



Vademecum

Open Data

Come rendere aperti i dati delle
pubbliche amministrazioni

Versione
Beta 2011

Il presente documento, realizzato nell'ambito delle attività finalizzate alla elaborazione delle Linee guida siti web delle pubbliche amministrazioni (previste dalla Direttiva n. 8 del 26 novembre 2009 del Ministro per la pubblica amministrazione e l'innovazione), è stato curato da *Formez PA*.

(Ottobre 2011)



Indice

Indice	5
Premessa	7
Obiettivi	10
Destinatari	12
Parte I – PA e dati aperti	14
1. Open Government	14
2. Open Data.....	17
3. Quadro Normativo.....	27
4. Come avviare il processo di apertura dei dati	37
Parte II – Come procedere per aprire i dati della PA.....	39
5. Organizzarsi per l’apertura dei dati	39
6. Identificazione dei dati	44
7. Analisi dei dataset	50

8. Pubblicazione del Dataset.....	55
9. Diffusione del Dataset	62
Glossario.....	64

Premessa

Con le “**Linee Guida per i siti Web della PA**”, pubblicate ai sensi dell’art. 4 **Direttiva del Ministro per la Pubblica Amministrazione e l’Innovazione del 26 novembre 2009, n. 8**, sono stati forniti alle amministrazioni pubbliche criteri e prescrizioni da seguire nell’attività di razionalizzazione dei siti web esistenti; in particolare, il documento detta principi in materia di caratteristiche tecniche, alla luce dei quali razionalizzare i siti web dei vari enti, dettando principi comuni in materia di realizzazione, gestione, sviluppo e aggiornamento dei contenuti pubblicati dalle Pubbliche Amministrazioni, nonché in tema di fornitura di servizi *on line*.

In base a quanto previsto dalla Direttiva n. 8/2009, l’adozione delle Linee Guida viene seguita dalla pubblicazione di *Vademecum* di approfondimento che hanno la finalità di fornire alle amministrazioni indicazioni operative su specifici temi innovativi, nonché sugli aspetti normativi, organizzativi e tecnologici ritenuti utili per dare attuazione a quanto previsto dalle Linee Guida.

Il tema della pubblicazione dei dati prodotti e detenuti dalla Pubblica Amministrazione e dell’importanza che tale patrimonio può avere in termini di efficienza del settore pubblico, di trasparenza dell’azione amministrativa nonché quale elemento per favorire la crescita economica e la competitività, è

affrontato nel capitolo 5 delle Linee Guida, “*Criteri di indirizzo e strumenti per il trattamento dei dati, della loro documentazione pubblica e per la loro reperibilità*”.

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

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

E' il fenomeno conosciuto come *Open Data*, la prassi amministrativa nell'ambito della quale alcune tipologie di dati sono rese liberamente accessibili a tutti sul Web, senza restrizioni di *copyright*, brevetti o altre forme di controllo che ne limitino la riproduzione.

Dal punto di vista normativo (su cui cfr. Parte Prima, §3), tale fenomeno rileva sotto un duplice profilo. In primo luogo in relazione all'**evoluzione del concetto di trasparenza**; come noto, la trasparenza, all'indomani del *D. Lgs. n. 150/2009*, è intesa come "*accessibilità totale (...) delle informazioni concernenti ogni aspetto dell'organizzazione, degli indicatori relativi agli*

andamenti gestionali e all'utilizzo delle risorse per il perseguimento delle funzioni istituzionali, dei risultati dell'attività di misurazione e valutazione" (articolo 11, comma 1). Si tratta di una nozione diversa – e ben più ampia – rispetto a quella contenuta negli articoli 22 e seguenti della Legge 7 agosto 1990, n. 241, che disciplina le specifiche fattispecie del diritto di accesso ai documenti amministrativi.

Come rilevato nella *Deliberazione Civit n. 105/2010*, "*l'accessibilità totale presuppone, invece, l'accesso da parte dell'intera collettività a tutte le "informazioni pubbliche", secondo il paradigma della «libertà di informazione» dell'open government di origine statunitense. Una tale disciplina è idonea a radicare, se non sempre un diritto in senso tecnico, una posizione qualificata e diffusa in capo a ciascun cittadino, rispetto all'azione delle pubbliche amministrazioni, con il principale scopo di favorire forme diffuse di controllo del rispetto dei principi di buon andamento e imparzialità"*.

In secondo luogo, nella società della conoscenza i dati pubblici diventano materia prima per l'elaborazione di ricerche, prodotti e servizi e anche il legislatore, con numerose disposizioni, ha deciso di **valorizzare e rendere fruibile l'informazione prodotta dal settore pubblico**.

A livello comunitario, al fine di agevolare il riutilizzo delle informazioni in possesso degli enti pubblici degli Stati membri, l'Unione Europea ha adottato la **Direttiva 2003/98/CE del 17 novembre 2003** (recepita dall'ordinamento italiano con il **Decreto Legislativo 24 gennaio 2006 n. 36**, "Attuazione della direttiva

2003/98/CE relativa al riutilizzo di documenti nel settore pubblico”) che ha attribuito a ciascuna Amministrazione la possibilità di autorizzare il riutilizzo delle informazioni che vengono raccolte, prodotte, e diffuse nell’ambito del perseguimento dei propri compiti istituzionali.

Successivamente, con il **Codice dell’Amministrazione Digitale** (D. Lgs. n. 82/2005) è stato introdotto l’importante principio di “*disponibilità dei dati pubblici*” (enunciato all’art. 2, comma 1, e declinato dall’art. 50, comma 1, dello stesso Codice) che consiste nella possibilità, per soggetti pubblici e privati, “*di accedere ai dati senza restrizioni non riconducibili a esplicite norme di legge*” (art.1, lett. o).

In tale ottica, con la riforma del Codice dell’Amministrazione Digitale (D. Lgs. n. 235/2010), il Legislatore ha inteso recepire espressamente la dottrina dell’*Open Data*, sollecitando le Amministrazioni ad aprire il proprio patrimonio informativo; nella sua attuale formulazione, infatti, l’art. 52, comma 1-bis, D. Lgs. n. 82/2005, prevede espressamente che “*le pubbliche amministrazioni, al fine di valorizzare e rendere fruibili i dati pubblici di cui sono titolari, promuovono progetti di elaborazione e di diffusione degli stessi anche attraverso l’uso di strumenti di finanza di progetto*”, utilizzando formati aperti che ne consentano il riutilizzo.

Obiettivi

Proprio in relazione ai più recenti interventi legislativi, si rende necessario illustrare i concetti base dell'*Open Data*, approfondendo in particolare gli aspetti tecnici, organizzativi e giuridici che gli Enti devono tenere presenti nel processo di apertura dei dati pubblici.

La legislazione vigente attribuisce al riutilizzo delle informazioni del settore pubblico un ruolo fondamentale, in quanto la gran mole dei dati pubblici rappresenta un enorme patrimonio in termini di conoscenza, trasparenza e sviluppo.

Questi dati, in alcuni casi, sono disponibili ma con una serie di vincoli di tipo tecnico (ad esempio, legati al formato utilizzato) o giuridico (relativi alla licenza di utilizzo) che ne impediscono o ne limitano il riutilizzo.

Obiettivo del presente *Vademecum* è supportare le Amministrazioni nell'applicazione delle summenzionate disposizioni e, in particolare, dell'art. 52, comma 1-bis, Codice dell'Amministrazione Digitale.

I siti web delle PA rappresentano potenzialmente il luogo preferibile per la pubblicazione dei dati pubblici da parte di ciascuna Amministrazione; bisogna, tuttavia, evitare che tale pubblicazione avvenga con modalità diverse e non sempre facilmente riconoscibili. Per questo motivo, appare particolarmente opportuno – anche per la grande quantità di informazioni a disposizione – definire principi e regole comuni per l'apertura dei dati e la loro pubblicazione.

La pubblicazione del presente *Vademecum* ha anche la finalità di armonizzare le modalità di pubblicazione dei dati delle diverse Pubbliche Amministrazioni in considerazione della realizzazione da parte del Ministero per la Pubblica Amministrazione e l'Innovazione del portale **www.dati.gov.it**. Il portale ha la funzione di condividere e diffondere gli obiettivi dell'iniziativa, i materiali e i documenti di riferimento e – soprattutto – il catalogo di dati resi pubblici da tutte le Amministrazioni italiane.

Destinatari

Il presente *Vademecum* è destinato ad amministratori, dirigenti e dipendenti pubblici, nonché a fornitori e consulenti delle pubbliche amministrazioni.

In particolare, il documento è concepito – in prima battuta – per supportare gli organi di governo nell'emanazione delle direttive generali in materia di *Open Data*; la **Prima Parte** del *Vademecum* contiene, infatti, un'introduzione ai concetti principali e l'indicazione degli elementi necessari ad avviare il processo di apertura dei dati pubblici.

Il documento è inoltre diretto a tutti coloro che, a vario titolo e con diverse responsabilità, partecipano alla gestione, manutenzione e pubblicazione del patrimonio informativo dell'Ente; la **Seconda Parte** contiene la descrizione degli aspetti tecnici, organizzativi e giuridici da affrontare prima di rendere disponibili i dati dell'Amministrazione.

Il *Vademecum* è corredato da un glossario e da schede di riepilogo sugli argomenti trattati; i testi normativi, i materiali e la documentazione di riferimento sono disponibili nella apposita sezione del sito <http://www.dati.gov.it/>.

Il presente documento viene periodicamente aggiornato per tenere conto dell'evoluzione delle tecnologie e del quadro normativo europeo e nazionale; a tal fine, il *Vademecum* è aperto ai contributi di amministrazioni, sviluppatori e associazioni.

Parte I – PA e dati aperti

1. Open Government

La dottrina dell'*Open Government* si basa sul principio per il quale tutte le attività dei Governi e delle Amministrazioni dello Stato devono essere aperte e disponibili per favorire azioni efficaci e garantire un controllo diffuso sulla gestione della cosa pubblica.

Tale dottrina, che pure ha origine come principio filosofico nell'Illuminismo, vede un concreto sviluppo nell'ultimo quinquennio, a valle dei provvedimenti legislativi presi dal Governo statunitense.

L'*Open Government* ridefinisce alla base il rapporto tra Pubblica Amministrazione e cittadino, spostando il *focus* della relazione da un approccio orientato all'erogazione di servizi in cui il cittadino è fruitore di prestazioni erogate dall'Amministrazione ad uno basato su un processo di collaborazione reale, in cui il cittadino partecipa alle scelte di governo.

In tal senso, l'*Open Government* si basa su tre elementi¹:

¹ B. Obama, Memorandum for the Heads of Executive Departments and Agencies on

- **Trasparenza:** La trasparenza favorisce e promuove la responsabilità fornendo ai cittadini le informazioni sulle attività dell'Amministrazione. Un'Amministrazione trasparente, per questo, è un'Amministrazione più controllata e nel contempo più aperta e affidabile. Per tale motivo le Amministrazioni che si muovono nella direzione dell'*Open Government* devono prendere tutti i provvedimenti legislativi e regolamentari necessari a far sì che le informazioni delle quali dispongono siano facilmente reperibili, riutilizzabili, aperte.
- **Partecipazione:** La partecipazione dei cittadini alle scelte della Pubblica Amministrazione aumenta l'efficacia dell'azione amministrativa e migliora la qualità delle decisioni dell'Amministrazione. I cittadini devono perciò essere coinvolti nei processi decisionali e potersi contribuire attivamente, anche grazie al ricorso alle tecnologie di comunicazione attualmente disponibili e diffuse.
- **Collaborazione:** La collaborazione vede un coinvolgimento diretto dei cittadini nelle attività dell'Amministrazione. Anche in tal senso l'*Open Government* traccia uno scenario nuovo nelle dinamiche di relazione tra l'Amministrazione e i suoi *stakeholder* (essendo questi ultimi i singoli cittadini, ma anche le organizzazioni non-profit e le imprese). Tale coinvolgimento non sottende certo una delega di responsabilità dell'Amministrazione nei riguardi del cittadino, contempla però la possibilità – per quest'ultimo – di monitorare la qualità del servizio pubblico in tutte le sue fasi di esecuzione.

Trasparenza, partecipazione e collaborazione hanno il risultato di mutare profondamente gli schemi operativi dell'Amministrazione Pubblica e i processi decisionali sui quali si basa. Tale cambiamento non riguarda soltanto le modalità e gli strumenti attraverso i quali si espleta la relazione con il cittadino, ma è relativo ad una mutazione più profonda, che attiene il modo in cui l'Amministrazione opera e la percezione del suo ruolo al servizio della comunità. Con l'*Open Government* si va quindi nella direzione di un'Amministrazione aperta, e per questo in grado di costruire una relazione con il cittadino basata sulla fiducia.

“La funzione pubblica è depositaria della fiducia del pubblico. I cittadini contano sull'impegno dei funzionari ad operare per l'interesse generale, dimostrando imparzialità e amministrando quotidianamente le risorse pubbliche in modo appropriato. Un'equa e affidabile pubblica amministrazione ispira fiducia al pubblico e crea un clima favorevole alle imprese, contribuendo quindi al buon funzionamento dei mercati e alla crescita economica”²

In questa logica l'Amministrazione che si muove nella direzione *dell'Open Government* mette, al centro del suo agire, trasparenza, collaborazione e partecipazione, si apre al dialogo e al confronto diretto con i privati e quindi focalizza i processi decisionali sulle effettive esigenze e necessità della comunità di riferimento. Centralità del cittadino, Amministrazione partecipata e

² Nota di sintesi dell'OCSE sulla Gestione Pubblica: Rafforzare l'Etica nella Pubblica Amministrazione, Settembre 2000 < <http://www.oecd.org/dataoecd/61/9/1899767.pdf>>

collaborativa, insieme a trasparenza, apertura e condivisione, quindi, sono i tratti distintivi dell'Open Government.

2. Open Data

Per attuare i principi di trasparenza, partecipazione e collaborazione propri della dottrina dell'*Open Government* è necessario mettere il cittadino nelle condizioni di disporre degli strumenti conoscitivi indispensabili per poter prendere decisioni o comunque valutare le decisioni prese dall'Amministrazione.

Tali strumenti sono essenzialmente costituiti dai dati e dalle informazioni dei quali la Pubblica Amministrazione dispone e che le sono indispensabili per la gestione dei processi che gestisce nell'assolvimento dei suoi compiti istituzionali e di servizio.

2.1 Dall'Open Government all'Open Data

Mettere a disposizione del cittadino e delle imprese l'insieme dei dati pubblici gestiti dall'Amministrazione in formato aperto rappresenta un passaggio culturale necessario per il rinnovamento delle istituzioni nella direzione di apertura e trasparenza proprie dell'*Open Government*, a tutti i livelli amministrativi.

Il libero accesso all'informazione pubblica si configura infatti come un aspetto fondamentale per tradurre il concetto di *Open Government* in un vero e proprio modello sostenibile, in quanto consente di:

- **Rendere l'amministrazione trasparente.** Il libero accesso a documenti, atti e saperi sul governo della *res publica* e sulle scelte politico-istituzionali compiute dalle Amministrazioni è un aspetto centrale per la trasparenza delle Istituzioni. Tale aspetto stimola e facilita i cittadini ad un controllo continuo e costante sull'operato e sui processi decisionali dei soggetti istituzionali. Attraverso l'attuazione di politiche di apertura reale delle informazioni e dei dati pubblici, i cittadini sono in condizione di verificare l'efficienza dell'apparato burocratico.
- **Rendere l'amministrazione aperta.** Distribuire i dati pubblici in un formato aperto e libero da restrizioni sia dal punto di vista dell'accesso che dell'integrazione e del riutilizzo, rappresenta il presupposto di base affinché possa svilupparsi un vero e proprio processo di collaborazione tra le istituzioni e la comunità dei cittadini sulle scelte di governo, anche la rielaborazione in forma nuova e diversa dei dati messi a disposizione. Mediante strategie di apertura dei dati della Pubblica Amministrazione i cittadini non sono più soltanto consumatori passivi di informazioni messe a disposizione dalle Amministrazioni. Hanno invece l'opportunità di riutilizzare e integrare i dati messi loro a disposizione, fino a sviluppare servizi e applicazioni a vantaggio dell'intera comunità di utenti, che vanno ad affiancarsi a quelli creati centralmente dalle Istituzioni. In questo modo i cittadini collaborano effettivamente con i soggetti istituzionali e partecipano attivamente alle azioni di governo della cosa pubblica.

Le politiche e le pratiche di apertura dei dati dell'Amministrazione sin qui descritte rientrano nella definizione del concetto di *Open Data*.

2.2 Cosa si intende per *Open Data*

Una definizione comunemente accettata di *Open Data* è quella fornita dall'*Open Data Manual*³, che definisce gli *Open Data* come

“dati che possono essere liberamente utilizzati, riutilizzati e redistribuiti, con la sola limitazione – al massimo – della richiesta di attribuzione dell'autore e della redistribuzione allo stesso modo (ossia senza che vengano effettuate modifiche)”

Un insieme di dati pubblicati prende il nome di *dataset*.

È bene differenziare sin da subito il concetto di trasparenza da quello di apertura. Il concetto di apertura include quello di trasparenza, ma non necessariamente è vero il contrario. In altri termini, non è sufficiente la trasparenza così come definita nel nostro ordinamento giuridico perché si possa parlare di *Open Data*.

Partendo dal concetto di conoscenza aperta così come delineato dalla *Open Knowledge Foundation*⁴, anche l'*Open Data* può essere caratterizzato dai seguenti principi⁵:

³ <http://opendatamanual.org/> – Documento redatto dall'Open Knowledge Foundation

⁴ Open Knowledge Foundation – OKF (Fondazione per la conoscenza aperta) è una fondazione non profit, fondata nel 2004, che ha l'obiettivo di promuovere la conoscenza aperta <<http://okfn.org/about/>>

⁵ Adattamento della descrizione del concetto di “Conoscenza Aperta” della Open Knowledge Foundation

- **Disponibilità e accesso:** i dati devono essere disponibili nel loro complesso, per un prezzo non superiore a un ragionevole costo di riproduzione, preferibilmente mediante scaricamento da Internet. I dati devono essere inoltre disponibili in un formato utile e modificabile.
- **Riutilizzo e redistribuzione:** i dati devono essere forniti a condizioni tali da permetterne il riutilizzo e la redistribuzione. Ciò comprende la possibilità di combinarli con altre basi di dati.
- **Partecipazione universale:** tutti devono essere in grado di usare, riutilizzare e redistribuire i dati. Non devono essere poste discriminazioni di ambiti di iniziativa in riferimento a soggetti o gruppi. Per esempio, il divieto di utilizzare i dati per scopi commerciali o le restrizioni che permettono l'uso solo per determinati fini (quale quello educativo) non sono contemplabili.

Per garantire i principi sopra elencati è necessario che i dati – per considerarsi aperti in base agli standard internazionali – siano⁶:

- **Completi.** I dati devono comprendere tutte le componenti (metadati) che consentano di esportarli, utilizzarli *on line* e *off line*, integrarli e aggregarli con altre risorse e diffonderli in rete.
- **Primari.** Le risorse digitali devono essere strutturate in modo tale che i dati siano presentati in maniera sufficientemente granulare, così che possano essere utilizzate dagli utenti per integrarle e aggregarle con altri dati e contenuti in formato digitale;
- **Tempestivi.** Gli utenti devono essere messi in condizione di accedere e utilizzare i dati presenti in rete in modo rapido e immediato,

⁶ Transparency International Georgia, Ten Open Data Guidelines
<http://transparency.ge/en/ten-open-data-guidelines>

massimizzando il valore e l'utilità derivanti da accesso e uso di queste risorse;

- **Accessibili.** I dati devono essere resi disponibili al maggior numero possibile di utenti senza barriere all'utilizzo, quindi preferibilmente attraverso il solo protocollo *Hypertext Transfer Protocol (HTTP)* e senza il ricorso a piattaforme proprietarie. Devono essere inoltre resi disponibili senza alcuna sottoscrizione di contratto, pagamento, registrazione o richiesta.
- **Leggibili da computer.** Per garantire agli utenti la piena libertà di accesso e soprattutto di utilizzo e integrazione dei contenuti digitali, è necessario che i dati siano *machine-readable*, ovvero processabili in automatico dal computer.
- **In formati non proprietari.** I dati devono essere codificati in formati aperti e pubblici, sui quali non vi siano entità (aziende o organizzazioni) che ne abbiano il controllo esclusivo. Sono preferibili i formati con le codifiche più semplici e maggiormente supportati.
- **Liberi da licenze che ne limitino l'uso.** I dati aperti devono essere caratterizzati da licenze che non ne limitino l'uso, la diffusione o la redistribuzione.
- **Riutilizzabili.** Affinché i dati siano effettivamente aperti, gli utenti devono essere messi in condizione di riutilizzarli e integrarli, fino a creare nuove risorse, applicazioni e servizi di pubblica utilità.
- **Ricercabili.** I dati devono essere facilmente identificabili in rete, grazie a cataloghi e archivi facilmente indicizzabili dai motori di ricerca.
- **Permanenti.** Le peculiarità fino ad ora descritte devono caratterizzare i dati nel corso del loro intero ciclo di vita.

2.3 Dal Raw Data Now al Linked Open Data

Per mettere a disposizione del pubblico i dati di un'Amministrazione può essere utilizzata una grande varietà di formati. Non tutti, tuttavia, rispondono ai prerequisiti elencati nel paragrafo precedente.

Qualora l'Amministrazione si trovi nella possibilità di scegliere il formato nel quale pubblicare i dati da aprire, è bene che identifichi formati che siano rispondenti al maggior numero di requisiti possibile. Se tuttavia la scelta fosse tra il pubblicare dati non completamente rispondenti ai requisiti indicati o non pubblicarli affatto, allora la logica dell'*Open Data* indirizza la risposta verso la prima soluzione (per questo si usa l'espressione *Raw Data Now*, a significare che nella peggiore delle ipotesi, quella di disporre di dati non aperti, è comunque preferibile distribuirli, anche se – appunto – in formato raw, cioè grezzo). L'auspicio è che se i dati pubblicati sono sufficientemente interessanti, la comunità degli utenti si preoccuperà di identificare un processo di conversione efficace (si parla, in questo caso, di *data scraping*).

Per distinguere i diversi formati utilizzabili nella codifica dei set di dati, è stato proposto in seno al W3C⁷ un modello di catalogazione che li classifica in base alle loro caratteristiche su una scala di valori da 1 (una stella) a 5 (cinque stelle):

- (★) **Una Stella.** È il livello base, costituito da file non strutturati: ad esempio un'immagine in formato grezzo (formati come .gif, .jpg, .png), un documento in formato Microsoft Word, un file in formato Adobe Pdf. Una sola stella indica la semplice disponibilità di una informazione e di

⁷ La proposta è di Tim Berners Lee, Presidente del W3C e ideatore del World Wide Web

un dato *on line*, in un formato qualsiasi, purché distribuito con licenza aperta. I dati distribuiti in questo formato sono leggibili e stampabili dagli utenti, possono essere conservati localmente su un PC e sono semplici da pubblicare. Tuttavia non sono un formato aperto in quanto non è possibile effettuare su di essi alcuna elaborazione.

- (★★) **Due Stelle.** Questo livello indica dati strutturati ma codificati con un formato proprietario. Ad esempio un documento in formato Microsoft Excel.

Due stelle indicano, oltre alle possibilità offerte dai dati contraddistinti da una sola stella, la possibilità di effettuare elaborazioni sui dati, a patto di disporre del *software* necessario a gestire un file codificato con un formato proprietario. I dati caratterizzati dalle due stelle non sono un formato aperto in quanto per elaborarli è necessario un *software* proprietario, tuttavia di norma possono essere convertiti – essendo dati strutturati – in dati aperti.

- (★★★) **Tre Stelle.** Questo livello indica dati strutturati e codificati in un formato non proprietario. Ad esempio il formato *.csv (Comma Separated Values)* al posto – ad esempio – del formato Microsoft Excel utilizzato nel caso precedente.

Tre stelle indicano, oltre alle possibilità offerte dai dati contraddistinti da due sole stelle, la possibilità di effettuare elaborazioni sui dati senza esser costretti ad utilizzare software proprietario. Quello caratterizzato dalle tre stelle è il formato più semplice di dati aperti.

- (★★★★) **Quattro Stelle.** Questo livello indica dati strutturati e codificati in un formato non proprietario che sono dotati di un URI⁸ che li rende indirizzabili sulla rete e quindi utilizzabili direttamente online, attraverso l'inclusione in una struttura basata sul modello RDF (*Resource Description Framework*).

Quattro stelle indicano quindi il fatto che il singolo dato di un dataset, disponibile *on line* in un formato aperto (tipicamente XML/RDF) può essere richiamato attraverso un'URL (*Uniform Resource Locator*) specifico. Ciò consente di puntare al dato o ad un insieme di dati da una applicazione o accedervi dall'interno di un programma che può poi elaborarlo in vari modi.

Si pensi, ad esempio, a un *dataset* contenente gli indirizzi dei monumenti di una città opportunamente codificati: da qualsiasi software – finanche dal *browser* – è possibile collegarsi all'URL che indica il singolo monumento, potendolo ad esempio georeferenziare su una mappa.

- (★★★★★) **Cinque Stelle.** Questo livello indica quelli che vengono definiti *Linked Open Data* (LOD). Quei dati aperti, cioè, che – dal punto di vista del formato – oltre a rispondere alle caratteristiche indicate al punto precedente (classificazione a quattro stelle) presentano anche, nella struttura del *dataset*, collegamenti ad altri *dataset*. In altri termini, grazie al ricorso al già citato modello di descrizione dei dati RDF, è possibile collegare dinamicamente tra loro più *dataset*, incrociando così

⁸ *Uniform Resource Identifier* – Identificatore Univoco di Risorsa, ossia di norma un indirizzo internet nel formato <http://www.nomedominio.it/>

informazioni provenienti da fonti diverse, eventualmente gestite da diverse Amministrazioni. Si pensi ad esempio al caso del *dataset* contenente gli indirizzi dei monumenti di una città alla quale si è fatto riferimento in precedenza. Tale *dataset*, pubblicato dall'Amministrazione A, potrebbe essere collegato al *dataset* di un altro Ente – l'Amministrazione B – che dispone dell'archivio delle opere d'arte presenti all'interno di ogni monumento. In questo caso un sistema *software* potrebbe, dopo aver chiesto all'utente i suoi gusti e le sue preferenze in fatto di arte, strutturare un percorso georeferenziando i monumenti a partire dalle informazioni presenti nel *dataset* disponibile presso l'Amministrazione A, dopo averli preselezionati in base alle opere d'arte in essi contenute, identificate nel *dataset* dell'Amministrazione B.

I *Linked Open Data*, quindi, consentono di combinare i contenuti di *dataset* diversi grazie a costrutti formali formulati secondo il modello RDF in uno dei diversi formati esistenti (XML/RDF, N3, ecc...). Ciò aumenta esponenzialmente il valore dei *dataset* reciprocamente correlati, consentendo il passaggio dal livello dei dati a quello dell'informazione e quindi a quello della conoscenza e fornendo così un quadro di contesto strutturato a partire dalla correlazione di informazioni provenienti da fonti diverse.

Una delle opportunità più importanti dell'*Open Data* è rappresentata dall'interoperabilità. Il valore dei dati, infatti, è tanto più alto quanto più è possibile effettuare correlazioni tra più *dataset* indipendenti l'uno dall'altro, ma

interoperabili nel formato e nel *data model*. Per questo motivo è auspicabile che i dati vengano aperti in modalità Linked Open Data rispetto alle altre tipologie di dati aperti, essendo i LOD la tipologia di dati aperti che consente il massimo livello di interoperabilità tra dataset diversi.

3. Quadro Normativo

La pianificazione e l'attuazione di un'operazione di *Open Data* in ambito pubblico, oltre alla conformità agli standard tecnici, deve avvenire nel rispetto della normativa nazionale vigente.

Di seguito, riassuntivamente, sono indicati i principali provvedimenti normativi rilevanti ai fini del presente *Vademecum*.

Provvedimento	Descrizione	Parole Chiave
Legge n. 633/1941	Legge a protezione del diritto d'autore e di altri diritti connessi al suo esercizio	<i>Copyright</i> , Diritto d'autore, Titolarità dei contenuti e dei dati pubblici, Licenze per il riutilizzo
Legge n. 241/1990	Nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti amministrativi	Trasparenza
D. Lgs. n. 196/2003	Codice in materia di protezione dei dati personali	Privacy, Sicurezza
Legge n. 4/2004	Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici	Accessibilità
D. Lgs. n. 82/2005	Codice dell'Amministrazione	Diritto dei cittadini all'uso delle tecnologie, E-

	Digitale (CAD)	<i>democracy, E-participation</i> , Titolarità dei dati pubblici, Formati aperti
D. Lgs. n. 36/2006	Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico	Riutilizzo delle informazioni del settore pubblico
D. Lgs. n. 163/2006	Codice dei contratti pubblici in materia di lavori, servizi e forniture	Contratti, Titolarità dei contenuti acquisiti dalle Amministrazioni
D. Lgs. n. 150/2009	Ottimizzazione della produttività del lavoro pubblico e di efficienza e trasparenza delle pubbliche amministrazioni	Trasparenza, valutazione e merito, Programma triennale per la trasparenza
D. Lgs. n. 32/2010	Attuazione della direttiva 2007/2/CE, che istituisce un'infrastruttura per l'informazione territoriale nella Comunità europea (INSPIRE)	Pubblicità, informazione territoriale, riutilizzo

Il complesso di questi provvedimenti detta alcuni principi fondamentali, la cui portata può essere così riassunta:

a) essere trasparenti in modo nuovo

Nell'ordinamento giuridico italiano, la trasparenza, e più in generale la disciplina del rapporto dialogico tra Pubblica Amministrazione e cittadino, è un

istituto di recente introduzione, essendo stata prevista solo dalla Legge n. 241/1990. Da quel momento, essa è diventata corollario del principio di buona amministrazione costituzionalmente garantito dall'art. 97 Costituzione: infatti, è previsto che *"l'attività amministrativa (...) è retta da criteri di economicità, di efficacia, di imparzialità, di pubblicità e di trasparenza"* (art. 1, Legge n. 241/1990).

Tuttavia, si tratta di un principio tutt'altro che statico: la trasparenza, infatti, costituisce uno dei gangli del diritto pubblico maggiormente soggetto all'evoluzione politica, sociale e tecnologica. Ed infatti, dopo l'introduzione dell'informatica nell'attività amministrativa e il ripensamento sul ruolo del settore pubblico, l'originario concetto di trasparenza - circoscritto al diritto di accesso agli atti e ai documenti per coloro che avessero specifico e concreto interesse - ha iniziato a dimostrarsi insufficiente.

La digitalizzazione dell'attività amministrativa conferisce nuovi significati al concetto di trasparenza. In proposito, è opportuno ricordare come il **Codice dell'Amministrazione Digitale (D. Lgs. n. 82/2005)**, prevede che *"le pubbliche amministrazioni nell'organizzare autonomamente la propria attività utilizzano le tecnologie dell'informazione e della comunicazione per la realizzazione degli obiettivi di efficienza, efficacia, economicità, imparzialità, trasparenza, semplificazione e partecipazione"* (art. 12, D. Lgs. n. 82/2005). Questo significa che le Amministrazioni sono tenute per legge ad assicurare *"la disponibilità, la gestione, l'accesso, la trasmissione, la conservazione e la fruibilità dell'informazione in modalità digitale"* (art. 2, D. Lgs. n. 82/2005). A tal fine è

inoltre previsto che *"i dati delle pubbliche amministrazioni sono formati, raccolti, conservati, resi disponibili e accessibili con l'uso delle tecnologie dell'informazione e della comunicazione che ne consentano la fruizione e riutilizzo da parte delle altre pubbliche amministrazioni e dai privati"* (art. 50, D. Lgs. n. 82/2005).

In tale mutato contesto normativo, il Legislatore è successivamente intervenuto, nel 2009, ponendo una nuova concezione normativa di trasparenza, intesa come *"accessibilità totale (...) delle informazioni concernenti ogni aspetto dell'organizzazione, degli indicatori relativi agli andamenti gestionali e all'utilizzo delle risorse per il perseguimento delle funzioni istituzionali, dei risultati dell'attività di misurazione e valutazione"*, così introducendo nel nostro ordinamento una nuova posizione qualificata in capo a ciascun cittadino, rispetto all'azione degli Enti, con il preciso *"scopo di favorire forme diffuse di controllo del rispetto dei principi di buon andamento e imparzialità"* (art. 11, D. Lgs. n. 150/2009).

In conclusione, alla luce dell'appena descritto processo legislativo, si è prodotta un'evidente maturazione del canone costituzionale della trasparenza, il quale non può essere più soltanto uno strumento finalizzato al controllo del procedimento amministrativo, bensì un vero e proprio risultato dell'azione amministrativa; non più, dunque, solo un criterio informatore, ma un obiettivo di essa.

Si può quindi affermare che non c'è vera trasparenza se l'Amministrazione non rende pubbliche *on line* le informazioni formate e raccolte nell'esercizio della propria attività: tale concetto si ricava, in maniera assai significativa, dalla **Delibera n. 105/2010 della Commissione per la Valutazione la Trasparenza e l'Integrità delle Amministrazioni Pubbliche** con la quale sono state adottate le *Linee guida per la predisposizione del programma triennale per la trasparenza*.

La Delibera n. 105/2010⁹ illustra in modo approfondito l'accennata evoluzione del concetto di trasparenza da una concezione definita statica, legata cioè alla pubblicità di determinate categorie di dati attinenti alle Pubbliche Amministrazioni, a quella dinamica che è correlata alla valutazione della *performance*. Si tratta di uno dei primi atti in cui si fa espresso riferimento alla pratica amministrativa dell'*Open Government*, che viene espressamente indicata come buona prassi che anche le Amministrazioni italiane dovrebbero seguire; all'interno del documento sono altresì individuati alcuni punti critici della normativa vigente (principalmente in materia di privacy), aspetti che – se non correttamente valutati ed affrontati – corrono il rischio di diventare un vero e proprio limite all'azione di trasparenza intrapresa dalle Amministrazioni.

b) incentivare il riutilizzo dei dati pubblici

Il principio di disponibilità dei dati pubblici, introdotto dal Codice dell'Amministrazione Digitale, non è rilevante solo sotto il profilo della

⁹ Disponibile all'indirizzo <http://www.civit.it/?p=2074>

trasparenza, ma anche per le sue implicazioni organizzative e di impulso al sistema economico–imprenditoriale¹⁰.

Infatti, per assicurare l’effettività di questa disposizione il D. Lgs. n. 82/2005 ha posto un vero e proprio obbligo per le Amministrazioni di attivarsi a livello tecnologico e organizzativo. La disponibilità del dato in formato digitale presuppone, infatti, che lo stesso dato sia formato, raccolto e conservato con l’uso delle tecnologie dell’informazione e della comunicazione che consentano la fruizione e riutilizzazione da parte di cittadini, imprese ed altri uffici pubblici.

Tale nuova finalità è stata da ultimo rafforzata dalla riforma del **Codice dell’Amministrazione Digitale** che, all’art. 52, comma–1–bis, ora prevede che le Pubbliche Amministrazioni devono promuovere “*progetti di elaborazione e di diffusione dei dati pubblici di cui sono titolari*”, nonché assicurarne la pubblicazione “*in formati aperti*”, al fine di “*valorizzare e rendere fruibili*” i dati stessi.

La disponibilità dei dati delle pubbliche amministrazioni è ormai espressamente preordinata alla “*riutilizzazione*” dei dati. Si tratta di un concetto introdotto nel nostro ordinamento dalla **Direttiva 2003/98/CE del 17 novembre 2003**

¹⁰ È importante precisare l’ambito di applicazione di questa disposizione. Il principio di disponibilità ha portata generale sia sotto il profilo oggettivo sia sotto quello soggettivo; la portata generalissima del principio di disponibilità è mitigata da alcune (tassative) eccezioni previste dall’art. 2, comma 6, D. Lgs. n. 82/2005, e relative all’esercizio delle attività e funzioni di ordine e di sicurezza pubblica, difesa e sicurezza nazionale e consultazioni elettorali.

(recepita nel nostro ordinamento con il D. Lgs. n. 36/2006). La normativa comunitaria si basa sulla considerazione che i dati detenuti dalle Pubbliche Amministrazioni costituiscono una vera e propria risorsa da sfruttare per la crescita economica e per la creazione di posti di lavoro e, pertanto, ne promuove il riutilizzo.

Per riutilizzo s'intende *“l'uso di documenti in possesso di enti pubblici da parte di persone fisiche o giuridiche a fini commerciali o non commerciali diversi dallo scopo iniziale nell'ambito dei compiti di servizio pubblico per i quali i documenti sono stati prodotti. Lo scambio di documenti tra enti pubblici esclusivamente in adempimento dei loro compiti di servizio pubblico non costituisce riutilizzo”* (art. 2, comma 1, lett. e) D. Lgs. n. 36/2006); la stessa normativa di recepimento della Direttiva comunitaria n. 2003/98/CE, ha regolato modalità, criteri e principi del riutilizzo.

c) rispettare la privacy

Disponibilità dei dati pubblici non significa però automatica condivisione di tutte le informazioni o accesso indiscriminato alle stesse. I limiti alla conoscibilità dei dati rimangono quelli previsti dalle leggi e dai regolamenti vigenti anche con riferimento alla riservatezza dei soggetti a cui i dati si riferiscono (che andrà garantita ai sensi del **D. Lgs. n. 196/2003, Codice in materia di protezione dei dati personali**).

Sotto questo profilo, appare importante sottolineare come fare *Open Data* non implichi che l'Ente debba rendere pubblici tutti i dati formati nell'esercizio delle proprie attività istituzionali!

Il rispetto della riservatezza degli individui e delle imprese è infatti una condizione per assicurare l'approvazione da parte dei cittadini per le operazioni di apertura dei dati pubblici, oltre che presupposto per conservare la fiducia degli individui nei confronti delle istituzioni.

In materia di privacy, numerosi ed importanti principi che possono guidare le Amministrazioni nella definizione delle soluzioni e delle modalità con cui procedere alla pubblicazione sono contenuti in una Deliberazione dell'Autorità Garante per la protezione dei dati personali con le quali sono state adottate le *"Linee guida in materia di trattamento di dati personali contenuti anche in atti e documenti amministrativi, effettuato da soggetti pubblici per finalità di pubblicazione e diffusione sul Web"* (Deliberazione n. 88/2011¹¹).

La Deliberazione n. 88/2011 ha lo scopo di definire un primo quadro unitario di misure e accorgimenti finalizzati a individuare opportune cautele che i soggetti pubblici sono tenuti ad applicare in relazione alle ipotesi di pubblicazione dei propri dati sul Web; come noto, questo documento indica tutti gli accorgimenti idonei ad assicurare che la pubblicazione delle informazioni sia conforme alla normativa dettata in materia di protezione dei dati personali.

Tuttavia, la privacy non deve essere vissuta come un ostacolo insormontabile nel processo di apertura delle informazioni del settore pubblico. Infatti, la gran parte dei dati pubblici (basti pensare alle cartografie, oppure alle informazioni relative all'inquinamento) non possono essere classificati come personali, in quanto - cioè - non riconducibili ad un soggetto. In tutti gli altri casi, la privacy può essere efficacemente tutelata pubblicando i dati in forma anonima o comunque adottando tutte le cautele idonee a evitare che i soggetti cui i dati si

¹¹ Disponibile all'indirizzo <http://www.garanteprivacy.it/garante/doc.jsp?ID=1793203>

riferiscono (siano essi individui, imprese, associazioni e Enti) possano essere identificati.

Tale impostazione è confermata da quanto affermato dal Garante Privacy nel provvedimento n. 88/2011 nel quale è confermato che il perseguimento della finalità di trasparenza dell'attività delle Pubbliche Amministrazioni può avvenire anche senza l'utilizzo di dati personali.

Secondo il Garante, infatti, *"non si ravvisa la necessità di adottare alcuna specifica cautela qualora le pubbliche amministrazioni ritengano di pubblicare sul sito web informazioni non riconducibili a persone identificate o identificabili (ad esempio dati quantitativi aggregati per uffici riguardanti i livelli retributivi ed accessori risultanti dai contratti collettivi o da atti interni di organizzazione; tassi di assenza e di maggiore presenza del personale; informazioni relative alla performance dell'amministrazione; obiettivi assegnati agli uffici insieme ai relativi indicatori e ai risultati complessivi raggiunti; l'ammontare complessivo dei premi collegati alla performance stanziati e di quelli effettivamente distribuiti; dati relativi al grado di differenziazione nell'utilizzo della premialità, informazioni concernenti la dimensione della qualità dei servizi erogati, notizie circa la gestione dei pagamenti e le buone prassi)".*

Di conseguenza, è opportuno che le Amministrazioni – nell'attività di apertura dei dati che detengono – valutino quali accorgimenti porre in essere al fine di evitare la diffusione di dati personali non consentita dalla legislazione vigente,

senza pregiudicare le finalità di trasparenza e di comunicazione alla base dell'*Open Data*.

d) fare attenzione al diritto d'autore

Ulteriore profilo normativo da affrontare nell'ambito del processo di liberazione delle informazioni del settore pubblico è quello relativo alla legislazione in materia di diritto d'autore (in particolare, **Legge n. 633/1941**).

Nel nostro ordinamento, il diritto d'autore tutela le opere dell'ingegno dotate di creatività; è molto ampia la latitudine delle opere coperte da questa tutela: dalle forme espressive di opere letterarie (come lettere, rapporti e guide), a quelle grafiche (come cartografie e grafici), ed anche le raccolte ordinate di dati e informazioni o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili mediante mezzi elettronici o in altro modo (banche dati o *database*)¹².

La normativa in materia di diritto d'autore si applica anche ai contenuti prodotti dal settore pubblico, salvo le eccezioni previste normativamente; ad esempio, l'art. 5 Legge n. 633/1941 dispone che *“le disposizioni di questa legge non si*

¹² Per le banche dati, il legislatore dispone una doppia tutela: 1) l'art. 64-quinquies, Legge n. 633/1941 concede all'autore il diritto esclusivo di eseguire o autorizzare la riproduzione permanente o temporanea, totale o parziale, con qualsiasi mezzo o forma, la traduzione, l'adattamento, e le modifiche, nonché qualsiasi forma di distribuzione al pubblico dell'originale o di copie; 2) l'art. 102-bis Legge n. 633/1941, a prescindere dall'esistenza di qualunque requisito di originalità, tutela il costituente (vale a dire il soggetto effettua investimenti rilevanti per la costituzione della banca di dati), assegnandogli il diritto ha diritto di vietare le operazioni di estrazione o reimpiego di parti o della totalità della banca dati (c.d. *diritto sui generis*).

applicano ai testi degli atti ufficiali dello stato e delle amministrazioni pubbliche, sia italiane che straniere”.

Particolare attenzione merita, inoltre, la norma dedicata alle opere realizzate dalle Amministrazioni con impegno di denaro pubblico: l’art. 11 Legge n. 633/1941 statuisce che *“alle amministrazioni dello stato, alle province ed ai comuni spetta il diritto d’autore sulle opere create e pubblicate sotto il loro nome e a loro conto e spese”.*

L'applicabilità della normativa in materia di diritto d'autore ai dati generati e gestiti dalle Amministrazioni può rappresentare un limite alla diffusione e al riutilizzo degli stessi sotto un duplice profilo:

- in primo luogo, l’Amministrazione dovrà porre particolare attenzione alla titolarità delle informazioni che intende pubblicare, evitando di ledere i diritti d’autore e/o altri diritti di privativa di soggetti terzi;
- in secondo luogo, nella pubblicazione l’Amministrazione dovrà adottare le licenze necessarie a evitare che l’esistenza del diritto d’autore e dei diritti connessi possa impedire, o comunque limitare, la possibilità utilizzare i dati stessi per scopi diversi da quelli istituzionali.

4. Come avviare il processo di apertura dei dati

Da quanto sin qui esposto, emerge come le norme dettino già i principi da rispettare nel processo di apertura dei dati pubblici.

Sotto il profilo tecnico-organizzativo, è auspicabile che l'organo di vertice dell'Amministrazione adotti un atto – preferibilmente un regolamento – in cui deliberi di pubblicare *on line* i propri dati in formato aperto, con le dovute eccezioni e mediante l'uso di licenze aperte che consentano la massima riutilizzazione.

Si tratta della scelta preferibile in quanto l'adozione di un regolamento fornisce un quadro chiaro del procedimento da seguire all'interno dell'Ente, uniformando gli accorgimenti utilizzati (ad esempio in materia di formazione dei *dataset*, di metadati e di licenze).

Tuttavia, la scelta di pubblicare i dati può essere compiuta anche in altri atti normativi e amministrativi. A titolo di esempio si citano:

- Leggi regionali di indirizzo in materia di riutilizzo dei dati pubblici;
- Dichiarazione politico-istituzionale;
- Inserimento delle azioni di *Open Data* all'interno del Programma triennale per la trasparenza previsto dall'art. 11 D. Lgs. n. 150/2009
- Previsione di obiettivi specifici nell'ambito del Piano Esecutivo di Gestione;
- Inserimento degli *Open Data* fra i parametri di valutazione della *performance* dirigenziale;
- Previsione di finanziamenti ad altri enti vincolati al raggiungimento di obiettivi prefissati di basi di dati resi pubblici, secondo determinate specifiche.

Parte II – Come procedere per aprire i dati della PA

5. Organizzarsi per l'apertura dei dati

Aprire i dati a disposizione dell'Amministrazione comporta una serie di passaggi che è opportuno strutturare in maniera esplicita attraverso un processo organizzato che prenda in considerazione le diverse variabili esistenti.

5.1 L'approccio per processi

L'approccio per processi proprio dei sistemi di gestione della Qualità¹³ che qui si suggerisce, declinato nell'attività di apertura dei dati dell'Amministrazione, sottolinea l'importanza di:

- aver compreso i requisiti di partenza dell'Amministrazione nelle diverse fasi connesse all'apertura dei dati;
- valutare i processi connessi all'apertura dei dati in termini di valore aggiunto per l'Ente e le comunità di riferimento¹⁴ derivante dalla loro apertura;
- conseguire risultati misurabili in termini di efficacia complessiva dell'azione;

¹³ L'approccio per processi è parte integrante delle norme UNI EN ISO 9001:2000 (Vision 2000)

¹⁴ Cittadini, aziende, organizzazioni

- ottenere un miglioramento continuo dei processi basato su misurazioni oggettive.

Un processo può essere definito come

“un insieme di attività correlate o interagenti che trasformano elementi in ingresso (input) in elementi in uscita (output) attraverso una fase di elaborazione”¹⁵

Uno dei maggiori vantaggi dell’approccio per processi, nei confronti di altri approcci, sta nella gestione e nel controllo delle interazioni tra processi e delle interfacce tra le gerarchie funzionali delle organizzazioni; il che lo rende particolarmente adatto al contesto della Pubblica Amministrazione.

Descrivere i processi connessi all’attività di apertura dei dati a disposizione di un’Amministrazione pubblica vuol dire quindi identificare e descrivere la sequenza strutturata di azioni che sottendono tale attività, a partire dall’identificazione del dato (elemento in entrata/input), passando per la sua elaborazione, per arrivare alla pubblicazione e diffusione attraverso i canali più idonei (elemento in uscita/output).

5.2 Il ciclo PDCA

Nella strutturazione dei processi connessi all’apertura dei dati della Pubblica Amministrazione, è utile fare riferimento al ciclo di Deming¹⁶ conosciuto anche

¹⁵ Adattamento da norma UNI EN ISO 9001:2000

come modello PDCA. Il **ciclo di Deming**, o *modello PDCA* (da *Plan*: Pianificare / *Do*: Eseguire / *Check*: Controllare / *Act*: Agire). Tale ciclo è alla base dell'organizzazione per processi dei sistemi di qualità orientati al miglioramento continuo. In sintesi, esso consente di strutturare processi complessi orientando la struttura stessa del processo così definito verso un modello che prevede momenti di analisi, verifica e cambiamento che ne consentono un'ottimizzazione continua.

La sua applicazione alla gestione dei dati da aprire nell'Amministrazione Pubblica permette di evidenziare la natura dinamica dei dati da essa gestiti, in un contesto in cui il dato non è da considerarsi come un elemento statico da conservare in un archivio, ma come un elemento dinamico – appunto – da valorizzare attraverso il suo utilizzo¹⁷. In altri termini, applicare il modello PDCA ai dati della Pubblica Amministrazione consente di considerarne il ruolo nel contesto, identificarne il valore, valutarne l'impatto sulla comunità di riferimento, definirne le modalità di gestione.

Riportare il modello PDCA all'apertura dei dati della Pubblica Amministrazione vuol dire prendere in considerazione il seguente percorso:

- **Plan – Pianificare**
 - *Determinare gli obiettivi*

Gli obiettivi della politica di *Open Data* vanno indicati in modo

¹⁶ Il ciclo di Deming è un modello studiato per il miglioramento continuo della qualità. Serve per promuovere una cultura della qualità che è tesa al miglioramento continuo dei processi e all'utilizzo ottimale delle risorse. È stato ideato da *Edwards Deming* in Giappone negli anni '50.

¹⁷ In tal senso, il valore del dato aumenta in maniera direttamente proporzionale alla frequenza d'uso e alla rapidità delle sue variazioni.

esplicito, così che sia più semplice procedere nelle fasi successive di identificazione dei *dataset* da aprire.

- o *Determinare i metodi e pianificare le azioni*

Definiti gli obiettivi da raggiungere, è necessario identificare le procedure – che siano razionali e facili da seguire – che consentano di perseguirli. Determinare un metodo vuol dire standardizzarlo e renderlo utile e accessibile. L'apertura dei dati di una Amministrazione Pubblica non è un'operazione che si svolge *una tantum*, ma va reiterata nel tempo (i dati mutano nel tempo e vanno aggiornati). È opportuno quindi standardizzare le procedure di apertura identificando sin da questa fase i momenti di confronto con i cittadini e la comunità degli sviluppatori per definire assieme le priorità.

- **Do – Svolgere il lavoro**

- o Attuare le procedure delineate vuol dire comprovarne l'efficacia e identificarne i punti critici, così da migliorarle progressivamente.

- **Check – Controllare gli effetti dell'apertura dei dati**

- o *Monitorare l'uso dei dati*

Per comprendere la reale efficacia dell'apertura dei dati è necessario verificare se la comunità di riferimento fa effettivamente uso dei dati aperti, ed in particolare di quali. In tal modo sarà possibile capire se la tipologia di dati aperti sia utile ai cittadini oppure se sia opportuno identificare altri *dataset*. Il numero di accessi, la quantità di servizi basati sui dati aperti, la quantità di attori che ne fa uso a vario titolo sono indicatori utili per valutare il successo dell'azione.

- **Act – Intraprendere azioni adeguate**

- *Ottimizzare il sistema*

Nessun processo nasce perfetto. Per questo motivo è importante valutare le attività delle fasi precedenti per individuarne le criticità e correggerle.

- *Far evolvere il sistema*

Così come i dati non sono elementi statici, anche la strategia sottesa alla loro apertura non può esserlo. Deve invece evolversi sulla base delle indicazioni acquisite nelle fasi precedenti. Solo in tal modo i dati diverranno un vero e proprio strumento di supporto nella costruzione di nuove forme di dialogo e partecipazione con l'utenza.

5.3 Le fasi del processo

Spostandosi da una visione complessiva e sistemica a una maggiormente operativa e analitica, a valle della decisione di aprire un insieme di dati in possesso dell'Amministrazione per metterli a disposizione del cittadino e della comunità è possibile identificare un percorso strutturato nelle seguenti fasi:

1. Identificazione dei dati
2. Analisi dei *dataset* identificati
3. Pubblicazione dei dataset
4. Diffusione dei dataset pubblicati

Per ogni fase è opportuno esplicitare:

- **Cosa:** azioni da compiere;

- **Quando:** tempi;
- **Come:** modalità operative;
- **Chi:** ruoli e responsabilità.

Sarà quindi utile, per ogni *dataset* individuato, compilare uno schema simile al seguente:

NOME DEL DATASET	Identificazione del dataset	Analisi del dataset	Pubblicazione del dataset	Diffusione del Dataset
Azioni da compiere				
Tempi				
Modalità Operative				
Ruoli e Responsabilità				

Segue una descrizione delle singole fasi, per ognuna delle quali sono indicati ed esplicitati i principali elementi da prendere in considerazione.

6. Identificazione dei dati

La fase di identificazione dei dati deve rispondere principalmente a tre domande:

- Dove sono i dati a disposizione dell'Amministrazione?

- Quali sono i dati a disposizione dell'Amministrazione?
- Quali sono i dati di interesse per la Comunità?

6.1 Dove sono i dati a disposizione dell'Amministrazione?

L'Amministrazione Pubblica, come qualsiasi organizzazione complessa, gestisce quotidianamente una grande quantità di dati che servono per garantire l'operatività e l'erogazione dei servizi al cittadino. Con lo sviluppo dell'*Information & Communication Technology* e la diffusione dei sistemi di *e-Government* tali dati sono ormai disponibili in formato digitale e possono quindi essere condivisi e diffusi. Per farlo è però necessario identificarli nell'ambito della struttura burocratica dell'Ente.

Decidere quali sono i dati di interesse per la Comunità detenuti dalla propria Amministrazione richiede quindi – come primo passaggio – una vera e propria **mappatura** dei dati complessivamente disponibili.

Definire dove sono i dati a disposizione vuol dire:

- In primo luogo, identificare quali sono le **strutture funzionali** dell'Amministrazione che detengono dati potenzialmente utili. Con riferimento alla Pubblica Amministrazione Centrale, ad esempio, essi saranno distribuiti nell'ambito delle diverse Direzioni. Con riferimento alla Pubblica Amministrazione Locale, invece, essi saranno gestiti dai diversi Uffici;
- in secondo luogo, identificate le strutture funzionali, è necessario comprendere – nel loro ambito – quali sono le **fonti dati** di rilievo. In altri

termini, è necessario identificare archivi strutturati, elenchi, basi dati prodotte da *software* in uso presso l'amministrazione.

La prima fase di mappatura, oltre ad essere utile alla definizione di una strategia di apertura dei dati, lo è anche per determinare e comprendere meglio il patrimonio informativo detenuto dalla propria Amministrazione.

6.2 Quali sono i dati a disposizione dell'Amministrazione?

Una volta mappati ed identificati i detentori di dati dell'Amministrazione è necessario procedere ad una ricognizione dei dati o delle tipologie di dato da essi effettivamente detenuti.

Le organizzazioni complesse non sempre sono a completa conoscenza dell'intero *corpus* di dati che producono o semplicemente che gestiscono. In molti casi, infatti, il dato è il risultato di un processo, o il sottoprodotto di un'elaborazione funzionale ad altri scopi.

In generale, è possibile identificare tre tipologie di dati in possesso dell'Amministrazione:

- **Dati necessari alla gestione delle attività**

I dati necessari alla gestione delle attività dell'Amministrazione sono tutti quei dati che servono all'Amministrazione per il corretto funzionamento dei suoi processi. Sono, insomma, i dati in ingresso (*input*) che alimentano i processi gestiti dall'Ente.

Ad esempio: dati anagrafici, provenienti da altri Enti, indispensabili alla gestione dei procedimenti amministrativi;

- **Dati prodotti come risultato dell'attività**

I dati prodotti come risultato dell'attività dell'Amministrazione sono tutti quei dati che l'Amministrazione produce nell'esercizio delle sue funzioni.

Sono, insomma, i dati in uscita (*output*) ottenuti come risultato finale dei processi e dei procedimenti gestiti dall'Ente.

Ad esempio: dati relativi all'inquinamento ambientale prodotti dalle centraline di controllo dislocate sul territorio; dati relativi all'incidenza della criminalità sul territorio prodotti dalle prefetture; dati inerenti i risultati scolastici prodotti dalle scuole e dai provveditorati; dati riferiti al mercato immobiliare; dati sul tessuto imprenditoriale prodotti dagli organismi camerali; dati economico finanziari e bilanci.

- **Dati prodotti nella gestione delle attività**

I dati prodotti nella gestione dell'attività sono tutti quei dati che l'Amministrazione ottiene come sottoprodotto (*by-product*) di un processo.

Ad esempio: I dati relativi all'attività dei parlamentari e dei consiglieri regionali, provinciali o comunali vengono registrati per certificare le azioni della macchina amministrativa, ma sono utili ai fini della trasparenza, fornendo indicazioni puntuali sull'attività dei singoli rappresentanti; I dati relativi alle prestazioni erogate dalle ASL sono

registrati per il controllo di spesa e la contabilizzazione delle attività delle singole Aziende Sanitarie. Per effettuare tale controllo le ASL registrano oltre al costo anche la tipologia di prestazione erogata. Questo dato consente di ottenere una indicazione relativa all'incidenza delle malattie nel corso del tempo. Ancora: il sistema di monitoraggio della qualità dell'acqua implementato per ovvi motivi di sicurezza produce come sottoprodotto un archivio contenente la variazione dei valori biochimici dell'acqua che può essere messo a disposizione della Comunità.

Come *output* dei primi due passaggi (identificazione delle strutture funzionali e delle fonti dati ed identificazione dei dati) si otterrà una prima mappa delle informazioni a disposizione dell'Amministrazione, per strutturare la quale si può procedere compilando una tabella simile alla seguente:

Struttura Funzionale	Fonte dati	Tipologia di dato
<i>Es. Azienda Regionale per l'ambiente</i>	<i>Es. Sistema di controllo delle acque</i>	<i>Es. Dati sulla qualità dell'acqua</i>
...
...

6.3 Quali sono i dati di interesse per la Comunità?

Nella definizione di una strategia di *Open Data* è bene identificare priorità relative all'apertura di dati che siano funzionali ai reali interessi della Comunità di riferimento. In tal senso, la stessa definizione delle priorità può diventare un'occasione di confronto con il cittadino e la comunità locale, che può essere interpellata e coinvolta nella definizione dei dati da aprire.

Tuttavia, nel momento in cui si avvia una strategia di apertura dei dati non è necessario identificare immediatamente tutti gli usi che verranno fatti di un dato. Spesso è la comunità degli utenti a identificare nei dati un valore che l'Amministrazione stessa non vi ha visto o non vi ravvede. In linea generale, quindi, quando un dato è disponibile senza che la sua estrazione rappresenti un costo significativo, è opportuno provvedere alla sua apertura anche nel momento in cui non se ne ravveda un'utilità immediata. Spesso, infatti, indicatori di interesse per la Comunità sono desumibili da elaborazioni su dati di base che non erano state previste da chi tale dato ha provveduto ad aprirlo.

Questo è uno degli elementi del valore aggiunto di una strategia di *Open Data*.

In linea generale, una recente analisi¹⁸ ha identificato tra gli interessi dei cittadini le seguenti priorità:

- Bilanci delle pubbliche amministrazioni
- Attività dei parlamentari e dei consiglieri regionali, provinciali e comunali
- Inquinamento ambientale
- Epidemiologie sanitarie
- Trasporti pubblici urbani e interurbani
- Criminalità
- Dati elettorali
- Dispersione scolastica

¹⁸ Analisi effettuata dall'Associazione Italiana per l'*Open Government*

- Mercato immobiliare

7. Analisi dei dati

Una volta realizzata la mappatura dei dati presenti nell'Amministrazione e definite delle priorità in base agli interessi della Comunità di riferimento è necessario passare ad una fase di analisi dei dati identificati, per valutare la fattibilità della loro apertura e le eventuali modalità per attuarla.

La fase di analisi dei dati deve rispondere principalmente a tre domande:

- Chi detiene i dati quali diritti ha sugli stessi?
- Quali sono le qualità dei dati?
- In che formato sono disponibili i dati?

7.1 Chi detiene i dati quali diritti ha sugli stessi?

Le Amministrazioni, nell'ambito dei procedimenti amministrativi ad esse affidati, trattano numerosi dati e informazioni che possono essere custoditi negli archivi, cartacei e digitali, dell'Ente.

A prescindere da altre problematiche (ad esempio relative alla riservatezza dei dati personali), bisogna precisare che non tutti questi dati possono essere legittimamente aperti dalle Amministrazioni.

Dal punto di vista giuridico, la prima operazione da compiere è relativa alla verifica della **titolarità del dato** che si intende pubblicare.

In proposito, l'art. 58, comma 1, Codice dell'Amministrazione Digitale detta un'importantissima disposizione in materia di fruibilità, affermando che *"il trasferimento di un dato da un sistema informativo a un altro non modifica la*

titolarità del dato". Questo significa che in capo all'Amministrazione titolare del dato rimanga la responsabilità sulla gestione del dato stesso e sulla sua esattezza e veridicità.

L'Amministrazione titolare del dato è quella che lo ha creato o comunque lo gestisce per fini istituzionali, mentre altre eventuali Amministrazioni che utilizzino tale dato non diventano titolari del dato medesimo. Ad esempio, la titolarità del dato anagrafico è del Comune, così come la titolarità del codice fiscale è dell'Amministrazione finanziaria. Altre amministrazioni che utilizzano tali dati, anche trasferendoli nei propri sistemi, sono comunque soltanto fruitrici.

Ulteriore disposizione che viene in rilievo è quella dettata dall'art. 11 Legge n. 633/1941 che prevede che *"alle amministrazioni dello stato, alle provincie ed ai comuni spetta il diritto di autore sulle opere create e pubblicate sotto il loro nome ed a loro conto e spese"*.

Di conseguenza, è possibile affermare che l'Amministrazione può ritenersi titolare del dato solo quando lo abbia creato direttamente oppure lo abbia commissionato ad un altro soggetto.

In tutti gli altri casi, qualora l'Ente – oltre ad usarli nell'ambito dei procedimenti amministrativi ad esso affidati – abbia intenzione di pubblicare dati formati da altre Amministrazioni o privati (fuori dall'esecuzione di specifici obblighi di legge), potrà legittimamente farlo solo ottenendo dal soggetto titolare apposita licenza che consenta tale pubblicazione.

7.2 Quali sono le qualità dei dati?

Verificato il fatto che i dati sotto il profilo giuridico siano distribuibili è utile valutarne le caratteristiche in termini qualitativi per definire l'opportunità della loro apertura verso l'utenza. Per farlo è opportuno prendere in considerazione una serie di fattori che attengono la dimensione qualitativa del singolo dato e l'incidenza di tali fattori rispetto all'insieme dei dati che compongono il dataset. Di seguito sono riportati i principali fattori¹⁹:

- **Accuratezza sintattica:** è definita come la vicinanza del valore del dato ad un valore nel dominio di definizione sintattico considerato corretto.

Ad esempio: un basso grado di accuratezza sintattica corrisponde ad errori in un database anagrafico nel quale "Stefano" è riportato come "Stfano".

- **Accuratezza semantica:** è definita come la vicinanza del valore del dato ad un valore nel dominio di definizione semantico considerato corretto.

Ad esempio: un basso grado di accuratezza semantica corrisponde al caso in cui "Stefano", in un database anagrafico, diventi "Stefania".

- **Attualità:** è definita come l'adeguatezza del valore del dato rispetto alle esigenze temporali del contesto d'utilizzo.

- **Completezza:** è definita come l'estensione con cui i valori dei dati coprono l'universo cui sono riferiti.

Ad esempio: un basso grado di completezza corrisponde al caso in cui nel dataset contenente gli indirizzi delle scuole di un territorio sia

¹⁹ Elementi tratti ed adattati da Norma ISO/IEC 25012 – *Data Quality Model* <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35736>

riportata una bassa percentuale delle scuole effettivamente presenti in quel territorio.

- **Consistenza interna:** indica il grado di coerenza dei dati presenti in un dataset riferiti ad una stessa entità.

Ad esempio: un basso grado di consistenza interna corrisponde ad errori in un database anagrafico in cui "Stefano", residente a "Roma", risulti in provincia di "Como".

- **Consistenza esterna:** indica il grado di coerenza tra dati diversi ma correlati presenti tra elementi di un dataset.

Ad esempio, un basso grado di consistenza esterna corrisponde ad errori in un database anagrafico in cui "Stefano" risulta sposato con "Francesca" mentre "Francesca" risulta "Nubile".

È importante evidenziare come la presenza di un alto numero di errori in alcuni elementi tra quelli sopra indicati possa sconsigliare l'apertura dell'insieme di dati (ad esempio la consistenza interna ed esterna) mentre in altri casi (ad esempio la completezza) la scelta di aprire il *dataset* possa essere funzionale a migliorarne la qualità, eventualmente anche attraverso processi di coinvolgimento della Comunità di riferimento (che può essere – ad esempio – chiamata ad arricchire il *dataset* con le informazioni delle quali dispone).

7.3 In che formato sono disponibili i dati?

Verificato il fatto che i dati che si intende aprire siano effettivamente distribuibili sotto il profilo giuridico e di qualità adeguata, è necessario effettuare una analisi del formato nel quale i dati sono disponibili.

Riassumendo e semplificando quanto già illustrato al par. 2.3 in merito ai tipi di dati aperti, ci si potrà quindi trovare di fronte ad una delle seguenti situazioni:

- **Dati grezzi** (una stella): sono dati disponibili in quei formati che – seppure disponibili su supporto informatico – non consentono un’extrapolazione immediata degli stessi.

Ad esempio: Immagini nei diversi formati grafici bitmap (es. .gif, .jpg, .bmp, ecc...), documenti in formato .pdf o Microsoft Word.

- **Dati strutturati** (da due a tre stelle): sono dati disponibili in formati – aperti o proprietari - che ne consentono l’elaborazione e la sistematizzazione in forma strutturata.

Ad esempio: documenti realizzati con fogli di calcolo come Microsoft Excel od OpenOffice Calc e salvati in formati come .xls, .csv, .sxc, ecc...

- **Dati presenti in database:** sono dati presenti all’interno dei database gestiti dall’Ente. Di norma di questi dati è possibile scegliere il formato di esportazione più idoneo (es. xml o altri formati aperti). L’esportazione può produrre file classificabili con quattro stelle, o Linked Open Data (cinque stelle).

È bene sottolineare ancora una volta come la logica dell'*Open Data* spinga comunque l'Amministrazione verso la distribuzione dei dati, anche nel caso in cui questi non fossero disponibili in un formato aperto. In altri termini, se un dato ritenuto utile non è disponibile in un formato strutturato o non si ha la possibilità di convertirlo facilmente in un formato aperto, è comunque meglio distribuirlo anche in forma grezza piuttosto che non distribuirlo affatto. Anche in questo caso, sarà poi la comunità di sviluppatori ad occuparsi della conversione del dato in un formato aperto, e quindi utilizzabile.

8. Pubblicazione del Dataset

8.1 In che formato pubblicare i dati?

Quando possibile e per quanto possibile compatibilmente con il formato di partenza dei dati (si veda par. 5.2), è opportuno pubblicare dati in un formato aperto. Ciò vuol dire – riferendosi alla classificazione per stelle del W3C – considerare i formati che vengono classificati con tre o più stelle.

- In caso di dati grezzi (una stella) è necessario valutare l'opportunità della conversione del dato dal formato disponibile a un formato aperto. Tale operazione è nota con il termine di *data scraping* e consiste nell'estrazione dei dati dal formato grezzo attuata attraverso particolari *software* (come ad esempio sistemi OCR²⁰). Tuttavia c'è da considerare

²⁰ *Optical Character Recognition*

che tale procedimento è raccomandabile solo in caso di dati particolarmente rilevanti o significativi, essendo onerosa in termini di risorse e tempi da dedicare all'operazione. Molto più utile, in tal senso e ove possibile, agire alla fonte, verificando la possibilità di estrarre i dati prodotti direttamente in un formato strutturato, e quindi più facilmente gestibile.

- In caso di dati strutturati ma in formato proprietario (due stelle), è quasi sempre possibile convertirli in un formato aperto, guadagnando così nella classificazione una terza stella. È bene quindi – quando si dispone di dati strutturati in formato proprietario – provvedere, prima di pubblicarli a trasformarli in file strutturati in formato aperto (tre stelle). La maggior parte dei dati aperti dall'Amministrazione Pubblica è classificabile con tre stelle.
- I *dataset* classificati con quattro o cinque stelle richiedono la costruzione di *dataset* basati su *Linked Open Data* (LOD) e quindi competenze tecniche specifiche. Di norma, comunque, i *dataset* in formato LOD sono prodotti da database relazionali o *software* specifico.

8.2 Come descrivere il dataset?

In fase di pubblicazione dei dati raccolti in un *dataset*, è opportuno descrivere il *dataset* corredandolo di informazioni utili a comprenderne il contenuto. Tali informazioni prendono il nome di metadati descrittivi²¹. Con il termine

21 Le Linee guida siti web PA nel Cap. 5– Criteri d'indirizzo e strumenti per il trattamento dei dati, della documentazione pubblica e per la loro reperibilità già raccomandano alle amministrazioni "di promuovere una organizzazione dei contenuti tale che possano essere oggetto di

metadato si intende l'informazione che descrive un insieme di dati. Nel caso specifico dei metadati descrittivi²² essi costituiscono un livello di informazioni descrittive che corredano il *dataset*, ne esplicitano alcune caratteristiche e ne rendono più semplice l'identificazione.

Per agevolare la reperibilità dei *dataset* e la loro interoperabilità è importante utilizzare elementi descrittivi come Titolo, Descrizione, Link, indicazione della licenza, Periodo di validità, Ente gestore, Formato ecc.

Un quadro completo dei metadati che descrivono un dataset è quello utilizzato dal Portale italiano dei dati aperti²³.

8.3 Come e dove pubblicare i dati?

Le modalità di pubblicazione dei dati aperti possono variare notevolmente in funzione della tipologia e del formato dei dati da pubblicare.

In generale, il livello di complessità della pubblicazione di dati aperti è direttamente proporzionale al numero di stelle nella classificazione proposta dal W3C.

- Nel caso di dati non strutturati, di dati strutturati in formato proprietario e di dati strutturati in formato aperto la soluzione più semplice consiste

condivisione, scambio, integrazione all'interno della stessa amministrazione e con altre amministrazioni" e "di utilizzare lo standard di metadati Dublin Core".
<http://www.innovazionepa.gov.it/lazione-del-ministro/linee-guida-siti-web-pa/indice/cap5-trattamento-dei-dati-documentazione-pubblica-reperibilita/classificazione-e-semantic.aspx>

²² Esistono anche altre tipologie di metadati, come i metadati amministrativi e gestionali ed i metadati strutturali

²³ <www.dati.gov.it>

nel caricare direttamente il *dataset* o i *dataset* sul *Web Server* della propria Amministrazione. L'operazione può essere compiuta inserendo direttamente il file contenente il *dataset* nella cartella di destinazione del server (tipicamente <http://www.nomesito.gov.it/dati/>). Oppure si può utilizzare il CMS²⁴ adottato per la gestione del proprio portale. In tal caso si procederà in modalità guidata, inserendo il *dataset* come allegato ad una pagina appositamente creata in una specifica sezione del sito.

- Nel caso di *Linked Open Data* la soluzione per la pubblicazione può presentare un livello di complessità maggiore. I *dataset* in formato LOD sono di norma prodotti attraverso database relazionali o per mezzo di applicazioni *software* che espongono delle API²⁵ appositamente concepite per produrre tali *dataset*. I *dataset* in formato *Linked Open Data* possono quindi essere dei file (ad esempio in formato RDF/XML) da pubblicare in una *directory* del proprio sito, analogamente a quanto descritto al punto precedente. Ma possono anche essere costruiti dinamicamente ad ogni richiesta da un database relazionale o da un'applicazione che esponga una API specifica. In questa situazione non esiste un file da pubblicare in una *directory*, ma il *dataset* viene ricostruito dinamicamente ad ogni richiesta effettuata dall'utente. In tal caso sarà necessario sviluppare un'architettura *software* specifica, da definire sulla base del tipo di

²⁴ *Content Management System*

²⁵ *Application Programming Interface*. Si tratta di un set di istruzioni standard attraverso le quali si può interagire con un software in modo da gestirne funzioni o elaborare dati.

dataset e, soprattutto, dei sistemi *software* a disposizione dell'Amministrazione.

8.4 Quale licenza utilizzare per la pubblicazione?

Alla luce delle disposizioni normative indicate nella **Parte Prima** del presente documento, è necessario per le Amministrazioni specificare quali sono le licenze con cui vengono pubblicati i dati resi disponibili sul Web. Ai sensi dell'art. 1, D. Lgs. n. 36/2006 per licenza si intende *"il contratto, o altro strumento negoziale, redatto ove possibile in forma elettronica, nel quale sono definite le modalità di riutilizzo dei documenti delle pubbliche amministrazioni o degli organismi di diritto pubblico"*.

Nella prassi, mutuando le categorie delle licenze del *software*, si usa distinguere tra due macrotipologie di licenze anche in relazione alla pubblicazione e diffusione di dati ed informazioni: le **licenze di tipo *closed*** (cioè chiuse) e le **licenze di tipo *open*** (ovvero aperte); tali licenze si distinguono in relazione ai diversi diritti concessi a chi fruisce di un'opera tutelata dal diritto d'autore (ai sensi della Legge n. 633/1941).

Le licenze di tipo chiuso rappresentano la soluzione tradizionale che consiste nel riservare tutti i diritti - utilizzando il simbolo © che tradizionalmente viene adoperato per indicare il titolare del *copyright* sull'opera. In tali casi, l'utente potrà limitarsi a fruire del documento ma - senza il consenso di colui che detiene i relativi diritti - non potrà copiare, ripubblicare o modificare i contenuti protetti dalla licenza.

Risulta evidente, dunque, che se il dato pubblico fosse reso disponibile con una licenza di tipo *closed*, non si rispetterebbe i principi che sono alla base dell'*Open Data*: i dati sarebbero soltanto liberamente consultabili, ma non anche manipolabili o riutilizzabili in altri contesti al di fuori di quello della mera consultazione.

A differenza delle licenze di tipo *closed*, le licenze *open* più che stabilire quali sono i limiti di utilizzabilità del dato, tendono a garantire una serie di diritti a chi entra in possesso delle informazioni.

Come innanzi accennato, in base alla *Open Definition*²⁶, un dato è aperto se tutti sono in grado di usarlo, riutilizzarlo e redistribuirlo liberamente, con l'unico vincolo di citare la fonte e redistribuirlo allo stesso modo; di conseguenza, le Amministrazioni devono scegliere ed utilizzare una delle licenze di tipo aperto, optando – preferibilmente – per una di quelle già sviluppate e disponibili²⁷.

Tra di esse si segnala la *Italian Open Data License (IODL)*, la licenza “aperta” creata allo scopo precipuo di dare a tutte le Amministrazioni uno strumento chiaro e certificato, in grado di facilitare la diffusione e il riutilizzo

²⁶ www.opendefinition.org. Si tratta di un insieme di definizioni della Open Knowledge Foundation che hanno lo scopo di definire il concetto di apertura in Open Data, Open Content e Open Services.

²⁷ Si ritiene opportuno ricorrere alle licenze già esistenti in quanto esse sono già conosciute da utenti e sviluppatori; il riutilizzo, presupponendo l'accettazione della licenza da parte del soggetto interessato, è quindi agevolato dall'uso delle licenze standard nel procedimento di apertura dei dati.

dell'informazione del settore pubblico²⁸. La IODL, il cui uso è già raccomandato nel cap. 5 delle Linee guida per i siti web della PA, prevede che l'utente possa liberamente:

- 1) consultare, estrarre, scaricare, copiare, pubblicare, distribuire e trasmettere le informazioni;
- 2) creare un lavoro derivato, per esempio attraverso la combinazione con altre informazioni (cosiddetto *mashup*), includendole in un prodotto o sviluppando un'applicazione informatica che le utilizzi come base dati.

In cambio, all'utente è chiesto di indicare la fonte delle informazioni e il nome del soggetto che fornisce il dato, includendo, se possibile, un *link* alla licenza; inoltre è richiesto all'utente di pubblicare e condividere gli eventuali lavori derivati con la stessa licenza o con altra licenza aperta, ritenuta compatibile.

Tra queste rientrano alcune licenze delle famiglie *Creative Commons*²⁹ e *Open Data Commons*³⁰.

²⁸ Il testo della IODL è disponibile all'indirizzo www.formez.it/iodl. Nata nell'ottobre del 2010 nell'ambito del progetto *MiaPA* (il primo sistema pubblico italiano che apre le pubbliche Amministrazioni al *social check-in*), ma progettata per permettere a tutte le pubbliche amministrazioni italiane di diffondere i propri dati, la *Italian Open Data Licence* è stata sviluppata e promossa da *FormezPA* con lo scopo di promuovere l'apertura e valorizzazione dei dati pubblici nel contesto dello sviluppo della società dell'informazione.

²⁹ Tali licenze nascono dall'esigenza di permettere la libera circolazione del materiale creativo protetto dal diritto d'autore dal momento che sono ispirate al modello "alcuni diritti riservati". Le licenze *Creative Commons* si strutturano idealmente in due parti: una prima parte indica quali sono le libertà che l'autore vuole concedere alla sua opera ed una seconda parte che chiarisce a quali condizioni è possibile utilizzare la stessa.

In particolare, le licenze *Creative Commons* compatibili con la *Italian Open Data License* sono :

9. Diffusione del Dataset

9.1 Come promuovere l'esistenza dei dati pubblicati?

Dopo aver pubblicato i *dataset* aperti è bene attivarsi perché altre Amministrazioni, cittadini e comunità di sviluppatori vengano a conoscenza dell'esistenza di dati aperti riferiti alla propria Amministrazione.

Per permettere un'adeguata diffusione dei dati aperti:

- a livello centrale è stato sviluppato il Portale nazionale dei dati aperti raggiungibile all'indirizzo <http://www.dati.gov.it/>, verso il quale far confluire i riferimenti dei dati aperti da ogni Amministrazione (URL e

– la licenza **Creative Commons, Attribuzione Condividi allo stesso modo** (CC-BY-SA) internazionale in versione 3.0 o successiva disponibile all'indirizzo <http://creativecommons.org/licenses/by-sa/3.0/legalcode>;

– licenza **Creative Commons, Attribuzione Condividi allo stesso modo** (CC-BY-SA), in versione 2.5 o successiva <http://creativecommons.org/licenses/by-sa/2.5/deed.it>.

Tra le altre licenze *Creative Commons* che nella prassi sono utilizzate per l'apertura dei dati pubblici si segnalano:

– la licenza CC BY 3.0 (disponibile in italiano all'indirizzo <http://creativecommons.org/licenses/by/3.0/it/legalcode>) che concede all'utente la possibilità di riprodurre, distribuire, trasmettere e adattare liberamente i dati, anche a scopi commerciali, a condizione che venga citata la fonte. Tale licenza si differenzia dalla IODL in quanto all'utente viene richiesto soltanto di citare la fonte.

– La licenza CC Zero (disponibile solo in lingua inglese all'indirizzo <http://creativecommons.org/publicdomain/zero/1.0/deed.it>) che è uno strumento attraverso il quale il titolare dei diritti sui contenuti rinuncia al proprio diritto d'esclusiva (e ad eventuali diritti connessi, compresi eventuali diritti sui generis). I dati pubblicati con la licenza CC 0 ricadono in pubblico dominio ed è pertanto consentito a tutti scaricarli, modificarli e riutilizzarli senza alcuna restrizione.

³⁰ Le licenze *Open Data Commons* rappresentano un progetto di *Open Knowledge Foundation* volto ad affermare l'uso di licenze "aperte" anche per i *database*.

In particolare, la licenza *Open Data Commons* compatibile con la *Italian Open Data Licence* è la **Open Database License (ODbL)**, in versione 1.0 o successiva il cui testo è disponibile solo in inglese all'indirizzo <http://opendatacommons.org/licenses/odbl/1.0/>.

metadati). Un'apposita sezione su detto Portale consente infatti di segnalare l'avvenuta pubblicazione di dati in formato aperto. Tali dataset, previa verifica, entrano a far parte del catalogo dei dati aperti delle Pubbliche Amministrazioni.

- alcune Regioni dispongono o stanno sviluppando portali regionali dei dati. Tali portali, in alcuni casi, consentono alle Amministrazioni del territorio di pubblicare anche i dati di queste ultime; è importante sottolineare come tale pubblicazione non sia sostitutiva di quella effettuata sul sito dell'Amministrazione titolare. In altri casi i portali regionali si configurano come cataloghi con i riferimenti agli indirizzi (URL) ove le singole Amministrazioni decidono di pubblicare il dato secondo le indicazioni fornite al par. 8.3.

Quando disponibile, quindi, l'Amministrazione può pubblicare i dati – oltre che nel suo sito – anche nel portale regionale secondo modalità e funzionalità previste da ciascun portale.

Tra il portale nazionale dei dati aperti e i portali regionali dei dati aperti si stabilisce un rapporto di reciproca collaborazione di scambio dei dati anche attraverso API che garantiscono l'aggiornamento permanente.

In mancanza di un portale regionale di riferimento, l'Amministrazione può comunicare l'avvenuta pubblicazione al Portale Nazionale favorendone in tal modo l'indicizzazione nel catalogo dei dati aperti delle Pubbliche Amministrazioni.

Glossario

API (Application Programming Interface)

Set di istruzioni standard attraverso le quali si può interagire con un software in modo da gestirne funzioni o elaborare dati.

CSV (Comma Separated Values)

Formato di file per insiemi di dati a base testuale, che rende particolarmente agevole l'importazione da o l'esportazione verso fogli di calcolo e database. Il formato CSV separa i valori attraverso virgole.

Dataset

Insieme di dati.

Data scraping

Operazione che consente la conversione del dato dal formato disponibile ad un formato aperto. Consiste nell'estrazione dei dati dal formato grezzo attuata attraverso particolari software (per esempio OCR - Optical Character Recognition).

Dati aperti (Open data)

Dati che possono essere liberamente utilizzati, riutilizzati e redistribuiti, con la sola limitazione – al massimo – della richiesta di attribuzione dell'autore e della redistribuzione allo stesso modo (ossia senza che vengano effettuate modifiche).

Dati collegati (Linked data)

Dati strutturati predisposti in modo tale da poter essere recuperati, interpretati e riusati in contesti e da soggetti diversi da quelli d'origine (marcatura semantica). I dati vengono rappresentati in maniera standard utilizzando **RDF**

(cfr.). Possono anche essere aperti e in tal caso si chiamano **Linked Open Data** – LOD.

Dati grezzi (Raw data)

Dati raccolti che non sono stati soggetti a nessun processo, aggregazione o manipolazione.

Dati strutturati

Dati la cui struttura è organizzata attraverso una logica comprensibile dall'elaboratore. Un foglio di calcolo, ad esempio, è articolato per righe e per colonne alle quali può essere attribuito un valore semantico preciso.

HTML (Hyper-Text Markup Language)

Linguaggio di marcatura che permette di descrivere la formattazione di un documento Web, definendone la visualizzazione attraverso il protocollo http (cfr.). Il linguaggio HTML permette di generare link a documenti differenti, consentendo l'organizzazione e la costruzione di **ipertesti** (cfr.).

HTTP (Hyper-Text Transfer Protocol)

Protocollo usato per lo scambio e il trasferimento di documenti ipertestuali nel Web.

IODL (Italian Open Data Licence)

Licenza "aperta" creata nel 2010 da FormezPA allo scopo precipuo di dare a tutte le Amministrazioni italiane uno strumento "certificato", in grado di facilitare la diffusione e il riutilizzo dell'informazione del settore pubblico.

Iper testo

Insieme di documenti costituenti un'unica struttura coerente e legati tra loro da parole chiave o link. La fruizione di un ipertesto può non seguire un percorso lineare ma essere discontinua e trasversale tra le pagine o i link.

Licenza

Un permesso di riprodurre o riusare materiale protetto da parte di chi detiene i diritti.

Licenze Creative Commons

Licenze che nascono dall'esigenza di permettere la libera circolazione del materiale creativo protetto dal diritto d'autore. Le licenze Creative Commons indicano generalmente quali sono le libertà che l'autore vuole concedere alla sua opera e a quali condizioni è possibile utilizzare la stessa.

Licenze Open Data Commons

Licenze che rappresentano un progetto di Open Knowledge Foundation (OKF) volto ad affermare l'uso di licenze "aperte" anche per i database.

ODBC (Open DataBase Connectivity)

Metodo di accesso standard a una base dati, indipendente dalla tipologia di database, dai linguaggi di programmazione e dai sistemi operativi utilizzati.

Open Source

Espressione inglese, rappresentativa di un movimento, che significa "sorgente aperta". Il sorgente in oggetto è quello dei software, ossia il codice al quale, stando a quanto in quest'ambito si teorizza e sostiene, si deve poter accedere liberamente. Ciò comporta una maggiore circolazione del software stesso, a cui si aggiunge la possibilità per chiunque di apportare miglioramenti.

Protocollo

Insieme di regole e convenzioni che permette a più computer di comunicare tra loro e scambiarsi dati e informazioni.

RDF (Resource Description Framework)

Linguaggio che definisce in che modo le informazioni devono essere rappresentate online, dal punto di vista semantico. Per fare questo, RDF associa dei metadati alle informazioni e ai documenti che circolano in Rete; i metadati permettono al computer di comprendere il contesto semantico e il significato delle informazioni, consentono inoltre l'interoperabilità semantica tra più applicazioni che si scambiano informazioni su internet.

Riuso/Riutilizzo

Qualsiasi uso dei dati diverso da quello per il quale sono stati prodotti o raccolti originariamente.

URI (Uniform Resource Identifier)

Stringa di caratteri che identifica univocamente una risorsa (pagina web, documento, immagine, file). L'URI localizza e nomina la risorsa, descrive il meccanismo da utilizzare per accedervi, specifica in quale computer e attraverso quale percorso al suo interno la risorsa può essere trovata.

URL (Uniform Resource Locator)

Indirizzo Web che identifica univocamente una risorsa su internet, ne specifica formalmente la collocazione in Rete e indica il protocollo da utilizzare per accedervi. L'URL è la tipologia più frequente e conosciuta di URI.

W3C (World Wide Web Consortium)

Consorzio internazionale fondato nel 1994 da Tim Berners Lee, con l'obiettivo di potenziare e diffondere il World Wide Web. L'organizzazione definisce e

rilascia i protocolli (cfr.) e i linguaggi per il Web per garantire l'interoperabilità, contribuendo allo sviluppo e al miglioramento del WWW.

Web Semantico

Evoluzione del Web da un contesto in cui le informazioni possono essere soltanto lette dai computer, ad uno in cui i dati sono processabili e comprensibili direttamente dai computer. Nel *Semantic Web* tecnologie semantiche e agenti intelligenti associano dei metadati a dati e informazioni, che permettono ai computer di comprenderne il contesto semantico e di interpretarne il significato. Il Web Semantico facilita l'interscambio e la cooperazione tra computer ed esseri umani.

XML (Extensible Markup Language)

Metalinguaggio che permette di creare e di definire sintatticamente dei linguaggi personalizzati di markup. Rappresenta un'evoluzione del linguaggio HTML, che permette soltanto di visualizzare una pagina web; XML, invece, è stato pensato per descrivere le informazioni di un documento o di un insieme di dati, dal punto di vista sintattico, attraverso dei marcatori all'interno di un testo. Ciascun documento o database viene rappresentato in modo gerarchico da una struttura ad albero, composta da tag o marcatori.