



INGEGNERI PER GLI OPEN DATA

Roma, 15 gennaio 2013

A cura di Gianmarco Guazzo, Mauro Di Giacomo, Clemente Tartaglione

INDICE

Introduzione : Più ingegneri per l'Open Government

1. Open Data: definizioni e tipologie

1.1 Dati

1.2 Open Data e Open Government Data: requisiti, qualità, valore

1.3 Dai dati grezzi (Raw Data) ai Linked Open Data: formati e tipologie di dati

1.4 Le tipologie di informazioni degli Open Data

2. Norme e iniziative governative per gli Open Data

2.1 Nel mondo

2.1.1 Open Government Directive (USA)

2.1.2 Il Freedom Of Information Act (FOIA)

2.3 In Italia

2.2.1 La trasparenza: il Codice dell'Amministrazione Digitale e la Legge 150/2009

2.2.2 Il riutilizzo dei dati pubblici: la Direttiva dell'Unione Europea 2003/98/CE

2.2.3 La Privacy: Deliberazione del Garante 88/2011

2.2.4 Il diritto d'autore: la Legge 633/1941

2.2.5 Le ultime norme in materia di Open Data

2.2.6 Iniziativa per l'adozione di un FOIA in Italia

2.3 Leggi Regionali sugli Open Data

2.4 Le iniziative italiane più recenti in materia di Open Data

2.4.1 Italian Open Data License

2.4.2 Agenda Digitale e Legge "Crescita 2.0"

2.4.3 Le nuove linee guida per l'interoperabilità dei Linked Open Data

3. Amministrazione Pubblica ed Enti privati: stato dell'arte ed esempi rilevanti

3.1 Open Data nel mondo

3.1.1 Regno Unito: il portale "data.gov.uk"

3.1.2 Unione Europea: il lancio del nuovo portale Open Data

3.1.3 America Latina e Caraibi

3.1.4 Open Data per il pubblico e il privato: alcuni esempi illuminanti

3.2 Open Data in Italia: esempi e casi di successo

3.2.1 *Lo stato dell'arte in Italia*

3.2.2 *Gli esempi italiani più rilevanti*

4. Il valore socio-economico degli Open Data: dati e prospettive di scenario

4.1 Esternalità positive

4.2 I Giacimenti informativi

4.3 Benefici "interni" alle amministrazioni

5. Il processo di apertura dei dati pubblici e le competenze tecniche richieste

5.1 Apertura dei dati pubblici: modelli e processi

5.1.1 *Il modello PDCA per la gestione delle fasi del processo*

5.1.2 *Le fasi del processo di apertura dei dati*

5.2 Ruoli e competenze tecniche per gli Open Data nella Pubblica Amministrazione

5.2.1 *Il Data Manager*

5.2.2 *Divulgare gli Open Data: Infografica, Data Journalism, applicazioni per smartphone*

6. Il valore strategico dei dati nello sviluppo di nuove applicazioni e nuovi modelli di business in campo tecnico

6.1 Tipologie di dati: valori strategici e ambiti specifici di applicazione

6.1.1 *Dati territoriali (geodati)*

6.1.2 *Dati ambientali*

6.1.3 *Dati della ricerca e competenze*

6.1.4 *Dati statistici e classificazione*

6.2 Modelli di business emergenti con gli Open Data

Glossario

INTRODUZIONE

Più ingegneri per l'Open Government

I dati detenuti dalle Pubbliche Amministrazioni stanno acquisendo un'importanza sempre più crescente in tutti i paesi: grazie all'uso delle tecnologie info-telematiche, è possibile, infatti, oggi utilizzare i dati pubblici sia per rendere l'Amministrazione più trasparente ed erogare servizi ancor più efficienti sia per riutilizzarli in ambiti differenti da quelli per i quali sono stati raccolti, non solo per attività di studio, ricerca e informazione e controllo ma anche e sempre più per applicazioni commerciali e sociali.

In dati pubblici che in passato dovevano rimanere relegati nell'ambito dei procedimenti amministrativi per i quali erano stati formati oggi tendono così grazie alle tecnologie a divenire sempre più accessibili aperti e quindi conoscibili anche ad altri soggetti.

Questa dinamica conosciuta come Open Data descrive la prassi in base alla quale alcune tipologie di dati pubblici 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).

L'apertura dei dati pubblici sta determinando, così, effetti importanti sulle amministrazioni pubbliche dei paesi avanzati producendo risparmi di tempo e semplificazione di processi ma sta producendo rilevanti impatti sull'economia complessiva: le stime internazionali segnalano vantaggi rilevanti con effetti ampiamente misurabili anche sul sistema economico-imprenditoriale per l'utilizzo di dati cartografici, dai sui trasporti, dati sui beni culturali, demografici, immobiliari, occupazionali e sociali per produrre applicazioni e servizi.

La Commissione Ue ha evidenziato grazie a uno studio commissionato nel 2011 un impatto economico dei dati pubblici (PSI – ovvero Public Service Information) che tra effetti diretti ed indiretti ha un valore nell'ordine di 140 miliardi di Euro annui.

Quasi tutte le informazioni pubbliche hanno, del resto, un valore intrinseco, che va oltre l'assolvimento dei compiti istituzionali dell'ente che le detiene. Anche quando la singola informazione ha un'utilità limitata, spesso la lettura in serie storica, la raccolta organizzata e la combinazione di più informazioni con fonti differenti consente di riattribuire un valore importante a quella informazione e generare un valore aggiunto notevole.

Così, ad esempio le informazioni geografiche sono alla base di cartine e mappe, le quali possono servire a fornire servizi online di ogni tipo; mentre i dati catastali possono servire a operatori immobiliari e istituti di credito o a provider informativi che vogliono fare incontrare la domanda e l'offerta di alloggi e per proporre soluzioni di finanziamento. O, ancora, i dati

meteorologici possono essere utilizzati sia per servizi a valore aggiunto in tempo reale (ad esempio messaggi inviati sul cellulare dell'abbonato con previsioni selettive del tempo), sia per alimentare i siti meteo più tradizionali. Non è tutto perché anche i dati sui trasporti possono servire per alimentare servi di infomobilità ma anche per decidere strategie di marketing mirate

I dati sui bilanci delle imprese, aggregati sulla base di criteri statistici, territoriali o merceologici potrebbero offrire importanti informazioni di mercato per fare analisi della concorrenza e studi di scenario micro e macro economico di assoluta utilità per i policy makers o per qualunque soggetto debba realizzare strategia di investimento.

Discorso analogo per i dati demografici retributivi e reddituali.

I dati e le informazioni pubbliche se resi disponibili e fruibili possono trasformarsi in contenuti "attivatori" di nuove e, anche, imprevedibili filiere di attività e servizi digitali a valore aggiunto per i potenziali utenti costituiti da cittadini, imprese e altre amministrazioni.

Nel caso dei dati meteorologici pubblici, a parte le nuove modalità di fruire delle classiche previsioni con dispositivi mobili si possono sviluppare servizi specializzati, ad esempio per l'agricoltura o il turismo.

I costi per l'apertura dei dati e la loro condivisione (peraltro decisamente molto più contenuti rispetto alle grandi infrastrutture e servizi on line del passato) facendo da volano a servizi a valore aggiunto sono in grado di generare un introito fiscale sicuramente maggiore rispetto anche a tariffe finalizzate a fare cassa. Gli studi Ue su questo fronte hanno evidenziato che la base tariffaria ottimale è attorno al costo marginale e quindi al costo davvero per il servizio di distribuzione su supporto o per il semplice invio dei dati.

L'elenco dei giacimenti informativi a disposizione della PA è davvero ampio ma è l'insieme degli utilizzi, delle applicazioni, dei software e delle modalità di fruizione, di produzione di contenuti derivati che appare davvero sconfinato.

La tabella che segue offre in sintesi una chiara indicazione della ricchezza di set informativi di cui le Pa e lo stato dispongono, evidenziando quelli che si caratterizzano per un pieno riuso (PSI) sul mercato, da quelli definiti come Public Service Content (PSC) che possono essere messi a valore anche per raggiungere obiettivi di inclusione sociale, obiettivi educativi, di istruzione e formazione e per promuovere modalità di cittadinanza attiva, fermo restando che anche questa distinzione appare ogni giorno sempre più labile dal momento che anche nel caso di servizi di tipo commerciale i provider di contenuti e servizi informativi tendono comunque ad ampliare il corredo di data set da mettere a disposizione dell'utenza comprendendo anche quelle informazioni di carattere culturale, educativo scientifico, ma anche legale o politico.

E' del tutto evidente che i metadati relativi a musei e archivi locali, o le immagini delle opere disponibili, o ancora gli archivi delle Tv pubbliche ecc. possono alimentare e arricchire i servizi e

gli applicativi vari finalizzati a promuovere l'offerta turistica-culturale oltre che necessariamente alimentare i percorsi educativi e di conoscenza aperta per tutti i cittadini.

Tavola 1

Anche in Italia ci si sta muovendo verso gli Open Data e come al solito si agisce soprattutto in termini di nuove norme: il nostro paese ha provveduto infatti tempestivamente ad adeguarsi alle indicazioni internazionali ed agli indirizzi europei come attesta anche l'ultimo decreto Crescita 2.0. Il nostro paese ha addirittura anticipato queste tematiche nel 2005 con il Codice delle Amministrazioni Digitali.

Nonostante le buone intenzioni del legislatore i dati delle nostre amministrazioni stentano però ad essere compiutamente e diffusamente aperti, perché per rendere aperti e fruibili i dati serve uno sforzo tecnico organizzativo importante da parte delle istituzioni di governo, sia a livello centrale che locale, impegno che deve essere adeguatamente sostenuto da risorse e soprattutto da personale con competenze tecnico specialistiche adeguate.

La Pa deve, perciò aprirsi a nuove competenze tecniche e soprattutto agli ingegneri per perseguire l'obiettivo dell'Open Governmente per promuovere una nuova economia degli open data: le competenze gestionali e organizzative e tecnico informatiche degli ingegneri possono rappresentare una straordinaria risorsa per strutturare e realizzare i processi connessi all'apertura dei dati di tutta la Pubblica Amministrazione oltre che per promuovere l'innovazione organizzativa ed una più efficace erogazione dei servizi sia tecnico specialistici che di quelli quotidiani destinati a tutti i cittadini.

Occorre sicuramente reindirizzare le poche risorse disponibili verso investimenti finalizzati ad una reale crescita di efficienza e trasparenza attraverso l'Ict e per promuovere le nuove dinamiche connesse alla messa in chiaro di tutti i dati a disposizione delle pubbliche amministrazioni.

Ma prima ancora si tratta di ripensare profondamente le scelte rispetto alle prossime assunzioni nella PA puntando, soprattutto oggi che i vincoli di spesa impongono e imporranno sempre di più il *downsizing* degli organici, su meccanismi di assunzione selettivi e mirati su profili di qualità e su specialismi tecnici.

Il Governo che verrà dovrà misurarsi allora con una esigenza prioritaria che è quella di riuscire a recuperare la capacità di attrarre al suo interno le giovani eccellenze, dotandosi per i prossimi anni di una classe dirigente capace e al servizio delle istituzioni e con competenze tecnico scientifiche e organizzative adeguate .

Lo sviluppo di un sistema pubblico aperto in cui le informazioni prodotte ed i dati siano tutti fruibili secondo i paradigmi dell'Open Data è, dunque, una straordinaria occasione per rilanciare un processo di modernizzazione del sistema delle amministrazioni pubbliche e con esso di tutto il paese, dando senso pieno ad una ipotesi di uscita dalla crisi e crescita del sistema Italia grazie alle tecnologie Ict più evolute e intelligenti.

Ma per fare ciò occorre dar seguito a quegli indirizzi che già il ministro della funzione pubblica Patroni Griffi, nel dicembre 2011, all'indomani del suo insediamento, aveva chiaramente esplicitato in una audizione alla camera: per cambiare la PA “servono ingegneri” e con essi le altre figure tecnico specialistiche, oltre che, come sempre, i bravi giuristi”.

1. OPEN DATA: DEFINIZIONI E TIPOLOGIE

1.1 Dati

Per spiegare meglio cosa sono i “dati aperti”, prima è necessario puntualizzare brevemente cosa intendiamo per “dati”. Partendo dalla definizione comune di dati come “singoli pezzi di informazione di varia natura”, possiamo fare riferimento alla definizione proposta dal **Digital Curation Centre**¹, organizzazione interuniversitaria britannica fondata nel 2004 sotto l’egida della Fondazione JISC (Joint Information Systems Committee)², che si occupa di promuovere l’utilizzo e la ricerca di tecnologie digitali a livello di istruzione superiore. Secondo tale definizione, i dati:

- descrivono fatti, o meglio: dati e fatti sono strettamente correlati;
- sono potenzialmente riproducibili senza ambiguità;
- sono parte di strutture/architetture informative più vaste;
- possono essere processati e archiviati in formati digitali.

Da ciò si può far discendere la seguente connessione logica, su cui torneremo più avanti:

DATI → INFORMAZIONI → CONOSCENZA → CONSAPEVOLEZZA

Il dato, di per sé neutro, diventa informazione quando viene creato, estratto, elaborato e utilizzato per determinati obiettivi. L’apparato di informazioni, spesso costituito da più dati dello stesso tipo o di tipo diverso (l’insieme di dati si definisce “dataset”), diventa conoscenza quando viene interpretato attraverso strumenti, applicazioni, metodi, incroci, indicatori etc. La conoscenza, infine, acquisisce il valore di consapevolezza – nel caso degli Open Data la si può senz’altro definire di tipo “collettivo”, nel senso più ampio di “bene comune” – quando ad essa viene attribuita la funzione di cambiamento e miglioramento della realtà (i fatti) interpretata e disvelata grazie ai dati.

1.2 Open Data e Open Government Data: requisiti, qualità, valore

La definizione oggi più largamente accettata di Open Data è quella fornita dalla **Open Knowledge Foundation – OKF**³, fondazione non profit internazionale fondata nel 2004 che ha

¹ <http://www.dcc.ac.uk>

² <http://www.jisc.ac.uk>

³ <http://okfn.org>

come obiettivo la promozione del sapere libero⁴. Secondo tale definizione, i dati aperti sono “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)”.

In particolare, quando si parla di “dati aperti” della Pubblica Amministrazione, e quindi di “Open Government Data”⁵, si intendono i dati pubblici in formato aperto e accessibili a tutti i cittadini, senza licenze di copyright o altri limiti legali all’utilizzo. Si capisce bene, dunque, come il concetto di Open Data sia strettamente correlato a quello di Open Government descritto nella nostra Introduzione. 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 cosa pubblica e sulle scelte politico-istituzionali compiute dalle Amministrazioni è un aspetto centrale per la trasparenza delle Istituzioni. 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 non solo un vero e proprio processo di collaborazione tra le istituzioni e la comunità dei cittadini sulle scelte di governo, ma anche la rielaborazione in forma nuova e diversa dei dati messi a disposizione. Mediante strategie di apertura dei dati della Pubblica Amministrazione, i cittadini hanno 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.

Tali presupposti richiamano direttamente i concetti-cardine dell’etica “open”⁶, vale a dire la trasparenza e la partecipazione diretta ai processi decisionali da parte di tutti i cittadini, soprattutto attraverso l’utilizzo delle tecnologie ICT.

⁴Per lo sviluppo dei suoi obiettivi, la community di Open Knowledge Foundation implementa e mantiene costantemente aggiornato il catalogo CKAN il quale, grazie a un potente software open-source di catalogazione dei dati, indicizza e rende fruibili i dataset utili reperibili in Internet (cfr. avanti).

⁵ Secondo la specificazione contenuta nella definizione fornita da Open Data Handbook (<http://opendatahandbook.org>), i dati aperti governativi, per essere tali, devono essere raccolti nel corso dell’attività amministrativa, non devono identificare soggetti specifici – cioè non devono contenere informazioni sui singoli individui – e non devono violare la “*commercial sensitivity*”. Il termine “Open Government Data” è stato elaborato per la prima volta da un gruppo di lavoro riunitosi a Sebastopoli (California) nel 2007. L’uso comune del concetto di “Open Government Data” inizia però nel 2009, quando diversi governi anglosassoni (Stati Uniti, Regno Unito, Canada, Nuova Zelanda) annunciano le prime iniziative per l’apertura dei propri dati pubblici.

⁶*Open Source*, *Open Access* e *Open Content* sono anch’essi concetti che rientrano in questa visione etica, che trova sempre più spazio non solo nelle comunità degli esperti, ma anche nei numerosi movimenti e associazioni di cittadini che promuovono il valore della conoscenza aperta e condivisa.

Concetti, questi, che si adattano perfettamente alla definizione dei principi della “conoscenza aperta” divulgati da Open Knowledge Foundation⁷ in relazioni ai contenuti e ai dati:

- **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.

Se si vuole andare più in dettaglio, soprattutto per quanto riguarda gli Open Data e la loro reperibilità sul/dal web, è possibile prendere in considerazione le “Dieci linee-guida per gli Open Data” proposte dall'organizzazione non profit della Georgia “**Transparency International**”⁸, dove si stabilisce che i dati aperti, per essere tali, devono essere:

- **completi:** devono cioè 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, 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;

⁷ “*Full Open Definition*”, Open Knowledge Foundation, cfr. Open Data Handbook.

⁸<http://transparency.ge/en/ten-open-data-guidelines>

- **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 qui descritte devono caratterizzare i dati nel corso del loro intero ciclo di vita.

Qualora venissero rispettati questi principi-guida, siano essi di carattere più generale come quelli proposti dalla OKF oppure più specifici come quelli indicati dalla ex Repubblica Sovietica⁹, i dati aperti potrebbero senz'altro contribuire non solo alla promozione e incoraggiamento della trasparenza nel rapporto tra amministratori pubblici e cittadini, ma anche alla produzione e incremento di valore sociale ed economico, soprattutto se pensiamo all'Era Digitale in cui tutti siamo immersi. Qualunque azione, sia essa di tipo "fisico", come localizzare l'ufficio postale più vicino, o di tipo "immateriale", come costruire un nuovo motore di ricerca su Internet, richiede l'accesso a un numero più o meno consistente di dati, molti dei quali sono appunto prodotti e rilasciati dalla Pubblica Amministrazione. Aprendo i propri dati, infatti, i Governi possono incoraggiare e aiutare efficacemente la creazione di nuovi e innovativi servizi, non solo di tipo pubblico ma anche di tipo commerciale, in grado quindi di produrre e alimentare valore economico e sociale.

Open Data non vuol dire però soltanto rendere pubblici i dati. Una caratteristica fondamentale degli Open Data è infatti quella del **riuso**. I dati resi pubblici dalle PA devono in altre parole poter essere riutilizzati senza restrizioni e per qualunque scopo, compresa quindi l'idea di consentire al mercato di avviare una qualche forma di business sui dati. Questo è un aspetto che viene messo in evidenza soprattutto dalla **Agenda Digitale Europea** e ripreso da tutte le

⁹ La Georgia ha investito molto, negli ultimi anni, nella trasparenza e nella lotta alla corruzione, tanto che – pur non potendo parlare di ritorno economico diretto grazie agli Open Data e alla trasparenza – oggi è diventata il quindicesimo Paese in cui risulta facile "fare business" nel mondo, grazie a una rinnovata fiducia di cittadini e investitori.

Agende Digitali nazionali e locali. In tutte viene dato ampio risalto agli Open Data ritenendoli un interessante elemento in grado di fungere efficacemente da stimolo per l'economia digitale. È stato stimato, per esempio, che il volume di affari teorico che potrebbe essere sviluppato complessivamente attorno ai dati del settore pubblico dell'intera Europa si aggira attorno ai **140 miliardi di Euro all'anno**¹⁰.

Va detto che gli Open Data sono una immensa risorsa ancora in gran parte inutilizzata, anche se non mancano esempi illuminanti in grado di dimostrare l'enorme potenzialità dei dati aperti, per i quali si rimanda al Capitolo 3.

1.3 Dai dati grezzi (*Raw Data*) ai Linked Open Data: formati e tipologie di dati

Per mettere a disposizione del pubblico i dati di una Pubblica Amministrazione o di un Ente Privato, è possibile utilizzare una grande varietà di formati. Non tutti, tuttavia, rispondono ai requisiti necessari a definire tali dati "aperti".

Il formato in cui le informazioni sono pubblicate, ossia la base digitale in cui sono memorizzate le informazioni, può essere infatti "aperto" o "chiuso". Un formato aperto è quello in cui le specifiche per il software sono a disposizione di chiunque, gratuitamente, in modo che chiunque possa utilizzarle nel proprio software senza alcuna limitazione di riuso imposto attraverso diritti di proprietà intellettuale. Quando il formato è "chiuso", può significare o che il formato è proprietario e le caratteristiche tecniche non sono pubblicamente disponibili, o che il formato del file è proprietario e, anche se le specifiche tecniche sono pubbliche, il riutilizzo è limitato.

La ragione fondamentale per cui è importante chiarire il significato di "aperto" e del perché utilizzare proprio questa definizione, può essere identificata in un termine: **interoperabilità**. L'interoperabilità è la capacità di diversi sistemi e organizzazioni di lavorare insieme; in questo caso, è la capacità di combinare una base di dati con altre. L'abilità di rendere ciascun dato un componente e di combinare insieme vari componenti è essenziale per la costruzione di sistemi sofisticati. In assenza di interoperabilità, ciò diventa quasi impossibile – come nel mito della Torre di Babele, in cui l'impossibilità di comunicare (e quindi di interoperare) dà luogo a un fallimento sistemico della costruzione della torre. Nel caso dei dati ci troviamo in una situazione simile: il punto cruciale di un bacino di dati accessibili e utilizzabili in modo condiviso è il fatto che potenzialmente possono essere liberamente "mescolati" con dati provenienti da fonti anch'esse aperte. L'interoperabilità è la chiave per realizzare il principale vantaggio pratico dell'apertura: aumenta in modo esponenziale la possibilità di combinare diverse basi di dati, e quindi sviluppare nuovi e migliori prodotti e servizi¹¹. Fornire una chiara

¹⁰ Cfr. [http://europa.eu/rapid/press-release MEMO-11-891_en.htm](http://europa.eu/rapid/press-release_MEMO-11-891_en.htm)

¹¹ Come vedremo più avanti, è proprio questo il motivo per cui è auspicabile che i dati vengano aperti in modalità *Linked Open Data (LOD)* 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.

definizione di apertura assicura che sia possibile combinare dataset aperti provenienti da fonti diverse, evitando una nostra “Torre di Babele”.

Per di più, il vantaggio dei formati di file aperto è evidente, in quanto permettono agli sviluppatori di produrre software e servizi multipli utilizzando tali formati. Questo riduce al minimo gli ostacoli per riutilizzare le informazioni in essi contenute. Usare per i file formati proprietari, per i quali le specifiche non sono disponibili pubblicamente, può creare invece dipendenza dal software di terzi o dai detentori della licenza del formato. Nel caso peggiore, ciò può significare che l'informazione può essere letta soltanto utilizzando un certo pacchetto software, che potrebbe avere un costo proibitivo o divenire obsoleto nel corso del tempo.

Ovviamente, dal punto di vista degli Open Government Data, è auspicabile che l'informazione venga rilasciata in formati di file aperti “*machine-readable*”. Tuttavia, qualora la scelta fosse tra il pubblicare dati non completamente rispondenti ai requisiti indicati o non pubblicarli affatto, la logica dell'Open Data indirizza la risposta verso la prima soluzione. In questo caso si parla di “**Raw Data Now**”: nella peggiore delle ipotesi (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*¹²).

Risulta utile, a questo punto, effettuare una carrellata dei principali formati, codici e contenitori di file utilizzati per la creazione, l'archivio e il rilascio di dati.

JSON

JSON è un formato di file semplice e molto facile da leggere per tutti i linguaggi di programmazione. La sua semplicità risiede nella maggiore facilità di elaborazione per i computer rispetto ad altri formati, quale l'XML.

XML (eXtensible Markup Language)

XML è un 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, mentre XML 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. XML è un formato/linguaggio flessibile ampiamente utilizzato per lo scambio dei dati, perché fornisce buone opportunità per

¹² Testualmente, “grattare i dati”. “Fare *scraping*” significa essenzialmente estrarre informazioni da un formato pensato per visualizzare i dati.

conservare la struttura dei dati, nonché consente agli sviluppatori di includere parte della documentazione insieme ai dati senza interferire con la loro lettura.

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 i 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. In altre parole, il formato RDF, fortemente raccomandato dal W3C¹³, consente di rappresentare i dati in una forma che rende facilmente integrabili dati provenienti da fonti diverse. I dati RDF possono essere memorizzati in XML e JSON, oltre che in altri formati. L'RDF incoraggia l'uso di URL come identificatori, provvedendo così un sistema utile per collegare direttamente le iniziative Open Data sul web. L'RDF è ancora poco diffuso, ma il suo uso sta diventando comune fra le iniziative di Open Government, compresi i progetti Linked Open Data dei governi britannico e spagnolo.

Fogli di calcolo

Molti soggetti posseggono informazioni archiviate in fogli di calcolo, come ad esempio Microsoft Excel. Questi dati possono essere utilizzati immediatamente assieme alla corretta descrizione di ciò che le varie colonne rappresentano. Tuttavia, in alcuni casi potrebbero esserci "macro"¹⁴ e formule nei fogli di calcolo scomodi da gestire. Si consiglia pertanto di documentare tali calcoli accanto al foglio elettronico, facilitando la loro accessibilità per gli utenti.

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. I file CSV possono risultare molto utili perché compatti e quindi adatti a trasferire grandi insiemi di dati con la stessa struttura. Tuttavia, il formato è così spartano che i dati, senza documentazione, sono spesso inutili, dal momento che diventa quasi impossibile identificare il significato delle diverse colonne. È pertanto fondamentale che sia fornita un'adeguata documentazione (metadati) insieme ai file in formato CSV. È inoltre essenziale che la struttura del file sia rispettata. L'omissione anche di un solo campo può impedire la lettura di

¹³ World Wide Web Consortium, cfr. avanti.

¹⁴ Una "macro" è una procedura ricorrente durante l'esecuzione di un programma. Concettualmente molto simile a una funzione, può essere richiamata da eventi ed essere parametrizzata. La macro può anche essere definita come un insieme di comandi.

tutti i restanti dati del file, senza alcuna possibilità di correzione, in quanto non è possibile interpretare il senso dei campi successivi.

Documento di testo

Documenti in formati classici come Word, ODF, OOXML, o PDF possono essere sufficienti per mostrare alcuni tipi di dati, ad esempio mailing list relativamente stabili o loro equivalenti. Possono essere condivisi a basso costo, essendo i formati in cui spesso i dati sono stati creati. Questi formati non aiutano in alcun modo a mantenere coerente la struttura, e quindi è spesso difficile inserire i dati in modo automatico. È necessario utilizzare, come base dei documenti, dei modelli che possano evidenziare i dati per un loro potenziale riutilizzo, in modo che almeno sia possibile estrarre le informazioni dai documenti. Generalmente si raccomanda di non proporre i dati in formato word processing, se i dati sono disponibili in formato diverso.

Testo semplice

I documenti di testo (.txt) sono molto facili da leggere per i computer. Tuttavia, generalmente non includono metadati strutturali all'interno del documento, con la conseguenza che gli sviluppatori devono creare un analizzatore in grado di interpretare ogni documento così come appare. Alcuni problemi possono nascere quando i file di testo sono scambiati tra sistemi operativi diversi. MS Windows, Mac OS X e altre varianti di Unix hanno un proprio modo di istruire il computer sul raggiungimento del fine-rigo.

Immagine acquisita (scanner)

Probabilmente la forma meno adatta per la maggior parte dei dati, ma sia TIF che JPEG sono in grado di aggiungere dettagli su ciò che è rappresentato nella foto (fino anche a inserire nell'immagine un contenuto con testo integrale del documento). Possono tuttavia essere utili per la visualizzazione in immagini di quei dati che non sono stati originati elettronicamente.

Formati proprietari

Alcuni sistemi specializzati hanno i propri formati di dati in cui possono salvare o esportare dati. A volte può anche essere sufficiente condividere i dati in questi formati, soprattutto se si prevede un uso successivo in un sistema simile a quello da cui provengono. Dovrebbe sempre essere indicato dove si possono trovare ulteriori informazioni su tali formati, ad esempio con un link al sito web del fornitore. In generale si consiglia, ove possibile, di condividere dati utilizzando formati non proprietari.

HTML (Hyper-Text MarkupLanguage)

Linguaggio di marcatura che permette di descrivere la formattazione di un documento Web, definendone la visualizzazione attraverso il protocollo http. Il linguaggio HTML permette di generare link a documenti differenti, consentendo l'organizzazione e la costruzione di ipertesti. Oggi una gran quantità di dati è disponibile nel formato HTML su vari siti. Ciò sarebbe senz'altro sufficiente se i dati fossero molto stabili e riguardassero un ambito limitato. In qualche caso sarebbe preferibile avere i dati in un formato più semplice da scaricare e manipolare, ma visto che è semplice ed economico fare riferimento a una pagina di un sito web, ciò potrebbe essere un buon punto di partenza per mostrare i dati. Nei documenti HTML sarebbe più opportuno utilizzare tabelle per contenere i dati, nonché visualizzare e identificare i vari campi di dati in modo che siano facili da trovare e manipolare.

Servizi web

Per i dati soggetti a frequenti modifiche, e dove ogni download è limitato nelle dimensioni, è utile esporre i dati attraverso *web services*. Ci sono diversi modi per creare un web service, ma alcuni dei più utilizzati sono SOAP e REST. Generalmente, SOAP su REST, servizi REST, ma sono molto facili da sviluppare e utilizzare, e si tratta pertanto di uno standard ampiamente utilizzato.

Database

Al pari dei *web services*, i database consentono dinamicamente un accesso diretto ai dati. I database hanno il vantaggio di poter consentire agli utenti di mettere insieme ed estrarre soltanto ciò a cui sono interessati. Vi sono tuttavia alcuni problemi di sicurezza nell'estrazione remota da un database. Peraltro, l'accesso al database risulta utile solo se la struttura del database e l'importanza delle singole tabelle e campi sono ben documentati.

Il rating del World Wide Web Consortium: le 5 stelle di Tim Berners-Lee

Per distinguere i diversi formati utilizzabili nella codifica dei set di dati, l'inventore del Web nonché Presidente del W3C, Tim Berners-Lee, ha proposto 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)¹⁵:

¹⁵ <http://5stardata.info>

★ 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 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 essere costretti ad utilizzare software proprietario. Quello caratterizzato dalle tre stelle è il formato più semplice di dati aperti.

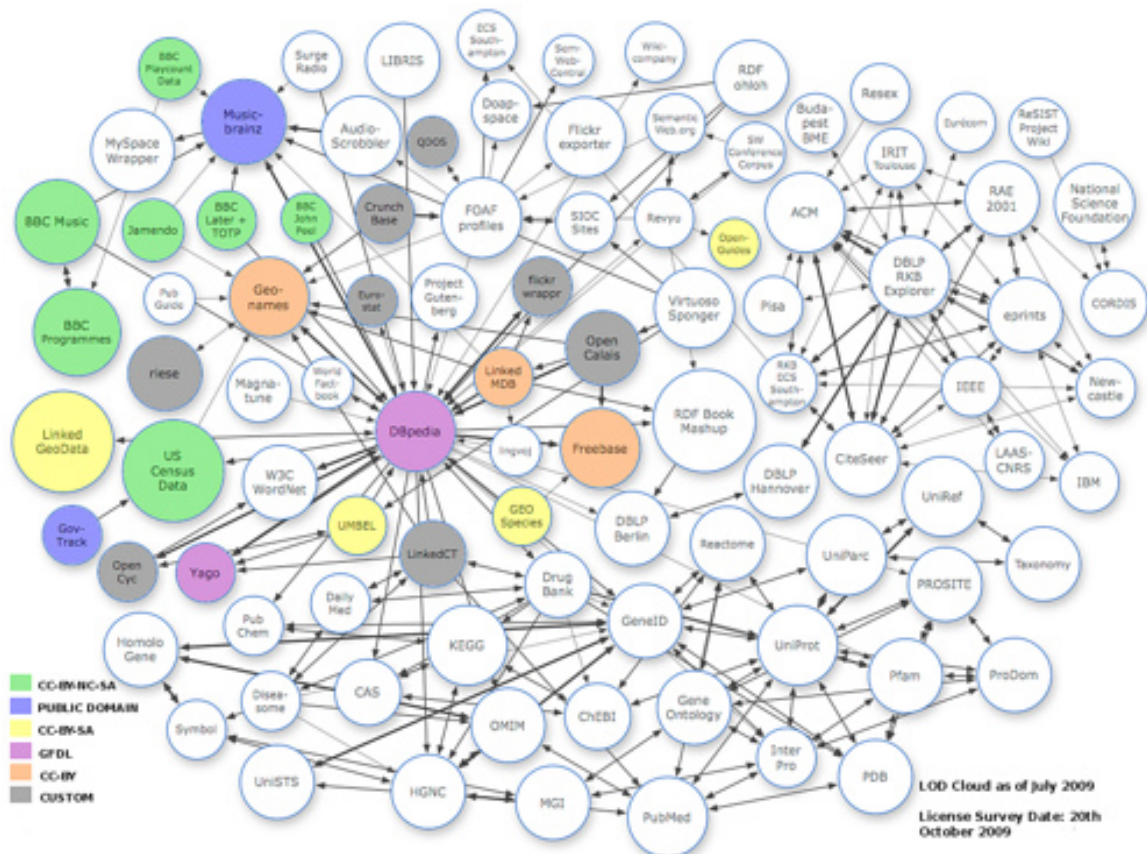
★★★★ Quattro Stelle. Questo livello indica dati strutturati e codificati in un formato non proprietario che sono dotati di un URI¹⁶ che li rende indirizzabili sulla rete e quindi utilizzabili direttamente online, attraverso l'inclusione in una struttura basata sul modello RDF (Resource Description Framework). 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 una 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 – finché dal browser – è possibile collegarsi alla 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, presentano anche, nella struttura del dataset,

¹⁶ *Uniform Resource Identifier* (URI, acronimo più generico rispetto ad "URL"): stringa che identifica univocamente una risorsa generica che può essere un indirizzo Web, un documento, un'immagine, un file, un servizio, un indirizzo di posta elettronica, ecc. L'URL è una URI, o più comunemente chiamato indirizzo web.

collegamenti ad altri dataset. In altri termini, grazie al ricorso al modello di descrizione dei dati RDF, è possibile collegare dinamicamente tra loro più dataset, incrociando così informazioni provenienti da fonti diverse, eventualmente gestite da diverse Amministrazioni o Enti Privati. 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, etc.). 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.

Fig. 1 – Esempio di Linked Open Data Cloud



SCHEDA

Linked Open Data: perché solo Open Data non basta, neppure in Italia

di Titti Cimmino¹⁷

Il Web dei documenti diventa il Web dei dati, questi descrivono "cose" che hanno "proprietà" a cui corrispondono determinati "valori" (da cui il nome "Web Semantico"). Immaginando una tabella: le righe sono le "cose", ogni colonna rappresenta le "proprietà", e l'intersezione rappresenta la proprietà della cosa. In sintesi tendiamo a pensare a dati in questo modo: "cosa", "proprietà", "valore". Ogni "cosa" può avere più proprietà e più cose possono essere in relazione. Dal punto di vista grafico, immaginando un grafo, i nodi sono le cose e gli archi le relazioni tra le cose. Precisa questione è quella della identificazione delle cose globalmente e univocamente dal punto di vista di un database. La chiave di volta dei Linked Data sono gli URI, che appunto consentono la identificazione di cui sopra. Gli URI identificano le cose che vengono descritte, piuttosto che azioni su quelle cose, e se due persone creano dati usando lo stesso URI, allora essi stanno descrivendo la stessa cosa rendendo facile il merging di dati provenienti da data sources distinte, con la possibilità di riconoscere la distinzione tra le risorse e le rappresentazioni di tali risorse: lo stesso URI potrebbe restituire una diversa rappresentazione della risorsa, come ad esempio HTML o XML o JSON. Quindi, se abbiamo intenzione di pubblicare i dati sul Web, abbiamo bisogno di uno standard per esprimerli in modo che un client ricevente i dati possa capire che cosa è una cosa, che cosa è una proprietà, che cosa è un valore e, dal momento che questo è il Web, anche cos'è un link. Questa è la norma fondamentale di cui abbiamo bisogno e questo è ciò che dà RDF: i dati espressi in formato RDF possono usare URI provenienti da differenti siti Web. Se due insiemi di dati utilizzano lo stesso URI poi è molto facile lavorare quando si parla della stessa cosa, ad esempio, permettendo di riunire le informazioni pubblicate da una scuola con le informazioni rilevate da indagini statistiche altrove pubblicate secondo lo standard, naturalmente. E la cosa grandiosa del modello RDF (che fa uso di URI per identificare le proprietà) è che quelle serie di dati possono essere combinate automaticamente, perché lo standard consente di sapere dove cercare le informazioni necessarie.

Usare URI HTTP facilita il recupero di un documento dal Web. Ciò consente di programmare, on demand, l'accesso alle informazioni. Gli sviluppatori non devono scaricare enormi database mentre sono interessati ad una piccola parte di quei dati. Come possiamo creare facilmente dati strutturati e riutilizzabili da formati Excel o (peggio) dai file PDF? Come affrontare i cambiamenti nel tempo, e registrare la provenienza delle informazioni che mettiamo a disposizione? Come possiamo rappresentare le informazioni statistiche? O informazioni sulla

¹⁷ www.titticimmino.com

localizzazione? Queste sono cose che si imparano mettendosi all'opera. È complicato cominciare ad adottare i Linked Data, sia per ragioni sociali, culturali che per motivi tecnologici. Non succederà nulla dall'ora alla mattina, ma a poco a poco ci saranno gli effetti di rete: URI più condivisi, più vocabolari condivisi, il che rende più facile da adottare i Linked Data patterns offrendo più vantaggi per tutti¹⁸.

1.4 Le tipologie di informazioni degli Open Data

Gli Open Data fanno di frequente riferimento a informazioni rappresentate in forma di database e riferite alla tematiche più disparate: cartografia, genetica, composti chimici, formule matematiche e scientifiche, dati medici e pratica, bioscienze, dati anagrafici, dati governativi, statistiche, economia ecc.

Più in generale, e a titolo esemplificativo, si possono distinguere alcune tipologie di dataset aperti secondo i seguenti macro-aggregati tematici:

Fig. 2 –Tipologie di Open Data



¹⁸ Il Web dei dati ha molti URIs equivalenti. Uno dei servizi on line più potenti per trovare le co-riferenze tra diversi dataset, nella logica dei Linked Open Data, è "SameAs", all'indirizzo web www.sameas.org.

Fonte: Open Knowledge Foundation

- **Geodati:** dati utilizzati per realizzare mappe, per esempio la localizzazione di strade e palazzi, la topografia, la visualizzazione dei confini, la georeferenziazione di esercizi commerciali etc.;
- **Cultura:** dati riferiti a opere e prodotti culturali (per esempio: titoli, autori etc.), e generalmente conservati da biblioteche, gallerie, archivi, musei;
- **Scienze:** dati prodotti come parte della ricerca scientifica, dall'astronomia alla zoologia;
- **Economia & Finanza:** dati relativi ai conti pubblici (entrate e spese), informazioni sui mercati finanziari (titoli, azioni, obbligazioni etc.);
- **Statistica:** dati prodotti da uffici e servizi statistici, indicatori sociali, economici, demografici etc.
- **Meteo:** i vari tipi di dati utilizzati per comprendere e prevedere il meteo e il clima;
- **Ambiente e Salute:** informazioni relativi all'ambiente (presenza e livello di fattori inquinanti, qualità delle acque, rifiuti.), ai tassi e cause di mortalità, all'incidenza di malattie in determinate zone etc.
- **Trasporti:** orari, percorsi, statistiche sui tempi di percorrenza etc.

2. NORME E INIZIATIVE GOVERNATIVE PER GLI OPEN DATA

2.1 Nel mondo

La “liberazione” dei dati che riguardano l’ambito pubblico rappresenta un’opportunità importante per le amministrazioni, affinché possano avvicinarsi e sfruttare pienamente le potenzialità dell’innovazione tecnologica e della digitalizzazione. Attraverso l’adozione di un formato “open” dei dati riguardanti il governo centrale e locale, le amministrazioni hanno la possibilità di superare gli schemi rigidi e burocratici di accesso ai dati e di gestione delle risorse informative, sia al loro interno, sia nei confronti della comunità locale. Al contempo, a partire dal libero accesso e dalla condivisione dei dati pubblici, Enti e istituzioni possono creare le condizioni organizzative e tecniche per adottare un nuovo modello di rapporto reciproco e bidirezionale, così come più trasparente ed efficiente, con i cittadini. In un’ottica di rinnovamento verso il modello Open Government, dunque, gli Enti e le istituzioni sono chiamati ad avvicinarsi alla pratica dei dati pubblici “liberi”, rendendo accessibili ai cittadini documenti, risorse e informazioni in loro possesso. Non solo. In un modello di governo e amministrazione “open”, le istituzioni devono adoperarsi affinché la modalità di accesso alle risorse in formato digitale sia efficiente. Ovvero, le PA – oltre a mettere a disposizione dei cittadini contenuti pubblici a titolo non oneroso e senza restrizioni di utilizzo – devono garantire l’accesso ai dati grezzi, per consentire ai privati di trattarli, integrarli e riutilizzarli. In questo senso si possono già registrare i primi significativi effetti legati alla diffusione di questi nuovi approcci in ambito pubblico.

Uno dei più considerevoli riguarda il recente rilascio della “Open Government License” (UK OGL), adottata dal National Archives in Gran Bretagna, licenza che ha dato una spinta significativa al riuso e alla valorizzazione delle banche dati istituzionali. Un’iniziativa simile – che non a caso ha fatto seguito alla UK OGL – è stata portata avanti anche in Italia (“Italian Open Data License” – IODL, cfr. avanti).

Inoltre, grazie allo sviluppo delle tecnologie digitali e al crescente livello di informatizzazione della Pubblica Amministrazione, in questi ultimi anni si è fatta strada la consapevolezza che i dati del settore pubblico (Public Sector Information - PSI) possano svolgere un ruolo importante non solo ai fini della trasparenza amministrativa e della partecipazione pubblica ma anche sul piano economico. Recenti studi europei e nazionali mirano a stimare il valore dell’informazione del settore pubblico in termini di partecipazione, trasparenza, e valore economico che si può ottenere grazie a un più facile accesso e a un più ampio uso di tale informazione. Le leggi approvate dalla Comunità Europea in tal senso si sono focalizzate inizialmente sul diritto del cittadino di accedere ai dati per una questione di trasparenza. Successivamente, l’attenzione si è estesa alla questione della partecipazione e della messa a disposizione dei dati secondo modalità che permettessero di riutilizzarli liberamente, in contesti diversi e innovativi rispetto a quelli nei quali essi sono stati raccolti. Questa tendenza, già presente in documenti strategici quali

l'Agenda Digitale Europea, è stata ulteriormente eripetutamente confermata. Le norme comunitarie, recepite dai vari Stati e dalle Amministrazioni locali, cominciano a essere finalmente messe in atto.

Ma se vogliamo stabilire l'inizio del "Big Bang" degli Open Data in ambito governativo, dobbiamo senz'altro fare riferimento al memorandum di Barack Obama del 2009 in materia di riutilizzo di dati pubblici e riassetto della PA. Tale esempio fornito a livello internazionale ha fatto sì che molte pubbliche amministrazioni abbiano superato alcune delle riserve importanti nei confronti degli Open Data.

2.1.1 Open Government Directive (USA)

Un impulso decisivo al dibattito sull'Open Government e alla diffusione mondiale del modello culturale di Open Government Data è stato fornito dall'Amministrazione USA, che ha tradotto il concetto in un vero e proprio sistema amministrativo. Si tratta del Memorandum "Open Government Directive" dell'8 dicembre 2009, provvedimento che codifica i principi della filosofia "open" all'interno delle istituzioni e delle amministrazioni, prescrive compiti, processi e modelli organizzativi che gli enti pubblici sono chiamati a seguire nel rispetto della Direttiva¹⁹. Sono tre i principi "open" ispiratori della dottrina, che gli enti pubblici americani devono rispettare:

- trasparenza. Le istituzioni sono chiamate a fornire ai cittadini dati e informazioni sulle decisioni prese e sul proprio operato. L'obiettivo del modello "open" USA è quello di creare un sistema di fiducia all'interno della comunità locale nei confronti dell'operato e delle scelte compiute dagli Enti americani;
- partecipazione. L'intelligenza collettiva e la collaborazione tra istituzioni e cittadini nei processi decisionali rappresenta uno dei nodi centrali del modello americano. In questo caso, la finalità da perseguire è il miglioramento della qualità delle scelte politico-amministrative degli enti pubblici, attraverso la proposta di interventi che siano effettivamente legati alle esigenze e necessità dei cittadini;
- collaborazione. Nel modello "open" d'oltreoceano le istituzioni non sono intese come strutture a se stanti, ma soggetti inseriti all'interno di una rete collaborativa e partecipata composta da Enti pubblici, organizzazioni no-profit e comunità di cittadini.

Sull'esempio americano molti Paesi si sono avvicinati al modello Open Government, avviando il processo di cambiamento dello scenario pubblico, verso una maggiore efficienza amministrativa, trasparenza e vicinanza con i cittadini. Di recente, anche l'Organizzazione delle Nazioni Unite è intervenuta a riguardo: attraverso il rapporto 2010 sullo stato dell'e-Government nel mondo ha raccomandato l'adozione di modelli amministrativi "open"²⁰.

¹⁹ Cfr. http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment

²⁰ Cfr. http://www2.unpan.org/egovkb/global_reports/10report.htm

2.1.2 Il Freedom Of Information Act (FOIA)

La legislazione che stabilisce l'accesso ai dati, più nota con l'acronimo inglese di FOIA (*Freedom of Information Act*) è considerata universalmente come uno degli strumenti più importanti delle democrazie mondiali.

Il FOIA è stato emanato negli Stati Uniti il 4 luglio 1966 dal presidente Lyndon B. Johnson, e impone alle amministrazioni pubbliche una serie di regole per permettere a chiunque di sapere come opera il Governo federale, comprendendo l'accesso totale o parziale a documenti classificati. Nel corso degli anni la legge è stata emendata, adeguandola ai tempi: nel 1974 viene approvato il Privacy Act, all'interno di uno schema legislativo che deve conciliare il diritto di cronaca e informazione con quello alla privacy; nel 1996 la legge viene ulteriormente estesa con regole per normare l'accesso ai documenti elettronici (Electronic Freedom of Information Act, E-FOIA); nel 2009, infine, la norma è stata migliorata ed estesa alla luce dei recenti sviluppi nel mondo degli Open Data ("*Open FOIA Act of 2009*").

Nel FOIA, il diritto dei cittadini rappresentati a controllare i governi rappresentanti attraverso l'accesso alle informazioni prodotte dalle pubbliche amministrazioni costituisce la pietra angolare del sistema di "*check and balances*" che consente la creazione di un circuito virtuoso nel rapporto tra cittadini e Stato.

Non è facile comparare la declinazione nazionale di ogni legislazione assimilabile al FOIA e i principi generali che una legge sulla trasparenza e sull'accesso ai dati pubblici deve avere, in quanto manca, ancora oggi, uno strumento di monitoraggio internazionale che consenta un raffronto sull'effettiva implementazione della legge e sul rapporto tra la qualità formale dell'atto legislativo e il suo reale impatto nei relativi Paesi.

Ogni sistema normativo riguardante la libertà di informazione nel settore pubblico deve infatti "dialogare" con il complesso delle legislazioni precedenti e successive in cui essa si colloca, soprattutto per quanto riguarda le leggi sulla privacy e le leggi sulla segretezza dei documenti, che sono in genere le attuazioni legislative che possono porre dei limiti al FOIA. In linea di massima, si può affermare che una legislazione FOIA:

- può prevedere la pubblicità obbligatoria di quasi tutti i documenti prodotti dall'amministrazione pubblica e tutelare anche il diritto all'informazione classificata, nel senso che non dovrebbe prevedere sanzioni per gli impiegati che rivelano dati coperti da confidenzialità. Tale tipo di legislazione trova la sua più esemplare incarnazione del FOIA svedese;
- può consentire l'accesso a tutti i documenti dello stato senza che vi sia necessità di specificare la ragione della richiesta o che vi debba essere un interesse personale per accedere all'informazione, ma solo per una finalità di controllo della PA da parte del

cittadino. È questo il caso del FOIA anglosassone (statunitense, irlandese, sudafricano, australiano, inglese);

- può consentire l'accesso ai dati solo previa presenza di un legittimo interesse di tipo giuridico (vale a dire una causa in corso o una contesa con l'amministrazione) e vietare l'utilizzo dell'accesso ai dati come modalità di controllo della PA. Questa tipologia di FOIA è molto restrittiva e viene considerata unanimemente il modello peggiore. Il dispositivo legislativo che incarna in maniera esemplare questo modello è il FOIA italiano, anche se negli ultimi due anni vi sono alcuni segnali di cambiamento.

2.2 In Italia

Le norme vigenti in Italia, obbligando le amministrazioni ad essere "trasparenti", consentono la divulgazione delle informazioni del settore pubblico: tale pubblicazione può avvenire legittimamente, nel rispetto della normativa in materia di riservatezza dei dati personali e attraverso l'adozione di licenze – già disponibili (cfr. avanti) – che consentono a cittadini e imprese di riutilizzare liberamente i dati pubblici, cogliendo appieno tutti i benefici degli Open Data.

Va comunque qui precisato che il concetto di "trasparenza" è diverso da quello di "apertura". Quest'ultimo include il concetto di trasparenza, ma non è necessariamente vero il contrario. In altri termini, non è sufficiente la trasparenza così come definita nel nostro ordinamento giuridico perché si possa parlare di Open Data²¹. La trasparenza, infatti, costituisce uno dei gangli del diritto pubblico maggiormente soggetto all'evoluzione politica, sociale e tecnologica: soprattutto 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 insufficiente²².

2.2.1 La trasparenza: il Codice dell'Amministrazione Digitale e la Legge 150/2009

Basti soltanto pensare ad alcuni principi, introdotti nel 2005 dal primotesto storico del Codice dell'Amministrazione Digitale (cosiddetto "CAD", D. Lgs. n. 82/2005), a cui devono oggi attenersi tutti gli enti pubblici: *"Le pubbliche amministrazioni, nell'organizzare autonomamente la propria*

²¹La trasparenza, e più in generale la disciplina del rapporto dialogico tra PA e cittadino, è una conquista giuridica di recentissima introduzione, essendo stata prevista solo 20 anni fa dalla Legge n. 241/1990. Da quel momento, è assunta a corollario del principio di buona amministrazione costituzionalmente garantito: *"L'attività amministrativa... è retta da criteri di economicità, di efficacia, di imparzialità, di pubblicità e di trasparenza"*, art. 1, comma 1, Legge n. 241/1990 sul processo amministrativo.

²² Cfr. *"Come si fa Open Data. Istruzioni per l'uso per Enti e Amministrazioni Pubbliche"*, a cura di Ernesto Belisario, Gianluigi Cogo, Stefano Epifani, Claudio Forghieri – Associazione Italiana per l'Open Government, Maggioli Editore, maggio 2011.

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, CAD). Inoltre, le PA sono obbligate per legge ad assicurare "la disponibilità, la gestione, l'accesso, la trasmissione, la conservazione e la fruibilità dell'informazione in modalità digitale" (art. 2, CAD). A tal fine viene esplicitato 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, CAD).

È per tali ragioni che il Legislatore è successivamente intervenuto, nel 2009 (D.Lgs. 150/2009, cosiddetta "Legge Brunetta"), introducendo 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" (articolo 11, comma 1, D.Lgs. n. 150/2009).

Tale nuova posizione giuridica del cittadino è stata poi rafforzata dall'entrata in vigore, il 25 gennaio 2011, del terzo decreto correttivo al CAD, il D.Lgs. n. 235/2010, che, nell'introdurre un nuovo obbligo in capo alla Pubblica Amministrazione, ha sostanzialmente e completato il citato art. 11, D.Lgs. n. 150/2009, facendolo assurgere a vero e proprio diritto soggettivo: il nuovo comma 1-bis dell'art. 52, infatti, afferma che le PA 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.

In conclusione, alla luce di questo processo legislativo, si può dire che siamo di fronte a una 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. In altri termini, alla luce del combinato disposto di cui all'art. 11 del D.Lgs. n. 150/2009, e del nuovo art. 52 comma 1-bis del CAD, ci troviamo in presenza di una nuova declinazione del principio di trasparenza della PA, secondo cui gli esiti dell'attività amministrativa (il provvedimento definitivo, il servizio reso, la banca dati elaborata) non sono più solo l'anello finale di un procedimento amministrativo, ma diventano il nuovo punto di partenza di successivi e autonomi percorsi virtuosi – poco importa se pubblici o privati – del tutto indipendenti dagli originari obiettivi pubblicistici dell'amministrazione precedente.

Si può quindi affermare che non c'è vera trasparenza se l'Amministrazione non rende pubbliche online le informazioni formate e raccolte nell'esercizio della propria attività: è questo il leitmotiv della legislazione degli ultimi anni, confermato, in maniera significativa, dalla Delibera

n. 105/2010 della Commissione per la Valutazione la Trasparenza e l'Integrità delle amministrazioni pubbliche (CIVIT) con la quale sono state adottate le "Linee guida per la predisposizione del programma triennale per la trasparenza"²³.

La legge 150/2009 recepisce inoltre un importante principio contenuto nelle raccomandazioni della Commissione GRECO (*Groupe d'Etats contre la Corruption*), e cioè la trasparenza amministrativa intesa come antidoto alla corruzione²⁴. Tale tema acquista rilievo all'interno della Legge Brunetta sotto diversi profili: la piena autonomia e responsabilità del dirigente, in quanto datore di lavoro pubblico, nella gestione delle risorse umane e nella individuazione dei profili professionali necessari allo svolgimento dei compiti istituzionali e al raggiungimento degli obiettivi, pena la decurtazione di una parte del trattamento economico accessorio, disegna un "argine" agli illeciti e alle distorsioni di una corretta azione amministrativa sufficientemente resistente. Tale disposizione legislativa è poi sfociata nella predisposizione dei "Patti di Integrità", un protocollo stabilito tra il Ministero per la Pubblica Amministrazione e l'Innovazione, ANCI e Transparency International, al fine di assicurare la predisposizione di specifici processi anti-corruzione nella gestione dei Comuni.

2.2.2 Il riutilizzo dei dati pubblici: la Direttiva dell'Unione Europea 2003/98/CE

La nuova declinazione del principio di trasparenza è peraltro estremamente coerente con alcune nuove esigenze della Pubblica Amministrazione emerse di recente un po' ovunque. In molti Paesi, infatti, simili strategie sono state attuate come vera e propria "misura anticrisi", grazie agli effetti positivi che la liberazione dei dati pubblici può arrecare al sistema economico-imprenditoriale. Le Pubbliche Amministrazioni, nell'esercizio della propria attività istituzionale, producono un'enorme mole di dati: si pensi, ad esempio, alle cartografie realizzate nel corso della formazione dei piani urbanistici o le informazioni raccolte dai Centri per l'impiego o dalle Aziende Sanitarie Locali. In passato questi documenti rimanevano 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 in modo che questi possano riutilizzarle. Anche in questo caso, non si tratta di una novità per il diritto: l'Unione Europea da tempo attribuisce al riutilizzo delle informazioni del settore pubblico un ruolo fondamentale per lo sviluppo economico e sociale del territorio. Proprio a tal fine, l'UE ha

²³La delibera n. 105/2010 CIVIT illustra in modo approfondito l'evoluzione del concetto di trasparenza da una concezione definita "statica", legata cioè alla pubblicità di determinate categorie di dati attinenti alle PA, a quella "dinamica" che è correlata alla valutazione della performance. Si tratta di uno dei primi atti in cui si fa espresso riferimento all'Open Government, che viene indicato 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 corrono il rischio di diventare un vero e proprio limite all'azione di trasparenza così faticosamente intrapresa.

²⁴Nel suo rapporto 2008, la Commissione GRECO indicava al governo italiano la necessità di aumentare il livello di trasparenza amministrativa al fine di rendere chiari i processi decisionali e gestionali e quindi l'affidabilità dei decisionmakers e dei responsabili amministrativi in genere.

adottato la Direttiva 2003/98/CE del 17 novembre 2003, recepita dall'Italia con il Decreto Legislativo 24 gennaio 2006 n. 36 "Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico" (*EU PSI-Public Sector Information Directive*), ulteriormente rafforzato poi dalla Legge n. 96 del 4 giugno 2010. La Direttiva UE attribuisce agli Stati Membri – o all'Ente pubblico interessato – la decisione di autorizzare il riutilizzo dei documenti che vengono raccolti, prodotti, riprodotti e diffusi nell'ambito del perseguimento dei propri compiti istituzionali. 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, con il cui termine si 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).

Anche il legislatore italiano ha preso coscienza della circostanza per cui la gran mole di dati pubblici rappresenta un enorme patrimonio comune di conoscenza che è auspicabile venga messo a disposizione degli utenti e, più in generale, di tutti i cittadini e le imprese. In proposito, il già citato Codice dell'Amministrazione Digitale ha introdotto l'importante principio di "disponibilità dei dati pubblici" (enunciato all'art. 2, comma 1, e declinato dal sopra citato 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). Da quanto sin qui esposto, emerge quindi che la liberazione dei dati pubblici è assolutamente legittima ai sensi della normativa vigente.

SCHEDA

Information Asset Register – IAR

La Direttiva 2003/98/CE sul riutilizzo di documenti nel settore pubblico (*EU PSI Directive*) riconosce inoltre l'importanza degli Information Asset Register (IAR) per i potenziali utilizzatori delle informazioni pubbliche. Gli IAR sono registri appositamente istituiti per immagazzinare e organizzare metadati sull'enorme quantità di informazioni detenute da amministrazioni ed enti pubblici. Uno IAR completo include set di dati, vecchi insiemi di file, file elettronici recenti, raccolte di statistiche, ricerche e così via. Secondo la Direttiva UE, gli Stati Membri dovrebbero mettere a disposizione dei potenziali utilizzatori di documenti pubblici strumenti e metodi in grado di facilitarne la ricerca, come per esempio liste, elenchi on line e portali web. Gli IAR possono includere informazioni detenute da enti pubblici, ma che non sono ancora state pubblicate, consentendo di conseguenza al pubblico di individuare informazioni esistenti e che possono essere richieste.

È importante che i metadati negli IAR siano completi, in modo che i motori di ricerca siano in grado di funzionare efficacemente. Nello spirito degli Open Government Data, gli enti pubblici dovrebbero rendere disponibili al pubblico i propri IAR offrendo dati grezzi (*raw data*) e con una licenza aperta, in modo che i cittadini-utenti possano utilizzare i dati, ad esempio per creare motori di ricerca, interfacce utente etc.

2.2.3 La Privacy: Deliberazione del Garante 88/2011

Disponibilità dei dati pubblici non significa automatica condivisione di tutte le informazioni o accesso indiscriminato alle stesse²⁵. 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 e importanti principi che possono guidare la PA nella definizione delle soluzioni e delle modalità con cui procedere alla pubblicazione sono contenuti nella **Deliberazione dell'Autorità Garante per la protezione dei dati personali n. 88/2011**, con la quale sono state adottate le “Linee guida in materia di trattamento di dati personali contenuti anche in atti ed documenti amministrativi, effettuato da soggetti pubblici per finalità di pubblicazione e diffusione sul Web”. La Deliberazione 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: il 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 (es. cartografie, informazioni relative all'inquinamento etc.) non possono essere classificati come personali, in quanto cioè non riconducibili a 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 riferiscono (siano essi individui, imprese, associazioni e Enti) possano essere identificati. Tale impostazione è confermata da quanto

²⁵ Per le informazioni riportate in questo paragrafo e nel successivo, cfr. “Linee Guida per i siti web della PA – Vademecum – Open Data: come rendere aperti i dati delle pubbliche amministrazioni”, versione Beta 2011, Foromez, ottobre 2011, Gangemi Editore.

affermato dal Garante Privacy nella sua Delibera, nel quale è confermato che il perseguimento della finalità di trasparenza dell'attività delle Pubbliche Amministrazioni può avvenire anche senza l'utilizzo di dati personali²⁶.

Di conseguenza, è opportuno che le Amministrazioni, nell'attività di apertura dei dati che detengono, valutino quali accorgimenti adottare 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 degli Open Data.

2.2.4 Il diritto d'autore: la Legge 633/1941

Un ulteriore profilo normativo da affrontare nell'ambito del processo di liberazione delle informazioni della Pubblica Amministrazione è quello relativo alla legislazione in materia di diritto d'autore, in particolare la **Legge n. 633/1941**. Nel nostro ordinamento, il diritto d'autore tutela le opere dell'ingegno dotate di creatività, ed 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), e 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).

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 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 stabilisce che *“alle amministrazioni dello Stato, alle Province e 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 all'utilizzo 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 proprietà di soggetti terzi; in secondo luogo, nella pubblicazione l'Amministrazione dovrà adottare le licenze necessarie a evitare che

²⁶Secondo 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)”*.

l'esistenza del diritto d'autore e dei diritti connessi possa impedire, o comunque limitare, la possibilità di utilizzare i dati stessi per scopi diversi da quelli istituzionali.

SCHEDA

I diritti sulle banche dati

Quando si parla di banche dati, la prima cosa che dobbiamo distinguere è la differenza tra struttura e contenuto di un database. Gli elementi strutturali includono elementi come i nomi di campo e un modello per i dati, l'organizzazione di questi campi e la loro interrelazione. In molte giurisdizioni, gli elementi strutturali di una banca dati potrebbero essere protetti dal diritto d'autore (ciò dipende dal livello di 'creatività' utilizzato nella creazione della struttura). Quando parliamo di "dati" è necessario fare attenzione, perché la parola non è particolarmente precisa: "dati" può significare pochi oggetti o anche un solo elemento (per esempio un singolo record bibliografico) oppure può significare una grande collezione (ad esempio, tutto il materiale nel database). Per evitare confusione ci riserviamo il termine "contenuto" per indicare i singoli elementi e dati per indicare la raccolta.

A differenza di un testo, una musica o un film, la situazione giuridica per i dati varia notevolmente da Paese a Paese. La distinzione fra il "contenuto" di una banca dati e la raccolta è fondamentale soprattutto per i database che raccolgono fatti, dal momento che non è giuridicamente valido concedere diritto di monopolio sui singoli fatti (il "contenuto"), anche se possono essere concessi diritti in merito alla raccolta. Come chiarimento, consideriamo il semplice esempio di una banca dati che elenca il punto di fusione di varie sostanze. Mentre il database nel suo complesso potrebbe essere protetto dalla legge in modo da non permetterne l'accesso, il reimpiego o la redistribuzione senza il permesso, ciò non impedisce di affermare il fatto che la sostanza X fonda alla temperatura Y.

2.2.5 Le ultime norme in materia di Open Data

Tra i mesi di giugno e ottobre 2012 in Italia sono state approvate due norme (entrambe relative alla strategia "Crescitalia" del ministro Passera) che riguardano specificamente gli Open Data:

- Decreto legge n. 83 (cosiddetto "Decreto Sviluppo"), convertito nella **Legge n. 134 del 7 agosto 2012**, che all'art. 18 ("Amministrazione aperta") sancisce per la prima volta in Italia l'obbligo di pubblicare online, entro il 1° gennaio 2013, i dati relativi a "concessione delle sovvenzioni, contributi, sussidi ed ausili finanziari alle imprese e l'attribuzione dei corrispettivi e dei compensi a persone, professionisti, imprese ed enti privati e comunque di vantaggi economici di qualunque genere [...]". Tali dati "devono essere resi di facile consultazione, accessibili ai motori di ricerca ed in formato tabellare

aperto che ne consente l'esportazione, il trattamento e il riuso ai sensi dell'articolo 24 del decreto legislativo 30 giugno 2003, n. 196".

- Decreto Crescita 179/2012 (cosiddetto "Decreto Crescita 2.0"), convertito il 13 dicembre 2012 nella **Legge 221/2012**, che indica le modifiche apportate al CAD relative al dato pubblico, definendo nello specifico cosa si intende per formato di dati di tipo aperto e identificando le caratteristiche dei dati di tipo aperto (ulteriori approfondimenti al paragrafo 2.4.2).

La tabella di seguito riassume sinteticamente i principali provvedimenti normativi descritti in questo Capitolo.

Tab. 1 – Quadro sinottico delle principali norme italiane sull'apertura dei dati pubblici

Provvedimento	Descrizione	Parole chiave
Legge 633/1941	Legge a protezione del diritto d'autore e di altri diritti connessi al suo esercizio	Copyright, Diritto d'autore, Titolarità dei contenuti e dei dati pubblici, Licenze per il riutilizzo
Legge 241/1990 (il legislatore ha già delegato il Governo ad adottare un nuovo Codice della PA che unifica tutte le norme, comprese quelle relative alla PA Digitale)	Nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti amministrativi	Trasparenza
DLgs 196/2003	Codice in materia di protezione dei dati personali	Privacy, Sicurezza
Legge 4/2004 (Legge Stanca)	Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici	Accessibilità
DLgs 82/2005	Codice dell'Amministrazione Digitale (CAD)	Diritto dei cittadini all'uso delle tecnologie, e-democracy, e-participation, Titolarità dei dati pubblici, Formati aperti, Disponibilità del dato pubblico
DLgs 36/2006	Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nella PA	Riutilizzo delle informazioni del settore pubblico
DLgs 163/2006	Codice dei contratti pubblici in materia di lavori, servizi e forniture	Contratti, Titolarità dei contenuti acquisiti dalle Amministrazioni

DLgs 150/2009 (Legge Brunetta)	Ottimizzazione della produttività del lavoro pubblico e di efficienza e trasparenza delle pubbliche amministrazioni	Trasparenza, valutazione e merito, Programma triennale per la trasparenza
DLgs 32/2010	Attuazione della direttiva 2007/2/CE, che istituisce una infrastruttura per l'informazione territoriale nella Comunità Europea (INSPIRE)	Pubblicità, informazione territoriale, riutilizzo
Legge 134/2012 (ex Decreto Sviluppo)	Misure urgenti per la crescita del Paese.	Pubblicazione e riuso dati aperti su finanziamenti PA
Legge 221/2012 (Crescita 2.0 – Legge Passera)	Ulteriori misure urgenti per la crescita del Paese	Agenda Digitale, definizione Open Data

2.2.6 Iniziativa per l'adozione di un FOIA in Italia

Tra aprile e maggio del 2012, esponenti di associazioni, giornalisti, politici e professori universitari che in questi ultimi anni si sono interessati al tema della libertà di informazione e trasparenza del settore pubblico, si sono incontrati e confrontati, giungendo alla determinazione di mettere insieme le loro esperienze per costituire una lobby che informi i cittadini del loro diritto a conoscere (“*the right to know*”) e dei modi per esercitarlo, aprendo un dibattito pubblico sull'esigenza di un maggiore riconoscimento del diritto all'informazione e promuovendo l'introduzione di una legge sulla libertà di informazione simile a quella introdotta negli Stati Uniti nel 1966 (FOIA) e da tempo esistente nei Paesi democratici.

Secondo i promotori del FOIA italiano²⁷, infatti, un confronto tra la legge italiana sulla trasparenza (241/1990) e quelle in vigore negli altri Paesi europei e in USA, mostra il ritardo dell'Italia dal punto di vista sia culturale sia legislativo, per quanto riguarda i diritti del cittadino. La legge italiana, secondo il movimento FOIA italiano, è l'unica in Europa a subordinare la richiesta della documentazione della pubblica amministrazione a un interesse diretto del singolo cittadino, e ad escludere esplicitamente la possibilità di un suo utilizzo come mezzo di controllo generalizzato sulla pubblica amministrazione. Nonostante il principio della “accessibilità totale” sia stato introdotto nella normativa italiana vigente (Legge 15/2009; 150/2009; 183/2010), esso resta appunto soltanto una mera affermazione di principio, non in grado di vincolare la pubblica amministrazione attraverso, ad esempio, un sistema di obbligo-sanzione. In Europa e negli USA, al contrario, il diritto all'accesso è garantito a chiunque, indipendentemente da ogni specifico interesse, e diventa quindi un vero e proprio strumento di controllo dell'attività amministrativa (esplicitamente esclusa dalle modifiche approvate alla legge italiana sulla trasparenza nel 2005) e di partecipazione dei cittadini ai meccanismi decisionali. Il principio del

²⁷ www.foia.it

Freedom of Information obbliga la pubblica amministrazione a rendere pubblici i propri atti e rende possibile a tutti i cittadini di chiedere conto delle scelte e dei risultati del lavoro amministrativo.

L'esperienza degli altri Paesi ha mostrato, inoltre, che una legge efficiente sul diritto di accesso ha effetti positivi sul funzionamento della pubblica amministrazione (favorendo il tasso di fiducia dei cittadini e dunque permettendo alla PA di operare al meglio), sulla semplificazione e ottimizzazione delle decisioni amministrative che attualmente ostacolano gli investimenti delle imprese, sulla lotta alla corruzione.

Il movimento FOIA italiano, in particolare, persegue due obiettivi prioritari:

- sensibilizzare l'opinione pubblica sull'importanza di un rapporto paritario tra cittadino e pubblica amministrazione;
- mettere in primo piano nella agenda parlamentare una revisione della legge del diritto di accesso.

I partecipanti al movimento si sono quindi costituiti in Comitato, hanno formato un gruppo di studio e hanno attivato un sito web (www.foia.it) in cui vengono raccolti e resi disponibili i materiali relativi al tema e viene lanciato un appello per raccogliere eventuali adesioni.

2.3 Leggi Regionali sugli Open Data

Per quanto riguarda lo stato dell'arte dei dispositivi normativi nazionali, mentre a livello di Governo Centrale non si riesce ancora ad andare oltre le disposizioni di principio che incentivano le PA a fare Open Data, le Regioni si stanno muovendo in tal senso, lavorando a leggi regionali che le vincolano a pubblicare i dati di cui sono titolari e a consentirne il riutilizzo.

All'indirizzo web <http://g.co/maps/xhz59> è possibile consultare una mappa ragionata, in continuo aggiornamento, con l'indicazione delle proposte già divenute legge e di quelle attualmente in discussione²⁸.

Al momento di compilazione del presente documento, i dati sono i seguenti: su 21 tra Regioni e Province Autonome, solo tre, il **Piemonte** (Legge Regionale n. 24 del 23 dicembre 2011), il **Lazio** (Legge Regionale 7/2012) e la **Puglia** (Legge 21/2012), hanno già approvato la propria legge regionale, mentre disegni e proposte di legge sono in discussione in Umbria, Lombardia, Sicilia, Campania, Provincia Autonoma di Trento, Basilicata, Marche.

Il Friuli–Venezia Giulia, ha adottato una disposizione di puro principio, nell'ambito della legge sul sistema informativo regionale.

²⁸ L'iniziativa è di Ernesto Belisario, avvocato esperto di Open Data, presidente dell'Associazione Italiana Open Government.

2.4 Le iniziative italiane più recenti in materia di Open Data

2.4.1 Italian Open Data License

Anche quando si parla di dati aperti è necessario descrivere alcune regole minime che indicano le modalità di rilascio degli stessi. Il primo intervento nel nostro Paese sul tema delle licenze dei dati prodotti e gestiti dalle amministrazioni pubbliche è la Italian Open Data License (IODL), una licenza italiana per banche dati pubbliche sviluppata da FormezPA e riveduta da un gruppo di lavoro di esperti (composto tra gli altri da Marco Ciurcina, Ernesto Belisario, Federico Morando e Stefano Costa)²⁹. La IODL ha l'intento di favorire il riuso dei dati pubblici, è modellata sulla Open Government Licence inglese ma al contrario della OGL è una licenza "share-alike", che richiede cioè la condivisione delle opere derivate secondo gli stessi termini.

Più in particolare, la IODL è un contratto di licenza che ha lo scopo di consentire agli utenti di condividere, modificare, usare e riusare liberamente la banca di dati, i dati e le informazioni con essa rilasciati, anche a fini commerciali, garantendo al contempo la stessa libertà per altri. La licenza mira a facilitare il riutilizzo delle informazioni pubbliche nel contesto dello sviluppo della società dell'informazione.

Le banche di dati, i dati e le informazioni sono protetti dalle leggi applicabili in materia di diritto d'autore (incluso il diritto *sui generis* del costituente di banche di dati) e/o dalle altre leggi applicabili. I diritti concessi dalla IODL non implicano alcun trasferimento di diritto di titolarità sulle banche di dati, sui dati e sulle informazioni pubbliche.

Attraverso la IODL, il licenziante concede una licenza per tutto il mondo, gratuita, perpetua, non revocabile e non esclusiva permettendo di riprodurre, distribuire al pubblico, concedere in locazione, presentare e dimostrare in pubblico, comunicare al pubblico, trasmettere e ritrasmettere in qualunque modo, eseguire, recitare, rappresentare, includere in opere collettive e/o composte pubblicare, estrarre e reimpiegare le Informazioni, a condizione di: 1) indicare la fonte delle Informazioni e il nome del licenziante, includendo, se possibile, una copia di questa licenza o un collegamento (link) ad essa; 2) non riutilizzare le Informazioni in un modo che suggerisca che abbiano carattere di ufficialità o che il licenziante approvi l'uso che fai delle Informazioni; 3) prendere ogni misura ragionevole affinché gli usi consentiti non traggano in inganno altri soggetti e le Informazioni medesime non vengano travisate.

La licenza IODL deve sempre ritenersi regolata dalla Legge italiana e deve, di conseguenza, essere interpretata applicando tale normativa. In particolare, la IODL non costituisce un'autorizzazione a violare la legislazione italiana in materia di diritto d'autore (Legge n. 633/1941) e di protezione dei dati personali (D. Lgs. n. 196/2003).

²⁹<http://www.formez.it/iodl>

La IODL è dichiarata compatibile con *Creative Commons Attribution – Share Alike* e con la *Open Database License*, permettendo la distribuzione di opere derivate nei termini di una delle due licenze (che, come è noto, non sono compatibili tra di loro), entrambe molto diffuse.

Ai sensi dell'art. 1 IODL, sono licenze compatibili:

- la licenza Creative Commons, Attribuzione Condividi allo stesso modo (CC-BY-SA), sia internazionale in versione 3.0 o successiva³⁰ che adattata a specifiche giurisdizioni, in versione 2.5 o successiva³¹;
- la licenza Open Data Commons, Open Database License (ODbL), in versione 1.0 o successiva³².

Nel marzo 2012 è stata rilasciata una seconda versione della Italian Open Data License, indicata come IODL 2.0, priva di clausole del tipo “condividi-allo-stesso-modo” (“*share-alike*”) e con la sola richiesta di attribuzione della fonte per il riutilizzo dei dati.

SCHEDA

Le licenze Creative Commons

Oltre alla IODL, le tipologie di licenze più utilizzate nella pubblicazione e condivisione di Open Data sono le licenze Creative Commons, e in particolare:

Creative Commons Zero (CC0). Esprime “la più ampia e libera utilizzazione gratuita, anche per fini commerciali e con finalità di lucro”. Apponendo su un documento la dichiarazione CC0 si rinuncia a tutti i diritti sul documento e sui suoi contenuti, dati inclusi, nella misura massima possibile prevista dalla legge. La Creative Commons Zero deve essere preceduta da una dichiarazione relativa alla provenienza del documento. Il è stato concesso da Regione Piemonte ai sensi della Legge regionale n. 24/2011 e s.m.i.. La Regione Piemonte, per esempio, ha autorizzato il riutilizzo della “Banca dati della rilevazione scolastica” secondo i termini della Dichiarazione Creative Commons - CC0 1.0 Universal, consentendo pertanto la libera e gratuita consultazione, estrazione, riproduzione, modifica e riutilizzo del documento e dei dati in esso contenuti da parte di chiunque vi abbia interesse per qualunque fine.

Creative Commons Attribuzione (CC-BY). Un'alternativa alla licenza CC0 è la licenza Creative Commons “Attribuzione o equivalente”, che permette al soggetto utilizzatore di riprodurre, distribuire, comunicare, esporre, rappresentare, nonché di modificare e usare un insieme di dati anche a fini commerciali con il solo obbligo di attribuire la paternità dell'opera. Anche questa licenza risulta essere, al pari della standard, espressione del principio della “più ampia e libera

³⁰<http://creativecommons.org/licenses/by-sa/3.0>

³¹<http://creativecommons.org/licenses/by-sa/2.5/deed.it>

³²<http://www.opendatacommons.org/licenses/odbl>

utilizzazione gratuita anche per fini commerciali e con finalità di lucro". In caso di uso della CC-BY però, l'unico obbligo imposto al licenziatario è quello di citare l'autore della base dati o del documento oggetto di riutilizzo, nel rispetto delle modalità indicate dall'autore stesso nella licenza.

Share Alike (SA): obbliga i lavori derivati a essere licenziati con la stessa licenza del lavoro originale.

Non Commercial (NC): consente la copia, la distribuzione e l'uso del lavoro (o dati) solo per scopi non commerciali.

No Derivative Works (ND): consente la copia, distribuzione e l'uso del lavoro, impedendola creazione di lavori derivati.

2.4.2 Agenda Digitale e Legge "Crescita 2.0"

L'Agenda Digitale è stata presentata dalla Commissione Europea nel maggio 2010 con lo scopo di sfruttare al meglio il potenziale delle tecnologie dell'informazione e della comunicazione per favorire l'innovazione, la crescita economica e la competitività. L'obiettivo principale dell'Agenda è ottenere vantaggi socio-economici sostenibili grazie a un mercato digitale unico basato su Internet veloce e superveloce e su applicazioni interoperabili. L'Agenda Digitale rappresenta una delle 7 iniziative faro individuate nella più ampia Strategia EU2020, finalizzata a una crescita inclusiva, intelligente e sostenibile dell'Unione. Con una maggiore diffusione e un uso più efficace delle tecnologie digitali, l'Europa potrà stimolare l'occupazione e affrontare le principali sfide a cui è chiamata, offrendo ai suoi cittadini una migliore qualità della vita, per esempio assicurando un migliore servizio sanitario, trasporti più sicuri ed efficienti, un ambiente più pulito, nuove possibilità di comunicazione e un accesso più agevole ai servizi pubblici e ai contenuti culturali. Tuttavia, i benefici che i cittadini potrebbero trarre dall'uso delle tecnologie digitali sono limitati da alcune preoccupazioni inerenti la riservatezza e la sicurezza e dalla mancanza o carenza di accesso a Internet, usabilità, capacità adeguate o accessibilità per tutti. L'Agenda Digitale individua i principali ostacoli che minano gli sforzi compiuti per sfruttare le ITC e indica la strategia unitaria a livello europeo volta al loro superamento individuando le aree d'azione che sono chiamati ad adottare gli Stati membri. Queste aree d'azione costituiscono i "pilastri" dell'Agenda Digitale relativi a: Mercato digitale unico, Internet veloce e superveloce, Interoperabilità e standard, Fiducia e sicurezza informatica, Ricerca e innovazione, Alfabetizzazione informatica, ICT per la società.

L'Italia, come ogni Paese Membro, è chiamata ad analizzare il contesto nazionale per elaborare una propria strategia di recepimento dell'Agenda digitale, individuando le priorità e le modalità di intervento.

Pertanto, il 1° marzo 2012, è stata istituita L'Agenda Digitale Italiana (ADI)³³, con decreto del Ministro dello Sviluppo Economico, di concerto con il Ministro per la Pubblica Amministrazione e la Semplificazione, il Ministro per la Coesione Territoriale, il Ministro dell'Istruzione, dell'Università e della Ricerca e il Ministro dell'Economia e delle Finanze.

Successivamente, il 18 ottobre 2012, è stato pubblicato in Gazzetta Ufficiale il Decreto Legge 179/2012 – "Ulteriori misure urgenti per la crescita del Paese" (cosiddetto "provvedimento Crescita 2.0"), convertito in Legge 221/2012 lo scorso 13 dicembre, in cui sono previste le misure per l'applicazione concreta dell'ADI. I principali interventi sono previsti nei settori: identità digitale, PA digitale/Open data, istruzione digitale, sanità digitale, divario digitale, pagamenti elettronici e giustizia digitale.

L'organo operativo dell'ADI, la Cabina di regia, è strutturata in 6 gruppi di lavoro a cui corrispondono 6 assi strategici: Infrastrutture e sicurezza; eCommerce; eGovernment Open Data; Alfabetizzazione Informatica – Competenze Digitali; Ricerca e Innovazione; Smart Cities and Communities. Per quattro mesi la Cabina di regia ha lavorato alla stesura dell'ADI, raccogliendo le indicazioni e i suggerimenti inviati da migliaia di cittadini e stakeholders. La Cabina di regia ha elaborato una serie di iniziative concrete con l'obiettivo di innovare ognuno dei sei ambiti tematici e di implementare gli indirizzi dell'Agenda Digitale Europea in campo nazionale.

Accanto alle iniziative, la cabina ha elaborato delle proposte normative e di riorganizzazione della macchina amministrativa, tenendo conto dei principi di efficienza e risparmio.

Riguarda agli Open Data, l'ADI ha posto l'attenzione su una *"necessaria azione di valorizzazione del patrimonio informativo pubblico, che renda sistematico e organico a tutti i livelli amministrativi il processo di pubblicazione dei dati per il riutilizzo, uniformi le iniziative normative e i progetti attivi a livello nazionale, incardini il processo di pubblicazione all'interno di ogni amministrazione, gestendo la governance di queste attività in modo organico ed efficace"*.

Tali raccomandazioni sono state recepite nel citato "Provvedimento Crescita 2.0", il cui art. 9 recita testualmente: *"I dati e le informazioni forniti dalla pubblica amministrazione dovranno essere obbligatoriamente pubblicati in formato aperto (cd. open data). In questo modo sarà possibile ampliare fortemente l'accesso a informazioni di pubblica utilità, favorendone il riutilizzo per analisi, servizi, applicazioni e soluzioni, con sensibili ricadute dal punto di vista della crescita economico-sociale. Tali dati avranno una licenza d'uso aperta e saranno dunque utilizzabili – in primis da persone affette da forme di disabilità sensoriali - senza alcun tipo di restrizione"*.

2.4.3 Le nuove linee guida per l'interoperabilità dei Linked Open Data

L'Agenda Digitale Europea prescrive, agli Stati Membri, di allineare i propri framework nazionali di interoperabilità a quello Europeo (EIF) entro il 2013. Come definito dal Codice

³³http://www.agenda-digitale.it/agenda_digitale

dell'Amministrazione Digitale (CAD), il framework italiano di interoperabilità è il Sistema Pubblico di Connettività e Cooperazione (SPC). Nell'ambito del programma ISA, che ha il compito di attuare a livello europeo gli obiettivi di interoperabilità del pilone 2 della Digital Agenda, il National Interoperability Framework Observatory (NIFO) ha valutato i vari framework nazionali di interoperabilità tra i quali SPC. La valutazione è stata effettuata sulla base di 5 ambiti: "principi", "governance", "modello concettuale", "accordi di interoperabilità", e "livelli di interoperabilità". Dall'analisi è emerso come l'Italia sia piuttosto bene allineata sui primi tre ambiti, evidenziando una conformità pari quasi al 100%, mentre i risultati più carenti sugli accordi di interoperabilità e, soprattutto, sui livelli di interoperabilità garantiti.

Tra i livelli di interoperabilità definiti dal modello EIF rientra l'interoperabilità semantica, ovvero la possibilità, offerta alle organizzazioni, di elaborare informazioni da fonti esterne o secondarie senza perdere il reale significato delle informazioni stesse nel processo di elaborazione. Questa definizione coglie il ruolo centrale che tale dimensione assume nella collaborazione e nell'interscambio di informazioni tra istituzioni, base quindi di ogni processo di e-government innovativo.

In tale ambito, nel novembre 2012 il Sistema Pubblico di Connettività e Cooperazione ha pubblicato le linee guida per la produzione di Open Data interoperabili³⁴, cogliendo gli obiettivi della Digital Agenda.

Il lavoro si è articolato partendo da un'attenta disamina dello scenario attuale, sia nazionale che internazionale, nell'ambito della gestione dei dati del settore pubblico. Da tale analisi è emerso come un profondo cambiamento in questo settore sia in atto, grazie allo sviluppo del nuovo paradigma degli Open Data. I dati delle Pubbliche Amministrazioni, tipicamente "nascosti" in applicazioni o basi di dati, e da loro gelosamente custoditi nella maggior parte dei casi, sono sempre più resi accessibili a chiunque con la consapevolezza che, di fatto, essi rappresentano un patrimonio della collettività, e non di singole istituzioni, e un importante strumento per la trasparenza, responsabilità e possibile sviluppo economico. Tuttavia, l'analisi dello scenario ha evidenziato altresì che, per sfruttare pienamente i suddetti benefici, è necessario favorire la facilità d'uso dei dati, così come il loro reperimento e consumo sia da parte degli esseri umani che, soprattutto, da parte dei software attivabili anche in maniera automatica.

Il gruppo di lavoro ha quindi analizzato le tipologie di dati in possesso delle Pubbliche Amministrazioni e lo stato dell'arte relativo ad alcune iniziative volte a raggiungere i suddetti obiettivi in tale ambito, e ha identificato nelle tecnologie standard del Web semantico, e in particolare nel modello dei Linked Open Data, gli strumenti imprescindibili per dare ai dati (aperti o non) un'identità e per renderli collegabili tra loro e soprattutto interoperabili. In altre parole, il gruppo di lavoro, all'unanimità, ha ritenuto che, per abilitare lo sviluppo di una

³⁴http://www.digitpa.gov.it/sites/default/files/allegati_tec/CdC-SPC-GdL6-InteroperabilitaSemOpenData_v2.0_0.pdf

concreta interoperabilità semantica tra Pubbliche Amministrazioni a livellonazionale e transfrontaliero sia necessario adottare il modello **Linked Open Data**.

A tale riguardo sono state definite unaserie di raccomandazioni, allo scopo di focalizzarel'attenzione su azioni concrete che l'esposizione dei concetti non rende così immediate.

A completamento dello studio, il gruppo ha posto particolare attenzione agli aspetti legali derivanti dallelicenze d'uso associate ai dati pubblicati, e ai modelli di business che possono essere abilitati, evidenziando aspetti di interoperabilità anche per le licenze, sostenibilità e governance dell'approccioLinked Open Data.

Lo studio è stato poi contestualizzato nell'ambito del framework nazionale d'interoperabilità SPC, al fine di individuare lo specifico ruolo che, sia le infrastrutture condivise sia i servizi e-government per laPA possono assumere per la concreta acquisizione e attuazione dell'approccio proposto e quindi dell'interoperabilità semantica. Dallo studio ne è risultato che SPC, e in particolare il servizio "CatalogoSchemi e Ontologie", come definito nel DPCM 1° aprile 2008 recante regole tecniche e di sicurezza perSPC, può essere profilato come il servizio Linked Open Data SPC (il Web dei Dati SPC), che consente di produrre Linked Data a partire da dati generati e scambiati in SPC, di collegare tali dati ad altri dati delle PA (centrali e locali), e di arricchire i dati con opportuni metadati semantici per stabilire unostandard di qualità a livello di pubblicazione, di utenza e di interoperabilità nella PA. Il servizio è pensato sia per la gestione dei meri dati pubblici che per la gestione di tutti quei dati relativi allefunzioni di back end delle PA.

Il documento evidenzia, infine, che l'elenco e le descrizioni delle tipologie di dati individuate e delletecnologie non vogliono essere esaustivi ma rappresentano il punto di convergenza dei membri delgruppo durante il periodo di operatività dello stesso. Il gruppo, infatti, ha convenuto che un ulteriorestudio nei prossimi mesi è necessario in questo ambito, che coinvolga altre tipologie di dati e checonsenta di aggiornare opportunamente le linee guida sulla base degli sviluppi di standardizzazione inquesto settore.

3. AMMINISTRAZIONE PUBBLICA ED ENTI PRIVATI: STATO DELL'ARTE ED ESEMPI RILEVANTI

3.1 Open Data nel mondo

Il primo e più celebre portale nazionale dei dati aperti, che ha sostanzialmente fatto da modello ai successivi, è stato il **data.gov** degli Stati Uniti d'America, lanciato nel maggio 2009 dal *Chief Information Officer* (CIO) del governo Obama con l'obiettivo principale di raccogliere in un unico portale tutte le informazioni rese disponibili dagli enti statunitensi (e non solo) in formato aperto. Poco tempo dopo, sempre il governo Obama ha impostato il "gold standard" degli Open Data, con la direttiva Open Government pubblicata nel dicembre 2009, dove si legge che "per quanto possibile e soggetti a valide restrizioni, gli enti dovrebbero pubblicare online le informazioni in un formato aperto in modo che possano essere recuperate, scaricate, indicizzate e ricercate da comuni applicazioni web di ricerca. Un formato aperto deve essere indipendente dalla piattaforma, machine-readable, e reso disponibile al pubblico senza restrizioni che ne impediscano il riuso". Di lì a pochi mesi anche il Regno Unito ha aperto il suo **data.gov.uk**, fortemente voluto e sponsorizzato da Tim Berners-Lee, l'inventore del World Wide Web. In pochi anni la pratica degli Open Data e dei data-store governativi si è estesa fino in Australia (**data.gov.au**), Canada (**data.gc.ca**), Norvegia (**data.norge.no**), Francia (**data.gouv.fr**), etc. In Italia, il portale di riferimento è **dati.gov.it** (cfr. avanti). Per una aggiornata ed esaustiva descrizione dello stato dell'arte sui portali governativi che pubblicano dati aperti nel mondo, basta accedere al sito web <http://www.data.gov/opendatasites> e navigare la mappa interattiva.

Fig. 3 – Mappa degli Open Data governativi (aggiornata a gennaio 2013)



Fonte: data.gov (USA)

3.1.1 Regno Unito: il portale “data.gov.uk”

Come dicevamo, subito dopo gli Stati Uniti, uno dei primi Governi a fare da traino per la promozione della trasparenza attraverso lo sviluppo dei dati aperti è stato il Regno Unito, soprattutto con la creazione del web portal istituzionale **data.gov.uk**, realizzato e pubblicato nel 2010 su impulso della Commissione per la Trasparenza (Transparency Board) e con il supporto di Tim Berners Lee. Sempre in Gran Bretagna, la Open Knowledge Foundation ha creato il CKAN (*Comprehensive Knowledge Archive Network*), dove sono catalogati tutti i dati del portale governativo data.gov.uk oltre che un numero crescente di registri Open Data internazionali.

Il portale data.gov.uk, recentemente aggiornato nella sua veste grafica e nell’interfaccia utente, è estremamente ricco di contenuti: oltre alla possibilità di scaricare circa 9.000 dataset, si possono consultare centinaia di case histories legate agli Open Data, consultare le applicazioni tecnologiche da essi scaturite (per esempio Apps per smartphone), esaminare nel dettaglio (fino al centesimo) la spesa pubblica del Governo, interagire a vari livelli con la redazione del portale.

3.1.2 Unione Europea: il lancio del nuovo portale Open Data

Con un certo ritardo rispetto ai Paesi anglosassoni (e non solo), l’Unione Europea ha lanciato, nel mese di gennaio 2013, il suo portale ufficiale Open Data, **open-data.europa.eu**³⁵.

Il portale offre l’accesso a oltre 5.800 dataset, perlopiù provenienti dai cataloghi Eurostat. Con questa iniziativa, l’Unione Europea intende promuovere con forza l’apertura e il riutilizzo libero dei dati dell’informazione del settore pubblico in Europa, come parte integrante della strategia Open Data europea. Nei mesi successivi al suo lancio, il portale mirerà a garantire l’accesso a tutti i portali Open Data nazionali, regionali e locali dei 27 Paesi Membri dell’Unione.

3.1.3 America Latina e Caraibi

Condividere l’informazione governativa in formati aperti, accessibili e strutturati potrebbe accrescere in maniera sostanziale la trasparenza e l’affidabilità del processo di elaborazione e rafforzamento delle politiche pubbliche latino-americane. Su questa base è nato il progetto denominato “Open Data for Development in Latin America and the Caribbean”, a cui collaborano W3C Brazil, la Commissione Europea per l’America latina e i Caraibi (ECLAC), l’Osservatorio per la Società dell’informazione in America Latina e Caraibi (OSILAC) e il Centro Internazionale di Ricerca e Sviluppo del Canada (IDRC).

³⁵ La versione italiana è reperibile al link <http://open-data.europa.eu/open-data/it>.

Il progetto ha 6 obiettivi specifici:

- tracciare le principali iniziative in America Latina e nei Caraibi sul piano della condivisione strutturata dei dati economici, sociali e ambientali e progettare un quadro metodologico per esaminare la relazione tra apertura dei dati e qualità delle politiche pubbliche;
- studiare e discutere strategie alternative per incentivare la formazione tecnica nelle agenzie governative e negli Osservatori della regione, attuando in tal modo l'apertura degli archivi per la progettazione, il controllo e la valutazione delle politiche pubbliche;
- sostenere le reti di ricerca in America Latina e nei Caraibi per la produzione di nuove informazioni e la creazione di applicazioni e servizi innovativi basati su dati aperti;
- esaminare la relazione tra sviluppo economico più inclusivo e trasparenza dei dati nei principali settori economici;
- aumentare la consapevolezza tra le comunità dei responsabili politici, funzionari pubblici e ricercatori del potenziale degli Open Data e di strategie operative appropriate per la sua efficace attuazione;
- valutare il potenziale di strategie Open Data nella progettazione e attuazione di politiche pubbliche volte a promuovere lo sviluppo economico e l'inclusione sociale nei paesi dell'America Latina e nei Caraibi.

3.1.4 Open Data per il pubblico e il privato: alcuni esempi illuminanti

Nell'ambito della trasparenza, progetti come il finlandese **Taxtree** ("albero delle tasse") e il britannico **WhereDoes My Money Go** ("dove vanno i miei soldi") permettono di identificare come i soldi delle tasse dei cittadini siano impiegati dal Governo.

Molti siti, tra cui il danese **folketsting.dk** e l'italiano **openparlamento.it**, tracciano le attività dei Parlamenti e il processo di formazione delle leggi, in modo da mostrare cosa succede esattamente e quali parlamentari sono coinvolti nelle varie attività.

I dati aperti governativi possono inoltre aiutare a prendere decisioni migliori nella nostra vita privata, o renderci più attivi nell'ambito della società civile.

In Danimarca, una sviluppatrice ha creato **findtoilet.dk**, che permette di accedere alla lista di tutti i bagni pubblici del Paese, così anche chi soffre di problemi di incontinenza ora si sente più rassicurato dovendo uscire di casa.

In Olanda il servizio **vervuilingsalarm.nl** avvisa l'utente quando la qualità dell'aria del suo quartiere raggiunge una soglia critica da lui definita.

Servizi come **mapumental** nel Regno Unito e **mapnificent** in Germania ci fanno capire dove possiamo andare ad abitare impostando i tempi massimi di percorrenza casa/ufficio, i prezzi delle case e la bellezza del quartiere.

Tutti questi esempi utilizzano dati aperti rilasciati dai governi.

Anche dal punto di vista economico i dati aperti hanno un'enorme importanza. Svariati studi hanno stimato il valore economico dei dati aperti in diverse decine di miliardi di euro ogni anno nella sola Europa. Nuovi prodotti e nuove aziende stanno riutilizzando dati aperti.

Il sito danese **husetsweb.dk**, per esempio, aiuta a trovare i modi migliori di risparmiare energia elettrica in casa, inclusa la pianificazione finanziaria e la possibilità di contattare gli artigiani che potranno eseguire il lavoro: funziona grazie al riutilizzo di dati catastali, a informazioni sugli incentivi governativi e al registro delle imprese locali.

Anche per il governo stesso i dati aperti hanno un grande valore, e possono aumentare l'efficienza della macchina amministrativa. Il Ministero olandese dell'Istruzione ha pubblicato online tutti i dati relativi al sistema educativo consentendone il riuso. Da allora, il numero di domande ricevute è sceso, riducendo il carico di lavoro e i costi, e anche per i dipendenti pubblici è diventato più facile rispondere alle domande residue, perché è chiaro dove possono essere trovati i dati che servono per rispondere.

Mentre ci sono numerosi esempi in cui i dati aperti stanno già creando vantaggi economici e sociali, ancora non sappiamo quali nuovi utilizzi saranno possibili in futuro. Nuove combinazioni di dati possono creare nuova conoscenza e nuove intuizioni, che possono portare a campi di applicazione inimmaginabili. Come accadde in passato, ad esempio, quando il dottor Snow scoprì la correlazione tra l'inquinamento dell'acqua potabile e il colera nella Londra dell'Ottocento, combinando i dati sui morti per colera con quelli sull'ubicazione dei pozzi: il fatto portò alla costruzione del sistema fognario a Londra, migliorando di molto le condizioni generali di salute della popolazione. Probabilmente vedremo di nuovo nascere intuizioni simili dalla combinazione di insiemi diversi di Open Data.

3.2 Open Data in Italia: esempi e casi di successo

In Italia si è cominciato a parlare di Open Data anche grazie al progetto **OpenStreetMap**, che si potrebbe definire la versione "aperta" di Google Map. Con il fine di creare uno stradario liberamente utilizzabile, tra il 2007 e il 2010 alcune amministrazioni locali, grazie all'iniziativa di volontari (tra cui Merano, Vicenza, Montecchio Maggiore e Schio), hanno pubblicato con licenza aperta i dati dei propri stradari e altri propri dati. Il Ministro per la Pubblica Amministrazione e l'Innovazione, Renato Brunetta, in una intervista rilasciata il 3 giugno 2010 all'allora direttore di "Wired Italia" Riccardo Luna, ha anticipato la realizzazione di un portale italiano di Open data sul modello dei datagov anglosassoni entro la fine dell'anno. Il 18 ottobre

2011 il portale **dati.gov.it** è stato messo on line, e rappresenta tuttora il portale di riferimento per gli Open Data in Italia.

Va però precisato che il primo data-store italiano è stato quello della Regione Piemonte, **dati.piemonte.it**, lanciato nel maggio 2010, all'interno del quale sono catalogati dati aperti riconducibili ai vari enti regionali (Comuni, Province, etc.): il sito resta al momento la più riuscita e strutturata esperienza nazionale sul tema dei dati aperti.

Dopo circa un anno dalla nascita del data-store piemontese, anche la Regione Emilia-Romagna, nell'ottobre 2011, ha pubblicato online il suo catalogo di dataset (**dati.emilia-romagna.it**), seguita dopo poco tempo dalla Regione Veneto.

Si può dire comunque che, a partire dal lancio del portale **dati.gov.it**, si è aperta una nuova stagione per l'innovazione e la trasparenza nella PA, una strada verso l'Open Data italiano³⁶, che ha visto protagonisti non solo alcune Regioni "virtuose", ma anche diversi Comuni (Firenze, Torino, Bologna, Udine, Roma, etc.) e altri enti locali (Provincia di Roma, Inps, Cnel etc.).

Oltre ai siti istituzionali, Open Knowledge Foundation Italia e il Centro NEXA su Internet & Società del Politecnico di Torino offrono il repository **it.ckan.net**, dove chiunque può segnalare i dataset italiani disponibili online.

Un servizio analogo è gestito dalla comunità del sito **Spaghetti Open Data (SOD)**³⁷ disponibile alla pagina "Dati". Anche sul portale **AppsForItaly**³⁸ è possibile trovare ulteriori elenchi di dataset aperti in Italia. Da settembre 2012 è disponibile inoltre **DatiOpen.it**³⁹, un'altra iniziativa indipendente che offre servizi gratuiti di segnalazione, caricamento e visualizzazione di dati. È necessario segnalare, infine, il portale di ricerca del web semantico **LOD Live**⁴⁰, un progetto sperimentale nato per divulgare la filosofia dei Linked Open Data e per creare uno strumento in grado di dimostrare che le risorse pubblicate secondo gli standard del W3C per il web semantico possono essere facilmente accessibili e comprensibili con strumenti adeguati.

3.2.1 Lo stato dell'arte in Italia

Va detto che i dati a disposizione in modo strutturato in Italia sono ancora pochi, ma è possibile trovare vari dati non strutturati. D'altra parte, il processo di gestione delle informazioni prodotte sta progressivamente evolvendo verso modelli più aperti, coinvolgendo in questo sempre più enti, sia pubblici che privati, che di conseguenza pubblicano i propri dati in formato "open".

³⁶ Per alimentare la discussione online sul portale Dati.gov.it, sulla licenza IODL e più in generale sul tema Open Data, è stato creato un gruppo su Innovatori PA, community promossa da Formez PA e Forum PA che raccoglie più di 4.000 operatori del settore pubblico. Su Innovatori PA vengono inoltre segnalate iniziative, convegni, seminari online dedicati ai temi dell'Open Data e dell'Open Government.

³⁷ www.spaghettiopendata.org

³⁸ www.appsforitaly.org

³⁹ www.datiopen.it

⁴⁰ www.lodlive.it

Per rimanere aggiornati su tutte le iniziative di Open Data avviate in Italia, il portale dati.gov.it ha creato un dataset che descrive i dati catalogati sul sito e li rende maggiormente interpretabili attraverso una interessante **infografica** il cui dataset viene aggiornato periodicamente⁴¹:

Fig. 4 –Localizzazione dei data-store in Italia (mappa aggiornata a gennaio 2013)



Fonte: dati.gov.it

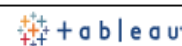
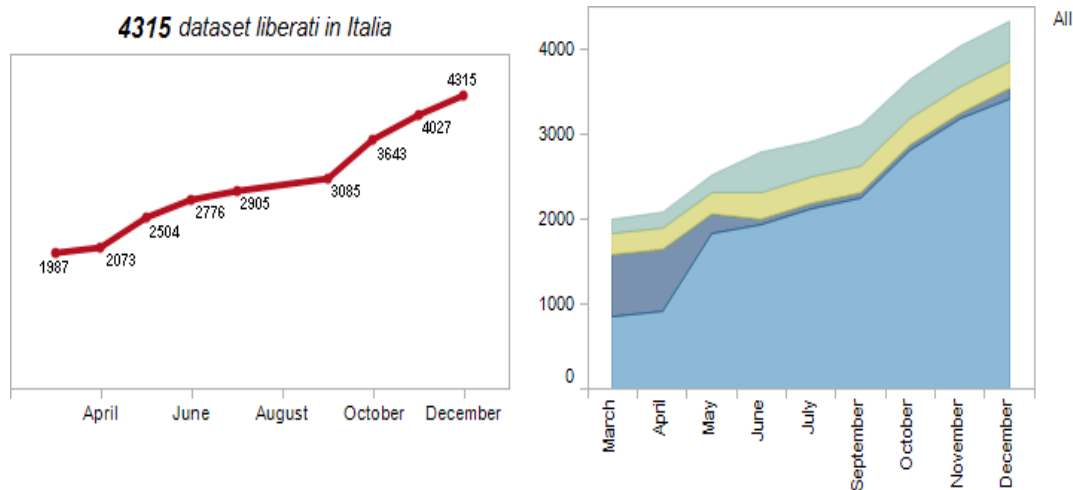
Come si può vedere, gli Open Data sono già una realtà in diverse Regioni e alcuni Comuni (concentrati soprattutto al Nord), e le proposte di legge regionali, come abbiamo visto, stanno aumentando di pari passo con l'aumento dei portali di regionali dedicati alla diffusione dei set di dati, anche se c'è ancora molto da fare.

Più precisamente, al gennaio 2013 in Italia sono disponibili **oltre 4.000 dataset** rilasciati in formato aperto. Le modalità di apertura adottate dalle amministrazioni che rilasciano open data sono diverse. Nella figura successiva, il quadrante di sinistra descrive il trend storico dei data-store italiani fino a dicembre 2012, mentre il quadrante di destra riporta una ripartizione dei dataset per livello di riusabilità, dove la gradazione dei colori rappresenta la scala di Tim

⁴¹[http://www.dati.gov.it/content/infografica#Quanti sono i dati aperti in Italia?](http://www.dati.gov.it/content/infografica#Quanti%20sono%20i%20dati%20aperti%20in%20Italia?)

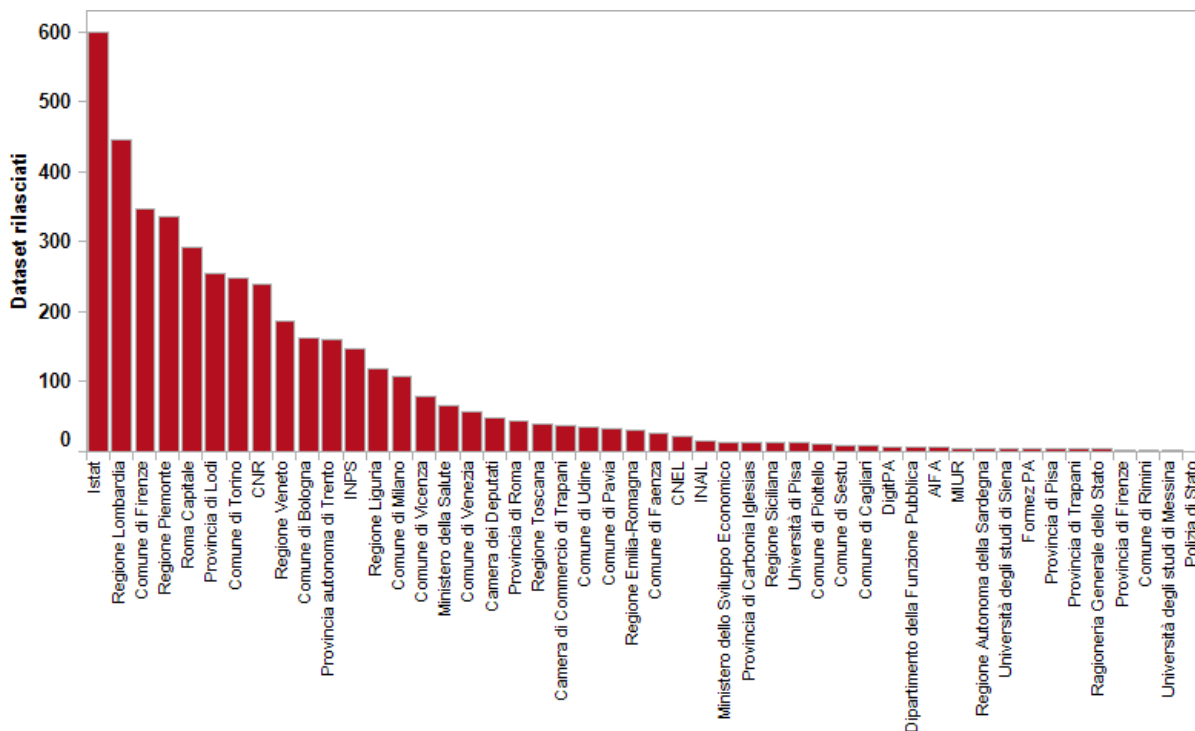
Berners-Lee: il blu e l'azzurro in basso riguardano i dataset meno utilizzabili (da 2 a 3 stelle), mentre l'ocra e il celeste in alto quelli più "usabili" (da 4 a 5 stelle).

Fig. 5 – Quanti sono i dataset in Italia: trend e tipologie (dati aggiornati a dicembre 2012)



Fonte: dati.gov.it

Per quanto riguarda i soggetti che rilasciano dataset, di seguito si riporta un confronto per istogrammi tra le PA che pubblicano i propri dati in formato aperto:



In cima alla lista figura Istat, seguito da Regione Lombardia, Comune di Firenze, Regione Piemonte e Roma Capitale.

A titolo esemplificativo, tra gli ultimi dataset pubblicati in Italia (al dicembre 2012) e catalogati su dati.gov.it, figurano il CNEL (“Contratti collettivi vigenti in ciascun settore pubblico e privato”, formato CSV), la Regione Liguria (“Siti contaminati, bonificati o sui quali è in corso l'intervento di bonifica”, formato shapefile), Provincia di Roma (“Il dataset del mosaico PRG – Piano Regolatore Generale”, formato KML), il Comune di Bologna (“Bilancio di previsione e piano esecutivo di gestione anno 2012”, formato CSV), la Provincia di Trapani (“Musei”, “Farmacie”, “Istituti Scolastici”).

3.2.2 Gli esempi italiani più rilevanti

Oltre ai contenitori e aggregatori di Open Data sin qui descritti (dati.gov.it, datiopen.it, CKAN Italia, LOD Live etc.), è doveroso citare qui alcuni tra gli esempi italiani più riusciti di pubblicazione e divulgazione di Open Data.

Enti Pubblici

- **Istat** (dati.istat.it): I.Stat è il datawarehouse delle statistiche prodotte dall'Istat, un patrimonio informativo completo ed omogeneo, unico per la statistica italiana. L'accesso è libero e gratuito per tutti. Le statistiche sono organizzate per temi strutturati in un albero gerarchico a due livelli. I dati sono presentati in forma aggregata in tavole multidimensionali, con la possibilità di comporre tabelle e grafici personalizzati, agendo sulle variabili, i periodi di riferimento e la disposizione di testate e fiancate. Il sistema è interrogabile secondo le seguenti modalità: per parola chiave, per tema e per territorio. Un ampio apparato di metadati agevola il reperimento e la comprensione delle statistiche da parte degli utenti. I contenuti di I.Stat sono costantemente aggiornati.
- **Camera dei Deputati** (dati.camera.it): piattaforma di pubblicazione e condivisione di Linked Open Data sull'attività e gli organi della Camera, da scaricare o interrogare liberamente. Il portale presenta un catalogo completo di dati e documenti digitali su tutte le legislature precedenti, dalla I del Regno di Sardegna alla XV della Repubblica, che alimentano il Portale storico della Camera dei deputati. Sono inoltre disponibili i dataset di deputati, organi e gruppi parlamentari della legislatura corrente. L'insieme dei dati viene periodicamente aggiornato ed esteso. Interessante la possibilità di ricercare e visualizzare i dati secondo la logica del web semantico (Linked Open Data).
- **Open Coesione** (opencoesione.gov.it): primo portale sull'attuazione degli investimenti programmati nel ciclo 2007-2013 da Regioni e amministrazioni centrali dello Stato con le risorse per la coesione. I dati sono pubblicati perché i cittadini possano valutare se i progetti corrispondono ai loro bisogni e se le risorse vengono impegnate in modo

efficace. Leggi qui per saperne di più. I progetti monitorati, al dicembre 2012, sono circa 590.000.

- **INPS Open Data** (inps.it): i dati INPS in formato aperto, la cui offerta è in via di miglioramento dopo una fase di stallo.
- **CNR** (data.cnr.it): le tecnologie del portale Open Data del Centro Nazionale Ricerche si basano sul formato RDF e seguono il paradigma dei Linked Data.
- **MIUR** (archivio.pubblicaistruzione.it): la sezione Open Data del MIUR, da cui è possibile scaricare, per esempio, tutte le informazioni sull'anagrafica scolastica, nonché visualizzare su mappa i geodati del progetto "Scuola in Chiaro".
- **Ministero della Salute** (dati.salute.gov.it): il portale Open Data del Ministero della Salute.
- **Regione Piemonte** (dati.piemonte.it): il primo portale di Open Data pubblicato in Italia, nonché uno dei più interessanti, curati e "userfriendly" in Italia e in Europa.
- **Comune di Firenze** (dati.comune.firenze.it): i dati aperti del Comune di Firenze. Il portale presenta molte funzionalità e sezioni, tra cui "Open Data per tutti", dove vengono visualizzati graficamente in maniera semplificata i dataset più significativi;
- **Comune di Bologna** (dati.comune.bologna.it): anche se ancora in versione "beta", il portale del Comune di Bologna si contraddistingue per facilità d'uso, aggiornamento costante e interazione con l'utente. Arricchisce il tutto una sezione con le ultime news dal mondo Open Data a Bologna e in Emilia-Romagna.
- **Comune di Milano** (dati.comune.milano.it): dati anagrafici, indagini sui consumi, servizi sociali e altro sul portale Open Data del capoluogo lombardo.
- **Comune di Udine** (comune.udine.it): uno tra i primi Comuni medio-piccoli a dedicare attenzione agli Open Data.
- **Provincia di Roma** (opendata.provincia.roma.it): in attesa della pubblicazione del portale Open Data della Regione Lazio (il dominio "dati.lazio.it" è stato appena registrato), la Provincia di Roma ha pubblicato (al dicembre 2012) 48 dataset che riguardano Lavoro e Formazione, Turismo, Territorio e Ambiente, Cultura, Trasparenza Amministrativa, Servizi al Cittadino.
- **Regione Veneto**: il portale Open Data della Regione Veneto.
- **Regione Lombardia** (dati.lombardia.it): i dati aperti della Regione Lombardia.

Iniziative indipendenti e private:

- **OpenGeoData** (opengeodata.it): La neonata Associazione OpenGeoData Italia opera per stimolare una politica di apertura dei dati geografici della Pubblica

Amministrazione, che sono i dati più richiesti in assoluto, affinché ognuno li possa usare per trarne nuove forme di conoscenza e per creare nuove attività imprenditoriali.

- **GFOSS.it**: il portale dell'Associazione Italiana per l'Informazione Geografica Libera offre una sezione di download e un elenco di risorse di geodati in formato aperto.
- **Decoro Urbano**: il sito del social network "Decoro Urbano", nato per favorire il dialogo tra le Pubbliche Amministrazioni e i cittadini desiderosi di contribuire personalmente alla cura della propria città, permette di segnalare le segnalazioni degli utenti in formato aperto.
- **OpenPolis** (openpolis.it): il portale dell'Associazione OnlusOpenpolis, che dal 2006 assolve alla funzione di "controllo" dell'attività politica italiana. Contiene al suo interno i dataset "openparlamento", "openmunicipio", "openpolitici".
- **Sardinia Open Data** (sardiniaopendata.org): questa recente iniziativa, avviata in Sardegna da un gruppo di giovani appassionati di Open Data, si pone l'obiettivo di promuovere i dati aperti presso le pubbliche amministrazioni e le associazioni. Ogni informazione e servizio prodotti da Sardinia Open Data sono correlati da dati aperti, riutilizzabili e disponibili a chiunque senta la necessità di usarli come base per un nuovo studio o un nuovo progetto che contribuisca alla crescita locale.
- **Enel** (data.enel.com): caso estremamente raro, in Italia, di un'azienda che rende disponibili (alcuni) suoi dati in formato aperto. Nel nostro Paese, infatti, al momento pochissime aziende hanno deciso di impegnarsi nel rendere disponibile una rendicontazione trasparente della loro attività, anche se l'apertura da parte di Enel potrebbe portare conseguenze molto importanti imponendo forse un nuovo standard. In Italia, Enel è stata la prima azienda a muoversi in questa direzione, rendendo pubblici nel sito web del gruppo i dati del Bilancio consolidato, della Relazione e Bilancio di Esercizio e del Bilancio di Sostenibilità, permettendo di riprodurre, distribuire, comunicare ed esporre al pubblico, modificare, anche a scopi commerciali (con il vincolo di riconoscere la paternità dell'opera all'autore nei modi indicati dall'autore stesso o da chi ha dato l'opera in licenza). Inoltre, si possono scaricare i dati messi a disposizione in formato XLS, CSV e XML. Enel Open Data adotta le licenze Creative Commons.

4. IL VALORE SOCIO-ECONOMICO DEGLI OPEN DATA: DATI E PROSPETTIVE DI SCENARIO

Grazie allo sviluppo delle tecnologie digitali alla diffusione nel livello e nella capacità di utilizzo dei cittadini e al crescente livello di informatizzazione della Pubblica Amministrazione, in questi ultimi anni si è fatta strada la consapevolezza che i dati del settore pubblico (Public Sector Information - PSI) possano svolgere un ruolo importante non solo ai fini della trasparenza amministrativa e della partecipazione pubblica ma anche sul piano economico.

Quasi tutte le informazioni pubbliche hanno, infatti, un valore intrinseco, che va oltre l'assolvimento dei compiti istituzionali dell'ente che le detiene. Spesso, la singola informazione ha un'utilità limitata, ma la lettura in serie storica, la raccolta organizzata e la combinazione di più informazioni con fonti differenti possono generare un valore aggiunto notevole. Così, ad esempio le informazioni geografiche sono alla base di cartine e mappe, le quali possono servire a fornire servizi online; mentre i dati catastali possono servire a operatori immobiliari e istituti di credito per fare incontrare la domanda e l'offerta di alloggi e per proporre soluzioni di finanziamento. I metadati relativi a musei e archivi locali, così come le immagini delle opere in essi disponibili, possono stimolare il turismo oltre alimentare i percorsi educativi e la conoscenza. O, ancora, i dati meteorologici possono essere utilizzati sia per servizi a valore aggiunto in tempo reale (ad esempio messaggi inviati sul cellulare dell'abbonato con previsioni selettive del tempo), sia per arricchire il dibattito pubblico sul cambiamento climatico.

I dati sui bilanci delle imprese aggregati sulla base di criteri statistici, territoriali o merceologici offrono importanti informazioni di marketing, di analisi della concorrenza e di analisi scenario micro e macro economico di assoluta utilità per i policy makers o per qualunque strategia di investimento.

Gli studi macroeconomici mirano a stimare il valore dell'informazione del settore pubblico in termini di partecipazione, trasparenza, e valore economico che si può ottenere grazie a un più facile accesso e a un più ampio uso di tale informazione.

Le leggi approvate dalla Comunità Europea si sono focalizzate inizialmente sul diritto del cittadino di accedere ai dati per una questione di trasparenza. Successivamente, l'attenzione si è estesa alla questione della partecipazione e della messa a disposizione dei dati secondo modalità che permettessero di riutilizzarli liberamente, in contesti diversi e innovativi rispetto a quelli nei quali essi sono stati raccolti. Questa tendenza, già presente in documenti strategici quali l'Agenda Digitale, è stata ulteriormente e ripetutamente confermata. Anche nel corso della conferenza stampa tenuta dalla commissaria Neelie Kroes il 12 dicembre 2011, è stato ribadito che i dati della PA sono in grado di generare ricchezza e posti di lavoro.

A tale riguardo è utile ricordare come uno studio del progetto MEPSIR (Measuring European Public Sector Information Resources) abbia quantificato in 28 miliardi di Euro il valore economico dei dati prodotti dalle amministrazioni pubbliche nel mercato del loro riutilizzo nel 2008. Le assunzioni di base di tipo conservativo prevedevano una crescita del 7% annuo il che proietterebbe il dato ad un livello vicino a 40 miliardi nel 2012.

Molto interessanti appaiono le proiezioni sugli effetti diretti e indiretti del sistema dei dati aperti. Uno studio del 2011 della Commissione europea realizzato da Graham Vickery ha evidenziato un impatto sul sistema economico in senso lato collegato all'economia dei dati pubblici aperti pari a 140 miliardi di euro annui. Lo studio evidenzia come il dato possa crescere in ragione del livello di disponibilità e apertura dei dati e anche delle scelte tariffarie che se gratis o al più al costo marginale possono produrre un forte impulso alla diffusione dei dati ed allo sviluppo di applicazioni alimentando quindi un mercato indotto collegato alla maggiore fruizione di informazioni che potrebbe arrivare a 200 miliardi annui.

4.1 Esternalità positive

Il valore economico dei dati e delle informazioni si sostanzia nella misura in cui i dati si trasformano in "attivatori" di nuove e, anche, imprevedibili filiere di attività e servizi digitali a valore aggiunto per i potenziali utenti costituiti da cittadini, imprese e altre amministrazioni ad un diverso livello di governo.

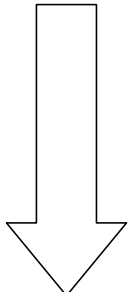
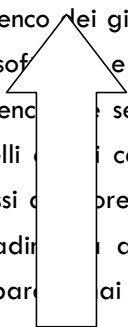
Tornando all'esempio dei dati meteorologici pubblici, non si pensi solo alle classiche previsioni del tempo, ma anche a servizi specializzati, ad esempio per l'agricoltura o il turismo invernale. In alcuni casi, gli introiti fiscali indiretti generati da tali attività possono essere pari o superiori a quelli che la PA otterrebbe facendo direttamente pagare per l'utilizzo dei dati;

Una delle ragioni più significative per rendere disponibili e liberamente riutilizzabili le informazioni detenute dalle PA risiede nella fantasia dei cittadini e degli imprenditori. Le PA hanno infatti la disponibilità di enormi quantità di dati, ma hanno anche molti compiti e risorse limitate e spesso gli incentivi e gli strumenti per la creazione o il miglioramento di servizi risultanti dall'elaborazione o aggregazione dei dati disponibili scarseggiano. Per questo, è fondamentale permettere ad altri di riutilizzare l'informazione del settore pubblico. In quest'ottica, l'informazione del settore pubblico dovrebbe essere vista come una piattaforma, le cui applicazioni sono ancora quasi tutte da scrivere. L'analogia con Internet come piattaforma di innovazione è calzante. Internet, di per sé, è "solo" una piattaforma, ma ha reso possibile la posta elettronica, il Web e Skype, lo streaming video e centinaia di altre applicazioni, declinate in migliaia di modi differenti. In modo simile, l'iPhone di Apple ha rapidamente catalizzato le capacità creative di migliaia di sviluppatori, i quali lo hanno dotato di innumerevoli applicazioni, molte

più di quante Apple avrebbe potuto sviluppare senza aprire (almeno in parte) la propria piattaforma alla creatività del resto del mondo.

Commercial re-use of PSI	4.2 I Giacimenti informativi	cartographic information
	Geographic Information	land use info (cadastral data)
		spatial data/geographical coordinates
		administrative and political boundaries
		topographical information
		elevation data
		oceanographic data
	Meteorological and Environmental Information	hydrographic data
		environmental (quality) data
		research data
		meteorological (weather) data
	Economic and Business Information	financial data
		company information
		industry and trade information
		geographic information
	Social Information	attitude surveys
		data on health/illness
		education and labour statistics
	Traffic and Transport Information	transport network information
		traffic information
		transport statistics
		car registration data
	Tourist and Leisure Information	hotel information
		tourism statistics
		entertainment (local and national)
	Agricultural, Farming, Forestry and Fisheries information	cropping/land use data
		farm incomes/use of resources
		fish farming/harvest information
		live stock data
	Natural Resource Information	biologic and ecologic information
		energy resource/consumption information
		geological and geophysical information
	Legal System Information	crime/conviction data
		laws
		information on rights and duties
		information on legislation
		information on judicial decisions
	Scientific Information and Research data	patent and trademark information
		university research
		publicly-funded research institutes
		governmental research
	Educational Content	academic papers and studies
		lecture material
	Political Content	governmental press releases
		local and national proceedings of governments
		green papers
	Cultural Content	museum material
		gallery material
		archeological sites
		library resources
		public service broadcast archives
		other public archives

L'elenco dei giacimenti informativi davvero ampio ma è il sistema di utilizzi e alle applicazioni, di software e modalità di fruizione, di contenuti derivati che appare davvero sconfinato. L'elenco segue da una chiara indicazione della ricchezza dei giacimenti informativi evidenziando quelli che li caratterizzano per un pieno riuso (PSI) sulle tracce di quelli che possono essere messi a disposizione anche rispetto a dimensioni di inclusione sociale, alla educazione e istruzione e di cittadini attiva (Public service Content) in una dimensione di fruizione che in realtà appare ai sempre più labile.



4.3 Benefici “interni” alle amministrazioni

Dal lato dei benefici interni alle amministrazioni nell'adozione di pratiche di Open Data, sembra utile concentrare l'attenzione sui risparmi di spesa conseguibili. Le Pubbliche amministrazioni sono esse stesse fruitori e utenti di informazioni provenienti da altri Enti, allo scopo di erogare un servizio al pubblico e/o per la chiusura dell'iter di un procedimento amministrativo interno. In tale prospettiva, la libera accessibilità a dati e informazioni di altri Enti consente di conseguire:

- miglioramenti di produttività ed efficienza nella gestione dei processi amministrativi interni grazie alla tempestiva e immediata disponibilità del dato (es. numero di giorni/uomo necessari all'adempimento di un'attività, ecc.);
- riduzione dei costi legati allo scambio delle informazioni (interoperabilità) delle basi dati nazionali (si segnala ad es. per il caso italiano il repertorio nazionale dei dati territoriali, indice nazionale delle anagrafi, banca dati nazionale dei contratti pubblici, casellario giudiziario, registro imprese, ecc.);
- monitoraggio della spesa pubblica sostenuta dalle diverse amministrazioni nell'esercizio della propria missione istituzionale, del suo funzionamento e per le diverse finalità di sviluppo territoriale sia di origine nazionale che comunitaria;
- risparmi di spesa derivanti dalla realizzazione e gestione di nuovi servizi digitali erogati da terze parti grazie all'utilizzo solo ed esclusivamente di dati (tecniche di *mash-up*) secondo l'approccio "*the State brings its data and they do the rest*";
- risparmi derivanti dalla riduzione dei “costi di transazione” per il reperimento dei dati di e da altre amministrazioni.
- Risparmi derivanti dal contenimento di extracosti e inefficienze connessi a fenomeni distortivi grazie alla maggiore trasparenza ed alla possibilità di valutare in modo più rapido e chiaro i risultati di scelte pubbliche, così da ridurre le discrezionalità che favoriscono fenomeni di clientela o peggio di corruzione

5. IL PROCESSO DI APERTURA DEI DATI PUBBLICI E LE COMPETENZE TECNICHE RICHIESTE

Con il vademecum “Open Data”, pubblicato dal Foromez a fine 2011 nell'ambito delle “Linee Guida per i siti Web della PA”⁴² e per supportare le amministrazioni pubbliche

⁴² Pubblicate ai sensi dell'art. 4 della Direttiva del Ministro per la Pubblica Amministrazione e l'Innovazione del 26 novembre 2009, n. 8.

nell'applicazione delle norme precedentemente descritte (in particolare dell'art. 52 comma 1-bis del Codice dell'Amministrazione Digitale), sono stati definiti e forniti alle PA alcuni principi e regole comuni da seguire nell'attività di apertura e pubblicazione dei dati pubblici, soprattutto per quanto riguarda i siti web della PA, che costituiscono ovviamente il luogo preferibile per la pubblicazione dei dati pubblici da parte di ciascuna Amministrazione.

Proprio nell'apertura dei dati pubblici, come vedremo in questo Capitolo, le competenze tecniche di tipo ingegneristico (soprattutto in ambito informatico e gestionale) possono giocare un ruolo fondamentale.

5.1 Apertura dei dati pubblici: modelli e processi

Ogni amministrazione pubblica italiana, nel processo di apertura dei dati in suo possesso, è chiamata a rispettare i principi dettati dalle norme precedentemente descritte (Cap. 2). Dal punto di vista tecnico-organizzativo, comunque, le "Linee Guida per i siti web delle PA" suggeriscono che l'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 utilizzazione. 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, come per esempio:

- 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.

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.

A tale proposito, il vademecum “Open Data” pubblicato dal Formez nell’ambito delle “Linee Guida per i siti web delle PA”⁴³ suggerisce di declinare il tipico approccio per “processi” dei sistemi di gestione della Qualità⁴⁴ nell’attività di apertura dei dati dell’Amministrazione, in quanto permette di:

- comprendere i requisiti di partenza dell’Amministrazione nelle diverse fasi connesse all’apertura dei dati;
- valutare i processi connesi all’apertura dei dati in termini di valore aggiunto per l’Ente e i suoi destinatari/utenti (cittadini, aziende, organizzazioni) derivante dalla loro apertura;
- conseguire risultati misurabili in termini di efficacia complessiva dell’azione;
- ottenere un miglioramento continuo dei processi basato su misurazioni oggettive.

Descrivere i processi connesi all’attività di apertura dei dati a disposizione di una PA vuol dire 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).

5.1.1 Il modello PDCA per la gestione delle fasi del processo

Nella strutturazione dei processi connesi all’apertura dei dati della Pubblica Amministrazione, il Formez propone di fare riferimento al ciclo di Deming, conosciuto anche come modello PDCA (da “Plan”: pianificare – “Do”: eseguire – “Check”: controllare – “Act”: agire), alla base dell’organizzazione per processi dei sistemi di qualità orientati al miglioramento continuo.

L’applicazione di questo modello alla gestione dei dati da aprire nell’Amministrazione Pubblica, secondo il vademecum Formez, 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, da valorizzare attraverso il suo utilizzo. 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.

Ecco come può apparire il percorso del modello PDCA applicato all’apertura dei dati della Pubblica Amministrazione:

1) Plan – Pianificare

⁴³ “Linee Guida per i siti web della PA – Vademecum – Open Data: come rendere aperti i dati delle pubbliche amministrazioni”, versione Beta 2011, Formez, ottobre 2011, Gangemi Editore.

⁴⁴ Norme UNI EN ISO 9001:2000, che definiscono un processo come “*un insieme di attività correlate o interagenti che trasformano elementi in ingresso (input) in elementi in uscita (output) attraverso una fase di elaborazione*”.

- *Determinare gli obiettivi.* Gli obiettivi della politica di Open Data vanno indicati in modo esplicito, così che sia più semplice procedere nelle fasi successive di identificazione dei dataset da aprire.
- *Determinare i metodi e pianificare le azioni.* Definiti gli obiettivi da raggiungere, è necessario identificare le procedure che consentano di perseguirli. Determinare un metodo vuol dire standardizzarlo e renderlo utile e accessibile. L'apertura dei dati di una PA va reiterata nel tempo (i dati vanno aggiornati), e quindi è opportuno 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.

2) Do – Svolgere il lavoro

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

3) Check – Controllare gli effetti dell'apertura dei dati

- *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.

4) Act – Intraprendere azioni adeguate

- *Ottimizzare il sistema.* 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.

Va precisato che quello appena descritto è da considerarsi uno dei modelli possibili, il cui miglioramento è senz'altro auspicabile soprattutto quando si parla di "controllo degli effetti dell'apertura dei dati". Secondo la comunità di innovatori ed esperti di Open Data, infatti, il concetto di apertura deve essere mantenuto quanto più "neutro" possibile, ciò significando che tutti i dataset sono da considerarsi potenzialmente utili agli utenti, i quali decidono come utilizzarli secondo i propri scopi e i propri tempi. Gli indicatori proposti in fase di "check", infatti, non sembrano tener conto del fatto che una delle grandi potenzialità degli Open Data, e quindi anche i servizi e le applicazioni che da essi possono scaturire, risiede anche nella sua "universalità temporale": un set di dati aperti, probabilmente, potrà rivelarsi più utile tra 10 anni piuttosto che oggi, a seconda delle evoluzioni sociali-tecniche-economiche, di nuove competenze

da parte degli utenti, di nuovi mercati etc. Piuttosto che “controllare” gli effetti dell’apertura dei dati per valutare il tipo di dati maggiormente utilizzati dagli utenti, insomma, sarebbe forse più opportuno promuovere l’utilizzo dei dati nel tempo e a ogni livello, tracciando magari un’analisi periodica dei trend di utilizzo in grado di stimolare l’apertura di nuovi dati e nuove forme del loro utilizzo.

5.1.2 Le fasi del processo di apertura dei dati

Dal punto di vista più strettamente operativo, il vademecum Formez elenca così le fasi del processo di apertura dei dati nelle PA:

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

Per ogni fase è opportuno esplicitare le azioni da compiere, i tempi, le modalità operative, i ruoli e le responsabilità dei soggetti coinvolti.

La matrice operativa che ne scaturisce è la seguente:

NOME DATASET	Identificazione	Analisi	Pubblicazione	Diffusione
Azioni				
Tempi				
Modalità operative				
Ruoli/Responsabilità				

Fonte: Formez, Vademecum Open Data

Identificazione dei dati

Va innanzitutto specificato **dove sono i dati a disposizione dell’Amministrazione.**

Con lo sviluppo dell’ICT e la diffusione dei sistemi di e-Government, i dati della PA 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 una vera e

propria mappatura dei dati complessivamente disponibili. Definire dove sono i dati a disposizione vuol dire:

- 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 da diversi Uffici;
- identificare le strutture funzionali e le fonti dati di rilievo. In altri termini, è necessario identificare archivi strutturati, elenchi, basi dati prodotte da software in uso presso l'amministrazione.

Questa prima fase di mappatura, oltre a essere utile alla definizione di una strategia di apertura dei dati, lo è anche per determinare e comprendere meglio il patrimonio informativo detenuto dall'Amministrazione.

Una volta mappati e identificati i detentori di dati dell'Amministrazione, è necessario procedere a 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.** Sono tutti quei dati che servono all'Amministrazione per il corretto funzionamento dei suoi processi, cioè i dati in ingresso (*input*) che alimentano i processi gestiti dall'Ente. Esempio: dati anagrafici, provenienti da altri Enti, indispensabili alla gestione dei procedimenti amministrativi;
- **Dati prodotti come risultato dell'attività.** Sono i dati che l'Amministrazione produce nell'esercizio delle sue funzioni, cioè i dati in uscita (*output*) ottenuti come risultato finale dei processi e dei procedimenti gestiti dall'Ente. 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 ai 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.** Sono tutti quei dati che l'Amministrazione ottiene come sottoprodotto (*by-product*) di un processo. Esempio: 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), 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).

Successivamente, è auspicabile stabilire **quali sono i dati di interesse per la Comunità**. Secondo il vademecum Open Data del Fornez, 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 aprire⁴⁵. 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. È proprio questo uno degli elementi del valore aggiunto di una strategia di Open Data.

Una recente analisi dell'Associazione Italiana per l'Open Government, inoltre, ha identificato tra gli interessi dei cittadini le seguenti priorità:

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

Analisi dei dati

Una volta realizzata la mappatura dei dati presenti nell'Amministrazione e definite alcune priorità in base agli interessi della Comunità di riferimento, è necessario passare a una fase di

⁴⁵ Così come ha fatto di recente, per esempio, il Comune di Bologna, attraverso la pubblicazione del dataset denominato "Open Data Index" (ODI), dove sono elencati a loro volta diversi dataset in via di apertura e per i quali si chiede ai cittadini/utenti di segnalare le effettive utilità.

analisi dei dati identificati, per valutare la fattibilità della loro apertura e le eventuali modalità per attuarla.

Per fare questo, prima di tutto bisogna definire **chi detiene i dati e quali diritti ha sugli stessi**. Le PA trattano numerosi dati e informazioni che possono essere custodite nei propri archivi cartacei e digitali. Non tutti questi dati, però, 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 del CAD detta un'importante 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 rimane 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 logistica 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 da considerare è quella stabilita dall'art. 11 Legge n.633/1941 che prevede che *"alle amministrazioni dello Stato, alle Provincie e ai Comuni spetta il diritto di autore sulle opere create e pubblicate sotto il loro nome e 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 a un altro soggetto.

In tutti gli altri casi, qualora l'Ente 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.

Successivamente, deve essere valutata la **qualità dei dati** per definire l'opportunità della loro apertura verso l'utenza. Per farlo va presa 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:

- Accuratezza sintattica, vale a dire la vicinanza del valore del dato a un valore nel dominio di definizione sintattico considerato corretto. Esempio: un basso grado di accuratezza sintattica corrisponde a errori in un database anagrafico nel quale "Carlo" è riportato come "Crlo";
- Accuratezza semantica, cioè la vicinanza del valore del dato a un valore nel dominio di definizione semantico considerato corretto. Esempio: un basso grado di accuratezza semantica corrisponde al caso in cui "Carlo", in un database anagrafico, diventi "Carla".

- Attualità, ossia l'adeguatezza del valore del dato rispetto alle esigenze temporali del contesto d'utilizzo;
- Completezza, cioè l'estensione con cui i valori dei dati coprono l'universo cui sono riferiti. Esempio: un basso grado di completezza corrisponde al caso in cui nel dataset contenente gli indirizzi delle scuole di un territorio sia riportata una bassa percentuale delle scuole effettivamente presenti in quel territorio;
- Consistenza interna, che indica il grado di coerenza dei dati presenti in un dataset riferiti a una stessa entità. Esempio: un basso grado di consistenza interna corrisponde a errori in un database anagrafico in cui "Carlo", residente a "Milano", risulta in provincia di "Torino";
- Consistenza esterna, che indica il grado di coerenza tra dati diversi ma correlati presenti tra elementi di un dataset. Esempio: un basso grado di consistenza esterna corrisponde a errori in un database anagrafico in cui "Carlo" risulta sposato con "Maria" mentre "Maria" risulta "Nubile".

La presenza di un alto numero di errori in alcuni elementi tra quelli qui indicati può scongiurare 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 può 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).

Fatte queste operazioni, è necessario effettuare una **analisi del formato nel quale i dati sono disponibili**. Riassumendo e semplificando quanto già riportato nel Capitolo 1 in merito ai tipi di dati aperti, ecco le situazioni di fronte a cui è possibile trovarsi:

- Dati grezzi ("raw data", 1 stella): sono dati disponibili in quei formati che – seppure disponibili su supporto informatico – non consentono un'estrapolazione immediata degli stessi. Esempio: immagini nei diversi formati grafici bitmap (es. .gif, .jpg, .bmp), documenti in formato .pdf o Microsoft Word;
- Dati strutturati (da 2 a 3 stelle): sono dati disponibili in formati – aperti o proprietari - che ne consentono l'elaborazione e la sistematizzazione in forma strutturata. Esempio: documenti realizzati con fogli di calcolo come Microsoft Excel od OpenOfficeCalc e salvati in formati come XLS, CSV etc.;
- Dati presenti in database: sono dati presenti all'interno dei database gestiti dall'Ente. Di norma è possibile scegliere il formato di esportazione più idoneo (es. xml o altri formati aperti). L'esportazione può produrre file classificabili con 4 stelle o Linked Open Data (5 stelle).

Giova ribadire che la logica dell'Open Data spinge la PA 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 (il cosiddetto principio del "Raw Data Now!"). Anche in questo caso sarà la comunità di sviluppatori a occuparsi della conversione del dato in un formato aperto, e quindi utilizzabile.

Pubblicazione dei dataset

Pubblicare dati in un formato aperto vuol dire – secondo la classificazione del W3C (cfr. Capitolo 1) – considerare i formati che vengono classificati con tre o più stelle. Più in particolare:

- in caso di **dati grezzi** (1 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 OCR⁴⁶). Tuttavia c'è da considerare 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 (2 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 (3 stelle). Come abbiamo visto in precedenza, la maggior parte dei dati aperti dalla PA è classificabile con 3 stelle;
- i dataset classificati con 4 o 5 stelle richiedono la costruzione di dataset basati su **Linked Open Data (LOD)** e quindi competenze tecniche specifiche (cfr. avanti). Di norma, comunque, i dataset in formato LOD sono prodotti da database relazionali o software specifico.

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 descrittivi**, intendendo con tale definizione l'informazione che descrive un insieme di dati. Nel caso specifico dei metadati descrittivi, essi costituiscono un livello di

⁴⁶Optical Character Recognition.

informazioni descrittive che corredano il dataset, ne esplicitano alcune caratteristiche e ne rendono più semplice l'identificazione⁴⁷.

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 etc.

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 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 CMS (*Content Management System*) adottato per la gestione del proprio portale. In tal caso si procederà in modalità guidata, inserendo il dataset come allegato a 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 API⁴⁸ 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 dataset e, soprattutto, dei sistemi software a disposizione della PA.

Prima di procedere alla pubblicazione definitiva, è necessario per le Amministrazioni specificare quali sono le **licenze** con cui vengono pubblicati i dati resi disponibili sul web. 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:

⁴⁷ Si può trovare un ottimo esempio di efficacia dei metadati descrittivi nei dataset aperti di Istat.it.

⁴⁸ *Application Programming Interface*: set di istruzioni standard attraverso le quali si può interagire con un software in modo da gestirne funzioni o elaborare dati. In ambito Open Data, le API maggiormente utilizzate riguardano i geodati e la geolocalizzazione di dati e informazioni su mappe (per esempio, su Google Map attraverso le Google API).

- licenze di tipo *closed* (cioè chiuse): 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;
- licenze di tipo *open* (ovvero aperte): le licenze open tendono a garantire una serie di diritti a chi entra in possesso delle informazioni. Le Amministrazioni devono scegliere e utilizzare una delle licenze di tipo aperto, optando – preferibilmente – per una di quelle già sviluppate e disponibili, come per esempio la già citata 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 dell’informazione del settore pubblico.

Diffusione dei dataset

Dopo aver pubblicato i dataset aperti, è naturalmente opportuno divulgarli verso altre Amministrazioni, cittadini e comunità di sviluppatori

Per permettere un’adeguata diffusione dei dati aperti, come abbiamo visto, a livello centrale è stato sviluppato il Portale nazionale dei dati aperti “dati.gov.it”, verso il quale far confluire i riferimenti dei dati aperti da ogni Amministrazione (URL emetadati): un’apposita sezione sul portale consente infatti di segnalare l’avvenuta pubblicazione di dati in formato aperto. Abbiamo anche visto come alcune Regioni dispongano o stiano sviluppando portali regionali Open Data. 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.

5.2 Ruoli e competenze tecniche per gli Open Data nella Pubblica Amministrazione

Per la PA l'ingresso nel mondo digitale ha comportato un processo lungo e di certo non privo di ostacoli e problemi⁴⁹. I primi sono arrivati con la formazione digitale dei dipendenti, che ha occupato molto tempo. Se, da una parte, le norme e i decreti obbligavano gli enti a digitalizzarsi, dall'altra solo con molta fatica si riuscivano a trovare persone – e soprattutto mezzi – per operare una vera e propria “rivoluzione digitale”. Occorre anche sottolineare che le norme e i decreti, in realtà, non hanno mai puntato in maniera decisa sulla sanzionabilità in caso mancata attuazione degli step digitali, e questo ha contribuito a rallentare il passaggio dalla carta al digitale. Oggi, invece, con l'avvio dell'Agenda Digitale e lo sviluppo sempre più rapido dei processi di apertura dei dati pubblici, occorre scegliere con cura chi si deve occupare di tali materie, per non far implodere i servizi ICT interni gravandoli di ulteriori servizi da gestire oltre a quelli strategici – come *cloudcomputing*, *disasterrecovery*, servizi on line e gestione delle reti – che già competono loro.

Abbiamo visto, nel paragrafo precedente, quali sono gli step necessari per “aprire”, pubblicare e divulgare i dati pubblici: tali indicazioni, ovviamente, si devono intendere anche quali requisiti di conoscenza fondamentali per quei soggetti – in particolare gli Ingegneri Informatici e Gestionali – che volessero partecipare alla sfida degli Open Data nel settore pubblico.

Sempre in relazione agli Open Government Data, esaminiamo adesso ulteriori competenze e specializzazioni che possono riguardare profili tecnici codificati a livello internazionale ma ancora poco presenti nella realtà della Pubblica Amministrazione italiana, utili a quella fase di “apertura dei dati” che rappresenta il punto di partenza dello sviluppo degli Open Data nelle amministrazioni pubbliche⁵⁰.

A tale riguardo, l'International Webmaster Association ha elaborato diversi *skillprofiles*⁵¹ che prevedono competenze, sia di base che avanzate, per lavorare nella PA digitale e aperta che ci aspetta. Quelli che qui ci preme esaminare sono i seguenti:

- **Data Manager:** il gestore dei dataset che, con gli Open Data, potrebbe diventare il vero coordinatore dell'attività di trasparenza e divulgazione di dati aperti delle PA;
- **Database Administrator:** si occupa di installare, configurare e gestire i sistemi di archiviazione dei dati e i sistemi Client/Server in rete. Definisce gli accessi al database e ne controlla la sicurezza, per preservare l'integrità dei dati contenuti nell'archivio. Cura la manutenzione del server e collabora alla definizione delle strategie dell'ente per lo sviluppo e l'utilizzo delle risorse che costituiscono il database. Ha conoscenza dell'architettura di un sistema operativo in una rete locale, dei concetti di sicurezza e implementazione di una

⁴⁹ Gran parte delle informazioni riportate in questo paragrafo sono desunte da “Piccola guida alle professioni del web nella PA”, a cura di Francesca Sensini, Quaderni ForumPA, luglio 2012.

⁵⁰ Per un approfondimento sul valore strategico delle competenze tecniche in ambito di specifiche tipologie di Open Data, si rimanda al Capitolo 6.

⁵¹ www.skillsprofiles.eu

banca dati. Il Db Administrator dovrebbe conoscere, tra le altre cose, i linguaggi XML e RDF. Di solito proviene da studi di "Ingegneria dell'Informazione";

- **Datawarehouse Developer:** si occupa di progettare e gestire particolari e complessi database, i Datawarehouse (o magazzini dei dati). Utilizza la tecnica del DataMining, che permette di analizzare a fondo grandi quantità di dati immagazzinati nei Datawarehouse, per scoprire le relazioni significative. Mentre il Db Administrator si limita a fornire i dati aggregati nei modi più diversi, il Datawarehouse Developer è in grado di rispondere a domande più complesse. Ha capacità manageriali, conoscenza dei sistemi informativi e competenza in materia di privacy dei dati trattati.

In particolare, risulta interessante approfondire la figura del Data Manager.

5.2.1 Il Data Manager

Come abbiamo visto, una Pubblica Amministrazione che si affaccia agli Open Data si troverà costretta ad affrontare scelte che riguardano i formati dei dataset, il software e le licenze di utilizzo. Tuttavia, prima di affrontare le scelte tecniche come il formato dei dati, è fondamentale, per un'amministrazione, non tralasciare gli **Open Services**, cioè quei servizi basati sui dati pubblici aperti che abilitano l'interazione (o addirittura la "transazione") con utenti od organizzazioni esterne all'amministrazione, per esempio tutto ciò che afferisce alla sfera di applicazione di metodi, strumenti e reti per le **Smart Cities**.

Da tutto ciò si può cogliere come ci si propongano orizzonti sempre più vasti e prospettive davvero interessanti che necessitano, però, di competenze specifiche, di un "operatore della tecnologia" dentro la PA che sappia riconoscere il valore aggiunto di un dato liberato in un certo formato rispetto a un altro, e che abbia la consapevolezza dell'importanza di utilizzare particolari standard tecnici, per altro alcuni già riconosciuti anche all'interno della vasta normativa legata ai siti web PA e alla trasparenza.

Questo operatore è appunto il Data Manager. Le sue competenze, però, non si fermano a quelle tecniche che gli *skills profiles* internazionali attribuiscono al DataBase Administrator o al Data Analyst. Per scegliere il prodotto giusto per il proprio ente, con lungimiranza, e favorire sempre e comunque la trasparenza potrebbero essere necessarie, infatti, ulteriori capacità.

Un Data Manager, infatti, dovrebbe avere un quadro molto più completo del mondo della gestione del dato, tra cui licenze, diritti e doveri, qualità e formati.

All'interno della PA operano già i responsabili dei sistemi informativi e i responsabili dei dati. Prima di tutto, dunque, serve avere una visione d'insieme dei vari attori coinvolti, quindi capire come potenziare le figure in essere, ma soprattutto abilitare il processo per la gestione dei nuovi flussi operativi. Se poi prendiamo in considerazione il *web of data*, ovvero i Linked Data, le cose cambiano ancora. Si pone la necessità sempre più impellente di dati di qualità nati da una

gestione interna avanzata, e di processi di qualità dovuti a una reale consapevolezza del quadro di insieme che si abilita. Si tratta di problematiche nuove, che necessitano di formazione e di una gestione ben precisa.

5.2.2 Divulgare gli Open Data: Infografica, Data Journalism, applicazioni per smartphone

Il concetto di Open Data si accompagna sempre di più a quello di divulgazione dei dati aperti. Diventa dunque importante, per una PA che voglia rendere effettivo il principio di trasparenza, dotarsi di ulteriori competenze in grado non solo di aprire e pubblicare i dataset nei cataloghi ufficiali a disposizione, ma anche di “visualizzarli” in formati grafici (e interattivi), interpretarli/raccontarli trasformarli in applicazioni per device mobili, in modo da descrivere in sintesi i dataset più rilevanti – soprattutto quelli che riguardano una vasta porzione di cittadinanza – e da rendere i dati fruibili in tempo reale e utili per la collettività.

Sebbene spesso non correlate con competenze tipicamente ingegneristiche, è il caso di focalizzare quindi l'attenzione su ulteriori profili tecnici in grado di aiutare le PA nei processi di apertura dei dati:

- **Esperto infografiche:** il perfetto esperto di infografiche dovrebbe saper ideare, creare ed elaborare disegni e illustrazioni elettroniche, sia statiche che dinamico-multimediali (2D e 3D), per prodotti cartacei e web, trasformare la grafica statica in progetti interattivi, selezionare e lavorare immagini utilizzando quali fonti di consultazione le principali banche dati, acquisire e trattare file video e audio, realizzare illustrazioni attraverso l'uso di programmi vettoriali, conoscere il pacchetto Adobe (Photoshop, Illustrator, Flash, InDesign, Premiere) e i linguaggi Html e Javascript, nonché possedere nozioni base di lavorazione per nuovi media (es. iPad, Mobile etc.)⁵².
- **Data Journalist:** pur non possedendo tutte le conoscenze descritte in relazione all'esperto di infografiche, un esperto di comunicazione che lavora presso un Ufficio Stampa di un ente pubblico, oppure anche un ingegnere con la passione per la comunicazione e la divulgazione scientifica, possono diventare esperti di “Data Communication”, ossia interpretare, elaborare e presentare i dati delle PA in forme e modi semplificati, anche avvalendosi dei numerosi software e programmi (perlopiù reperibili on line) in grado di trasformare i dati in informazioni grafiche, come per esempio Tableau Public, ManyEyes, Google Fusion Tables etc.
- **Mobile Application Developer:** potrebbe inoltre essere utile, nel processo di liberazione e divulgazione dei dati pubblici, una figura professionale in grado di realizzare/codificare soluzioni applicative per periferiche mobili a partire dai dataset di riferimento. Un buon

⁵² Al link web <http://datavisualization.ch/tools> è possibile consultare una nutrita selezione di strumenti e applicazioni utili per realizzare infografiche a partire dai dati.

Mobile Application Developer dovrebbe conoscere le metriche connesse allo sviluppo delle applicazioni mobili, i metodi per lo sviluppo delle applicazioni mobili e la loro logica (prototipazione, metodi agili, *reverse engineering* etc.), i prototipi per supportare la validazione dei requisiti di sviluppo, le modalità di progettazione e realizzazione di interfacce di dialogo delle periferiche mobili, uno o più linguaggi di programmazione per ambienti mobili (ObjectiveC, Java, etc.), le modalità di pubblicazione delle applicazioni mobili. Tutto ciò dovrebbe inoltre essere accompagnato da una competenza nell'interazione con applicazioni di database e nella portabilità dell'applicazione mobile su diverse piattaforme (es. iOS). Un buon sviluppatore di Apps avrebbe inoltre con sé la qualifica "MDICD – Mobile Development Institute Certified Developer".

SCHEDA

Alcuni profili di ingegneri "Open Data" ricercati all'estero

Facendo un giro per il mondo sul web, capita di imbattersi in alcuni casi assai significativi, nonché estremamente illuminanti, di ricerca e selezione di personale tecnico per Open Data.

Per esempio, l'associazione onlus *Open Data For Africa* – al cui sito opendataforafrica.org vale la pena dare uno sguardo per la grande attenzione al tema e per la ricchezza dei progetti che porta avanti – cerca un *Web Development Engineer*, un *Mobile Development Engineer*, un *Software Development Engineer* e un *Data Engineer*. Tutte figure professionali, queste, strettamente attinenti ai profili di Ingegneria Informatica, Ingegneria delle Telecomunicazioni, Ingegneria dell'Informazione.

La società di consulenza statunitense di analisi statistiche *Kaggle* (www.kaggle.com), invece, va ancora più avanti, spingendosi a selezionare un "Linked Data Engineer", livello "Senior", in grado di portare avanti i progetti di web semantico e LOD Cloud della compagnia. Per far questo, l'ingegnere dei Linked Data deve possedere le giuste competenze in Java, ontologia web, standard W3C ma soprattutto in tecnologie di web semantico quali RDF, RDFS, OWL, SPARQL.

Come possiamo vedere da questi due rapidi esempi, l'attenzione verso le competenze e i profili tecnici in ambito Open Data va crescendo a livello internazionale, anche se in Italia, ad oggi, probabilmente non essendo stata condotta alcuna riflessione specifica riguardo a tali fabbisogni professionali da/nella PA ed enti privati, manca ancora un mercato in tal senso. Eppure, per dirla con Jonathan Grey, il Presidente di Open Knowledge Foundation, "*i dataset non cambiano il mondo, perché il mondo lo cambiano le persone*". Sulle persone, e quindi con particolare riguardo verso i profili tecnici e ingegneristici, si dovrà investire al più presto, per non rischiare di attardarsi laddove altri, in USA come in Kenya, stanno già esplorando nuovi mondi.

6. IL VALORE STRATEGICO DEI DATI APERTI NELLO SVILUPPO DI NUOVE APPLICAZIONI E NUOVI MODELLI DI BUSINESS IN CAMPO TECNICO

Sin dall'inizio del movimento Open Government, l'utilizzo dei dati esposti dalle PA è statopromosso e stimolato anche attraverso l'organizzazione di competizioni e "civichackathons", eventiche riuniscono aziende IT, programmatori e progettisti nel campo dello sviluppo di software chiamati asfidarsi per realizzare le applicazioni più interessanti.

Si possono poi innescare meccanismi per cui icittadini stessi possono proporre "Apps", o elaborazioni e interpretazioni originali dei dati,potenzialmente di grande interesse e rilevanza per la società.

Si possono, inoltre, implementare nuovi servizi in grado di agevolare il rapporto tra il settore pubblico e il settore privato, nonché servizi in grado di connettere realtà ed esperienze diverse con l'obiettivo di migliorare le condizioni di vita dei cittadini – come per esempio sta accadendo nello sviluppo delle idee di *Smart Cities* e *Open Services*.

Tutto questo è possibile proprio a partire da una nuova considerazione dell'importanza strategica che assumono oggi i dati aperti, e in particolare i Linked Open Data.

6.1 Tipologie di dati: valori strategici e ambiti specifici di applicazione

Le Linee Guida per la produzione di Open Data interoperabili, pubblicate nel novembre 2012 dal Sistema Pubblico di Connettività e Cooperazione nell'ambito delle attività indicate dall'Agenda Digitale Europea⁵³, fornisce una serie di informazioni e strumenti per far sì che i dati della PA italiana entrino a far parte del cosiddetto *Web of Data*(Web dei dati), attraverso appunto la liberazione di Linked Open Data (5 stelle secondo la classificazione W3C) in grado di far diventare il web uno "spazio dati distribuito" in cui molteplici informazioni sono connesse fra di loro tramite collegamenti semantici, così dafacilitare lo sviluppo di applicazioni innovative e il supporto a interrogazioni che coinvolgono unamolteplicità di basi di dati distribuite.

Il gruppo di lavoro del Sistema Pubblico di Connettività e Cooperazione ha individuato alcune tipologie di dati pubblici con cui le PA possono rispondere efficacemente airequisiti di trasparenza, responsabilità e di sviluppo di applicazioni e servizi. Tali tipologie sono strettamente correlate con le funzioni e le competenze tecniche dei soggetti chiamati a rendere i dati aperti e interoperabili tra loro – in modo particolare gli Ingegneri Informatici, Civili, delle Telecomunicazioni, dell'Ambiente e Territorio, Gestionali – e pertanto risulta estremamente utile,

⁵³ Commissione di Coordinamento SPC, "Linee Guida per l'Interoperabilità Semantica attraverso i Linked Open Data", Agenzia per l'Italia Digitale, ver. 2.0, 12 novembre 2012. Da questo documento è ricavata gran parte delle informazioni riportate nel presente Capitolo.

ai nostri fini, effettuare una disamina dettagliata, in particolare per quanto riguarda i dati territoriali (geodati), i dati ambientali, i dati della ricerca e i dati statistici.

6.1.1 Dati territoriali (geodati)

I dati territoriali sono fondamentali in tutte le attività di pianificazione e gestione del territorio, nonché di coordinamento. Esempi di dati territoriali includono: le cartografie tematiche (geologica, idrogeologica, geomorfologica), i limiti amministrativi (regionali, provinciali, comunali, municipali, unità urbanistiche, sezioni di censimento), le carte delle aree soggette a vincoli o carte del rischio (aree d'interfaccia boschiva, aree percorse dal fuoco, aree a rischio incendio, aree esondabili e allagate, zone a rischio frane), lavori pubblici, barriere architettoniche⁵⁴. I dati territoriali, oltre a consentire lo sviluppo di nuovi servizi, concorrono a formare decisioni utili per lo sviluppo di quasi tutte le attività economiche. Per quanto riguarda la condivisione e l'accesso ai dati territoriali, inoltre, risultano importanti gli aspetti legati alla trasparenza, alla partecipazione e alla democrazia che derivano dalla possibilità, da parte dei cittadini, di conoscere gli elementi e gli aspetti fondamentali del proprio territorio e di divenire soggetti attivi nelle politiche e nei processi decisionali e di pianificazione urbanistica e territoriale.

Se negli ultimi anni le PA hanno considerevolmente sviluppato i sistemi per la gestione dei propri dati territoriali, la loro azione è oggi maggiormente orientata a rendere tale patrimonio informativo accessibile e disponibile in modo generalizzato e interoperabile. Non mancano tuttavia le difficoltà: la complessità dei dati territoriali, gli elevati costi associati ai processi di acquisizione dei dati stessi e alla realizzazione di sistemi evoluti di fruizione costituiscono sicuramente un fattore critico che rallenta l'apertura. Inoltre, i dati territoriali, anche simili ma trattati da amministrazioni diverse, per la realizzazione, ad esempio, di carte tecniche, piani regolatori, carte di vincolo, sono spesso disomogenei per contenuto, precisione, formati e documentazione. I potenziali utenti, infine, quali cittadini, piccole imprese, professionisti e anche enti locali di piccole dimensioni, spesso non sanno bene a chi rivolgersi per ottenerli o come interpretare i dati ricevuti.

C'è da dire che le problematiche italiane, riconducibili a mancanza di disponibilità, qualità, organizzazione, accessibilità e condivisione dei dati territoriali⁵⁵, sono comuni anche agli altri Paesi UE, tanto che nel 2007 è stata emanata la citata Direttiva INSPIRE per la creazione

⁵⁴Un elenco delle tipologie di dati territoriali è fornito nell'allegato 1 del DM 10 novembre 2011 relativo alle regole tecniche del Repertorio Nazionale dei Dati Territoriali (RNDT).

⁵⁵Per ovviare alla situazione, in Italia, il recente decreto sulla "spending review" n 95, convertito in legge n. 135 del 7 agosto del 2012, ha regolamentato la fruibilità di dati geospaziali acquisiti con risorse pubbliche. Il Codice dell'Amministrazione Digitale, inoltre, ha istituito il Repertorio Nazionale dei Dati Territoriali (RNDT) allo scopo di "agevolare la pubblicità dei dati di interesse generale, disponibili presso le pubbliche amministrazioni a livello nazionale, regionale e locale". Il Repertorio rappresenta il catalogo nazionale dei metadati su dati territoriali e relativi servizi.

di un'infrastruttura per l'informazione territoriale finalizzata a garantire l'interoperabilità attraverso l'armonizzazione, l'accesso e il riuso dei dati prodotti e/o gestiti dalle PA.

Un apposito Regolamento, emanato in applicazione della Direttiva, fornisce le indicazioni per l'interoperabilità degli insiemi di dati e dei servizi. Tale Regolamento indica che:

- per garantire l'interoperabilità e l'armonizzazione tra le categorie tematiche di dati territoriali, è opportuno rispettare i requisiti in materia di tipi di dati comuni, individuazione degli oggetti territoriali, metadati per l'interoperabilità, modello generico di rete;
- al fine di garantire l'interoperabilità e l'armonizzazione all'interno di una categoria tematica di dati territoriali, è opportuno utilizzare le classificazioni e le definizioni degli oggetti territoriali, i relativi attributi chiave e relazioni, i tipi di dati, gli ambiti dei valori e le norme specifiche che si applicano alla categoria tematica di dati territoriali interessata⁵⁶.

Garantire l'interoperabilità di dati territoriali e dei relativi servizi implica anche incrementarne e migliorarne l'utilizzo, soprattutto nella prospettiva di renderli aperti. In linea di massima, infatti, tutti i dati territoriali possono essere resi aperti, ma, per aumentarne il valore aggiunto, sempre in termini di opportunità di riutilizzo e di integrazione nelle infrastrutture e/o sistemi di gestione, la traduzione anche in Linked Data implica una valutazione e una selezione dei dati in funzione delle informazioni, anche non esplicitamente territoriali, che a essi possono essere relazionate. Infatti il dato territoriale, per sua natura, si presta a essere utilizzato per "collegare", secondo il paradigma dei Linked Data, informazioni provenienti da fonti diverse, arricchendole, rendendole interoperabili e più facilmente fruibili nel contesto del Web Semantico, anche in sistemi che prevedono comunicazioni tra macchine.

Attualmente, nell'ambito dei dati territoriali il formato più diffuso è lo "shapefile" che, pur essendo nato come formato proprietario, è diventato di fatto uno standard, essendo ormai utilizzato in modo aperto. Anche altri formati come KML, GML, GeoRSS, GeoJSON, e GeoTIFF sono utilizzati, aumentando il ventaglio di soluzioni applicative possibili che ne fanno uso. Sono ancora poco numerose le iniziative mirate all'apertura in formato "Linked" di tale tipologia di dati. Alcuni esempi di rilievo si registrano sempre in campo inglese⁵⁷, mentre in Italia sono state da poco avviate iniziative in tale contesto. In particolare, in Emilia-Romagna sono stati creati componenti software e ontologie che rendono disponibili come Linked Open Data dati e metadati gestiti dalla propria infrastruttura di dati territoriali tramite un'applicazione riusabile.

⁵⁶ Inoltre, per ciascuna categoria tematica è stata elaborata una specifica dei dati coerente con la struttura definita nello Standard ISO 19131 *"Geographic Information – Data Product Specifications"*. Le specifiche comprendono la documentazione tecnica dello schema di applicazione, i tipi di oggetti territoriali con le loro proprietà e altre indicazioni sulle categorie tematiche, utilizzando sia il linguaggio naturale sia il linguaggio formale dello schema concettuale.

⁵⁷ Goodwin, C. Dolbear, G. Hart, *"Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web"*, 2008; Goodwin, C. Dolbear, G. Hart, *"Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web"*, 2008.

SCHEDA

Dati territoriali come Linked Open Data: GetLOD e il modello Emilia-Romagna⁵⁸

Un ruolo rilevante nella Smart City è rivestito dalle informazioni geografiche e dai sistemi a supporto dell'integrazione di fonti dati esterne e interne alla PA con il territorio. I dati geografici pubblicati dagli enti locali sono gli Open Data più utilizzati dalla rete, perché permettono la geolocalizzazione di altre informazioni pubbliche, così come di dati provenienti da banche dati private. Per contribuire alla realizzazione del modello di Smart City, la Regione Emilia-Romagna, in collaborazione con Sinergis, società di Dedagroup ICT Network, e Planetek Italia, ha sviluppato **GetLOD**: una soluzione innovativa che consente, in maniera automatica, di rendere disponibili, come Open Data e come Linked Open Data, dati e metadati gestiti dall'infrastruttura di dati territoriali, facilitandone così l'aggregazione con altri dati, l'elaborazione e l'analisi.

Il modello di Pubblica Amministrazione digitale che si sta consolidando a livello internazionale è quello che fa leva sulla standardizzazione di Open Services e Open Data come mezzo per l'integrazione fra le informazioni delle PA e dei privati e quindi sull'accesso a dati certificati sempre aggiornati e affidabili sui quali è possibile realizzare applicazioni innovative per la Smart City. L'architettura GetLOD permette alle pubbliche amministrazioni, anche in un contesto di risorse scarse, di compiere i primi passi verso gli Open Services quale fattore abilitante di una città intelligente che utilizza in modo più efficiente le proprie risorse, a cominciare dai dati geografici, ossatura portante della Smart City.

L'iniziativa rientra nell'ambito di un progetto di sviluppo evolutivo dell'Infrastruttura Geografica del sistema degli enti del territorio emiliano-romagnolo, ed è progettata per essere portabile e poter essere calata in realtà differenti, in quanto interoperava con l'Infrastruttura Geografica attraverso servizi standard (OGS WFS, OGC CSW) ed è perfettamente integrata con il portale Open Data Emilia-Romagna.

GetLOD nasce infatti dall'esigenza di massimizzare il valore dei dati geografici che, resi accessibili in maniera conforme ai formati standard di riferimento per gli Open Data, anche se pubblicati in modo indipendente da diversi soggetti, possono essere incrociati liberamente da terze parti, pubbliche e private. Qualsiasi applicazione informatica, grazie al formato aperto e standardizzato, può così accedere a questi dati e riutilizzarli in modo creativo sfruttando, per i propri obiettivi, le relazioni (link) fra informazioni eterogenee.

Un vantaggio concreto che può consentire a un ente, ad esempio, di mappare sul territorio le proprie banche dati anagrafiche, a un'azienda che distribuisce il gas di correlare i propri contatori ai fabbricati serviti, a un'azienda di trasporti di correlare i percorsi dei propri mezzi al

⁵⁸ <http://saperi.forumpa.it/story/69629/dati-territoriali-come-linked-open-data-il-modello-emilia-romagna>. Si veda anche: <http://blog.planetek.it/2012/06/20/verso-i-linked-open-data-geografici>.

grafo delle strade regionale. Allo stesso tempo, la soluzione si presta a supportare lo sviluppo commerciale di un'azienda, permettendo di mappare sul territorio i propri clienti e ottimizzare l'analisi dei bacini di vendita.

GetLOD garantisce la pubblicazione di dati geospaziali partendo dai tradizionali webservice cartografici, come dati aperti e linkabili. Rendendo fruibili le informazioni geografiche sotto forma di Open Data, con particolare attenzione per il formato RDF/XML, assicura l'indicizzazione dei dati sui motori di ricerca di Open Data e l'integrazione con portali di dati aperti o con il Comprehensive Knowledge Archive Network (CKAN), il catalogo di dataset e progetti liberi.

Si tratta di una soluzione "open" e riusabile, che si integra in modalità standard a Geoportali cartografici, Portali Open Data e Infrastrutture di Dati Territoriali basate sugli standard per l'interoperabilità definiti dall'Open Geospatial Consortium (OGC), e consente di pubblicare i dati geografici aperti sia in formato RDF (Linked Open data per l'appunto) sia in altri formati di interscambio non linkabili (ad esempio shapefile e GML).

Geoportali

Con il termine geoportale si intende un sito Web realizzato in modo tale da costituire un punto di accesso unico (gateway) ai servizi relativi a dati e risorse spaziali, che non devono necessariamente risiedere all'interno del sito stesso ma che possono invece essere distribuiti. Soggetti, siano essi pubblici, privati o generiche comunità di utenti che realizzano un geoportale, permettono l'accesso a informazioni territoriali per mezzo di un'interfaccia web e attraverso l'utilizzo di *web services*. Dall'altro lato, gli utenti Web che hanno la necessità di utilizzare informazioni territoriali, siano essi utenti generici, professionisti o PA, utilizzano i geoportali per ricercare, accedere e utilizzare i dati.

I geoportali possono rappresentare inoltre un elemento fondamentale delle infrastrutture di dati spaziali (SDI - Spatial Data Infrastructure), di cui l'esempio più significativo a livello nazionale e europeo è la SDI definita e regolata dalla direttiva INSPIRE (Direttiva 2007/2/CE). Per garantire la ricerca e l'accesso alle informazioni territoriali, un geoportale offre all'utente una serie di servizi che, pur essendo spesso fruibili tramite interfaccia grafica per mezzo di un comune browser Web, devono essere implementati, all'interno di un'architettura di tipo Service Oriented (SOA), come veri e propri web-services realizzati in modo conforme a una serie di standard che ne garantiscono l'interoperabilità.

Per quanto riguarda il panorama *open source*, le soluzioni software più utilizzate per la realizzazione di geoportali, sia in ambito pubblico che privato, possono essere suddivise in tre categorie principali:

- applicazioni server che implementano e consentono la pubblicazione di web services per la visualizzazione e il download di dati territoriali;
- applicazioni server che implementano e consentono la pubblicazione di web services per la gestione dei metadati relativi ai dati territoriali;
- librerie software API per la fruizione tramite browser web di dati territoriali.

Un'ulteriore modalità di pubblicazione dei dati è quella dell'"*opendata-as-a-service*" dove anziché mettere a punto una piattaforma *in-house* si sfruttano piattaforme di *cloud* (pubblico o privato).

GIS

Un *Geographic Information System* (GIS) è un sistema adatto per catturare, immagazzinare, manipolare, analizzare, gestire e rappresentare tutti i tipi di dati geografici. L'acronimo GIS è spesso usato per significare la scienza o gli studi sulle informazioni geografiche (dette anche geospaziali). In termini semplici, con il GIS si possono unire cartografie, eseguire analisi statistiche e gestire i dati attraverso tecnologie database. La traduzione in italiano di GIS richiama il Sistema informativo territoriale e viene spesso usata erroneamente come sinonimo, ma per correttezza con GIS si individuano solo i software che sono sviluppati con questa logica.

Di seguito, a titolo di esempio, si offre un elenco di risorse web che offrono GIS liberi ("open"):

- Geodati Regioni: una tabella riassuntiva dei geodati resi disponibili dalle Regioni italiane attraverso i siti web. L'indagine di base è completa, mancano solo alcune informazioni sulle licenze (in ogni caso proprietarie);
- Confini amministrativi (Regioni, Province, Comuni) in formato Shapefile: <http://www.istat.it/ambiente/cartografia>;
- http://www.rigacci.org/wiki/doku.php/tecnica/gps_cartografia_gis/download: una buona collezione di dati sull'Italia;
- Portale Cartografico Nazionale (PCN), mantenuto dal Ministero dell'Ambiente;
- GeoPortale dell'informazione Geografica, mantenuto dal Sistema Informativo Territoriale della Provincia di Treviso: offre multiservizi CSW, WMS, WFS, KML, GeoTIF, download di shapefile e meta-documentazione;
- Ministero dell'Ambiente: perimetrazioni in shapefile e schede pdf FTP;
- Server geodati di FEM – Fondazione Edmund Mach.

Vi sono inoltre diversi software GIS "open source", tra cui i più utilizzati sono Quantum GIS, GRASS GIS, gvSIG, Orfeo toolbox, OpenJUMP, Mapguide.

SCHEDA

Il Geo-catalogo semantico della Provincia Autonoma di Trento

Uno degli esempi più interessanti, che coniuga l'apertura di dati territoriali con la logica dei Linked Open Data è costituito dal Geo-catalogo semantico sviluppato dalla Provincia Autonoma di Trento. L'obiettivo del progetto è favorire l'interoperabilità e la collaborazione tra le strutture che si occupano di pianificazione e gestione economico-produttiva e ambientale del territorio trentino attraverso la pubblicazione ordinata e aperta di dati territoriali linkabili. Nello specifico, il progetto offre un servizio che consente di interrogare il dataset dei dati territoriali attraverso il Portale Geocartografico Trentino.

Il Geocatalogo semantico implementa il paradigma LOD ed è conforme alle indicazioni contenute nella Direttiva europea INSPIRE, e alle direttive nazionali IntesaGIS e RNDT. L'infrastruttura è in questo ambito intesa come un insieme di politiche, accordi istituzionali, tecnologie, dati e persone che rende possibile una condivisione efficiente ed efficace dei dati geo-referenziati.

I benefici attesi dal servizio sono, da un lato, il miglioramento dell'accuratezza dei risultati di ricerca delle geo-informazioni e, dall'altro, l'avanzamento verso la società della conoscenza attraverso lo sviluppo della cultura dell'uso di geo-informazioni nella vita quotidiana.

OpenStreetMap: una storia "open" di successo

OpenStreetMap (OSM) è un progetto collaborativo finalizzato a creare mappe a contenuto libero del mondo. Il progetto punta a una raccolta mondiale di dati geografici, con lo scopo principale di creare mappe e cartografie. La caratteristica fondamentale dei dati geografici presenti in OSM è che possiedono una licenza libera. È cioè possibile utilizzarli liberamente per qualsiasi scopo con il solo vincolo di citare la fonte e usare la stessa licenza per eventuali lavori derivati dai dati di OSM. Tutti possono contribuire arricchendo o correggendo i dati.

Le mappe sono create usando come riferimento i dati registrati da dispositivi GPS portatili, fotografie aeree e altre fonti libere. Sia le immagini renderizzate che i dati vettoriali, oltre che lo stesso database di geodati sono rilasciati sotto licenza Open Database License.

Di base, i dati per la realizzazione delle mappe sono ricavati da schizzi realizzati da volontari che intraprendono sistematici rilievi sul territorio muniti di unità GPS portatili capaci di registrare il percorso, insieme a smartphone, computer portatili o registratori vocali; le informazioni raccolte vengono introdotte via computer nel database, in formato vettoriale. Molti collaboratori usano programmi come GPSTabel per convertire i dati GPS dal formato grezzo o da formati proprietari al formato GPX (un'applicazione dell'XML). Più recentemente, la disponibilità di ortofoto aeree e di altri dati da enti commerciali e fonti governative ha incrementato la velocità

e l'efficacia di questo lavoro, lasciando ai volontari il tempo per l'acquisizione di ulteriori dati legati al territorio. La conversione e l'importazione dei dati da grandi basi di dati, quando queste vengono rese disponibili con licenza libera, è affidata a un team tecnico.

I rilievi sul territorio vengono effettuati da volontari (autodefinitisi “mappers”) a piedi, in bici, auto, treno o qualsiasi altro mezzo di trasporto (benché la bicicletta rimanga il mezzo di trasporto preferito dai mappers nelle aree urbane), usando un'unità GPS, unitamente in alcuni casi a computer portatili, registratori vocali digitali e fotocamere digitali per la raccolta di dati (in alcuni casi si è rivelato utile anche la raccolta di informazioni mediante domande ai passanti). Alcuni contributori concentrano la propria attività nella propria città di residenza, raccogliendo dati in maniera sistematica. Inoltre, periodicamente vengono organizzati dai volontari dei mapping party, campagne di mappatura concentrate in un luogo preciso a cui i volontari partecipano divisi in squadre; queste attività possono durare anche più di un giorno nel caso di “mappature intensive”.

OpenStreetMap è stato ispirato da siti come Wikipedia: la pagina in cui la mappa è consultabile espone in evidenza un'etichetta “modifica” per procedere con la modifica dei dati. Il progetto è accompagnato da un archivio storico delle modifiche (cronologia e log). Gli utenti registrati possono caricare nei database del progetto tracce GPS e modificare i dati vettoriali usando gli editor forniti.

Alcune agenzie di governo hanno rilasciato i propri dati ufficiali con licenze appropriate per l'importazione in OpenStreetMap. Varie altre autorità locali hanno donato i propri dati fotografici per mezzo di licenze compatibili attraverso OpenAerialMap.

La più grande importazione di dati in tal senso è stata effettuata negli Stati Uniti d'America, dove per legge tali dati non possono essere coperti da diritto d'autore.

In Italia, Istat mette a disposizione le coordinate dei centri abitati e i confini di Regioni, Province e Comuni. I comuni di Merano, Schio, Vicenza e Storo hanno assunto atti amministrativi formali per dichiarare liberi i dati geografici in loro possesso. Nel 2010 il Ministero dell'Ambiente e della Tutela del Territorio e del Mare ha autorizzato l'uso delle fotografie aeree ortorettificate, disponibili sul Portale Cartografico Nazionale, per ricalcarle e ricavarne dati vettoriali da inserire in OpenStreetMap. Molti dei dati rilasciati sui portali Open Data regionali (in particolare: dati.piemonte.it, dati.emilia-romagna.it e dati.veneto.it) con licenze CC0 o CC-BY o IODL, sono stati inseriti sotto l'occhio vigile della comunità.

OpenStreetMap è stato fondato nel luglio 2004 da Steve Coast, e nel 2006 è diventata una Fondazione “non a scopo di lucro, dedita a incoraggiare la crescita, lo sviluppo e la distribuzione di dati geospaziali liberi e a fornire dati geospaziali per l'utilizzo e la condivisione con tutti”. Nel luglio 2007, quando la Fondazione OpenStreetMap ha organizzato la prima conferenza internazionale di OSM (nominata “The State of the Map”), il progetto contava 9.000 utenti registrati. Tra gli sponsor dell'evento figuravano Google, Yahoo! e Multimap. Nel

dicembre dello stesso anno l'Università di Oxford è stata la prima importante organizzazione a utilizzare i dati di OpenStreetMap sul proprio sito web. Nell'agosto 2008, poco dopo la seconda conferenza internazionale "The State of the Map", gli utenti registrati sono saliti a 50.000, con oltre 5.000 contributori attivi. Nel marzo 2009 sono stati superati i 100.000 utenti. Il 5 e 6 giugno 2009 si è tenuta a Trento la prima convention italiana di OSM, denominata "OSMit 2009". Il 3 e 4 giugno 2010 è stata effettuata la seconda edizione di OSMit a Genova. La terza si è tenuta il 7 e 8 ottobre 2011 a Legnaro(PD), all'interno del festival delle libertà digitali. Si è deciso, da parte della comunità, di utilizzare il sito <http://conf.openstreetmap.it> come punto di riferimento per le edizioni a venire.

6.1.2 Dati ambientali

I dati ambientali sono alla base dell'informazione ambientale.

La Direttiva 2003/4/CE definisce come "informazione ambientale" qualsiasi informazione disponibile riguardante lo stato degli elementi dell'ambiente (aria, atmosfera, acqua, suolo, territorio, paesaggio), ma anche parametri di misura della qualità dell'ambiente (sostanze chimiche di origine non naturale, energia, rumore, emissioni inquinanti) che incidono, o possono incidere, sugli elementi dell'ambiente. Alcuni esempi di dati ambientali sono:

- le rilevazioni sulla qualità dell'aria, effettuate dalle centraline dislocate nel territorio, riguardanti concentrazioni di agenti inquinanti, come gli ossidi di azoto, l'ozono e le polveri sottili (es. PM10, PM2,5);
- la quantità di acque prelevate, immesse nella rete idrica e consumate, nell'anno solare;
- i consumi idrici civili totali, nell'anno solare, distinti in utenze domestiche e di servizio;
- la depurazione delle acque reflue, coi dettagli su eventuali fermi d'impianto, recanti i valori della domanda chimica di ossigeno (COD).

In materia di informazione ambientale, in Italia nel 2005 è stato emanato il D. Lgs. n. 195/2005, che recepisce la Direttiva 2003/4/CE, con l'obiettivo di "garantire il diritto d'accesso all'informazione ambientale detenuta dalle autorità pubbliche (...) e garantire, ai fini della più ampia trasparenza, che l'informazione ambientale sia sistematicamente e progressivamente messa a disposizione del pubblico e diffusa, anche attraverso i mezzi di telecomunicazione e gli strumenti informatici, in forme o formati facilmente consultabili, promuovendo a tale fine, in particolare, l'uso delle tecnologie dell'informazione e della comunicazione". Per raggiungere tale obiettivo, la norma individua anche alcuni strumenti:

- cataloghi pubblici dell'informazione ambientale contenenti l'elenco delle tipologie dell'informazione ambientale detenuta;

- banche dati elettroniche facilmente accessibili al pubblico tramite reti di telecomunicazione pubbliche, da aggiornare annualmente, in cui, oltre a relazioni, trattati, accordi, convenzioni, autorizzazioni e studi, siano inseriti tutti i dati o le sintesi di dati ricavati dal monitoraggio di attività che incidono o possono incidere sull'ambiente.

Da ciò si evince che l'apertura dei dati ambientali va nella direzione dell'applicazione delle norme indicate e, quindi, nella direzione di garantire il diritto all'accesso e la diffusione dell'informazione ambientale. Va comunque precisato che il valore dei dati ambientali è strettamente legato alla contestualizzazione degli stessi, poiché gli indicatori dello stato dell'ambiente sono realmente significativi solo quando vengono correlati a dati territoriali.

L'adozione di formati Linked Open Data nel settore dei dati ambientali è quindi fondamentale a tale scopo in quanto, oltre a contestualizzare gli stessi dati in una particolare prospettiva, ne arricchisce di fatto il loro significato. Sarebbe quindi possibile, ad esempio, porre in relazione le rilevazioni di fattori inquinanti con le suddivisioni territoriali dell'amministrazione (quartieri e contrade di un Comune), evidenziando le zone più bisognose di interventi migliorativi. Ugualmente interessanti sarebbero le relazioni di dati ambientali relativi alla qualità dell'aria con i dati territoriali rappresentanti il grafo stradale comunale, dato particolarmente utile all'amministrazione per adottare, ed eventualmente giustificare, determinate scelte sulla viabilità urbana (attivazione di ZTL, ordinanze di limitazione eccezionale del traffico urbano etc.). Infine, un altro processo notevole di contestualizzazione consisterebbe nel mettere in relazione gli indicatori di inquinamento, come ad esempio quelli dell'atmosfera, con i dati relativi alla dislocazione nel territorio di attività agricole, industriali, minerarie o di smaltimento dei rifiuti, evidenziando il reale impatto delle stesse sulle rilevazioni ottenute.

Indipendentemente dalle politiche che potranno essere adottate in futuro dalle singole PA per la raccolta dei dati, e che appare plausibile possano orientarsi sempre più verso l'adozione di formati standard atti a presentare dati di tipo LOD, la problematica principale da affrontare oggi è rappresentata sicuramente dalla difficoltà di convertire i dati ambientali già raccolti, presenti attualmente in una varietà di formati:

- talvolta aperti e "machine readable", ma non identificati da un URI (ad esempio file CSV);
- spesso "machine readable", ma strutturati in formati chiusi (ad esempio file Excel);
- spesso chiusi e non strutturati, come nel tipico esempio dei file PDF, sebbene siano ancora disponibili, nella maggioranza dei casi, i documenti in formati strutturati da cui sono stati originariamente esportati e memorizzati.

Per promuovere l'apertura dei dati ambientali e la realizzazione di nuove applicazioni in tale ambito, il 26 ottobre 2012 si è tenuta a Roma l'incontro intitolato "Open Data: Mobile Apps & Green Solutions"⁵⁹, nell'ambito del quale sono state presentate diverse iniziative di

⁵⁹ www.greenopendata.it

partecipazione civica attraverso l'utilizzo di Open Data di tipo ambientale (tra cui citiamo i progetti "Decoro Urbano"⁶⁰ e "eutocosacivedi.it – Progettazione Urbana Aumentata"⁶¹).

6.1.3 Dati della ricerca e competenze

I dati della ricerca comprendono sia le informazioni anagrafiche dei professori, dei ricercatori, degli assegnisti e dei borsisti, dei tecnici e degli amministrativi della ricerca pubblica, sia i dati relativi alle strutture (ad esempio, istituti, dipartimenti, laboratori) e ai prodotti della ricerca (pubblicazioni, brevetti, progetti, ecc.). Con riferimento a questi dati è anche possibile definire o estrarre le competenze dei ricercatori e strutture, l'impatto della ricerca, la sua distribuzione geografica e il collegamento con le richieste del mercato e delle istituzioni.

Gli utenti potenziali dei dati includono:

- imprese e istituti finanziari interessati a tecnologie, brevetti, soluzioni di ricerca;
- laboratori, ricercatori e studenti per la ricerca di competenze in un ambito di ricerca;
- decisori interessati all'analisi della ricerca e delle potenzialità di trasferimento tecnologico.

Più in generale, i dati possono essere usati anche da organizzazioni che producono o utilizzano dati eterogenei come enti, imprese e consulenti – soprattutto in ambito tecnico e scientifico – che vogliono far incontrare domanda e offerta di qualsiasi natura: "expert finding", "competence matching", centri interinali, oppure sviluppatori di prodotti e servizi per l'analisi organizzativa (ad es. cruscotti aziendali, conoscenza adattiva).

I vantaggi nell'apertura di tali tipi di dati sono molteplici: accessibilità e trasparenza dei dati della ricerca, sostegno alle attività di trasferimento tecnologico e in generale della conoscenza interna a un'organizzazione, arricchimento automatico dei dati già disponibili mediante analisi automatica del testo e inferenze automatiche, interoperabilità con dati già disponibili in Italia e nel resto del mondo.

6.1.4 Dati statistici e classificazione

I dati relativi a classificazioni ufficiali sono particolarmente importanti per supportare l'interoperabilità semantica di diversi sistemi, anche in un'ottica integrata e/o trasversale alle altre tipologie di dati precedentemente illustrate. Alcune classificazioni rilevanti sono l'Attività Economica (Ateco 2007), le Professioni (CP2011), i Titoli di Studio (Istat-MIUR), Unità istituzionali del settore pubbliche amministrazioni, i codici di comuni, province e regioni.

⁶⁰ www.decorourbano.org

⁶¹ www.etucosacivedi.it

Oltre alle classificazioni ufficiali, una tipologia di dati per la quale i requisiti di accessibilità e fruibilità sono particolarmente importanti è costituita dai dati pubblicati nell'ambito della statistica ufficiale, ovvero dati aggregati che costituiscono il prodotto finale delle rilevazioni e delle elaborazioni condotte nell'ambito del Programma statistico nazionale. La fornitura di questi dati in formato aperto ha diversi vantaggi: tra questi, ad esempio, la possibilità di integrare fonti statistiche distinte (ad esempio a livello europeo), la possibilità di confrontare in modo diretto tali dati con elaborazioni condotte su microdati disponibili presso altri soggetti, ecc. Questa tipologia è prodotta da Istat ed è diffusa a vari livelli di aggregazione e principalmente in modalità CSV, SDMX e JSON-stat.

6.2 Modelli di business emergenti con gli Open Data

Come abbiamo visto attraverso la disamina sin qui condotta, i dati aperti possono senza dubbio alimentare una miriade di attività destinate a creare valore. Il paradigma degli Open Data, infatti, genera un ecosistema in cui i vari attori, per loro convenienze specifiche, possono condividere in modalità aperta parte dei loro dati e riutilizzare dati di altri attori, attraverso un modello in cui l'informazione rappresenta un elemento di valore condiviso su cui sviluppare servizi a valore aggiunto per cittadini e imprese.

Per lo sviluppo dell'ecosistema degli Open Data è necessario, però, che il flusso informativo che lo alimenta (proveniente da PA, comunità e aziende) sia continuo, e che le informazioni e i dati pubblicati siano costantemente aggiornati.

Anche se una componente importante della domanda di dati aperti nei confronti della PA arriva dalle comunità di cittadini (di cui fanno parte, per esempio, tecnici assai preparati), ciò che può effettivamente fare la differenza sono i servizi a valore aggiunto sviluppati dalle imprese sulla base delle informazioni e dei dati messi a disposizione in modalità aperta.

Diventa estremamente importante, dunque, analizzare la domanda delle imprese che, in prima battuta, può essere caratterizzata in funzione della dimensione geografica dei dati richiesti e delle opportunità che le imprese possono cogliere da essi.

I geodati possono avere carattere locale, regionale, sovranazionale, nazionale o sovranazionale. Le opportunità che l'impresa può cogliere vertono su:

- sviluppo del core business (ad esempio, sviluppo di servizi a valore aggiunto basati sui dati aperti, identificazione di mercati potenziali, posizionamento di punti di vendita o di assistenza, quantificazione e distribuzione di potenziali clienti profilati);
- promozione dell'innovazione di prodotto o di processo (ad esempio, programmi di ricerca e sviluppo, *open innovation*);

- soluzione di problematiche legate alla logistica (ad esempio, localizzazione magazzini o centri distributivi, ottimizzazione dei mezzi di trasporto e delle tratte, sicurezza dei trasporti) o alla gestione della supply chain (ad esempio, identificazione di possibili fornitori alternativi di beni commodity).

Nell'ottica della comprensione delle modalità attraverso cui la creazione e l'appropriazione di valore possono aver luogo, è necessario effettuare una riflessione sulle differenti logiche di business che gli attori possono adottare. Ecco una panoramica dei modelli di business delineabili attraverso i dati aperti⁶²:

Prodotti/servizi “premium”. In questo modello di business il riusatore “canonico” offre al mercato finale un prodotto o un servizio, presumibilmente caratterizzato da elevato valore intrinseco, in cambio di un pagamento effettuato in contropartita al bene scambiato. Le modalità di pagamento generalmente sono a *la carte* o previa sottoscrizione: mentre la prima, assimilabile a un *pay-per-use*, prevede la corresponsione di un importo per ciascuna unità di prodotto acquistata, la seconda contempla una tariffa “tutto compreso” che, per un arco temporale prestabilito, consente l'utilizzo di determinate funzionalità in conformità alle pattuizioni contrattuali. Alla luce del meccanismo di *pricing* e dell'elevato valore intrinseco, la clientela target è per lo più di tipo B2B e viene gestita mediante logiche di medio-lungo termine che vanno al di là del mero approccio transattivo.

Prodotti/servizi “freemium”. In questo modello di business il riusatore “canonico” propone al mercato un prodotto o un servizio attraverso un'offerta apparentemente conveniente agli occhi del cliente finale. Il meccanismo di fissazione del prezzo, infatti, si articola secondo l'approccio *freemium*, il quale prevede la fruizione gratuita delle funzionalità di base mentre, per funzionalità più avanzate, è richiesta la corresponsione di una tariffa. Nel settore pubblico, questo paradigma si declina attraverso limitazioni applicate ai prodotti e servizi offerti: pagamenti *ad hoc* possono essere infatti necessari per accedere a funzionalità supplementari o a set di dati aggiuntivi. In questo caso, la clientela appartiene tendenzialmente al segmento B2C; la gestione del cliente segue logiche di breve-medio termine e non prevede, se non in rari casi, una vera personalizzazione del servizio. L'erogazione della proposta di valore si avvale generalmente dei canali web e mobile (attraverso App), dalla cui sinergia è possibile raggiungere un numero cospicuo di basi installate.

Simil Open Source. Il modello di business “open” prevede l'offerta di prodotti, servizi o set di dati non “confezionati”, senza il pagamento di alcuna tariffa. Le voci di ricavo per il riusatore “canonico” che adotta questa logica di business possono essere legate a pagamenti per servizi a valore aggiunto o per modifiche alla licenza originaria. Si attua così una “sovvenzione incrociata”, in cui le entrate da ascrivere a queste linee di business supplementari, comunque legate al settore pubblico, vengono impiegate per coprire i costi ingenerati

⁶² Cfr. Osservatorio ICT Regione Piemonte, “*Modelli di Business nel Riuso dell'Informazione Pubblica*”.

dalle linee di business offerte a titolo gratuito: in altre parole, agli occhi del cliente finale il dato è offerto gratuitamente in formati aperti e rielaborabili senza barriere di natura tecnica (da qui il riferimento al mondo dell'Open Source). La flessibilità che contraddistingue il modello di business rende questa logica adatta a una variegata platea di clienti gestiti attraverso relazioni divolte in volta definite a seconda dell'interlocutore.

“Rasoio e lamette”. Gli attori abilitatori fattori di questa logica agiscono come intermediari che facilitano l'accesso a risorse del settore pubblico da parte di sviluppatori orientati al profitto o di scienziati/tecnici non guidati da intenti commerciali. Il nome conferito a questo modello deriva dalla strategia comunemente definita come *“razor&blades”*: inizialmente un prodotto viene venduto ad un prezzo molto basso al fine di alimentare il successivo acquisto ripetuto di un bene complementare (generalmente un consumabile), avente una domanda piuttosto inelastica, sul quale è invece possibile realizzare marginalità significative. Nel caso del settore pubblico, il predetto modello viene implementato posizionando set di dati pubblici su piattaforme di *cloud computing* e rendendoli accessibili via API da chiunque, tariffando poi il solo utilizzo della capacità computazionale richiesta per il processamento dei dati stessi. Anche in questo caso è ravvisabile una sovvenzione incrociata, poiché i profitti realizzati dalla messa a disposizione di capacità elaborativa *“on-demand”* vanno a coprire i costi ascrivibili allo *storage* e all'organizzazione del dato.

Piattaforme orientate alla domanda. Questo modello ambisce a fornire agli sviluppatori consistenti facilitazioni nell'accesso a risorse del settore pubblico, le quali vengono memorizzate su server intermedi proprietari a elevata affidabilità, catalogate mediante metadati, armonizzate in termini di formati ed esposte attraverso API, rendendo agevole il reperimento dinamico dei dati in maniera *“on-demand”*. Una cospicua gamma di criticità intrinseche al dato grezzo vengono quindi rese ininfluenti grazie al ricorso a piattaforme di accesso ai dati capaci di tradurre dataset in *“datastream”*, contribuendo sensibilmente alla *“commoditization”* e alla *“democratizzazione”* del dato. Tali strutture adibite all'intermediazione offrono così agli sviluppatori un ampio ventaglio di dati secondo un approccio *“onestop shopping”* teso a minimizzare il costo di ricerca da parte del compratore: quest'ultimo, rivolgendosi a un solo fornitore, accede attraverso API standardizzate a una molteplicità di risorse informative, senza doversi preoccupare dell'interfaccia di accesso a ciascuna delle fonti originali. In termini di costo, poiché un bene che nasce libero e aperto non può essere commercializzato a fini di lucro in assenza di elaborazioni su di esso, gli abilitatori che impiegano questo modello di business sull'informazione del settore pubblico ricevono flussi finanziari in entrata in cambio di servizi avanzati o di set di dati raffinati, configurando un modello di costo orientato al paradigma *“freemium”* in cui l'importo viene modulato in base a limitazioni funzionali.

Piattaforme orientate all'offerta. Questo modello prevede nuovamente la presenza di un attore intermediario fornitore di servizi infrastrutturali; l'impresa qualificata come abilitatore,

però, in questocaso non applica alcuna tariffa agli sviluppatori, ricaricando invece le Pubbliche Amministrazioni detentrici di informazione. Osservando lo scenario nel suo complesso, si nota la presenza di un mercato a due versanti, costituito dagli “holder” dell’informazione del settore pubblico da una lato e dagli sviluppatori dall’altro lato. Seguendo leregole proprie dei mercati multi-versante, l’abilitatore modula il prezzo in funzione del grado di esternalità indiretta positiva che ciascun versante è in grado di esercitare: vengono così azzerate le barriere per gli sviluppatori, i quali accedono gratuitamente a dati ben strutturati, e vengono applicate delle tariffe agli enti pubblici che diventano conseguentemente titolari delle piattaforme di gestione dei dati. A livello tecnico, infatti, vengono predisposte soluzioni tese a facilitare l’esposizione di dati dell’informazione del settore pubblico che siano capaci di adattarsi alle esigenze degli enti pubblici desiderosi di avviare programmi di apertura dei dati. Generalmente tali piattaforme garantiscono lo “storage”, il rapido aggiornamento di nuovi dataset da parte del personale pubblico, la catalogazione mediante metadati, la standardizzazione dei formati e l’esposizione verso l’esterno dei dati sia attraverso interfacce applicative (API) sia mediante interfacce grafiche (GUI). Le Pubbliche Amministrazioni che aderiscono ai programmi avviano quindi una relazione di lungo periodo con i provider e sono tenute alla corresponsione periodica di tariffe in funzione della sofisticazione delle soluzioni acquistate e di alcuni parametri tecnici.

Promozione del marchio. I promotori con interessi indiretti offrono agli utilizzatori finali prodotti o servizi basati sulla informazione del settore pubblico senza realizzare diretto ricavo a partire da essi; l’attività promozionale condotta da questi attori economici si candida a generare ricadute positive in senso lato, andando potenzialmente a favorire i risultati economici registrabili su linee di business del tutto estranee all’informazione del settore pubblico. Gli “economics” sottesi a questa pratica sono connotati da una forte spinta pubblicitaria, che porta l’impresa a considerare i costi non coperti da relativi ricavi alla stregua di investimenti promozionali compresi nel “marketing mix”, o dalla presenza di costi marginali nulli, situazione che si verifica allorché i costi di distribuzione e di utilizzo siano non significativi, rendendo l’attività non coperta da ricavi non gravosa a livello di centro di costo. Le linee di business incentrate sull’informazione del settore pubblico, dal canto loro, diventano quindi strumenti per porre in essere il cosiddetto “service advertising”, modello promozionale emergente che sostituisce l’esposizione di comunicazioni visuali atte a influenzare in maniera intenzionale e sistematica gli atteggiamenti e le scelte degli individui in relazione al consumo di beni e all’utilizzo di servizi, con la fornitura di servizi di pubblica utilità che, in maniera più indiretta ma anche più incisiva, tentano di evidenziare una condotta “illuminata” dell’azienda promotrice. Il pubblico target di riferimento è ampio e può essere raggiunto sia attraverso il “tradizionale” canale web sia mediante il canale mobile. Un’efficace implementazione del modello di promozione del marchio si incentra sul binomio costituito da competenze tecniche e competenze creative: quando esse non albergano entrambe sotto lo stesso tetto, può diventare indispensabile esternalizzare in toto o parzialmente la produzione dei servizi promozionali che assurgono a “strumenti di attrazione”.

Sviluppo per terze parti. Le officine pubblicitarie entrano in gioco allorché i promotori con interessi indiretti decidono di esternalizzare lo sviluppo dei servizi di utilità collettiva forniti con finalità promozionali. Lo sviluppo per conto di terze parti si rivolge quindi ad aziende desiderose di promuoversi attraverso l'informazione del settore pubblico, con le quali le officine pubblicitarie instaurano relazioni di medio-lungo periodo orientate alla personalizzazione. Alla luce del ruolo vitale ricoperto dal marchio dei promotori con interessi indiretti, le officine pubblicitarie adottano la strategia "whitelabel", in virtù della quale viene eclissato il brand dell'intermediario per conferire piena visibilità al marchio dell'attore economico che affronta il mercato finale. Il committente, in cambio della soluzione chiavi in mano così sviluppata, corrisponde pagamenti fissi o periodici, a seconda che il bene confezionato si configuri come prodotto o servizio.

GLOSSARIO

Anonymisation

Il processo di adattare i dati in modo che gli individui non possano essere individuati da esso.

API – Application Programming Interface

Un modo per i programmi di comunicare tra loro, ossia un set di istruzioni standard attraverso le quali si può interagire con un software in modo da gestirne funzioni o elaborare dati.

Attribution License

Una licenza che richieda che venga citata la fonte originale del materiale distribuito (attribuzione).

BitTorrent

BitTorrent è un protocollo per la distribuzione di file di grandi dimensioni attraverso computer che contribuiscono a loro volta al trasferimento. Invece di scaricare un file da una fonte singola, BitTorrent permette di scaricare anche attraverso altre fonti in parallelo.

Cloud Computing

Un insieme di tecnologie che permettono, tipicamente sotto forma di un servizio offerto da un provider al cliente, di memorizzare/archiviare e/o elaborare dati (tramite CPU o software) grazie all'utilizzo di risorse hardware/software distribuite e virtualizzate in Rete.

Copyright

Diritto per i realizzatori di opere creative a circoscrivere l'utilizzo di quei lavori da parte di altri. Chi possiede il copyright ha titolo per decidere come altri possono utilizzare quel lavoro.

DAP – Data Access Protocol

Un sistema che permette a soggetti esterni di accedere a set di dati senza sovraccaricare il proprio sistema o quello dove i dati sono ospitati.

Data Protection Legislation

La legislazione per la protezione dei dati non riguarda il proteggere i dati, ma il proteggere il diritto dei cittadini a vivere senza il timore che informazioni sulle loro vite private possano diventare pubbliche. La legge protegge la privacy (si pensi alle informazioni sulla posizione economica, la salute e le posizioni politiche di una persona) ed altri diritti quali il diritto alla libertà di movimento e di riunione. Per esempio, in Finlandia un sistema di abbonamenti veniva utilizzato per registrare tutti i casi in cui l'abbonamento veniva mostrato al lettore automatico sulle differenti linee di trasporto pubblico. Ciò sollevò un dibattito dal punto di vista della libertà

di movimento e la raccolta dei dati degli abbonamenti fu abbandonata sulla base della legislazione per la protezione dei dati.

Database Rights

Diritto di impedire ad altri di estrarre e riutilizzare la totalità o una parte sostanziale dei contenuti di un database. Esiste soprattutto nelle giurisdizioni europee.

Dataset

Una collezione di dati, generalmente riguardanti una stessa organizzazione, che vengono erogati e gestiti congiuntamente; insieme di dati strutturati in forma relazionale.

Data Scraping

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

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.

EU PSI Directive

La direttiva sul riutilizzo dell'informazione del settore pubblico (PSI – Public Sector Information), 2003/98/CE "si occupa di come gli enti del settore pubblico massimizzino il riutilizzo delle proprie risorse informative".

Information Asset Register – IAR

Gli IAR sono registri appositamente istituiti per immagazzinare e organizzare metadati sull'enorme quantità di informazioni detenute da amministrazioni ed enti pubblici. Uno IAR completo include set di dati, vecchi insiemi di file, file elettronici recenti, raccolte di statistiche, ricerche e così via.

Intellectual Property Rights

Diritti esclusivi attribuiti a soggetti su opere creative a carattere intellettuale.

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.

Machine-readable

Le informazioni che sono leggibili a macchina - machine readable - sono informazioni che possono essere estratte da programmi informatici. I documenti PDF non sono machine readable. I computer sono in grado di visualizzare il testo, ma hanno grande difficoltà a comprendere il contesto che circonda il testo.

Mashup

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

Open Data

I dati aperti (Open Data) possono essere utilizzati per qualsiasi scopo. Maggiori informazioni sono reperibili presso opendefinition.org.

Open Government Data

Open data prodotti dal governo. È comunemente accettato riferirsi ai dati raccolti nel corso dell'attività amministrativa, che non identificano soggetti specifici e che non violano la "commercial sensitivity". Il concetto di "dati governativi aperti" è un sottoinsieme della nozione di Public Sector Information, che ha una portata più ampia. Cfr. <http://opengovernmentdata.org> per ulteriori dettagli.

Open Standards

Generalmente intesi come norme tecniche svincolati da restrizioni di licenza. Possono anche essere interpretati nel senso che gli standard sono sviluppati in modo da essere neutrali rispetto al produttore.

PSI – Public Sector Information

Nella direttiva Europea 2003/98/EC è definita come l'informazione della Pubblica Amministrazione. L'informazione pubblica si caratterizza per essere di tipo statico o di tipo dinamico. L'informazione statica è rappresentata dal contenuto informativo in possesso della Pubblica Amministrazione (ad esempio gli archivi dei beni culturali); l'informazione dinamica è invece prodotta dalle istituzioni pubbliche nello svolgimento dei propri compiti istituzionali (ad esempio i dati di bilancio di un ente).

Riuso

Utilizzo di contenuti al di fuori della loro finalità originaria.

Share-alike License

Una licenza che richiede che un'opera derivata sia protetta alle stesse condizioni o termini simile a quelle dell'opera originale.

SQL (Structured Query Language)

SQL è un linguaggio standardizzato per database basati sul modello relazionale (RDBMS) progettato per creare e modificare schemi di database (DDL - Data Definition Language), inserire, modificare e gestire dati memorizzati (DML - Data Manipulation Language), interrogare i dati memorizzati (DQL - Data Query Language), creare e gestire strumenti di controllo ed accesso ai dati (DCL - Data Control Language).

SPARQL (Sparql Protocol And RDF Query Language)

È un linguaggio con una sintassi simile a quella SQL per l'interrogazione di dati RDF e un protocollo di comunicazione basato su HTTP.

SPC – Sistema Pubblico di Connettività e Cooperazione

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 82/2005 (artt. 72-87).

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.

Web API

Una API che è stata progettata per funzionare su Internet.

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.