza dalle persone. Essa riguarda dati associati tra loro o inseriti in un determinato contesto, tali da assumere un significato unico e condiviso.

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, il contratto, o altro strumento negoziale, con il quale si regolano i diritti morali e patrimoniali del software e dei dati nonché le condizioni d'uso e di distribuzione degli stessi.

MASHUP: termine utilizzato nell'ambito delle opere creative per cui un'opera è costituita interamente da parti di altre opere, tra loro integrate. Esso è stato poi esteso al contesto informatico ad indicare un processo in cui si integrano contenuti, dati e informazioni provenienti da fonti differenti per fornire nuovi servizi.

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

OPEN SOURCE: termine che indica software di cui è disponibile il codice sorgente e una licenza open source che consenta la concessione a terzi dei diritti patrimoniali di uso, copia, modifica, distribuzione di copie anche modificate, del software. Altrimenti detto “software libero”.

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

PSI (Public Sector Information): è l'informazione in possesso degli enti pubblici, ivi compresi gli organismi di diritto pubblico (cfr. dir. n. 2003/98/CE relativa al riutilizzo dell'informazione del settore pubblico ss.mm.ii. e il d.lgs. 36/2003). 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 della 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.

TITOLARE DEL DATO: la pubblica amministrazione o l'organismo di diritto pubblico che ha originariamente formato per uso proprio o commissionato ad altro soggetto pubblico o privato il documento che rappresenta il dato (d.lgs. 36/2003).

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 4.0 (CC-BY-SA).



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.

Le presenti linee guida sono incrementalmente aggiornate 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, raccomandare un possibile coordinamento tra il livello centrale e locale per lo scambio e pubblicazione di analoghe tipologie di dati aperti, nonché 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 a dare supporto alle pubbliche amministrazioni (i) nella definizione di elementi di provenienza dei dati importanti per “certificare” il titolare del dato, (ii) nella scelta della licenza per i dati di tipo aperto da associare ai dati, (iii) nell’analisi di eventuali aspetti di costo dei dati di tipo aperto, (iv) nella pubblicazione dei dati nei portali Open Data per una maggiore standardizzazione di tale processo, e (v) nell’individuazione di possibili clausole da inserire nei capitolati di gara per il trattamento dei dati pubblici.

Destinatari. Il presente documento è destinato, secondo quanto previsto dal Codice dell’Amministrazione Digitale (art. 2, commi 2 e 4) per l’applicazione del Capo V, a tutte le pubbliche amministrazioni, le società interamente partecipate da enti pubblici o con prevalente capitale pubblico inserite nel conto economico consolidato della pubblica amministrazione, come individuate dall’ISTAT ai sensi dell’art. 1, co. 5, della L. 311/2004 nonché, con riferimento alle disposizioni concernenti l’accesso ai documenti informatici e la fruibilità delle informazioni digitali di cui al capo V del CAD, ai gestori di servizi pubblici e agli organismi di diritto pubblico.

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 a 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 introduce il concetto di titolarità del dato, presentando un esempio concreto nel quale tale elemento viene rappresentato; la sezione 8 descrive alcune licenze disponibili per i dati di tipo aperto, guidando nella scelta di quella più opportuna in base agli obiettivi di riutilizzo e accesso ai dati. La sezione 9 analizza alcuni aspetti legati ai costi e alla tariffazione per i dati di tipo aperto. La sezione 10 fornisce indicazioni operative per la pubblicazione dei dati di tipo aperto nei portali Open Data e infine la sezione 11 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 da tre appendici: un'appendice che illustra alcuni esempi di utilizzo di formati per i dati, ontologie per dati considerabili comuni a tutte le pubbliche amministrazioni (e.g., contratti), una che presenta un esempio di applicazione del framework PROV per la rappresentazione delle informazioni relative alla provenienza del dato e, infine un'appendice che include una "check list" di verifica giuridica delle fonti necessaria per valutare efficacemente eventuali limitazioni nell'apertura di dati.



3. INTRODUZIONE

In coerenza con il ruolo che le norme hanno attribuito all’Agenzia, intesa come un abilitatore dell’economia digitale e non solo come ente di regolazione o di indirizzo, l’Agenzia ha il compito di promuovere le politiche nazionali per la valorizzazione del patrimonio informativo pubblico. A tal riguardo, l’Agenzia ha pubblicato, in collaborazione con la Presidenza del Consiglio dei Ministri, un’agenda nazionale per la valorizzazione del patrimonio pubblico relativa all’anno 2014 [94] nella quale è stato individuato un percorso di valorizzazione articolato nelle tre seguenti direttrici:

- Basi di dati di interesse nazionale
- Convenzioni aperte tra pubbliche amministrazioni
- Dati di tipo aperto

Per tutte e tre le direttrici, l’agenda individua e dettaglia inoltre principi di fruibilità, interoperabilità, qualità, accessibilità, gratuità e riutilizzabilità.

Per poter attuare le politiche dell’agenda, assicurando il pieno rispetto dei suddetti principi, l’Agenzia predispone linee guida e/o regole tecniche per le pubbliche amministrazioni allo scopo di guidarle nella scelta delle azioni da intraprendere e degli standard tecnici da adottare.

Basi di dati di interesse nazionale. Nell’ambito delle basi di dati di interesse nazionale, la definizione degli standard tecnici e delle regole per garantire interoperabilità nello scambio tra queste e le altre basi di dati delle pubbliche amministrazioni è demandata al contesto del Sistema Pubblico di Connettività, dove il framework della Cooperazione Applicativa [97] è applicato ai sensi dell’art. 60 del Codice dell’Amministrazione Digitale (CAD).

In tale ambito è inoltre importante riferirsi alle regole tecniche sulle basi di dati critiche [128] dove lo standard ISO/IEC 25012 su “Data quality model” [52] è adottato come standard di riferimento per assicurare il raggiungimento di determinati livelli di qualità dei dati gestiti.

Convenzioni aperte tra le pubbliche amministrazioni. Nell’ambito delle convenzioni aperte tra pubbliche amministrazioni, per dare attuazione alle disposizioni previste agli articoli 50 e 58 del CAD, l’Agenzia, d’intesa con il Garante per la protezione dei dati personali, ha adottato con determinazione commissariale n. 126 del 24 luglio 2013, le “*linee guida (versione 2.0) per la stesura di convenzioni per la fruibilità di dati delle pubbliche amministrazioni*” [98] connesse al Sistema Pubblico di Connettività.

Le linee guida indirizzano le amministrazioni nella scelta delle modalità attraverso cui rendere fruibili gratuitamente ad altre amministrazioni (o ad altri organismi privati che operano per finalità pubbliche), che ne necessitano per finalità istituzionali, i dati in loro possesso ivi compresi eventuali dati personali. In tal senso, è di particolare rilevanza la semplificazione del procedimento introdotta dalla seconda versione delle linee guida, per cui le convenzioni-quadro che hanno ad oggetto l’accesso a dati personali possono ritenere soddisfatto l’adempimento previsto di “richiesta di parere al Garante per la protezione dei dati personali” se la convenzione è redatta in piena conformità alle linee guida stesse e se tale conformità è opportunamente “autocertificata”.



Queste linee guida, quindi, guidano le amministrazioni nella *condivisione dei dati* e non devono essere confuse con quelle relative al rilascio di dati pubblici secondo il paradigma dei dati aperti, così come definito dall'art. 68 del CAD e oggetto principale del presente documento.

E' infatti importante notare in questo ambito la netta distinzione tra i concetti di *condivisione dei dati* e *dati di tipo aperto*.

Mentre la condivisione di dati riguarda solitamente contesti ristretti (e.g., tra pubbliche amministrazioni o enti con finalità pubbliche) e può agire sulla base di uno determinato scopo di condivisione e su un insieme di dati specifici, inclusi anche dati personali, i dati aperti sono invece tipicamente (i) non riferibili a singole persone e (ii) disponibili gratuitamente per l'uso, il riutilizzo e la distribuzione da parte di chiunque anche per finalità commerciali, soggetti al massimo alla richiesta di indicare la fonte di provenienza dei dati e di riutilizzarli secondo gli stessi termini per cui sono stati licenziati originariamente.

Dati di tipo aperto. I dati aperti quindi rappresentano una direttrice importante per l'attuazione delle politiche di valorizzazione del patrimonio informativo pubblico; le restanti sezioni delle presenti linee guida, alle quali le pubbliche amministrazioni si *uniformano* ai sensi dell'art. 52 comma 7 del CAD, si focalizzano in particolare su tale direttrice.

Già nel 2013, l'Agenzia ha emanato linee guida allo scopo di supportare le pubbliche amministrazioni in un corretto processo di apertura dei dati. Dopo quasi un anno dalla pubblicazione della versione 1.0 e a ormai diversi anni dalla nascita del portale open data in Italia, si nota come un maggior numero di enti rendano disponibili online dati pubblici, con un incremento conseguente dei dataset esposti, e un maggior numero di eventi emergono per promuovere iniziative di apertura e di coinvolgimento delle comunità. Tuttavia, a ciò non corrisponde ancora un ampio riutilizzo di tali dati, anche per finalità commerciali, ovvero i reali benefici in termini di sviluppo economico derivanti dal paradigma dei dati aperti non sono ancora pienamente tangibili. In generale, questo può imputarsi a diversi fattori tra cui anche quello della qualità del dato, spesso trascurata sia in termini di aggiornamento secondo la frequenza a cui i dati sono soggetti, sia in termini di scarsa documentazione del dato (metadattazione), sia in termini di rilascio in formati non aperti e che richiedono un significativo investimento umano e tecnico per poter consentire uno sviluppo efficace di applicazioni e servizi.

Nella pratica poi, si tende ancora spesso a confondere il rilascio dei dati aperti con la messa a disposizione online di dati pubblici in un portale, trascurando gli aspetti di disaggregazione, neutralità tecnologica, metadattazione completa, e legali richiesti per la conformità al paradigma Open Data.

Per superare tali ostacoli, il presente documento di linee guida estende il precedente proponendosi come supporto alle amministrazioni nelle scelte, sotto diversi aspetti organizzativi, tecnici, legali ed economici, necessarie per una produzione (anche intrinsecamente prevista nella realizzazione dei sistemi informativi delle pubbliche amministrazioni) e un rilascio di dati aperti sempre più di qualità, interoperabili, omogenei e quindi facilmente riutilizzabili da chiunque, nel rispetto della normativa italiana vigente e delle recenti disposizioni previste dalla direttiva europea per il trattamento dell'informazione del settore pubblico [3].



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, come in materia di protezione dei dati personali.

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 D.lgs 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.

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 D.lgs. 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 D.lgs. 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 fermi restando i divieti di cui all'art.9”*, che prevede la tutela del segreto statistico, e quindi anche la protezione dei dati personali, consentendo la comunicazione e la diffusione unicamente di dati aggregati o comunque secondo modalità che non rendano identificabili gli interessati.

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 D.lgs 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 nel rispetto della disciplina in materia di protezione dei dati personali; 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 G8 tenutosi a Londra a giugno 2013 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 basilari 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. Tali impegni sono stati recentemente ribaditi e rafforzati nell'ambito dei lavori di preparazione in vista del prossimo G8.

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

3.1.1. Le relazioni tra trasparenza e Open Data

Il D.lgs. 33 del 14 marzo 2013, c.d. decreto trasparenza, ha introdotto importanti disposizioni che da



un lato rafforzano l'applicazione del principio dell'Open Data by default, come riportato precedentemente, ma che dall'altro possono definire una diversa chiave di lettura. Il D.lgs. 33/2013 infatti nasce dall'esigenza di coordinare tre azioni legislative orientate a definire parametri misurabili dell'efficienza delle pubbliche amministrazioni, favorire il contrasto alla corruzione all'interno della servizio pubblico mediante la trasparenza e la pubblicità degli atti. Tre sono le principali norme che trattano di questi temi e che sono state coordinate in questo decreto legislativo: la Legge n. 69 del 19 giugno 2009; il Decreto legislativo n. 150 del 27 ottobre 2009; la Legge n. 190 del 6 novembre 2012.

Questi interventi normativi partono da esigenze diverse rispetto all'open data, per poi convergere in diversi punti, ma senza mai veramente confluire in un corpo organico. Ciò è confermato dall'art. 1 del "decreto trasparenza" che definisce le finalità:

«1. La trasparenza è intesa come accessibilità totale delle informazioni concernenti l'organizzazione e l'attività delle pubbliche amministrazioni, allo scopo di favorire forme diffuse di controllo sul perseguimento delle funzioni istituzionali e sull'utilizzo delle risorse pubbliche.»

e ancora nel comma 3: le disposizioni del presente decreto *«integrano l'individuazione del livello essenziale delle prestazioni erogate dalle amministrazioni pubbliche a fini di trasparenza, prevenzione, contrasto della corruzione e della cattiva amministrazione.»*

Un ulteriore chiarimento è contenuto nell'art. 2 che definisce gli obiettivi:

«1. Le disposizioni del presente decreto individuano gli obblighi di trasparenza concernenti l'organizzazione e l'attività delle pubbliche amministrazioni e le modalità per la sua realizzazione.»

2. Ai fini del presente decreto, per pubblicazione si intende la pubblicazione, in conformità alle specifiche e alle regole tecniche di cui all'allegato A, nei siti istituzionali delle pubbliche amministrazioni dei documenti, delle informazioni e dei dati concernenti l'organizzazione e l'attività delle pubbliche amministrazioni, cui corrisponde il diritto di chiunque di accedere ai siti direttamente ed immediatamente, senza autenticazione ed identificazione.»

Alla luce di tali disposizioni, il decreto trasparenza ha quindi lo scopo di dare pubblicità ai documenti delle pubbliche amministrazioni che per legge devono essere resi accessibili senza ulteriori barriere (accesso civico ex art. 5 del decreto in parola) salvo le tutele in tema di dati personali, segreto di stato, segreto d'ufficio, segreto statistico, diritti industriali (cfr. artt. 1 e 4) e diffondere così la conoscibilità di informazioni per attuare *«il principio democratico e i principi costituzionali di eguaglianza, di imparzialità, buon andamento, responsabilità, efficacia ed efficienza nell'utilizzo di risorse pubbliche, integrità e lealtà nel servizio alla nazione»* (art. 1). Lo scopo è anche fornire ai cittadini documenti in forme fruibili e comprensibili ovvero più orientate alla lettura umana che al "machine-readable", anche per onorare la *«conformità ai documenti originali in possesso dell'amministrazione»* (art. 6). Per questi motivi, spesso tali informazioni sono erogate in forma di documento PDF, ovvero la forma più utilizzata per la pubblicità di atti amministrativi e quindi inclusivi di una certa visione interpretata dalle pubbliche amministrazioni stesse.

Ciò non di meno il punto di contatto con l'open data è fortemente incentivato grazie al formato aperto imposto nell'art. 7 laddove si dispone che *«1. I documenti, le informazioni e i dati oggetto di pubblicazione obbligatoria ai sensi della normativa vigente, resi disponibili anche a seguito dell'accesso civico di cui all'articolo 5, sono pubblicati in formato di tipo aperto ai sensi dell'articolo 68 del Codice dell'amministrazione digitale, di cui al decreto legislativo 7 marzo 2005, n. 82, e sono riutilizzabili ai sensi del decreto legislativo 24 gennaio 2006, n. 36, del decreto legislativo 7 marzo 2005, n. 82, e del decreto legislativo 30 giugno 2003, n. 196, senza ulteriori restrizioni diverse dall'obbligo di citare la fonte e di rispettarne l'integrità.»* Il formato aperto tuttavia è una condizione necessaria ma non sufficiente per poter definire un dato di tipo aperto; infatti quest'ultimo



è definito sulla base di ulteriori elementi caratteristici quali i metadati, le licenze che ne consentono il riutilizzo anche per finalità commerciali, le tecnologie neutre che consentono di riutilizzare realmente tali documenti/dati.

In virtù di tali osservazioni, si evince che i due concetti svolgono due ruoli informativi e funzionali diversi e ortogonali. In termini di trasparenza, alcuni documenti resi pubblici a seguito dell'applicazione del d.lgs 33/2013 nella sezione "Amministrazione Trasparente" del sito web istituzionale devono essere rimossi dopo aver svolto la loro funzione (dopo 5 anni e, per alcuni casi, dopo 3 anni - cfr. art. 14 e 15) e quindi non possono essere propriamente considerati Open Data, per i quali tali restrizioni temporali non si applicano. Inoltre, esistono dati delle pubbliche amministrazioni che assumono un ruolo importante nell'ecosistema degli Open Data e nella creazione di nuove forme di partecipazione (e.g. edifici, farmacie, musei, turismo, etc.) ma che non risultano nell'elenco dei dati obbligatori da pubblicare ai sensi del D.lgs n. 33/2013 (allegato A).

Infine occorre menzionare le "Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati" [130] per ricordare che anche ove *"l'amministrazione riscontri l'esistenza di un obbligo normativo che impone la pubblicazione dell'atto o del documento nel proprio sito web istituzionale è necessario selezionare i dati personali da inserire in tali atti e documenti, verificando, caso per caso, se ricorrono i presupposti per l'oscuramento di determinate informazioni."*

Le pubbliche amministrazioni infatti sono tenute a ridurre al minimo i dati personali e i dati identificativi in conformità ai principi di finalità, necessità, pertinenza e non eccedenza con gli scopi da perseguire e in ogni caso per motivi di trasparenza solo per espressa disposizione di legge o regolamento (art. 19, comma 3, del Codice).

Per motivi di efficienza, efficacia, corretto uso delle risorse pubbliche, buona comunicazione verso i cittadini, nonché sostenibilità nel tempo di un livello informativo coerente e di qualità, si raccomanda quindi di mettere in atto una serie di azioni volte a coordinare, ove possibile, le attività relative alla trasparenza e all'Open Data, seguendo le raccomandazioni delle presenti linee guida proposte nel modello operativo (sezione 5.1).

3.1.2. 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, 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.

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 D.lgs 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,

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'"*. Di prossima approvazione una nuova Agenda Digitale Lombarda che include un intervento sulla valorizzazione del patrimonio informativo pubblico della Regione. In corso di definizione un nuovo documento, aperto al contributo della comunità, con proposte di dati che i Comuni potrebbero pubblicare, con i relativi tracciati.

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 allegati".



Toscana⁸, Umbria⁹ e Provincia autonoma di Trento 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:

- 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;

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 Disciplinary dei criteri generali per l'open data in Umbria.



- 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 1 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 riutilizzo dei dati, 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 autonomia	Legge Regionale	Linee guida	Portale Open Data	Condivisio ne altri enti/riuso	Licenza	Organizzazione processi di liberazione dati
Basilicata			opendata.regione.basilicata.it/opendata/home.jsp			
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, CC-BY e IODL 2.0	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.
Friuli Venezia Giulia	LR 42. del 09/04/2014	In corso di predisposizione	Attivo dal 5/5/2014 https://www.dati.friuliven		IODL 2.0	



Regione Provincia autonom a	Legge Regionale	Linee guida	Portale Open Data	Condivisio ne altri enti/riuso	Licenza	Organizzazione processi di liberazione dati
			eziagiulia.it			
Lazio ¹⁰	LR n. 7/2012 - "Disposizioni in materia di dati aperti e riutilizzo di informazioni e dati pubblici e iniziative connesse"		Aggiudicata gara Filas per la realizzazione del portale dati.lazio.it			
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
Lombardia	LR n. 7/2012 art. 52 (Accessibilità e valorizzazione e del patrimonio informativo pubblico della Regione)	DGR n. 2904 del 11/01/2012	Attivo da 3/2012 il portale dati.lombardia.it	Portale aperto agli enti del territorio, come definito in DRG n. 4324 del 26/10/2012	IODL 2.0	Decreto DG Semplificazione e Digitalizzazione n. 6115 del 10/07/2012 -approvazione «piano di governance Open Data»
Marche			Sezione dedicata nel portale regionale www.regione.marche.it/Home/OpenData.aspx			

¹⁰ Le informazioni sono state aggiunte sulla base delle notizie disponibili allo stato dell'arte.



Regione Provincia autonomia	Legge Regionale	Linee guida	Portale Open Data	Condivisio ne altri enti/riuso	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 responsabilizzazione delle singole strutture dell'ente - processo di richiesta formale dati dall'esterno, vincolante per l'ente
Provincia autonoma di Trento	L.P. n. 16/2012	DGR n. 2858 del 27/12/2012 - in corso di aggiornament o annuale	Attivo da 03/ 2013 dati.trentino.i t Nel febbraio 2012 aperti 160 tematismi geografici in modalità linked data Attivo da 03/2013 dati.trentino.i t A febbraio 2014 aperti 619 dataset.	Punto unico di accesso per il territorio. Presenti dati aperti del Comune di Trento, Comune di Rovereto, APT Val di Fiemme, Informatica Trentina spa.	CC0 e CC-BY 2.5 - previsto a breve il passaggio a CC-BY 4.0	Gruppo di progetto dedicato, coadiuvato dal sistema della ricerca locale: TrentoRISE - FBK. Programmazione annuale per Dipartimento allineata alla pianificazione ICT e innovazione. Nomina di 11 referenti dei dipartimenti e individuazione di referenti tecnici nei servizi per la pubblicazione dei dati. Gruppo di lavoro PaT trasversale su tematiche statistiche, semplificazione, informatica, privacy, affari legislativi - coordinamento strutture referenti per anticorruzione - trasparenza e partecipazione. Attività di formazione/informazione interne all'amministrazione provinciale, ad utenti e professionisti esterni (School of Data) Collaborazione con Comuni e Comunità di Valle. Coinvolgimento delle PMI del territorio: Trentino OD Challenge. Coinvolgimento community e pubblico attraverso social media @DatiTrentinoit - Trentino Open Data (FB).



Regione Provincia autonomia	Legge Regionale	Linee guida	Portale Open Data	Condivisio ne altri enti/riuso	Licenza	Organizzazione processi di liberazione dati
						Monitoraggio avanzamento processo attraverso portale istituzionale PaT ¹¹
Puglia	LR n. 20/2012	DGR n. 2183 del 31/10/2010	Attivo il portale www.dati.puglia.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
Sicilia			Sezione dedicata nel portale regionale pti.regione.sicilia.it/portal/page/portal/PIR_PORTALE/PIR_Iniziativa/PIR_OpenDataSicilia	la Regione Sicilia ha scelto di riusare il software della piattaforma open data della Regione Piemonte.		
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 responsabilizzazione che coinvolge tutte le strutture interne. Costituzione di un gruppo di lavoro
Umbria	LR n. 8/2011 sviluppo amministrazione digitale; prevede all'art.15 un repertorio regionale dei dati pubblici aperti	DGR n. 914/2013	Attivo dal 15/04/2013 dati.umbria.it	Il repertorio è aperto a tutti gli enti del territorio, 3 enti pubblicano dati	CC-BY	Secondo le modalità previste nel disciplinare di cui alla DGR 914/2013 e nel Piano digitale regionale triennale di cui alla DGR 1555/2013

¹¹ <http://www.innovazione.provincia.tn.it/opendata>.



Regione Provincia autonom a	Legge Regionale	Linee guida	Portale Open Data	Condivisio ne altri enti/riuso	Licenza	Organizzazione processi di liberazione dati
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 amministrazi oni pubblicano dati	IODL 2.0 CC- BY	Costituzione di un gruppo di lavoro (CDG-OD)

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 (si pensi ad esempio ai casi dei Comuni di Palermo e Matera che hanno rilasciato recentemente dei piani di azione sui dati aperti che recepiscono, in alcune parti, le raccomandazioni incluse nel presente documento, o al caso del Comune di Guidonia Montecelio che, seppur in presenza del nuovo articolo 52 del CAD e delle linee guida nazionali per la valorizzazione del patrimonio informativo pubblico (2013), ha deciso di adottare un modello per la gestione degli Open Data¹², commissionato all'inizio del 2013, che in alcune parti risulta comunque allineato alle principali indicazioni contenute nelle linee guida nazionali).

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 cogenza 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.

¹² Comune di Guidonia Montecelio, "Sintesi del Modello e Standard per la Gestione degli Open Data", Novembre 2013, <http://www.guidonia.org/portale/comune/modules.php?name=Atti&op=documento&id=3057&siteid=21>



3.2. Aspetti di sussidiarietà

Le azioni necessarie per abilitare una concreta valorizzazione del patrimonio informativo pubblico sono strettamente correlate ai principi di sussidiarietà sia verticale che orizzontale.

Sussidiarietà verticale. 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.

E' possibile pensare a forme di riutilizzo e di condivisione di strumenti tecnici e di conoscenze grazie alla valorizzazione della comunità nazionale del sistema delle PA in merito agli Open Data

Sussidiarietà orizzontale. Per sussidiarietà orizzontale si intende l'insieme di azioni che coinvolgono il rapporto tra i poteri e i ruoli del servizio pubblico e delle organizzazioni della società. La sussidiarietà orizzontale applicata ai dati permette di individuare momenti diversi di attività per il servizio pubblico e per il sistema delle imprese di mercato innescando filiere di qualità e valore.

In particolare, spettano al settore pubblico tutte quelle azioni che coinvolgono la creazione, la certificazione, l'aggiornamento, e più in generale la qualità del dato, così come è di competenza del pubblico soddisfare la domanda di innovazione sociale, fornire servizi innovativi ai cittadini, sulla base della conoscenza dei contesti sociali e delle necessarie politiche di governo. In altre parole, il settore pubblico è responsabile di creare, valorizzare e proteggere il bene comune "dato" dando così nuovo valore ai saperi della pubblica amministrazione che vede reinterpretata la funzione pubblica: da erogatore di servizi a garante della salvaguardia e della produzione del valore collettivo legato alle informazioni. Questa nuova funzione pubblica trova strumenti di premialità per garantirne la diffusione.

Di contro, spetta al privato la parte di uso e la creazione di plusvalore sul dato. Sono le imprese grandi, medie e/o piccole, le associazioni di cittadini e singoli cittadini stessi a innescare idee innovative su quel patrimonio collettivo messo a disposizione dal servizio pubblico, producendo servizi per la collettività a partire dai dati di tipo aperto messi a disposizione. In questa attività, la pubblica amministrazione agisce come *driver* di innovazione spingendo su politiche tese al trasferimento delle conoscenze dai centri di ricerca alle imprese; dai contesti a maggior densità di conoscenza a quelli meno avvantaggiati, con progetti, bandi, e con la convergenza di strumenti di programmazione anche in ottica di acquisizione di nuove conoscenze, di nuovi modelli di business rispetto alle aziende che già collaborano con le pubbliche amministrazioni.

Questa specializzazione in merito alla filiera della produzione di valore sui dati deve tendere a produrre territori ad alta densità di cultura del dato. Per fare questo le PA si dotano di strumenti di conoscenza standardizzati sui principi base di organizzazione delle informazioni in modalità strutturata.



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;”*

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, i dati personali), oggetto delle presenti linee guida è il dato pubblico, cioè il dato della pubblica amministrazione conoscibile da chiunque e non soggetto a restrizioni temporali (e.g., diritto all'oblio).

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., D.lgs n. 196/2003 e Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati [130])

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."* Laddove norme specifiche consentono la pubblicazione di certe tipologie di informazioni, come ad esempio le norme sulla trasparenza come prima descritto (D.lgs. n. 33/2013) deve comunque essere assicurato il rispetto della disciplina in materia di protezione dei dati personali e delle Linee guida in materia di trattamento di dati personali emanate dal Garante per la protezione dei dati personali [130]. Si deve inoltre porre attenzione a non esporre quasi-identificatori (e.g., data di nascita, domicilio, residenza, sesso, razza, etnia, composizione nucleo familiare, status giuridico, ecc.) che possono con poco sforzo re-identificare i soggetti che si intende invece tutelare o che addirittura hanno una tutela speciale perché appartenenti a fasce particolarmente protette (e.g., testimoni giudiziari, profughi, rifugiati, pentiti, ecc.).

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 formato aperto¹³ 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.

¹³ 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.



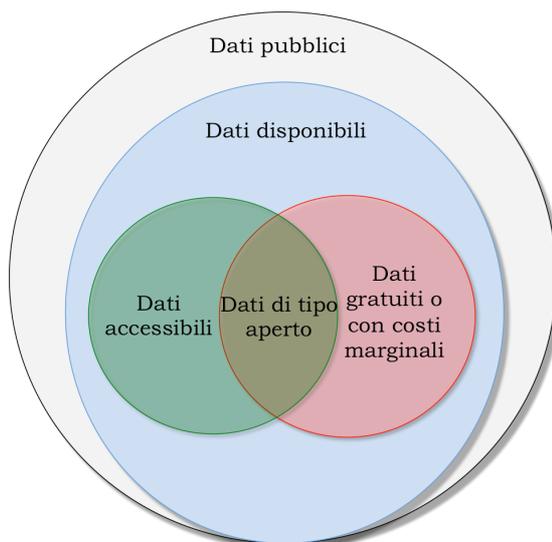


Figura 2: Tipi di dato delle pubbliche amministrazioni

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.

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.

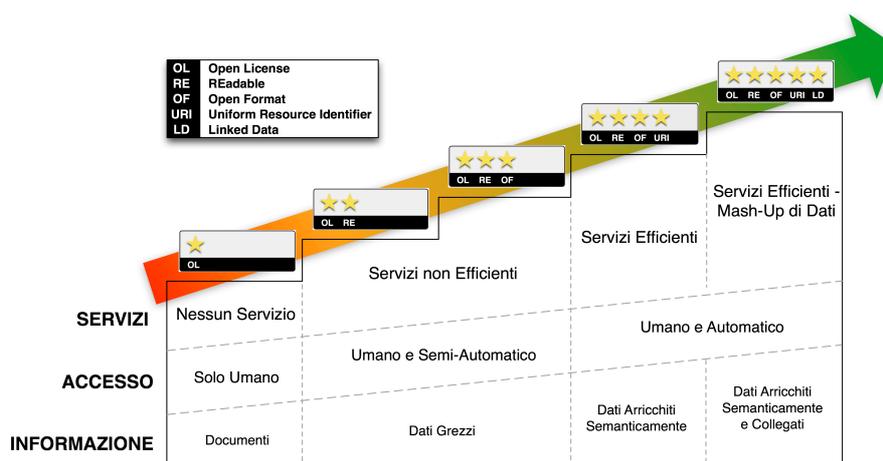


Figura 3: Modello a cinque livelli per la produzione e il rilascio di dati di tipo aperto

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 (e.g., CSV, geoJSON).



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 sinteticità del modello delle cinque stelle così come proposto da Tim Berners-Lee ha spesso portato a interpretazioni confuse dello stesso. Ad esempio, accade sovente che un file di dati avente come formato una delle serializzazioni RDF sia classificato come dato di livello cinque anziché di livello tre. A tale scopo, l'Agenda Nazionale propone un arricchimento dello stesso modello, come in mostrato in Figura 3, che fornisce ulteriori elementi per comprendere il livello di qualità offerto ai fini del massimo riutilizzo degli stessi. In particolare, La Figura 3 aggiunge alla figura nota e disponibile sul Web del modello delle cinque stelle [90] tre differenti dimensioni: *informazione*, *accesso* e *servizi implementabili*. 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 descritte 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 grazie a un intervento umano (livello 1);
- *dati grezzi (o semi-strutturati)*: 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 e programmi 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*: i programmi 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*: i programmi sono in grado di conoscere l'ontologia di riferimento e pertanto di elaborare i dati quasi senza ulteriori interventi umani (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*: servizi ad-hoc che usano i dati. Questi servizi devono incorporare al loro interno i dati per consentire un accesso diretto via Web agli stessi (livelli 2 e 3);
- *servizi e apps efficienti*: servizi, 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*: servizi, 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), facilitando il mashup di dati.

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 (principio 1). Conseguentemente tale URI può essere utilizzata per effettuare accessi diretti alle informazioni relative a quella entità (principio 2). I dati sono detti "linked" per la possibilità di riferenziarsi (i.e., "collegarsi") tra loro (principio 3). Nel riferenziarsi, 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 (principio 4). In questo modo, i Linked Data sono un metodo elegante ed efficace per risolvere problemi di identità e provenienza, semantica, integrazione e interoperabilità. In sostanza, triple RDF i cui URI non siano utilizzabili da un agente Web per recuperare le informazioni ad essi associati, non possono essere considerati Linked Data.



In sostanza, 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.

In Italia, l'utilizzo dei Linked Open Data sta continuamente crescendo con la presenza attualmente di diversi esempi virtuosi quali quelli, tra gli altri, della Camera dei Deputati [44], del Senato della Repubblica [99] e dell'Archivio Centrale dello Stato [100].

Per un approfondimento sui Linked Open Data, si può far riferimento 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.

Differentemente dai modelli esistenti, viene proposto un modello di classificazione della qualità dei metadati che è maggiormente rivolto allo sviluppo di servizi, quindi all'accessibilità dei metadati e al loro livello di granularità. 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.

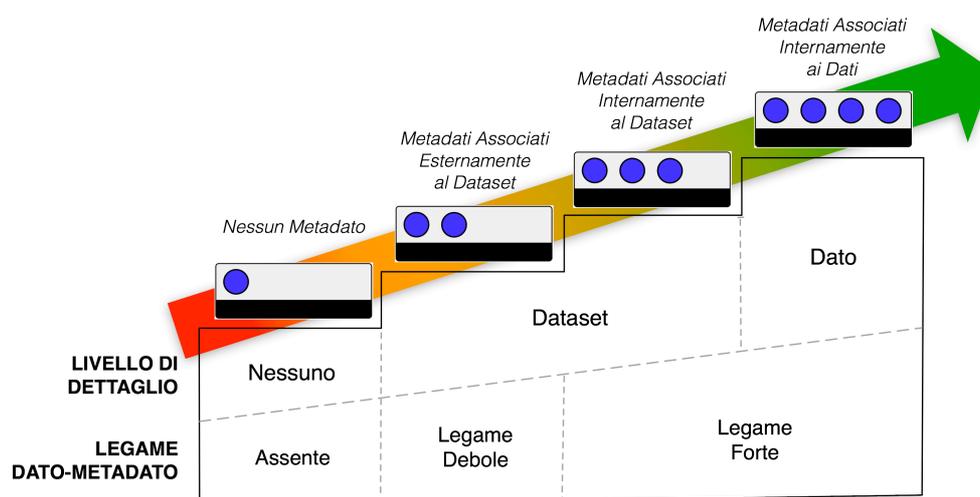


Figura 4: Modello per la definizione dei metadati



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 che li descrivono (livello 3 e 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 generali, e non dipendenti da specifiche tipologie di dati, che le pubbliche amministrazioni italiane dovranno utilizzare per accompagnare i dati che pubblicano. Nel caso specifico, ad esempio, dei dati territoriali/geografici, va considerato esclusivamente il provvedimento della Presidenza del Consiglio dei Ministri del 10/11/2011 con il quale è stato definito il profilo nazionale di metadati basato sugli Standard ISO 19115¹⁴, 19119¹⁵ e TS19139¹⁶, e le regole tecniche definite nell'ambito della Direttiva

¹⁴ http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020



INSPIRE [22]. Detto profilo nazionale è implementato attraverso il Repertorio Nazionale dei Dati Territoriali [29]; per un maggiore approfondimento si può fare riferimento alla documentazione normativa e tecnica disponibile nella sezione "Archivio documenti" del portale RNDT.

Si noti che le pubbliche amministrazioni possono estendere l'elenco dei metadati qui riportati con ulteriori metadati scelti sulla base di esigenze specifiche di documentazione dei dati rilasciati.

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. I metadati sono stati scelti a partire da metadati ben noti a livello internazionale. In particolare, 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 6.3.3 mentre esempi concreti sull'uso dei metadati inclusi in Tabella 2 e Tabella 3 sono inclusi in Appendice I.

Campo	Provenienza	Descrizione e note
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.
rightsHolder	DCMI	Indica il soggetto o l'organizzazione che detiene e gestisce i diritti sul dataset.
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. Il formato della data segue le specifiche dello standard ISO 8601, i.e., <i>yyyy-mm-dd</i>
accrualPeriodicity	DCMI	Indica la frequenza di aggiornamento dei dati.
license	CC	Indica la licenza utilizzata.
keyword	DCAT	Indica le parole chiave, separate da virgole, che descrivono il dataset

Tabella 2: Metadati obbligatori

¹⁵ http://www.iso.org/iso/catalogue_detail.htm?csnumber=39890

¹⁶ http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557



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.
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: solo attraverso il raggiungimento di un buon livello di ottimizzazione, le attività relative all'Open Data sono in grado di auto-sostenersi nel tempo;
- 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 omogeneo, auto-sostenibile e 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. Si noti che, affinché il lavoro del Team Open Data possa essere incisivo all'interno dell'amministrazione, è comunque importante far sì che tale team si confronti con il livello più politico dal quale ottenere le necessarie "spinte" per avviare un percorso concreto di apertura dei dati pubblici.

Responsabile Open Data (o Data Manager). All'interno del team Open Data è nominato un responsabile. Tale figura permette da un lato di localizzare le competenze necessarie alla gestione delle attività Open Data entro un sistema autonomo di comunicazione e funzionamento, e dall'altro di integrare i processi relativi alle attività di trasparenza in modo parallelo e non seriale. Il responsabile Open Data deve quindi possedere sia le capacità operative di controllo di tale sistema, sia quelle amministrative di coordinamento con i processi 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 e si coordina con il Responsabile della Trasparenza (quest'ultimo istituito ai sensi del D.lgs. n. 33/2013) al fine di rafforzare vicendevolmente gli obiettivi da un lato di massimo riutilizzo dei dati pubblici di tipo aperto e dall'altro di trasparenza. Il responsabile Open Data quindi fornisce le indicazioni:

- sulle risorse informative a disposizione dell'amministrazione definendo insieme al responsabile trasparenza una *roadmap* di produzione e pubblicazione condivisa dei dati per aree di sovrapposizione (e.g. bilanci, bandi, concorsi, organigrammi, spesa pubblica, ecc.);
- tecnico-operative (e.g., "formati machine-readable", modalità di gestione della pubblicazione dei dati) per il rilascio dei dati della trasparenza in formato Open Data, coordinandosi con il responsabile trasparenza per identificare tutti quei dati per i quali tale paradigma è applicabile;



- su eventuali difficoltà tecnico-organizzative nel reperimento e nel rilascio dei dati.

Responsabile 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 responsabile 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 Responsabile della Banca Dati l'approvazione per eventuali azioni correttive strutturali sul dataset.

Referente tematico della banca dati. Si tratta di un esperto di dominio che conosce in modo approfondito l'ufficio e la storia dei dati su cui l'ufficio opera. Spesso propone nuovi dataset da esporre a partire dal sistema gestionale corrispondente e cura eventuali valutazioni di dominio o relative al significato dei dati. Ha anche la possibilità di compiere bonifiche e semplici adeguamenti sulla banca dati, su segnalazione di cittadini o su valutazioni proprie. Riferisce invece al Responsabile della Banca dati la necessità di eventuali variazioni strutturali al sistema gestionale che insiste sui dati.

Ufficio Statistica. E' un anello importante dell'intera catena 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, comprese le problematiche di privacy, di finalità del dataset e di trattamento del dato personale ove presente nei dataset.

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.

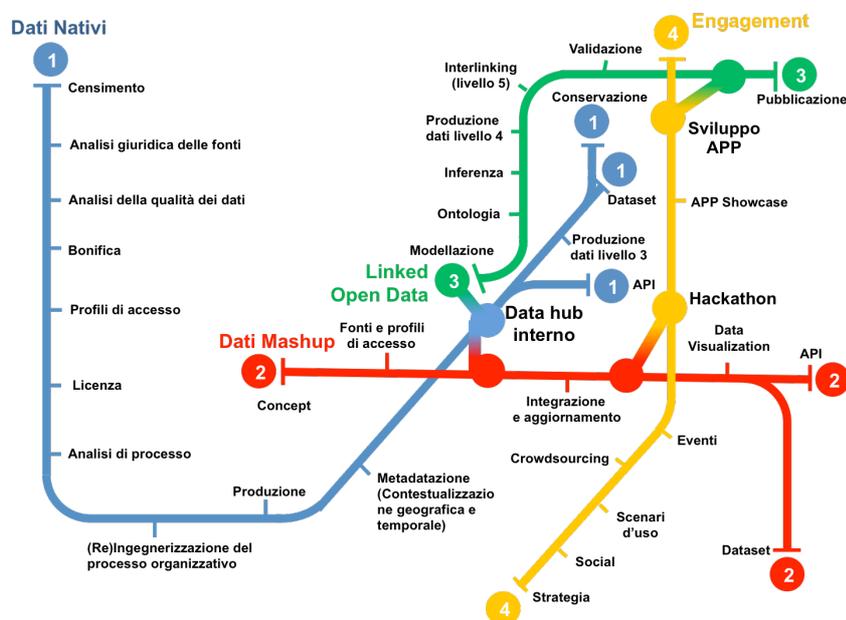


Figura 5: Modello operativo per la produzione e gestione dei dati aperti nella PA

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 dei dati esistenti generati dalle amministrazioni. Tutta la pubblica amministrazione, centrale o locale che sia, genera continuamente dati. Questi sono essenzialmente dati amministrativi, dati cioè che sono



prodotti dai vari uffici durante l'adempimento delle proprie funzioni istituzionali. Tanti di questi dati possono, senza particolari problemi, essere pubblicati come dati aperti e il loro rilascio costituisce un'ineludibile fonte potenziale di sviluppo per il territorio e per l'intero sistema Paese cui occorre dar seguito tempestivamente, attraverso uno specifico processo produttivo.

Poiché la produzione di dati è una conseguenza del fatto che un determinato ufficio adempie a determinate funzioni e gestisce uno o più processi amministrativi, è proprio all'interno dei singoli uffici o dei vari settori dell'amministrazione che vanno ricercate quelle che si possono chiamare "basi di dati primarie" oggetto della ricognizione. Di fatto si tratta di individuare quegli uffici che generano, mantengono e sono responsabili delle specifiche tipologie di dati che si vogliono rendere aperti.

Esempio: Lo sportello unico per le attività produttive (SUAP) del comune è lo strumento che va a semplificare tutti gli adempimenti connessi alla creazione, l'avvio, la modifica e la cessazione delle imprese per la produzione di beni e servizi. Sarà pertanto l'ufficio SUAP che gestisce e mantiene l'archivio con i dati di tutte le imprese del territorio.

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, alla luce della normativa vigente, in collaborazione con i responsabili delle basi di dati, al fine di individuare l'insieme di dati esistenti pubblicabili in formato aperto.

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' interessante notare che in quelle 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 giuridica delle fonti. Alla fase di censimento fa seguito l’analisi giuridica delle fonti del dato. L’analisi giuridica è un passo fondamentale per garantire sostenibilità nel tempo del processo di produzione e pubblicazione dei dati e creare un servizio equilibrato nel rispetto della funzione pubblica e dei diritti dei singoli individui. Come riportato dal D.lgs. n. 33/2013 la trasparenza deve creare la «condizione di garanzia delle libertà individuali e collettive, nonché dei diritti civili, politici e sociali, integra il diritto ad una buona amministrazione e concorre alla realizzazione di una amministrazione aperta, al servizio del cittadino». L’analisi giuridica delle fonti rappresenta il primo passo per valutare questi delicati equilibri, evidenziando limitazioni d’uso, finalità di competenza e per la determinazione dei diritti e dei termini di licenza. In sostanza, tale tappa del modello operativo risponde alle domande, fra le altre, “Di chi è il dato?” e “Quali sono le regole e i diritti che ne governano l’esposizione?”, “Quali limiti normativi devo tenere presente?”.

In Appendice III si riporta una breve “check list”, utile per verificare se tutti gli aspetti giuridici sono stati valutati dal responsabile della banca dati in collaborazione con il responsabile Open Data.

In generale, 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.

Analisi della qualità dei dati. All'analisi giuridica delle fonti segue l'analisi della qualità dei dati. Per la definizione del concetto generale di qualità si può ricorrere alla norma ISO 9000:2005¹⁷, secondo cui la qualità è la totalità degli elementi e delle caratteristiche di un prodotto o servizio che concorrono alla capacità dello stesso di soddisfare esigenze espresse o implicite.

Dalla stessa norma possono essere ricavate altre importanti relazioni inerenti ad una visione generale di gestione della qualità (quality management), data dalla sinergia tra il controllo e la garanzia di qualità. In particolare per controllo della qualità (quality control) si intende l'insieme delle tecniche operative e delle attività utilizzate per soddisfare i requisiti di qualità, mentre per garanzia di qualità (quality assurance) si intendono tutte le azioni pianificate e metodiche necessarie per fornire un'adeguata affidabilità che un prodotto o un servizio siano in grado di soddisfare i requisiti di qualità.

La quality assurance ha un'accezione prettamente preventiva, mentre il quality control verifica la conformità del prodotto durante e/o al termine del processo produttivo.

Per determinare la bontà dei dati è necessario definire delle metriche attraverso le quali misurare la qualità dei dati. Tuttavia, è molto difficile definire delle metriche universalmente valide in quanto la correttezza dei dati è profondamente legata ai singoli contesti operativi. La qualità del dato è un concetto multidimensionale la cui valutazione implica la definizione di metriche soggettive, adattabili ad un particolare contesto di business. E' comunque possibile tentare di definire delle metriche universali indipendenti dal contesto di utilizzo dei dati. Pertanto si possono individuare due tipologie di valutazione :

- Metriche di valutazione indipendenti dal contesto o oggettive: metriche che riflettono lo stato dei dati senza considerare come e dove vengono utilizzati;
- Metriche di valutazione dipendenti dal contesto o soggettive: misurazioni che tengono in considerazione il contesto di utilizzo, regole, caratteristiche e vincoli del business di riferimento.

Dei possibili indicatori per accertare la qualità dei dati indipendentemente dal contesto di utilizzo sono proposti da Thomas Redman. Redman propone due semplici indicatori in grado di determinare il

¹⁷ http://www.iso.org/iso/catalogue_detail?csnumber=42180



livello di correttezza di un insieme di dati:

- Correttezza a livello di attributi
- Correttezza a livello di record

Secondo Redman il livello di correttezza a livello di record è un buon indicatore di qualità della base di dati in quanto permette di identificare la percentuale di record che contengono degli errori. Tuttavia, senza tenere conto del contesto di utilizzo dei dati tali misurazioni potrebbero risultare falsate. Altre tipologie di metriche oggettive fanno uso di tecniche matematico-statistiche per determinare il livello di completezza e correttezza dei dati. Ad esempio è possibile utilizzare l'analisi dell'andamento temporale dei dati che permette di determinare gli scostamenti dal valore atteso e di identificare eventuali problematiche.

La definizione di metriche in grado di considerare il contesto passa dalla definizione di dimensioni attraverso cui valutare la qualità dei dati che includono, tra gli altri, l'*accuratezza*, la *completezza*, e l'*attualità*.

L'accuratezza consente di catturare errori sintattici (ad esempio un cognome italiano digitato in modo errato e non presente nel vocabolario dei cognomi italiani) o semantici (ad esempio un cognome errato perché, pur esistendo nel vocabolario dei cognomi italiani, viene attribuito ad un soggetto che in realtà ha un altro cognome).

La completezza è riconducibile all'assenza di valori, laddove invece ne è prevista obbligatoriamente la presenza. Non tutti i valori assenti, infatti, comportano incompletezza; ad esempio un indirizzo email può essere mancante e non essere causa di incompletezza, purché non etichettato come obbligatorio, oppure qualora il soggetto cui riferisce semplicemente non è in possesso di un indirizzo di email.

L'attualità cattura il grado di aggiornamento del dato, rappresentando pertanto l'accuratezza temporale del dato stesso. L'aggiornamento è un processo che può essere gestito in maniera manuale o automatizzata dipendentemente dalla tipologia di dato che si sta considerando: per dati altamente dinamici e variabili (e.g., dati provenienti da sensori) l'aggiornamento automatico è sicuramente da preferire, al contrario con dati di tipo statico (e.g., classificazioni ufficiali) è bene valutare lo sforzo rispetto ad un aggiornamento manuale. Da un punto di vista sostanziale la gestione di tale dimensione di qualità del dato può portare ad avere un processo di qualità a sistema: il dato, a partire dal lavoro di ordinaria amministrazione di qualsiasi ufficio, è incanalato in un processo di utilizzo trasversale da parte delle varie strutture dell'amministrazione, per poi confluire, compatibilmente con la suddetta analisi giuridica delle fonti, fra gli Open Data.

Le dimensioni elencate costituiscono un sottoinsieme, seppur significativo, delle tante dimensioni introdotte per caratterizzare il concetto di qualità del dato. 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.

Nonostante la presenza di uno standard ISO, 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"), che come l'attualità è legata alla cattura di aspetti temporali della qualità del dato, è definita in termini di "disponibilità di dati in tempo", oppure di "grado di aggiornamento dei dati", quindi secondo modalità evidentemente differenti (nel



secondo caso in modo coincidente con l'attualità).

Da un punto di vista operativo, tipicamente si seleziona un insieme di dimensioni di qualità ritenute “rilevanti” per la caratterizzazione di uno specifico dataset.

Una volta individuate tali dimensioni di qualità, il passo successivo è quantificarle in termini di misure, ed individuare delle soglie che consentano di discriminare la bontà o meno di un dato rispetto alla dimensione in esame.

Esempio: Riprendendo l'esempio di una città d'arte dove sono ubicati musei. Se il dataset che corrisponde alla lista dei musei è completo al 99%, la completezza può essere considerata accettabile se la lista deve essere pubblicata ai fini di una reperibilità turistica dei musei. Invece, se la lista deve essere usata per la distribuzione di fondi, naturalmente solo una completezza totale (100%) può essere ritenuta accettabile.

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

Bonifica. Il risultato dell'analisi della qualità del dato può comportare una 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.

L'apertura dei dati può essere uno stimolo importante per la conduzione di attività mirate di bonifica. Oltre all'utilizzo del dato all'interno dell'amministrazione, infatti, si possono aprire nuovi scenari di utilizzo che giustificano maggiormente gli investimenti connessi al processo di bonifica.

Una possibile classificazione di tali processi distingue processi di bonifica “basati sui dati” da processi di bonifica “basati sui processi”.

I processi di bonifica “basati sui dati” prevedono che il dataset sia corretto in uno dei due seguenti modi: (i) confronto con il mondo reale (anche con attività economicamente onerose come contattare direttamente i soggetti che presentano errori di qualità nella base di dati e correggerli con loro) e (ii) confronto incrociato (matching) con altri dataset. Tali processi hanno il vantaggio di poter pervenire in termini relativamente brevi al risultato, ma lo svantaggio di non risolvere il problema alla radice. Infatti, in un arco temporale medio-lungo, i dataset potranno nuovamente presentare i problemi di qualità risolti.

I processi di bonifica “basati sui processi” hanno invece la caratteristica di analizzare le cause che hanno portato alla scarsa qualità del dato e di rivedere i processi di produzione del dato di modo tale da garantirne la qualità nel tempo. Se ad esempio si riscontra che la scarsa accuratezza di una base di dati deriva da un processo di data entry manuale, si può intervenire sul processo prevedendo una fase



di acquisizione automatica che minimizzi la possibilità di errore di acquisizione.

L'adozione di processi di bonifica "basati sui processi" ha dunque il consistente vantaggio di essere una strategia risolutiva. Tuttavia, è interessante notare come le politiche di apertura dei dati abilitino fortemente anche i processi di bonifica "basati sui dati".

La possibilità di confrontare dati tra varie amministrazioni o tra amministrazioni ed altri soggetti è un'importante opportunità per la bonifica dei dati. Se un dataset è (totalmente o parzialmente) fornito da più soggetti, le attività di riconciliazione dei valori del dataset possono essere condotte in modo più agevole.

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 8 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) solo in forma anonima, ossia ad un livello di aggregazione tale da impedire di identificare le persone cui i dati si riferiscono. 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.

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 fare in modo che la produzione di quel dato sia consolidata e diventi stabile, secondo la frequenza di aggiornamento e le modalità di rilascio adottate. Vanno quindi individuati non solo i dati nativi "grezzi" di partenza ma anche gli attori che concorrono alla prima produzione del dato, distinguendo chi è responsabile e titolare dello stesso e chi invece aggiunge altri elementi informativi nel processo produttivo. 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 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 di 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.

In generale, 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.

Si consiglia, in particolare, di porre attenzione ai seguenti due aspetti della metadattazione:

- **Contestualizzazione geografica.** Ove 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.
- **Contestualizzazione temporale.** La contestualizzazione temporale del dato è uno dei punti di maggior valore del processo Open Data. Spesso la pubblicazione nei siti non si stratifica nel tempo e viene sovrascritto. L’Open Data permette al contrario di creare serie storiche dello stesso dataset che possono fornire importanti informazioni sui fenomeni sociologici e demografici. Per questo occorre porre molta cura nel creare i metadati temporali facendo attenzione a differenziare accuratamente il momento di creazione del dataset, il momento di pubblicazione e l’intervallo temporale al quale il dataset si riferisce.



Esempio: Un dataset relativo a fatti storici può essere creato oggi, pubblicato fra un mese, dopo un adeguato controllo di qualità, e riferirsi al periodo storico del quarantennio scorso (e.g. eventi storici bloccati da accesso per motivi giudiziari). I metadati devono fotografare con rigore queste informazioni temporali per consentire agli utilizzatori di orientarsi adeguatamente nelle fonti.

L'esposizione corretta dei parametri temporali del dataset ha anche risvolti giuridici legati alla tempestività di correzione o di aggiornamento dei dati esposti per evitare possibili problemi. Per questo è altrettanto importante indicare la frequenza di aggiornamento e la data dell'ultima modifica effettuata, metadati infatti definiti come obbligatori dal modello dei metadati proposto (Tabella 2).

Data hub interno, produzione di livello 3, e pubblicazione. 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 dell'amministrazione nella loro versione rilasciata ufficialmente. Questa infrastruttura, una volta attivata e messa a regime, viene a contenere lo stato dell'arte del patrimonio informativo e costituisce un potente punto di riferimento, accessibile da parte delle autorità di accesso, secondo diverse modalità (a "tag" o "query").

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 a 3 stelle 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.

Ed è a seguito della produzione che attività di verifica sulla presenza di tutti gli elementi di accompagnamento dei dataset (descrizioni, licenze, note, ulteriori metadati, ecc.) possono essere condotte. Per facilitare tale verifica, recentemente l'Istituto Open Data inglese (ODI) ha lanciato l'iniziativa relativa ai certificati Open Data [101]. I certificati rappresentano uno strumento utile ad ottenere un'auto-certificazione sulla qualità dei dati prodotti e pubblicati. I certificati sono stati tradotti anche in italiano dal nodo dell'ODI di Trento¹⁸.

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

Le risposte alle domande del questionario producono un livello di certificazione che si distingue in quattro livelli: (i) grezzo (raw), che rappresenta un'ottima base per iniziare il processo di apertura dei

¹⁸ <http://digitalcommonslib.fbk.eu>



dati; (ii) pilota (pilot), dove gli utilizzatori dei dati possono ricevere maggior supporto; (iii) standard, che rappresenta il livello in cui i dati aperti che vengono pubblicati sono regolarmente corredati da un robusto supporto a cui rivolgersi, (iv) esperto (expert), ovvero un eccezionale esempio di infrastruttura informativa. Ciascun livello di certificazione offre poi dei suggerimenti per il miglioramento della qualità dei dati.

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, oltre che 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.

E' da notare che elementi di qualità del dato, come ad esempio il livello di aggiornamento richiesto, influenza le modalità di gestione del dato stessa, sulla pubblicazione come API o come dataset, sul lavoro di backoffice e sull'impatto sui sistemi informatici dai quali vengono erogati i dati o servizi. Le API hanno un apparente minor "effort" di gestione in quanto il dato che forniscono è sempre aggiornato, mentre la pubblicazione di dataset necessita di aggiornamento secondo una frequenza prestabilita. Per dati in tempo reale, è probabilmente opportuno pensare a un'esposizione tramite API: in questo caso anche la più piccola dilazione temporale può vanificare l'attendibilità del dato. Tuttavia, l'effort è sull'infrastruttura utilizzata per implementare l'API che potenzialmente può essere oggetto di decine di migliaia di accessi (anche attacchi) simultanei, e che quindi dovrà essere opportunamente dimensionata per evitare degrading o cadute del servizio dovuti all'eccesso di richieste.

In sostanza, per tenere conto di tutte le esigenze e le criticità è consigliabile utilizzare un approccio



misto di pubblicazione dataset/API e pubblicare come API sicuramente i dataset che necessitano di un aggiornamento in tempo reale e sviluppare una serie di meccanismi che automatizzano l'aggiornamento per dataset variabili e dinamici (ad aggiornamento periodico o comunque piuttosto frequente), alleviando dall'onere dell'aggiornamento manuale.

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 graduale 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, 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 10.3 e 10.4 per l'identificazione degli ulteriori requisiti da soddisfare per l'impiego di una soluzione Linked Open Data.

5.2.4. Linea 4: Engagement (Coinvolgimento)

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 prestabilite.

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.

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

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



★ ★ Inserire dati nel contesto – accompagnare i dati con una ricca documentazione ne permette un facile riutilizzo. Porli nel corretto contesto amplifica tale possibilità. E' il caso di una rappresentazione grafica dei dati che permette di veicolare l'informazione in maniera più efficace. Un ottimo esempio di implementazione di strategia di engagement di livello 2 viene dal progetto "OpenCoesione" del Dipartimento per lo Sviluppo e la Coesione Sociale (DPS). Il portale OpenCoesione presenta una grafica, corredata da una mappa e diagrammi, che permette di prendere visione, in maniera efficace, della distribuzione dei fondi sociali europei sul territorio italiano. L'applicazione permette inoltre di scaricare i dati sia nella loro totalità, sia nello specifico caso dei progetti presentati o nelle loro aggregazioni per categoria o amministrazione comunale/provinciale/regionale.

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

★ ★ ★ ★ Creare capacità, competenze e reti – in questo livello rientra la fase “scenari d'uso” prima introdotta: nel presentare i dati attraverso infografiche interattive si fornisce la possibilità ad un utente di capire al meglio i dati. Rimane importante però stimolare il riutilizzo organizzando, ove possibile, incontri formativi volti a spiegare i dati e/o a mostrare strumenti di pulizia, analisi, e visualizzazione. Tra gli esempi virtuosi di tali pratiche rientrano “School of Data” dell'Open Knowledge Foundation, i datalab promossi da ISTAT e “A scuola di OpenCoesione” del DPS-MIUR.

★ ★ ★ ★ ★ Collaborare su dati come una risorsa comune – in questo livello rientra la fase “eventi, hackaton e app showcase” prima introdotta: il rilascio dei dati, infatti, prevede cicli di feedback con una comunità di riferimento (spesso quella da cui si è partiti per aprire i dati) da cui trarne delle considerazioni e produrre nuovi dati e strumenti. Nuovamente, l'esempio di OpenCoesione fornisce iniziative virtuose di engagement a cinque stelle quali hackaton organizzati con la comunità e il progetto monithon.it dove, attraverso segnalazioni partendo dai progetti presentati nel sito di OpenCoesione, chiunque può riportare informazioni aggiuntive per stimolare evoluzioni dei progetti finanziati). Ulteriori esempi sono quelli forniti da OpenStreetMap (OSM) e Wikidata.

Utilizzando i ruoli identificati precedentemente, per ciascuna linea del modello operativo è possibile tener traccia di una matrice di Ruoli e Responsabilità (RACI)¹⁹ che garantisca con chiarezza il flusso e la comunicazione tra le diverse figure identificate. Tale matrice è rappresentata nella seguente tabella:

¹⁹ A Guide to the Project Management Body of Knowledge (PMBOK Guide). PMI Standards Committee, Project Management Institute. 2010. ISBN 1-933890-66-5.



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

Tabella 4: Matrice RACI

Dove la denominazione dei ruoli è riportata come segue:

Responsible (R): Coloro che lavorano per eseguire un determinato compito. Esiste almeno un ruolo di responsabile.

Accountable (A): Il solo che può approvare il corretto completamento di un compito e che delega il lavoro a coloro responsabili. Può esistere un solo ruolo accountable per uno specifico compito.

Consulted (C): Coloro che possono essere consultati in quanto esperti di dominio e con i quali instaurare una comunicazione bidirezionale.

Informed (I): Coloro che devono essere tenuti aggiornati sui progressi del processo, spesso al termine dello stesso.

5.3. Coordinamento tra livello nazionale e livello locale

Diverse pubbliche amministrazioni centrali, al fine di adempiere a specifici obblighi normativi a loro assegnati, hanno necessità di raccogliere dati provenienti dal livello di governo locale. Si pensi ad esempio al caso del Ministero dell'Istruzione, Università e Ricerca (MIUR) che deve ottenere i dati sulla sicurezza degli edifici scolastici dalle diverse Regioni Italiane, o il caso dell'ISTAT per le rilevazioni relative ai censimenti o ai numeri civici. Vi sono inoltre altre amministrazioni che operano quasi esclusivamente sulla base di tale modello, come il caso del Dipartimento della Protezione Civile della Presidenza. Lo stesso scenario si può presentare anche in presenza di adempimenti locali che coinvolgono un'interazione tra Regioni e Comuni.

In questi casi, se il processo di produzione e rilascio di Open Data non è opportunamente coordinato tra i diversi livelli di governo, il rischio è di vedere una pubblicazione disomogenea degli stessi dati, vanificando i possibili benefici relativi al rilascio secondo determinati livelli di qualità che tipicamente agevolano il riutilizzo.



Per evitare tali situazioni, nei casi di analoghe tipologie di dati raccolti e gestiti sia dal livello centrale che locale, si raccomanda di avviare attività di coordinamento, con un eventuale supporto tecnico dell'Agazia su richiesta, che mirino, tra gli altri, a:

- identificare schemi di dati comuni per il livello centrale e locale, tenendo in considerazione il riuso di vocabolari/ontologie note a livello internazionale (sezione 6), evitando di ridefinire, per dati di uso diffuso, schemi che già sono stati definiti e standardizzati;
- identificare l'insieme minimo di dati rilasciati dal livello centrale e i dati che il livello locale può ulteriormente dettagliare per cogliere le specificità della propria realtà locale, abilitando ove possibile meccanismi di collegamento tra i due insiemi. Questo consentirebbe di avere una vista nazionale e un unico punto di accesso centrale ai dati, e una vista locale e più specializzata offerta dal governo locale. Si noti che il paradigma dei Linked Open Data può essere particolarmente conveniente in questi casi in quanto il collegamento degli URI consentirebbe un'agevole navigazione, da parte di programmi, nel Web dei dati così prodotto;
- documentare sia a livello centrale che locale i dati secondo il modello per i metadati proposto in sezione 4.2, specificando le diverse frequenze di aggiornamento dei dati che potrebbero variare a seconda della vista nazionale o locale esposta. Ove possibile, è consigliabile fornire tutte quelle informazioni necessarie per certificare la provenienza del dato (tutte le attività di raccolta e gestione dei dati e i soggetti preposti a tali attività, così come descritto nella successiva sezione 7) al fine di agevolare i possibili utilizzatori nel comprendere le diverse fasi di gestione complessiva del dato.

Fermo restando quanto indicato precedentemente, è auspicabile il riutilizzo coordinato dei dati tra amministrazioni centrali e locali attraverso il paradigma dell'Open Data, dove ciò riguardi strettamente dati pubblici e non riferibili a singole persone. Tale pratica consente, nel medio-lungo periodo, di abilitare nuove forme di cooperazione tra pubbliche amministrazioni realizzando una concreta efficacia dell'azione amministrativa.



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 comunque 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 e riferibili da tutti.

Tecnicamente, RDF è un framework concettuale che consente, sfruttando l'identificazione delle risorse con la modalità appena illustrata, di descrivere tali risorse e metterle in relazione tra loro. RDF ha un solo costrutto informativo di base, la cosiddetta tripla <oggetto> <predicato> <oggetto>. 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 le risorse vengono descritte tramite delle relazioni aventi un significato ben preciso e inserite in un particolare contesto fatto di dati, metadati e altre risorse.

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, focalizzata maggiormente sul contenuto informativo, è detta Web Semantico.

RDF è un framework concettuale e può essere implementato tramite diverse forme sintattiche, quali RDF/XML, Notation3, N-Triple, Turtle e JSON-LD (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 inter-traducibili. Esiste inoltre la possibilità di poter includere informazioni RDF all'interno delle pagine Web mediante il formalismo RDFa [102] (RDF in Attributes).

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 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 i cosiddetti ragionatori automatici per le logiche descrittive 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. È altresì possibile specificare operazioni di manipolazione dei dati, come ad esempio istruzioni di insert, update e delete.

SPARQL non è solo un linguaggio di interrogazione ma è un protocollo completo per l'accesso ai dati in quanto definisce anche le modalità con cui le query possono essere convogliate via Web (appoggiandosi al protocollo HTTP) e come i risultati devono essere restituiti all'utente. I servizi Web che implementano il protocollo SPARQL sono detti SPARQL endpoint.

Web (appoggiandosi al protocollo HTTP) le interrogazioni verso servizi Web detti SPARQL endpoint.

SDMX (Statistical Data and Metadata eXchange) [16]. È uno standard ISO per lo scambio di dati statistici basato su sintassi XML. Esso implementa al suo interno un modello dati per la



rappresentazione di dati multidimensionali, 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

I formati ampiamente usati per i dati di tipo aperto sono illustrati sinteticamente nel seguito. Alcuni di questi sono approfonditi attraverso un esempio concreto in Appendice I.

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 più specifici per l'annotazione di documenti (e.g., XBRL per la rappresentazione dei bilanci, Normattiva per la rappresentazione di documenti informatici in ambito giuridico, ecc). 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 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 esulano dall'uso stretto 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.

N-Triples [58]²⁰. N-Triples è una serializzazione di RDF in cui ogni tripla è espressa interamente (i.e., non in modo sintetico) e indipendentemente dalle altre. La concatenazione delle triple di un dataset RDF secondo N-Triples avviene mediante il carattere punto (i.e., <oggetto1> <predicato1> <oggetto1> . <oggetto2> <predicato2> <oggetto2>).

²⁰ N-Triples è un subset di Turtle e di N3: ogni riga rappresenta una tripla del tipo "<oggetto> <predicato> <oggetto>.". In questo modo il formato però perde in parte di leggibilità, ma risulta particolarmente semplice da utilizzare in streaming (l'ordine di trasmissione delle triple può essere in molti casi ignorato poiché la risorsa è correttamente ricostruita all'interno del triplestore), o per import/export massivi di dataset.



JSON (JavaScript Object Notation) [57]. E' un formato aperto per la rappresentazione e lo scambio di dati semi-strutturati che risulta leggibile anche dagli umani e che mantiene, rispetto a formati simili come l'XML, una sintassi poco prolissa. Questo aspetto ne fa un formato flessibile e compatto. Esso nasce dalla rappresentazione di strutture dati semplici nel linguaggio di programmazione JavaScript, ma mantiene indipendenza rispetto ai linguaggi di programmazione. Ha avuto ampia diffusione per il fatto di essere flessibile e compatto.

JSON-LD [104]. E' un formato di serializzazione per RDF, standardizzato recentemente dal W3C, che fa uso di una sintassi JSON. Viene proposto come formato per Linked Data, mascherando di proposito la sua natura di serializzazione di RDF per ragioni di diffusione del formato. Il gruppo di lavoro che l'ha definito ha posto come obiettivo, oltre quello di mettere a disposizione un'ulteriore funzionalità al framework RDF, anche quello di avvicinare il mondo dello sviluppo Web e degli utilizzatori dei sistemi di gestione dati NoSQL (in particolare dei document store) al Web Semantico. L'idea alla base del lavoro su tale formato è quella di favorire l'engagement verso diverse tipologie di utenti sviluppatori che finora sono risultati un po' scettici sull'uso dei Linked Data, per ridurre il gap esistente nell'integrazione di RDF in contesti più generali.

Da un punto di vista pratico è possibile rilasciare dati RDF utilizzando questo "dialetto" JSON nelle situazioni in cui non siano inizialmente sostenibili le risorse per dotarsi di tecnologie ad-hoc come i Triple stores. Allo stesso tempo, con JSON-LD si fornisce uno strumento standard che consente il collegamento di documenti JSON che per loro natura sono unità di informazione indipendenti. Si noti che data la recente proposta dello standard, i tool esistenti che mettono a disposizione un servizio di SPARQL endpoint, ancora non supportano nativamente JSON-LD.

E' stato messo a disposizione anche un sito di esempi e risorse per JSON-LD [103]. Il formato prevede dei "node object" utili all'identificazione di elementi chiave per RDF come "context", "type" e "namespace". Si riporta a titolo di esempio un documento JSON-LD:

```
{
  "@context": {
    "name": "http://xmlns.com/foaf/0.1/name",
    "homepage": {
      "@id": "http://xmlns.com/foaf/0.1/homepage",
      "@type": "@id"
    }
  },
  "name": "Manu Sporny",
  "homepage": "http://manu.sporny.org/"
}
```

CSV (Comma Separated Values). È 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, tipicamente, come indica il nome stesso, la virgola. Il CSV non è uno standard vero e proprio ma la sua modalità d'uso è descritta nell'RFC 4180 [17]. Il W3C ha recentemente istituito un gruppo "CSV on the web" [105] al quale partecipano, tra gli altri, esperti di Google e dell'Open Data Institute (ODI). L'iniziativa mira a produrre entro il 2015 uno standard attraverso cui formalizzare un'opportuna sintassi per l'introduzione dei metadati all'interno dei CSV, e fornire suggerimenti per la conversione verso vari formati, anche RDF, arrivando quindi a un analogo "Linked CSV", come proposto da Jeni Tennison di Open Knowledge Foundation.

6.2.2. Formati per i dati geografici

Shapefile [59]. È il formato standard de-facto per la rappresentazione dei dati dei sistemi informativi geografici (GIS). I dati sono di tipo vettoriale. Lo shapefile è stato creato dalla società privata ESRI che rende comunque pubbliche le sue specifiche. L'apertura delle specifiche ha consentito lo sviluppo di diversi strumenti in grado di gestire e creare tale formato.

Seppur impropriamente ci si riferisca a uno shapefile, nella pratica si devono considerare almeno tre file: un .shp contenente le forme geometriche, un .dbf contenente il database degli attributi delle forme geometriche e un file .shx come indice delle forme geometriche. A questi tre si deve anche accompagnare un file .prj che contiene le impostazioni del sistema di riferimento.

È importante notare che il dibattito su dove collocare tale formato rispetto al modello a stelle proposto in sezione 4 è ancora aperto; la comunità geomatica, seppur conscia del fatto che il formato è prodotto da una società privata, lo considera in ogni caso aperto (e quindi coerente con la definizione introdotta dall'art. 68 del CAD), di livello tre secondo il modello per i dati proposto dal presente documento²¹. Questo è dovuto al fatto che i dati possono essere gestiti attraverso una serie di strumenti non necessariamente confinati a determinate tipologie software (grazie alle specifiche tecniche dello shapefile aperte e pubbliche rese disponibili da ESRI). Altre comunità lo identificano come un formato proprietario e quindi di livello due secondo il modello proposto in sezione 4.

Il gruppo di lavoro, tenuto conto anche dell'ampio uso di tale formato per la rappresentazione dei dati geografici ha ritenuto importante includerlo in questo elenco, supportando più la visione, prima descritta, della comunità geomatica.

KML [106]. È un formato basato su XML per rappresentare dati geografici. Nato con Google, è diventato poi uno standard OGC. Le specifiche della versione 2.2 presentano una serie di entità XML attraverso cui archiviare le coordinate geografiche che rappresentano punti, linee e poligoni espressi in coordinate WGS84 e altre utili a definire gli stili attraverso cui i dati andranno visualizzati. Eventuali attributi delle geometrie va espressa invece attraverso la personalizzazione di alcune entità. Molti strumenti di conversione non si occupano tuttavia di creare questa struttura dati e delegano gli attributi delle geometrie allo stile di visualizzazione. Si consiglia pertanto di distribuire questo dato prestando attenzione o, eventualmente, accompagnando il dataset assieme ad un altro formato aperto per i dati geografici (ad esempio, .shp, .geojson).

Spesso è distribuito in modalità compressa attraverso file con estensione .kmz.

²¹ <http://gfoss.it/drupal/opendata>



GeoJSON [107]. E' un formato aperto per la rappresentazione e l'interscambio dei dati territoriali in forma vettoriale, basato su JSON (JavaScript Object Notation). Ogni dato è codificato come oggetto che può rappresentare una geometria, una feature o una collezione di feature. Ad ogni oggetto è associato un insieme di coppie nome/valore (membri). I principali nomi di membri che rappresentano le caratteristiche dei dati geografici sono: "type" che serve ad indicare il tipo di geometria (punto, linea, poligono o insieme multi-parte di questi tipi); "coordinates" attraverso cui sono indicate le coordinate dell'oggetto in un dato sistema di riferimento; "bbox" attraverso cui sono indicate le coordinate di un riquadro di delimitazione geografica; "crs" (opzionale) per l'indicazione del sistema di riferimento. È possibile inoltre associare all'oggetto specifici attributi attraverso il membro con nome "properties". Si tratta di un formato molto diffuso e supportato da diversi software, ampiamente utilizzato in ambito di sviluppo web.

GML (Geography Markup Language) [108]. E' una grammatica XML che rappresenta un formato di scambio aperto per i dati territoriali. Essa, definita originariamente da OGC e diventata lo Standard ISO 19136:2008, fornisce la codifica XML (schemi XSD) delle classi concettuali definite in diversi Standard ISO della serie 19100 e di classi aggiuntive appositamente definite: geometrie, oggetti topologici, unità di misura, tipi di base, riferimenti temporali, feature, sistemi di riferimento, copertura.

GeoPackage [129]. E' un formato aperto per la rappresentazione di dati geografici che può essere considerato un'alternativa al formato shapefile prima introdotto. Esso supporta Spatialite ovvero un'estensione dello schema del database SQLite. Il principale vantaggio offerto da GeoPackage è quello di poter rappresentare in un unico file diversi dati geografici, sia di tipo vettoriali che raster, che possono essere gestiti anche tramite apposite query SQL. Lo standard è riconosciuto dall'Open Geospatial Consortium (OGC)²².

6.2.3. 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 evitare di utilizzare formati come PDF o ODT ma si consiglia di 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. Nel caso di documenti in generale, si raccomanda, ove esistano, standard XML documentali internazionali o nazionali.

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

²² <http://www.opengeospatial.org/standards/geopackage>



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.).

Akoma Ntoso [109]. E' un linguaggio basato su XML per la rappresentazione di documenti giuridici. Attualmente è in fase di approvazione presso il consorzio OASIS. È utilizzato dal Parlamento Europeo e dalla Commissione Europea come standard documentale per i documenti legislativi, giuridici e allegati tecnici.

6.3. Ontologie e vocabolari

Questa sezione presenta una selezione non esaustiva 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'OKFN²³).

L'uso consolidato di tali ontologie e vocabolari, noti e condivisi a livello internazionale/europeo, ha l'obiettivo di far sì che, all'interno della pubblica amministrazione, dataset differenti ma di una stessa tipologia abbiano uno schema comune e siano pertanto rappresentati in maniera omogenea.

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

²³ <http://lov.okfn.org/dataset/lov/>



potenzialmente è possibile usare uno stesso software) e dataset differenti possono essere inseriti all'interno di uno stesso database senza operazioni di conversione intermedie. Si noti che le stesse ontologie vengono fornite in modalità Open Data.

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

Per ciascuna ontologia si propone, ove applicabile, l'indicazione delle basi di dati, in particolare di interesse nazionale, in cui la stessa può essere riutilizzata.

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.). Tale ontologia si presta quindi a essere utilizzata nel caso, per esempio, di apertura di parte della base di dati di interesse nazionale sui contratti pubblici.

PAYMENT ONTOLOGY [62]. E' un'ontologia sviluppata per rappresentare informazioni di spesa relative ad organizzazioni (non necessariamente pubbliche amministrazioni). Il nucleo dell'ontologia è rappresentato dal concetto di Pagamento che è associato a un soggetto che paga (e.g., un'autorità



locale che deve effettuare un pagamento), uno che riceve (e.g., il fornitore che riceve il pagamento a seguito dell'erogazione di una prestazione) e la data. Il pagamento può inoltre ulteriormente descritto attraverso una serie di proprietà opzionali che includono, la specifica unità organizzativa che eroga il pagamento, l'insieme di prodotti e/o servizi oggetto del pagamento e eventuali link a risorse che documentano il processo.

Tale ontologia si presta quindi a essere utilizzata nel caso di rappresentazione di informazioni che riguardano le spese e i pagamenti delle pubbliche amministrazioni per esempio.

FOAF (Friend of a Friend) [63]. E' un'ontologia che consente di rappresentare dati riguardanti le persone e le loro relazioni. Essa può quindi essere utilizzata per la rappresentazione di alcuni dati inclusi in basi di dati di interesse nazionale che riguardano le persone quali "Anagrafe della Popolazione Residente – ANPR", l'Anagrafe Nazionale dell'Assistiti (ANA), cataloghi della piattaforma delle comunità intelligenti.

WAI (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]. E' stato altresì proposto al Comitato Tecnico delle comunità intelligenti per la rappresentazione delle informazioni relative al catalogo dei servizi della piattaforma nazionale delle comunità intelligenti.

CLV (Core Location Vocabulary) [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. Tale vocabolario si presta quindi ad essere utilizzato per rappresentare le principali informazioni che costituiscono per esempio l'Archivio Nazionale dei Numeri Civici e delle Strade Urbane (art. 3 del DL 179/2012).



CPV (Core Person Vocabulary) [67]. E' un modello dati per la rappresentazione delle informazioni sulle persone. Nasce per essere ben integrato con gli altri vocabolari "core" della commissione europea e fa largo riuso di schema.org e FOAF. Come nel caso di FOAF, tale vocabolario può essere utilizzato per rappresentare alcune informazioni presenti nella base di dati di interesse nazionale "Anagrafe della Popolazione Residente – ANPR" e "Anagrafe Nazionale Assistiti - ANA". Attualmente il vocabolario è utilizzato all'interno del progetto europeo e-CODEX – 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 ampiamente utilizzata nella versione Linked Open Data dell'Indice nazionale della Pubblica Amministrazione italiana (IPA). Il vocabolario è multilingua ed è stato recentemente tradotto anche in italiano dall'Agenzia per l'Italia Digitale.

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 come vocabolario base per la rappresentazione dei dati della base dati di interesse nazionale "Registro delle imprese" (art. 60 del CAD).

OWL-S (Semantic Markup for Web Service) [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 IETF²⁴ (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

²⁴ <http://www.ietf.org/rfc/rfc2426.txt>



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 SKOS²⁵. Strettamente connesso al tesoro Eurovoc è il progetto Eur-lex²⁶ per l'accesso al diritto nell'Unione Europea. Di tale progetto esiste una versione RDF interrogabile via SPARQL²⁷.

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 principalmente sulla base di alcuni dei settori di competenza delle amministrazioni coinvolte nel rilascio di dataset chiave individuati nell'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico.

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.

GEMET (General Multilingual Environmental Thesaurus) [95]. E' un dizionario terminologico multilingue realizzato a livello europeo dall'agenzia europea per l'ambiente "EEA" che si avvale della rete "EIONET" per rendere fruibile il contenuto dei documenti relativi alle attività ambientali. GEMET nasce con la funzione principale di strumento di gestione informativa del Centro Tematico Europeo per il Catalogo delle Fonti Dati) e ha origine dalla fusione del contenuto dei migliori tesauri realizzati in ambito europeo al fine di ottenere una base condivisa per la descrizione delle informazioni. Si tratta di un database terminologico specificatamente orientato ai temi del territorio e dell'ambiente. è dotato di struttura semantica interna ed è multilingue (ad oggi 33 lingue). Le tavole di definizione (e i dati) sono scaricabili²⁸.

ThIST (Thesaurus Italiano di Scienze della Terra) [96]. Il tesoro è il risultato di una pressoché totale revisione del tesoro nato dall'integrazione tra la banca dati terminologica dell'allora Biblioteca

²⁵ <http://eurovoc.europa.eu/drupal/?q=it>

²⁶ <http://eur-lex.europa.eu/content/welcome/about.html>

²⁷ <http://eur-lex.publicdata.eu/sparql>

²⁸ <http://www.eionet.europa.eu/gemet/rdf?langcode=it>



del Servizio geologico d'Italia (oggi Biblioteca APAT) e il tesoro pubblicato dal CNR nel 1997. È strettamente connesso al Multilingual Thesaurus of Geosciences (MTG), curato dal Gruppo di lavoro internazionale MTG, all'interno del quale l'Italia è rappresentata dall'APAT. Viene già utilizzato per l'indicizzazione del materiale biblio-cartografico (monografie, articoli di periodici, carte) posseduto dalla Biblioteca e il recupero delle informazioni relative.

Allo stato attuale il ThIST presenta 10350 descrittori, 14 collegati tra loro da 103850 relazioni di tipo gerarchico, associativo e di equivalenza. Ad ogni descrittore è affiancata la corrispondente traduzione in inglese così da rendere il ThIST un tesoro bilingue.

I descrittori sono generalmente sostantivi o sintagmi con valore di sostantivo; gli aggettivi, i verbi e gli avverbi sono esclusi dal corpo del tesoro, in ottemperanza alla norma ISO 278815. Si è preferita la forma singolare, a eccezione di alcuni casi specifici in cui è utilizzato il plurale. Al contrario, per i descrittori in lingua inglese si è utilizzata la forma plurale in conformità al modello GeoRef TH, da cui molti di essi provengono. Quanto alla struttura, si è mantenuta al momento quella del tesoro CNR di provenienza, basata su una classificazione di tipo tematico-disciplinare comprendente 56 categorie.

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 per gestire i requisiti specifici per la pubblicazione, da parte degli Istituti Nazionali di Statistica, dei dati della produzione di statistica ufficiale. Ad esempio, XKOS introduce i "livelli" delle classificazioni statistiche: tali livelli, identificando i livelli di aggregazione nelle statistiche pubblicate, sono fondamentali sia per garantire la consistenza dei dati prodotti sia per controllare i rischi di violazione della riservatezza.

In Appendice II è incluso un esempio di applicazione di XKOS per la rappresentazione in RDF della classificazione ATECO2007.

6.3.2.3. Dati culturali

Tesoro PICO [76]. E' il vocabolario controllato utilizzato nel portale della cultura italiana²⁹ ma è sufficientemente generale per essere usato nella descrizione di dati relativi al dominio dei dati culturali.

²⁹ <http://www.culturaitalia.it/>



Tesaurus UNESCO [77]. E' un tesaurus per i documenti e le pubblicazioni sulle attività dell'UNESCO. In particolare contiene i termini usati negli ambiti dell'istruzione, cultura, scienze naturali, scienze sociali, comunicazione e informazione. Ne esiste una versione modellata attraverso SKOS³⁰.

CIDOC-CRM [78]. E' un modello dati di riferimento per rappresentare le informazioni su concetti e relazioni nell'ambito della cultura (e.g., documenti di musei, archivi, biblioteche). Esiste già una versione in OWL³¹.

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.

6.3.2.4. Dati contabili

XBRL (eXtensible Business Reporting Language) [110]. E' un linguaggio basato su XML per la rappresentazione e lo scambio di informazioni contabili e finanziarie.

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]³². 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 standardizzato dal W3C utilizzato per metadattare i cataloghi di dati, i dataset contenuti in un catalogo e le diverse modalità di reperimento e accesso ai dati. E' stato impiegato per caratterizzare i profili applicativi dei portali di dati (DCAT-AP [111]). Attualmente è il vocabolario utilizzato dal catalogo nazionale dati.gov.it.

Dublin Core [81]. La versione "core" consiste di un vocabolario minimo di metadati (15 elementi)

³⁰ <http://skos.um.es/unescothes/>

³¹ <http://erlangen-crm.org/>

³² E' possibile descrivere graficamente il riutilizzo di vocabolari tramite una rappresentazione grafica della "linked data cloud", che può essere generata a partire da VOID, utilizzando strumenti come void-graph (<http://lmatteis.github.io/void-graph/>)



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 framework standardizzato dal W3C per la rappresentazione della provenienza di artefatti. Consente di modellare anche il processo di generazione di un artefatto in maniera quasi analoga ai sistemi di controllo versione. Per ulteriori dettagli sul modello e sull'uso concreto dello stesso si rimanda alla sezione 7 e all'Appendice II dove un esempio dell'uso del framework è proposto per la rappresentazione delle informazioni di provenienza relative alla classificazione ATECO2007 recentemente pubblicata in RDF dall'Agenzia per l'Italia Digitale in collaborazione con l'ISTAT.

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 e annotazioni su vocabolari. In particolare, esso è utilizzato per arricchirli con esempi e note di utilizzo.

VOAF (Vocabulary of a Friend) [112]. E' un vocabolario che consente di descrivere i vocabolari usati coi Linked Data. Ad esempio, esso consente di mettere in relazione diversi vocabolari o di indicare quali dataset ne fanno uso.

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. Data la sua natura enciclopedica, viene "linkato" parecchio all'interno del Web dei dati tanto da essere considerato l'hub principale.

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.

OpenStreetMap – OSM [113]. E' un progetto collaborativo, ispirato a Wikipedia, che ha il fine di creare una base dati georeferenziata. Gli output di OSM sono molteplici: mappe digitali online, oggetti classici del mondo della cartografia stampata, servizi di routing, geocodifica, mappe tematiche (es. ciclismo, escursionismo, accessibilità motoria, archeologia), ecc. Il progetto tutela la propria banca dati attraverso la licenza ODbL (si veda la sezione 8) che prevede l'obbligo di continuare a tenere aperta la banca dati citando la fonte e usando la frase "Copyright OpenStreetMap and contributors" nei prodotti derivati.



I dati sono disponibili attraverso varie modalità: download diretto dell'intera banca dati in formato XML, download di alcune classi convertite nei formati GIS più noti, download tramite API attraverso interrogazioni per area o categoria. La categorizzazione dei dati avviene tramite un processo bottom-up governato da una pagina wiki dove gli attributi sono definiti con più tag. Maggiori dettagli su applicazioni, tassonomie, api, si trovano nella sezione del wiki dedicato al progetto³³.

6.4.3. Dati sulla medicina

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://biportal.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.

³³ http://wiki.openstreetmap.org/wiki/Main_Page



7. TITOLARITÀ E PROVENIENZA DEI DATI

La pubblicazione dei dati sul Web, oltre a tutti gli evidenti vantaggi in termini di efficienza ed efficacia nella fornitura dei servizi ai cittadini, porta con sé un problema relativo all'attendibilità del dato, infatti chiunque può pubblicare dati, qualunque dato, sulla rete.

Nasce quindi la necessità di corredare i dati con dei metadati che ne descrivano la provenienza certificandone così, indirettamente, l'affidabilità, la qualità e l'attendibilità. Nel modello dei metadati proposto, si sono identificati già alcuni metadati che consentono di documentare la provenienza del dato di tipo aperto pubblicato (come ad esempio i metadati "publisher", "creator" e "rightsHolder"). Tuttavia, è possibile fornire ulteriori dettagli sulle entità e la filiera di attività che hanno portato alla generazione e pubblicazione del dato al fine di certificarne in maniera più accurata la reale provenienza.

A tal fine è importante fornire alcune definizioni utili per caratterizzare la provenienza e titolarità dei dati. Tali definizioni sono riportate nel seguito.

Responsabile del dato: il responsabile del dato è la persona/Ente/Istituzione/Amministrazione che gestisce/genera/manipola il dato. In alcuni casi, tale responsabile è individuato a livello normativo.

Dato certificato: un dato è certificato se è stato pubblicato dal responsabile del dato stesso.

Provenienza (provenance) dei dati: insieme di informazioni dettagliate relative alle entità e ai processi coinvolti nella produzione e nella pubblicazione dei dati. Le informazioni possono riguardare ogni elemento che ha impatto sulla risorsa dati pubblicata.

Entità: "oggetti" di cui si vuole descrivere la provenienza. Il termine "oggetto" comprende diversi concetti tra cui oggetti digitali, come file, pagine web, oggetti fisici, come montagne, case, oggetti astratti come idee e concetti.

Agente: qualcuno o qualcosa responsabile rispetto ad una attività in corso, rispetto all'esistenza di una entità o alla attività di un altro agente.

Attività: qualcosa che accade in un periodo di tempo e che agisce su o con delle entità. Tra le attività sono incluse il consumo, il processamento, la trasformazione, la modifica, l'utilizzo o la generazione di entità.

7.1. PROV Framework

Il PROV Framework [80] è uno standard W3C che consente di descrivere in maniera strutturata la provenienza di artefatti e quindi anche di dati che si intende pubblicare. Consente di modellare anche il processo di generazione di un artefatto in maniera quasi analoga ai sistemi di controllo versione.

In dettaglio, il PROV Framework è costituito da una famiglia di specifiche articolate nei seguenti componenti:

PROV-DM: descrive il modello concettuale dei dati ed è quindi il nucleo centrale della famiglia di



specifiche. Esso non fa riferimento ad uno specifico dominio ma è corredato da estensioni per domini più specifici. PROV-DM definisce quindi il vocabolario dei seguenti concetti base: (1) entità, attività e data di creazione, utilizzo o fine dell'artefatto, (2) derivazione di entità da entità, (3) agenti responsabili delle entità generate o delle attività avvenute, (4) notazione per la gestione della provenance dei dati di provenance, (5) proprietà per la connessione di entità che fanno riferimento allo stesso oggetto, (6) definizione di strutture logiche tra gli elementi (collezioni).

PROV-O: anche detto PROV Ontology, definisce l'ontologia OWL2 del PROV-DM in modo da poter essere utilizzato direttamente nel contesto del Web Semantico e dei Linked Data.

PROV-N: definisce una notazione fruibile da un utente umano per i dati di provenance creati col PROV.

PROV-XML: propone una serializzazione in XML del PROV-DM.

PROV Constraints: definisce un insieme di vincoli che devono essere soddisfatti affinché la propria istanza di dati di provenance sia consistente; tale verifica assicura il mantenimento del significato dei dati in caso di ragionamenti automatici sui dati o di altri tipi di analisi.

PROV Sem: definisce la semantica del PROV-DM in termini di logica del primo ordine.

PROV-Dictionary: propone un'estensione di PROV-DM per la descrizione dei dati di provenance nel caso di aggregazione delle informazioni in collezioni/dizionari.

PROV-Links: propone un'estensione del modello dei dati (PROV-DM) per la descrizione dei dati di provenance nel caso in cui si debba tracciare la provenienza di fonti di dati multiple.

PROV-AQ: propone alcune modalità di accesso ed interrogazione dei dati di provenance.

PROV-DC: propone la mappatura tra l'ontologia Dublin Core (DC) e l'ontologia PROV-O, consentendo di accedere contemporaneamente a dati di provenance espressi sia tramite DC che tramite PROV.

7.2. Un esempio di utilizzo del framework PROV

Nel seguito si riporta un esempio concreto di utilizzo del framework PROV nel contesto di una classificazione ufficiale recentemente pubblicata da AgID insieme all'ISTAT.

Eurostat definisce la Classificazione Ufficiale delle Attività Economiche (NACE) la quale è recepita dall'Istat nella Classificazione Ufficiale Italiana delle Attività Economiche (ATECO). Se viene apportata una modifica alla classificazione, viene generata una nuova "versione" ufficiale del documento delle classificazioni (i.e., ATECO2002, ATECO2007).

Da una ricerca sul Web si può osservare che oltre al sito dell'Istat esistono numerosi altri link contenenti tale classificazione. Ciononostante, l'Istat rimane il solo responsabile del dato e pertanto è opportuno che al momento della pubblicazione della classificazione ne certifichi la provenienza consentendo così all'utente di stabilire l'affidabilità, la qualità e l'attendibilità del dato a cui ha acceduto.

L'Agenzia e l'Ufficio Sviluppo Sistemi Informatici per la Statistica (RST/B) dell'Istat hanno pubblicato l'Ateco2007 in formato Linked Open Data sul portale SPCData [27] lavorando insieme alla modellazione in LOD di tale classificazione. In particolare, la Figura 6 mostra uno schema secondo il



modello PROV dell'ATECO2007 delle attività dei due soggetti coinvolti e le relative entità considerate e generate.

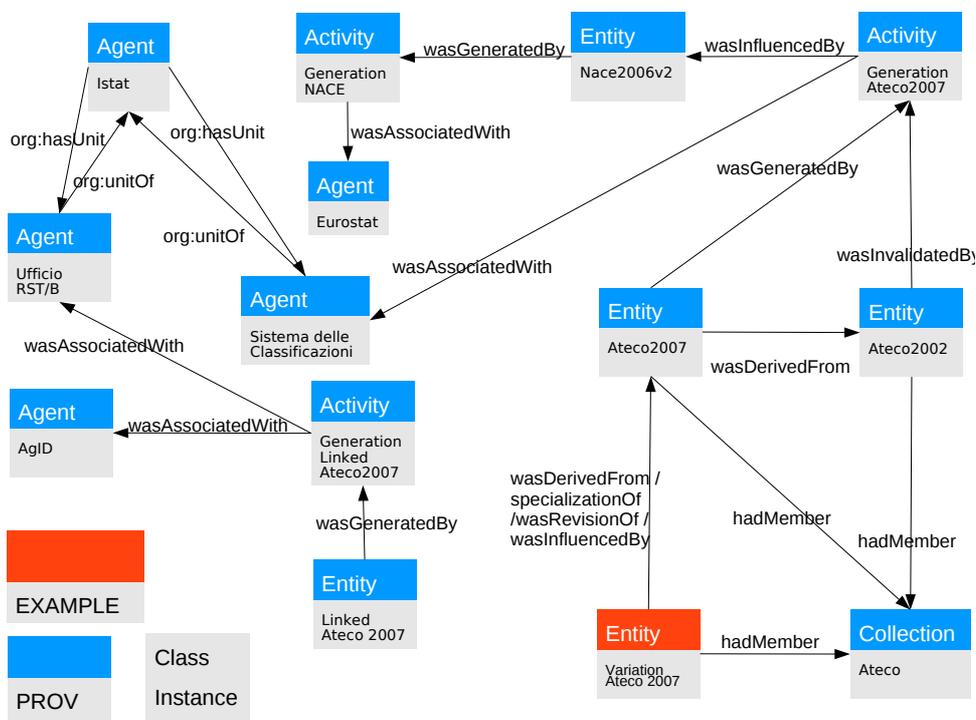


Figura 6: Il framework PROV per la generazione e pubblicazione della classificazione ATECO2007

In Appendice II è possibile trovare l'esempio completo dell'uso di PROV e di ulteriori descrizioni semantiche impiegate (i.e., DCAT, ADMS, DC e SKOS/XSKOS) per la rappresentazione in LOD della classificazione Ateco2007.

8. LICENZE ASSOCIATE AL DATO

Una fase importante del modello operativo riguarda la scelta della licenza da associare al dato. Tale scelta è cruciale per poter attuare pienamente il paradigma dei dati aperti in quanto i termini stabiliti dalla licenza definiscono l'uso che gli utenti possono fare dei dati, la paternità degli stessi e le modalità di creazione di eventuali lavori derivati. La licenza, e la relativa versione utilizzata, rientra quindi tra i metadati obbligatori minimi da fornire in fase di pubblicazione di dataset aperti (un esempio di specifica RDF dei metadati relativi alla licenza è presentata in Appendice I).

Facendo riferimento alla definizione Open Data fornita dall'Open Knowledge Foundation (OKF) per cui un dato è aperto se è "usabile, riutilizzabile e ridistribuibile liberamente da chiunque anche per finalità commerciali, soggetto al massimo alla richiesta di attribuzione e condivisione allo stesso modo", le sole licenze ammesse per abilitare l'effettivo paradigma dell'Open Data sono classificate come mostrato in Figura 7³⁴. Come evidenziato in figura, tutte le altre licenze che non consentono lavori derivati, anche per finalità commerciali (i.e., licenze che riportano chiaramente clausole Non Commercial - NC e/o Non Derivative - ND e/o ogni altra clausola che limita la possibilità di riutilizzo e redistribuzione dei dati) non possono essere ritenute valide per identificare dati di tipo aperto.

La Figura 7 mostra inoltre come tra gli strumenti legali per l'Open Data vi sia una chiara distinzione tra licenze e pubblico dominio.

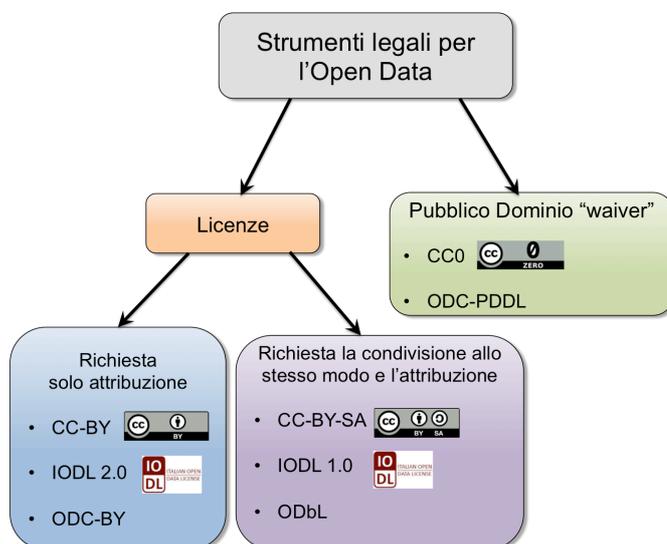


Figura 7: Strumenti legali per l'Open Data

Pubblico Dominio. Il pubblico dominio o “waiver” rappresenta uno strumento legale per l'Open

³⁴ La Figura 7 è una rivisitazione anche grafica dello schema proposto in [40].

Data attraverso il quale il dichiarante “*apertamente, pienamente, permanentemente, irrevocabilmente e 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 (includendo espressamente le pretese presenti come quelle future) relativa all'Opera*”. Pertanto, con il “waiver” è consentito:

- condividere, e quindi copiare, distribuire ed esporre al pubblico i dati;
- modificare e quindi trasformare ed eseguire “mashup” dei dati per generare lavori derivati

senza nessun tipo di restrizione e senza dover richiedere nessuna autorizzazione.

Rientra in tale categoria la CC0 della famiglia delle licenze internazionali Creative Commons [6] e la Open Data Commons – Public Domain Dedication License (ODC-PDDL) per i dati/database [117].

Un esempio di applicazione della licenza CC0 è fornito in [9].

Occorre tuttavia prestare attenzione nell'utilizzo della CC0, in particolare se i dati provengono da fonti che, a loro volta, hanno già definito una licenza. Non si può, infatti, disporre di diritti superiori rispetto alla licenza di partenza (e.g., non si può attribuire una CC0 ad un dataset ottenuto da una fonte a cui è associata una licenza di attribuzione). Infine, occorre ricordare che alla maggior parte dei dati e dei documenti necessari per lo svolgimento delle funzioni tipiche delle pubbliche amministrazioni non è opportuno applicare la CC0, in quanto questa prevede il rilascio dei diritti morali che sono inalienabili, indisponibili, imprescrittibili secondo le norme nazionali ed europee. In particolare poi tutti i beni del demanio culturale (art. 10 e art. 53, Codice dei beni culturali, D.lgs. 22 gennaio 2004, n. 42) per loro natura sono indisponibili e quindi è richiesta esplicitamente la menzione di informazioni “di attribuzione” (si veda sotto).

Licenze per l'Open Data. Sono ritenute valide nel contesto dei dati di tipo aperto le licenze con richiesta di attribuzione e con richiesta di condivisione allo stesso modo.

Le licenze con richiesta di attribuzione consentono di:

- condividere, e quindi copiare, distribuire, esporre al pubblico i dati;
- modificare, e quindi trasformare ed eseguire il “mashup” dei dati per la creazione di lavori derivati

anche per fini commerciali, con il solo vincolo di attribuire in maniera opportuna la paternità dei dati (si consiglia di indicare a tal proposito il nome dell'organizzazione unitamente all'URL della pagina Web dove si trovano i dati/contenuti da licenziare), fornendo un link alla licenza e indicando eventuali modifiche apportate.

Come illustrato in Figura 7, rientrano in tale categoria la licenza CC-BY della famiglia delle licenze internazionali Creative Commons [6], la IODL (Italian Open Data License) nella sua versione 2.0 [114] e la Open Data Commons Attribution License (ODC-BY) per dati/database [115].

Le licenze con richiesta di condivisione allo stesso modo consentono di:

- condividere, e quindi copiare, distribuire, esporre al pubblico i dati;



- modificare, e quindi trasformare ed eseguire il “mashup” dei dati per la creazione di lavori derivati

anche per finalità commerciali, nel rispetto comunque dei seguenti due vincoli:

- distribuire eventuali lavori derivati con la stessa identica licenza che governa il lavoro originale;
- attribuire in maniera opportuna la paternità dei dati, fornendo un link alla licenza e indicando eventuali modifiche apportate.

Come illustrato in Figura 7, rientrano in tale categoria la licenza CC-BY-SA della famiglia delle licenze internazionali Creative Commons [6], la IODL (Italian Open Data License) nella sua precedente versione 1.0 e la Open Data Commons Open Database License (ODbL) [32] ovvero la licenza utilizzata dal progetto OpenStreetMap – (OSM)³⁵.

E' importante notare che, recentemente, la famiglia delle licenze internazionali Creative Commons si è ulteriormente arricchita introducendo una nuova versione, la 4.0, che apporta importanti modifiche alle versioni precedenti. Tali modifiche possono essere riassunte come segue. La licenza CC 4.0 assume una valenza internazionale: sebbene si preveda la sua traduzione in molte lingue, essa non acquisisce più una giurisdizione locale e nazionale a seconda della lingua, ma mantiene una giurisdizione internazionale.

La CC 4.0 introduce inoltre il diritto *sui generis*, ossia il diritto che nel quadro europeo (direttiva 96/9/CE) tutela il costituente della banca dati (D.lgs. 169/1996 recepito poi nella L. 633/1941 art. 102 *bis* e art. 102 *ter*). Il costituente è definito dall'art. 102 *bis* L. 633/1941 come «*il soggetto che effettua investimenti rilevanti per la realizzazione di una banca di dati, ovvero per la sua verifica o presentazione, impegnando a tal fine mezzi finanziari tempo o lavoro*» e che ha il diritto di regolare il trasferimento e l'estrazione parziale o totale della stessa. Il diritto del costituente non esaurisce la titolarità del contenuto originale della banca dati. Pertanto, se il dataset elaborato da una pubblica amministrazione tratta dati di un altro ente, due diritti si sovrappongono: quello del costituente che ha investito tempo e risorse a creare sia il dataset che i metadati e quello dell'ente che ha prodotto i dati. Il diritto del costituente ha una durata di 15 anni a partire dal mese di gennaio successivo rispetto alla creazione del database e viene rinnovato ogni qual volta vi sia un significativo incremento dei dati. Poiché i dataset sono assimilabili alle banche dati, l'introduzione nella CC 4.0 di questo istituto rappresenta una svolta importante per attribuire pieni diritti di regolazione del riuso a tutte le pubbliche amministrazioni titolari del dataset che effettuano un affinamento, pulizia, conversione, qualificazione dei dati.

Infine, sempre all'interno della famiglia delle Creative Commons, sia per ciò che concerne le licenze con richiesta di attribuzione sia per quelle che prevedono la condivisione allo stesso modo, esiste una licenza appositamente creata per organizzazioni intergovernative (CC IGO) [116]. Tali organizzazioni tipicamente godono di particolari immunità rispetto alla giurisdizione nazionale; pertanto, la CC IGO estende le licenze CC 3.0 Unported aggiungendo alle autorizzazioni già presenti con la versione 3.0, le seguenti due specificità: “*quando il licenziante del materiale è un'organizzazione intergovernativa, eventuali dispute e violazioni alle condizioni della licenza saranno affidate agli strumenti della mediazione e dell'arbitrato; inoltre (come introdotto anche nella versione 4.0 delle licenze CC), in caso di violazioni alla licenza è prevista una procedura di reintegrazione automatica nel diritto ad utilizzare i materiali, nel caso tali violazioni siano sanate entro trenta giorni*” [116].

³⁵ per maggiori informazioni sull'uso della ODbL ci si può riferire a <http://openstreetmap.it/2014/03/in-quali-casi-vengono-attivate-le-clausole-di-condivisione-allo-stesso-modo-share-alike/>



Compatibilità tra licenze. Un altro aspetto di grande rilievo per ciò che riguarda le licenze Open Data è la definizione delle regole di compatibilità. Tale tema è allo studio di diversi specialisti del settore; un'indicazione di compatibilità può essere considerata come riportata nella seguente tabella³⁶:

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

Tabella 5: Compatibilità tra licenze



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



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



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

Ulteriori raccomandazioni.

In generale, si vuole qui ribadire l'importanza di associare ai dati pubblici una licenza aperta (secondo quanto indicato in Figura 7) che consenta di rispettare requisiti di (i) interoperabilità, anche transfrontaliera (non limitandosi a selezionare una licenza che sia

³⁶ Lo schema proposto in tabella è tratto principalmente da: Federico Morando, "Interoperabilità giuridica: rendere i dati (pubblici) aperti compatibili con imprese e comunità online", J LIS.it Italian Journal of Library and Information Science, Gennaio 2013, <http://leo.cineca.it/index.php/jlis/article/download/5461/7928> e modificato secondo i recenti aggiornamenti alle licenze considerate.



valida e nota solo entro i confini nazionali) e, per quanto possibile, (ii) di massimo riutilizzo dei dati. Quest'ultimo principio è espressamente indicato anche nella recente direttiva PSI all'art. 8.

Le amministrazioni possono prevedere casi di applicazione di licenze che limitino il riutilizzo dei dati se e solo se ciò si renda necessario per il rispetto di altre normative (e.g., norme in materia di protezione dei dati personali, norme sul diritto d'autore) e comunque motivando opportunamente la scelta.

In particolare, laddove il dato pubblico contenga riferimenti o sia collegato a dati personali, come riportato in sezione 3.3, si rammenta che trova applicazione il comma 5 dell'articolo 2 del CAD, il Codice della Privacy e le recenti "Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati". Pertanto, in riferimento al tema delle licenze "*per garantire il rispetto dei diritti degli interessati da parte degli utilizzatori, i termini delle licenze per il riutilizzo dovrebbero contenere una clausola di protezione dei dati sia quando il riuso riguardi dati personali, sia quando riguardi dati anonimi derivati da dati personali*".

In ogni caso, il rilascio dei dati con una licenza che ne consente il massimo riutilizzo non modifica la titolarità dei dati stessi e non preclude il titolare a mantenere i dati aggiornati.

In conclusione, ai sensi dell'art. 52 del CAD, 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).

In tal caso, secondo quanto anche definito dall'agenda nazionale sulla valorizzazione del patrimonio informativo pubblico, la mancata indicazione della licenza implica che i dati siano pubblicati 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 4.0.



9. 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 [3] 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.”

9.1. 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.

In aggiunta ai due casi di rilascio a titolo oneroso, si può prevedere inoltre una politica di eliminazione della tariffa una volta che i costi di produzione siano stati coperti.

Uso a titolo gratuito (nessuna tariffa). I costi di produzione e di rilascio dei dati sono interamente a carico dell'amministrazione. Questi potrebbero essere notevolmente 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 8.

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 documenti.

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

9.2 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.

Più recentemente è stato pubblicato uno studio in sede di Commissione Europea sui modelli di business nell'ambito dei Linked Open Government Data ricavato dall'analisi delle iniziative di 37 attori, tra cui anche l'Agenzia per l'Italia Digitale [118].

Lo studio è stato strutturato a partire dagli elementi del Business Model Canvas proposto da



Alexander Osterwalder. Per ogni elemento sono stati estrapolati gli aspetti chiave emersi nelle interviste. I benefici riscontrati sono sia di tipo tangibile (e in questo caso anche monetizzabili) che intangibile, principalmente dovuti alla flessibilità dell'integrazione dei dati, all'aumento della qualità dei dati, alla creazione di nuovi servizi e alla riduzione dei costi.

Un ulteriore studio dei modelli di business è stato condotto dalla London Business School i cui risultati sono pubblicati in [119].

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.

Sviluppo di applicazioni in-house. L'amministrazione può sviluppare applicazioni (ad esempio mobile) con i suoi dati aperti per sfruttare indirettamente un modello di business legato all'applicazione. In questo caso può essere utile lo studio sui modelli di business delle applicazioni sviluppate con l'Open Data presentato in [120].

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].



10. ALCUNE INDICAZIONI OPERATIVE PER LA PUBBLICAZIONE DI DATI APERTI

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 fornisce suggerimenti per gli elementi architettonici e tecnologici comuni nelle soluzioni Open Data.

Nel rispetto delle scelte già effettuate dalle varie amministrazioni, la sezione 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”; piuttosto lo scopo è indirizzare le amministrazioni tra le diverse opportunità architettoniche e tecnologiche adottabili per implementare soluzioni Open Data.

Gli elementi dell'architettura sono descritti attraverso una proposta di architettura generale. Tuttavia si segnala che il W3C sta lavorando alla standardizzazione di un'architettura logica per una soluzione Open Data evoluta, i.e., la Linked Data Platform [121]. Essa fornisce una tassonomia delle risorse che la piattaforma deve gestire e indicazioni sulle operazioni che devono essere implementate per poter mettere a disposizione le suddette risorse in modo standardizzato. Esistono già soluzioni tecnologiche completamente conformi alla Linked Data Platform, un esempio è rappresentato da Apache Marmotta³⁷.

In modo trasversale, l'architettura può considerare elementi di federazione, ovvero piattaforme e interfacce di accesso istanziate singolarmente che sono legate logicamente in modo da fornire ad un utente un singolo punto di accesso.

Infine, la sezione illustra 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. Sono poi discusse alcune modalità generiche di pubblicazione dei dati all'interno di portali Web.

10.1. Elementi architettonici per la pubblicazione

Questa sezione introduce i principali livelli architettonici che compongono una soluzione per la pubblicazione e interrogazione di Open Data. Le componenti dei livelli architettonici possono essere istanziate in diverso modo a seconda della qualità dei dati che si intende soddisfare. Nel seguito, si illustrano i principali requisiti che le infrastrutture di esposizione devono prevedere per rendere disponibili dati secondo i vari livelli dei modelli proposti in sezione 4.

LIVELLO PRESENTAZIONE. Tale livello include due macro componenti: il portale o la sezione di un portale Web, e un'interfaccia di accesso via Web.

Un portale Web o una sezione in un portale esistente questa componente rappresenta l'interfaccia, tipicamente “user-friendly”, per supportare l'interazione con gli utenti finali. E' lo strumento che ha come obiettivo quello di dare massima visibilità ai dataset disponibili, di interagire con l'utente finale per raccogliere statistiche utili a capire quali dati sono di interesse, nuove richieste, feedback ecc. La successiva sezione illustra alcune modalità per la realizzazione di tale componente

³⁷ <https://marmotta.apache.org/>



architetturale.

L'interfaccia di accesso via Web integra la componente portale o sezione di portale consentendo di accedere puntualmente ai dati e ai metadati memorizzati. Ha come obiettivo quello di aumentare l'iterazione machine-to-machine attraverso il dispiegamento di una piattaforma di esposizione dati, basata su API di servizio (o OpenDataServices) che agisce invocando i dati e metadati memorizzati. Nel caso di dati esposti secondo i livelli 4 e 5 del modello in sezione 4.1 l'interfaccia di accesso via Web è rappresentata da uno SPARQL endpoint. In generale, si noti che tale interfaccia deve essere fruibile sia da parte degli utenti finali che da programmi.

LIVELLO INFRASTRUTTURALE. Il livello model è rappresentato dall'infrastruttura che ospita i dati e i metadati.

Tenuto conto della natura intrinseca dei dati aperti, ovvero dati tipicamente non riferibili a singole persone e per i quali solitamente non si richiede il soddisfacimento di specifici requisiti di protezione dei dati personali, moderne tecnologie basate sul paradigma del cloud computing pubblico (o di comunità come il cloud del Sistema Pubblico di Connettività) possono essere facilmente impiegabili al fine di ospitare le infrastrutture per la pubblicazione di dati aperti.

A tale scopo potrebbero essere di ausilio i risultati ottenuti nel contesto del progetto europeo OpenDAI [122] che ha affrontato tali tematiche con l'obiettivo di liberare grandi quantità di dati presenti nei database pubblici per renderli fruibili al vasto pubblico di potenziali utenti, "rompendo" i silos applicativi verticali tramite strumenti di virtualizzazione. Tali strumenti, infatti, non prevedono la replicazione del dato ma la creazione di una vista virtualizzata dello stesso all'interno del cloud eliminando così i costi e gli oneri derivanti dal mantenimento di copie dello stesso dato.

L'approccio descritto ha quindi il vantaggio di spostare su cloud "potenziali data hub" sfruttando le potenzialità di questo ambiente e lasciando i dati all'interno dei silos legacy afferenti alle singole PA.

10.2. Soluzioni Open Data per i portali Web

Si consiglia la predisposizione di un portale Web per accompagnare l'esposizione dei dati aperti, indipendentemente dal livello di apertura esposto ed eventualmente integrandolo con una serie di altri contenuti utili (ad esempio moduli per i feedback, app showcase, ecc.).

Le soluzioni per la creazione del portale sono molteplici; nel seguito si riportano alcune scelte che possono essere considerate.

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. La soluzione può ospitare dati e metadati in un qualsiasi formato e/o livello di classificazione. Si raccomanda tuttavia di mettere in evidenza eventuali strumenti di interrogazione dei dati (e.g., SPARQL endpoint, API) 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 IKS³⁸ 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 Stanbol³⁹ che fornisce una serie di strumenti utili per le integrazioni con i CMS esistenti. Si segnalano per esempio i portali del 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 piattaforme esterne. Esistono casi in cui, per la pubblicazione dei dati, si fa affidamento a strumenti e piattaforme preconfezionate. I vantaggi in questo caso sono quelli di usufruire di una serie di funzionalità già sviluppate per la gestione della fase di pubblicazione. In alcuni casi è possibile far uso di queste piattaforme esternalizzando la memorizzazione e la storicizzazione dei dati, in questo caso anche in modalità cloud computing. 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). Questi strumenti possono essere facilmente integrati con portali già esistenti.

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:

- assegnare ai dataset nomi autoesplicativi che consentano di comprendere il principale contenuto;
- pubblicare i responsabili del dataset con i relativi contatti;
- fornire delle interrogazioni (query) d'esempio per facilitare l'utente nel comprendere l'ontologia in uso;
- produrre e mostrare statistiche di uso, accesso e produzione;
- mettere in evidenza la presenza di un eventuale SPARQL endpoint, pubblicandone il link di accesso;
- mettere in evidenza la licenza in uso in forma “human-readable” e “machine-readable” per favorire la negoziazione fra diversi dataset;
- nel caso di uso di molteplici licenze diverse con obbligo di attribuzione (si veda Figura 7), 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;
- fare, ove possibile, descrizioni testuali esaustive sui dataset.

10.3. 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

38 <http://www.iks-project.eu/>

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



caratteristiche.

Presenza di uno SPARQL endpoint. Pubblicare solo dataset in formato RDF non è sufficiente. La maggior parte dei database per dati RDF forniscono questo tipo di servizio. Quest'aspetto è ampiamente discusso nelle sezioni precedenti.

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.

Nei casi di amministrazioni molto piccole o amministrazioni che non siano nelle condizioni di poter fornire un tale servizio, si consiglia o di implementare azioni di sussidiarietà verticale utilizzando lo SPARQL endpoint dell'amministrazione di riferimento, ove presente (ad esempio, nel caso di piccoli comuni questi possono riferirsi alla Regione di appartenenza), o di sfruttare quello offerto nel contesto delle infrastrutture nazionali del Sistema Pubblico di Connettività (i.e., SPCData).

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.

Seppur non necessario avere URI con una semantica intrinseca, rispettare una *naming convention* può essere utile a chi scrive le query e facilita l'implementazione di una politica di URI persistenti. Infatti, le 10 regole per le URI persistenti proposte tramite un lavoro della commissione europea considerano sia aspetti tecnici che di naming convention [89]. Per autoconsistenza si riportano le 10 regole nel seguito. I lettori interessati possono riferirsi a [89] per ulteriori esempi di buone pratiche in alcuni paesi europei per la definizione di URI persistenti.

Al fine di facilitare la creazione di URI persistenti sono da evitare URI che contengano:

- Nome dell'organizzazione/progetto che detiene la risorsa in modo da evitare i problemi derivanti dalla fine del progetto stesso o da azioni di fusione e chiusura dell'organizzazione
- Numeri di versione
- Identificatori esistenti che in passato sono stati utilizzati per identificare risorse differenti
- Riferimenti generati in modo automatico e incrementale a meno che non ci sia la garanzia che il processo non venga mai più ripetuto o, se ripetuto, generi sicuramente gli stessi identificatori per gli stessi dati di input
- Stringhe rappresentanti “query” a database
- Estensione del file



Al fine di facilitare la creazione di URI persistenti sono, invece, da ritenersi buone pratiche da seguire i seguenti punti:

- Per poter gestire URI persistenti, si raccomanda di definire il formato dell'URI come segue:


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

 dove gli elementi che compongono la URI sono i seguenti:
 - **Dominio:** il dominio Web su cui reperire la risorsa.
 - **Tipo:** è l'elemento che specifica il tipo di risorsa che viene identificato dalla URI. Dovrebbe poter assumere un numero limitato di valori come “doc” se la risorsa identificata è un documento descrittivo, “set” se la risorsa è un dataset, “id” o “item” se la risorsa è un oggetto del mondo reale.
 - **Concetto:** potrebbe essere il tipo di un oggetto del mondo reale
 - **Riferimento:** è lo specifico elemento, termine o concetto che rappresenta la risorsa
- Costruire URI per più formati al fine di identificare al meglio la risorsa
- Collegare tra loro le rappresentazioni multiple della stessa risorsa
- Implementare il codice di risposta 303 per gli oggetti del mondo reale
- Utilizzare servizi dedicati

Per quanto concerne la pubblicazione ed il mantenimento di URI persistenti, è possibile anche utilizzare un server PURL⁴⁰. Questo presenta alcuni vantaggi:

1. la struttura delle url è disaccoppiata dai servizi reali, che possono essere mantenuti senza mandare offline le risorse. Operativamente può bastare che il servizio PURL sia esposto (o comunque raggiungibile) via porta 80, così che l'accesso agli altri servizi sarà poi gestito tramite esso. I dati possono essere forniti direttamente, o può essere predisposto un redirect opportuno (303).
2. ogni server PURL può essere configurato all'interno di un network di federated service, così che la rete è in grado di distribuire il carico delle richieste. Ogni risorsa sarà accessibile tramite ciascuno dei proxy, distribuendo il carico (in modo di favorire il buy-in da parte di amministrazioni dotate di meno risorse, ad esempio piccoli comuni).
3. ogni server PURL può decidere di accettare utenti sottoscrittori esterni: ad esempio piccole amministrazioni dotate di meno risorse potrebbero non voler installare un proprio server PURL, ma di gestire il proprio spazio dei nomi tramite il server ospitato presso una PA centrale.

Dereferenziazione delle URI e “content negotiation”. Facendo uso di URI HTTP per identificare

⁴⁰ <http://purl.oclc.org/docs/index.html>



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 la risoluzione dell'URI 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.

Nel caso si intenda fornire all'utente pagine dedicate alle singole entità, sarà necessario predisporre meccanismi di traduzione trasparenti per gli strumenti di pubblicazione. Ad esempio, per rendere le risorse linked deferenziabili è preferibile l'utilizzo di navigatori basati sulle Linked Data API, quali Pubby⁴¹ o ELDA⁴², che integrano nativamente la "content-negotiation", e possono quindi esporre le risorse richieste sia in formato HTML (accesso normale tramite browser alla URL di pubblicazione della risorsa), sia in qualche serializzazione di RDF.

E' possibile far uso di strumenti automatici per verificare la corretta configurazione della dereferenziazione e della content negotiation come ad esempio Vapour⁴³.

10.4. Recupero e accesso ai dati

Il recupero e l'accesso puntuale ai dati è necessario per garantire una buona fruibilità dell'informazione, consentendo di superare un processo di "discovery" che potrebbe richiedere: (i) il download del file contenente il dataset, (ii) il caricamento dello stesso in memoria e (iii) la ricerca dei dati di interesse all'interno del dataset.

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.

⁴¹ <http://wifo5-03.informatik.uni-mannheim.de/pubby/>

⁴² <https://code.google.com/p/elda/>. Il repository aggiornato per il progetto Elda è <https://github.com/epimorphics/elda>

⁴³ <http://validator.linkeddata.org/vapour>



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).

In caso di esposizione di API, possono essere utilizzati strumenti che raccolgono in maniera semplice e intuitiva la lista dei servizi e delle API disponibili, fornendo al tempo stesso la relativa documentazione. Swagger⁴⁴, Apiary⁴⁵, o Hydra⁴⁶ sono esempi di tali strumenti.

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. Occorre segnalare al sistema la presenza del proprio SPARQL endpoint.

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.

Data Discovery. Possono essere utili da fornire delle soluzioni di data discovery, ovvero strumenti visuali di accesso ed esplorazione dei dati. Alcuni di questi possono essere utilizzati con qualsiasi tipo

⁴⁴ <http://swagger.wordnik.com>

⁴⁵ <http://apiary.io/>

⁴⁶ <http://www.markus-lanthaler.com/hydra/>



di dato ma ovviamente in caso di dati non di livello 4 e 5, è richiesto l'uso di una infrastruttura di backend che possa gestire anche in memoria volatile tutti i dati disponibili. Esempi di questi strumenti sono ad esempio QlikView⁴⁷, Tableau⁴⁸ e Linkurious⁴⁹.

In caso di dati di livello 4 e 5 si possono sfruttare soluzioni più snelle come LODLive, spiegata nel seguito.

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>.

10.5. Elementi di federazione

L'idea di federazione nasce per offrire all'utente finale la possibilità di effettuare in un unico punto, ricerche e accessi su piattaforme differenti.

La federazione di piattaforme differenti presuppone, in genere, un accordo preliminare tra i soggetti che intendono federare le loro soluzioni e quindi considera aspetti puramente organizzativi, che tuttavia non sono trattati in questa sezione.

Nel seguito si illustrano diverse possibilità di federazione, legate al livello di classificazione dei dati e dei metadati.

Condivisione di una piattaforma. Un primo livello di federazione tra amministrazioni può essere quello in cui un'amministrazione mette a disposizione di altre la propria soluzione. Ad esempio, un'amministrazione regionale dotata di una piattaforma Open Data può mettere a disposizione dei comuni della regione tale piattaforma.

Federazione cataloghi di dati. Questa soluzione è adottata per fornire uno strumento che consente di effettuare in modo trasparente una ricerca su molteplici cataloghi. E' una soluzione largamente utilizzata per la pubblicazione di dati di livello fino al 3.

Le piattaforme CKAN e Socrata forniscono funzionalità di questo tipo. Ad esempio, la federazione di cataloghi CKAN è stata adottata dal portale dati.gov.it (per abilitare la federazione con dati.gov.it, il dipartimento della funzione pubblica ha redatto recentemente un documento tecnico dove è possibile trovare indicazioni operative per abilitare tale federazione [124]), mentre un esempio di federazione mediante l'uso di Socrata è il portale della Regione Lombardia dati.lombardia.it.

In questi casi la federazione è basata tipicamente da un "accordo" di tipo tecnico su cui eventualmente istanziare una base comune di metadati.

⁴⁷ <http://www.qlik.com/us>

⁴⁸ <http://www.tableausoftware.com/>

⁴⁹ <http://linkurio.us/>



Dal punto di vista della federazione tramite metadati, il progetto Homer [123] propone un indice federato di metadati esposti da diversi portali Open Data, anche di diversa natura tecnologica. Una delle sfide del progetto è quella di superare il problema del multilinguismo grazie ad un'integrazione con il tesoro Eurovoc.

Federazione di SPARQL endpoint. Il W3C ha definito uno standard per federare SPARQL endpoint [125]. Lo standard prevede una sintassi aggiuntiva per SPARQL in grado di considerare, in una stessa "query", dati provenienti da SPARQL endpoint differenti. Lo standard prevede inoltre modalità in cui molteplici SPARQL endpoint risultano totalmente trasparenti all'utente; in questo caso, è il sistema che gestisce l'invio della "query" a più endpoint o la scomposizione della "query" e la ricomposizione dei frammenti del risultato finale⁵⁰. Nell'ambito della federazione di SPARQL endpoint, si segnala anche il progetto SemaGrow [126] che indica un'architettura per abilitare tale federazione.

⁵⁰Un'alternativa alla federazione di SPARQL endpoint si basa sul modello dei "Linked Data Fragments" (<http://linkeddatafragments.org/>)



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

L'art. 52 del CAD introduce al comma 3 una disposizione che intende regolamentare la produzione di dati di tipo aperto in tutte quelle gare per la fornitura di prodotti e/o servizi che comportino la raccolta e la gestione di dati pubblici.

Per poter attuare tale disposizione, il gruppo di lavoro ha ritenuto importante dare alle pubbliche amministrazioni indicazioni operative per la definizione di possibili clausole che possano essere inserite nella documentazione tecnica di gara.

A tal riguardo, si raccomanda in primo luogo di includere nello schema di contratto una clausola simile a quella seguente, che richiami la suddetta disposizione:

“Ai sensi dell’articolo 52 del Codice dell’Amministrazione Digitale (CAD), nella fornitura di prodotti e/o servizi che comportino la raccolta e la gestione di dati pubblici, tali dati, i relativi metadati, gli schemi delle strutture di dati e delle relative banche dati, devono poter essere acceduti telematicamente, nel rispetto del principio di neutralità tecnologica, e predisposti per essere riutilizzati da parte di persone fisiche e giuridiche secondo le modalità tecniche specificate nella documentazione di gara allegata e comunque nel rispetto dei principi e delle raccomandazioni dell’agenda e delle linee guida nazionali per la valorizzazione del patrimonio informativo pubblico rilasciate dall’Agenzia per l’Italia Digitale”.

A livello di capitolato tecnico di gara, diversi casi sono possibili dipendentemente dalle peculiarità dell'appalto e comunque dalle specificità dei prodotti e/o servizi oggetto della gara. In questo senso, supportati anche dall'analisi della relazione interpretativa [127] che accompagna la disposizione precedentemente riportata, emerge come il legislatore abbia voluto in ogni caso riferirsi a un ampio spettro di forniture che comportino la realizzazione non solo di prodotti e/o servizi che abbiano direttamente come oggetto la gestione di dati pubblici, ma anche quelli che indirettamente producano dati pubblici.

Alla luce di tale analisi, si possono distinguere due diversi casi:

- l'oggetto della fornitura riguarda prodotti e/o servizi che comportano indirettamente una produzione di dati pubblici; in questo caso, l'amministrazione o la stazione appaltante può inserire all'interno del capitolato tecnico una clausola che obblighi l'aggiudicatario della gara a:
 - fornire all'amministrazione tali dati pubblici in forma disaggregata e, ove possibile, tabellare consentendole di acquisire piena titolarità dei dati; i dati da richiedere all'aggiudicatario possono essere specificati dall'amministrazione stessa in una clausola prevedendo le eventuali restrizioni normative (e.g., norme sulla protezione dei dati personali) e/o di carattere industriale applicabili ai dati;
 - le modalità di acquisizione dei dati da parte dell'amministrazione (e.g., tempistiche e meccanismi di trasferimento dall'aggiudicatario all'amministrazione);
 - una descrizione puntuale dei dati pubblici prodotti ivi incluse le attività ed entità che hanno dato origine alla produzione dei dati.

L'amministrazione, una volta acquisite tali informazioni, si assume l'onere di renderle pienamente disponibili sotto forma di dati di tipo aperto, secondo le raccomandazioni delle presenti linee guida e gli obiettivi individuati dall'agenda nazionale per la valorizzazione del



patrimonio informativo pubblico, al fine di incentivare il massimo riutilizzo da parte di persone fisiche e giuridiche;

- l'oggetto stesso della fornitura riguarda prodotti e/o servizi che prevedono la gestione diretta di dati pubblici (come ad esempio la gestione di uno specifico sistema informativo); in questo caso, l'amministrazione può inserire all'interno del capitolato tecnico una serie di clausole che obbligano l'aggiudicatario a produrre, a partire dall'oggetto della gara, anche dati di tipo aperto il più possibile disaggregati, nel rispetto comunque di eventuali restrizioni normative (e.g., normativa sulla protezione dei dati personali) e/o di carattere industriale applicabili. In questo caso, il capitolato può essere dettagliato per richiedere espressamente:
 - la produzione di dati in formato aperto minimo di livello tre, secondo il modello proposto dalle presenti linee guida e tenuto conto dei principi di fruibilità, qualità, interoperabilità individuati dall'agenda nazionale per la valorizzazione del patrimonio informativo pubblico;
 - la produzione di metadati che accompagnano i dati aperti prodotti, secondo il modello per i metadati proposto dalle presenti linee guida, includendo anche le indicazioni sulla specifica licenza aperta da applicare ai dati ed ogni altra attività e/o entità che descriva come tali dati sono stati originati (provenance);
 - la pubblicazione dei dati aperti secondo specifiche indicazioni dell'amministrazione e comunque tenendo conto delle raccomandazioni delle presenti linee guida;
 - l'aggiornamento dei dati aperti, secondo la frequenza di aggiornamento individuata dall'amministrazione;
 - l'esportazione in formato aperto di ogni altro eventuale dato di reportistica generato.

In entrambi i casi, l'amministrazione o la stazione appaltante individuerà i parametri dimensionali riguardanti le attività di fornitura di dati pubblici e produzione di dati aperti, rispettivamente, al fine della determinazione della valutazione economica del prodotto e/o servizio oggetto della gara.



12. 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 - 2014”, http://www.agid.gov.it/sites/default/files/leggi_decreti_direttive/regolamenti_e_direttive_locali_open_data_regioni_agg_cisis.pdf, 2014.
- [6] Creative Commons Italia, <http://www.creativecommons.it/Licenze>, 2014.
- [7] Agenzia per l'Italia Digitale, “Obiettivi di accessibilità per l'anno 2014”, http://www.agid.gov.it/sites/default/files/documentazione/obiettivi_accessibilita_agid_2014.pdf, Marzo 2014.
- [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/>, 2014.
- [12] W3C, RDF Working Group, http://www.w3.org/2011/rdf-wg/wiki/Main_Page, 2014.
- [13] W3C, RDF Vocabulary Description Language 1.0: RDF Schema, <http://www.w3.org/TR/rdf-schema/>, 2014.
- [14] W3C, OWL Working Group, http://www.w3.org/2007/OWL/wiki/OWL_Working_Group, 2014.
- [15] W3C, SPARQL Working Group, http://www.w3.org/2009/sparql/wiki/Main_Page, 2014.
- [16] SDMX - Statistical Data and Metadata Exchange, <http://sdmx.org/>, 2014.
- [17] IETF, “RFC 4180 – Common Format and MIME Type for Comma Separated Values (CSV) Files”, <http://tools.ietf.org/html/rfc4180>, 2014.
- [18] OASIS, “Open Document Format for Office Applications”, https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office, 2014.
- [19] W3C, SKOS Simple Knowledge Organization System - Home Page,



- <http://www.w3.org/2004/02/skos/>, 2014.
- [20] European Commission, “Core public service vocabulary”,
http://joinup.ec.europa.eu/asset/core_public_service/news/core-public-service-vocabulary-released-public-review, 2014.
- [21] SPOCS project, <http://www.eu-spocs.eu/>, 2014.
- [22] European Commission, “INSPIRE – Infrastructure for Spatial Information in the European Community”, <http://inspire.jrc.ec.europa.eu/>, 2014.
- [23] W3C, RDF Data Cube, <http://www.w3.org/TR/vocab-data-cube/>, 2014.
- [24] W3C, Data Catalog Vocabulary (DCAT), <http://www.w3.org/TR/vocab-dcat/>, 2014.
- [25] DBpedia, <http://dbpedia.org/About>, 2014.
- [26] DBpedia Italiana, <http://it.dbpedia.org/>, 2014.
- [27] SPCData, <http://spcdata.digitpa.gov.it>, 2014.
- [28] GeoNames, <http://www.geonames.org/>, 2014.
- [29] Agenzia per l'Italia Digitale, “Repertorio nazionale per i dati territoriali – RNDT”,
<http://www.rndt.gov.it/RNDT/home/index.php>, 2014.
- [30] ISPRA, SINAnet, <http://www.mais.sinanet.isprambiente.it/ost/>, 2014.
- [31] OpenStreetMap – The free Wiki World Map, <http://www.openstreetmap.org/>, 2014.
- [32] Open Data Commons, Open Database License (ODbL),
<http://opendatacommons.org/licenses/odbl/>, 2014.
- [33] W3C Draft Tag Finding, Deferencing http URI, <http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14>, 2014.
- [34] CKAN – The Data Hub, “Linking Open Data Cloud”, <http://thedatahub.org/group/lodcloud>, 2012.
- [35] Socrata, <http://www.socrata.com/>, 2014.
- [36] OpenLab, Microsoft, <https://github.com/openlab>, 2014.
- [37] Creative Commons, “Describing Copyright in RDF”, <http://creativecommons.org/ns>, 2014.
- [38] Dublin Core Metadata Initiative, <http://dublincore.org/documents/dcmes-xml/>, 2014.
- [39] W3C, Asset Description Metadata Schema (ADMS), <http://www.w3.org/ns/adms>, 2014.
- [40] Simone Aliprandi, Legal tools for Open Data,
http://www.ifosslr.org/public/opendata_graph.pdf.
- [41] Sindice – The Semantic Web Index, <http://sindice.com/>, 2014.
- [42] Squin – Query the Web of Linked Data, <http://squin.sourceforge.net/index.shtml>, 2014.
- [43] Lodlive – browsing the Web of Data, <http://en.lodlive.it/>, 2014.
- [44] Camera dei deputati, <http://dati.camera.it/it/>, 2014.
- [45] W3C, Cool URIs for the Semantic Web –<http://www.w3.org/TR/cooluris/>, 3 Dicembre 2008.



- [46] Paul Uhler, “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%20on%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>, 2014.
- [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/>, 2014.
- [56] W3C, “Turtle - Terse RDF Triple Language”, <http://www.w3.org/TeamSubmission/turtle/>, 2014.
- [57] IETF, RFC 4627 - The application/json Media Type for JavaScript Object Notation (JSON), <http://www.ietf.org/rfc/rfc4627.txt>, 2014.
- [58] W3C, “N-Triples”, <http://www.w3.org/TR/n-triples/>, 2014.
- [59] ESRI, “Shapefile Technical Description”,
<http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, 2014.
- [60] GoodRelations Language Reference,
<http://www.heppnetz.de/ontologies/goodrelations/v1.html>, 2014.
- [61] Public Contracts Ontology, <https://code.google.com/p/public-contracts-ontology/>, 2014.
- [62] Guide to the Payments Ontology, <http://data.gov.uk/resources/payments>, 2014.
- [63] FOAF Vocabulary Specification, <http://xmlns.com/foaf/spec/>, 2014.
- [64] Who Am I! Vocabulary Specification, <http://vocab.ctic.es/wai/wai.html>, 2014.
- [65] SIOC Core Ontology Specification, <http://rdfs.org/sioc/spec/>, 2014.
- [66] European Commission, Core Location Vocabulary,
https://joinup.ec.europa.eu/asset/core_location/description, 2014.
- [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/>, 2014.
- [69] W3C, “Working Group Note. Registered Organization Vocabulary”, <https://dvcs.w3.org/hg/gld/raw-file/default/legal/index.html>, 2014.
- [70] W3C, “OWL-S: Semantic Markup for Web Services”, <http://www.w3.org/Submission/OWL-S/>, 2014.
- [71] Description of a Project, <https://github.com/edumbill/doap/wiki>, 2014.
- [72] W3C, “vCard Ontology”, <http://www.w3.org/TR/vcard-rdf/>, 2014.
- [73] Eurovoc, <http://eurovoc.europa.eu/>, 2014.
- [74] FAO, Agrovoc, <http://aims.fao.org/standards/agrovoc/about>, 2014.
- [75] UNICE, EUROSTAT, OECD, “Extended Knowledge Organization System (XKOS)”, <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP10.pdf>, 2014.
- [76] Thesaurus PICO, http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml, 2014.
- [77] UNESCO Thesaurus, <http://databases.unesco.org/thesaurus/>, 2014.
- [78] The CIDOC Conceptual Reference Model, http://www.cidoc-crm.org/official_release_cidoc.html, 2014.
- [79] Europeana Data Model (EDM) Documentation, <http://pro.europeana.eu/edm-documentation>, 2014.
- [80] W3C Interest Group Note, “Describing Linked Datasets with the VOID Vocabulary”, <http://www.w3.org/TR/void/>, 2014.
- [81] Dublin Core Metadata Element Set, Version 1.1, <http://dublincore.org/documents/dces/>, 2014.
- [82] Dublin Core Metadata Initiative, <http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#>, 2014.
- [83] W3C Working Group Note, “PROV-Overview An Overview of the PROV Family of Documents”, <http://www.w3.org/TR/prov-overview/>, 2014.
- [84] Schema.org, <http://schema.org/>, 2013.
- [85] VANN: A vocabulary for annotating vocabulary descriptions, <http://vocab.org/vann/.html>, 2014.
- [86] U.S. National Library of Medicine, “Medical Subject Headings”, <http://www.nlm.nih.gov/mesh/>, 2014.
- [87] SNOMED CT, <http://www.ihtsdo.org/snomed-ct/>, 2014.
- [88] European Commission, “e-Government Core Vocabularies”, https://joinup.ec.europa.eu/community/core_vocabularies/description, 2014.
- [89] European Commission, “10 Rules for Persistent URIs”, <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>, 2014.
- [90] 5 star open data, <http://5stardata.info/>, 2014.



- [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>, 2014.
- [94] Agenzia per l'Italia Digitale, "Agenda Nazionale per la valorizzazione del patrimonio informativo pubblico (anno 2014)", http://www.agid.gov.it/sites/default/files/documenti_indirizzo/agendanazionalepatrimoniopubblico2014.pdf, 2014.
- [95] EIONET, GEMET Thesaurus, <http://www.eionet.europa.eu/gemet/index.html?>, 2014.
- [96] ISPRA, ThIST Thesaurus, <http://www.isprambiente.gov.it/contentfiles/00003600/3687-thist.pdf/view>, 2014.
- [97] Agenzia per l'Italia Digitale, "SPCoop – Servizio pubblico di cooperazione applicativa", <http://www.agid.gov.it/infrastrutture-sicurezza/servizi-interoperabilita-evoluta>, 2014
- [98] Agenzia per l'Italia Digitale, "Linee guida per la stesura di convenzioni per la fruibilità di dati delle pubbliche amministrazioni (versione 2.0)", http://www.agid.gov.it/sites/default/files/linee_guida/linee_guida_convenzioni_fruibilita_dati_delle_pa_art_58_cad_0.pdf, 2014
- [99] Senato della Repubblica, <http://dati.senato.it/>, 2014.
- [100] Archivio Centrale dello Stato, <http://dati.acs.beniculturali.it/>, 2014.
- [101] ODI, "Open Data Certificates", <https://certificates.theodi.org/>, 2014
- [102] RDFa, "Linked Data in HTML", <http://rdfa.info/>, 2014.
- [103] JSON-LD, "JSON for Linkind Data", <http://json-ld.org/>, 2014.
- [104] W3C, "JSON-LD 1.0", <http://www.w3.org/TR/json-ld/>, 2014.
- [105] W3C, "CSV on the Web", <http://www.w3.org/2013/05/lcsv-charter.html>, 2014.
- [106] OGC, "KML", <http://www.opengeospatial.org/standards/kml>, 2014.
- [107] geoJSON, <http://geojson.org/>, 2014.
- [108] OGC, "Geography Markup Language – GML", <http://www.opengeospatial.org/standards/gml>, 2014.
- [109] Akoma Ntoso, "XML for parliamentary, legislative & judiciary documents", <http://www.akomantoso.org/>, 2014.
- [110] XBRL, <http://www.xbrl.org/>, 2014.
- [111] European Commission, "DCAT application profile for open data portals in Europe", https://joinup.ec.europa.eu/asset/dcat_application_profile/description, 2014.
- [112] VOAFA, <http://lov.okfn.org/vocab/voaf/v2.3/index.html>, 2014.
- [113] OpenStreetMap, <http://www.openstreetmap.org/#map=5/42.088/12.564>, 2014.



- [114] Italian Open Data License 2.0, <http://www.dati.gov.it/iodl/2.0/>, 2014.
- [115] Open Data Commons Attribution, <http://opendatacommons.org/category/odc-by/>, 2014.
- [116] Creative Commons, Intergovernmental Organizations - CC-IGO, http://wiki.creativecommons.org/Intergovernmental_Organizations, 2014.
- [117] Open Data Commons, Public Domain Dedication and License, <http://opendatacommons.org/licenses/pddl/>, 2014.
- [118] P. Archer, M. Dekkers, S. Goedertier, N. Loutas, “Study on Business Models for Linked Open Government Data”, 2013
- [119] Carla Bonina, “New business models and the value of open data: definitions, challenges and opportunities”, 2013.
- [120] M. Zarreh, “Apps for Europe - Business and Policy”, 2014.
- [121] W3C, “Linked Data Platform”, <http://www.w3.org/TR/ldp>, 2014.
- [122] OPEN-DAI - Open Data Architectures & Infrastructure, <http://open-dai.eu/>, 2014.
- [123] Homer project, <http://homerproject.eu/>, 2014.
- [124] FormezPA, “Guida sintetica per lo scambio di dati con il catalogo nazionale dati.gov.it”, https://docs.google.com/document/d/1niBkBRJ-rxAKVJpntnDkf5xfqeMDtV_94ViGXMIBRQM/edit?pli=1#heading=h.gjdgxs, 2014.
- [125] W3C Recommendation, “SPARQL 1.1 Federated Query”, Marzo 2013.
- [126] SemaGrow project, <http://www.semagrow.eu/>, 2014.
- [127] Relazione interpretativa DL 179/2012 articolo 9, <http://static.ilssole24ore.com/DocStore/Professionisti/AltraDocumentazione/body/13500001-13600000/13584007.pdf>, 2014.
- [128] Agenzia per l'Italia Digitale, Determinazione Commissariale n. 68 del 10 maggio 2013, “Regole tecniche per le basi dati critiche”, http://www.agid.gov.it/sites/default/files/circolari/dt_cs_n.68_-_2013dig_-_regole_tecniche_basi_dati_critiche_art_2bis_dl_179-2012_sito.pdf, 2013.
- [129] OGC, GeoPackage <http://www.geopackage.org/>, 2014.
- [130] Garante per la Protezione dei Dati Personali, “Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati”, <http://194.242.234.211/documents/10160/0/Linee+guida+trasparenza+2014.pdf>, 15 maggio 2014.
- [131] Agenzia per l'Italia Digitale, “Risultati della consultazione pubblica relativa al le linee guida per la valorizzazione del patrimonio informativo pubblico”, http://www.agid.gov.it/sites/default/files/linee_guida/reportconsultazionepubblica.pdf, 2014.



APPENDICE I

Lo scopo di questa appendice è quello di fornire esempi concreti di utilizzo degli standard e delle ontologie di riferimento. In particolare, si mostra come uno stesso insieme di dati RDF può essere serializzato con diversi formati e come poter utilizzare alcune ontologie per dati trasversali e per metadati.

Infine l'appendice presenta un esempio di utilizzo del framework PROV integrato con altre ontologie. L'esempio riporta il caso concreto di pubblicazione della classificazione ufficiale ATECO2007 da parte di ISTAT con il supporto di AgID.

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/Turtle, 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:geonames="http://www.geonames.org/ontology#"
  xmlns:org="http://www.w3.org/ns/org#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:spcdata="http://spcdata.digitpa.gov.it/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
<rdf:Description rdf:about="http://spcdata.digitpa.gov.it/Amministrazione/agid">
  <locn:address rdf:resource="http://spcdata.digitpa.gov.it/Indirizzo/agid"/>
  <geonames:locatedIn rdf:resource="http://spcdata.digitpa.gov.it/Comune/H501"/>
  <org:classification
    rdf:resource="http://spcdata.digitpa.gov.it/CategoriaAmministrazione/32"/>
  <foaf:homepage>www.agid.gov.it</foaf:homepage>
  <org:identifier>97735020584</org:identifier>
  <spcdata:acronimo_amministrazione>AGID</spcdata:acronimo_amministrazione>
  <rdfs:label>Agenzia per L'Italia Digitale</rdfs:label>
  <rdf:type rdf:resource="http://spcdata.digitpa.gov.it/Amministrazione"/>
```



```
</rdf:Description>
<rdf:Description rdf:about="http://spcdata.digitpa.gov.it/Indirizzo/agid">
  <locn:postCode>00144</locn:postCode>
  <locn:fullAddress>Viale Liszt 21, 00144, Roma</locn:fullAddress>
  <rdf:type rdf:resource="http://www.w3.org/ns/locn#Address"/>
</rdf:Description>
</rdf:RDF>
```

NOTATION 3 - TURTLE

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

<http://spcdata.digitpa.gov.it/Amministrazione/agid>
  a spcdata:Amministrazione ;
  rdfs:label "Agenzia per L'Italia Digitale" ;
  spcdata:acronimoAmministrazione "AGID" ;
  geonames:locatedIn <http://spcdata.digitpa.gov.it/Comune/H501> ;
  locn:address <http://spcdata.digitpa.gov.it/Indirizzo/agid> ;
  org:classification <http://spcdata.digitpa.gov.it/CategoriaAmministrazione/32> ;
  org:identifier "97735020584" ;
  foaf:homepage "www.digitpa.gov.it" .

<http://spcdata.digitpa.gov.it/Indirizzo/agid>
  a locn:Address ;
  locn:fullAddress "Viale Liszt 21, 00144, Roma" ;
  locn:postCode "00144" .
```



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/2000/01/rdf-schema#label> "Agenzia per L'Italia Digitale" .

<http://spcdata.digitpa.gov.it/Amministrazione/agid>
<http://www.w3.org/ns/org#classification>
<http://spcdata.digitpa.gov.it/CategoriaAmministrazione/32> .

<http://spcdata.digitpa.gov.it/Amministrazione/agid>
<http://xmlns.com/foaf/0.1/homepage> "www.agid.gov.it" .

<http://spcdata.digitpa.gov.it/Amministrazione/agid>
<http://www.w3.org/ns/org#identifier> "97735020584" .

<http://spcdata.digitpa.gov.it/Amministrazione/agid>
<http://spcdata.digitpa.gov.it/acronimo_amministrazione> "AGID" .

<http://spcdata.digitpa.gov.it/Amministrazione/agid>
<http://www.geonames.org/ontology#locatedIn>
<http://spcdata.digitpa.gov.it/Comune/H501> .

<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> "00144" .

<http://spcdata.digitpa.gov.it/Indirizzo/agid> <http://www.w3.org/ns/locn#fullAddress> "Viale Liszt 21, 00144, 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 specifico 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#> .
@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_aool> org:headOf
<http://spcdata.digitpa.gov.it/Amministrazione/abftrm> .

```

ESEMPIO DI METADATAZIONE E UTILIZZO DI ONTOLOGIE PER I METADATI

Nel seguito si mostra l'uso delle ontologie per i metadati individuate in sezione 6.3.3. L'esempio è reale e basato sulla metadatazione del dataset Linked Open IPA di SPCData dove si utilizzano i metadati obbligatori, i metadati obbligatori condizionatamente e altri metadati supplementari. L'URI del dataset preso come esempio è <http://spcdata.digitpa.gov.it/loipa> che viene distribuito attraverso molteplici modalità, identificate appunto tramite la classe Distribution di DCAT.

Nell'esempio sono indicati, attraverso commenti, i metadati obbligatori e obbligatori condizionatamente. Il solo metadato (obbligatorio condizionatamente) temporal di Dublin Core non è



stato specificato in quanto, in questo caso, la condizione non ne impone l'obbligatorietà.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@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: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:loipadump rdf:type dcat:Distribution .
spcdata:loipawod rdf:type dcat:Distribution .
spcdata:loipa dcat:distribution spcdata:loipadump .
spcdata:loipa dcat:distribution spcdata:loipawod .
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:License cc:permits cc:Reproduction .
spcdata:License cc:permits cc:Distribution .
```



spcdata:License cc:permits cc:Sharing .

spcdata:License cc:permits cc:DerivativeWorks .

#METADATO OBBLIGATORIO: publisher

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

#METADATO OBBLIGATORIO: creator

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

#METADATO OBBLIGATORIO: rightsHolder

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

#METADATO OBBLIGATORIO: title

<http://spcdata.digitpa.gov.it/loipa> <http://purl.org/dc/terms/title> "Linked Open IPA (Indice della Pubblica Amministrazione)" .

#METADATO OBBLIGATORIO: description

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

#METADATO OBBLIGATORIO: modified

<http://spcdata.digitpa.gov.it/loipa> <http://purl.org/dc/terms/modified> "2014-01-15" .

#METADATO OBBLIGATORIO: license

spcdata:loipa cc:license spcdata:License .

#METADATO OBBLIGATORIO: keyword

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

#METADATO OBBLIGATORIO: accrualPeriodicity

spcdata:loipa dcmi:accrualPeriodicity "monthly" .



#METADATO OBBLIGATORIO CONDIZIONATAMENTE: identifier

spcdata:loipa dcmi:identifier "http://spcdata.digitpa.gov.it/loipa" .

#METADATO OBBLIGATORIO CONDIZIONATAMENTE: spatial

spcdata:loipa dcmi:spatial <http://spcdata.digitpa.gov.it/Nazione/Italia> .

#METADATO OBBLIGATORIO CONDIZIONATAMENTE: language

spcdata:loipa dcmi:language "it" .

#METADATO OBBLIGATORIO CONDIZIONATAMENTE: byteSize

spcdata:loipadump dcat:byteSize "179200" .

#METADATO OBBLIGATORIO CONDIZIONATAMENTE: accessURL

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

#METADATO OBBLIGATORIO CONDIZIONATAMENTE: downloadURL

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



APPENDICE II

ESEMPIO DI UTILIZZO DEL PROV FRAMEWORK

Il processo che porta alla standardizzazione di una classificazione è spesso articolato e coinvolge numerosi attori tra cui istituti sia nazionali che internazionali. In questo processo possono nascere delle varianti della classificazione stessa e delle evoluzioni che determinano un aggiornamento di versione. Il framework PROV (Sezione 7) cattura ogni aspetto che serve per descrivere questo flusso informativo.

Nel seguito sono riportati alcuni diagrammi che illustrano come il framework PROV è stato utilizzato in un caso reale, i.e., la pubblicazione in versione Linked Open Data della classificazione ufficiale ATECO 2007, avvenuta a seguito di un'attività congiunta tra ISTAT e AgID.

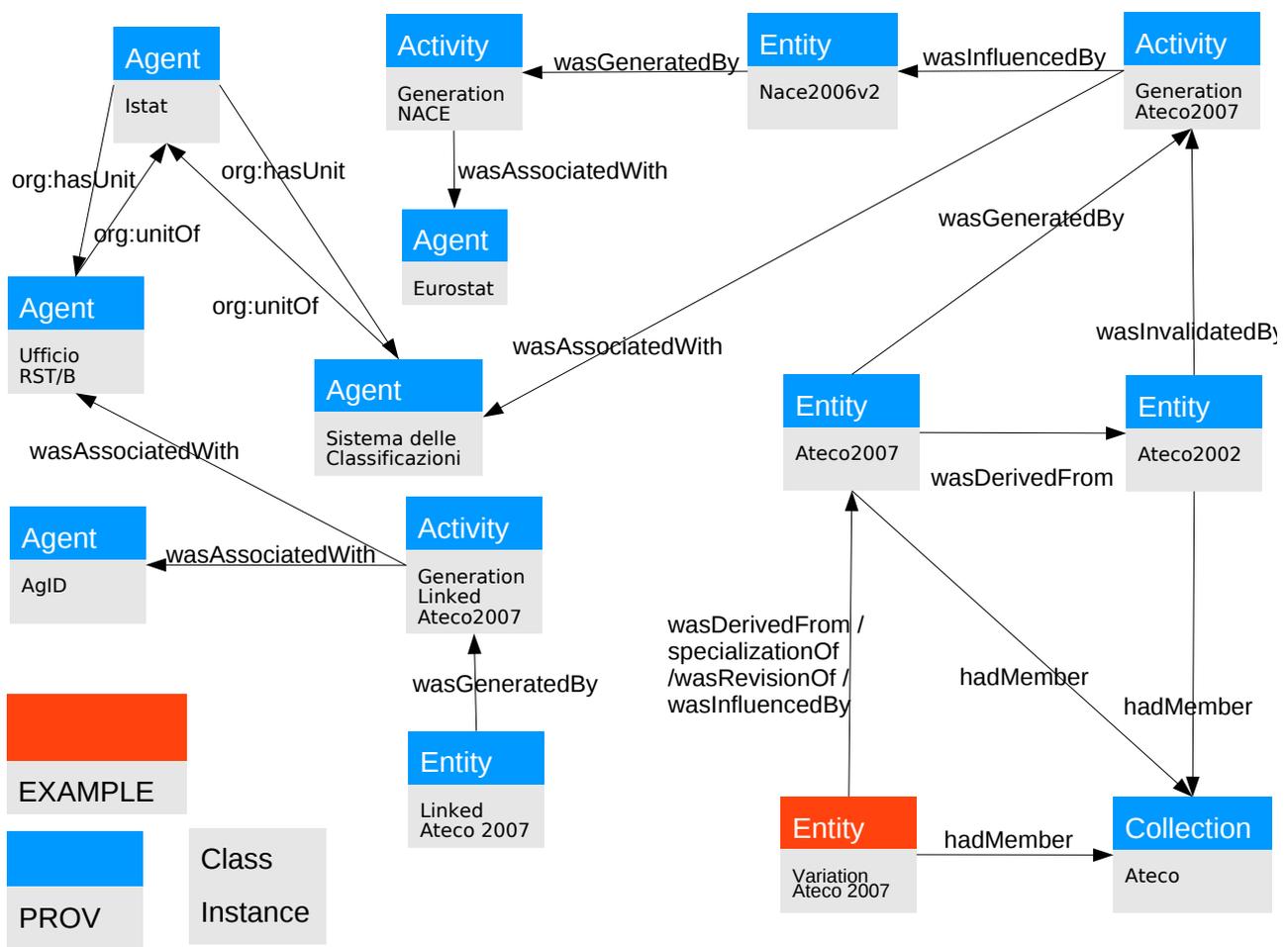


Figura 8: Provenance Linked Open Data Ateco 2007

La Figura 8 descrive la provenance della LinkedAteco2007, versione Linked Open Data della

classificazione Ateco2007. Dal modello mostrato in figura si evince come la definizione dell'Ateco2007 sia stata influenzata dalla classificazione NACE2006v2 dell'Eurostat a sua volta derivante dalla classificazione ATECO2002. Nel modello è riportata anche una variante d'esempio dell'Ateco 2007. Tutte le classificazioni Ateco sono raggruppate attraverso una Collection. La generazione della versione LinkedAteco2007 è rappresentata da un'attività specifica (Generation Linked Ateco2007) che vede coinvolti due attori (Agent), nello specifico l'Ufficio RST/B (i.e., unità organizzativa DCIT dell'ISTAT) e AgID. Il modello specifica inoltre che AgID è il pubblicatore (publisher), mentre l'Ufficio RST/B è il creatore (creator).

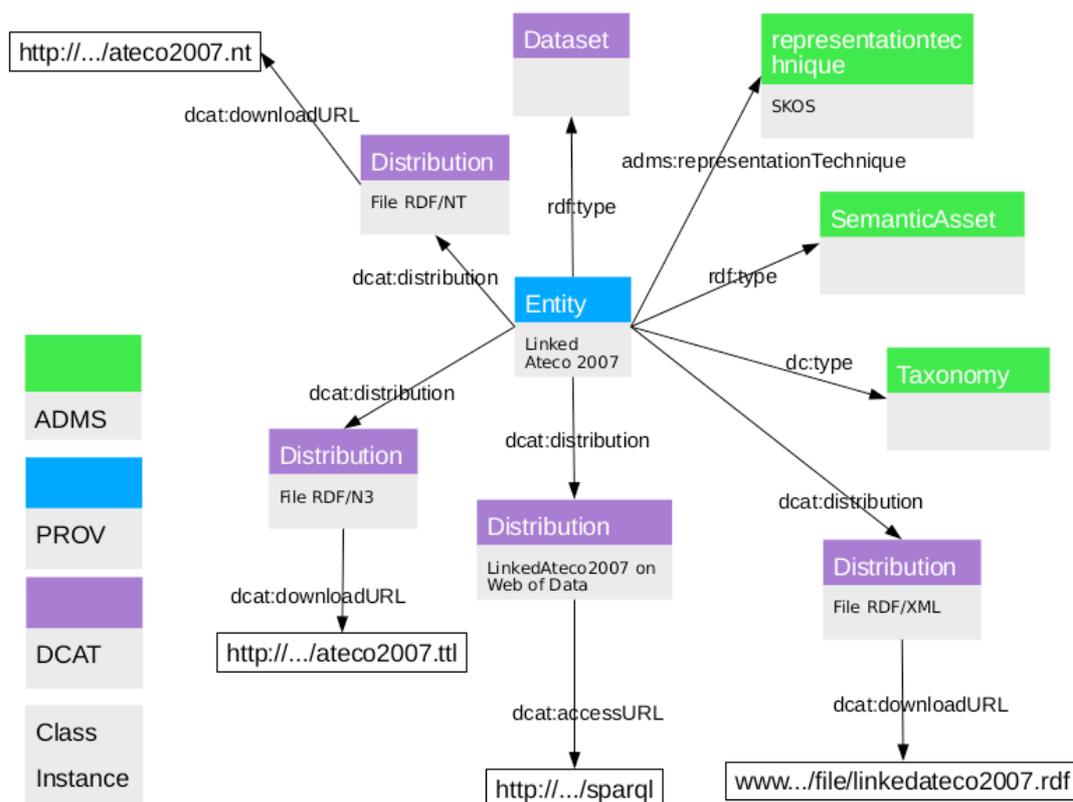


Figura 9: Dettaglio dell'entità LinkedAteco2007

La Figura 9 caratterizza ulteriormente l'entità LinkedAteco2007 attraverso le ontologie DCAT e ADMS. Nello specifico, la LinkedAteco2007 è sia un Dataset che un SemanticAsset di tipo tassonomico (Taxonomy) rappresentato mediante l'uso di SKOS. Inoltre l'entità è distribuita in diversi formati (Distribution), alcuni accessibili via Web e altri scaricabili agli indirizzi specificati.



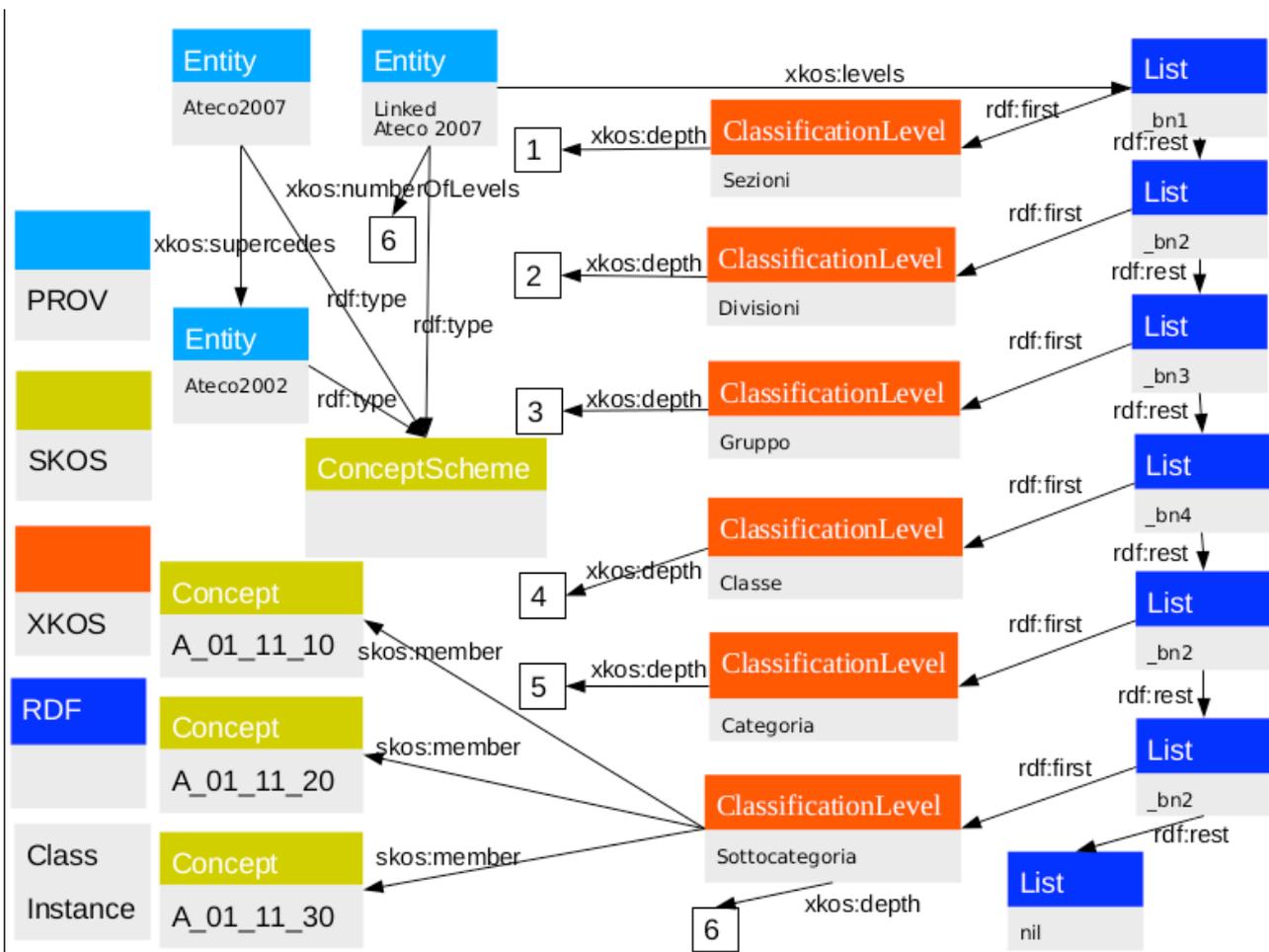


Figura 10: Ulteriore modello di dettaglio della classificazione LinkedAteco2007

La Figura 10 illustra la semantica della classificazione mediante l'uso di SKOS e XSKOS. In particolare è evidente che la classificazione è un `ConceptScheme` formato da vari livelli di classificazione, ordinati gerarchicamente (si ricorda che questa classificazione è una tassonomia). Ogni livello è caratterizzato dalla sua profondità e ha come membri gli elementi della classificazione.

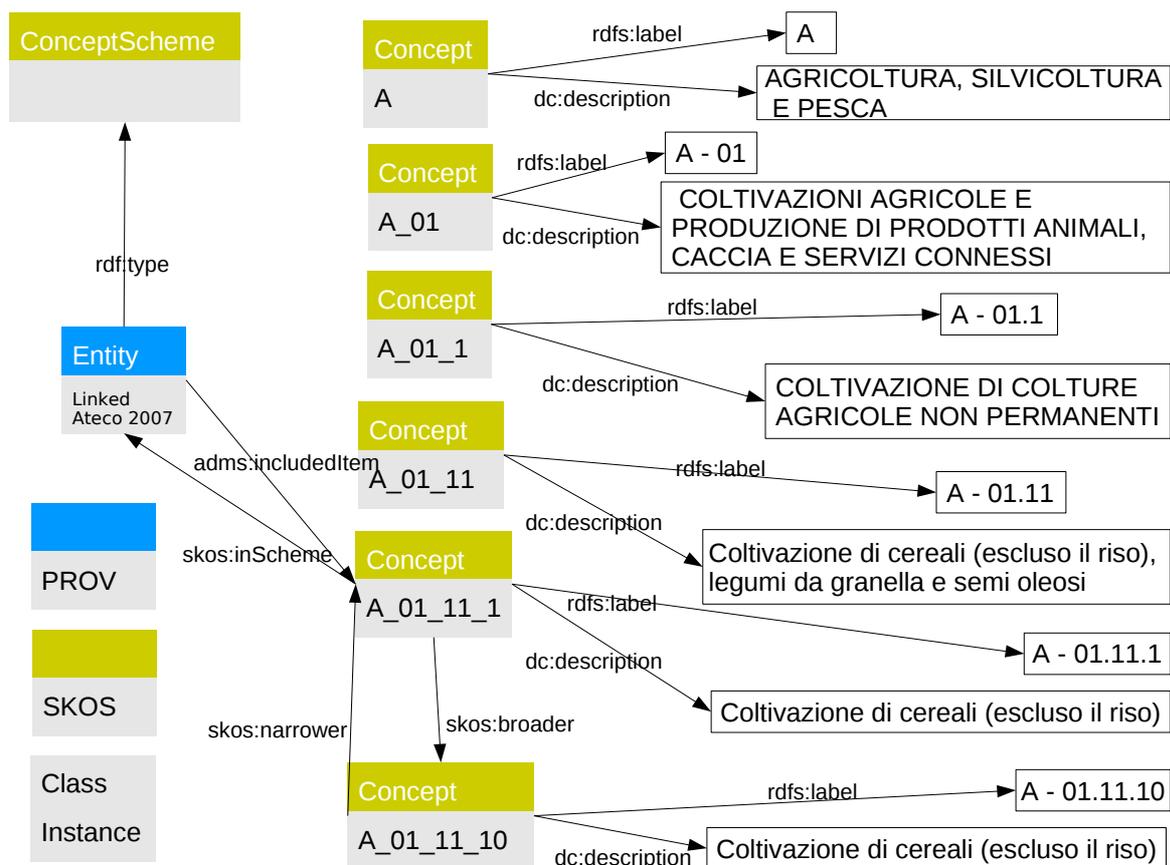


Figura 11: istanziazione di alcuni elementi della LinkedAteco2007

Infine la Figura 11 mostra come ogni elemento della classificazione ha una notazione, un'etichetta e delle specifiche testuali come note e commento. Ogni elemento della classificazione è anche messo esplicitamente in relazione con il relativo elemento del livello superiore di cui è una specializzazione e con i relativi elementi del livello inferiore di cui è una generalizzazione.

APPENDICE III

Privacy		Si/No
	i dati sono liberi da ogni informazione personale che possa identificare in modo diretto l'individuo (nome, cognome, indirizzo, codice fiscale, patente, telefono, email, foto, descrizione fisica, etc.)?	
	i dati sono liberi da ogni informazione indiretta che possa identificare l'individuo (caratteristiche personali che possono identificare facilmente il soggetto)? In caso negativo queste informazioni sono autorizzate per legge?	
	i dati sono liberi da ogni informazione sensibile che può essere ricondotta all'individuo? In caso negativo queste informazioni sono autorizzate per legge?	
	i dati sono liberi da ogni informazione relativa al soggetto che incrociata con dati comunemente reperibili nel web (e.g. google maps, linked data, etc.) possa identificare l'individuo? In caso negativo queste informazioni sono autorizzate per legge?	
	i dati sono liberi da ogni record relativo a profughi, protetti di giustizia, vittime di violenze o in ogni caso categorie protette?	
	hai usato un tool per calcolare il rischio di de-anonimizzazione del tuo dataset prima di pubblicarlo?	
	esponi dei servizi di ricerca tali da poter filtrare i dati in modo da ottenere un solo record geolocalizzato?	
Proprietà intellettuale della sorgente		



	hai creato tu i dati?	
	sei il proprietario dei dati anche se non li hai creati tu?	
	sei sicuro di non usare dati per i quali vi è una licenza o un brevetto di terzi?	
	se i dati non sono tuoi hai un accordo o una licenza che ti autorizzi a pubblicarli?	
Licenza di rilascio		
	stai rilasciando i dati di cui possiede la proprietà accompagnati da una licenza?	
	hai incluso anche la clausola di salvaguardia "In ogni caso, i dati non possono essere utilizzati al fine di identificare nuovamente gli interessati" ?	
Limiti alla pubblicazione		
	hai verificato che non vi siano impedimenti di legge o contrattuali che impediscano la pubblicazione dei dati?	
Segretezza		
	hai verificato se non vi siano motivi di ordine pubblico o di sicurezza nazionale che ti impediscono la pubblicazione dei dati?	
	hai verificato se non vi siano motivi legati al segreto d'ufficio che impediscono la pubblicazione dei dati?	
	hai verificato se non vi siano motivi legati al segreto di stato che impediscono la pubblicazione dei dati?	
Condizioni economiche		
	hai verificato di poter rilasciare in modo gratuito i dati senza infrangere qualche norma di finanza pubblica ?	



	se hai imposto condizioni economiche per l'utilizzo dei dati sei sicuro di aver imposto un prezzo per coprire i soli costi marginali?	
Temporalizzazione		
	i dati sono soggetti per legge a restrizioni temporali di pubblicazione?	
	i dati sono aggiornati frequentemente in modo da sanare eventuali informazioni lesive di persone o organizzazioni?	
	i dati hanno dei divieti di legge o giurisprudenziali che impediscono la loro indicizzazione da parte di motori di ricerca?	
Trasparenza		
	i dati rientrano nella lista dell'allegato A del d.lgs. 33/2013? se sì come sono stati trattati dal responsabile della trasparenza nella sezione "Amministrazione trasparente"?	
	In che forma si possono creare sinergie la sezione "Amministrazione Trasparente" e quella "open data" per minimizzare le inconsistenze e favorire la trasparenza del dato aperto?	

Tabella 6: Check-list di supporto all'analisi giuridica delle fonti del modello operativo

